Interpretation and consequences of copy number variants

A thesis submitted to The University of Manchester for the degree of

Doctor of Clinical Sciences

in the Faculty of Biology, Medicine and Health

2021

George J. Burghel

School of Health Sciences

TABLE OF CONTENTS

Word count: ~27,000 words excluding references, table of contents and appendices.

Table of Figures

Table of Tables

List of Abbreviations

aCGH	array comparative genome hybridisation
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal Dominant
ADHD	Attention Deficit Hyperactivity Disorder
AR	Autosomal Recessive
ASD	autistic spectrum disorder
BRWS1	Baraitser-Winter syndrome-1
CDCBM3	Cortical Dysplasia, Complex, with other Brain Malformations 3
ClinGen	the Clinical Genome Resource
CNG	Copy Number variant Gain
CNL	Copy Number variant Loss
CNV	Copy Number Variant
dbVar	Database of structural Variants
DClinSci	Doctorate in Clinical Sciences
DD	Developmental Delay
DDG2P	Development Disorder Genotype - Phenotype Database
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
DGV	Database of Genomic Variants
D-HI	Decipher Haploinsufficiency
DNA	Deoxyribonucleic Acid
DR	Digenic Recessive
FISH	Fluorescence *In Situ* Hybridization
FRCPath	Fellowship of the Royal College of Pathology
GDPR	General Data Protection Regulation
GoF	Gain of Function
GTEX	The Genotype-Tissue Expression
HGMD	Human Gene Mutation Database
HGNC	HUGO gene nomenclature committee
HNPP	Hereditary Neuropathy with liability to Pressure Palsies
HSST	Higher Specialist Scientific Trainee
ID	Intellectual Disability
IDDBCS	Intellectual Developmental Disorder with Behavioral abnormalities and Craniofacial dysmorphism with or without Seizures
IMDR	Index of Multiple Deprivation Rank
ISCA	International Standards for Cytogenomic Arrays
KLEFS2	Kleefstra Syndrome-2
LoF	Loss of Function
LOEUF	Loss-of-function Observed/Expected Upper bound Fraction
MACID	Macrocephaly with impaired intellectual development
MCGM	Manchester Centre for Genomic Medicine
MRD-30	Mental Retardation 30
MRD-39	Mental Retardation 39
MRD-50	Mental Retardation 50
NDHSAL	Neurodevelopmental Disorder with Hypotonia, Seizures, and Absent Language
NGS	Next Generation Sequencing
NW-GLH	North West Genomic Laboratory Hub
OBHD	Obesity Hyperphagia, and Developmental delay syndrome
OCNDS	Okur-Chung neurodevelopmental syndrome
OGT	Oxford Gene Technology
OMIM	Online Mendelian Inheritance In Man
pLI	Probability of being Loss-of-function Intolerant
RVIS	Residual Variation Intolerance Score
SES	Socio Economic Status
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
STISS	Stankiewicz-Isidor Syndrome
TAD	Topologically Associating Domain
VUS	Variants of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Abstract

Background: Copy Number Variants (CNVs) are an important cause for human diseases. The overall aim of this DClinSci research project was to improve the classification and understanding of the impact of CNVs. **Objectives**: (1) Current approaches to clinical interpretation leave a significant proportion of CNVs with uncertain clinical significance (class 3). Periodic re-analysis of all Class 3 CNVs in a diagnostic laboratory is not feasible. We decided to devise a strategy to facilitate large-scale re-analysis for a sub-group of class 3 CNVs. (2) Socioeconomic status (SES) is a major determinant of health and related outcomes. Biological factors including genetic variants that may influence SES are only beginning to be understood. We decided to study the correlation between pathogenic CNVs and SES. **Methods and Results:** We curated a large database of CNVs identified at the Manchester Centre of Genomic Medicine over a period of ~7 years via array comparative genomic hybridisation performed in ~16,000 patients. We harmonised clinical class terminology as per current guidance and identified ~7,000 Class 3 CNVs. From these CNVs we chose to systematically re-analyse 2,173 class 3 losses (CNLs). Then using a gene-focussed approach, with up-to-date disease association information and haploinsufficiency scores we generated a shortlist of 204 class 3 CNLs that encompassed gene(s) with autosomal dominant disease association predicted to be haploinsufficient or have a loss of function disease mechanism. Clinical scientist manual review of these shortlisted CNLs led to reclassification of 13 class 3 CNLs (~6.4% of the shortlisted cohort) as (likely)pathogenic. We then performed a detailed *in silico* analyses of a single case with class 3 16p13.3 CNL and showed that haploinsufficiency of *ATP6V0C* probably underlies the pathology of this condition. We studied the Index of Multiple Deprivation Rank (IMDR) of 473 individuals with (likely)pathogenic autosomal CNVs and known inheritance status. The IMDR distribution of families with (likely)pathogenic CNVs was significantly different from the general population. Families with inherited CNVs were significantly more likely to be living in areas of higher deprivation when compared with families that had individuals with *de novo* CNVs. **Conclusions:** We present an efficient re-analysis strategy for Class 3 CNLs for diagnostic laboratories. Although the value of re-analysis of next generation sequencing data is well established, this is the first such study of CNV re-analysis. Our study on correlations between pathogenic CNVs and SES provide unique insights into biological determinants of SES. As pathogenic CNVs are relatively frequent in the general population, these results have important medical and policy consequences.

DECLERATION

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

COPYRIGHT STATEMENT

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

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

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

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

Acknowledgements

I would like to start by giving my deepest appreciation to my supervisor, Dr Siddharth Banka, for his guidance, assistance, advice, patience, encouragement and help throughout the last five years and especially for the work presented in this thesis.

I am very thankful to many colleagues and friends who supported this work in various ways. I would especially like to thank Dr Jamie Ellingford for his amazing bioinformatics skills, Dr Wei-Yu Lin for his statistical magnificence and for Ronnie Wright, Jonathan Edgerley, Chris Watt, Jake Miller, Algy Taylor, Sash Palaniswamy and Unzela Khan for their help with data analysis. I would also like to thank all the clinical colleagues who saw and referred patients for aCGH testing and all the technical and scientific colleagues who ran tests, analysed and interpreted CNV data over the years. Special thanks go to Dr Andrew Wallace, Dr Lorraine Gaunt and Dr David Gokhale for their invaluable guidance throughout my HSST. I am also very grateful to Dr Helene Schlecht and Marta Pereira from our cancer team who provided great support during the writing of this thesis. I am also thankful to the National School of Healthcare Science (NSHCS), Manchester Academy for Healthcare Scientist Education (MAHSE) and the University of Manchester for this amazing opportunity.

To my beloved family, my brother Aboud and his family, Hanan, Joseph and George, and my brother Rami, and my parents-in-law, Keith (OBE) and Sue, I can never thank you enough for your amazing love and support. You are all brilliant.

My beautiful wife Heather, thank you for your never-ending support and encouragement, I can't do what I am doing without you. My children, Diana and Elias ("the Twins") and my baby Joseph who were all born during my HSST, you are the joy of my life and my absolute inspiration.

Finally, I would like to dedicate this work to my parents, Baba Joseph and Mama Diana, who are sadly not with us today but are somewhere better. Throughout my life, your support and love has provided me with the strength and the drive to keep on going. I miss you so much.

The author

George Burghel, BSc in Biotechnology and Genetic Engineering, MSc in Molecular Medicine (Genetics), MSc in Clinical Sciences (Genetics), PhD in Cancer Genetics, is a Principal Clinical Scientist in Genomics and the deputy lead for the cancer genomics programme at the North West Genomic Laboratory Hub. The author's main research interest, apart from the work presented here, is focussed around the application of ctDNA in different cancer types including ovarian cancer and melanoma, in addition to, understanding the implication of genetic variation in cancer genes. Th author also has special interest in patient and public engagement. His work was recognised through two recent awards; 1) St Mary's Hospital Rising Start award 2019 and 2) Manchester University NHS foundation trust excellence award 2020.

List of published/(in process of publication) work during the period of HSST 2015-2020

1. Morgan, R.D., Burghel, G.J., Flaum, N., Bulman, M., Smith, P., Clamp, A.R., Hasan, J., Mitchell, C.L., Woodward, E., Lalloo, F., Crosbie E.J., Edmondson R.J., Jayson, G.C., Wallace, A.J., Evans, G.R., "Mainstream somatic BRCA1/2 testing in high-grade serous ovarian cancer: an observational study". under review, **BJOG, 2021.**
2. Batiha O*, **Burghel GJ***, Alkofahi A, Alsharu E, Smith H, Alobaidi B, Al-Smadi M, Awamlah N, Hussein L, Abdelnour A, Sheth H, Veltman JA, "Molecular characterization of a cohort of azoospermic Jordanian males". In press. **Hum Fertil, 2021. *Equal Contribution**
3. Molina-Ramírez L, Kyle C, Ellingford JM, Wright R, Taylor A, Campbell C, Jackson H, Fairclough A, Rousseau A, **Burghel GJ**, Dutton L, Banka S, Briggs T, Clayton-Smith J, Douzgou S, Jones EA, Kongston HM, Kerr B, Ealing J, Somarathi S, Chandler K, Stuart HM, Burkitt-Wright E, Newman WG, Bruce IA, Black G, Gokhale D., "Personalised virtual gene panels reduce interpretation workload and maintain diagnostic rates of proband-only clinical exome sequencing for rare disorders". In press. **J Med Genet, 2021.**
4. Evans DG, van Veen EM, Byers HJ, Evans SJ, **Burghel GJ**, Woodward ER, Harkness EF, Eccles DM, Greville-Heygate S, Ellingford JM, Bowers NL, Pereira M, Wallace AJ, Howell SJ, Howell A, Lalloo F, Newman WG, Smith MJ, "High likelihood of actionable pathogenic variant detection in breast cancer genes in women with very early onset breast cancer". **J Med Genet, 20021.**
5. Barriuso J, Nagaraju RT, Belgamwar S, Chakrabarty B, **Burghel GJ**, Schlecht H, Foster L, Kilgour E, Wallace AJ, Braun M, Dive C, Evans DG, Bristow RG, Saunders MP, O'Dwyer ST, Aziz O. "Early adaptation of colorectal cancer cells to the peritoneal cavity is associated with activation of 'stemness' programs and local inflammation". **Clin Cancer Res, 2021.**
6. Garrett A, Durkie M, Callaway A, **Burghel GJ**, Robinson R, Drummond J, Torr B, Cubuk C, Berry IR, Wallace AJ, Ellard S, Eccles DM, Tischkowitz M, Hanson H, Turnbull C; CanVIG-UK, "Combining evidence for and against pathogenicity for variants in cancer susceptibility genes: CanVIG-UK consensus recommendations". **J Med Genet, 2020.**
7. Tinker RJ, **Burghel GJ**, Garg S, Steggall M, Cuvertino S, Banka S, "Haploinsufficiency of ATP6V0C possibly underlies 16p13.3 deletions that cause microcephaly, seizures, and neurodevelopmental disorder". **Am J Med Genet A, 2020.**
8. Terradas M, Mur P, Belhadj S, Woodward E.R, **Burghel GJ**, Munoz-Torres PM, Quintana I et al. "TP53, a gene for colorectal cancer predisposition in the absence of Li-Fraumeni-associated phenotypes." ***Gut*, 2020.**
9. Flaum, N., Morgan, R.D., **Burghel, G.J**., Bulman, M., Clamp, A.R., Hasan, J., Mitchell, C.L., Badea, D., Moon, S., Hogg, M. and Hadjiyiannakis, D., "Mainstreaming germline BRCA1/2 testing in non-mucinous epithelial ovarian cancer in the North West of England". ***Eur J Hum Genet*, 2020.**
10. Garrett, A., Callaway, A., Durkie, M., Cubuk, C., Alikian, M., **Burghel, G.J.,** Robinson, R., Izatt, L., Talukdar, S., Side, L. and Cranston, T., "Cancer Variant Interpretation Group UK (CanVIG-UK): an exemplar national subspecialty multidisciplinary network". ***J Med Genet*, 2020.**

11. Nelson L, Tighe A, Golder A, Littler S, Bakker B, Moralli D, Murtuza Baker S, Donaldson IJ, Spierings DCJ, Wardenaar R, Neale B, **Burghel GJ**, Winter-Roach B, Edmondson R, Clamp AR, Jayson GC, Desai S, Green CM, Hayes A, Foijer F, Morgan RD, Taylor SS., "A living biobank of ovarian cancer ex vivo models reveals profound mitotic heterogeneity". ***Nat Commun*, 2020.**
12. **Burghel, G.J.,** Khan, U., Lin, W.Y., Whittaker, W., Banka, S., "Presence of pathogenic copy number variants is correlated with socioeconomic status. ***J Med Genet,* 2020**.
13. Forde C, Lim DK, Alwan Y, **Burghel G,** Butland L, Cleaver R, Dixit A, Vialard L, Hanson H, Lalloo F, Oliveira P, Wallis Y Maher ER, Woodward ER. "Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC): Clinical, molecular features and surveillance features in a cohort of 185 affected individuals". ***Eur Urol Oncol*, 2019.**
14. **Burghel, G.J.,** Quinn A.M., Smith P, Harris J, Banks C, Wallace A. "The importance of neoplastic cell content assessment and enrichment by macrodissection in cancer pharmacogenetic testing". **J *Clin Pathol*, 2019.**
15. Morgan R.D., **Burghel G.J.,** Flaum N, Crosbie E.J., Edmondson R.J., Woodward E.R., Lalloo F.I., Wallace A.J., Jayson G.C., Evans D.G.R. "Prevalence of germline pathogenic *BRCA1/2* variants in sequential epithelial ovarian cancer cases". ***J Med Genet,* 2019.**
16. Smith MJ, Woodward ER, **Burghel GJ,** Banks C, Morgan RD, Wallace AJ, Turnbull C, Evans DG. "Rapid reversal of clinical down-classification of a BRCA1 splicing variant avoiding psychological harm". ***Clin Genet,* 2019.**
17. **Burghel, G.J.,** Abu-Dayyeh, I., Babouq, N., Wallace, A., Abdelnour, A., "Mutational screen of a panel of tumour genes in a case report of mucoepidermoid carcinoma of the breast from Jordan". ***Breast J,* 2018.**
18. Silva, S., Danson, S., Teare, D., Taylor, F., Bradford, J., McDonagh, A.J., Salawu, A., Wells, G., **Burghel, G.J.,** Brock, I. and Connley, D., 2018. Genome-Wide Analysis of Circulating Cell-Free DNA Copy Number Detects Active Melanoma and Predicts Survival. ***Clin Chem*, 2018**.
19. Batiha, O., Haifawi, S., Al-Smadi, M., **Burghel, G.J.,** Naber, Z., Elbetieha, A.M., Bodoor, K., Al Sumadi, A., Swaidat, S., Jarun, Y. and Abdelnour, A., "Molecular analysis of CAG repeat length of the androgen receptor gene and Y chromosome microdeletions among Jordanian azoospermic infertile males". ***Andrologia*, 2018.**
20. Hurst, C.D., Alder, O., Platt, F.M., Droop, A., Stead, L.F., Burns, J.E., **Burghel, G.J.,** Jain, S., Klimczak, L.J., Lindsay, H. and Roulson, J.A., "Genomic subtypes of non-invasive bladder cancer with distinct metabolic profile and female gender bias in KDM6A mutation frequency". ***Cancer Cell*, 2017**
21. Innes, J., Reali, L., Clayton-Smith, J., Hall, G., Lim, D.H., **Burghel, G.J.,** French, K., Khan, U., Walker, D., Lalloo, F. and Evans, D.G.R., 2017. "CNVs affecting cancer predisposing genes (CPGs) detected as incidental findings in routine germline diagnostic chromosomal microarray (CMA) testing". ***J Med Genet*, 2017.**
22. Daniels, S.L., **Burghel, G.J.,** Chambers, P., Al-Baba, S., Connley, D.D., Brock, I.W., Cramp, H.E., Dotsenko, O., Wilks, O., Wyld, L. and Cross, S.S., "Levels of DNA methylation vary at CpG sites across the BRCA1 promoter, and differ according to triple negative and "BRCA-like" status, in both blood and tumour DNA". ***PloS one*, 2016.**
23. Rushton, S., **Burghel, G.,** Wallace, A. and Nonaka, D., "Immunohistochemical detection of BRAF V600E mutation status in anaplastic thyroid carcinoma". ***Histopathology*, 2016**.
24. Camp, N.J., Lin, W.Y., Bigelow, A., **Burghel, G.J.,** Mosbruger, T.L., Parry, M.A., Waller, R.G., Rigas, S.H., Tai, P.Y., Berrett, K. and Rajamanickam, V., "Discordant haplotype sequencing identifies functional variants at the 2q33 breast cancer risk locus". ***Cancer Res*, 2016**
25. **Burghel, G.J.,** Hurst, C.D., Watson, C.M., Chambers, P.A., Dickinson, H., Roberts, P. and Knowles, M.A., 2015. Towards a next-generation sequencing diagnostic service for tumour genotyping: a comparison of panels and platforms. ***BioMed Res Int*, 2015**

1 INTRODUCTION

1.1 DNA variation in the human genome

Human genetic and genomic variations reflect differences in the DNA sequence between different individuals (Frazer et al., 2009; Ku et al., 2010). There are various types of variants in the human genome, including single nucleotide variants (SNVs), polymorphic tandem repeats, insertions and deletions (indels) of a few nucleotides, larger deletions or gains that change the copy number of a segment of DNA (copy number variations (CNVs)), in addition to copy-neutral structural chromosomal rearrangements such as inversions and translocations and aneuploidy involving the gain and loss of whole chromosomes (Frazer et al., 2009; Ku et al., 2010; Strachan and Read, 2011).

These different types of genetic variants can range in size from single nucleotides to thousands and millions of bases and can occur within coding regions, non-coding regions or across both (Ku et al., 2010). SNVs are the most common of all and as the name indicates they involve a change in a single nucleotide at a specific locus. On the other hand, CNVs and structural variants (inversions and translocations) affect larger segments of DNA ranging from 50 nucleotides to millions of bases. In between these extremes of sizes, tandem repeats and indels affect a range from few nucleotides (more than one) to a thousand or so (Ku et al., 2010; Strachan and Read, 2011). Genetic variants occur through different mechanisms, but generally they are a result of errors and mistakes of DNA replication and recombination that occurred during cell division (Ku et al., 2010; Strachan and Read, 2011). Genetic variants can also be classified as germline or somatic variants (Ku et al., 2010).

Genetic variants contribute to human diversity and play a central role in making individuals unique. They also have important implications in health and disease (Ku et al., 2010; Zarrei et al., 2015). The rest of this introduction will focus on CNVs, their role in human disease and methods to identify them and determine their clinical significance in diagnostic genetic and genomic laboratories.

1.2 Copy number variation

CNVs are the most common of all structural variants in the human genome (Stankiewicz and Lupski, 2010). They are unbalanced rearrangements that result in increase (gains) or decrease (deletions) in the DNA content and therefore alter the diploid status of the genome (Spielmann et al., 2018; Zarrei et al., 2015). CNVs involve more polymorphic base pairs than any other type of genetic variation including SNVs (Rice and McLysaght, 2017; Zarrei et al., 2015). On average, each individual has more than a 1,000 CNVs greater than 450bp in size, most of which are inherited, but some occur *de novo* (Conrad et al., 2010; Rice and McLysaght, 2017). Up to 10% of the human genome shows structural variation in the general population, which is mainly caused by CNVs (Kearney et al., 2011; Lupiáñez et al., 2015; Zarrei et al., 2015).

Although other forms of human genome variation (SNVs and cytogenetically recognisable structural variants) have been known for long time, the large-scale presence of CNVs across the human genome was only recognised relatively recently (Iafrate et al., 2004; Sebat et al., 2004; Zarrei et al., 2015). This was mainly due to technical developments and the use of array comparative genome hybridisation (aCGH) (Iafrate et al., 2004; Sebat et al., 2004). Initially, CNV size was defined as larger or equal to 1Mb (Iafrate et al., 2004), however as microarray technology further advanced and resolution improved, CNVs are now defined as any unbalanced structural variants larger than 50bp, whereas smaller elements are known as insertions or deletions (indels) (Zarrei et al., 2015).

1.2.1 CNV and human disease

Although CNVs are largely polymorphic, pathogenic disease-causing CNVs with variable penetrance have also been identified (Hegele, 2007; Ibn-Salem et al., 2014; Rice and McLysaght, 2017; Riggs et al., 2020; Zarrei et al., 2015; Zhang et al., 2009). This can range from Mendelian disease to sporadic and complex disorders such as cancer, autism, schizophrenia and various other neurodevelopmental disorders (Coughlin et al., 2012; Shaikh et al., 2009; Xu et al., 2008; Zarrei et al., 2015).

It is estimated that approximately 5% of children are affected by developmental disorders of varying severity that result in various combinations of developmental delays, intellectual disability, neurological or behavioural problems and congenital malformations (McRae et al., 2017). CNVs are implicated in ~14% of cases of developmental disorders in children and therefore contribute significantly to childhood morbidity (Cooper et al., 2011; Sansovi et al., 2017). Hence, genome-wide testing of CNVs (losses and gains) is recommended as a first-tier test in the investigations of individuals with congenital abnormalities, developmental delays, autism spectrum disorders and intellectual disabilities (Riggs et al., 2020)

1.2.2 CNV detection and diagnosis of disease

The introduction of high resolution chromosomal microarray technologies, such as aCGH and single nucleotide polymorphism (SNP) arrays, into clinical practice have revolutionised investigations for patients with intellectual disability, developmental delays, autism spectrum disorders, epilepsy and multiple congenital abnormalities (Conrad et al., 2010; Hehir-Kwa et al., 2010; Miller et al., 2010; Rice and McLysaght, 2017; Riggs et al., 2020; Schaaf et al., 2011).

Microarrays, especially aCGH, have now been used for more than a decade as a first-line diagnostic test for the above patient cohorts (Ahn et al., 2013; Miller et al., 2010; Riggs et al., 2020). Due to the significantly higher resolution and the associated greater diagnostic yield, aCGH was considered a major improvement from the use of more traditional cytogenetic techniques such as karyotyping and fluorescence in situ hybridization (FISH) (Ahn et al., 2013).

1.2.3 CNVs – mechanisms of pathogenesis

There are several ways by which structural variants, including CNVs, can be pathogenic and cause or predispose to a disease phenotype. These include disruption of the copy number of dosage sensitive genes, gene interruption, gene fusion, spatial organisation and position effect which includes interfering with regulatory elements, disruption of chromosomal structures, and unmasking of recessive alleles (for example, being *in trans* with a recessive disease causing SNV) (Boone et al., 2013; Rice and McLysaght, 2017; Zhang et al., 2009).

Copy number disruption of dosage sensitive genes is thought to be the most common mechanism of CNV pathogenicity (Rice and McLysaght, 2017). Dosage disruption can happen through deletion of the gene, which may result in underexpression, or duplication of the gene and possible overexpression. This is supported by the positive correlation between copy number of genes and their expression levels in various settings (Henrichsen et al., 2009). The *PMP22* gene is a typical example of a dosage sensitive gene. Deletion of *PMP22* causing underexpression is associated with hereditary neuropathy with liability to pressure palsies (HNPP) (OMIM #162500) (Zhang et al., 2009). On the other hand, *PMP22* duplication causing overexpression is associated with Charcot-Marie-Tooth Neuropathy Type 1A (CMT1A) (OMIM # 118220) (Zhang et al., 2009). Gene interruption, is another mechanism of CNV pathogenesis and can happen when structural variants' breakpoints are located within functional genes, examples include intragenic *DMD* gene deletions and duplications associated with Duchene and Becker muscular dystrophies (Hegde et al., 2008).

Interfering with regulatory elements, such as silencers, promoters, enhancers and Topologically Associating Domains (TADs), is another mechanism by which CNVs are known to cause human disease (Lupiáñez et al., 2015; Zarrei et al., 2015; Zhang and Lupski, 2015). This is especially true when CNVs occur in non-coding regions. Examples include 8kb deletion upstream *ROU3F4* associated with deafness and deletion/duplications upstream and downstream of *SOX9* resulting in multiple syndromes including sex reversal and Pierre Robin sequence (Benko et al., 2009; de Kok et al., 1996; Zhang and Lupski, 2015). This area is gaining more attention now, as our understanding of the human genome regulation is improving.

1.3 Clinical interpretation of CNVs

In order to be useful in a clinical setting, pathogenic disease causing CNVs need to be differentiated from benign ones. In diagnostic genomic laboratories in the UK, CNVs are mainly detected using aCGH and then systematic clinical interpretation is performed in order to differentiate disease-causing CNVs from benign ones (de Leeuw et al., 2012; Riggs et al., 2020; South et al., 2013). In order to interpret CNVs accurately and consistently, clinical laboratories follow the American College of Medical Genetics (ACMG) standards and guidelines published in 2011 (Kearney et al., 2011) and recently updated by the publication of the joint consensus recommendation of the ACMG and the Clinical Genome Resource (ClinGen) (Riggs et al., 2020). These guidance are based on multiple criteria including genomic content, CNV size, comparison to CNVs in internal and external databases, published literature and familial studies.

1.3.1 The five-class system

As recommended by the ACMG and ClinGen/ACMG guidelines, the clinical significance of CNVs is classified into one of five categories: benign (class 1), likely benign (class 2), variants of uncertain significance (VUS – class 3), likely pathogenic (class 4) and pathogenic (class 5) (Kearney et al., 2011; Riggs et al., 2020) (Figure 1.1).

Parameter	Benign (Class 1)	Likely benign (Class 2)	Variants of uncertain significance (Class 3)	Likely pathogenic (Class 4)	Pathogenic (Class 5)
Point value (Riggs et al 2019)	≤-0.99	-0.9 to -0.98	-0.89 to 0.89	0.9 to 0.98	≥0.99
Size	← Tend to be smaller		Varies	Tend to be larger →	
Gene(tic) content	← Few genes or none		Some with genes Some in non-coding regions	Several genes, disease genes →	
Reported on databases?	← Frequently seen		Not common	Listed on Disease databases →	
Internal databases	← Frequently seen		Not common	Not seen/reported in other patients →	
Published literature			Varies	Published →	

Figure 1.1 A simplified guide to CNV clinical interpretation

A simplified overview of the different CNV clinical classes based on local and published guidelines (Kearney et al., 2011; Riggs et al., 2020). The criteria are simplified here to represent the concept. The different resources are usually used in combination, which varies based on the available/published evidence.

1.3.2 CNV classes and clinical reporting

According to the local policy, variants that are classified as pathogenic (class 5) or likely pathogenic (class 4) and are thought to be contributing or responsible for the patient's phenotype are summarised on the front page of the clinical report. On the other hand, variants that are considered benign (class 1) or likely benign (class 2) are not reported back to the referring clinicians as they are considered part of the normal population variation. Class 3 VUSs are mostly summarised on the back of the clinical report unless there is a degree of doubt that they maybe likely pathogenic and need to be brought to the attention of the referring clinician. In these circumstances, class 3 VUS will be summarised at the front of the report with possible follow up and family studies' recommendations.

1.3.3 Criteria and resources for clinical interpretation

1.3.3.1 CNV size and genomic content

Genomic content is the most informative criteria used in clinical interpretation of CNVs (Kearney et al., 2011; Riggs et al., 2020). Clinical scientists investigate whether the CNV region contains or affects any genes or not. If genes are present/affected, they are usually scrutinised for any published association with clinical disorders. This includes literature search and checking relevant databases such as Online Mendelian Inheritance in Man (OMIM). It is also prudent to check if the genes are reported to be dosage sensitive including the use of ClinGen Dosage Sensitivity Map. If dosage sensitive, it is important to correlate the CNV type (deletion/duplication) and predicted effects (loss/gain) on gene dose with the patient's phenotype based on published information (Conrad et al., 2010; Kearney et al., 2011; Lupiáñez et al., 2016; Riggs et al., 2020; Zarrei et al., 2015). For example deletion of a gene, in which only gain-of-function pathogenic variants cause disease, is unlikely to be deleterious (Kearney et al., 2011; Riggs et al., 2020).

1.3.4 Interrogation of curated public and local databases

Another important tool in defining CNV pathogenicity is through comparing them to those found in apparently healthy individuals or those previously detected in similarly affected patients

(Conrad et al., 2010; Hehir-Kwa et al., 2010; Kearney et al., 2011; Lupiáñez et al., 2016; Riggs et al., 2020; Zarrei et al., 2015). The Database of Genomic Variants (DGV) is one of the main resources that provide an up to date curated catalogue of human genomic structural variants identified in DNA samples from healthy controls (MacDonald et al., 2014). More recently, gnomAD introduced gnomAD SV based on high resolution SVs created from ~15,000 genomes sequenced from diverse populations and providing a catalogue of >433,000 SVs (Collins et al., 2020). The DECIPHER database (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources), on the other hand, provides a catalogue of genomic variants found in affected patients in addition to the associated phenotype (Firth et al., 2009). These are the most commonly used databases for the interpretation of CNVs. Most genomic laboratories will also have local databases whereby any previously identified and classified CNVs would have been recorded.

1.4 Class 3 CNVs – the known that is unknown

Growing number of variants of uncertain significance pose a major challenge for the interpretation of CNVs in a clinical setting (Zarrei et al., 2015). Locally, they comprise ~8% of all detected CNVs (7,198 variants of uncertain significance identified in ~7 years) and affect ~27% of all of our referrals (carry a least one class 3 CNV). In general, VUSs pose challenges to the scientists, clinicians and importantly to the patients (Hoffman-Andrews, 2017; Lovato et al., 2019). For scientists, the process of classifying variants that end up as VUS is usually lengthy and time consuming. Moreover, there are also challenges of deciding which VUSs should be reported and what kind of follow up testing to be recommended (Ellard et al., 2020). For clinicians, issues include decisions on results disclosure, counselling and follow up studies (Hoffman-Andrews, 2017). For the patients, challenges include understanding VUSs and dealing with uncertainties (Hoffman-Andrews, 2017). A recent meta-analysis have shown that patients who receive a VUS are more likely to have genetic test-specific concerns than patients with negative results, without any benefits to their medical management (Mighton et al., 2020).

As part of the DClinSci innovation proposal (C1), I sent a questionnaire to all of the consultants' clinical geneticists in the developmental delays team who refer to the aCGH service. Nine consultants responded and 8 out of the 9 agreed that improving the classification of CNVs and reducing the incidence of VUSs will aid in the diagnosis and clinical management of patients. In summary, there is a real clinical need to improve the interpretation and classification of CNVs.

1.5 Socioeconomic status

Socioeconomic status (SES) is a measure of an individual's or family's economic and social status based on multiple factors including income, education, and occupation (Bradley and Corwyn, 2002). SES is a major determinant of health and related outcomes (Marioni et al., 2014). For example, lower SES confers increased risk for multifactorial disorders like stroke, and cardiovascular disease and plays a key role in child health and development (Clark et al., 2009; Cox et al., 2006). Early life adversity negatively impacts child health and produces lasting and deleterious effects on developmental outcomes (Hackman and Farah, 2009). It was previously shown that SES is positively associated with health and that low SES can impact an individual's physical and mental health (Bradley and Corwyn, 2002). Also, SES is considered an essential predictor of cognitive regression in later age (Singh-Manoux et al., 2005).

Existence of SES health gradient across populations and diseases is well recognised and much research has gone into unravelling the multiple pathways by which SES determines health (Wang and Geng, 2019). However, biological factors, including genetic variants, that may influence SES are only beginning to be understood.

1.6 Aims and Objectives

The overall aim of the research described here was to improve the classification and understanding of the impact of CNVs.

There were two main hypotheses that directed our work:

1. Systematic re-analysis of class 3 CNVs can improve CNV classification and help in deciphering biological mechanisms underlying disease pathogenicity
2. Pathogenic CNVs and SES are correlated.

Our objectives were:

1. Devise a strategy to facilitate large-scale re-analysis for a sub-group of class 3 CNVs.
    a. Create and curate a large data set of CNVs and their clinical classifications
    b. Annotate all CNVs with gene-centred information including names of genes affected by CNVs, their haploinsufficient scores and any associated syndromes
    c. Prioritise a shortlist of class 3 CNVs that could undergo a full reclassification
2. Study the correlation between pathogenic CNVs and SES.
    a. Annotate our curated data set with inheritance information and postcode information
    b. Retrieve the index of multiple deprivation rank (IMDR), and its seven constituent domains for all of the postcodes
    c. Annotate all the CNVs with the IMDR and constitute domain information
    d. Study the correlation between pathogenic CNVs and SES

1.7 Thesis overview

1.7.1 Main Thesis

This thesis forms module C2 of the University of Manchester qualification for the professional Doctorate of Clinical Sciences (DClinSci) and is presented in the chapters described below.

Please note that this is thesis is submitted as a journal format thesis as chapters 3 and 4 have already been peer-reviewed and published (Burghel et al., 2020; Tinker et al., 2020). Chapter 2, a systematic re-analysis of copy number losses of uncertain clinical significance, will be submitted for publication soon after the completion of this thesis. Contributions of co-authors and colleagues are summarised at the end of each of the results chapters 2-4.

Please note that the thesis does not contain a separate methods chapter. The reason behind this is that each of the results chapters 2-4 contain a detailed description of all the methods used for that specific part of the work. Rewriting a separate methods chapter would have either replicated the detailed methods already described in the published chapters 3 and 4 or would have involved considerable changes to the otherwise a representation of already published work. Hence, for chapter 2, which is yet to be published, but written in a scientific paper style, detailed methods were also described and included within the chapter itself. Chapter 5 discusses the overall picture of the work presented in this thesis, some of the limitations and ideas for future work.

In addition to the work described here, the appendices include references to certificates and outcomes of other requirements to complete the DClinSci.

1.7.2 Appendices

Appendix A: PGDip (DClinSci Module A) results

Module A included five taught units of leadership and management in healthcare sciences from The University of Manchester Alliance Manchester Business School which were examined through the submission of two written essays for each unit. The final marks as ratified by the University of Manchester Board of Examiners are presented in this appendix. An overall distinction was achieved for module A of the DClinSci.

Appendix B: FRCPath exam results (DClinSci module B)

Importantly, completion of the DClinSci module B (and HSST training) in genomics require passing relevant Royal College of Pathologists (RCPath) examinations (Module B). The author

passed RCPath Part One (theory) in 2016 (Appendix B1), FRCPath Part One practical in 2017 (Appendix B2) and FRCPath Part Two oral examination in 2019 (Appendix B3).

Appendix C: DClinSci Module C1 examination feedback

HSST C1 innovation business case: the research innovation Module C1 was presented to Manchester University scientific and lay staff as part of the HSST training and DClinSci C1 examination. The feedback for the C1 innovation module is presented here. The C1 literature review itself formed the basis of the introduction to the thesis therefore it is not repeated here. As noted in the appendix, I received excellent feedback and the assessor noted that there were no improvements to be made.

Appendix D: Further Qualifications and HSST relevant courses

The author also achieved Level 5 Certificate in Leadership from the institute of Leadership and Management (Appendix D1) and completed a competitive Health Education England funded leading transformational culture change course (Appendix D2). The author also completed two relevant Genomic Medicine Level 7 courses (University of Manchester) in Economics of Genomics and Precision Medicine (Appendix D3) and Ethical, Legal and Social Issues in Applied Genomics (Appendix D4).

2 A SYSTEMATIC RE-ANALYSIS OF COPY NUMBER LOSSES OF UNCERTAIN CLINICAL SIGNIFICANCE FROM A REGIONAL GENOMIC DIAGNOSTIC LABORATORY

2.1 Introduction

CNVs are the commonest of all structural variants in the human genome (Stankiewicz and Lupski, 2010). They are unbalanced rearrangements that result in increase (gains) or decrease (deletions) in the DNA content (Zarrei et al., 2015). Although largely polymorphic and a source of variation and diversity, pathogenic CNVs implicated in human diseases have also been identified (Rice and McLysaght, 2017; Zarrei et al., 2015). This can range from Mendelian disease to sporadic and complex syndromes such as cancer, autism, schizophrenia and various other neurodevelopmental disorders (Coughlin et al., 2012; Shaikh et al., 2009; Xu et al., 2008; Zarrei et al., 2015). It is estimated that approximately 5% of children are affected by developmental disorders of varying severity that result in various combinations of developmental delays, intellectual disability, neurological or behavioural problems and congenital malformations (McRae et al., 2017). CNVs are implicated in ~14% of cases of developmental disorders in children and therefore contribute significantly to childhood morbidity (Cooper et al., 2011; Sansovi et al., 2017). The diagnostic yield of array-CGH in diagnostic laboratories have been estimated as ~12% (Miller et al., 2010).

Due to the size and complexity of many of the identified CNVs, clinical interpretation remains challenging and results in frequent classification of variants of uncertain significance (VUS) (Riggs et al., 2020). The latter remains a major issue in the field of clinical genomics and create issues for patients and healthcare professionals (Hoffman-Andrews, 2017). A recent meta-analysis have shown that patients who receive a VUS are more likely to have genetic test-specific concerns than patients with negative results, without any benefits to their medical management (Mighton et al., 2020).

We have previously shown that deeper interrogation of this data can improve clinical interpretation, lead to discovery of new genetic disorders and provide novel insights into fundamental biology and disease mechanisms (Banka et al., 2015; Kasher et al., 2016; Tinker et al., 2020). Previous studies have shown that periodic reassessment of whole exome sequencing (WES) and whole genome sequencing (WGS) data helps in the reclassification of class 3 variants into more informative categories (benign or pathogenic) and improve diagnostic

yield. The level of improvement varied a lot depending on the used methodology (multi-gene panels, WES, WGS), time elapsed since initial analysis and cohort phenotype with a variety of reported rates of reclassification: 0% (Tan et al., 2020), 2-6% (Hiatt et al., 2018; Salfati et al., 2019; Sun et al., 2019), around 10% (Bruel et al., 2019; Costain et al., 2018; Das et al., 2014) and up to 13% (Wright et al., 2018). However, so far there is no study to demonstrate the value of CNVs identified via aCGH.

