

ARTICLE OPEN



Reanalysis of genomic data doubles the diagnostic yield for Welsh patients recruited to the UK 100,000 Genomes Project

Jana Jezkova^{1,16}, Martin A. McClatchey^{2,12,16}, Sophie Shaw¹, Rhys Vaughan², Joseph Halstead¹, Iris Egner^{2,13}, Sharon D. Whatley^{1,14}, Arveen Kamath¹, Oliver Murch¹, Vinod Varghese¹, Rachel Irving¹, Jennifer F. Gardner¹, Ayesha Ahmed¹, Ian Tully¹, Vani Jain¹, Mark T. Rogers^{1,2}, Francis H. Sansbury^{1,2}, Angus J. Clarke^{1,2}, Caroline Pottinger¹, Maribel Verdesoto Rodriguez¹, Johann te Water Naudé³, Aimee Bettridge², Alexandra C. Martin-Geary^{1,4,5}, Nils Wagner⁶, Julien Gagneur^{6,7,8}, Marcela Votruba^{9,10}, David J. Bunyan¹¹, Kevin Ashelford², Peter Giles², Hywel J. Williams², Sian Morgan^{1,15}, Julian R. Sampson^{1,2} and Andrew E. Fry^{1,2}✉

© The Author(s) 2026

The 100,000 Genomes Project (100KGP) undertook genome sequencing of patients with rare diseases and cancer to study the role that genes play in disease, and to integrate genomics into UK healthcare. To contribute to 100KGP, the Wales Genomic Medicine Centre (a partnership between the NHS All Wales Medical Genomics Service (AWMGS), Cardiff University and Genomics England) recruited 438 individuals from 154 families that had been subject to pre-genomic genetic testing without reaching a diagnosis. The majority of probands (64%, 98/154) had neurological or neurodevelopmental phenotypes. Genome sequencing, variant calling, gene-based filtering and variant prioritisation were performed by Genomics England. AWMGS undertook clinical interpretation, validation and reporting of variants. Initial diagnostic yield was 20.8% (32/154) with variants of uncertain significance (VUS) reported for 11 more families. Most of the initial findings (83.7%, 36/43) could have been detected by clinical exome sequencing which was standard of care in Wales at the time. Reanalysis of the 100KGP Wales data using updated variant prioritisation tools, expanded gene lists, re-phenotyping and segregation studies, has increased diagnostic yield to 42.2% (65/154) (an improvement of 103%). RNA analysis was used to clarify the clinical significance of VUS in *COQ4*, *ENPP1*, *GATAD2B*, *NCAPD2*, and *THOC2*. These findings demonstrate the clinical utility of genome sequencing, RNA analysis, and