Re-analysis of aCGH CNV data is timely for many reasons; hundreds of new disease-genes have been defined since the start of using aCGH as a first tier test (Bamshad et al., 2019), approaches to interpretation have evolved and are more standardised (Riggs et al., 2020), the relatively recent computation of reliable haploinsufficiency scores and the availability of rich data sets such as Decipher (Huang et al., 2010; Lek et al., 2016; Petrovski et al., 2013). However, re-analysis of Class 3 CNVs detected in a diagnostic laboratory can be challenging. For example, our centre performed diagnostic array Comparative Genomic Hybridisations (aCGH) testing in more than 16,000 cases between 2010 and 2017 and at least one class 3 CNV were identified in 27% of the cases (unpublished local data). Case-by-case analysis of a total of >4,000 cases will therefore, require significant resources and is impractical in a busy diagnostic laboratory. We, therefore, decided to devise an approach for efficient systematic reanalysis of class 3 CNVs detected in our diagnostic laboratory. For this study we focussed on re-analysis of Copy Number Loss (CNL) because, consequences of deletions could be assumed by loss of one or more haploinsufficient genes within the deleted region (Collins et al., 2021; Rice and McLysaght, 2017). Whereas, interpretation of copy number gains (CNGs) of uncertain significance remains challenging and less amenable to group-based analysis (Collins et al., 2021; Hurles et al., 2008).

2.2 Material and Methods

2.2.1 Ethical approval

This project was designated as service improvement project. Reanalysis of diagnostic genetic data is an accepted routine practice and does not require additional patient consent.

Furthermore, the Ethics and General Data Protection Regulation (GDPR) team at the University of Manchester reviewed the project proposal and considered that this project raises no particular research ethics issue and therefore does not require approval from the Research Ethics Committee.

2.2.2 CNV categorisation

We anonymised and curated a departmental database of results from over 16,000 postnatal clinical array-CGH testing performed at MCGM between 2010 and 2017 (data exported 01/06/2017). The database included information on chromosomal location, clinical classification, size, loss or gain status and where available, the inheritance status (*de novo* or inherited from a parent) for each CNV. Historically, different terms have been used in MCGM to describe the various CNV classes (Table 2.1). Following manual examination of various previous terms, they were all re-assigned to one of the five current classes (benign, likely benign, variant of uncertain significance, likely pathogenic and pathogenic) (Kearney et al., 2011; Riggs et al., 2020).

Table 2.1 CNV categories

ACMG CNV class	Selection of historical terms found in the database
Benign	Benign; CNV seen in normal individuals; Normal CNV; Not reported (not pathogenic)
Likely Benign	Likely Benign; Unlikely significant; Reported (unlikely significance); Likely Rare CNV; Not reported (Not plausibly pathogenic)
Uncertain significance	Uncertain Significance; Unclassified; Unknown significance; Reported (Uncertain significance), CNV with unknown significance
Likely pathogenic	Likely pathogenic
Pathogenic	Pathogenic; Pathogenic - Causative; Pathogenic - Incidental; Pathogenic - Susceptibility Locus; Neurosusceptibility pathogenic; Reported (Pathogenic)

CNVs were then split according to the type of imbalance (loss/gain) and the work reported here only focussed on CNLs. These were then clustered into three groups based on reported clinical classification; Clinically pathogenic (classes 4 and 5), Clinically benign (classes 1 and 2) and VUSs (Class 3).

We excluded any CNLs that were greater than 4.8Mb in size. This is ~10% of chromosome 21 (smallest chromosome) and around the limit of microscopically visible CNVs on karyotyping.

These CNLs are likely to have large number of genes which will complicate and confound the analysis as previously described (Rice and McLysaght, 2017). To further reduce the number of genes to analyse, we trimmed all pathogenic CNLs and class 3 CNLs of any genomic regions that they shared with benign ones. Haploinsufficiency of genes found within regions shared with benign CNLs are less likely to be responsible for disease.

2.2.3 CNV annotation

Based on genomic location, we identified all the genes that overlapped with the three CNL groups (minimal CNL coordinates (Human genome build hg19)). In order to identify any genes of interest that may drive class 3 CNL reclassification, we annotated all of the genes (based on the HUGO gene nomenclature committee (HGNC) name), with information from the online mendelian inheritance in man database (OMIM); OMIM number, associated syndrome and mode of inheritance (if available/applicable) and also the mutation consequence as per the Development Disorder Genotype - Phenotype Database (DDG2P).

A recent genome-wide analysis has shown that genes residing within regions that are exclusively affected by pathogenic CNVs are more likely to be haploinsufficient compared to those that occur in regions that are exclusively affected by benign CNVs (Rice and McLysaght, 2017). This biased distribution provided evidence for the potential application of such scores in the clinical interpretation and reinterpretation of CNVs. Multiple statistical predictions of haploinsufficiency are available including the probability of being loss-of-function (LoF) intolerant (pLI), Residual Variation Intolerance Score (RVIS) and the Decipher haploinsufficiency (D-HI) score (Huang et al., 2010; Lek et al., 2016; Lessard et al., 2016; Petrovski et al., 2013). We also annotated all of the genes with their dosage sensitivity scores (pLI, RVIS and Decipher HI) where available.

2.2.4 Statistical analysis

Chi squared tests of independence and trend were performed using an online calculator (https://www.socscistatistics.com/tests/chisquare/) and statistical significance was set at p-value <0.05. The tests were performed to examine the associations between three haploinsufficiency prediction scores and the clinical classification.

2.2.5 Gene-centric analysis and shortlisting

Genomic content is the main cause of CNV pathogenicity (Riggs et al., 2020). Therefore, we devised a systematic approach to short list genes of interest within class 3 CNLs that may be pathogenic upon loss and possibly drive CNL reclassification. All genes within cases that have pathogenic CNLs alongside the class 3 CNLs were excluded as well as any genes that occurred within common CNLs (occurring in ≥100 cases). These common CNLs are likely to be either artefactual or (likely)benign (Figure 2.1). Genes were shortlisted for further analysis if they were associated with an OMIM syndrome with autosomal dominant inheritance and/or a loss of function consequence as defined by DDG2P. Genes were further shortlisted if they were predicted to be haploinsufficient with pLI score ≥0.9 and RVIS score ≤0.01 (Figure 2.1).

2.2.6 Analysis of shortlisted genes & reclassification

Genes were further shortlisted manually by a Clinical Scientist. This was based on relevance of associated phenotype (Relevant phenotypes included developmental delays, congenital anomalies, autistic spectrum disorders etc), mechanism of disease (Loss of function), manual review of the array calls (where possible) and any similar CNVs on Decipher. Any CNLs that included shortlisted genes after this manual review had a full clinical interpretation following local procedures, the recent ClinGen/ACMG guidance and the ClinGen CNV pathogenicity calculator (https://cnvcalc.clinicalgenome.org/cnvcalc/) (Riggs et al., 2020).

2.3 Results

2.3.1 CNV categorisation and shortlisting

Our database export included 83,993 CNVs (46,984 Gains (CNGs) and 37,009 Losses (CNLs)) from 16,358 postnatal cases. We then focussed on CNLs and after standardising terminology of classification (as per Table 2.1), we then subdivided them by class; 33,857 benign CNLs, 2,173 class 3 CNLs and 979 pathogenic CNLs (Figure 2.1). We then applied a size filter and excluded any CNLs that were greater than 4.8Mb in size. This left 33,857 benign CNLs, 2,130 CNLs of uncertain significance and 707 pathogenic CNLs (Figure 2.1). The CNLs were distributed across the genome (Supplementary Figure 2.3).

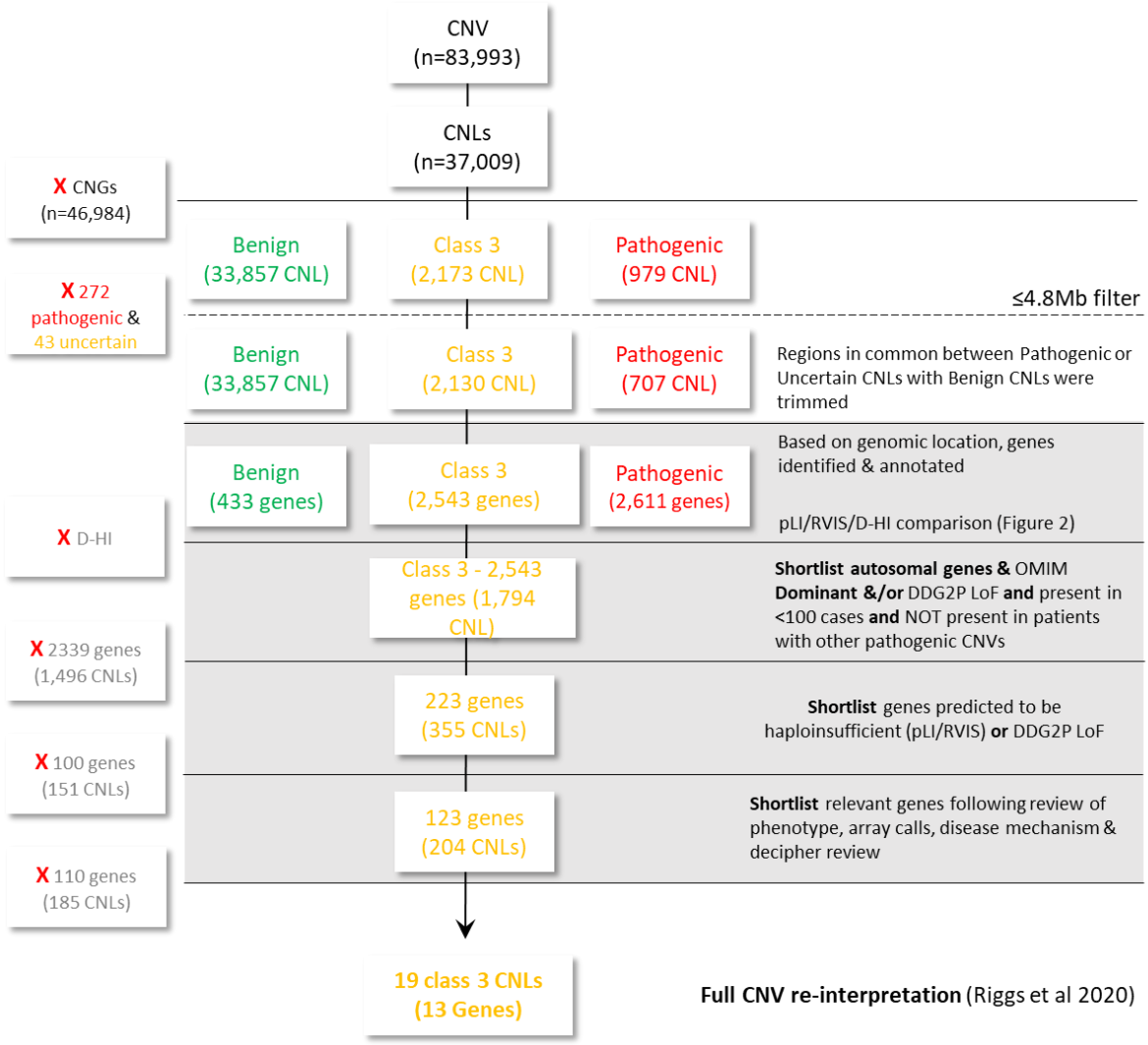


Figure 2.1 Systematic CNL analysis

2.3.2 Gene-based shortlisting

Based on genomic location, we trimmed all pathogenic and class 3 CNLs of any regions that they shared with benign CNLs. This was performed to exclude regions/genes in common with benign CNLs as they are considered less likely to be pathogenic (Rice and McLysaght, 2017). We then identified all the genes within all categories (trimmed coordinates for class 3 and pathogenic CNLs). This resulted in 2,611 genes within regions affected by pathogenic CNLs and 2,543 genes within regions affected by class 3 CNLs and 433 genes in benign CNLs. All genes (based on the HGNC gene name) were then annotated for OMIM information, DDG2P information and haploinsufficiency scores (November 2018).

2.3.3 Landscape of haploinsufficiency predictions

We analysed the patterns of haploinsufficiency within each of the CNL groups using the pLI, RVIS and D-HI scores as proxy for haploinsufficiency. We found that 78-96% of pathogenic CNLs included at least one haploinsufficient gene, while 7-49% of benign CNLs were found to have at least one haplosinufficient gene. For class 3 CNLs, 34-55% were found to have at least one haploinsufficient gene (Table 2.2).

Table 2.2 CNLs with at least one haploinsufficient gene

		Predicted Haploinsufficient genes		
CNL class (Total No of CNLs)	**Unique CNLs (n)**	**pLI (%)**	**RVIS (%)**	**D-HI (%)**
Pathogenic (707)	410	345 (84.1)	394 (96.1)	321 (78.3)
Class 3 (2,130)	1,489	509 (34.2)	812 (54.5)	516 (34.7)
Benign (3,3857)	1,314	95 (7.2)	174 (13.2)	643 (48.9)

When the three CNL categories were compared against each other, benign CNLs were found to be significantly less likely to contain one or more haploinsufficient genes compared to pathogenic CNLs. This difference was significant across the three haploinsufficiency scores; pLI ~84% (Pathogenic) compared to ~7% (Benign), RVIS ~96% (Pathogenic) versus ~13% (benign) and D-HI ~78% (Pathogenic) compared to ~49% (Benign) (P=<0.00001 X^2-Test) (Figure 2.2). These differences in distribution of haploinsufficient genes were also significant for class 3 CNLs but only for pLI (~34% (class 3) vs ~7% (benign)) and RVIS (~54% (class 3) vs ~13% (benign) (P=<0.00001 X^2-Test), but not the D-HI score. The latter predicted that benign CNLs were more likely to have ≥1 haploinsufficient gene (~49%) compared to class 3 CNLs (~35%) (P=<0.00001 X^2-Test) (Figure 2.2).

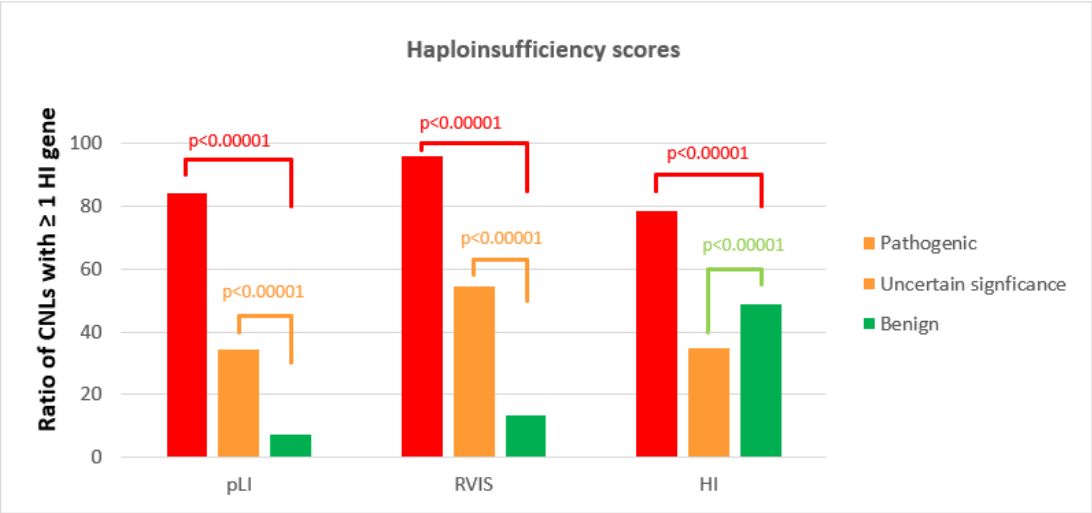


Figure 2.2 Haploinsufficiency scores enrichment

Comparison of CNLs with ≥ 1 Haploinsufficient genes across the 3 CNL categories (Pathogenic, uncertain significance and benign). Please note that ratios are presented here, while X^2-Test p-value is based on total numbers. Details Supplementary Table 2.4.

Overall, out of all the genes deleted only in pathogenic CNLs, ~19% were considered to be haploinsufficeint by pLI and ~21% by RVIS. Similar patterns were observed for class 3 CNLs (pLI ~19% and RVIS ~24%). Using pLI and RVIS, both pathogenic and class 3 CNLs had higher levels of haploinsufficient genes compared to benign CNLs (~12% (PLI) and ~16% (RVIS)) (Supplementary Table 2.4). In contrast D-HI seems to predict more haploinsufficient genes within benign CNLs (~79%) compared to pathogenic CNLs (~55%) and class 3 CNLs (~51%) (Supplementary Table 2.4).

Due to the overprediction of haploinsufficient genes using D-HI within benign CNLs, this score was excluded from further analysis. On the other hand, enrichment of haploinsufficient genes as predicted by pLI and RVIS in pathogenic CNLs confirmed that it is reasonable to use them to predict haploinsfficient genes within class 3 CNLs.

2.3.4 Shortlisting of genes within CNLs of uncertain significance

A total of 2,543 genes (1,794 class 3 CNLs) were annotated for information from OMIM, DDG2P in addition to haploinsufficiency scores (pLI and RVIS) (Figure 2.1). Autosomal genes were included if they have OMIM disease association (with dominant inheritance) and/or LoF

mutation consequence by DDG2P, present in less than 100 cases and not present in patients who had other pathogenic CNVs. This resulted in a shortlist of 223 genes (355 CNLs). Following that, genes were further shortlisted based on being haploinsufficient as predicted by both pLI (score ≥0.9) and RVIS (score ≤0.01) scores but also including DDG2P mutation consequence of LoF. This resulted in a list of 123 genes (204 CNLs) (Figure 2.1).

This list was then assessed by clinical scientists and this included; review of original aCGH data (where possible) to exclude any likely-artefactual CNLs and non-coding CNLs; examining phenotypic information of associated syndrome(s) to include relevant phenotype (e.g. developmental delays, congenital anomalies, autistic spectrum disorders etc) and exclude any non-relevant/late age of onset phenotypes (such as cancer) and likely mechanism of pathogenicity (LoF vs gain of function (GoF)) in addition to similar and overlapping Decipher cases. The clinical scientist review resulted in a priority shortlist of 13 genes (19 CNLs).

2.3.5 Re-classifications

The 19 CNLs had a full CNV interpretation as per the local procedures and latest ACMG/ClinGen criteria (Riggs et al., 2020). This resulted in a total of 13 CNL reclassifications, all of which were due to genes and disease association (or disease-causing mechanism) described post the identification and the original classification and reporting of the CNV. Our reclassification rate is 6.4% of the shortlisted CNLs (204 CNLs) and ~0.6% of the starting cohort of 2,173 CNLs. All the genes (and reclassified CNLs) are listed in Table 2.3 and described in the following sections.

2.3.5.1 Shortlisted genes and reclassified cases

2.3.5.1.1 NAA15

Loss of function variants in *NAA15* cause autosomal dominant mental retardation 50 (MRD50, OMIM #617787) and has a ClinGen haploinsufficiency score of 2 (https://dosage.clinicalgenome.org/; reviewed February 2019). Our case was identified in 2016 with a 28kb deletion affecting exons 2-11 of *NAA15* (Table 2.3). Referral reasons included

severe learning difficulties with mother and brother both affected with severe learning difficulties. In 2017, several cases with developmental delays and intellectual disabilities and loss of function pathogenic variants in *NAA15* were published as part of a large study into the biology of neurodevelopmental disorders (Stessman et al., 2017). This was followed by another study that described 39 cases, mostly from unrelated families, with intellectual disabilities, congenital abnormalities and autistic spectrum disorder associated with loss of function pathogenic variants in *NAA15* (Cheng et al., 2018). Based on the above evidence, the CNL was reclassified as likely pathogenic (Riggs et al., 2020).

2.3.5.1.2 ZMYND11

Loss of function variants in *ZMYND11* cause autosomal dominant mental retardation 30 (MRD30, OMIM #616083) has a ClinGen haploinsufficiency score of 3 (https://dosage.clinicalgenome.org/; reviewed March 2018). Two cases were identified in this study in 2014 with deletions affecting *ZMYND11*; one patient was referred for autistic spectrum disorder, attention deficit hyperactivity disorder and moderate learning difficulties and had a 90kb CNL affecting the first 2 exons of *ZMYND11* (Table 2.3). The 2nd case was referred with developmental delay and poor language skills and had a 776kb deletion affecting the whole gene (and 2 more protein coding genes). Later in the same year, *ZMYND11* had the first report linking its loss to developmental delays, autism, aggression and complex neuropsychiatric features (Coe et al., 2014). Based on the available evidence, both cases were reclassified as likely pathogenic (Riggs et al., 2020).

2.3.5.1.3 ACTB

Pathogenic variants in *ACTB* cause autosomal dominant Baraitser-Winter syndrome-1 characterised by developmental delay, short stature and various dysmorphic features (BRWS1, OMIM #243310) and no ClinGen dosage sensitivity score to date (https://dosage.clinicalgenome.org/; last reviewed February 2015). Our case was identified in 2011 with a 556kb CNL resulting in the loss of *ACTB* (and 6 other protein coding genes). Patient had features including short stature, microcornea, inguinal hernia, abnormality of the kidney, recurrent infections and intellectual disability (Table 2.3). Although missense *ACTB* variants were linked to BRWS1 for few years, it was not until 2017, when loss of function of *ACTB*

variants were also shown to lead to a developmental disorders (Cuvertino et al., 2017). This CNL was therefore reclassified as pathogenic (Riggs et al., 2020).

2.3.5.1.4 KMT2C

Loss of function variants in *KMT2C* cause autosomal dominant Kleefstra syndrome-2 characterised by developmental delays, intellectual disability and mild dysmorphic features (KLEFS2, OMIM #617768) and has a ClinGen haploinsufficiency score of 3 (https://dosage.clinicalgenome.org/; reviewed January 2018). Our case was identified in 2014 with a *de novo* 2,971kb CNL resulting in the loss of *KMT2C* (and 44 other protein coding genes). Patient had features including hypotonia and growth delays (Table 2.3). Loss of function pathogenic variants of *KMT2C* were shown in 2017 and 2018 to be responsible for KLEFS2 (Faundes et al., 2018; Koemans et al., 2017). Based on this evidence, this CNL was reclassified as pathogenic (Riggs et al., 2020).

2.3.5.1.5 PHF21A

Loss of function variants in *PHF21A* cause autosomal dominant intellectual developmental disorder with behavioural abnormalities and craniofacial dysmorphism with or without seizures (IDDBCS, OMIM #618725) and has a ClinGen haploinsufficiency score of 2 (https://dosage.clinicalgenome.org/; reviewed May 2020). Our case was identified in 2015 with a *de novo* 1,446kb CNL resulting in the loss of *PHF21A* (and 34 other protein coding genes). Patient had delayed gross motor development and multiple congenital anomalies (Table 2.3). Loss of function truncating variantss of *PHF21A* were reported in 2019 in association with IDDBCS (Hamanaka et al., 2019; Kim et al., 2019). This CNL was reclassified as likely pathogenic (Riggs et al., 2020).

2.3.5.1.6 MYT1L

Loss of function pathogenic variants in *MYT1L* cause autosomal dominant mental retardation 39 (MRD-39, OMIM #613084) and has a ClinGen haploinsufficiency score of 3 (https://dosage.clinicalgenome.org/; reviewed March 2017). Our case was identified in 2015 with a *de novo* 687kb CNL resulting in the loss of exons 1-8 of *MYT1L*. Patient was referred due to unclear speech, truncal obesity, joint hyperextensibiulity and decreased muscle tone (Table 2.3). Prior to 2015, there was only one case reported in the literature with a *de novo*

splice-site variant in *MYT1L* associated with mental retardation (de Ligt et al., 2012). Another case was reported in 2015 (De Rocker et al., 2015) and then few more cases in 2017 (Blanchet et al., 2017). MRD39 is associated with developmental delays (including impaired language), hyperphagia and obesity, hypotonia and autistic features (Blanchet et al., 2017; De Rocker et al., 2015). This CNL was therefore reclassified as likely pathogenic (Riggs et al., 2020).

2.3.5.1.7 PSMD12 and BPTF

Loss of function pathogenic variants in *PSMD12* cause autosomal dominant Stankiewicz-Isidor syndrome (STISS, OMIM #617516) characterised by developmental delays, behavioural disorders, congenital abnormalities and facial dysmorphism (Küry et al., 2017) while pathogenic variants in *BPTF* are associated with autosomal dominant neurodevelopmental disorder with dysmorphic facies and distal limb anomalies (NEDDFL, OMIM #617755) (Stankiewicz et al., 2017). Both genes are awaiting review by the ClinGen dosage sensitivity curation team (https://dosage.clinicalgenome.org/; accessed March 2021). Our case was identified in 2014 with a 528kb CNL deleting exon 1 of *PSMD12*, and exons 1-4 of *BPTF* (and 2 more protein coding genes). Patient referral reasons included poor weight gain, cleft lip and heart murmur (Table 2.3). Several cases were reported with deletions affecting both *PSMD12* and *BPTF* with features including failure to thrive, developmental delays, dysmorphic features and cardiac and skeletal anomalies (Stankiewicz et al., 2017). Based on the above evidence, this CNL was reclassified as pathogenic (Riggs et al., 2020).

2.3.5.1.8 NFIB

Loss of function pathogenic variants in *NFIB* cause autosomal dominant acquired macrocephaly with impaired intellectual development (MACID, OMIM #618286) and no ClinGen dosage sensitivity score to date (https://dosage.clinicalgenome.org/; accessed March 2021). Three cases (case 9 and 11 are related) were identified in our laboratory between 2013-2016 (Table 2.3). The CNL in cases 9 and 11 affects exons 1-2 while in case 10 deletes the whole gene. Patients had a variety of referral reasons including learning difficulties, developmental delay and dysmorphism. Deletions and truncating variants in the *NFIB* gene were reported in 2018 to cause MACID with features including macrocephaly, developmental

delay, muscular hypotonia and behavioural and psychiatric abnormalities (Schanze et al., 2018). These CNLs was therefore reclassified as pathogenic (Riggs et al., 2020).

2.3.5.1.9 CSNK2A1

Pathogenic variants in *CSNK2A1* cause autosomal dominant Okur-Chung neurodevelopmental syndrome (OCNDS, OMIM #617062) and no ClinGen dosage sensitivity score to date (https://dosage.clinicalgenome.org/; accessed March 2021). Two cases were identified in our laboratory (2012 and 2014) (Table 2.3). Case 12 has a 421kb CNL which affect exons 1-4 of CSNK2A1 (and 14 more protein coding genes). Patient was referred with leaning difficulties, dysmorphism and short stature. Case 13 has a 622kb CNL which results in the loss of the whole gene (and 17 more protein coding genes). Patient was referred with developmental delay with short stature. Pathogenic variants (missense and truncating) in *CSNK2A1* were initially reported in 2016 (Okur et al., 2016) but several more publications followed on (Chiu et al., 2018; Owen et al., 2018; Trinh et al., 2017). Features associated with OCNDS include intellectual disabilities, dysmorphic features and short stature (Okur et al., 2016; Owen et al., 2018). These CNLs was reclassified as pathogenic (Riggs et al., 2020).

2.3.5.2 Other shortlisted genes

Although made it to the final list, three more genes (*NTRK2, KIF2A* and *HECW2*) affecting 6 cases (all four of *NTRK2* cases are related) remain as class 3. Although variants in these genes are linked with relatively related phenotypes; *NTRK2* with autosomal dominant obesity, hyperphagia, and developmental delay syndrome (OBHD), *KIF2A* with autosomal dominant cortical dysplasia, complex, with other brain malformations 3 (CDCBM3) and *HECW2* with autosomal dominant neurodevelopmental disorder with hypotonia, seizures, and absent language (NDHSAL) (Table 2.3), the evidence supporting the pathogenicity of CNLs was not conclusive (Riggs et al., 2020).

Table 2.3 Prioritised Genes and the associated CNLs that have undergone reclassification

Gene	Case	CNV - hg19 (size)	Symptoms	OMIM phenotype (#)	Revised classification
NAA15	Case 1	Chr4:140251306-140278920 (28kb)	Severe Learning Difficulties, Not Dysmorphic, Mum and Brother with Severe Learning Difficulties	MRD50 (#617787)	Likely Pathogenic
ZMYND11	Case 2	chr10:136145-226212 (90kb)	Autistic spectrum disorder, attention deficit hyperactivity disorder, moderate learning difficulties	MRD30 (#616083)	Likely Pathogenic
	Case 3	chr10:136145-912555 (776kb)	Developmental delay, poor language skills, ?Fragile X.		Likely Pathogenic
ACTB	Case 4	chr7:5057992-5617869 (556kb)	Growth parameters 20.4th centile (short stature), small for gestational age, microcornea, ?immune dysfunction, inguinal hernia, abnormality of the kidney, Lacrimal duct aplasia, recurrent infections, intellectual disability	BRWS1 (#243310)	Pathogenic
KMT2C	Case 5	chr7:150017071-152988408 (2,971kb)	Hypotonia, small for Age	KLEFS2 (#617768)	Pathogenic
PHF21A	Case 6	chr11:45936922-47383330 (1,446kb)	Preauricular pit, Abnormality of the musculature, Generalized hypotonia, Abnormality of the nervous system, Delayed gross motor development, Partial agenesis of the corpus callosum, Abnormality of the skeletal system, Relative macrocephaly, Prominent metopic ridge, Hyperextensibility of the finger joints, Tapered finger , Plagiocephaly , Butterfly vertebrae	IDDBCS (#618725)	Likely Pathogenic
MYT1L	Case 7	Chr2:1973125-2659861 (687kb)	Unclear speech, truncal obesity. joint hyperextensibiulity. Decreased muscle tone	MRD39 (#616521)	Likely Pathogenic
PSMD12	Case 8	Chr17:65356977-65884868 (528kb)	Poor weight gain. Cleft lip in utero repaired. Heart murmur. ?22q11 deletion.	STISS #(617516)	Pathogenic
BPTF				NEDDFL (#617755)	
NFIB	Case 9	Chr9:13974415-14286259 (312kb)	Mild learning difficulties, microcephaly	MACID (618286)	Pathogenic
	Case 10	Chr9:14074214-14650748 (578kb)	Global developmental delay. Facial dysmorphism, crowding, high arched palate, Bilateral short fourth metatarsals, bilateral clinodactyly		Pathogenic
	Case 11	Chr9:13974415-14286259 (312kb)	Learning difficulties, developmental delay, frequent chest infections and drooling		Pathogenic
CSNK2A1	Case 12	Chr20:60734-481631 (421kb)	Leaning difficulties, dysmorphism, short stature, mother also has Leaning difficulties and dysmorphism	OCNDS (#617062)	Pathogenic
	Case 13	Chr20:60734-683016 (622kb)	Developmental delay with short stature		Pathogenic
NTRK2	Case 14	Chr9:87353048-88124421 (771kb)	Developmental delay, in particular gross motor, decreased tone.	OBHD (#613886)	Class 3
	Case 15	Chr9:87353048-88124421 (771kb)	Brother has 9q21.33 deletion and similar symptoms - motor delay, behavioural difficulties, mild learning difficulties.		Class 3
	Case 16	Chr9:87353048-88124421 (771kb)	Family history of 9q21.33 deletion in brothers and father		Class 3
	Case 17	Chr9:87353048-88124421 (771kb)	Global developmental delay and feeding difficulties		Class 3
KIF2A	Case 18	Chr5:60905946-61754821 (849kb)	Significant dysparxia(2nd centile).concerns regarding social communication and interaction. overlapping of 2nd and 3rd bilateral toes.	CDCBM3 (#615411)	Class 3
HECW2	Case 19	Chr2:193360118-197153119 (3,793kb)	Moderate/severe learning disability, obesity.	NDHSAL (#617268)	Class 3

2.4 Discussion

Reassessment of pre-existing genomic data is an accepted valid and clinically useful approach. All the previous publications of reassessment of genomic data has focussed on NGS data including exome and genome re-analysis (Bruel et al., 2019; Costain et al., 2018; Das et al., 2014; Hiatt et al., 2018; Sun et al., 2019; Wright et al., 2018). Although CNV analysis has been a first-tier genomic diagnostic test for more than decade in the UK, no such re-analysis has been reported for CNV data generated by aCGH.

Many of the CNVs identified in a diagnostic setting are usually rare and unique (Riggs et al., 2020). Most of diagnostic laboratories will only revisit class 3 variants reactively, for example if this was indicated by a clinician or another similar variant was identified in another patient (El Mecky et al., 2019). Here, we devised an approach to efficiently facilitate systematic re-analysis of class 3 CNLs at a large-scale.

We used an updated OMIM annotations in addition to haploinsufficiency scores to short list class 3 CNLs of interest. Our haploinsufficiency score landscape was akin to previously published data. Our analysis of pathogenic CNLs (after excluding overlap with benign CNLs) have shown that 19.4% of genes within these regions are predicted to be haploinsufficient by pLI scores while only 12.4% in the benign CNLs. This is comparable to the results from an analysis performed on CNLs obtained from the database of structural variants (dbVar) (3,269 pathogenic CNLs and 3,699 benign CNLs), whereby they found that 19.5% of the genes exclusively observed in pathogenic CNV regions to be haploinsufficient (by pLI) compared to 12% in benign regions. RVIS scores have shown similar patterns (21.1% for pathogenic CNLs and 15.9% for benign ones). On the other hand, D-HI scores indicate haploinsufficient genes to be more prevalent in benign CNLs (79%) compared to pathogenic CNLs (55%). D-HI scores were predicted based on an evolutionary and functional similarities to ~300 established haploinsufficient genes (Huang et al., 2010). In contrast, pLI and RVIS scores were calculated for all genes based on comparing observed protein truncating variants with expected ones based on gene size (Lek et al., 2016; Petrovski et al., 2013) . These differences in computing

these scores could account for the better performance of pLI and RVIS in comparison to D-HI. Based on this, we excluded the D-HI scores from further analyses and confirmed that pLI and RVIS were robust scores for predicting haploinsufficient genes in our cohort. Further comparison of haplosinsufficiency scores has also shown enrichment of haploinsufficient genes in class 3 CNLs compared to benign ones. This indicated the increased likelihood of the presence of pathogenic deletions amongst the variants of uncertain significance (Class 3 CNLs).

We then followed a systematic filtering and shortlisting process (Figure 2.1) that resulted in a short list of 123 genes. These genes were interrogated by trained clinical scientists reviewing associated phenotypes, disease mechanism and array calls. This resulted in a final short list of 13 genes (19 CNLs) that warranted full reinterpretation according to the latest guidance (Riggs et al., 2020). This resulted in 13 CNLs (10 genes) reclassifications to (likely)pathogenic. All the reclassifications were due to genes and disease association described post the identification and reporting of the class 3 CNLs, apart from *ACTB,* whereby disease association was known prior to the class 3 CNL reporting, however, the loss of function disease mechanism was not confirmed until 2017 (post class 3 CNL reporting). The final reclassification rate was ~0.6% of the total number of class 3 CNLs identified at the start of this study but importantly ~6.4% of the shortlisted 204 CNLs.

The reclassification rate is lower than reported rates of reclassification from most NGS based studies (multi-gene panels, WES and WGS) (Bruel et al., 2019; Costain et al., 2018; Das et al., 2014; Hiatt et al., 2018; Sun et al., 2019; Wright et al., 2018). However, there are no similar studies from aCGH data for comparison. This lower rate could be attributed to the more targeted nature of aCGH compared to exome and genome analysis. Most aCGH designs include backbone coverage and targeted coverage for mainly known disease-causing regions/genes, such as the ISCA (International Standards for Cytogenomic Arrays) based arrays used in our department. This targeted nature may be responsible for the lower reclassification rate as most targeted genes have been included due to known phenotype association.

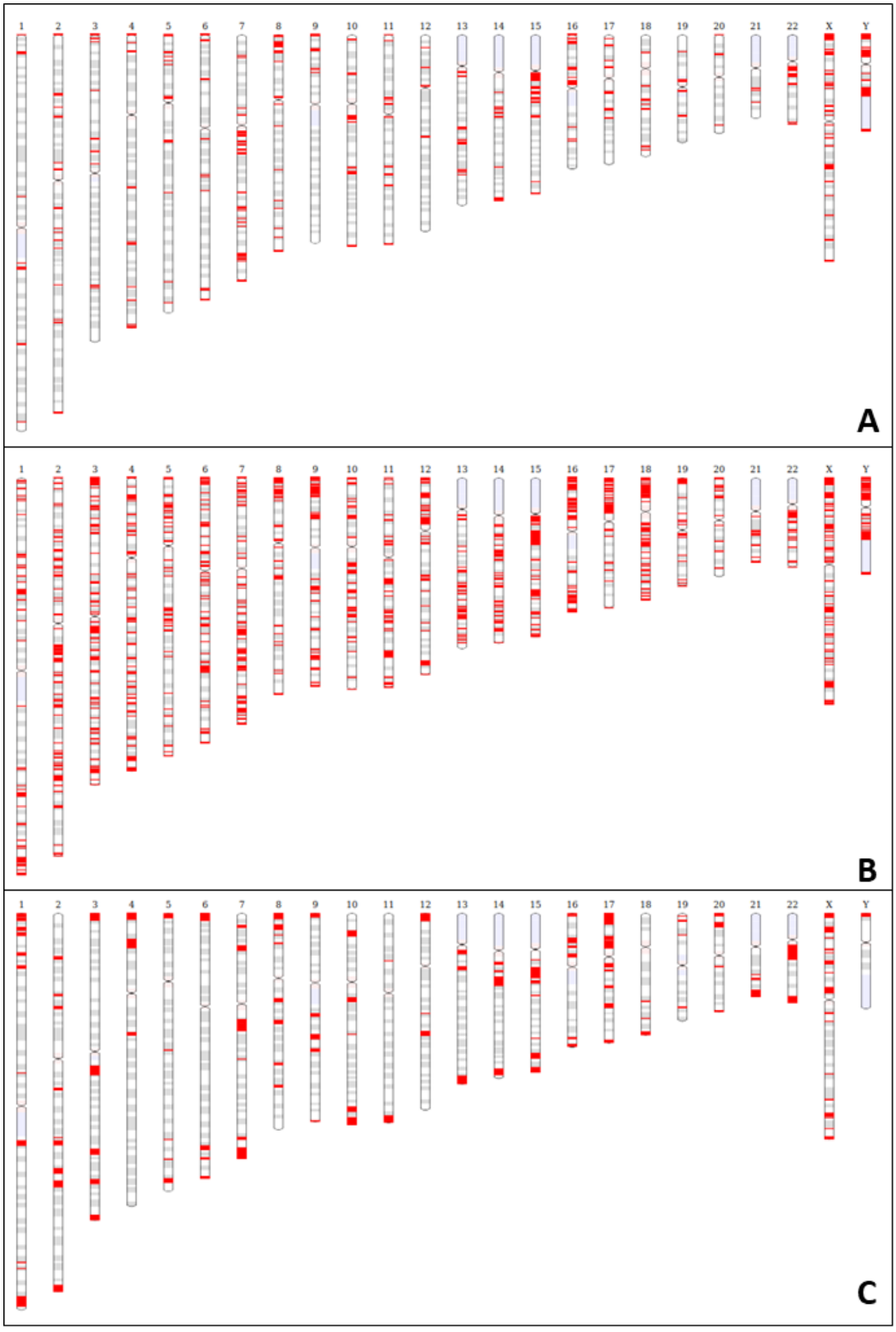
Apart from the limitation of the aCGH design, this study has several limitations that may impact reclassification rates. The shortlisting of genes was mainly reliant on OMIM phenotype associations and haploinsufficient score predictions. The OMIM phenotype associations could miss some interesting genes with recent publications compared to literature searches, however, the latter is not suitable for automating the process. Haploinsufficient score predictions were used as proxy for haploinsufficiency and although are reasonably predictive they are not accurate for all genes. An example of which is discussed in Chapter 3. The *ATP6V0C* gene within the class 3 CNL described in chapter 3 has been excluded in this analysis due to the stringent haploinsufficiency filtering criteria. Nevertheless, our detailed analysis of the reclassified CNL strongly suggested that haploinsufficiency of *ATP6V0C* is likely to underlie the pathogenicity of the variant and the associated 16p13.3 deletion syndrome (Chapter 3). Moreover, we haven't attempted the identification of contiguous gene disorders nor reclassification of class 3 CNLs that may disrupt regulatory regions and TADs. We also did not attempt the reclassification of class 3 CNVs into benign categories using updated population databases. Finally, we have created an equivalent list of genes within the 3 categories of our CNGs with similar annotations as for CNLs. We will perform a reclassification of class 3 CNGs, mainly focussed on up-to-date disease associations. All of these extra analyses could potentially increase the reclassification rate.

In conclusion, we have shown that a systematic bulk reanalysis of Class 3 CNLs identified in a diagnostic laboratory provided new diagnoses to some patients.

2.5 Contribution statement

The study was conceived by Dr Siddharth Banka; the research plan was devised by myself and Dr Banka; data collection was done by me; data cleaning was done by me; data analysis was done by me and Dr Jamie Ellingford (bioinformatics analysis); manual review of CNVs was done by me, Jake Miller, Jonathan Edgerley and Ronnie Wright; and the paper was drafted by me and reviewed by Dr Banka.

2.6 Chapter 2 Supplementary Material



Supplementary Figure 2.3 Ideograms of CNLs

A) Class 1 and 2 CNLs, B) Class 3 CNLs, C) Class 4 and 5 CNLs

Supplementary Table 2.4 Haploinsufficiency scores and genes according to CNL category

	Number of genes with available scores			Predicted Haploinsufficient genes (%)		
CNL class (Total no of genes)	**pLI**	**RVIS**	**Decipher HI**	**pLI (%)**	**RVIS (%)**	**Decipher HI (%)**
Pathogenic (2,611)	2,396	2,338	2,554	464 (19.37)	493 (21.09)	1,403 (54.93)
Class 3 (2,543)	2,287	2,300	2,481	430 (18.80)	559 (24.30)	1,257 (50.67)
Benign (433)	274	251	409	34 (12.41)	40 (15.94)	322 (78.73)

Supplementary Table 2.5: List of the prioritised 123 genes that have undergone scientist review

Gene	OMIM ID (*)	Number of cases	pLI ≥0.9 (Yes/No)	RVIS ≤ 0.01 (Yes/No)	DDG2P consequence	OMIM Phenotype # (Inheritance pattern)
NAA15	608000	9[$]	Yes	Yes	LoF	617787 (AD)
MBD5	611472	6	N/A	Yes	LoF	156200 (AD)
PIEZO2	613629	4	No	Yes	LoF	114300 (AD);108145(AD); 617146 (AR), 248700 (AD)
TP63	603273	3	Yes	Yes	LoF; Uncertain	103285 (AD); 604292 (AD); 106260 (AD); 603543 (AD); 129400 (AD); 129400 (AD); 605289 (AD)
SIX1	601205	3	Yes	No	LoF; all missense/in frame	608389 (AD); 605192 (AD)
ZIC2	603073	3	N/A	N/A	LoF	609637 (AD)
FOXG1	164874	3	N/A	N/A	LoF	613454 (AD)
FREM1	608944	3	N/A	N/A	LoF	608980; 248450 (AR); 614485 (AD)
NF1	613113	3	Yes	Yes	LoF	607785 (AD); 162210 (AD); 162200 (AD); 601321 (AD); 193520 (AD)
ZMYND11	608668	3	Yes	Yes	LoF	616083 (AD)
TBX1	602054	3	Yes	No	LoF	217095 ; 188400 (AD); 187500 (AD); 192430 (AD)
SLC2A2	138160	3	No	No	LoF	125853 (AD); 227810 (AR)
MITF	156845	2	Yes	Yes	LoF; Uncertain	617306 (AR); 614456; 103500 (AD); 193510 (AD); 103470 (AD)
CDH23	605516	2	No	Yes	LoF; all missense/in frame	601386 (AR); 617540 (AD); 601067 (AR/DR); 601067 (AR/DR)
MYH8	160741	2	No	No	LoF; all missense/in frame	608837; 158300 (AD)
RERE	605226	2	Yes	Yes	LoF	616975 (AD)
AUTS2	607270	2	Yes	Yes	LoF	615834 (AD)
SATB2	608148	2	Yes	Yes	LoF	612313 (AD)
ERCC6	609413	2	No	Yes	LoF	214150 (AR);133540 (AR); 278800 (AR); 211980 (AR); 613761 ; 616946 (AD); 600630 (AR)
FOXP2	605317	2	N/A	Yes	LoF	602081 (AD)
ATR	601215	2	No	Yes	LoF	614564 (AD); 210600 (AR)
ABCC6	603234	2	No	Yes	LoF	614473 (AR); 264800 (AR); 177850 (AD)
MYH11	160745	2	Yes	Yes	LoF	132900 (AD)
ANKRD11	611192	2	Yes	Yes	LoF	148050 (AD)
PIEZO1	611184	2	No	No	LoF	194380 (AD); 616843 (AR)
GNAS	139320	2	Yes	Yes	Activating; LoF	219080; 174800; 166350 (AD); 617686; 103580 (AD); 603233 (AD); 612462 (AD); 612463 (AD)
MYH9	160775	1	Yes	Yes	LoF; Uncertain	603622 (AD); 153650 (AD); 153640 (AD); 600208 (AD); 155100 (AD); 605249 (AD)
CDKN1C	600856	1	No	No	LoF; GoF	130650 (AD); 614732 (AD)
ITPR1	147265	1	Yes	Yes	LoF; DN; LoF; All missense/in frame	206700 (AR, AD); 606658 (AD); 117360 (AD)

GJA1	121014	1	No	No	LoF; All missense/in frame; Uncertain	600309 (AD); 218400 (AR); 617525 (AD), 241550 (AR),164200 (AD); 257850 (AR); 104100 (AD); 186100 (AD)
KIF1A	601255	1	N/A	Yes	LoF; All missense/in frame	614255 (AD); 614213 (AR); 610357 (AR)
PLCG2	600220	1	Yes	Yes	LoF; All missense/in frame	614878 (AD); 614468 (AD)
ZFPM2	603693	1	Yes	Yes	LoF; All missense/in frame	610187; 187500 (AD); 616067 (AD)
TBC1D24	613577	1	No	No	LoF; All missense/in frame	220500 (AR); 614617 (AR); 616044 (AD); 615338 (AR); 605021 (AR)
ACTB	102630	1	Yes	No	LoF; All missense/in frame	243310 (AD); 607371 (AD)
KMT2C	606833	1	Yes	N/A	LoF	617768 (AD)
EHMT1	607001	1	Yes	Yes	LoF	610253 (AD)
CTNNB1	116806	1	N/A	Yes	LoF	114500;617572 (AD); 114550; 155255; 615075 (AD); 167000; 132600
TCF4	602272	1	Yes	Yes	LoF	613267 (AD); 610954 (AD)
DSTYK	612666	1	No	Yes	LoF	610805 (AD); 270750 (AR)
TGFB2	190220	1	Yes	Yes	LoF	614816 (AD)
LRP4	604270	1	No	Yes	LoF	212780 (AR); 616304 (AR); 614305 (AR/AD)
MYT1L	613084	1	Yes	Yes	LoF	616521 (AD)
PARN	604212	1	No	Yes	LoF	616353 (AR); 616371 (AD)
POU1F1	173110	1	No	Yes	LoF	613038 (AR/AD)
DSG1	125670	1	Yes	Yes	LoF	615508 (AR); 148700 (AD)
IGF1R	147370	1	No	Yes	LoF	270450 (AR/AD)
TTN	188840	1	No	Yes	LoF	604145; 613765 (AD); 608807 (AR); 603689; 611705 (AR); 600334 (AD)
RYR1	180901	1	No	Yes	LoF	117000 (AR/AD); 145600 (AD); 145600 (AD); 255320 (AR); 117000 (AR/AD)
PSMD12	604450	1	Yes	Yes	LoF	617516 (AD)
BPTF	601819	1	Yes	Yes	LoF	617755 (AD)
SYNGAP1	603384	1	Yes	Yes	LoF	612621 (AD)
LTBP3	602090	1	Yes	Yes	LoF	601216 (AR); 617809 (AD)
CHD2	602119	1	Yes	Yes	LoF	615369 (AD)
RELN	600514	1	Yes	Yes	LoF	616436 (AD); 257320 (AR)
EPHB4	600011	1	Yes	Yes	LoF	617300 (AD)
LEMD3	607844	1	Yes	Yes	LoF	166700 (AD); 166700 (AD)
SOX5	604975	1	Yes	Yes	LoF	616803 (AD)
EXT2	608210	1	No	No	LoF	133701 (AD); 616682 (AR)
TWIST1	601622	1	N/A	No	LoF	123100 (AD); 180750 (AD); 101400 (AD); 617746 (AD)
MEF2C	600662	1	N/A	No	LoF	613443 (AD); 613443 (AD)

EVC	604831	1	No	No	LoF	225500 (AR); 193530 (AD)
EVC2	607261	1	No	No	LoF	225500 (AR); 193530 (AD)
TGIF1	602630	1	N/A	No	LoF	142946 (AD)
LMX1B	602575	1	No	No	LoF	161200 (AD)
CRB1	604210	1	No	No	LoF	613835;172870 (AD); 600105 (AR)
ATP8B1	602397	1	No	No	LoF	243300 (AR); 147480 (AD); 211600 (AR)
NR2F2	107773	1	N/A	No	LoF	615779 (AD)
CD96	606037	1	N/A	No	LoF	211750 (AD)
TGFBR2	190182	1	No	No	LoF	614331; 133239; 610168 (AD)
TNFRSF13B	604907	1	No	No	LoF	240500 (AR/AD); 609529
GATA4	600576	1	No	No	LoF	607941 (AD); 614430 (AD);615542 (AD);187500 (AD); 614429 (AD)
SMAD6	602931	1	No	No	LoF	614823 (AD); 617439 (AD)
SMAD3	603109	1	No	No	LoF	613795 (AD)
SIX3	603714	1	No	No	LoF	157170 (AD); 269160
PTHLH	168470	1	Yes	No	Increased gene dosage; LoF	613382 (AD)
NRXN2	600566	1	Yes	Yes	LoF	N/A
NFIB	600728	3	Yes	Yes	LoF	N/A
CTNNA2	114025	1	Yes	Yes	LoF	N/A
PHF21A	608325	1	Yes	Yes	LoF	N/A
OTUD7A	612024	8	Yes	No	LoF	N/A
NRXN3	600567	2	Yes	Yes	LoF	N/A
RYR3	180903	1	Yes	Yes	LoF	N/A
CACNA1C	114205	7	Yes	Yes	Activating	611875; 601005 (AD)
NR3C2	600983	4	Yes	Yes	N/A	605115 (AD); 177735 (AD)
NTRK2	600456	3	Yes	Yes	All missense/in frame	617830 (AD); 613886 (AD)
RANBP2	601181	3	Yes	Yes	All missense/in frame	608033 (AD)
ERBB4	600543	3	Yes	Yes	N/A	615515 (AD)
ABL1	189980	2	Yes	Yes	Activating	617602 (AD)
CSNK2A1	115440	2	Yes	Yes	Activating	617062 (AD)
RYR2	180902	2	Yes	Yes	N/A	600996 (AD); 604772 (AD)
PRKAG2	602743	2	Yes	Yes	N/A	600858 (AD); 261740 (AD); 194200 (?AD)
BMPR2	600799	2	Yes	Yes	N/A	178600 (AD);178600 (AD); 265450 (AD)
CFH	134370	2	Yes	Yes	N/A	126700 (AD); 609814 (AR/AD); 235400 (AR/AD); 610698

CACNA1B	601012	2	Yes	Yes	N/A	614860 (AD)
AKAP10	604694	2	Yes	Yes	N/A	115080 (AD)
COL4A1	120130	1	Yes	Yes	DN	611773 (AD); 607595 (AD); 614519; 175780 (AD);180000 (AD);269160
KIF2A	602591	1	Yes	Yes	DN	615411 (AD)
COL11A2	120290	1	Yes	Yes	All missense/in frame; DN	601868 (AD); 609706 (AR);614524 (AR)/AD); 184840 (AD); 215150 (AR)
SMARCA2	600014	1	Yes	Yes	All missense/in frame	601358 (AD)
HECW2	617245	1	Yes	Yes	All missense/in frame	617268 (AD)
NEDD4L	606384	1	Yes	Yes	All missense/in frame	617201 (AD)
RBPJ	147183	1	Yes	Yes	All missense/in frame	614814 (AD)
KIF5C	604593	1	Yes	Yes	All missense/in frame	615282 (AD)
DYNC1H1	600112	1	Yes	Yes	All missense/in frame	614228 (AD);614563 (AD); 158600 (AD)
AKT3	611223	1	Yes	Yes	All missense/in frame	615937 (AD)
PACS1	607492	1	Yes	Yes	Activating	615009 (AD)
IGF2BP2	608289	1	Yes	Yes	N/A	125853 (AD)
ADAM10	602192	1	Yes	Yes	N/A	615590; 615537 (AD)
F2	176930	1	Yes	Yes	N/A	613679 (AR); 613679 (AR); 614390 (AD); 601367 ; 188050 (AD)
DPP6	126141	1	Yes	Yes	N/A	616311; 612956 (AD)
KCNH2	152427	1	Yes	Yes	N/A	613688 (AD); 613688 (AD); 609620
DNM2	602378	1	Yes	Yes	N/A	160150 (AD); 606482 (AD); 606482 (AD); 615368 (AR)
ETV6	600618	1	Yes	Yes	N/A	601626; 616216 (AD)
ACTN4	604638	1	Yes	Yes	N/A	603278 (AD)
SLITRK1	609678	1	Yes	Yes	N/A	137580 (AD), 613229 (AD)
LMNB2	150341	1	Yes	Yes	N/A	616540 (AR), 608709 (AD)
STK11	602216	1	Yes	Yes	N/A	175200 (AD); 273300
MEN1	613733	1	Yes	Yes	N/A	131100 (AD)
JPH3	605268	1	Yes	Yes	N/A	606438 (AD)
SNRNP200	601664	1	Yes	Yes	N/A	610359 (AD)
FLCN	607273	1	Yes	Yes	N/A	135150 (AD); 114500; 173600 (AD); 144700
WNK1	605232	1	Yes	Yes	N/A	201300 (AR); 614492 (AD)

Above is a list of 123 genes that were prioritised for scientist review (Please refer to Figure 2.1)
Genes in **blue bold** font are those that were selected to undergo full CNV reclassification (Table 2.3)
[$] 8 out of the 9 cases with *NAA15* loss were deemed artefactual on aCGH data inspection
LoF: Loss of Function, GoF: Go of Function, AD: Autosomal Dominant, AR: Autosomal Recessive, DR: Digenic Recessive

3 HAPLOINSUFFICIENCY OF *ATP6V0C* POSSIBLY UNDERLIES 16P13.3 DELETIONS THAT CAUSE MICROCEPHALY, SEIZURES AND NEURODEVELOPMENTAL DISORDER[1]

[1] This chapter is adapted from Tinker RJ, **Burghel GJ**, Garg S, Steggall M, Cuvertino S, Banka S. Haploinsufficiency of *ATP6V0C* possibly underlies 16p13.3 deletions that cause microcephaly, seizures, and neurodevelopmental disorder. Am J Med Genet A. 2020 Epub ahead of print.

3.1 Abstract

Introduction: We recently described eight individuals with a novel condition caused by 16p13.3 microdeletions encompassing *TBC1D24*, *ATP6V0C*, and *PDPK1* and resulting in epilepsy, microcephaly and neurodevelopmental problems. The phenotypic spectrum, the minimum overlapping region and the underlying disease mechanism for this disorder remain to be clarified.

Results: Here we report a 3.5-year-old male, with microcephaly, autism spectrum disorder and a *de novo* 16p13.3 microdeletion. We performed detailed *in silico* analysis to show that the minimum overlapping region for the condition is ~80Kb encompassing five protein coding genes. Analysis of loss of function constraint metrics, transcript-aware evaluation of the population variants, GeVIR scores, analysis of reported pathogenic point variants, detailed review of the known functions of gene products and their animal models showed that the haploinsufficiency of *ATP6V0C* likely underlies the phenotype of this condition. Protein-protein interaction network, gene phenology and analysis of topologically associating domain showed that it was unlikely that the disorder has an epistatic or regulatory basis.

Conclusions: 16p13.3 deletions encompassing *ATP6V0C* cause a neurodevelopmental disorder. Our results broaden the phenotypic spectrum of this disorder and clarify the likely underlying disease mechanism for the condition.

3.2 Introduction

Copy number variants (CNVs) are an important cause for neurodevelopmental disorders (Coe et al., 2014; Cooper et al., 2011; Mefford, 2014). Disease causing CNVs result in diverse phenotypes and have significant medical and socioeconomic impact (Burghel et al., 2020; Coppola et al., 2019). Variability of CNVs can make their clinical correlation challenging. Individuals with overlapping CNVs can help in determining the minimum overlapping regions for specific disorders (Kasher et al., 2016). A significant proportion of copy number losses are driven by haploinsufficiency of dosage sensitive genes (Rice and McLysaght, 2017). Studying CNVs can also provide insights into genetic basis of single gene (Cuvertino et al., 2017; Yagi et al., 2003) or complex disorders (Banka et al., 2015). Other mechanisms by which chromosome deletions can convey phenotypes include contiguous gene deletions, gene interruption, generation of fusion genes, unmasking of deleterious variants resulting in recessive phenotypes, epistasis, position effects such as disruption of the regulatory regions of the genome and disruption of topologically associating domains (TADs) (Lupski and Stankiewicz, 2005; Spielmann et al., 2018; Stankiewicz and Lupski, 2002).

We recently described 8 individuals with epilepsy, microcephaly, neurodevelopmental problems and overlapping 205 kb to 504 kb 16p13.3 microdeletions and proposed that it is a novel genetic condition (Table 3.1) (Mucha et al., 2019). The defined minimum critical region of ~112 kb (GRCh37:Chr16:2,530,000–2,642,000) that included seven known genes - *TBC1D24, ATP6V0C, AMDHD2, CEMP1, MIR3178, PDPK1* and *DQ577714*. Since our publication, there has been no other peer-reviewed publication on this condition. The phenotypic spectrum, the minimum overlapping region and the underlying disease mechanism for this disorder remain to be clarified. Here we report an additional individual with an overlapping 16p13.3 deletion and a similar phenotype, which further establishes this microdeletion syndrome as a distinct entity. We also present *in silico* work to understand the underlying disease mechanism.

Table 3.1 Summary of phenotype of patients with 16p13.3 deletions

Feature	Current Patient (1)	Previously reported patients (8)
Sex	Male	Females (2), Males (6)
Age	2	Mean: 13.8
Neurodevelopment and behaviour	Moderate autistic spectrum disorder	Developmental delay (6) Attention deficit hyperactivity disorder (3) bipolar disorder (1)
Seizures	Febrile tonic clonic and absence seizures	Generalized tonic–clonic in (6)
Microcephaly	Yes (0.4th centile)	6 (ranging from 0.4th - 5th centile)
Additionally features (N/8)	Recurrent infections	Hypotonia (2) Short stature (2) Strabismus (2) Beaked nose (1) Failure to thrive (1) Hearing loss (1) Hypotonia (1) Nystagmus (1) Pointed chin (1) Posteriorly rotated ears (1) Prognathism (1) Tapering fingers (1) Strabismus (1) Tubular nose (1) Vision loss (1) Insomnia (1/8)
Brain Imagining	Normal	Cerebral & cerebellar atrophy (1) Small stable venous anomaly (1) Thickening of calvarium (1)

3.3 Methods and Results

3.3.1 Case Report

The proband is a 3.5-year-old male, born to non-consanguineous parents with no relevant family history. His mother's pregnancy was complicated by antenatal bleeding and placenta praevia. He was born via Caesarean section at 38 weeks gestation with a birth weight of 2.41kg (-2SD). The immediate postnatal period was uneventful apart from a brief period of hypothermia.

At the age of 10 months, the child presented with a tonic-clonic seizure during a febrile illness. Subsequently he had multiple febrile tonic-clonic seizures. 24-hour electroencephalograms did not show any epileptic activity. A brain magnetic resonance imaging scan at 1 year did not identify any abnormality. He has not been formally diagnosed with epilepsy and no anti-epileptic treatment has been required to be initiated so far. He first rolled over at the age of 10 months but then did not roll over again until several months later. He attained independent walking at 18 months of age. An Early Social cognitive Battery assessment performed at the age of 38 months showed his early social interaction and play skills to be impaired. An Autism Diagnostic Observation Schedule (ADOS Module 2) performed at 38 months of age showed him to be on moderate autism spectrum. Social affect score was 8 and restricted repetitive behaviours score of 1 with calibrated severity score of 5, meeting the criteria for autism spectrum disorder. Weschsler Preschool and Primary Scale of Intelligence – Fourth UK edition (WPPSI-IV) performed at 36 months of age gave a Full Scale Intelligence Quotient of 90 within the average range (25th centile). His verbal comprehension index was 91 on the 27th centile; (scaled scores: Receptive vocabulary = 10; Information = 7) and Visual spatial index was 97 on the 42nd centile (scaled scores: Block Design = 10; Object Assembly = 9). Working memory and vocabulary index could not be computed due to lack of compliance but he scored within low average range for Picture memory (scaled score of 7) and Picture Naming (scaled score 7).

The child also has a history of frequent infections including recurrent upper respiratory tract infections, several episodes of glue ear and one episode of enteroviral meningitis. He also required tonsillectomy.

At age 2 years, the child's occipital frontal head circumference was 45 cm (-3SD), weight was 10.2 kg (-1.8SD) and height was 84.6 cm (-0.25SD). A neurological examination at this age was unremarkable. Facially the child is not dysmorphic.

3.3.2 Array comparative genomic hybridisation

Array comparative genomic hybridisation (a-CGH) analysis using OGT CytoSure Constitutional v3 Array (6x60K) revealed a 220kb loss of chromosome 16p13.3 ([GRCh37] 16p13.3(2,415,389_2,635,921)x1) (Figure 3.1A).The deletion encompassed eight known protein coding genes (*CCNF*, *C16orf59*, *NTN3*, *TBC1D24, ATP6V0C, AMDHD2, CEMP1 and PDPK1)* and two non-protein coding genes (*MIR3178* and *DQ577714).*Targeted a-CGH performed on DNA samples of both parents revealed normal results indicating that the proband's deletion had occurred *de novo*. Comparison of the clinical features of our individual, with the phenotypes of previously published individuals (Table 3.1) showed remarkable overlap (Mucha et al., 2019). We, therefore, classed this deletion as pathogenic based on the joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen) (Riggs et al., 2020).

3.3.3 *In silico* analysis and literature review

Combining our data with the previously published individuals revealed the minimum overlapping region for the condition to be ~80Kb (GRCh37:chr16:2,555,682-2,635,921) encompassing five protein coding (*TBC1D24, ATP6V0C, AMDHD2, CEMP1* and *PDPK1)* and two non-protein coding genes (*MIR3178* and *DQ577714)* (Figure 3.1A).

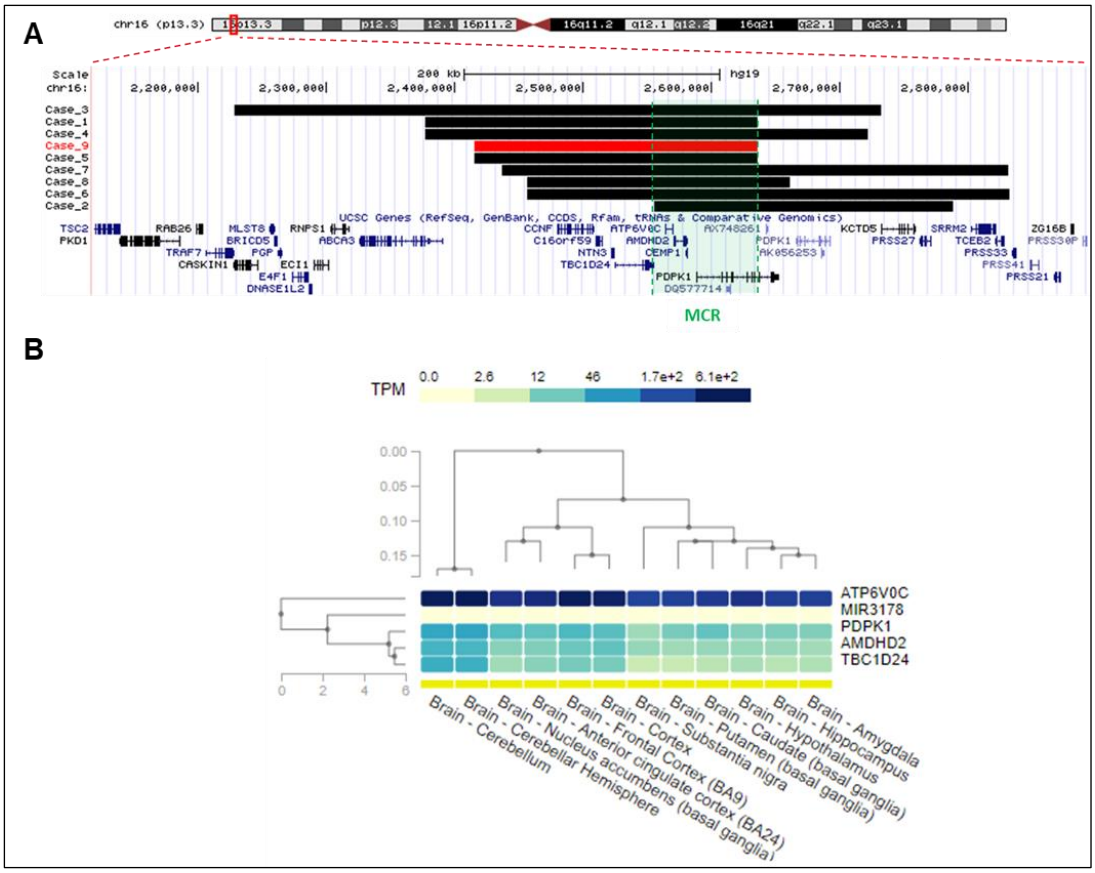


Figure 3.1 The genetic basis of 16p13.3 deletion syndrome

A) Schematic representation of the reported 16p13.3 microdeletions. Cases 1 to 8 are previously reported. Case 9 is the present case. B) Quantified expression of 5 of the deleted genes in the central nervous system (Transcripts Per Million (TPM) with dendrograms representing different groups of central nervous system tissue.

To investigate the possible disease mechanism, we first compiled the pLI scores of the seven genes in the deleted region (Table 3.2). *PDPK1* and *ATP6V0C* were found to have the highest pLI scores of 0.98 and 0.74 respectively. Closer inspection of the gnomAD data revealed that the low pLI score of *ATP6V0C* is, at least in part, likely attributable to the relatively small size of this gene (only 3 exons encoding a protein of 155 amino acids). Furthermore, the only observed loss of function variant is located in part of its exon 2 that is not included in two out of four known transcripts of the gene (Supplementary Figure 3.2).

Table 3.2: Summary of the pLI and GeVIR scores of the genes within 16p13.3 microdeletions

Gene	pLI Score	GeVir scores								
		GeVIR%	LOEUF%	VIRLoF%	GeVIR AD	LOEUF% AD	VIRLoF AD	GeVIR AR	LOEUF AR	VIRLoF AR
TBC1D24	0	25.12	61.99	41.78	1.37	0.56	0.72	1.15	1.36	2.01
ATP6VOC	0.74	2.03	34.48	16.03	4.23	0.82	1.71	0.24	1.56	0.68
AMDHD2	0	71.28	38.87	54.66	0.39	0.7	0.56	0.85	1.65	1.73
CEMP1	Not Known	97.8	98.21	99.54	0.11	0.2	0.12	0.11	0.15	0.05
MIR3178	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known
PDPK1	0.98	21.16	13.21	15	1.53	2.06	1.89	1.02	0.6	0.63
DQ577714	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known	Not Known

Additionally, we compiled loss-of-function observed/expected upper bound fraction scores (LOEUF) for the five deleted protein coding genes from gnomAD v2.1.1 (https://gnomad.broadinstitute.org/) (Karczewski et al., 2020). Scores for non-coding genes and *CEMP1* were not available. The results were in keeping with the previous pLI scores with both *ATP6V0C* and *PDPK1* showing sensitivity to haploinsufficiency with scores of 0.67 and 0.29 respectively. Next, we obtained Gene Variation Intolerance Ranking (GeVIR) scores for the genes in this region (Table 3.2) (Abramovs et al., 2020). GeVIR is also able to prioritize short genes, for which loss of function constraint cannot be estimated with confidence. *ATP6V0C* was found to be the highest ranked autosomal dominant disorder gene in GeVIR, followed by *PDPK1*. Next, we quantified the levels of gene expression in selected brain regions using GTEX portal (Figure 3.1B) (Supplementary Table 3.3) (Supplementary Figure 3.2) (The GTex Consortium, 2015). Expression of *ATP6V0C* was found to be highest across multiple brain tissues. Our analyses thus far indicated that the haploinsufficiency of *PDPK1* or *ATP6V0C* or both possibly underlies 16p13.3 deletions that cause microcephaly, seizures and neurodevelopmental problems.

Next, we investigated the known functions, animal models and pathogenic point variants in these two genes in more detail. *PDPK1* (also known as *PDK1*) encodes a protein kinase that is involved in intracellular signalling pathway facilitating cell migration (Bergfeld et al., 2012). KO of PDPK1 in murine is incompatible with life but mice with reduced residual PDK1 activity are smaller but fertile and without any obvious neurological phenotype (Alvarez-Pérez et al., 2006). No *PDPK1* point variants were detected in ClinVar (Landrum et al., 2016). Three possible disease causing missense *PDPK1* variants were listed in Human Gene Mutation Database (HGMD) (Stenson et al., 2003). However, the associated phenotypes were congenital heart defect (Russell et al., 2019), premature ovarian insufficiency and modification of phenotype in Williams syndrome (Masson et al., 2019). None of these phenotypes overlap significantly with what has been observed in individuals with 16p13.3 deletions. *ATP6V0C* (ATPase, H+ transporting, lysosomal 16 kDa, V0 subunit C) encodes for a component of vacuolar ATPase (V-ATPase), a subunit enzyme that acidifies the intracellular organelles of

eukaryotes (Bayascas et al., 2008). This process maintains endocytic and exocytic pathways (Bayascas et al., 2008). The subunit is present in synaptic vesicles, endosomes, lysosomes, clathrin-coated vesicles, and the Golgi complex (Abbas et al., 2020). ATP6V0C has a neuron-specific expression at presynaptic vesicles in zebrafish embryos and facilities neurotransmitter storage (Wullschleger et al., 2011). ClinVar lists a single missense *ATP6V0C* variant of uncertain significance without any associated clinical data. HGMD listed one *de novo* frameshift *ATP6V0C* variant in an individual with Dravet syndrome (Carvill et al., 2014). These analyses indicated that haploinsufficiency of *ATP6V0C* is more likely to explain the phenotype caused by 16p13.3 deletions.

Next we considered the possibility of an epistatic model to explain the phenotype of 16p13.3 deletions (Veltman and Brunner, 2010). We first looked for known interactions between the protein products of these genes. Protein-protein interaction network analysis was conducted using STRING and the gene phenology was studied using PANTHER (Mi et al., 2019). These analyses did not reveal any known interactions or obvious functional overlaps (Supplementary Table 3.4- Supplementary Table 3.5). Finally, we looked at the possibility of this deletion disrupting any TAD boundaries. Analysis of the deleted region by ClinTAD (https://www.clintad.com/, accessed April 2020) (Lupiáñez et al., 2016) showed that the breakpoints did not overlap with known TAD boundaries.

3.4 Discussion

We present a boy with microcephaly and autistic spectrum disorder (ASD) with a *de novo* chromosome 16p13.3 deletion, which spans between 2,415,389 and 2,635,921 [GRCh37]. The similarity of the phenotype of our individual with our eight recently described individuals and overlap of this deletion with the previously defined minimum critical region are strongly indicative that this deletion is causal (Mucha et al., 2019). Our observations further establish chromosome 16p13.3 deletion syndrome microcephaly, seizures and neurodevelopmental problems as a distinct genetic disorder. Of note, this condition should not be confused with other non-overlapping 16p13.3 deletions in this region resulting in alpha-thalassemia-

intellectual disability syndrome (Babbs et al., 2020) or Rubinstein-Taybi syndrome (Breuning et al., 1993; Hennekam et al., 1993). Epilepsy (including tonic clonic and myoclonic seizures) and microcephaly are the most frequent features in all individuals with this syndrome (Mucha et al., 2019). Notably, although our individual had history of febrile seizures, he has not been diagnosed as having epilepsy. The individual described in this study was diagnosed with ASD. Reported neuropsychiatric developmental disorders in the disorder include developmental delay (DD), Attention deficit hyperactivity disorder (ADHD), Intellectual disability (ID) and bipolar disorder (Mucha et al., 2019). Of note, the assessments in our patient suggest social communication impairments, but with average range cognitive ability. Additionally, the individual in the current individual study and individual 5 of the previous study have an identical deletion but sustainable phenotypic variation (Mucha et al., 2019). Individual 5 of the previous study has facial dysmorphia, short stature and a venous anomaly on magnetic resonance imaging. This is in contrast with the current patient who has none of these features (Mucha et al., 2019). This adds further evidence to a phenotypic spectrum the current syndrome of *ATP6V0C* deletions.

Our patient had history of repeated infections. None of the previous individuals have been reported to with this phenotype. At present it is not possible to be certain if this observation is linked to the 16p13.3 deletion or if it is coincidental. Of interest, PDPK1 has been reported to regulate macrophages metabolism, T cells differentiation and B cells haemostasis (Yang et al., 2016).

Previously the phenotype of this condition was attributed to haploinsufficiencies of *TBC1D24*, *ATP6V0C* and *PDPK1*. Our *in-silico* analyses indicated that haploinsufficiency of either *ATP6V0C* or *PDPK1* or both the genes may underlie the microcephaly, seizures and neurodevelopmental disorders in individuals with these 16p13.3 deletions. Of the two genes haploinsufficiency of *ATP6V0C* provides a more compelling explanation for the phenotype of this condition. It is because of its known function in acidification of synaptic vesicles (Bayascas et al., 2008), its high level of expression in brain tissues (Fig 1B), high GeVIR ranking (Table 3.2) and previous description of an individual with a frameshift variant and epilepsy (Carvill et

al., 2014) and a frameshift mutation in *ATP6V0C* resulting in neurodevelopmental disability in the Deciphering Developmental Disorders study (McRae et al., 2017)

Furthermore on assessment we assessed for *de novo* mutations using the denovo-db database (https://denovo-db.gs.washington.edu/denovo-db/) and identified a reported de-novo frameshift mutation in *ATP6V0C* that has been shown to in the Deciphering Developmental Disorders study to results in a neurodevelopmental disorder. Of note, the pLI score of *ATP6V0C* is only 0.74, which is lower than expected for a gene, loss of function variants in which may cause a severe early-onset developmental disorder. However, our transcript-aware evaluation of population variants demonstrates that the deletions or loss of function variants that affect all the transcripts of this gene may be more deleterious than those variants which affect only some transcripts. *TBC1D24* encodes a member of the Tre2-Bub2-Cdc16 (TBC) domain-containing RAB-specific GTPase-activating protein family (Falace et al., 2010). Bi-allelic *TBC1D24* loss-of-function variants cause Deafness (OMIM 614617) (Rehman et al., 2014), DOORS syndrome (OMIM 220500) (Campeau et al., 2014), Rolandic epilepsy with paroxysmal exercise-induced dystonia and writer's cramp (OMIM 608105) (Lüthy et al., 2019), Early Infantile Epileptic encephalopathy (OMIM 615338) (Duru et al., 2010) and infantile myoclonic epilepsy (OMIM 605021) (Falace et al., 2010). Heterozygous carriers of *TBC1D24* loss-of-function variants do not show clinical features such as epilepsy, microcephaly or developmental delay. Of note, a heterozygous missense p.(S178L) *TBC1D24* variant has been proposed to cause autosomal dominant deafness-65 (OMIM #616044) (Azaiez et al., 2014; Zhang et al., 2014). However, the proposed mechanism in this variant was gain of function or a dominant-negative. Heterozygous loss of *TBC1D24* is unlikely to be responsible for the phenotype seen with these deletions. Our analyses also indicate that epistatic or regulatory dysregulation are unlikely to explain the phenotypes of this condition. However, it is impossible to rule out the contribution to the phenotype from loss of other genes within the deletions without further studies.

In summary, we describe an additional individual with a rare chromosome 16p13.3 deletion that establishes the condition as a distinct entity and clarifies its phenotypic spectrum. Our analysis suggests that the haploinsufficiency of *ATP6V0C* is likely to underlie the pathology of this

condition. Further data and experiments will be needed to prove these assertions conclusively. The establishment of a phenotypic and genetic cohorts of patients with *ATP6V0C* loss of function mutations and neurodevelopmental pathology will be required to further elucidate the phenotypic spectrum of this emerging condition.

3.5 Contribution statement

The study was conceived by Dr Siddharth Banka; the research plan was devised by me and Dr Banka; the clinical data was collected by Dr Banka, Rory J Tinker, Dr Shruti Garg and Dr Maggie Steggall; *in silico* analyses were performed by me and Dr Sara Cuvertino; and the manuscript was drafted by Rory J Tinker and me and reviewed by all co-authors. Rory J Tinker (medical student and first author) was co-supervised by me and Dr Banka.

3.6 Chapter 3 Supplementary information

Supplementary Table 3.3 Quantified expression of the deleted genes across neuronal tissue (Transcripts per million) (TPN).

	Cerebellum	Cerebellar	Nucleus accumbens	Anterior cingulate cortex	Frontal Cortex	Cortex	Substantia nigra	Putamen	Caudate	Hypothalamus	Hippocampus	Amygdala
ATP6V0C	600.0	613.0	312.2	371.1	585.4	531.8	203.7	220.2	256.5	316.6	230.4	225.5
MIR3178	0	0	0	0	0	0	0	0	0	0	0	0
TBC1D24	27.8	26.6	6.5	9.9	11.4	11.2	3.199	3.4	4.5	5.9	4.6	5.3
AMDHD2	23.6	22.4	8.9	9.9	12.9	15.4	7.650	6.7	7.5	7.3	7.3	6.4
PDPK1	33.1	34.4	18.2	15.4	20.2	17.3	6.605	11.0	14.9	9.6	10.3	9.9
MCEMP1	0.18	0.18	0.13	0.1	0.2	0.2109	0.1663	0.1	0.2	0.1	0.1	0.1

Colour scheme: TPN >100 green; 10<TPN<99.9 amber; 1<TPN<9.9 red and TPN<1 uncoloured

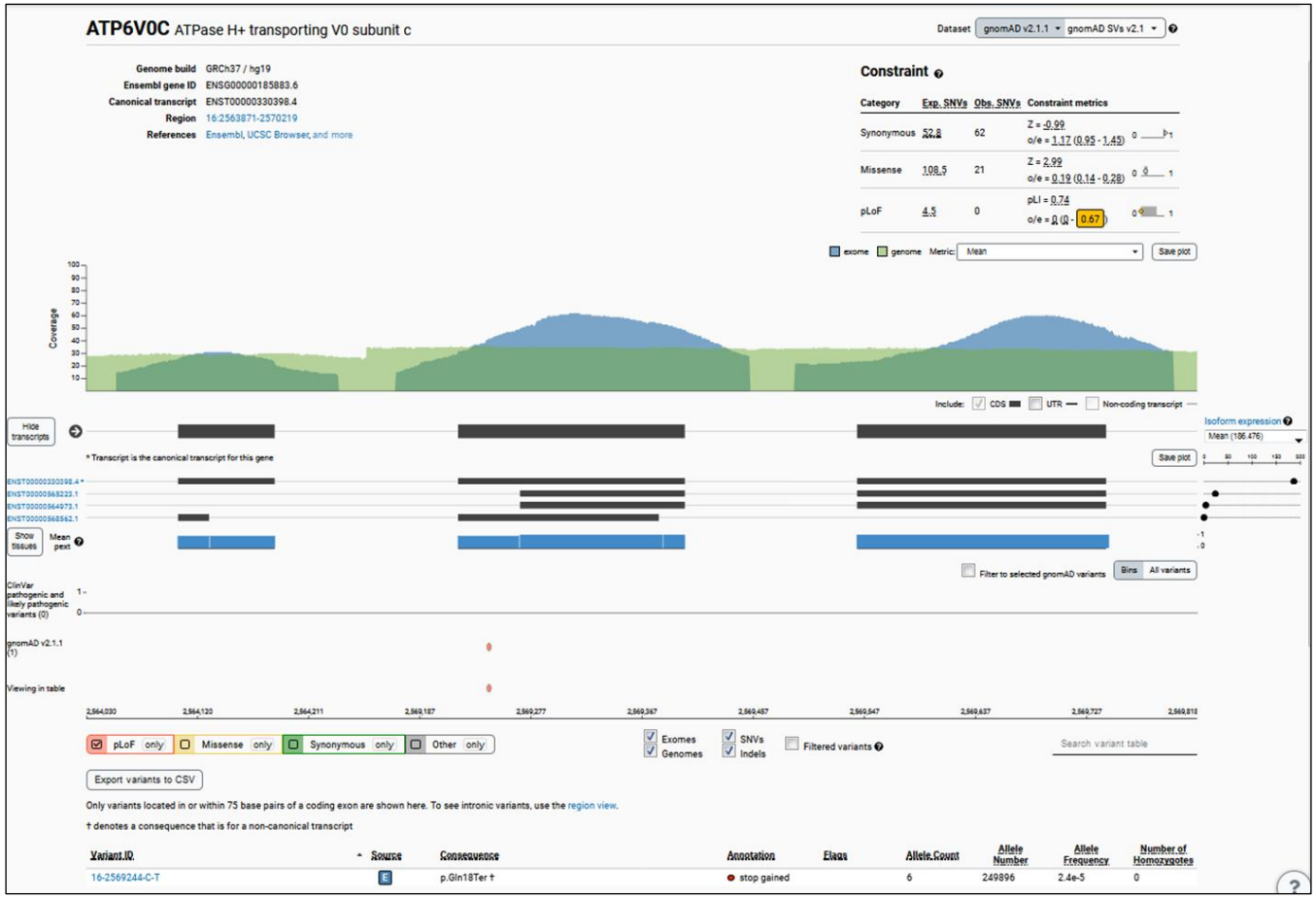
Supplementary Table 3.4: Phenological sub categorical analysis of the deleted genes biological roles.

Panther category	Category (N)
Go Slim Molecular Function	Catalytic activity (3) Transporter Activity (1) None function (1)
Go Slim Biological Process	Metabolic Process (2) Biological Regulation (1) Cellular Process (1) Response to stimuli (1) Signalling (1)
Go Slim Cellular Component	Membrane part (1) Membrane (1)
Go Slim Protein Class	Metabolite interconversion enzyme (1) Protein Modifying enzyme (1) Protein Binding activity Modulator (1) Transporter (1)
Go Slim Panther Pathways	P53 pathway (2) CCKR signalling map (1) inflammation Mediated by chemokine and cytokine signalling pathways (1) Insulin /IGF Pathway – protein kinase B signalling cascade (1) Interleukin signalling pathway (1) N-acetylglucosamine metabolism (1) PDGF signalling Pathway (1) P13 kinase pathway (1) Ras Pathway (1)

Supplementary Table 3.5: Gene and phenology specific analysis of the deleted genes.

Gene	PANTHER Family/Subfamily	PANTHER Protein Class	PANTHER GO-Slim Molecular Function	PANTHER GO-Slim Biological Process	PANTHER GO-Slim Cellular Component	Pathway
ATP6V0C	V-TYPE PROTON ATPASE 16 KDA PROTEOLIPID SUBUNIT (PTHR10263:SF5)	ATP synthase	ATPase activity ATPase-coupled transmembrane transporter activity proton transmembrane transporter activity active ion transmembrane transporter activity		Integral component of membrane	
TBC1D24	TBC1 DOMAIN FAMILY MEMBER 24 (PTHR23353:SF6)	GTPase-activating protein	-			N-acetylglucosamine metabolism->N-acetylglucosamine-6-phosphate deacetylase
AMDHD2	N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE (PTHR11113:SF14)	deacetylase	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides deacetylase activity	amino sugar catabolic process		
CEMP1	-	-	-			
PDPK1	3-PHOSPHOINOSITIDE-DEPENDENT PROTEIN KINASE 1-RELATED (PTHR24356:SF163)	Non-receptor serine/threonine protein kinase	protein serine/threonine kinase activity	peptidyl-serine phosphorylation intracellular signal transduction -		CCKR signaling map->PDPK1 PI3 kinase pathway->Activated p110 Insulin/IGF pathway-protein kinase B signaling cascade->3-phosphoinositide-dependent protein kinase 1 and 2 Interleukin signaling pathway->3-phosphoinositide-dependent protein kinase 1 and 2 p53 pathway feedback loops 2->3-phosphoinositide-dependent protein kinase 1 and 2

						Inflammation mediated by chemokine and cytokine signaling pathway->3-phosphoinositide-dependent protein kinase 1 Ras Pathway->3-phosphoinositide-dependent protein kinase PDGF signaling pathway->3-phosphoinositide-dependent protein kinase 1 and 2 p53 pathway->3-phosphoinositide-dependent protein kinase 1 and 2 PI3 kinase pathway->3-phosphoinositide-dependent protein kinase 1, active PI3 kinase pathway->3-phosphoinositide-dependent protein kinase 1



Supplementary Figure 3.2: Loss of function *ATP6V0C* variant on gnomAD

The low pLI score may be due to the relatively small size of this gene (3 exons encoding a protein of 155 amino acids). The only observed loss of function variant in this gene is located in part of exon 2 that is not included in 2 out of 4 known transcripts of the gene .

4 PRESENCE OF PATHOGENIC COPY NUMBER VARIANTS (CNVs) IS CORRELATED WITH SOCIOECONOMIC STATUS[2]

[2]This chapter is adapted from **Burghel GJ**, Khan U, Lin WY, Whittaker W, Banka S. Presence of pathogenic copy number variants (CNVs) is correlated with socioeconomic status. J Med Genet. 2020 Jan;57(1):70-72.

- Paper selected as the J Med Genet Editor's Choice.
- I presented this work at the ESHG conference 2019 (poster – Appendix E1).
- A layperson summary was published on the websites of the Journal of Medical Genetics (Appendix E2) and Manchester University NHS foundation trust (Appendix E3).

4.1 Abstract

Socioeconomic status (SES) is a major determinant of health. We studied the Index of Multiple Deprivation Rank (IMDR) of 473 families with individuals with pathogenic autosomal copy number variants (CNVs) and known inheritance status. The IMDR distribution of families with pathogenic CNVs was significantly different from the general population. Families with inherited CNVs were significantly more likely to be living in areas of higher deprivation when compared to families that had individuals with *de novo* CNVs. These results provide unique insights into biological determinants of SES. As CNVs are relatively frequent in the general population, these results have important medical and policy consequences.

4.2 Introduction

Socioeconomic status (SES) is a measure of an individual's or family's economic and social status based on factors such as income, education and occupation. SES is a major determinant of health and related outcomes (Bradley and Corwyn, 2002; Clark et al., 2009; Cox et al., 2006; Hanscombe et al., 2012). Lower SES confers increased risk for multifactorial disorders like stroke, and cardiovascular disease and plays a key role in child health and development (Clark et al., 2009; Hanscombe et al., 2012). Early life adversity negatively impacts child health and produces lasting and deleterious effects on developmental outcomes (Hackman and Farah, 2009). Biological factors including genetic variants that may influence SES are only beginning to be understood. Recently an association between pathogenic copy number variants (CNVs) and lower SES in clinically unaffected adults from the United Kingdom (UK) Biobank was described (Kendall et al., 2019). However, this study examined a limited spectrum of common CNVs and there was no information available on the inheritance of these CNVs (Kendall et al., 2019). The aim of our study was to investigate the correlation of a wide range of unselected pathogenic and likely-pathogenic CNVs, and their inheritance pattern, with SES.

4.3 Methods

4.3.1 Database

Manchester Centre for Genomic Medicine (MCGM) is a regional genomic diagnostic laboratory and receives referrals for genomic testing from the North-West (NW) of England (http://www.mangen.co.uk, 2019). MCGM offers array-comparative genomic hybridisation (aCGH) as first line of investigation to individuals with developmental disorders and congenital malformations to identify disease causing copy number variants (CNVs).

CNVs are the most common type of structural variation in the human genome and involve more base pairs than any other type of genetic variation (Stankiewicz and Lupski, 2002). CNVs can be classed as losses (deletions) or gains (e.g. duplications or triplications). Depending on their phenotypic effect, CNVs can be classed as benign (class 1), likely benign (class 2), variant of uncertain significance (class 3), likely pathogenic (class 4) and pathogenic (class 5) (Kearney et al., 2011). Likely pathogenic (class 4) and pathogenic (class 5) CNVs have been implicated in a wide range of human disorders including intellectual disabilities and neurodevelopmental delays (Rice and McLysaght, 2017). These disease-causing class 4 and 5 CNVs, like any other CNVs, can be inherited from a parent or they may arise *de novo*. Inherited CNVs generally show reduced penetrance, meaning that the individual carrying the CNVs may not be clinically classed as affected. *De novo* pathogenic CNVs are generally, associated with more severe phenotype when compared with the phenotypes of inherited pathogenic CNVs (Veltman and Brunner, 2012).

The current aCGH platform used at MCGM is the OGT v3 8x60K array platform with a backbone resolution of 180kb and is validated prior to diagnostic use. Analysis is performed using the the CytoSure™ Interpret Software (v4.9). This tool contains tracks linking information from the Database of Genomic Variants, DECIPHER, Online Mendelian Inheritance of Man and the local MCGM patient database (https://www.ogt.com, 2019). Interpretation and classification of CNVs are conducted using the laboratory standard operating procedures and published guidelines (Kearney et al., 2011).

We curated an anonymised departmental database of results from over 16,000 postnatal clinical array-CGH testing performed at MCGM between 2010 and 2017. Our database included information on each identified CNV, its clinical classification, size, loss or gain status, the inheritance status (*de novo* or inherited from a parent) where available and IMDR information extracted based on postcode from original data. Details of CNVs are in Supplementary Table 4.2.

4.3.2 Data cleaning and filtering

From the database, we selected all pathogenic (Class 5) or likely pathogenic (Class 4) CNVs. From these pathogenic/likely-pathogenic CNVs, we identified CNVs in which complete inheritance status was available (*De novo* vs inherited). From these CNVs we included only one proband from each family (to avoid sample bias by double counting). We also excluded derivative CNVs resulting from parental balanced translocation (as these CNVs cannot be classed as *de novo* or inherited in the traditional sense) and sex chromosomes CNVs (because their phenotype effect is sex dependant). Cases for which IMDR data was unavailable were excluded.

4.3.3 Determination of SES

A number of measures and methodologies to study SES in health-related contexts exist (Shavers, 2007). English Indices of Deprivation are one such widely used SES measure in health research. They combine seven domains of deprivation for small geographical areas in England (referred to as Lower-Layer Super Output Areas or LSOAs) into an overall weighted aggregation index of multiple deprivation or IMD. These seven domains and their associated weightings are: income deprivation (22.5%), employment deprivation (22.5%), health deprivation and disability (13.5%), education, skills and training deprivation (13.5%), barriers to housing and services (9.3%), crime (9.3%) and living environment deprivation (9.3%) (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015). Each of the LSOAs is scored and ranked from the most deprived (rank 1) to the least deprived (rank

32,844). LSOAs are categorised into 10 equal groups whereby the first and the tenth deciles include the most and the least deprived 10% LSOAs respectively. Each postcode in England falls within an LSOA (http://imd-by-postcode.opendatacommunities.org/). For each CNV, we retrieved the index of multiple deprivation rank (IMDR), and its seven constituent domains, using the English indices of deprivation 2015 (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015).

4.3.4 Statistical analyses

All the statistical tests were performed using R 3.5.2, unless otherwise specified. Statistical significance was set at $p < 0.05$. Chi-squared tests of independence and trend were performed to examine the associations between IMDR deciles (and its seven constituent domains) and CNV inheritance (*de novo*/inherited). To investigate if there was a correlation between age of diagnosis and SES, Jonckheere-Terpstra test was used. To explore the joint effect of the CNV type (losses/gains) and inheritance on IMDR, we further categorised samples into 4 groups based on CNV inheritance and type. Kruskal-Wallis rank sum test was used to test if IMDR are different among 4 groups. Mann Whitney U-tests were used as post-hoc testing for Kruskal-Wallis rank sum test and FDR (false discovery rate) was used to control the inflation of type I error rates. Pairwise fdr-adjusted p values were reported. CNV sizes are defined as the intervals between CNV start and end in the million base scale (Mb). To investigate if CNV sizes differ by IMDR group, Kruskal-Wallis rank sum test was used.

4.4 Results

We interrogated an anonymised departmental database of results from >16,000 postnatal, mostly paediatric (98%), clinical array comparative genomic hybridisations (aCGH) performed at the Manchester Centre for Genomic Medicine between 2010 and 2017. This database included information on each identified CNV, its clinical pathogenic classification, size, type (loss or gain), and where available inheritance status (*de novo* or inherited from a parent), and postcode IMDR of the patient (Supplementary Methods).

We identified 473 unique unrelated cases of pathogenic or likely-pathogenic autosomal CNVs (Table 4.1 and Supplementary Table 4.2) with known inheritance status ($n_{inherited}$=218; n_{denovo}= 255) and available postcode information (Supplementary Methods). We then obtained Index of Multiple Deprivation Rank (IMDR) (and the seven constituent domains) associated with each of these postcodes (https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015). English Indices of Deprivation is a widely used measure of SES in health research (Lang et al., 2016; White et al., 2016) and reflects the SES of the households of the affected individuals in this study (Supplementary Table 4.2).

Table 4.1: Frequency of pathogenic copy number variants identified in this study

Cytogenetic location of CNVs	*De novo* loss	*De novo* gain	Inherited loss	Inherited gain	Total
16p11.2	12	2	18	17	49
1q21.1q21.2	5	1	13	21	40
22q11.21	30	2	3	0	35
15q13.2q13.3	0	0	27	1	28
15q11.2q13.1	11	2	1	6	20
16p13.11	3	0	8	7	18
17q12	4	2	1	8	15
7q11.23	7	4	1	1	13
16p12.2	1	0	12	0	13
17p12	1	0	6	4	11
15q13.3	0	0	8	1	9
17p11.2	5	3	0	1	9
3q29	3	0	0	5	8
2p16.3	3	0	4	0	7
17p13.3	2	3	1	0	6
22q13.33	5	0	1	0	6
21q11.2q22.3	0	4	0	0	4
4p16.3	1	0	2	0	3
6q27	2	0	1	0	3
9p24.3p23	3	0	0	0	3
16p13.11p12.3	1	0	2	0	3
20p13	1	0	1	1	3
22q11.21q11.22	3	0	0	0	3
22q11.23q13.33	0	3	0	0	3
2q37.1q37.3	1	1	0	0	2
5q14.3	2	0	0	0	2
6p25.3p25.2	1	0	1	0	2
7p22.1	2	0	0	0	2

8p23.1	1	0	0	1	2
9p24.2p23	1	0	1	0	2
9p24.3p13.1	0	2	0	0	2
10q26.2q26.3	2	0	0	0	2
10q26.3	1	0	1	0	2
11q24.2q25	2	0	0	0	2
13q12.3q13.1	2	0	0	0	2
15q11.1q13.1	1	1	0	0	2
16p13.3	1	0	1	0	2
17q21.31	2	0	0	0	2
18p11.32p11.21	0	2	0	0	2
18p11.32p11.31	0	0	1	1	2
22q11.1q11.21	0	2	0	0	2
Other CNVs with single instance only	71	28	23	5	127

Pathogenic copy number variants (CNVs) identified in this study are listed in the descending order of their total frequencies in our cohort. Note that the individual CNVs grouped to generate frequencies have been grouped according to their chromosomal location and may include overlapping CNVs with different breakpoints. Only chromosomal locations with at least 2 CNVs in the study have been individually listed here. Full details of all the CNVs and their specific locations are provided in Supplementary Table 4.2.

We found that IMDR composition for families with pathogenic or likely-pathogenic CNVs were significantly different when compared to the IMDR of the general population of the North-West England ($p_{x^2 \text{goodness of fit}} = 1.8 \times 10^{-8}$). Furthermore, families with inherited pathogenic and likely-pathogenic CNVs were significantly more likely to be living in areas of higher deprivation when compared to families that have individuals with *de novo* pathogenic and likely-pathogenic CNVs ($p_{x^2 \text{trend}} = 2.1 \times 10^{-6}$) or with the general population of North-West England ($p_{x^2 \text{trend}} = 4.3 \times 10^{-14}$) (Figure 4.1 and Supplementary Figure 4.2) (Supplementary Results).

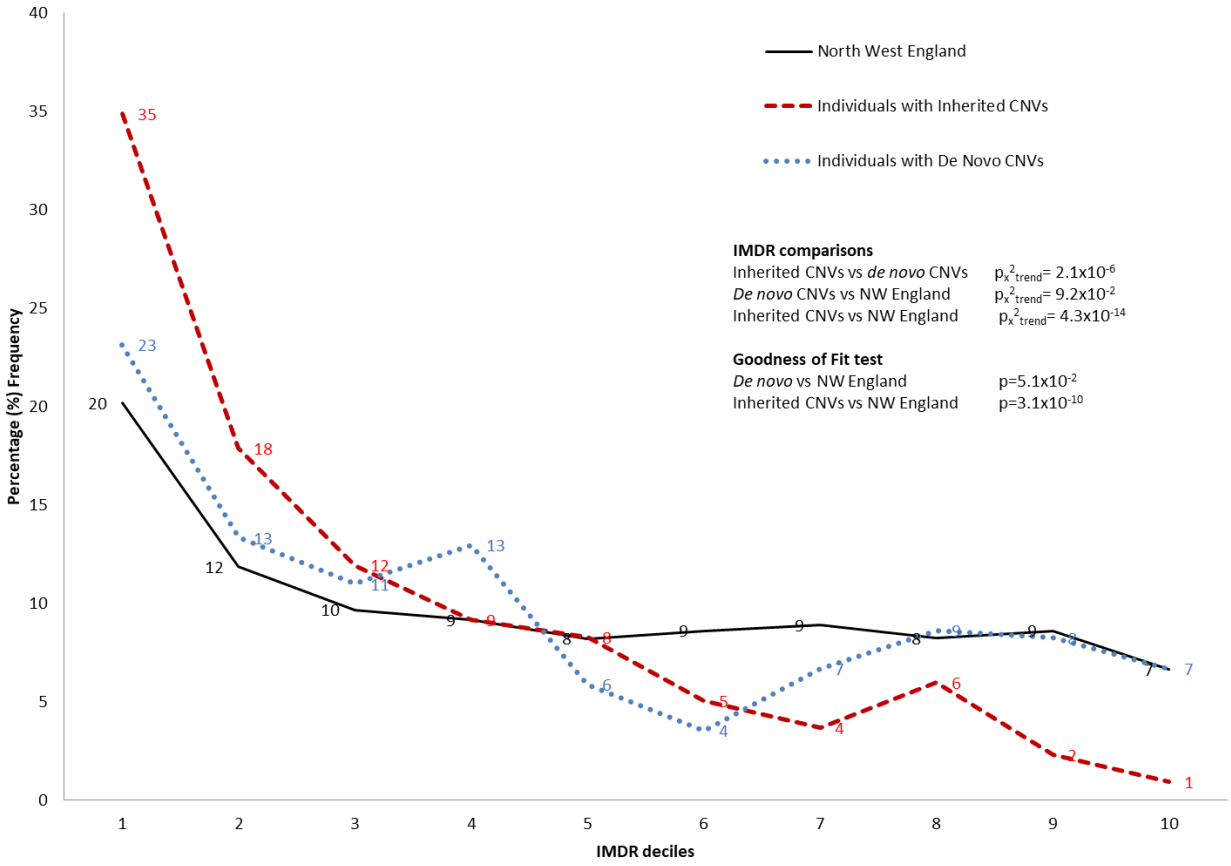


Figure 4.1: Presence of pathogenic or likely pathogenic copy number variants is correlated with SES and is primarily driven by inherited CNVs

This difference was significant across the following deprivation domains of IMDR – income; employment; health; education, skills and training (Supplementary Figure 4.3 - Supplementary Figure 4.7). There were not significant differences in the following domains - barriers to housing and services; and living environment (Supplementary Figure 4.8 - Supplementary Figure 4.9). We also performed the Jonckheere-Terpstra test in both *de novo* and inherited CNVs to see if there was any correlation between SES and age of diagnosis. No significant trends were identified (p-value = 0.0615 for *de novo* CNVs and p-value = 0.1615 for inherited CNVs) (Supplementary Figure 4.10). IMDR comparisons according to CNV type did not reveal any significant difference between CNV Losses and Gains (p=0.52) (Supplementary Results) (Supplementary Figure 4.11). There was no evidence of an effect of CNV size on IMDR (Supplementary Results) (p-Kruskal-Wallis rank sum test=0.48) (Supplementary Figure 4.12). Our data shows that the presence of pathogenic and likely pathogenic CNVs is correlated with SES. Notably in our cohort this correlation seems to be driven by partially penetrant inherited

CNVs, of low effect-size rather than *de novo* CNVs that are usually associated with more severe and more penetrant phenotypes (Rosenfeld et al., 2013).

4.5 Discussion and conclusions

The vast majority of the probands in our cohort are children (98%), and therefore, the IMDR dataset reflects the SES of the household in which they are growing up. The parents of children with *de novo* CNV do not carry the CNV and are not affected by the condition of their children. On the other hand, at least one of the parents of individuals with inherited CNVs will be carrying the same CNV. Almost 51% (n=112) of inherited CNVs in our cohort can be classed as recurrent (Supplementary Table 4.2) (Rosenfeld et al., 2013). The penetrance of these recurrent CNVs has been estimated to range between 10%-62% (Rosenfeld et al., 2013). Based on reported estimated penetrance of these recurrent CNVs, majority of the carrier parents of our index cases are likely to be classed as medically unaffected. The high level of deprivation observed in our cohort suggests that being a carrier of a low penetrant negatively impacts SES even in absence of a medical phenotype. Our observations suggest that there are likely sub-clinical effects in individuals who are medically non-penetrant carriers of milder pathogenic and likely pathogenic CNVs. This agrees with the recent findings from the UK biobank study whereby carriers of pathogenic CNVs had lower levels of household income and higher deprivation in the absence of neurodevelopmental disorders (Kendall et al., 2019). Lower SES in families with medically relevant inherited pathogenic and likely pathogenic CNVs with milder phenotype could therefore be due to cumulative multi-generational consequences of these sub-clinical effects.

These results demonstrate that in addition to the primary phenotypes of the pathogenic and likely pathogenic CNVs, their secondary socioeconomic, and resultant medical, consequences need to be studied especially in families with inherited CNVs. The combined frequency for a subset of these pathogenic CNVs in the general population is estimated to be at least 3.8% (Crawford et al., 2019) and therefore are significant in the context of public health. The correlation of SES with inheritance patterns of pathogenic and likely-pathogenic CNVs,

therefore provide further unique insights into biological determinants of SES and has important implications for planning of medical and social services.

4.6 Contribution statement

The study was conceived by Dr Siddharth Banka; the research plan was devised by me and Dr Banka; data collection and cleaning was done by me and Unzela Khan; data analysis was performed by me, Dr Wei-Yu Lin (statistical analysis) and Dr William Whittaker (IMDR analysis); and the manuscript was drafted by me and reviewed by all co-authors. Unzela Khan (MSc student) was co-supervised by me and Dr Banka.

4.7 Supplementary Results

4.7.1 Database

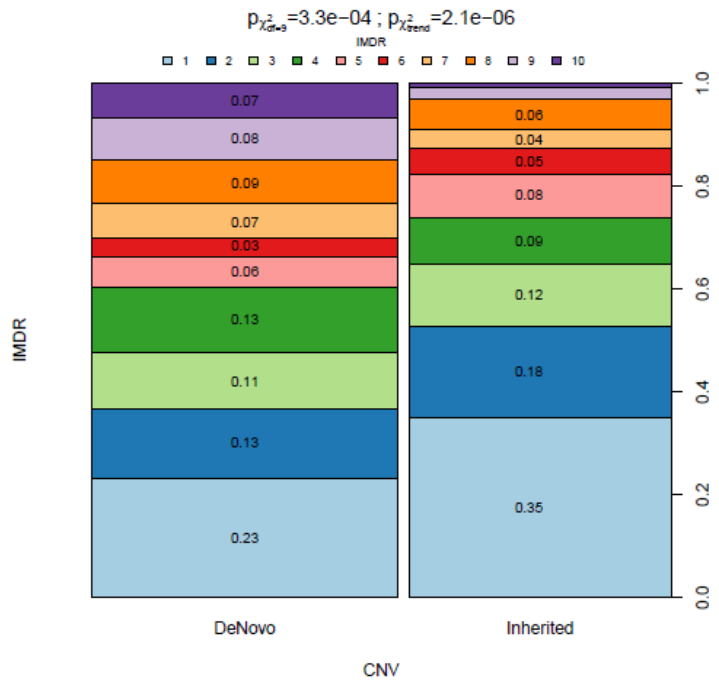
We identified 1,567 pathogenic (Class 5) or likely pathogenic (Class 4) CNVs in our database. Complete inheritance status was available for 614 CNVs (324 *de novo* and 290 inherited). Cleaning and filtering of our data left us a final set of 473 IMDR datasets corresponding to unique individuals with class 4 and 5 autosomal CNVs with full inheritance information. This included 218 inherited (138 losses and 80 gains) ranging in size between 0.002Mb-13.75Mb (median 1.33Mb), 255 *de novo* (193 losses, 62 gains) ranging in size between 0.006Mb-13.93Mb (median 3.45Mb) (Supplementary Table 4.2).

4.7.2 Detailed IMDR comparisons

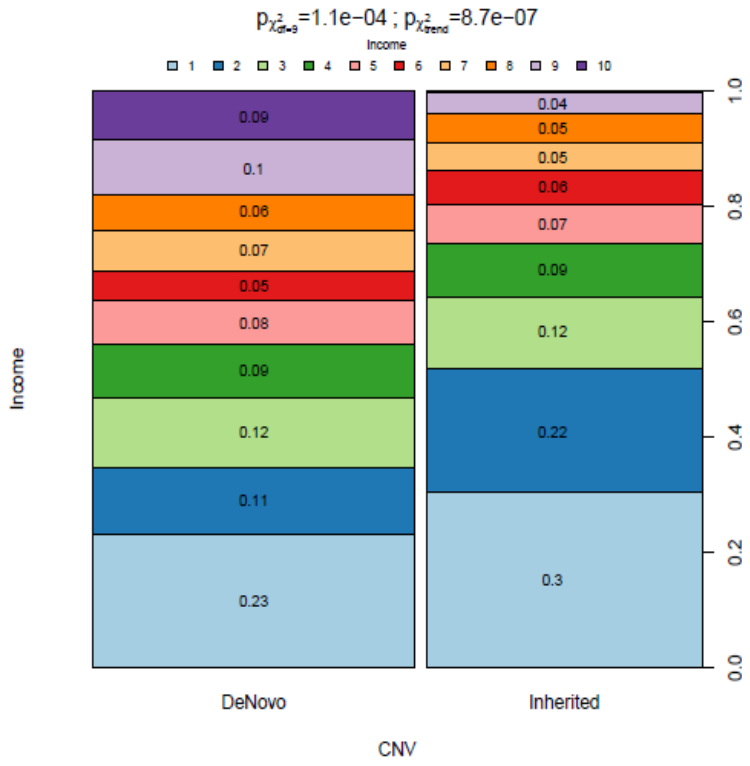
IMDR comparisons of pathogenic CNVs against the general population: Chi-squared goodness of fit test showed that the IMDR spread of our patients was significantly different from that of the NW of England population ($p=1.8x10^{-8}$) (Figure 4.1).

IMDR comparisons according to inheritance: The difference above is driven by inherited CNVs rather than *de novo* CNVs; the IMDR of patients with *de novo* CNVs were not too different to that of the NW of England population ($p=5.1x10^{-2}$) while the IMDR of patients with inherited CNVs were significantly different to that in the NW of England ($p=3.1x10^{-10}$) (Figure 4.1). We also found significant differences in IMDR scores between inherited and *de novo* pathogenic CNVs ($p_{x^2_{df=9}}=3.3x10^{-4}$) (Figure 4.1). Relative to *de novo* CNVs, inherited CNVs are 2.06 times more likely to be living in areas of high deprivation (relative risk ratios = 2.06, $p= 1.1x10^{-5}$) (Figure 4.1) (Supplementary Figure 4.2). This difference was significant across following deprivation domains of IMDR – income; employment; health; education, skills and training (Supplementary Figure 4.3 - Supplementary Figure 4.7). There were not significant differences in the following domains - barriers to housing and services; and living environment (Supplementary Figure 4.8 - Supplementary Figure 4.9).

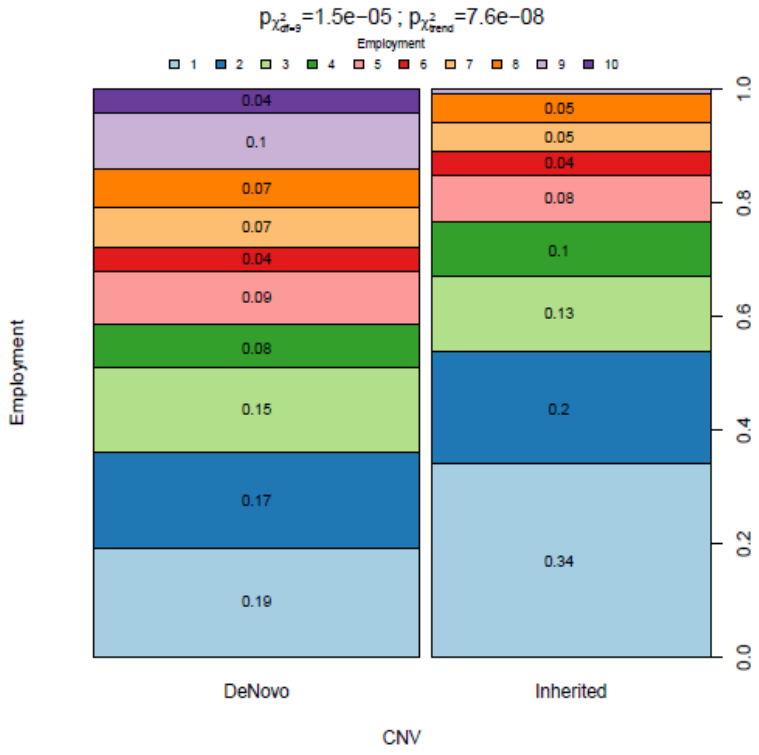
Note for Supplementary Figure 4.2 - Supplementary Figure 4.9: The widths of the bars are proportional to the number of CNVs. Colours represent decile and number on the stacked bar is the proportion of each of the deciles. Number is absent if it is 0:02. P values for two independent and trend tests are shown



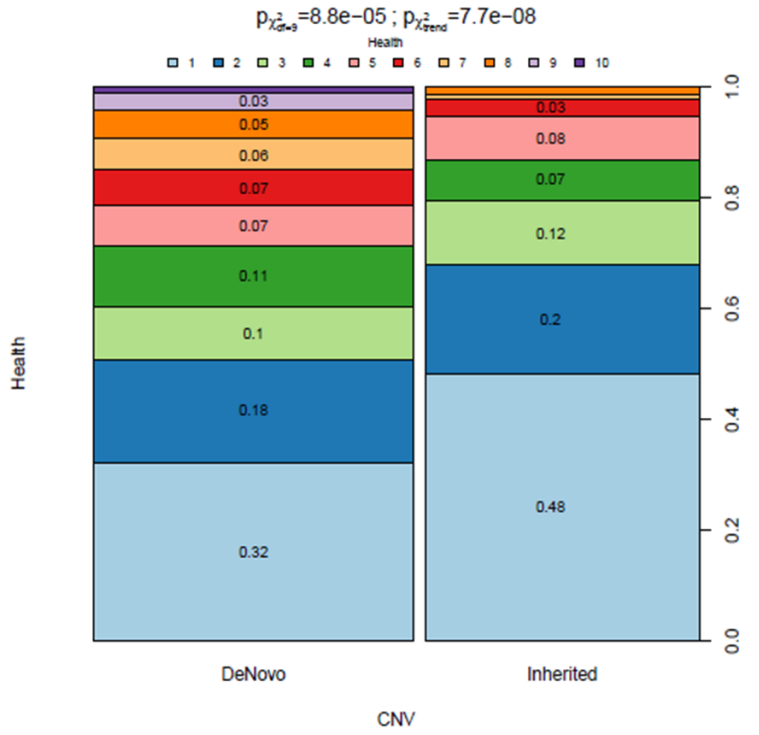
Supplementary Figure 4.2 Spineplot of IMDR and CNVs



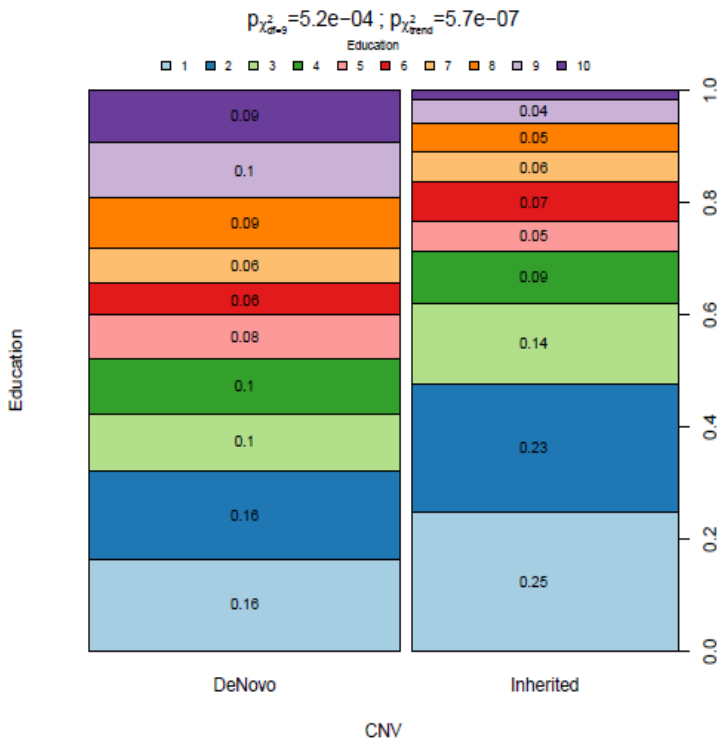
Supplementary Figure 4.3: Spineplot of Income deprivation and CNVs



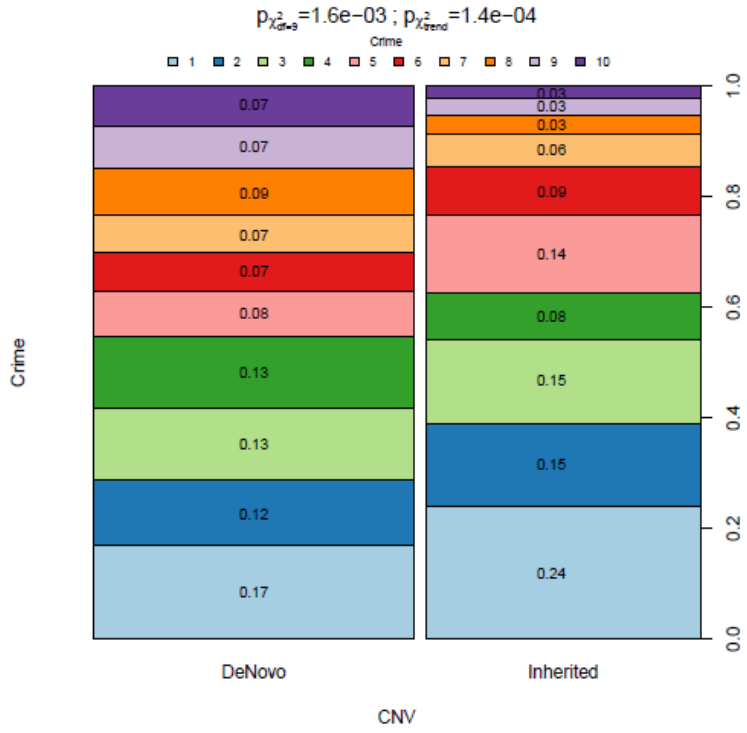
Supplementary Figure 4.4 Spineplot of Employment deprivation and CNV



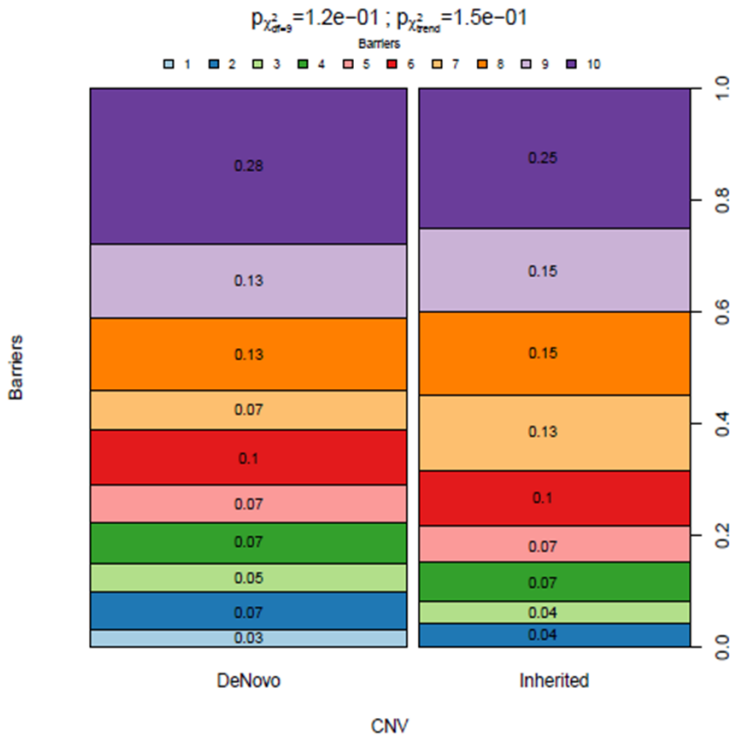
Supplementary Figure 4.5: Spineplot of Health deprivation and disability and CNVs



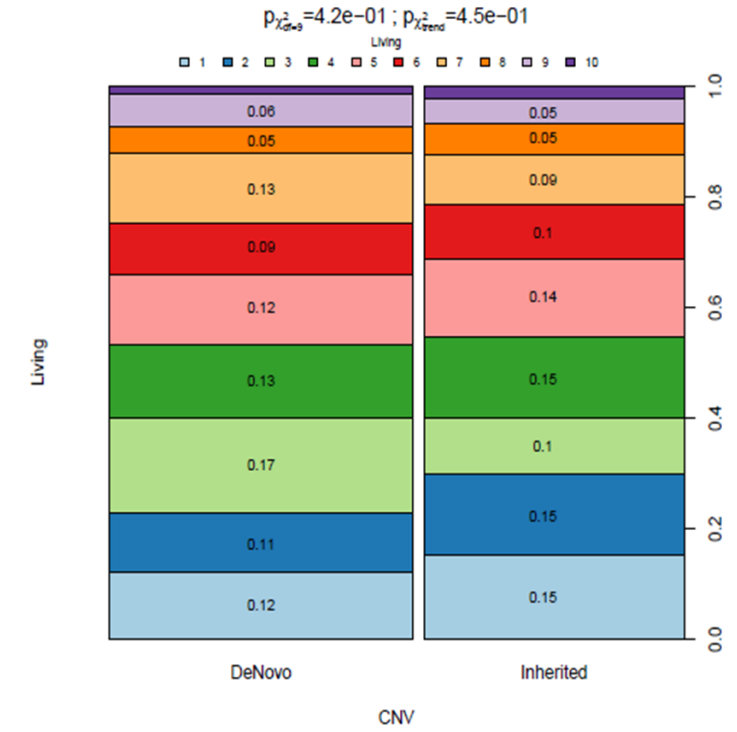
Supplementary Figure 4.6 Spineplot of Education, skills and training deprivation and CNVs



Supplementary Figure 4.7 Spineplot of Crime and CNVs

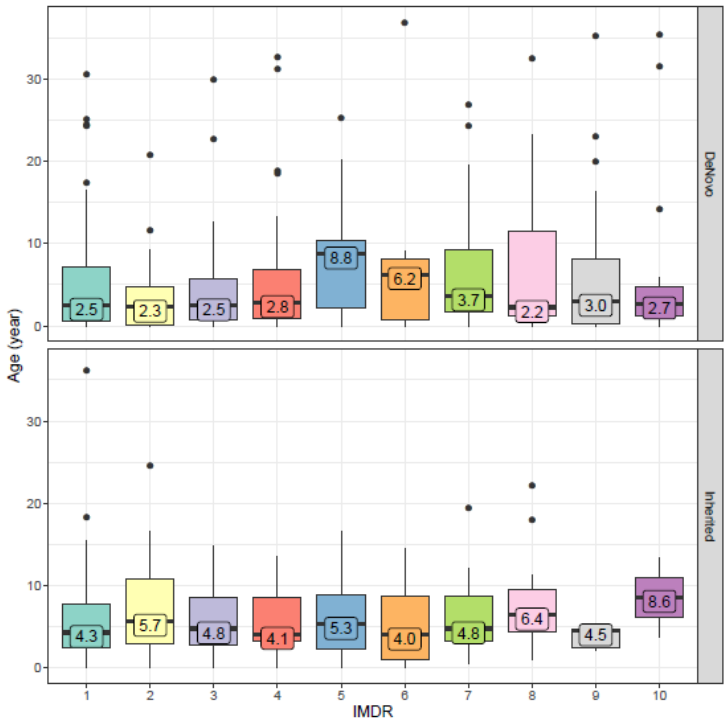


Supplementary Figure 4.8: Spineplot of Barriers to housing and services and CNVs



Supplementary Figure 4.9 Spineplot of Living environment deprivation and CNVs

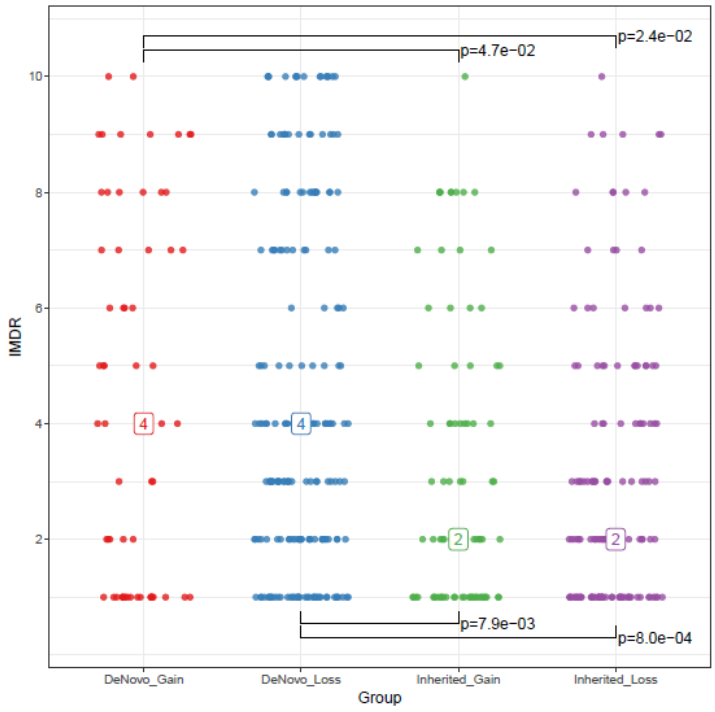
IMDR comparisons according to age: There was no significant age trend across the different IMDR deciles for *de novo* CNVs (p=0.0615) and for inherited CNVs (p=0.1615) (Supplementary Figure 4.10).



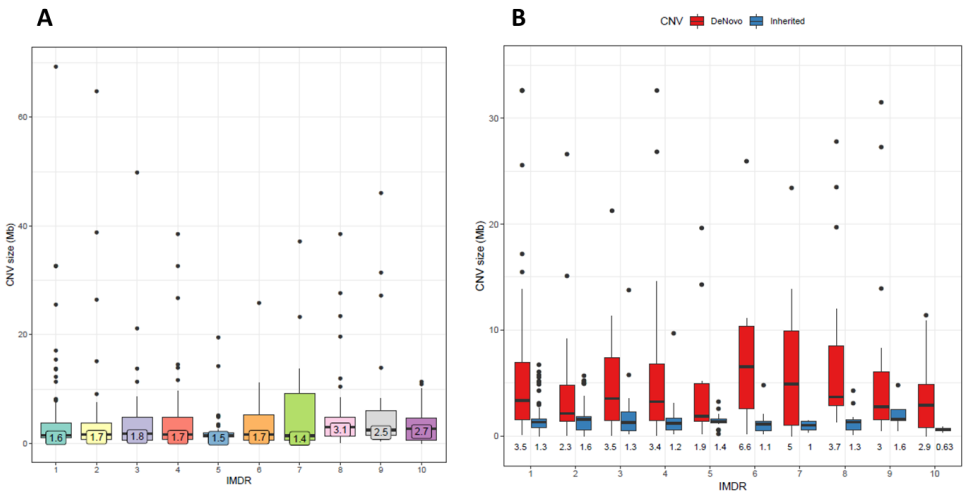
Supplementary Figure 4.10 : IMDR comparison by age by Jonckheere-Terpstra test
This analysis did not reveal significant age trend across the different IMDR deciles for *de novo* CNVs (p=0.0615) and for inherited CNVs (p=0.1615).

IMDR comparisons according to CNV type: IMDR was not significantly different between the two CNV types (Losses and Gains) (p=0.52). Of note, *de novo* CNVs have higher percentages of losses (76%) compared to inherited CNVs (63%) (p=0.005). To check the joint effect of the CNV type and inheritance on IMDR, we further categorised samples based on inheritance (*de novo*/inherited) and type (losses/gains). There are significant differences in IMDR among 4 groups ($p_{\text{Kruskal-Wallis rank sum test}}=2.1\text{x}10^{-4}$) (Supplementary Figure 4.11). Post-hoc pairwise comparisons showed that both inherited losses and gains were significantly associated with lower IMDR in comparison to *de novo* losses and gains indicating that the effect mainly resulted from mode inheritance rather than CNV type ($p_{\text{false discovery rate}} < 0.05$, Supplementary Figure 4.11).

IMDR comparisons according to CNV size: Median CNV sizes ranged from 1.4 to 3.1 Mb across IMDR (Supplementary Figure 4.12A). There was no evidence of an effect of CNV size on IMDR ($p_{Kruskal\text{-}Wallis\ rank\ sum\ test}$=0.48) (Supplementary Figure 4.12B).



Supplementary Figure 4.11: Pairwise comparisons using Wilcoxon rank sum test



Supplementary Figure 4.12: IMDR and CNV size

A) Median CNV size across IMDR. B) IMDR comparisons according to CNV inheritance and CNV size

Supplementary Table 4.2 All CNVs identified in this study

Chr	Cytogenetic Location	Start	Stop	Size (Kb)	Gain / Loss	Inheritance	Age at Diagnosis (Months)	Index of Multiple Deprivation Decile	Income Decile	Employment Decile	Education and Skills Decile	Health and Disability Decile	Crime Decile	Barriers to Housing and Services Decile	Living Environment Decile
1	p36.33p36.32	120840	3235952	3115.11	Loss	*De Novo*	149	4	3	3	5	2	7	9	3
1	p36.33	564512	1450830	886.32	Loss	Inherited	0	1	1	1	1	1	1	2	7
1	p36.33p36.23	779733	8068387	7288.65	Loss	*De Novo*	10	3	3	2	4	1	8	3	5
1	p36.33p36.21	779733	14745082	13965.35	Loss	*De Novo*	0	4	9	7	6	8	9	1	1
1	p36.21p36.12	15422198	21235203	5813.01	Loss	*De Novo*	74	4	4	4	4	4	3	3	7
1	p35.3p34.3	28529620	36537372	8007.75	Gain	*De Novo*	28	1	1	2	3	1	1	4	2
1	p34.3	36094966	38501807	2406.84	Loss	*De Novo*	79	9	10	9	9	7	7	10	6
1	p21.3p12	97154750	120520297	23365.55	Gain	*De Novo*	44	7	8	7	8	5	8	5	5
1	q21.1q21.2	143700143	149754257	6054.11	Gain	Inherited	2	1	1	1	1	1	1	9	1
1	q21.1	145415156	145799615	384.46	Loss	Inherited	45	10	9	9	10	8	7	7	9
1	q21.1q21.2	145415156	147441119	2025.96	Gain	Inherited	23	5	4	5	7	4	6	8	2
1	q21.1q21.2	145415156	147510582	2095.43	Gain	Inherited	42	6	6	5	7	4	5	10	5
1	q21.1q21.2	145415156	148545664	3130.51	Loss	*De Novo*	0	3	4	5	5	2	1	4	3
1	q21.1q21.2	145899359	147824212	1924.85	Loss	*De Novo*	131	1	1	1	1	1	6	5	5
1	q21.1q21.2	145899359	147824212	1924.85	Loss	*De Novo*	48	1	1	2	3	1	1	2	3
1	q21.1q21.2	145899359	147824212	1924.85	Loss	Inherited	57	1	1	2	2	1	1	9	1
1	q21.1q21.2	145899359	147824212	1924.85	Loss	Inherited	9	2	3	3	4	1	1	10	2
1	q21.1q21.2	146155929	147379627	1223.70	Loss	Inherited	75	1	1	1	2	2	3	3	4
1	q21.1q21.2	146155929	147379954	1224.03	Gain	Inherited	72	1	1	1	2	1	5	3	5
1	q21.1q21.2	146155929	147379954	1224.03	Gain	Inherited	34	2	2	2	3	1	1	10	1
1	q21.1q21.2	146155929	147379954	1224.03	Gain	Inherited	49	1	2	1	1	1	3	9	6
1	q21.1q21.2	146155929	147380106	1224.18	Gain	Inherited	67	5	4	3	6	3	5	8	4
1	q21.1q21.2	146155929	147380942	1225.01	Gain	Inherited	145	2	2	1	2	1	2	6	8
1	q21.1q21.2	146155929	147380942	1225.01	Gain	Inherited	49	4	4	4	5	2	3	9	2
1	q21.1q21.2	146155929	147380942	1225.01	Gain	Inherited	38	3	2	2	2	1	5	8	5
1	q21.1q21.2	146155929	147380942	1225.01	Gain	Inherited	9	1	1	1	3	1	2	7	3
1	q21.1q21.2	146155929	147824212	1668.28	Loss	*De Novo*	18	4	4	4	4	2	4	10	3
1	q21.1q21.2	146155929	147824212	1668.28	Loss	*De Novo*	2	4	4	3	5	3	4	10	3

1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	116	3	4	3	2	2	4	7	4
1	q21.1q21.2	146155929	147824212	1668.28	Loss	Inherited	48	4	3	4	6	2	2	8	4
1	q21.1q21.2	146155929	147824212	1668.28	Loss	Inherited	60	5	5	4	5	3	3	10	7
1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	61	2	3	2	2	1	5	9	1
1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	38	2	2	1	1	1	3	7	6
1	q21.1q21.2	146155929	147824212	1668.28	Loss	Inherited	6	4	4	4	4	2	5	7	2
1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	25	1	1	1	1	1	1	10	1
1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	34	1	2	1	3	1	1	10	2
1	q21.1q21.2	146155929	147824212	1668.28	Gain	Inherited	33	2	1	2	1	2	3	10	3
1	q21.1q21.2	146155929	147824212	1668.28	Loss	Inherited	0	2	3	2	2	1	3	5	7
1	q21.1q21.2	146155929	147887713	1731.78	Gain	Inherited	107	8	7	7	6	5	7	10	9
1	q21.1q21.2	146507649	147380106	872.46	Gain	Inherited	87	1	2	2	1	1	2	5	4
1	q21.1q21.2	146507649	147380942	873.29	Gain	*De Novo*	46	9	9	9	9	6	8	6	9
1	q21.1q21.2	146507649	147381479	873.83	Gain	Inherited	161	10	10	8	9	6	10	10	7
1	q21.1q21.2	146507649	147381479	873.83	Gain	Inherited	101	1	1	1	1	1	4	6	6
1	q21.1q21.2	146507649	147824212	1316.56	Loss	Inherited	0	3	4	3	3	2	3	10	4
1	q21.1q21.2	146571394	147379954	808.56	Loss	Inherited	53	4	5	4	3	2	4	9	6
1	q21.1q21.2	146571394	147379954	808.56	Loss	Inherited	46	1	1	1	1	1	1	9	7
1	q21.1q21.2	146571394	147379954	808.56	Loss	Inherited	13	1	1	2	1	1	1	7	3
1	q21.1q21.2	146571394	147824212	1252.82	Loss	Inherited	50	1	1	1	1	1	2	6	6
1	q21.1q21.2	146571394	147824212	1252.82	Loss	Inherited	47	1	1	1	3	1	2	2	6
1	q21.3q23.1	154537654	158597824	4060.17	Gain	*De Novo*	109	6	5	5	6	4	7	8	7
1	q41q42.12	219134834	225090355	5955.52	Loss	*De Novo*	195	9	9	8	8	5	6	10	6
1	q42.12q42.2	225764891	231962511	6197.62	Loss	*De Novo*	276	9	8	7	9	7	9	10	9
1	q42.2q42.3	231024479	236037015	5012.54	Loss	*De Novo*	28	2	2	2	2	2	3	7	4
1	q43	239940368	243128788	3188.42	Loss	*De Novo*	48	1	1	1	1	1	2	4	5
1	q43q44	243309051	247124412	3815.36	Loss	*De Novo*	2	8	9	8	8	5	10	6	5
2	p23.3	25204172	27557157	2352.99	Loss	*De Novo*	89	1	1	1	2	1	3	3	6
2	p16.3	50620243	51037303	417.06	Loss	Inherited	124	2	2	1	2	2	3	6	8
2	p16.3	50822818	50978515	155.70	Loss	*De Novo*	322	7	7	6	9	6	5	10	2
2	p16.3	50855476	51122200	266.72	Loss	Inherited	35	2	2	2	3	2	5	10	4
2	p16.3	50909746	51251569	341.82	Loss	Inherited	51	3	3	3	3	2	2	10	2
2	p16.3	51013656	51158275	144.62	Loss	Inherited	53	8	8	7	7	3	6	10	8

2	p16.3	51083418	51251569	168.15	Loss	*De Novo*	2	1	1	1	2	1	1	4	3
2	p16.3	51166648	51712211	545.56	Loss	*De Novo*	45	4	5	5	6	3	1	10	2
2	p16.1p15	60882964	62775266	1892.30	Loss	*De Novo*	21	6	6	7	6	4	5	10	1
2	p12p11.2	78068777	87295729	9226.95	Loss	*De Novo*	44	7	6	6	8	4	7	5	7
2	q11.1q11.2	95529037	101549409	6020.37	Loss	*De Novo*	97	9	9	10	9	8	10	2	8
2	q13	111442175	113065741	1623.57	Loss	Inherited	8	1	1	1	2	1	5	5	7
2	q22.3	145201122	145719056	517.93	Loss	*De Novo*	17	1	1	2	2	1	1	3	3
2	q24.2q24.3	160219774	165409878	5190.10	Loss	*De Novo*	107	5	5	4	7	4	4	10	3
2	q24.3q31.1	164571160	170645266	6074.11	Loss	*De Novo*	7	4	4	3	5	3	5	10	6
2	q24.3q31.1	164961109	173644660	8683.55	Gain	*De Novo*	1	6	7	6	7	4	4	10	4
2	q31.1	173582189	177950555	4368.37	Loss	*De Novo*	11	1	1	1	1	1	1	4	5
2	q32.3q33.1	195763510	202021245	6257.74	Loss	*De Novo*	0	1	1	1	1	1	1	2	5
2	q33.3q34	205122897	212885545	7762.65	Loss	*De Novo*	152	3	4	3	1	3	2	5	4
2	q37.1q37.3	234087038	239195978	5108.94	Gain	*De Novo*	130	5	7	5	3	5	6	2	6
2	q37.1q37.3	235356485	243087748	7731.26	Loss	*De Novo*	103	3	3	2	3	2	4	7	5
3	p25.3	10191433	10193448	2.02	Gain	Inherited	8	1	1	1	1	1	1	4	5
3	q13.13q13.32	109679257	118225716	8546.46	Loss	*De Novo*	9	3	3	3	2	2	4	8	6
3	q23q24	141144001	146628329	5484.33	Loss	*De Novo*	90	2	1	2	2	1	4	10	3
3	q24	143152382	148863186	5710.80	Loss	Inherited	163	2	2	2	2	1	6	8	3
3	q25.1q25.31	151342335	155574097	4231.76	Loss	*De Novo*	58	3	6	8	2	1	2	2	2
3	q26.32	175863108	178552811	2689.70	Loss	*De Novo*	7	2	2	3	4	1	1	2	2
3	q29	195521763	196554028	1032.27	Loss	*De Novo*	32	3	3	2	3	2	4	7	4
3	q29	195684950	197317103	1632.15	Gain	Inherited	84	2	2	2	4	1	3	5	5
3	q29	195684950	197317103	1632.15	Gain	Inherited	31	1	1	1	1	1	2	7	5
3	q29	195740402	197317103	1576.70	Loss	*De Novo*	58	2	4	3	1	2	2	3	4
3	q29	195740402	197317103	1576.70	Gain	Inherited	162	4	4	4	5	3	3	7	3
3	q29	195740402	197317103	1576.70	Gain	Inherited	63	1	1	1	2	1	1	7	2
3	q29	195740402	197317103	1576.70	Gain	Inherited	57	1	1	1	2	1	1	8	2
3	q29	195804645	197837069	2032.42	Loss	*De Novo*	22	5	5	5	9	5	4	9	2
4	p16.3	72320	1333296	1260.98	Loss	Inherited	185	1	1	1	2	1	1	4	4
4	p16.3	72320	4028412	3956.09	Loss	*De Novo*	2	9	9	8	8	7	9	7	9
4	p16.3	75647	564526	488.88	Loss	Inherited	30	2	2	1	2	1	2	7	8
4	p16.1p11	9814844	48609760	38794.92	Gain	*De Novo*	0	2	2	2	2	1	2	10	5

4	p15.33p15.31	14062788	18362091	4299.30	Loss	Inherited	39	1	1	1	1	1	1	10	1
4	q12q13.1	53850058	65512919	11662.86	Loss	*De Novo*	291	7	9	7	9	4	7	6	3
4	q27q31.21	122266843	143502993	21236.15	Gain	*De Novo*	65	3	3	3	3	2	3	10	3
4	q32.3q34.3	169088100	179139609	10051.51	Loss	*De Novo*	36	10	10	10	9	7	10	7	6
4	q35.1q35.2	184717961	190469515	5751.55	Loss	Inherited	77	1	1	2	3	1	2	2	2
5	p15.33	22149	3239621	3217.47	Loss	Inherited	17	5	6	5	3	4	7	8	5
5	p15.33p15.1	22149	15151743	15129.59	Loss	*De Novo*	27	2	1	1	1	2	3	7	7
5	p15.33p14.3	22149	19606209	19584.06	Gain	*De Novo*	6	5	9	9	7	8	10	1	1
5	p15.33p14.1	22149	27788732	27766.58	Gain	*De Novo*	15	8	10	10	10	9	9	1	3
5	p15.33p11	22149	46115173	46093.02	Gain	*De Novo*	25	9	8	8	10	5	10	9	6
5	p15.33p13.3	2274696	29103194	26828.50	Gain	*De Novo*	16	4	5	5	5	2	9	10	1
5	p13.2	36877363	37586235	708.87	Gain	*De Novo*	0	1	1	1	1	1	1	5	3
5	q12.1q13.1	62167701	67066730	4899.03	Loss	*De Novo*	69	8	7	6	10	4	8	9	7
5	q14.3	88062497	88556592	494.10	Loss	*De Novo*	39	1	1	2	2	1	1	9	1
5	q14.3	88193289	88450318	257.03	Loss	*De Novo*	22	2	2	2	2	2	1	9	4
5	q15q23.1	95864493	115563383	19698.89	Loss	*De Novo*	205	8	7	7	8	7	6	10	4
5	q21.3q23.2	108714679	125896994	17182.32	Loss	*De Novo*	180	1	1	1	1	1	4	9	7
5	q34	161309644	161681646	372.00	Loss	Inherited	433	1	1	1	1	1	1	9	1
5	q35.2q35.3	174897894	177013956	2116.06	Loss	*De Novo*	0	2	2	3	2	3	6	8	1
5	q35.3	176625772	176694826	69.05	Loss	*De Novo*	11	4	7	7	9	3	3	6	1
6	p25.3p25.2	132025	3243890	3111.87	Loss	Inherited	0	1	1	1	2	1	1	8	2
6	p25.3p25.2	132025	4078678	3946.65	Loss	*De Novo*	30	1	1	1	1	2	2	3	3
6	p25.3p24.3	189601	7133658	6944.06	Loss	*De Novo*	78	3	3	3	3	2	2	10	1
6	p25.3p25.1	2155698	5875603	3719.91	Loss	*De Novo*	15	4	3	3	4	3	4	10	5
6	p25.2p25.1	2380260	6062956	3682.70	Loss	*De Novo*	28	8	8	8	9	7	4	9	5
6	p25.1p22.3	7019804	20740949	13721.15	Gain	*De Novo*	208	1	1	1	1	1	2	9	5
6	p21.33p21.31	31075418	34427330	3351.91	Gain	*De Novo*	3	8	7	7	7	7	8	10	5
6	q14.1	80499981	82577828	2077.85	Gain	Inherited	34	1	1	1	1	1	1	8	4
6	q25.3	156115120	157531075	1415.96	Loss	*De Novo*	44	9	9	9	10	9	8	4	7
6	q27	165033413	170923504	5890.09	Loss	*De Novo*	0	1	1	1	1	1	2	7	3
6	q27	169450930	170906766	1455.84	Loss	Inherited	72	2	2	2	2	1	4	8	4
6	q27	170103409	170906766	803.36	Loss	*De Novo*	379	10	10	9	10	8	10	10	7
7	p22.1	5151574	6745570	1594.00	Loss	*De Novo*	6	8	8	7	8	6	7	10	3

7	p22.1	5370613	6296827	926.21	Loss	*De Novo*	33	5	5	5	5	4	3	10	2
7	p14.1p12.1	41518844	51980756	10461.91	Gain	*De Novo*	278	8	8	6	8	8	8	10	7
7	p12.3p11.2	46523414	56174815	9651.40	Gain	Inherited	145	4	4	4	2	4	6	8	2
7	q11.23	72549388	76214107	3664.72	Loss	*De Novo*	23	1	1	1	1	1	1	4	7
7	q11.23	72643724	74142342	1498.62	Gain	Inherited	77	8	7	6	8	5	7	10	8
7	q11.23	72645480	74142342	1496.86	Gain	*De Novo*	52	4	4	4	10	2	2	6	3
7	q11.23	72645480	74142342	1496.86	Gain	*De Novo*	36	9	8	9	9	5	6	8	8
7	q11.23	72645840	74142342	1496.50	Gain	*De Novo*	24	2	2	2	3	1	3	6	3
7	q11.23	72649950	74142342	1492.39	Loss	*De Novo*	303	5	5	4	4	4	4	10	4
7	q11.23	72649950	74142342	1492.39	Loss	Inherited	43	1	1	1	1	1	1	7	3
7	q11.23	72649950	74193397	1543.45	Loss	*De Novo*	0	2	3	3	4	1	1	6	3
7	q11.23	72766312	74142342	1376.03	Gain	*De Novo*	301	1	1	2	1	1	1	4	1
7	q11.23	72766312	74142342	1376.03	Loss	*De Novo*	33	2	2	3	2	3	1	9	2
7	q11.23	72766312	74142342	1376.03	Loss	*De Novo*	22	1	1	1	1	1	2	3	8
7	q11.23	72766312	74142342	1376.03	Loss	*De Novo*	0	2	2	2	3	2	1	5	2
7	q11.23	72766312	74142342	1376.03	Loss	*De Novo*	7	3	4	4	4	2	1	10	2
7	q21.11	79393141	85145078	5751.94	Loss	Inherited	41	3	4	5	5	2	1	4	3
7	q21.11q21.3	83852376	93039621	9187.25	Loss	*De Novo*	61	2	1	2	2	1	4	10	3
7	q21.3q36.3	94257592	159124141	64866.55	Gain	*De Novo*	29	2	3	2	2	2	1	10	4
7	q21.3	95325118	96142267	817.15	Loss	*De Novo*	226	4	3	3	5	3	3	10	4
7	q22.1q31.1	102579883	110065384	7485.50	Loss	*De Novo*	31	2	2	3	2	2	2	2	1
7	q31.1q31.31	109958035	117946370	7988.34	Loss	*De Novo*	273	3	3	4	2	2	2	4	2
7	q32.1q36.3	127632552	159128555	31496.00	Gain	*De Novo*	0	9	9	9	7	4	8	9	8
7	q35q36.1	146405016	148100456	1695.44	Loss	Inherited	68	2	2	2	2	2	3	6	6
7	q36.3	155965227	159124141	3158.91	Loss	*De Novo*	61	1	1	1	1	1	2	7	1
8	p23.3p23.1	61749	11723203	11661.45	Loss	*De Novo*	14	4	5	3	6	4	6	3	4
8	p23.1	7074596	12334393	5259.80	Loss	*De Novo*	0	1	1	2	2	1	1	2	2
8	p23.1	8130572	11181487	3050.92	Gain	Inherited	88	8	8	8	9	6	6	6	5
8	p23.1p11.22	12586532	39133799	26547.27	Gain	*De Novo*	4	2	3	2	3	1	2	9	3
8	p22q21.2	17145513	86548194	69402.68	Gain	*De Novo*	26	1	1	1	1	1	2	8	8
8	p21.3p21.2	21779623	23325848	1546.23	Loss	*De Novo*	54	1	1	1	2	1	1	9	2
8	q11.23q12.1	54909137	58653480	3744.34	Loss	*De Novo*	249	2	2	2	2	1	3	6	1
8	q24.11	117857424	117868530	11.11	Loss	*De Novo*	55	7	9	6	6	4	3	10	5

9	p24.3	204090	1803749	1599.66	Loss	Inherited	55	9	8	8	8	8	8	10	6
9	p24.3p23	204090	9817292	9613.20	Loss	*De Novo*	160	4	5	4	4	2	4	8	7
9	p24.3p23	204090	10126183	9922.09	Loss	*De Novo*	35	7	9	5	7	4	6	10	7
9	p24.3p23	204090	10126183	9922.09	Loss	*De Novo*	0	7	6	6	6	4	9	9	7
9	p24.3p13.1	204090	38815471	38611.38	Gain	*De Novo*	152	8	8	7	7	6	5	10	6
9	p24.3p13.1	204090	38815471	38611.38	Gain	*De Novo*	50	4	4	3	2	3	7	4	8
9	p24.3p22.3	948026	15260598	14312.57	Loss	*De Novo*	105	5	6	7	7	2	3	10	5
9	p24.2p23	3063252	11573652	8510.40	Loss	*De Novo*	26	8	7	7	8	6	10	10	6
9	p24.2p23	3662381	9195389	5533.01	Loss	Inherited	0	1	1	1	1	1	2	4	5
9	p24.1p23	4856332	12984040	8127.71	Gain	*De Novo*	8	1	1	1	1	1	1	4	1
9	q21.12q21.13	72464822	75835811	3370.99	Loss	*De Novo*	33	4	4	3	6	2	4	10	3
9	q22.2q22.31	93301497	95959417	2657.92	Loss	Inherited	62	2	2	2	2	1	4	8	4
9	q22.2q22.32	93686087	96782492	3096.41	Gain	Inherited	45	4	4	3	3	2	9	9	7
9	q34.2q34.3	136005837	140378774	4372.94	Gain	*De Novo*	44	10	10	9	9	9	9	10	10
9	q34.3	140469159	141073897	604.74	Loss	*De Novo*	3	3	2	2	4	2	4	8	3
10	p14	7127004	11755025	4628.02	Loss	*De Novo*	222	4	4	4	4	3	5	10	7
10	p11.22p11.1	34167350	38408613	4241.26	Gain	*De Novo*	234	7	10	10	9	4	10	8	1
10	q11.22q11.23	49430978	51031850	1600.87	Loss	Inherited	186	2	2	2	3	2	3	6	4
10	q11.23q23.2	52020150	89246167	37226.02	Gain	*De Novo*	23	7	6	5	7	6	6	10	9
10	q24.31q24.32	102969334	103453028	483.69	Gain	*De Novo*	240	9	9	10	9	8	10	5	4
10	q24.32q25.3	104752664	116131723	11379.06	Loss	*De Novo*	0	10	10	10	10	7	8	8	9
10	q26.2q26.3	129167026	135434113	6267.09	Loss	*De Novo*	52	2	4	2	3	1	8	10	1
10	q26.2q26.3	129391404	135434113	6042.71	Loss	*De Novo*	0	4	4	4	7	3	1	6	4
10	q26.3	131197502	135434113	4236.61	Loss	*De Novo*	7	1	1	1	2	1	5	5	5
10	q26.3	131197502	135434113	4236.61	Loss	Inherited	14	8	8	8	9	5	6	4	4
11	p15.1p14.1	18300114	30323877	12023.76	Loss	*De Novo*	25	8	10	8	10	9	8	2	4
11	p13	31825724	31831382	5.66	Loss	*De Novo*	15	10	10	9	10	8	10	10	7
11	q22.3q23.3	105289128	116644381	11355.25	Loss	*De Novo*	14	3	3	3	4	3	2	8	3
11	q24.1q25	123748005	134868420	11120.42	Loss	*De Novo*	8	6	7	5	5	3	9	10	3
11	q24.2q25	123950746	134868420	10917.67	Loss	*De Novo*	21	10	10	9	10	8	10	10	7
11	q24.2q25	124540302	134868420	10328.12	Loss	*De Novo*	75	6	6	5	7	4	5	10	4
11	q24.3q25	128293973	134868420	6574.45	Loss	*De Novo*	198	1	1	1	1	1	1	5	4
12	p13.33p13.2	230437	11646534	11416.10	Gain	*De Novo*	22	1	1	1	1	1	1	10	1

12	p13.33p13.32	244335	4736827	4492.49	Loss	*De Novo*	47	1	1	1	2	1	3	3	6
12	p12.1p11.22	21536624	28105323	6568.70	Loss	*De Novo*	97	6	5	4	6	3	9	6	7
13	q12.11q12.12	20407270	23948573	3541.30	Loss	*De Novo*	360	3	3	3	4	1	4	4	4
13	q12.3q13.1	30848032	32452613	1604.58	Loss	*De Novo*	57	3	2	2	3	1	7	9	9
13	q12.3q13.1	30848032	32797505	1949.47	Loss	*De Novo*	126	1	2	2	1	1	2	5	2
13	q13.1q14.13	32753843	46566127	13812.28	Loss	*De Novo*	20	7	7	7	8	4	7	7	9
13	q13.1q31.1	33786863	83585442	49798.58	Loss	*De Novo*	0	3	4	4	4	1	2	10	2
13	q14.11q21.1	40814667	55433232	14618.57	Loss	*De Novo*	107	4	4	3	4	3	5	7	4
13	q33.2q34	106223621	115092581	8868.96	Loss	*De Novo*	34	4	4	4	3	4	4	9	4
13	q33.3q34	108700491	113028081	4327.59	Loss	*De Novo*	0	4	3	3	5	3	7	6	3
14	q11.2	21819565	23529619	1710.05	Loss	*De Novo*	81	2	3	2	5	1	3	10	2
14	q13.1q21.1	34572408	38115804	3543.40	Loss	*De Novo*	149	3	3	4	4	2	1	9	2
14	q13.2q21.1	35700740	39240633	3539.89	Loss	*De Novo*	30	5	6	5	5	4	6	5	5
14	q21.1q21.2	38428480	43557024	5128.54	Loss	Inherited	108	2	2	1	2	1	8	9	6
14	q32.2q32.33	101218538	105407712	4189.17	Gain	*De Novo*	32	10	10	10	10	10	10	8	7
14	q32.32q32.33	103245881	105797981	2552.10	Loss	*De Novo*	17	9	9	9	10	9	8	4	7
15	q11.1q13.1	20686203	28526437	7840.23	Loss	*De Novo*	13	1	1	2	2	1	1	2	1
15	q11.1q13.1	20686203	28859765	8173.56	Gain	*De Novo*	41	1	2	2	2	1	1	10	1
15	q11.1q13.2	20686203	30366138	9679.94	Gain	*De Novo*	392	4	4	5	6	2	4	10	3
15	q11.2q13.1	22652047	28134489	5482.44	Loss	*De Novo*	0	8	8	8	10	5	9	7	4
15	q11.2q13.1	22729339	28559437	5830.10	Loss	*De Novo*	2	2	2	2	2	1	3	8	2
15	q11.2q13.1	23455146	28559437	5104.29	Gain	Inherited	73	1	1	2	1	1	1	4	1
15	q11.2q13.1	23586243	28520316	4934.07	Gain	Inherited	146	2	3	2	3	1	1	5	1
15	q11.2q13.1	23586243	28520316	4934.07	Gain	Inherited	48	2	2	2	3	2	5	9	1
15	q11.2q13.1	23586243	28526437	4940.19	Gain	*De Novo*	16	3	3	3	3	2	3	10	3
15	q11.2q13.1	23656064	28520316	4864.25	Loss	*De Novo*	15	8	8	7	8	5	4	8	9
15	q11.2q13.1	23656064	28520316	4864.25	Loss	*De Novo*	0	5	5	5	4	4	4	10	4
15	q11.2q13.1	23656064	28520316	4864.25	Loss	*De Novo*	5	7	7	5	5	5	9	9	9
15	q11.2q13.1	23656064	28520316	4864.25	Loss	Inherited	32	1	1	1	2	2	1	6	4
15	q11.2q13.1	23656064	28559437	4903.37	Loss	*De Novo*	13	1	1	2	2	1	1	2	2
15	q11.2q13.1	23656064	28559437	4903.37	Loss	*De Novo*	26	10	10	9	10	8	10	10	7
15	q11.2q13.1	23656064	28559437	4903.37	Loss	*De Novo*	0	10	10	8	10	7	7	9	7
15	q11.2q13.1	23656064	28606275	4950.21	Loss	*De Novo*	2	7	6	5	8	5	5	10	9

15	q11.2q13.1	23656064	28795723	5139.66	Gain	Inherited	135	1	1	1	1	1	1	2	3
15	q11.2q13.1	23707435	28520316	4812.88	Loss	*De Novo*	2	4	5	4	5	4	2	10	2
15	q11.2q13.1	23707435	28520316	4812.88	Loss	*De Novo*	0	1	1	1	1	1	2	10	1
15	q11.2q13.1	23707435	28520316	4812.88	Gain	Inherited	174	6	5	6	5	5	5	8	4
15	q11.2q13.1	23707435	28726651	5019.22	Gain	Inherited	139	2	2	3	2	3	1	9	2
15	q11.2q13.1	23782549	28520316	4737.77	Gain	*De Novo*	73	2	2	2	2	1	3	8	3
15	q13.2q13.3	30389992	32861612	2471.62	Loss	Inherited	106	3	3	4	2	3	4	9	2
15	q13.2q13.3	30389992	32899558	2509.57	Gain	Inherited	41	1	1	1	2	1	1	7	1
15	q13.2q13.3	30419801	32861612	2441.81	Loss	Inherited	82	1	1	1	1	1	2	4	7
15	q13.2q13.3	30419801	32899558	2479.76	Loss	Inherited	39	3	3	2	3	2	6	7	5
15	q13.2q13.3	30419801	32899558	2479.76	Loss	Inherited	26	9	6	7	10	8	9	8	10
15	q13.2q13.3	30858990	32432125	1573.14	Loss	Inherited	143	5	5	4	4	3	5	9	4
15	q13.2q13.3	30888776	32426904	1538.13	Loss	Inherited	45	2	1	2	1	1	3	7	2
15	q13.2q13.3	30888776	32443563	1554.79	Loss	Inherited	93	1	1	1	1	1	1	4	5
15	q13.2q13.3	30888776	32445920	1557.14	Loss	Inherited	26	1	1	1	2	1	5	3	5
15	q13.2q13.3	30888776	32509932	1621.16	Loss	Inherited	74	2	2	4	4	2	3	3	2
15	q13.2q13.3	30888776	32702923	1814.15	Loss	Inherited	145	2	2	2	1	1	3	4	4
15	q13.2q13.3	30888776	32702923	1814.15	Loss	Inherited	109	4	3	3	8	2	5	8	4
15	q13.2q13.3	30888776	32702923	1814.15	Loss	Inherited	32	1	1	1	1	1	1	8	4
15	q13.2q13.3	30954724	32443124	1488.40	Loss	Inherited	78	1	2	1	1	1	1	10	1
15	q13.2q13.3	30954724	32443124	1488.40	Loss	Inherited	19	5	6	4	6	4	2	6	7
15	q13.2q13.3	30954724	32443563	1488.84	Loss	Inherited	40	4	4	3	4	2	4	10	4
15	q13.2q13.3	30954724	32443563	1488.84	Loss	Inherited	43	3	3	4	6	2	3	10	1
15	q13.2q13.3	30954724	32443563	1488.84	Loss	Inherited	16	1	1	1	1	1	2	6	6
15	q13.2q13.3	30954724	32445230	1490.51	Loss	Inherited	67	1	1	1	1	1	2	8	5
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	160	5	4	4	6	2	4	10	5
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	154	3	3	2	3	2	4	6	4
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	104	1	1	1	1	1	4	5	7
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	76	8	9	8	8	6	6	6	8
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	59	1	1	1	2	1	1	10	1
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	41	1	1	1	1	1	1	2	9
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	41	2	3	3	2	2	1	8	1
15	q13.2q13.3	30954724	32509932	1555.21	Loss	Inherited	29	9	9	9	9	6	5	5	9

15	q13.2q13.3	30954724	32702923	1748.20	Loss	Inherited	199	2	2	1	2	1	2	10	3
15	q13.3	31261891	32702923	1441.03	Gain	Inherited	91	7	6	5	8	5	5	10	9
15	q13.3	31972643	32509932	537.29	Loss	Inherited	60	3	2	3	2	4	4	5	3
15	q13.3	31972643	32635957	663.31	Loss	Inherited	0	6	6	5	7	5	4	9	6
15	q13.3	32064983	32443563	378.58	Loss	Inherited	295	2	2	2	3	2	2	7	2
15	q13.3	32064983	32443563	378.58	Loss	Inherited	11	1	1	1	1	1	1	5	5
15	q13.3	32064983	32450670	385.69	Loss	Inherited	46	7	9	7	9	5	10	2	6
15	q13.3	32064983	32509932	444.95	Loss	Inherited	94	3	4	3	4	2	3	10	3
15	q13.3	32064983	32509932	444.95	Loss	Inherited	58	6	7	6	6	3	7	7	5
15	q13.3	32064983	32509932	444.95	Loss	Inherited	31	3	3	3	3	3	8	10	2
15	q24.1q26.3	75148259	102383479	27235.22	Gain	*De Novo*	0	9	8	8	7	9	9	9	5
15	q26.3	98907958	100390515	1482.56	Loss	Inherited	68	6	5	5	6	6	7	6	5
16	p13.3	72768	1195490	1122.72	Loss	Inherited	108	1	1	1	1	1	3	4	3
16	p13.3	2108399	2132544	24.15	Loss	*De Novo*	139	2	2	3	3	1	4	5	3
16	p13.3p12.3	2180265	17676029	15495.76	Gain	*De Novo*	1	1	1	1	1	1	3	6	5
16	p13.12p13.11	14762141	16525374	1763.23	Loss	Inherited	74	5	5	5	8	2	3	7	3
16	p13.11	14910213	16194575	1284.36	Loss	Inherited	32	5	7	6	5	3	5	7	4
16	p13.11	14910213	16194575	1284.36	Gain	Inherited	26	5	6	4	5	4	5	7	5
16	p13.11	14910213	16420691	1510.48	Loss	*De Novo*	119	9	8	8	9	7	9	9	6
16	p13.11	14910213	16420691	1510.48	Loss	Inherited	53	1	2	1	2	1	2	4	1
16	p13.11	14910213	16420691	1510.48	Loss	Inherited	36	2	4	4	3	1	1	10	1
16	p13.11	14910213	16525374	1615.16	Loss	*De Novo*	17	2	2	2	3	2	2	4	2
16	p13.11	14910213	16525374	1615.16	Loss	Inherited	14	4	5	5	6	3	3	10	1
16	p13.11	15022142	16420691	1398.55	Loss	Inherited	18	3	3	4	7	1	2	10	3
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	142	1	1	1	1	1	5	7	7
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	133	1	1	1	1	1	1	9	5
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	114	8	9	8	9	5	5	7	7
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	96	1	2	1	1	1	1	10	1
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	49	5	6	5	7	3	2	5	7
16	p13.11	15048732	16194575	1145.84	Gain	Inherited	35	4	4	3	3	2	6	8	5
16	p13.11	15404476	16194575	790.10	Loss	Inherited	102	2	2	2	2	1	3	8	3
16	p13.11p12.3	15404476	18306841	2902.37	Loss	*De Novo*	25	8	7	8	8	4	9	10	3
16	p13.11p12.3	15404476	18306841	2902.37	Loss	Inherited	136	1	2	2	2	1	1	4	4

16	p13.11p12.3	15404476	18306841	2902.37	Loss	Inherited	95	3	3	2	5	2	2	10	3
16	p13.11	15551062	16194575	643.51	Loss	*De Novo*	111	2	2	2	2	2	5	10	1
16	p13.11	15551062	16311080	760.02	Loss	Inherited	48	1	2	1	2	1	4	10	2
16	p13.11	15551062	16311080	760.02	Loss	Inherited	55	1	1	1	1	1	1	10	1
16	p12.3	16861655	18141015	1279.36	Loss	Inherited	10	6	5	6	6	3	7	6	6
16	p12.2	21838419	22467007	628.59	Loss	*De Novo*	423	9	9	9	7	5	6	7	7
16	p12.2	21926361	22467007	540.65	Loss	Inherited	200	5	5	4	4	3	5	9	9
16	p12.2	21926361	22467007	540.65	Loss	Inherited	146	7	7	5	7	4	8	7	10
16	p12.2	21926361	22467007	540.65	Loss	Inherited	140	6	6	5	6	5	6	8	6
16	p12.2	21926361	22467007	540.65	Loss	Inherited	136	8	8	8	8	4	10	5	6
16	p12.2	21926361	22467007	540.65	Loss	Inherited	69	3	3	3	3	2	4	10	1
16	p12.2	21926361	22467007	540.65	Loss	Inherited	54	9	9	7	10	7	6	7	10
16	p12.2	21926361	22467007	540.65	Loss	Inherited	34	1	1	1	2	1	3	3	8
16	p12.2	21926361	22467007	540.65	Loss	Inherited	20	3	3	3	3	2	2	9	2
16	p12.2	21959950	22407951	448.00	Loss	Inherited	74	1	1	1	1	1	2	6	7
16	p12.2	21959950	22407951	448.00	Loss	Inherited	33	1	1	1	2	1	3	3	6
16	p12.2	21959950	22467007	507.06	Loss	Inherited	152	3	2	2	3	2	9	9	10
16	p12.2	21959950	22467007	507.06	Loss	Inherited	107	4	3	3	5	2	7	9	3
16	p11.2	28455394	29182200	726.81	Loss	*De Novo*	30	2	2	1	1	2	4	8	3
16	p11.2	28455394	29182200	726.81	Loss	Inherited	59	7	7	7	4	5	9	10	3
16	p11.2	28682192	29076269	394.08	Loss	*De Novo*	58	10	10	9	10	9	10	10	7
16	p11.2	28843754	29031071	187.32	Loss	*De Novo*	68	5	7	5	8	6	10	1	3
16	p11.2	28843754	29031071	187.32	Loss	*De Novo*	0	2	1	2	2	2	5	4	2
16	p11.2	28843754	29031071	187.32	Loss	Inherited	119	5	5	5	7	4	6	9	2
16	p11.2	28843754	29031071	187.32	Loss	Inherited	49	4	5	4	5	4	2	10	2
16	p11.2	28843754	29031071	187.32	Loss	Inherited	60	3	4	3	3	2	9	10	1
16	p11.2	28843754	29031071	187.32	Loss	Inherited	26	2	3	2	4	1	1	6	4
16	p11.2	29478060	30190593	712.53	Gain	*De Novo*	292	1	2	2	1	1	6	8	1
16	p11.2	29478060	30190593	712.53	Gain	Inherited	54	3	2	2	2	3	5	9	6
16	p11.2	29592791	30190593	597.80	Gain	*De Novo*	38	1	1	1	1	1	2	6	7
16	p11.2	29592791	30190593	597.80	Loss	Inherited	215	8	7	7	7	7	10	10	4
16	p11.2	29592791	30190593	597.80	Gain	Inherited	125	1	1	1	1	1	5	7	7
16	p11.2	29592791	30190593	597.80	Gain	Inherited	118	3	3	2	2	2	3	9	7

16	p11.2	29592791	30190593	597.80	Gain	Inherited	81	1	1	1	1	1	1	5	6
16	p11.2	29592791	30190593	597.80	Gain	Inherited	26	8	9	7	9	5	5	6	8
16	p11.2	29592791	30190593	597.80	Gain	Inherited	78	3	2	2	2	1	5	8	5
16	p11.2	29592791	30190593	597.80	Gain	Inherited	45	4	3	3	3	3	6	6	8
16	p11.2	29592791	30190593	597.80	Gain	Inherited	32	2	2	3	2	2	1	7	3
16	p11.2	29592791	30190593	597.80	Gain	Inherited	4	1	1	1	1	1	1	9	3
16	p11.2	29592791	30190593	597.80	Gain	Inherited	15	7	7	6	7	5	5	10	8
16	p11.2	29592791	30190593	597.80	Gain	Inherited	2	4	3	3	5	4	6	5	5
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	171	10	10	10	9	9	9	8	9
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	160	5	5	5	5	3	8	5	4
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	46	7	6	6	8	5	8	8	7
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	27	3	4	3	4	2	1	10	6
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	38	3	3	2	3	2	3	9	6
16	p11.2	29673967	30190593	516.63	Loss	*De Novo*	34	4	5	3	3	2	3	10	7
16	p11.2	29673967	30190593	516.63	Loss	Inherited	177	2	2	2	4	1	2	7	4
16	p11.2	29673967	30190593	516.63	Loss	Inherited	115	3	3	2	3	2	5	3	9
16	p11.2	29673967	30190593	516.63	Gain	Inherited	103	4	5	5	3	3	4	10	1
16	p11.2	29673967	30190593	516.63	Loss	Inherited	55	1	1	2	1	1	2	10	1
16	p11.2	29673967	30190593	516.63	Loss	Inherited	73	4	4	3	4	2	3	10	4
16	p11.2	29673967	30190593	516.63	Gain	Inherited	90	2	2	2	2	1	2	10	5
16	p11.2	29673967	30190593	516.63	Gain	Inherited	63	8	8	6	7	5	9	6	10
16	p11.2	29673967	30190593	516.63	Loss	Inherited	22	2	2	2	2	2	3	7	3
16	p11.2	29673967	30190593	516.63	Gain	Inherited	36	1	1	1	1	1	5	3	6
16	p11.2	29673967	30190593	516.63	Gain	Inherited	23	1	2	1	2	1	3	10	1
16	p11.2	29673967	30190593	516.63	Loss	Inherited	15	6	6	6	6	3	7	10	2
16	p11.2	29673967	30190593	516.63	Gain	Inherited	0	4	4	3	2	3	5	8	4
16	p11.2	29673967	30264952	590.99	Loss	*De Novo*	126	1	1	1	2	1	4	4	6
16	p11.2	29673967	30264952	590.99	Loss	Inherited	96	5	5	5	4	4	4	10	4
16	p11.2	29673967	30264952	590.99	Loss	Inherited	78	2	2	2	3	1	4	10	4
16	p11.2	29673967	30264952	590.99	Loss	Inherited	21	1	1	1	1	1	2	4	7
16	p11.2	29673967	30264952	590.99	Loss	Inherited	0	2	1	2	1	1	3	7	2
16	p11.2	29673967	30264952	590.99	Loss	Inherited	0	1	1	1	1	1	5	8	5
16	p11.2	29673967	30332569	658.60	Loss	*De Novo*	85	2	3	2	4	1	5	3	4

16	p11.2	29673967	30332569	658.60	Loss	Inherited	35	1	1	1	1	1	1	8	9
16	q24.2q24.3	87845657	89547614	1701.96	Loss	*De Novo*	133	1	1	1	1	1	5	7	7
17	p13.3	1656	948473	946.82	Loss	Inherited	88	4	5	3	5	2	2	8	5
17	p13.3	1656	1965782	1964.13	Loss	*De Novo*	4	9	8	7	8	6	10	7	9
17	p13.3	1184539	1602513	417.97	Gain	*De Novo*	39	1	2	1	2	1	1	9	2
17	p13.3	1248825	1422430	173.61	Gain	*De Novo*	86	6	7	5	6	4	3	7	8
17	p13.3	2456924	3112133	655.21	Gain	*De Novo*	155	7	7	9	8	8	10	2	3
17	p13.3	2502877	2588761	85.88	Loss	*De Novo*	5	3	3	3	2	2	3	8	4
17	p12	14111754	15442178	1330.42	Loss	*De Novo*	55	1	1	1	1	1	4	6	6
17	p12	14111754	15442178	1330.42	Gain	Inherited	232	7	7	7	8	4	7	7	9
17	p12	14111754	15442178	1330.42	Gain	Inherited	265	8	9	8	9	5	8	4	7
17	p12	14111754	15442178	1330.42	Loss	Inherited	178	3	3	3	4	3	3	8	2
17	p12	14111754	15442178	1330.42	Loss	Inherited	155	1	1	1	1	1	1	5	6
17	p12	14111754	15442178	1330.42	Loss	Inherited	121	1	2	1	3	1	1	9	1
17	p12	14111754	15442178	1330.42	Loss	Inherited	95	1	2	1	1	1	1	10	1
17	p12	14111754	15442178	1330.42	Loss	Inherited	99	5	6	5	4	3	7	10	2
17	p12	14111754	15442178	1330.42	Loss	Inherited	109	5	5	4	4	3	7	10	4
17	p12	14111754	15442178	1330.42	Gain	Inherited	48	6	6	5	8	4	6	6	5
17	p12	14111754	15442178	1330.42	Gain	Inherited	11	8	8	8	6	6	8	9	5
17	p11.2	16637872	20219455	3581.58	Loss	*De Novo*	22	2	4	4	3	1	1	10	1
17	p11.2	16637872	20294010	3656.14	Loss	*De Novo*	46	1	1	1	2	1	1	8	2
17	p11.2	16637872	20294010	3656.14	Loss	*De Novo*	1	9	9	9	10	6	8	10	5
17	p11.2	16762073	20219455	3457.38	Gain	*De Novo*	14	8	10	9	10	6	2	9	5
17	p11.2	16782547	20294010	3511.46	Gain	*De Novo*	170	1	1	1	1	1	2	3	5
17	p11.2	16782547	20294010	3511.46	Gain	*De Novo*	20	1	1	1	1	1	1	5	5
17	p11.2	16782547	20294010	3511.46	Loss	*De Novo*	40	8	10	9	10	5	3	4	8
17	p11.2	16782547	20294010	3511.46	Gain	Inherited	2	3	3	3	3	1	5	8	8
17	p11.2	17375712	19925867	2550.16	Loss	*De Novo*	98	8	8	8	9	5	6	8	9
17	q11.2	28652130	31021253	2369.12	Loss	*De Novo*	228	8	9	9	8	8	7	9	2
17	q11.2	29033863	30326952	1293.09	Loss	*De Novo*	9	8	9	8	10	6	5	8	1
17	q12	34652173	36326362	1674.19	Loss	*De Novo*	31	2	2	1	3	1	1	9	2
17	q12	34850785	36248926	1398.14	Gain	Inherited	55	7	8	6	7	5	6	6	5
17	q12	34850785	36248926	1398.14	Gain	Inherited	11	1	1	1	1	1	1	9	1

17	q12	34856305	36248926	1392.62	Gain	*De Novo*	118	5	5	4	4	5	5	9	5
17	q12	34856305	36248926	1392.62	Gain	*De Novo*	0	5	8	8	8	7	4	1	1
17	q12	34856305	36248926	1392.62	Gain	Inherited	219	1	2	1	1	1	2	8	1
17	q12	34856305	36248926	1392.62	Gain	Inherited	110	1	1	1	2	1	1	9	2
17	q12	34856305	36248926	1392.62	Gain	Inherited	57	1	1	1	1	1	4	9	7
17	q12	34856305	36248926	1392.62	Loss	Inherited	37	1	2	2	2	1	1	10	1
17	q12	34856305	36248926	1392.62	Gain	Inherited	28	5	4	4	6	3	7	8	8
17	q12	35076189	36101737	1025.55	Loss	*De Novo*	114	3	3	3	4	1	2	10	1
17	q12	35199438	36326362	1126.92	Gain	Inherited	97	1	1	1	2	1	1	9	2
17	q12	35199438	36326422	1126.98	Gain	Inherited	163	6	8	8	8	5	6	2	2
17	q21.31	42815171	43032872	217.70	Loss	*De Novo*	20	10	10	10	9	8	7	7	8
17	q21.31	43706895	44694311	987.42	Loss	*De Novo*	22	1	1	1	1	1	1	5	5
17	q23.1q23.2	58172677	60395815	2223.14	Loss	Inherited	103	4	4	4	4	3	3	10	3
18	p11.32p11.31	149089	5406436	5257.35	Loss	Inherited	43	2	2	2	2	1	2	10	5
18	p11.32p11.21	149089	14081934	13932.85	Gain	*De Novo*	0	9	9	8	9	6	9	9	3
18	p11.32p11.31	605175	4391041	3785.87	Gain	Inherited	77	2	2	3	1	1	3	8	5
18	p11.32p11.21	1754415	14081934	12327.52	Gain	*De Novo*	294	1	1	1	2	1	1	8	2
18	q21.2	53050512	53172293	121.78	Loss	Inherited	122	1	1	1	1	1	3	2	6
18	q21.31q23	54532547	78012819	23480.27	Loss	*De Novo*	139	8	9	8	7	7	5	6	7
18	q22.1q23	64165310	78012819	13847.51	Loss	*De Novo*	92	1	1	1	1	1	1	8	4
20	p13	60734	1474792	1414.06	Loss	Inherited	6	7	8	7	9	4	8	10	2
20	p13	60734	1935292	1874.56	Loss	*De Novo*	242	5	7	5	8	6	10	1	3
20	p13	60734	2798659	2737.93	Gain	Inherited	29	3	3	3	4	3	2	8	2
20	p12.3p12.2	5281784	10070297	4788.51	Loss	Inherited	57	9	8	8	10	6	6	10	9
20	q11.22	33898988	34141829	242.84	Loss	Inherited	10	6	7	7	8	5	10	2	2
20	q13.33	61704302	62382472	678.17	Loss	Inherited	20	1	1	1	2	1	1	6	1
21	q11.2q21.3	15485038	29238260	13753.22	Loss	Inherited	0	3	3	3	4	1	6	8	7
21	q11.2q22.3	15485038	48090352	32605.31	Gain	*De Novo*	23	4	5	4	5	3	3	10	3
21	q11.2q22.3	15485038	48090352	32605.31	Gain	*De Novo*	4	1	1	1	3	1	1	8	1
21	q11.2q22.3	15485038	48090352	32605.31	Gain	*De Novo*	0	1	1	1	1	1	3	6	7
21	q11.2q22.3	15485038	48090352	32605.31	Gain	*De Novo*	0	1	3	5	2	1	1	1	1
21	q22.13q22.3	39111103	45815113	6704.01	Gain	Inherited	138	1	1	1	2	1	2	4	1
21	q22.2q22.3	40558279	48090352	7532.07	Loss	*De Novo*	53	3	4	2	5	1	6	6	4

22	q11.1q11.21	16940617	18848020	1907.40	Gain	*De Novo*	2	3	3	3	4	2	3	5	3
22	q11.1q11.21	17397633	18706059	1308.43	Gain	*De Novo*	134	8	8	7	9	5	7	10	5
22	q11.21	18628147	21661435	3033.29	Loss	*De Novo*	12	3	3	2	3	2	3	9	6
22	q11.21	18661699	20311733	1650.03	Loss	*De Novo*	83	1	1	1	1	1	2	2	6
22	q11.21	18661699	21457610	2795.91	Loss	*De Novo*	63	4	4	4	6	2	8	6	7
22	q11.21	18661699	21661435	2999.74	Loss	*De Novo*	48	1	1	1	1	1	4	6	5
22	q11.21	18661699	21661435	2999.74	Loss	*De Novo*	31	9	9	9	10	6	8	10	5
22	q11.21	18661699	21661435	2999.74	Loss	*De Novo*	9	10	10	10	10	10	8	6	10
22	q11.21	18661699	21661435	2999.74	Loss	*De Novo*	0	1	1	1	1	1	3	6	7
22	q11.21	18661699	21808997	3147.30	Loss	*De Novo*	1	2	2	2	2	2	5	6	6
22	q11.21	18765102	21457610	2692.51	Loss	*De Novo*	53	10	10	9	9	7	5	10	7
22	q11.21	18765102	21661435	2896.33	Loss	*De Novo*	425	10	9	9	9	8	7	7	8
22	q11.21	18765102	21661435	2896.33	Loss	*De Novo*	1	2	2	1	2	1	4	4	3
22	q11.21	18818376	21540347	2721.97	Loss	*De Novo*	135	1	1	1	1	1	4	9	8
22	q11.21	18818376	21661435	2843.06	Loss	*De Novo*	375	4	2	3	5	4	7	7	5
22	q11.21	18818376	21883930	3065.55	Loss	*De Novo*	389	8	7	7	8	6	9	8	4
22	q11.21	18847961	21440515	2592.55	Loss	*De Novo*	2	4	3	3	4	3	5	9	5
22	q11.21	18847961	21440515	2592.55	Loss	*De Novo*	0	1	1	2	2	1	2	1	5
22	q11.21	18847961	21457610	2609.65	Loss	*De Novo*	34	4	3	3	4	2	6	9	6
22	q11.21	18847961	21468411	2620.45	Loss	*De Novo*	2	1	1	1	3	1	2	2	4
22	q11.21	18847961	21468411	2620.45	Loss	Inherited	3	1	1	1	1	1	1	9	5
22	q11.21	18894820	20279641	1384.82	Gain	*De Novo*	119	5	6	5	6	3	3	9	3
22	q11.21	18894820	20279641	1384.82	Gain	*De Novo*	4	4	5	5	4	3	7	2	3
22	q11.21	18894820	21025719	2130.90	Loss	*De Novo*	367	1	1	1	2	1	3	6	6
22	q11.21	18894820	21025719	2130.90	Loss	*De Novo*	71	10	10	10	10	10	10	3	10
22	q11.21	18894820	21025719	2130.90	Loss	*De Novo*	0	2	2	2	1	2	3	8	3
22	q11.21	18894820	21407681	2512.86	Loss	*De Novo*	2	2	2	3	5	1	1	2	1
22	q11.21	18894820	21440515	2545.70	Loss	*De Novo*	56	3	2	3	3	3	4	8	1
22	q11.21	18894820	21440515	2545.70	Loss	Inherited	0	5	6	5	6	2	9	10	2
22	q11.21	18894820	21457610	2562.79	Loss	*De Novo*	443	6	5	5	7	4	6	10	5
22	q11.21	18894820	21457610	2562.79	Loss	*De Novo*	7	7	5	6	9	5	8	8	3
22	q11.21	18894820	21457610	2562.79	Loss	*De Novo*	0	3	3	2	3	2	5	9	4
22	q11.21	18894820	21457610	2562.79	Loss	Inherited	0	2	2	2	3	2	5	9	1

22	q11.21q11.23	19058830	25528036	6469.21	Gain	*De Novo*	112	7	6	6	8	4	9	8	9
22	q11.21	19399499	20311733	912.23	Loss	*De Novo*	140	4	4	3	5	2	6	10	6
22	q11.21	19747275	21457610	1710.34	Loss	*De Novo*	32	9	9	9	9	6	9	4	9
22	q11.21	21081284	21457610	376.33	Loss	*De Novo*	33	1	1	1	2	1	1	6	1
22	q11.21	21081284	21759580	678.30	Loss	*De Novo*	82	4	3	3	8	2	5	8	4
22	q11.21q11.22	21759521	22905025	1145.50	Loss	*De Novo*	59	7	7	7	6	4	6	10	4
22	q11.21q11.22	21759521	22905025	1145.50	Loss	*De Novo*	1	10	10	9	10	7	8	6	10
22	q11.21q11.22	21759521	22905025	1145.50	Loss	*De Novo*	10	1	1	1	1	1	2	2	6
22	q11.23q13.33	24685145	50241148	25556.00	Gain	*De Novo*	7	1	1	1	1	1	3	3	8
22	q11.23q13.33	24754813	50681161	25926.35	Gain	*De Novo*	0	6	6	6	7	3	8	10	3
22	q13.2q13.33	42924458	51178213	8253.76	Loss	*De Novo*	74	9	9	9	9	6	8	5	4
22	q13.2q13.33	44073850	51178213	7104.36	Loss	*De Novo*	15	1	1	2	1	2	1	8	1
22	q13.31q13.33	45096595	51178213	6081.62	Loss	*De Novo*	21	3	3	3	3	2	4	10	4
22	q13.33	50241089	51178213	937.12	Loss	*De Novo*	43	2	1	2	1	2	3	8	6
22	q13.33	51121858	51178213	56.36	Loss	*De Novo*	32	4	4	3	5	3	4	10	4
22	q13.33	51123497	51134355	10.86	Loss	Inherited	135	2	2	2	4	1	1	3	4

5 FINAL DISCUSSION

CNVs are a common cause of human disease, especially neurodevelopmental disorders, and have been investigated in diagnostic setting for more than a decade (Coughlin et al., 2012; Rice and McLysaght, 2017; Riggs et al., 2020; Zhang et al., 2009). This has profoundly improved investigations for a wide range of patients including those with intellectual disability, developmental delays and congenital abnormalities (Conrad et al., 2010; Hehir-Kwa et al., 2010; Riggs et al., 2020; Schaaf et al., 2011). Nevertheless, the pathogenic mechanisms of CNVs are complex with variable expressivity and incomplete penetrance associations and so we are far from understanding their true and full impact (Collins et al., 2021; Riggs et al., 2020; Zarrei et al., 2015; Zhang et al., 2009). The overall aim of this work was to improve the interpretation of CNVs and our understanding of their impact.

5.1 Improving interpretation of CNVs

In Chapter 2, I have presented an efficient re-analysis strategy for Class 3 CNLs in diagnostic laboratories. Although the value of re-analysis of next generation sequencing data is well established, this is the first such study of CNV re-analysis. The overall rate of class 3 CNLs reclassification in this study (~0.6%), may appear low but, there are no comparable published studies of reclassification of aCGH data. Of note, our strategy prioritised ~200 CNLs for manual reanalysis from an overall list of ~2,000 CNLs. This reduction demonstrates the utility of the approach presented here and the level of reclassification in the priority short list was ~6.4%.

Notably, this work focussed on deletions only (CNLs). This is because the consequences of deletions could be assumed by loss of one or more haploinsufficient genes within the deleted region (Collins et al., 2021). However, the impact of gains is more difficult to assess in a group-based analysis (Collins et al., 2021; Hurles et al., 2008). However, triplosensitive genes/regions are being curated by the ClinGen dosage sensitivity group (https://dosage.clinicalgenome.org/, accessed March 2021). In future, this, in addition to up-to-date disease associations, will make a strategy similar to what is presented here, feasible for assessing gains as well.

Most of CNV pathogenicity can be explained by their effect on dosage-sensitive disease genes (Rice and McLysaght, 2017) which led us to a gene-focussed approach of reclassifying class 3 CNLs. However, CNVs are known to cause human disease by affecting regulatory regions such as promoters, and TADs and their boundaries (Lupiáñez et al., 2015; Zhang and Lupski, 2015). A systematic reinterpretation of 922 deletion cases from Decipher database has shown that up to 11% of deletions effect could be attributed to effects on TADs (Spielmann and Mundlos, 2013). Assessment for disruption of TADs boundaries could improve clinical interpretation of CNVs (Lupiáñez et al., 2016).

Published evidence in the literature from reviewing WGS and WES data have shown that downgrading class 3 variants into benign categories form a considerable part of reclassification rates (Das et al., 2014; Hiatt et al., 2018). We expect a substantial proportion of Class 3 CNVs to be benign, but at present there is insufficient evidence to downgrade them. Mapping against up-to-date population databases, such as database of genomic variants (DGV) (MacDonald et al., 2014) and gnomAD SV (Collins et al., 2020) can help to downgrade any CNVs that are now known to be common in the general population.

In this work, we focussed only on autosomal CNVs because of extra considerations needed for reviewing X-linked class 3 CNVs including the sex of the cases and also the lower sensitivity of constraint scores, such as pLI, to predict X-linked haploinsufficiency. These X-linked constraint scores limitations relate to multiple factors such as the presence of heathy female heterozygote carriers of (likely)pathogenic variants in the control populations and the presence of less X-chromosomes (~25% less assuming relatively equal male/female representation) compared to autosomes (Havrilla et al., 2019; Ziegler et al., 2019). However, incorporating X-chromosome CNVs in future re-analyses could further improve the reclassification rates.

Our study focussed on coding CNLs affecting genes with known OMIM phenotype associations. This may miss interesting genes with recent publications. Moreover, although haploinsufficient scores are reasonable predictors of haploinsufficiency, they are not perfect for all genes and they could be less accurate for shorter ones (Karczewski et al., 2020). Our work with *ATP6V0C*,

a relatively small gene with 3 exons (155 codons), illustrates this issue. Other than that, using cut-off scores, pLI and RVIS create a binary divide of haploinsufficient and non-haploinsufficient genes (Lek et al., 2016; Petrovski et al., 2013). Although useful for characterising a set of genes as haploinsufficient, binary divisions do not reflect the reality of a continuous spectrum of the selective pressure against dosage variation (Karczewski et al., 2020). More recent haploinsufficiency scores, such as the loss-of-function observed/expected upper bound fraction (LOEUF), provides a continuous measure of intolerance to LoF variants and is based on the largest cohort for a score of its type (Karczewski et al., 2020). It is therefore more accurate and allows the classification of genes on a spectrum of intolerance to LoF variation (Karczewski et al., 2020). Follow up work is underway looking at genes with no current OMIM associations that also include LOEUF scores to help in shortlisting these genes. As part of this work we already identified 12 extra genes that now have relevant OMIM disease associations that have been published since our OMIM data interrogation in November 2018. These genes and their associated CNLs are now undergoing full reclassification. The rest of this work and the deeper interrogation of the data may help in identifying novel genetics syndromes and/or may provide insights into disease mechanisms (Banka et al., 2015; Cuvertino et al., 2017; Kasher et al., 2016; Tinker et al., 2020). Extra analysis, including more complex methodologies that account for regulatory and epistatic gene to gene interactions, could be applied to try and identify novel contiguous gene disorders.

Our work on the *ATP6V0C* case described in chapter 3 exemplifies the potential of this extra work. This case was originally reported in our laboratory as a class 3 CNL and was excluded as part of the work in chapter 2 due to low pLI score (0.74) and lack of OMIM disease association (*108745). Although interesting, our interpretation of the role of *ATP6V0C* in the novel 16p13.3 microdeletion syndrome is mainly based on *in silico* analysis which also indicates that point variants in *ATP6V0C* could be disease causing. We have now identified three missense SNVs in *ATP6V0C* in patients recruited as part of the 100,000 Genomes project. *In silico* analysis using multiple prediction tools indicated a possibly damaging effect of these variants. Upon contacting submitting clinicians, we collected further information that reveals similar but more severe phenotype in the three patients including severe intellectual disability

(3/3) and epilepsy (2/3). We hypothesis that these *ATP6V0C* missense variants could result in a dominant negative effect that may be more deleterious than haploinsufficiency. Studies of *in vitro* and *in vivo* models are ongoing to investigate the potential functional impact of our *ATP6V0C* variants, understand the disease mechanism and investigate potential treatments.

5.2 Beyond clinical implications

Our work in chapter 4 has found that individuals with inherited (likely)pathogenic CNVs were significantly more likely to be living in areas of higher deprivation when compared to individuals with *de novo* (likely)pathogenic CNVs or with the general population of North-West England. This indicates that the inheritance of (likely)pathogenic CNVs and lower SES are correlated. This may appear counterintuitive as *de novo* CNVs are often associated with more severe phenotypes in comparison to inherited CNVs (Vulto-van Silfhout et al., 2013). More than half of the inherited CNVs in our cohort can be classed as recurrent with penetrance estimated to range between 10%-62% (Rosenfeld et al., 2013). Our cohort is almost completely paediatric (~98%) and based on reported penetrance, up to 72% of their carrier parents would be classed as not demonstrating a medical phenotype. SES of our paediatric cases reflects that of their parents and the high level of deprivation observed in our cohort suggests that being a carrier of a low penetrant CNV negatively impacts SES even in absence of a medical phenotype (or a milder one).

We speculate that even clinically unaffected individuals may have relatively lower cognitive abilities and therefore it is possible that the lower SES in families with medically relevant inherited (likely)pathogenic CNVs with milder phenotype is likely to be due to cumulative multi-generational consequences of relatively lower intellectual disabilities.

Early life deprivation can have a significant impact on the future life of young people (Belsky et al., 2019; Woods et al., 2005), which is in addition to any impact caused by the (likely)pathogenic CNV and its associated phenotype. The combined frequency of recurrent

CNVs in the UK is estimated to be at least 3.8% (Crawford et al., 2019) which means that these results could be of significance in the context of public health.

Our results demonstrate the need to quantify the secondary social and medical consequences of lower SES of individuals with (likely)pathogenic CNVs. These results have important implications for planning of medical and social services and provide unique insights into biological determinants of SES.

Fundamentally, the results of this work do not establish a causal relationship between CNVs and SES but only presents a correlation. The study is limited by the relatively small cohort and the limitation of a single geographic location (North west of England). It will be very interesting to try and expand this work and replicate the analysis in other regions within the UK. If the findings are replicated on bigger population within the UK (data from other Genomic Laboratory Hubs (GLHs)), this may indeed provide significant evidence for further considering the secondary social impact of CNVs. This study can also be replicated in other parts of the world. The relationship between genetic factors and SES is complex and evidence correlating various genetic variants with cognitive abilities, educational attainment, other social aspects and SES is variable and circular (Belsky et al., 2019; Figlio et al., 2017; Marioni et al., 2014; Queirós et al., 2015; Trzaskowski et al., 2014; von Stumm and Plomin, 2015; Woods et al., 2005) similar to the age-old debate of ‘nature v nurture’ .

Our novel results in this area demonstrate the need for larger studies to uncover the pattern of multi-morbidities in individuals with (likely)pathogenic CNVs.

6 References

Abbas, Y.M., Wu, D., Bueler, S.A., Robinson, C.V., Rubinstein, J.L., 2020. Structure of V-ATPase from the mammalian brain. Science 367, 1240–1246. https://doi.org/10.1126/science.aaz2924

Abramovs, N., Brass, A., Tassabehji, M., 2020. GeVIR is a continuous gene-level metric that uses variant distribution patterns to prioritize disease candidate genes. Nat. Genet. 52, 35–39. https://doi.org/10.1038/s41588-019-0560-2

Ahn, J.W., Bint, S., Bergbaum, A., Mann, K., Hall, R.P., Ogilvie, C.M., 2013. Array CGH as a first line diagnostic test in place of karyotyping for postnatal referrals - results from four years' clinical application for over 8,700 patients. Mol. Cytogenet. 6, 16. https://doi.org/10.1186/1755-8166-6-16

Alvarez-Pérez, M.A., Narayanan, S., Zeichner-David, M., Rodríguez Carmona, B., Arzate, H., 2006. Molecular cloning, expression and immunolocalization of a novel human cementum-derived protein (CP-23). Bone 38, 409–419. https://doi.org/10.1016/j.bone.2005.09.009

Azaiez, H., Booth, K.T., Bu, F., Huygen, P., Shibata, S.B., Shearer, A.E., Kolbe, D., Meyer, N., Black-Ziegelbein, E.A., Smith, R.J.H., 2014. TBC1D24 mutation causes autosomal-dominant nonsyndromic hearing loss. Hum. Mutat. 35, 819–823. https://doi.org/10.1002/humu.22557

Babbs, C., Brown, J., Horsley, S.W., Slater, J., Maifoshie, E., Kumar, S., Ooijevaar, P., Kriek, M., Dixon-McIver, A., Harteveld, C.L., Traeger-Synodinos, J., Wilkie, A.O.M., Higgs, D.R., Buckle, V.J., 2020. ATR-16 syndrome: mechanisms linking monosomy to phenotype. J. Med. Genet. 57, 414–421. https://doi.org/10.1136/jmedgenet-2019-106528

Bamshad, M.J., Nickerson, D.A., Chong, J.X., 2019. Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am. J. Hum. Genet. 105, 448–455. https://doi.org/10.1016/j.ajhg.2019.07.011

Banka, S., Cain, S.A., Carim, S., Daly, S.B., Urquhart, J.E., Erdem, G., Harris, J., Bottomley, M., Donnai, D., Kerr, B., Kingston, H., Superti-Furga, A., Unger, S., Ennis, H., Worthington, J., Herrick, A.L., Merry, C.L.R., Yue, W.W., Kielty, C.M., Newman, W.G., 2015. Leri's pleonosteosis, a congenital rheumatic disease, results from microduplication at 8q22.1 encompassing GDF6 and SDC2 and provides insight into systemic sclerosis pathogenesis. Ann. Rheum. Dis. 74, 1249–1256. https://doi.org/10.1136/annrheumdis-2013-204309

Bayascas, J.R., Wullschleger, S., Sakamoto, K., García-Martínez, J.M., Clacher, C., Komander, D., van Aalten, D.M.F., Boini, K.M., Lang, F., Lipina, C., Logie, L., Sutherland, C., Chudek, J.A., van Diepen, J.A., Voshol, P.J., Lucocq, J.M., Alessi, D.R., 2008. Mutation of the PDK1 PH domain inhibits protein kinase B/Akt, leading to small size and insulin resistance. Mol. Cell. Biol. 28, 3258–3272. https://doi.org/10.1128/MCB.02032-07

Belsky, D.W., Caspi, A., Arseneault, L., Corcoran, D.L., Domingue, B.W., Harris, K.M., Houts, R.M., Mill, J.S., Moffitt, T.E., Prinz, J., Sugden, K., Wertz, J., Williams, B., Odgers, C.L., 2019. Genetics and the geography of health, behaviour and attainment. Nat. Hum. Behav. 3, 576–586. https://doi.org/10.1038/s41562-019-0562-1

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., McBride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M.M., Picard, A., Temple, I.K., Thomas, P., Vazquez, M.-P., Vekemans, M., Roest Crollius, H., Hastie, N.D., Munnich, A., Etchevers, H.C., Pelet, A., Farlie, P.G., Fitzpatrick, D.R., Lyonnet, S., 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat. Genet. 41, 359–364. https://doi.org/10.1038/ng.329

Bergfeld, A.K., Pearce, O.M.T., Diaz, S.L., Pham, T., Varki, A., 2012. Metabolism of Vertebrate Amino Sugars with N-Glycolyl Groups. J. Biol. Chem. 287, 28865–28881. https://doi.org/10.1074/jbc.M112.363549

Blanchet, P., Bebin, M., Bruet, S., Cooper, G.M., Thompson, M.L., Duban-Bedu, B., Gerard, B., Piton, A., Suckno, S., Deshpande, C., Clowes, V., Vogt, J., Turnpenny, P., Williamson, M.P., Alembik, Y., Clinical Sequencing Exploratory Research Study Consortium, Deciphering Developmental Disorders Consortium, Glasgow, E., McNeill,

A., 2017. MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus. PLoS Genet. 13, e1006957. https://doi.org/10.1371/journal.pgen.1006957

Boone, P.M., Campbell, I.M., Baggett, B.C., Soens, Z.T., Rao, M.M., Hixson, P.M., Patel, A., Bi, W., Cheung, S.W., Lalani, S.R., 2013. Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. Genome Res. 23, 1383–1394.

Bradley, R.H., Corwyn, R.F., 2002. Socioeconomic Status and Child Development. Annu. Rev. Psychol. 53, 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Breuning, M.H., Dauwerse, H.G., Fugazza, G., Saris, J.J., Spruit, L., Wijnen, H., Tommerup, N., van der Hagen, C.B., Imaizumi, K., Kuroki, Y., van den Boogaard, M.J., de Pater, J.M., Mariman, E.C., Hamel, B.C., Himmelbauer, H., Frischauf, A.M., Stallings, R., Beverstock, G.C., van Ommen, G.J., Hennekam, R.C., 1993. Rubinstein-Taybi syndrome caused by submicroscopic deletions within 16p13.3. Am. J. Hum. Genet. 52, 249–254.

Bruel, A.-L., Nambot, S., Quéré, V., Vitobello, A., Thevenon, J., Assoum, M., Moutton, S., Houcinat, N., Lehalle, D., Jean-Marçais, N., Chevarin, M., Jouan, T., Poë, C., Callier, P., Tisserand, E., Philippe, C., Them, F.T.M., Duffourd, Y., Faivre, L., Thauvin-Robinet, C., 2019. Increased diagnostic and new genes identification outcome using research reanalysis of singleton exome sequencing. Eur. J. Hum. Genet. 27, 1519–1531. https://doi.org/10.1038/s41431-019-0442-1

Burghel, G.J., Khan, U., Lin, W.-Y., Whittaker, W., Banka, S., 2020. Presence of pathogenic copy number variants (CNVs) is correlated with socioeconomic status. J. Med. Genet. 57, 70–72. https://doi.org/10.1136/jmedgenet-2019-106292

Campeau, P.M., Kasperaviciute, D., Lu, J.T., Burrage, L.C., Kim, C., Hori, M., Powell, B.R., Stewart, F., Félix, T.M., van den Ende, J., Wisniewska, M., Kayserili, H., Rump, P., Nampoothiri, S., Aftimos, S., Mey, A., Nair, L.D.V., Begleiter, M.L., De Bie, I., Meenakshi, G., Murray, M.L., Repetto, G.M., Golabi, M., Blair, E., Male, A., Giuliano, F., Kariminejad, A., Newman, W.G., Bhaskar, S.S., Dickerson, J.E., Kerr, B., Banka, S., Giltay, J.C., Wieczorek, D., Tostevin, A., Wiszniewska, J., Cheung, S.W., Hennekam, R.C., Gibbs, R.A., Lee, B.H., Sisodiya, S.M., 2014. The genetic basis of DOORS syndrome: an exome-sequencing study. Lancet Neurol. 13, 44–58. https://doi.org/10.1016/S1474-4422(13)70265-5

Carvill, G.L., Weckhuysen, S., McMahon, J.M., Hartmann, C., Møller, R.S., Hjalgrim, H., Cook, J., Geraghty, E., O'Roak, B.J., Petrou, S., Clarke, A., Gill, D., Sadleir, L.G., Muhle, H., von Spiczak, S., Nikanorova, M., Hodgson, B.L., Gazina, E.V., Suls, A., Shendure, J., Dibbens, L.M., De Jonghe, P., Helbig, I., Berkovic, S.F., Scheffer, I.E., Mefford, H.C., 2014. GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. Neurology 82, 1245–1253. https://doi.org/10.1212/WNL.0000000000000291

Cheng, H., Dharmadhikari, A.V., Varland, S., Ma, N., Domingo, D., Kleyner, R., Rope, A.F., Yoon, M., Stray-Pedersen, A., Posey, J.E., Crews, S.R., Eldomery, M.K., Akdemir, Z.C., Lewis, A.M., Sutton, V.R., Rosenfeld, J.A., Conboy, E., Agre, K., Xia, F., Walkiewicz, M., Longoni, M., High, F.A., van Slegtenhorst, M.A., Mancini, G.M.S., Finnila, C.R., van Haeringen, A., den Hollander, N., Ruivenkamp, C., Naidu, S., Mahida, S., Palmer, E.E., Murray, L., Lim, D., Jayakar, P., Parker, M.J., Giusto, S., Stracuzzi, E., Romano, C., Beighley, J.S., Bernier, R.A., Küry, S., Nizon, M., Corbett, M.A., Shaw, M., Gardner, A., Barnett, C., Armstrong, R., Kassahn, K.S., Van Dijck, A., Vandeweyer, G., Kleefstra, T., Schieving, J., Jongmans, M.J., de Vries, B.B.A., Pfundt, R., Kerr, B., Rojas, S.K., Boycott, K.M., Person, R., Willaert, R., Eichler, E.E., Kooy, R.F., Yang, Y., Wu, J.C., Lupski, J.R., Arnesen, T., Cooper, G.M., Chung, W.K., Gecz, J., Stessman, H.A.F., Meng, L., Lyon, G.J., 2018. Truncating Variants in NAA15 Are Associated with Variable Levels of Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies. Am. J. Hum. Genet. 102, 985–994. https://doi.org/10.1016/j.ajhg.2018.03.004

Chiu, A.T.G., Pei, S.L.C., Mak, C.C.Y., Leung, G.K.C., Yu, M.H.C., Lee, S.L., Vreeburg, M., Pfundt, R., van der Burgt, I., Kleefstra, T., Frederic, T.M.-T., Nambot, S., Faivre, L., Bruel, A.-L., Rossi, M., Isidor, B., Küry, S., Cogne, B., Besnard, T., Willems, M., Reijnders, M.R.F., Chung, B.H.Y., 2018. Okur-Chung neurodevelopmental syndrome:

Eight additional cases with implications on phenotype and genotype expansion. Clin. Genet. 93, 880–890. https://doi.org/10.1111/cge.13196

Clark, A.M., DesMeules, M., Luo, W., Duncan, A.S., Wielgosz, A., 2009. Socioeconomic status and cardiovascular disease: risks and implications for care. Nat. Rev. Cardiol. 6, 712–722. https://doi.org/10.1038/nrcardio.2009.163

Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., Schuurs-Hoeijmakers, J.H., Hoischen, A., Pfundt, R., Krumm, N., Carvill, G.L., Li, D., Amaral, D., Brown, N., Lockhart, P.J., Scheffer, I.E., Alberti, A., Shaw, M., Pettinato, R., Tervo, R., de Leeuw, N., Reijnders, M.R.F., Torchia, B.S., Peeters, H., O'Roak, B.J., Fichera, M., Hehir-Kwa, J.Y., Shendure, J., Mefford, H.C., Haan, E., Gécz, J., de Vries, B.B.A., Romano, C., Eichler, E.E., 2014. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. 46, 1063–1071. https://doi.org/10.1038/ng.3092

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., Watts, N.A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C.W., Huang, Y., Brookings, T., Sharpe, T., Stone, M.R., Valkanas, E., Fu, J., Tiao, G., Laricchia, K.M., Ruano-Rubio, V., Stevens, C., Gupta, N., Cusick, C., Margolin, L., Taylor, K.D., Lin, H.J., Rich, S.S., Post, W.S., Chen, Y.-D.I., Rotter, J.I., Nusbaum, C., Philippakis, A., Lander, E., Gabriel, S., Neale, B.M., Kathiresan, S., Daly, M.J., Banks, E., MacArthur, D.G., Talkowski, M.E., 2020. A structural variation reference for medical and population genetics. Nature 581, 444–451. https://doi.org/10.1038/s41586-020-2287-8

Collins, R.L., Glessner, J.T., Porcu, E., Niestroj, L.-M., Ulirsch, J., Kellaris, G., Howrigan, D.P., Everett, S., Mohajeri, K., Nuttle, X., Lowther, C., Fu, J., Boone, P.M., Ullah, F., Samocha, K.E., Karczewski, K., Lucente, D., Consortium, E., Gusella, J.F., Finucane, H., Matyakhina, L., Aradhya, S., Meck, J., Lal, D., Neale, B.M., Hodge, J.C., Reymond, A., Kutalik, Z., Katsanis, N., Davis, E.E., Hakonarson, H., Sunyaev, S., Brand, H., Talkowski, M.E., 2021. A cross-disorder dosage sensitivity map of the human genome. medRxiv 2021.01.26.21250098. https://doi.org/10.1101/2021.01.26.21250098

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., 2010. Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., 2011. A copy number variation morbidity map of developmental delay. Nat. Genet. 43, 838–846.

Coppola, A., Cellini, E., Stamberger, H., Saarentaus, E., Cetica, V., Lal, D., Djémié, T., Bartnik-Glaska, M., Ceulemans, B., Helen Cross, J., Deconinck, T., Masi, S.D., Dorn, T., Guerrini, R., Hoffman-Zacharska, D., Kooy, F., Lagae, L., Lench, N., Lemke, J.R., Lucenteforte, E., Madia, F., Mefford, H.C., Morrogh, D., Nuernberg, P., Palotie, A., Schoonjans, A., Striano, P., Szczepanik, E., Tostevin, A., Vermeesch, J.R., Van Esch, H., Van Paesschen, W., Waters, J.J., Weckhuysen, S., Zara, F., Jonghe, P.D., Sisodiya, S.M., Marini, C., 2019. Diagnostic implications of genetic copy number variation in epilepsy plus. Epilepsia 60, 689–706. https://doi.org/10.1111/epi.14683

Costain, G., Jobling, R., Walker, S., Reuter, M.S., Snell, M., Bowdin, S., Cohn, R.D., Dupuis, L., Hewson, S., Mercimek-Andrews, S., Shuman, C., Sondheimer, N., Weksberg, R., Yoon, G., Meyn, M.S., Stavropoulos, D.J., Scherer, S.W., Mendoza-Londono, R., Marshall, C.R., 2018. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. Eur. J. Hum. Genet. 26, 740–744. https://doi.org/10.1038/s41431-018-0114-6

Coughlin, C.R., Scharer, G.H., Shaikh, T.H., 2012. Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. Genome Med. 4, 80.

Cox, A.M., McKevitt, C., Rudd, A.G., Wolfe, C.D., 2006. Socioeconomic status and stroke. Lancet Neurol. 5, 181–188. https://doi.org/10.1016/S1474-4422(06)70351-9

Crawford, K., Bracher-Smith, M., Owen, D., Kendall, K.M., Rees, E., Pardiñas, A.F., Einon, M., Escott-Price, V., Walters, J.T.R., O'Donovan, M.C., Owen, M.J., Kirov, G., 2019. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. J. Med. Genet. 56, 131–138. https://doi.org/10.1136/jmedgenet-2018-105477

Cuvertino, S., Stuart, H.M., Chandler, K.E., Roberts, N.A., Armstrong, R., Bernardini, L., Bhaskar, S., Callewaert, B., Clayton-Smith, J., Davalillo, C.H., Deshpande, C., Devriendt, K., Digilio, M.C., Dixit, A., Edwards, M., Friedman, J.M., Gonzalez-Meneses, A., Joss, S., Kerr, B., Lampe, A.K., Langlois, S., Lennon, R., Loget, P., Ma, D.Y.T., McGowan, R., Medt, M.D., O'Sullivan, J., Odent, S., Parker, M.J., Pebrel-Richard, C., Petit, F., Stark, Z., Stockler-Ipsiroglu, S., Tinschert, S., Vasudevan, P., Villa, O., White, S.M., Zahir, F.R., Woolf, A.S., Banka, S., 2017. ACTB Loss-of-Function Mutations Result in a Pleiotropic Developmental Disorder. Am. J. Hum. Genet. 101, 1021–1033. https://doi.org/10.1016/j.ajhg.2017.11.006

Das, J., Ingles, J., Bagnall, R.D., Semsarian, C., 2014. Determining pathogenicity of genetic variants in hypertrophic cardiomyopathy: importance of periodic reassessment. Genet. Med. 16, 286.

de Kok, Y.J., Vossenaar, E.R., Cremers, C.W., Dahl, N., Laporte, J., Jia Hu, L., Lacombe, D., Fischel-Ghodsian, N., Friedman, R.A., Parnes, L.S., 1996. Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. Hum. Mol. Genet. 5, 1229–1235.

de Leeuw, N., Dijkhuizen, T., Hehir-Kwa, J.Y., Carter, N.P., Feuk, L., Firth, H.V., Kuhn, R.M., Ledbetter, D.H., Martin, C.L., van Ravenswaaij-Arts, C., 2012. Diagnostic interpretation of array data using public databases and internet sources. Hum. Mutat. 33, 930–940.

de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B.A., Brunner, H.G., Veltman, J.A., Vissers, L.E.L.M., 2012. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 367, 1921–1929. https://doi.org/10.1056/NEJMoa1206524

De Rocker, N., Vergult, S., Koolen, D., Jacobs, E., Hoischen, A., Zeesman, S., Bang, B., Béna, F., Bockaert, N., Bongers, E.M., de Ravel, T., Devriendt, K., Giglio, S., Faivre, L., Joss, S., Maas, S., Marle, N., Novara, F., Nowaczyk, M.J.M., Peeters, H., Polstra, A., Roelens, F., Rosenberg, C., Thevenon, J., Tümer, Z., Vanhauwaert, S., Varvagiannis, K., Willaert, A., Willemsen, M., Willems, M., Zuffardi, O., Coucke, P., Speleman, F., Eichler, E.E., Kleefstra, T., Menten, B., 2015. Refinement of the critical 2p25.3 deletion region: the role of MYT1L in intellectual disability and obesity. Genet. Med. Off. J. Am. Coll. Med. Genet. 17, 460–466. https://doi.org/10.1038/gim.2014.124

Duru, N., Iseri, S.A.U., Selç, N., Tolun, A., 2010. Early-Onset Progressive Myoclonic Epilepsy With Dystonia Mapping to 16pter-p13.3. J. Neurogenet. 24, 207–215. https://doi.org/10.3109/01677063.2010.514368

El Mecky, J., Johansson, L., Plantinga, M., Fenwick, A., Lucassen, A., Dijkhuizen, T., van der Hout, A., Lyle, K., van Langen, I., 2019. Reinterpretation, reclassification, and its downstream effects: challenges for clinical laboratory geneticists. BMC Med. Genomics 12, 170. https://doi.org/10.1186/s12920-019-0612-6

Ellard, S., Baple, E.L., Callaway, A., Berry, I., Forrester, N., Turnbull, C., McMullan, D.J., 2020. ACGS best practice guidelines for variant classification in rare disease. Assoc Clin Genomic Sci ACGS.

Falace, A., Filipello, F., La Padula, V., Vanni, N., Madia, F., De Pietri Tonelli, D., de Falco, F.A., Striano, P., Dagna Bricarelli, F., Minetti, C., Benfenati, F., Fassio, A., Zara, F., 2010. TBC1D24, an ARF6-Interacting Protein, Is Mutated in Familial Infantile Myoclonic Epilepsy. Am. J. Hum. Genet. 87, 365–370. https://doi.org/10.1016/j.ajhg.2010.07.020

Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., Horton, R., Kerr, B., Kumar, D., Lehman, A., McKee, S., Morton, J., Parker, M.J., Rankin, J., Robertson, L., Temple, I.K., Clinical Assessment of the Utility of Sequencing and Evaluation as a Service (CAUSES) Study, Deciphering Developmental Disorders (DDD) Study, Banka, S., 2018. Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am. J. Hum. Genet. 102, 175–187. https://doi.org/10.1016/j.ajhg.2017.11.013

Figlio, D.N., Freese, J., Karbownik, K., Roth, J., 2017. Socioeconomic status and genetic influences on cognitive development. Proc. Natl. Acad. Sci. 114, 13441–13446. https://doi.org/10.1073/pnas.1708491114

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., Carter, N.P., 2009. DECIPHER: database of chromosomal

imbalance and phenotype in humans using ensembl resources. Am. J. Hum. Genet. 84, 524–533.
Frazer, K.A., Murray, S.S., Schork, N.J., Topol, E.J., 2009. Human genetic variation and its contribution to complex traits. Nat. Rev. Genet. 10, 241–251.
Hackman, D.A., Farah, M.J., 2009. Socioeconomic status and the developing brain. Trends Cogn. Sci. 13, 65–73. https://doi.org/10.1016/j.tics.2008.11.003
Hamanaka, K., Sugawara, Y., Shimoji, T., Nordtveit, T.I., Kato, M., Nakashima, M., Saitsu, H., Suzuki, T., Yamakawa, K., Aukrust, I., Houge, G., Mitsuhashi, S., Takata, A., Iwama, K., Alkanaq, A., Fujita, A., Imagawa, E., Mizuguchi, T., Miyake, N., Miyatake, S., Matsumoto, N., 2019. De novo truncating variants in PHF21A cause intellectual disability and craniofacial anomalies. Eur. J. Hum. Genet. EJHG 27, 378–383. https://doi.org/10.1038/s41431-018-0289-x
Hanscombe, K.B., Trzaskowski, M., Haworth, C.M.A., Davis, O.S.P., Dale, P.S., Plomin, R., 2012. Socioeconomic Status (SES) and Children's Intelligence (IQ): In a UK-Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0030320
Havrilla, J.M., Pedersen, B.S., Layer, R.M., Quinlan, A.R., 2019. A map of constrained coding regions in the human genome. Nat. Genet. 51, 88–95. https://doi.org/10.1038/s41588-018-0294-6
Hegde, M.R., Chin, E.L., Mulle, J.G., Okou, D.T., Warren, S.T., Zwick, M.E., 2008. Microarray-based mutation detection in the dystrophin gene. Hum. Mutat. 29, 1091.
Hegele, R.A., 2007. Copy-number variations and human disease. Am. J. Hum. Genet. 81, 414–415.
Hehir-Kwa, J.Y., Wieskamp, N., Webber, C., Pfundt, R., Brunner, H.G., Gilissen, C., de Vries, B.B., Ponting, C.P., Veltman, J.A., 2010. Accurate distinction of pathogenic from benign CNVs in mental retardation. PLoS Comput Biol 6, e1000752.
Hennekam, R.C., Tilanus, M., Hamel, B.C., Voshart-van Heeren, H., Mariman, E.C., van Beersum, S.E., van den Boogaard, M.J., Breuning, M.H., 1993. Deletion at chromosome 16p13.3 as a cause of Rubinstein-Taybi syndrome: clinical aspects. Am. J. Hum. Genet. 52, 255–262.
Henrichsen, C.N., Chaignat, E., Reymond, A., 2009. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18, R1–R8.
Hiatt, S.M., Amaral, M.D., Bowling, K.M., Finnila, C.R., Thompson, M.L., Gray, D.E., Lawlor, J.M.J., Cochran, J.N., Bebin, E.M., Brothers, K.B., East, K.M., Kelley, W.V., Lamb, N.E., Levy, S.E., Lose, E.J., Neu, M.B., Rich, C.A., Simmons, S., Myers, R.M., Barsh, G.S., Cooper, G.M., 2018. Systematic reanalysis of genomic data improves quality of variant interpretation. Clin. Genet. 94, 174–178. https://doi.org/10.1111/cge.13259
Hoffman-Andrews, L., 2017. The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. J. Law Biosci. 4, 648.
Huang, N., Lee, I., Marcotte, E.M., Hurles, M.E., 2010. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6, e1001154.
Hurles, M.E., Dermitzakis, E.T., Tyler-Smith, C., 2008. The functional impact of structural variation in humans. Trends Genet. TIG 24, 238–245. https://doi.org/10.1016/j.tig.2008.03.001
Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. Nat. Genet. 36, 949–951.
Ibn-Salem, J., Köhler, S., Love, M.I., Chung, H.-R., Huang, N., Hurles, M.E., Haendel, M., Washington, N.L., Smedley, D., Mungall, C.J., 2014. Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol. 15, 423.
Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A.,

Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kasher, P.R., Schertz, K.E., Thomas, M., Jackson, A., Annunziata, S., Ballesta-Martinez, M.J., Campeau, P.M., Clayton, P.E., Eaton, J.L., Granata, T., 2016. Small 6q16. 1 deletions encompassing POU3F2 cause susceptibility to obesity and variable developmental delay with intellectual disability. Am. J. Hum. Genet. 98, 363–372.

Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., South, S.T., 2011. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. 13, 680–685. https://doi.org/10.1097/GIM.0b013e3182217a3a

Kendall, K.M., Bracher-Smith, M., Fitzpatrick, H., Lynham, A., Rees, E., Escott-Price, V., Owen, M.J., O'Donovan, M.C., Walters, J.T.R., Kirov, G., 2019. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. Br. J. Psychiatry 214, 297–304. https://doi.org/10.1192/bjp.2018.301

Kim, H.-G., Rosenfeld, J.A., Scott, D.A., Bénédicte, G., Labonne, J.D., Brown, J., McGuire, M., Mahida, S., Naidu, S., Gutierrez, J., Lesca, G., des Portes, V., Bruel, A.-L., Sorlin, A., Xia, F., Capri, Y., Muller, E., McKnight, D., Torti, E., Rüschendorf, F., Hummel, O., Islam, Z., Kolatkar, P.R., Layman, L.C., Ryu, D., Kong, I.-K., Madan-Khetarpal, S., Kim, C.-H., 2019. Disruption of PHF21A causes syndromic intellectual disability with craniofacial anomalies, epilepsy, hypotonia, and neurobehavioral problems including autism. Mol. Autism 10, 35. https://doi.org/10.1186/s13229-019-0286-0

Koemans, T.S., Kleefstra, T., Chubak, M.C., Stone, M.H., Reijnders, M.R.F., de Munnik, S., Willemsen, M.H., Fenckova, M., Stumpel, C.T.R.M., Bok, L.A., Sifuentes Saenz, M., Byerly, K.A., Baughn, L.B., Stegmann, A.P.A., Pfundt, R., Zhou, H., van Bokhoven, H., Schenck, A., Kramer, J.M., 2017. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genet. 13, e1006864. https://doi.org/10.1371/journal.pgen.1006864

Ku, C.S., Loy, E.Y., Salim, A., Pawitan, Y., Chia, K.S., 2010. The discovery of human genetic variations and their use as disease markers: past, present and future. J. Hum. Genet. 55, 403–415.

Küry, S., Besnard, T., Ebstein, F., Khan, T.N., Gambin, T., Douglas, J., Bacino, C.A., Craigen, W.J., Sanders, S.J., Lehmann, A., Latypova, X., Khan, K., Pacault, M., Sacharow, S., Glaser, K., Bieth, E., Perrin-Sabourin, L., Jacquemont, M.-L., Cho, M.T., Roeder, E., Denommé-Pichon, A.-S., Monaghan, K.G., Yuan, B., Xia, F., Simon, S., Bonneau, D., Parent, P., Gilbert-Dussardier, B., Odent, S., Toutain, A., Pasquier, L., Barbouth, D., Shaw, C.A., Patel, A., Smith, J.L., Bi, W., Schmitt, S., Deb, W., Nizon, M., Mercier, S., Vincent, M., Rooryck, C., Malan, V., Briceño, I., Gómez, A., Nugent, K.M., Gibson, J.B., Cogné, B., Lupski, J.R., Stessman, H.A.F., Eichler, E.E., Retterer, K., Yang, Y., Redon, R., Katsanis, N., Rosenfeld, J.A., Kloetzel, P.-M., Golzio, C., Bézieau, S., Stankiewicz, P., Isidor, B., 2017. De Novo Disruption of the Proteasome Regulatory Subunit PSMD12 Causes a Syndromic Neurodevelopmental Disorder. Am. J. Hum. Genet. 100, 352–363. https://doi.org/10.1016/j.ajhg.2017.01.003

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., Maglott, D.R., 2016. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44, D862-868. https://doi.org/10.1093/nar/gkv1222

Lang, S.-J., Abel, G.A., Mant, J., Mullis, R., 2016. Impact of socioeconomic deprivation on screening for cardiovascular disease risk in a primary prevention population: a cross-sectional study. BMJ Open 6, e009984. https://doi.org/10.1136/bmjopen-2015-009984

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

Lessard, S., Manning, A.K., Low-Kam, C., Auer, P.L., Giri, A., Graff, M., Schurmann, C., Yaghootkar, H., Luan, J., Esko, T., 2016. Testing the role of predicted gene knockouts in human anthropometric trait variation. Hum. Mol. Genet. ddw055.

Lovato, D.V., Herai, R.R., Pignatari, G.C., Beltrão-Braga, P.C.B., 2019. The Relevance of Variants With Unknown Significance for Autism Spectrum Disorder Considering the Genotype–Phenotype Interrelationship. Front. Psychiatry 10. https://doi.org/10.3389/fpsyt.2019.00409

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025.

Lupiáñez, D.G., Spielmann, M., Mundlos, S., 2016. Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. 32, 225–237.

Lupski, J.R., Stankiewicz, P., 2005. Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. PLOS Genet. 1, e49. https://doi.org/10.1371/journal.pgen.0010049

Lüthy, K., Mei, D., Fischer, B., De Fusco, M., Swerts, J., Paesmans, J., Parrini, E., Lubarr, N., Meijer, I.A., Mackenzie, K.M., Lee, W.-T., Cittaro, D., Aridon, P., Schoovaerts, N., Versées, W., Verstreken, P., Casari, G., Guerrini, R., 2019. TBC1D24-TLDc-related epilepsy exercise-induced dystonia: rescue by antioxidants in a disease model. Brain J. Neurol. 142, 2319–2335. https://doi.org/10.1093/brain/awz175

MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., Scherer, S.W., 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986–D992.

Marioni, R.E., Davies, G., Hayward, C., Liewald, D., Kerr, S.M., Campbell, A., Luciano, M., Smith, B.H., Padmanabhan, S., Hocking, L.J., Hastie, N.D., Wright, A.F., Porteous, D.J., Visscher, P.M., Deary, I.J., 2014. Molecular genetic contributions to socioeconomic status and intelligence. Intelligence 44, 26–32. https://doi.org/10.1016/j.intell.2014.02.006

Masson, J., Demily, C., Chatron, N., Labalme, A., Rollat-Farnier, P.-A., Schluth-Bolard, C., Gilbert-Dussardier, B., Giuliano, F., Touraine, R., Tordjman, S., Verloes, A., Testa, G., Sanlaville, D., Edery, P., Lesca, G., Rossi, M., 2019. Molecular investigation, using chromosomal microarray and whole exome sequencing, of six patients affected by Williams Beuren syndrome and Autism Spectrum Disorder. Orphanet J. Rare Dis. 14, 121. https://doi.org/10.1186/s13023-019-1094-5

McRae, J.F., Clayton, S., Fitzgerald, T.W., Kaplanis, J., Prigmore, E., Rajan, D., Sifrim, A., Aitken, S., Akawi, N., Alvi, M., 2017. Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433–433.

Mefford, H.C., 2014. CNVs in Epilepsy. Curr. Genet. Med. Rep. 2, 162–167. https://doi.org/10.1007/s40142-014-0046-6

Mi, H., Muruganujan, A., Ebert, D., Huang, X., Thomas, P.D., 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47, D419–D426. https://doi.org/10.1093/nar/gky1038

Mighton, C., Shickh, S., Uleryk, E., Pechlivanoglou, P., Bombard, Y., 2020. Clinical and psychological outcomes of receiving a variant of uncertain significance from multigene panel testing or genomic sequencing: a systematic review and meta-analysis. Genet. Med. 1–12. https://doi.org/10.1038/s41436-020-00957-2

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. 86, 749–764.

Mucha, B.E., Banka, S., Ajeawung, N.F., Molidperee, S., Chen, G.G., Koenig, M.K., Adejumo, R.B., Till, M., Harbord, M., Perrier, R., Lemyre, E., Boucher, R.-M., Skotko, B.G., Waxler, J.L., Thomas, M.A., Hodge, J.C., Gecz, J., Nicholl, J., McGregor, L., Linden, T., Sisodiya, S.M., Sanlaville, D., Cheung, S.W., Ernst, C., Campeau, P.M., 2019. A new microdeletion syndrome involving TBC1D24, ATP6V0C , and PDPK1 causes epilepsy, microcephaly, and developmental delay. Genet. Med. 21, 1058–1064. https://doi.org/10.1038/s41436-018-0290-3

Okur, V., Cho, M.T., Henderson, L., Retterer, K., Schneider, M., Sattler, S., Niyazov, D., Azage, M., Smith, S., Picker, J., Lincoln, S., Tarnopolsky, M., Brady, L., Bjornsson, H.T., Applegate, C., Dameron, A., Willaert, R., Baskin, B., Juusola, J., Chung, W.K., 2016. De novo mutations in CSNK2A1 are associated with neurodevelopmental

abnormalities and dysmorphic features. Hum. Genet. 135, 699–705. https://doi.org/10.1007/s00439-016-1661-y

Owen, C.I., Bowden, R., Parker, M.J., Patterson, Jo, Patterson, Joan, Price, S., Sarkar, A., Castle, B., Deshpande, C., Splitt, M., Ghali, N., Dean, J., Green, A.J., Crosby, C., Deciphering Developmental Disorders Study, Tatton-Brown, K., 2018. Extending the phenotype associated with the CSNK2A1-related Okur-Chung syndrome-A clinical study of 11 individuals. Am. J. Med. Genet. A. 176, 1108–1114. https://doi.org/10.1002/ajmg.a.38610

Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S., Goldstein, D.B., 2013. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet 9, e1003709.

Queirós, F.C., Wehby, G.L., Halpern, C.T., 2015. Developmental Disabilities and Socioeconomic Outcomes in Young Adulthood. Public Health Rep. 130, 213–221.

Rehman, A.U., Santos-Cortez, R.L.P., Morell, R.J., Drummond, M.C., Ito, T., Lee, K., Khan, A.A., Basra, M.A.R., Wasif, N., Ayub, M., Ali, R.A., Raza, S.I., Nickerson, D.A., Shendure, J., Bamshad, M., Riazuddin, Saima, Billington, N., Khan, S.N., Friedman, P.L., Griffith, A.J., Ahmad, W., Riazuddin, Sheikh, Leal, S.M., Friedman, T.B., 2014. Mutations in TBC1D24, a Gene Associated With Epilepsy, Also Cause Nonsyndromic Deafness DFNB86. Am. J. Hum. Genet. 94, 144–152. https://doi.org/10.1016/j.ajhg.2013.12.004

Rice, A.M., McLysaght, A., 2017. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. Nat. Commun. 8, 14366. https://doi.org/10.1038/ncomms14366

Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., Pineda-Alvarez, D., Aradhya, S., Martin, C.L., 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. 22, 245–257. https://doi.org/10.1038/s41436-019-0686-8

Rosenfeld, J.A., Coe, B.P., Eichler, E.E., Cuckle, H., Shaffer, L.G., 2013. Estimates of penetrance for recurrent pathogenic copy-number variations. Genet. Med. Off. J. Am. Coll. Med. Genet. 15, 478–481. https://doi.org/10.1038/gim.2012.164

Russell, M.W., Moldenhauer, J.S., Rychik, J., Burnham, N.B., Zullo, E., Parry, S.I., Simmons, R.A., Elovitz, M.A., Nicolson, S.C., Linn, R.L., Johnson, M.P., Yu, S., Sampson, M.G., Hakonarson, H., Gaynor, J.W., 2019. Effect of parental origin of damaging variants in pro-angiogenic genes on fetal growth in patients with congenital heart defects: Data and analyses. Data Brief 25. https://doi.org/10.1016/j.dib.2019.104311

Salfati, E.L., Spencer, E.G., Topol, S.E., Muse, E.D., Rueda, M., Lucas, J.R., Wagner, G.N., Campman, S., Topol, E.J., Torkamani, A., 2019. Re-analysis of whole-exome sequencing data uncovers novel diagnostic variants and improves molecular diagnostic yields for sudden death and idiopathic diseases. Genome Med. 11, 83. https://doi.org/10.1186/s13073-019-0702-2

Sansovi, I., Ivankov, A.-M., Bobinec, A., Kero, M., Barii, I., 2017. Chromosomal microarray in clinical diagnosis: a study of 337 patients with congenital anomalies and developmental delays or intellectual disability. Croat. Med. J. 58, 231–238.

Schaaf, C.P., Wiszniewska, J., Beaudet, A.L., 2011. Copy number and SNP arrays in clinical diagnostics. Annu. Rev. Genomics Hum. Genet. 12, 25–51.

Schanze, I., Bunt, J., Lim, J.W.C., Schanze, D., Dean, R.J., Alders, M., Blanchet, P., Attié-Bitach, T., Berland, S., Boogert, S., Boppudi, S., Bridges, C.J., Cho, M.T., Dobyns, W.B., Donnai, D., Douglas, J., Earl, D.L., Edwards, T.J., Faivre, L., Fregeau, B., Genevieve, D., Gérard, M., Gatinois, V., Holder-Espinasse, M., Huth, S.F., Izumi, K., Kerr, B., Lacaze, E., Lakeman, P., Mahida, S., Mirzaa, G.M., Morgan, S.M., Nowak, C., Peeters, H., Petit, F., Pilz, D.T., Puechberty, J., Reinstein, E., Rivière, J.-B., Santani, A.B., Schneider, A., Sherr, E.H., Smith-Hicks, C., Wieland, I., Zackai, E., Zhao, X., Gronostajski, R.M., Zenker, M., Richards, L.J., 2018. NFIB Haploinsufficiency Is Associated with Intellectual Disability and Macrocephaly. Am. J. Hum. Genet. 103, 752–768. https://doi.org/10.1016/j.ajhg.2018.10.006

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., 2004. Large-scale copy number polymorphism in the human genome. Science 305, 525–528.

Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'arcy, M., 2009. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. 19, 1682–1690.

Shavers, V.L., 2007. Measurement of socioeconomic status in health disparities research. J. Natl. Med. Assoc. 99, 1013–1023.

Singh-Manoux, A., Richards, M., Marmot, M., 2005. Socioeconomic position across the lifecourse: how does it relate to cognitive function in mid-life? Ann. Epidemiol. 15, 572–578. https://doi.org/10.1016/j.annepidem.2004.10.007

South, S.T., Lee, C., Lamb, A.N., Higgins, A.W., Kearney, H.M., 2013. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genet. Med. 15, 901–909.

Spielmann, M., Lupiáñez, D.G., Mundlos, S., 2018. Structural variation in the 3D genome. Nat. Rev. Genet. 1.

Spielmann, M., Mundlos, S., 2013. Structural variations, the regulatory landscape of the genome and their alteration in human disease. Bioessays 35, 533–543.

Stankiewicz, P., Khan, T.N., Szafranski, P., Slattery, L., Streff, H., Vetrini, F., Bernstein, J.A., Brown, C.W., Rosenfeld, J.A., Rednam, S., Scollon, S., Bergstrom, K.L., Parsons, D.W., Plon, S.E., Vieira, M.W., Quaio, C.R.D.C., Baratela, W.A.R., Acosta Guio, J.C., Armstrong, R., Mehta, S.G., Rump, P., Pfundt, R., Lewandowski, R., Fernandes, E.M., Shinde, D.N., Tang, S., Hoyer, J., Zweier, C., Reis, A., Bacino, C.A., Xiao, R., Breman, A.M., Smith, J.L., Katsanis, N., Bostwick, B., Popp, B., Davis, E.E., Yang, Y., 2017. Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and Dysmorphic Features. Am. J. Hum. Genet. 101, 503–515. https://doi.org/10.1016/j.ajhg.2017.08.014

Stankiewicz, P., Lupski, J.R., 2010. Structural variation in the human genome and its role in disease. Annu. Rev. Med. 61, 437–455.

Stankiewicz, P., Lupski, J.R., 2002. Genome architecture, rearrangements and genomic disorders. Trends Genet. TIG 18, 74–82. https://doi.org/10.1016/s0168-9525(02)02592-1

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S.T., Abeysinghe, S., Krawczak, M., Cooper, D.N., 2003. Human Gene Mutation Database (HGMD): 2003 update. Hum. Mutat. 21, 577–581. https://doi.org/10.1002/humu.10212

Stessman, H.A.F., Xiong, B., Coe, B.P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., Vives, L., Lin, J., Turner, T.N., Santen, G., Ruivenkamp, C., Kriek, M., van Haeringen, A., Aten, E., Friend, K., Liebelt, J., Barnett, C., Haan, E., Shaw, M., Gecz, J., Anderlid, B.-M., Nordgren, A., Lindstrand, A., Schwartz, C., Kooy, R.F., Vandeweyer, G., Helsmoortel, C., Romano, C., Alberti, A., Vinci, M., Avola, E., Giusto, S., Courchesne, E., Pramparo, T., Pierce, K., Nalabolu, S., Amaral, D.G., Scheffer, I.E., Delatycki, M.B., Lockhart, P.J., Hormozdiari, F., Harich, B., Castells-Nobau, A., Xia, K., Peeters, H., Nordenskjöld, M., Schenck, A., Bernier, R.A., Eichler, E.E., 2017. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat. Genet. 49, 515–526. https://doi.org/10.1038/ng.3792

Strachan, T., Read, A., 2011. Human Molecular Genetics (New York: Garland Science, Taylor & Francis Group).

Sun, Y., Xiang, J., Liu, Y., Chen, S., Yu, J., Peng, J., Liu, Z., Chen, L., Sun, J., Yang, Yun, Yang, Yaping, Zhou, Y., Peng, Z., 2019. Increased diagnostic yield by reanalysis of data from a hearing loss gene panel. BMC Med. Genomics 12, 76. https://doi.org/10.1186/s12920-019-0531-6

Tan, N.B., Stapleton, R., Stark, Z., Delatycki, M.B., Yeung, A., Hunter, M.F., Amor, D.J., Brown, N.J., Stutterd, C.A., McGillivray, G., Yap, P., Regan, M., Chong, B., Fanjul Fernandez, M., Marum, J., Phelan, D., Pais, L.S., White, S.M., Lunke, S., Tan, T.Y., 2020. Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review. Mol. Genet. Genomic Med. 8. https://doi.org/10.1002/mgg3.1508

The GTex Consortium, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110

Tinker, R.J., Burghel, G.J., Garg, S., Steggall, M., Cuvertino, S., Banka, S., 2020. Haploinsufficiency of ATP6V0C possibly underlies 16p13.3 deletions that cause microcephaly, seizures, and neurodevelopmental disorder. Am. J. Med. Genet. A. https://doi.org/10.1002/ajmg.a.61905

Trinh, J., Hüning, I., Budler, N., Hingst, V., Lohmann, K., Gillessen-Kaesbach, G., 2017. A novel de novo mutation in CSNK2A1: reinforcing the link to neurodevelopmental abnormalities and dysmorphic features. J. Hum. Genet. 62, 1005–1006. https://doi.org/10.1038/jhg.2017.73

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., Dale, P.S., Plomin, R., 2014. Genetic influence on family socioeconomic status and children's intelligence. Intelligence 42, 83–88. https://doi.org/10.1016/j.intell.2013.11.002

Veltman, J.A., Brunner, H.G., 2012. De novo mutations in human genetic disease. Nat. Rev. Genet. 13, 565–575. https://doi.org/10.1038/nrg3241

Veltman, J.A., Brunner, H.G., 2010. Understanding variable expressivity in microdeletion syndromes. Nat. Genet. 42, 192–193. https://doi.org/10.1038/ng0310-192

von Stumm, S., Plomin, R., 2015. Socioeconomic status and the growth of intelligence from infancy through adolescence. Intelligence 48, 30–36. https://doi.org/10.1016/j.intell.2014.10.002

Vulto-van Silfhout, A.T., Hehir-Kwa, J.Y., van Bon, B.W.M., Schuurs-Hoeijmakers, J.H.M., Meader, S., Hellebrekers, C.J.M., Thoonen, I.J.M., de Brouwer, A.P.M., Brunner, H.G., Webber, C., Pfundt, R., de Leeuw, N., de Vries, B.B.A., 2013. Clinical significance of de novo and inherited copy-number variation. Hum. Mutat. 34, 1679–1687. https://doi.org/10.1002/humu.22442

Wang, J., Geng, L., 2019. Effects of Socioeconomic Status on Physical and Psychological Health: Lifestyle as a Mediator. Int. J. Environ. Res. Public. Health 16. https://doi.org/10.3390/ijerph16020281

White, J., Rehkopf, D., Mortensen, L.H., 2016. Trends in Socioeconomic Inequalities in Body Mass Index, Underweight and Obesity among English Children, 2007-2008 to 2011-2012. PloS One 11, e0147614. https://doi.org/10.1371/journal.pone.0147614

Woods, L.M., Rachet, B., Riga, M., Stone, N., Shah, A., Coleman, M.P., 2005. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. J. Epidemiol. Community Health 59, 115–120. https://doi.org/10.1136/jech.2003.013003

Wright, C.F., McRae, J.F., Clayton, S., Gallone, G., Aitken, S., FitzGerald, T.W., Jones, P., Prigmore, E., Rajan, D., Lord, J., Sifrim, A., Kelsell, R., Parker, M.J., Barrett, J.C., Hurles, M.E., FitzPatrick, D.R., Firth, H.V., 2018. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. Genet. Med. 20, 1216–1223. https://doi.org/10.1038/gim.2017.246

Wullschleger, S., Wasserman, D.H., Gray, A., Sakamoto, K., Alessi, D.R., 2011. Role of TAPP1 and TAPP2 adaptor binding to PtdIns(3,4)P2 in regulating insulin sensitivity defined by knock-in analysis. Biochem. J. 434, 265–274. https://doi.org/10.1042/BJ20102012

Xu, B., Roos, J.L., Levy, S., Van Rensburg, E., Gogos, J.A., Karayiorgou, M., 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat. Genet. 40, 880–885.

Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., Ichida, F., Joo, K., Kimura, M., Imamura, S., Kamatani, N., Momma, K., Takao, A., Nakazawa, M., Shimizu, N., Matsuoka, R., 2003. Role of TBX1 in human del22q11.2 syndrome. Lancet Lond. Engl. 362, 1366–1373. https://doi.org/10.1016/s0140-6736(03)14632-6

Yang, Y., Kong, W., Xia, Z., Xiao, L., Wang, S., 2016. Regulation mechanism of PDK1 on macrophage metabolism and function. Cell Biochem. Funct. 34, 546–553. https://doi.org/10.1002/cbf.3235

Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W., 2015. A copy number variation map of the human genome. Nat. Rev. Genet. 16, 172–183.

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R., 2009. Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. 10, 451–481.

Zhang, F., Lupski, J.R., 2015. Non-coding genetic variants in human disease. Hum. Mol. Genet. 24, R102–R110.
Zhang, L., Hu, L., Chai, Y., Pang, X., Yang, T., Wu, H., 2014. A Dominant Mutation in the Stereocilia-Expressing Gene TBC1D24 is a Probable Cause for Nonsyndromic Hearing Impairment. Hum. Mutat. 35, 814–818. https://doi.org/10.1002/humu.22558
Ziegler, A., Colin, E., Goudenège, D., Bonneau, D., 2019. A snapshot of some pLI score pitfalls. Hum. Mutat. 40, 839–841. https://doi.org/10.1002/humu.23763

7 Appendices

7.1 Appendix A: PGDip (DClinSci Module A) results

This thesis forms module C2 of the University of Manchester Higher Specialist Scientific Training (HSST). Module A included five taught units which were examined through the submission of two written essays for each unit. The final marks as ratified by the university of Manchester Board of Examiners are presented below.

MANCHESTER 1824
The University of Manchester
Alliance Manchester Business School

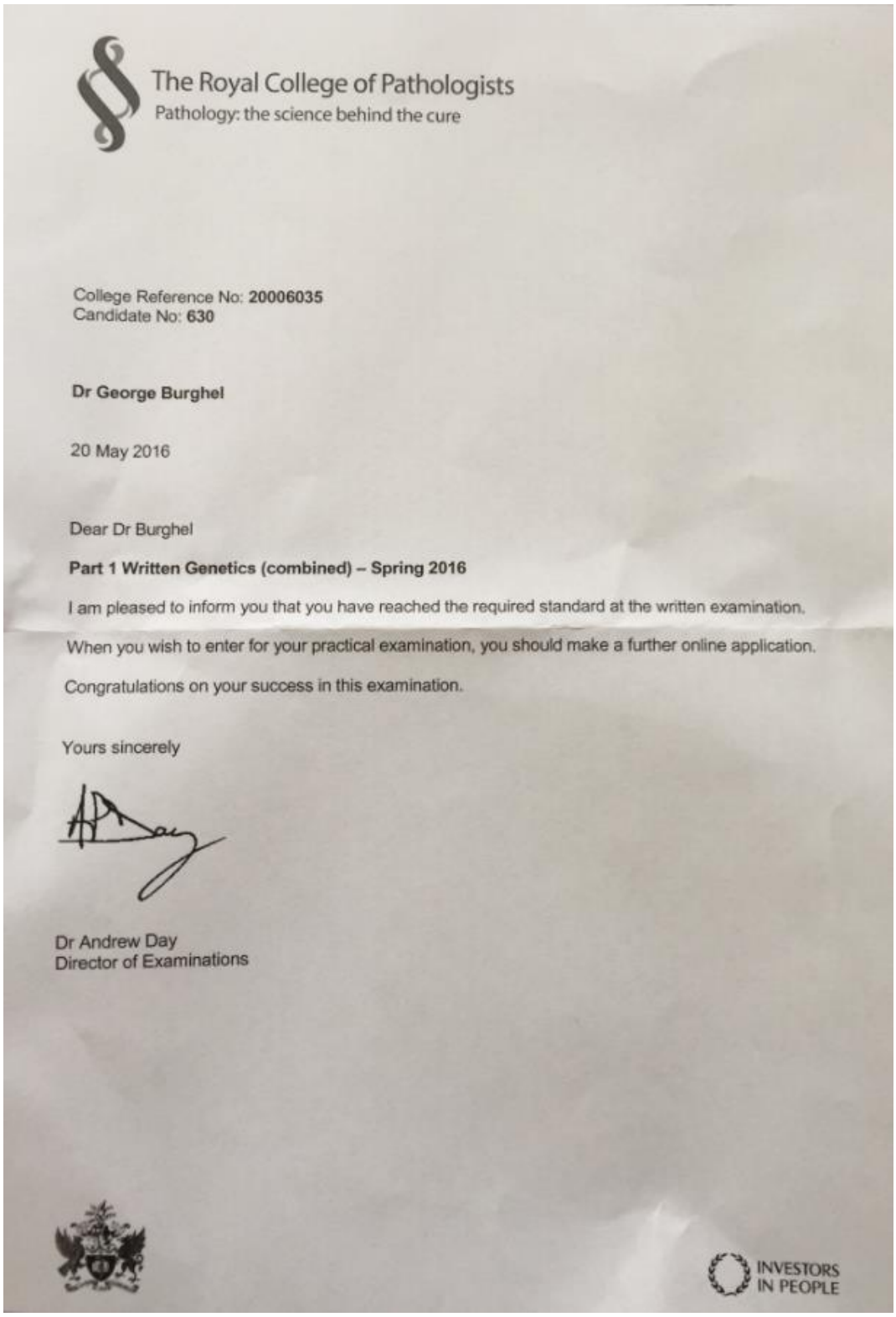
PGDip Leadership & Management in the Healthcare Sciences
Unit marks ratified by Board of Examiners, November 2017

Trainee name: George Burghel
Student ID: 9844218
Award: PG Credit

Unit	Unit Title	Mark	Credits
BMAN73511	Unit A1 Professionalism and Professional Development in the Healthcare Environment	69% **Pass**	**30**
BMAN73522	Unit A2 Theoretical Foundations of Leadership	82% **Pass**	**20**
BMAN73531	Unit A3 Personal and Professional Development to Enhance Performance	70% **Pass**	**30**
BMAN73542	Unit A4 Leadership and Quality Improvement in the Clinical and Scientific Environment	84% **Pass**	**20**
BMAN73550	Unit A5 Research and Innovation in Health and Social Care	75% **Pass**	**20**
			120 / 120

7.2 Appendix B: FRCPath exam results (DClinSci module B)

7.2.1 Appendix B1: FRCPath Part 1 theory



The Royal College of Pathologists
Pathology: the science behind the cure

College Reference No: **20006035**
Candidate No: **630**

Dr George Burghel

20 May 2016

Dear Dr Burghel

Part 1 Written Genetics (combined) – Spring 2016

I am pleased to inform you that you have reached the required standard at the written examination.

When you wish to enter for your practical examination, you should make a further online application.

Congratulations on your success in this examination.

Yours sincerely

Dr Andrew Day
Director of Examinations

INVESTORS IN PEOPLE

7.2.2 Appendix B2: FRCPath Part 1 practical

The Royal College of Pathologists
Pathology: the science behind the cure

College Reference No: **20006035**
Candidate No: **504**

George Burghel

24 November 2017

Dear Dr Burghel

FRCPath Part 1 Practical Examination in Genetics (Combined)– Autumn 2017

I am pleased to inform you that you have satisfied the Examiners in the Part 1 Examination. You have therefore completed this part of the examination and you are now eligible to become an Associate of the College.

Council will formally ratify the nominations to Associateship on **1 February 2018**. Please complete and return the enclosed consent form and address slip to the Membership Department, in the envelope provided, as soon as possible and no later than **Monday 8 January 2018.**

The subscription year runs from January and once you have returned your consent form an invoice will be sent out for your subscription due. This will take into account any fee you may have already paid and pro-rata rates for the year. For current rates please refer to our website(https://www.rcpath.org/discover-pathology/membership/subscription-fees-2017.html)

After Council has ratified the results, your certificate will be sent out to you with your full name as displayed at the top of this letter by the Membership Department. If there is an error with your name please notify the Examinations Department in writing, at exams@rcpath.org, as soon as possible.

Please note the Associate grade holds no 'post-nominals'. If you do not wish to become an Associate of the College please indicate on the enclosed form or e-mail the Membership department on membership@rcpath.org

Details of the Part 2 Examinations can be found on the Examinations page of our website.

Congratulations on your success in this examination.

Yours sincerely

Dr Andrew Day
Director of Examinations

Enc: Consent form and address slip
Envelope addressed to the Membership Department

INVESTORS IN PEOPLE

7.2.3 Appendix B3: FRCPath Part 2 oral exam

The Royal College of Pathologists
Pathology: the science behind the cure

College Reference Number: **20006035**
Candidate Number: **183**

George Burghel

24 May 2019

Dear Dr Burghel

FRCPath Part 2 Oral Examination in Genetics (Old Format) – Spring 2019

I am pleased to inform you that you have satisfied the Examiners in the Part 2 Examination.

However, as you are aware, you are not yet eligible to become a Fellow of The Royal College of Pathologists as your Part 2 Project has not been approved.

We look forward to receiving the project in due course. If you have any queries about your project please contact exams@rcpath.org.

Congratulations on your success in this examination.

Yours sincerely

Dr Andrew Day
Clinical Director of Examinations

7.3 Appendix C: DClinSci Module C1 examination feedback

MANCHESTER 1824

The University of Manchester

C1 Oral Presentation Feedback

Name of Trainee: George Burghel

Date of assessment: 30/09/17

Specialism: ____ Genetics___

Innovation Title: Copy number variants

Assessment criteria:

- Quality and clarity of explanation of the innovation for a lay audience (awareness of the use of jargon, scientific language and acronyms)
- Synthesis of relevant scientific evidence for a lay audience
- Ability to persuade a lay audience of the merits (or otherwise) of the innovation and its potential role in healthcare science services
- Style of presentation (slides, delivery; body language, eye contact, voice, confidence) and appropriateness for a lay audience
- Demonstration of values, attitudes and behaviours expected of a leader in clinical science

Summary comments for feedback:

What was good:
This was an excellent presentation. You were able to explain complex concepts clearly and without slipping into jargon. Your slides were well paced and very clear, you made good use of diagrams, we particularly liked your opening slide of the genome which was a clear visual representation. We were impressed with both the style and content of your delivery and enjoyed your talk very much. You answered the questions confidently and knowledgeably and although your innovation is on a more scientifically focused topic you clearly demonstrated your patient-centred focus and leadership qualities. All assessors noted your passion and enthusiasm for your work which came across very well.

What could be improved:
Nothing noted by any assessor.

Overall:
An excellent talk that met all of the above criteria

Recommendation: Pass

7.4 Appendix D: Further Qualifications and HSST relevant courses

7.4.1 Appendix D1 – Level 5 Certificate in Leadership

ilm

Level 5 Certificate in Leadership

600/2260/7

Is awarded to

George Burghel

Through

Phoenix Consultancy USA Inc

Awarded by
The City and Guilds of London Institute

Awarded 26 October 2016

Chris Jones
Director-General
The City and Guilds of London Institute

John Yates
Group Director
ILM

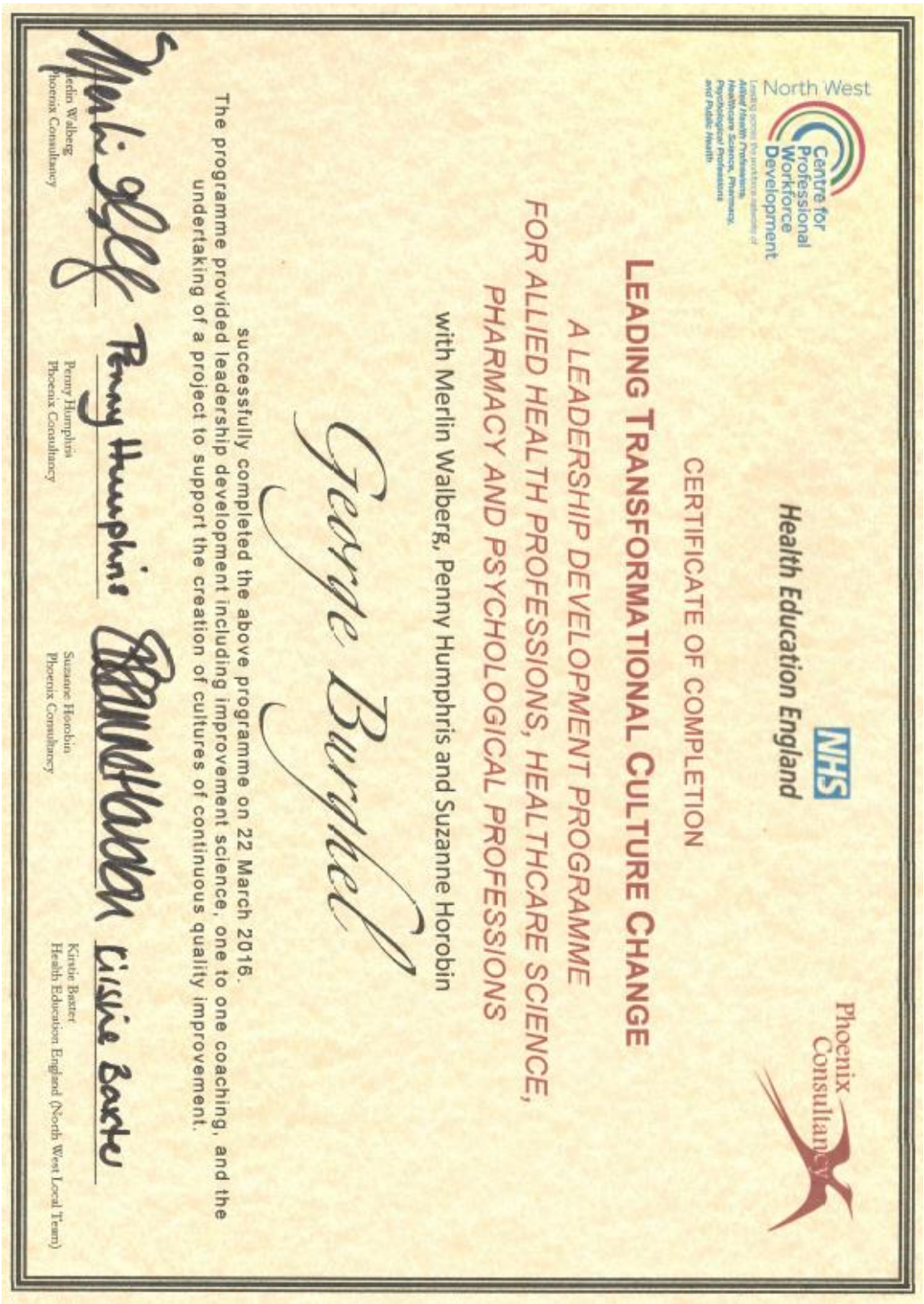
261016/8322-90/821690/LCV2949/M/15/11/82
5501842436/20

The City and Guilds of London Institute founded 1878 and incorporated by Royal Charter 1900.
ILM is a City & Guilds Group business.

Regulated by
Ofqual

Cardiff dyfarnu cydnabyddedig
Cymwysterau Cymru Qualifications Wales
Recognised awarding body

7.4.2 Appendix D2 – Leading Transformational Culture Change



North West Centre for Professional Workforce Development

Leading across the workforce agenda of Allied Health Professions, Healthcare Science, Pharmacy, Psychological Professions and Public Health

NHS Health Education England

Phoenix Consultancy

CERTIFICATE OF COMPLETION

LEADING TRANSFORMATIONAL CULTURE CHANGE

A LEADERSHIP DEVELOPMENT PROGRAMME FOR ALLIED HEALTH PROFESSIONS, HEALTHCARE SCIENCE, PHARMACY AND PSYCHOLOGICAL PROFESSIONS

with Merlin Walberg, Penny Humphris and Suzanne Horobin

George Burghel

successfully completed the above programme on 22 March 2016.

The programme provided leadership development including improvement science, one to one coaching, and the undertaking of a project to support the creation of cultures of continuous quality improvement.

Merlin Walberg
Phoenix Consultancy

Penny Humphris
Phoenix Consultancy

Suzanne Horobin
Phoenix Consultancy

Kirstie Baxter
Health Education England (North West Local Team)

7.4.3 Appendix D3 – Economics of Genomics and Precision Medicine

MANCHESTER 1824
The University of Manchester

The University of Manchester

Faculty of Biology, Medicine and Health

Certificate of Completion of Continuing Professional Development Programme

This is to certify that

George Burghel

Has successfully completed the Economics of Genomics and Precision Medicine course with 15 Credits at Level 7

Kelly Osgood
Deputy Head of School Operations

Kelly Salimian
Education Support Manager

7.4.4 Appendix D4 – Ethical, Legal and Social Issues in Applied Genomics

MANCHESTER 1824
The University of Manchester

The University of Manchester

Faculty of Biology, Medicine and Health

Certificate of Completion of Continuing Professional Development Programme

This is to certify that

George Burghel

Has successfully completed the Ethical, Legal and Social Issues in Applied Genomics course with 15 Credits at Level 7

Kelly Osgood
Deputy Head of School Operations

Kelly Salimian
Education Support Manager

7.5 Appendix E: Further presentations from Chapter 2

7.5.1 Appendix E1: ESHG 2019 poster presentation

MANCHESTER 1824 The University of Manchester

Copy number variants correlate with socio economic status

NHS Manchester University NHS Foundation Trust

George Burghel[1,2], Unzela Khan[1], Wei-Yu Lin[3], William Whittaker[4] and Siddharth Banka[1,2]

1. Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, UK; 2. Faculty of Biology, Medicine and Health, University of Manchester, UK; 3. Northern Institute for Cancer Research, Medical School, Newcastle University, UK; 4. Division of Population Health, Health Services Research & Primary Care, University of Manchester, UK.

Introduction: Socioeconomic status (SES) is a measure of an individual's or family's economic and social status. SES is a major determinant of health and related outcomes[1,2]. Biological factors including genetic variants that may influence SES are only beginning to be understood[3]. We investigated the correlation of properties of pathogenic CNVs with SES.

Methods:

- Manchester Centre for Genomic Medicine
- 2010 - 2017
- >17,000 results with pathogenicity classification

Postnatal array-CGH results

Information for each pathogenic CNV

- Inheritance status
- Type (loss or gain)
- Size
- Postcode of the patient

- Index of multiple deprivation rank (IMDR) for each pathogenic CNV extracted from post codes.
- Each IMDR has seven domains - income; employment; health and disability; education, skills and training; barriers to housing and services; crime; and living environment.

SES estimation

Results:

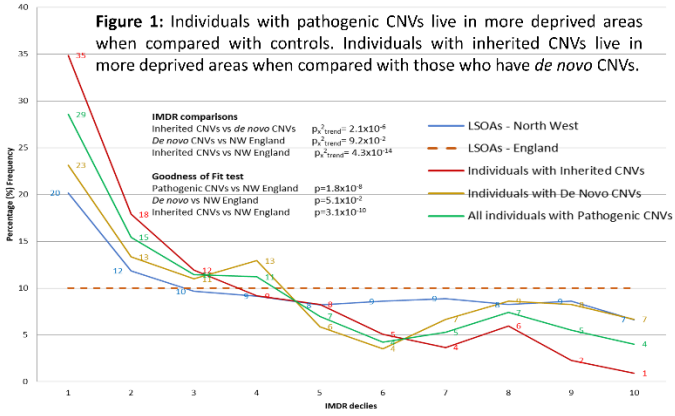
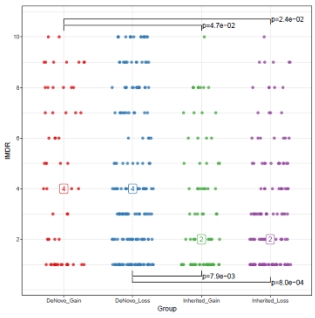


Figure 1: Individuals with pathogenic CNVs live in more deprived areas when compared with controls. Individuals with inherited CNVs live in more deprived areas when compared with those who have *de novo* CNVs.

Figure 2: There is no significant difference in the SES of individuals with pathogenic CNV gains or losses.



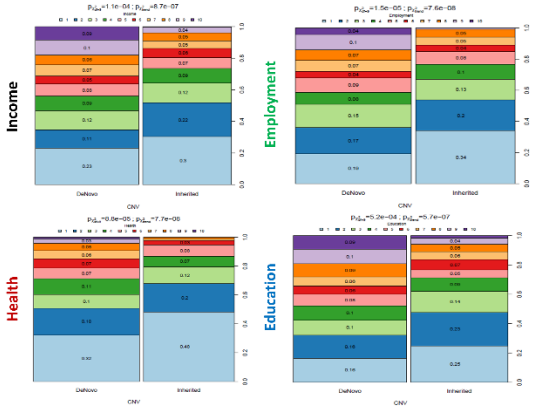
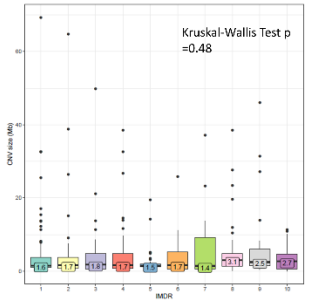


Figure 4: There was significant difference between inherited vs *de novo* CNVs groups across **income**; **employment**; **health and disability**; and **education, skills and training** IMDR domains.

Figure 3: There is no significant difference in the SES of individuals according to the size of the pathogenic CNVs



Conclusions:

1. Pathogenic CNVs and SES are correlated.
2. Individuals with inherited pathogenic CNVs tend to live in areas of lower SES.
3. The secondary social and medical consequences of lower SES of individuals with CNVs needs to be studied.
4. These results have important implications for planning of medical and social services.
5. This study provides an insight into biological determinants of SES.

References: **(1)** Clark AM *et al. Nat Rev Cardiol* 2009; 6: 712–22. **(2)** Hanscombe, K.B. *et al.*. *PLoS One* 2012; 7: e30320. **(3)** Kendall, K.M. et al.. Br J Psychiatry 2019; 1-8

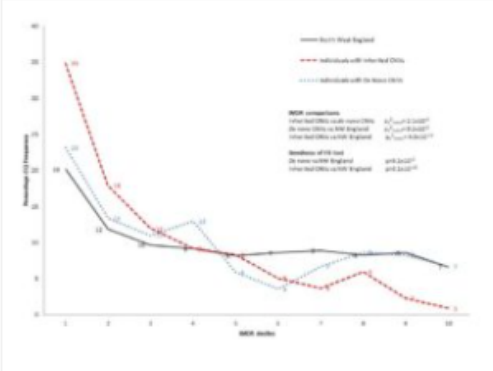
Manchester Centre for Genomic Medicine

7.5.2 Appendix E2 – SES publication layperson blog (Journal of Medical Genetics)

Presence of pathogenic copy number variants (CNVs) is correlated with socioeconomic status

Posted on September 25, 2019 by hqqu

Socio-economic status (SES) has profound effects on health outcomes. This paper shows that people with disease-causing small chromosomal variations (CNVs) tend to have lower SES. Interestingly, SES of individuals with milder inherited CNVs tends to be even lower than of individuals with more severe CNVs without any family history. These findings mean that future studies on the effect of CNVs must take into account the impact of lower SES of affected individuals. As disease-causing CNVs are common in the population, these findings have important implications for policy making. These results also provide unique insights into how genes shape our society. (By Dr. George Burghel, https://jmg.bmj.com/content/early/2019/08/21/jmedgenet-2019-106292)



https://blogs.bmj.com/jmg/2019/09/25/presence-of-pathogenic-copy-number-variants-cnvs-is-correlated-with-socioeconomic-status/

7.5.3 Appendix E3 – SES work publication on MFT website

MFT researchers find chromosomal changes implicated in disease linked to social and economic disadvantage

Posted: Wednesday Sep 25th 2019

Researchers from the Manchester Centre for Genomic Medicine (MCGM) have found that chromosomal changes implicated in disease are linked to social and economic disadvantage.

Dr George Burghel, Principal Clinical Scientist at the MCGM is the first author of the paper published in the BMJ Journal of Medical Genetics yesterday. Dr Siddharth Banka, Honorary Consultant Clinical Geneticist at MCGM, based at Saint Mary's Hospital at Manchester University NHS Foundation Trust (MFT), is the corresponding author.

The paper summarises the findings from 473 families, and Dr Burghel and Dr Banka, along with their colleagues, say the results demonstrate that chromosome changes which cause diseases are also correlated with social and economic disadvantage.

Dr Banka, who is also a Clinical Senior Lecturer at the University of Manchester, said: "Chromosomes are the thread-like structures found in the nuclei of all living cells. They carry genetic information in the form of genes.

> *It has long been thought that genetic conditions may affect the social and economic status of patients and their families, but few studies have tested this theory – until now.*

Dr Burghel added: "It's great to see the results of our research published in such a prestigious medical journal, and to be able to add to the evidence base in this important, developing area of research.

> *We hope that this work and other studies can in the future positively inform policy makers planning medical and social services.*

You can read more about the study and its findings on the BMJ website.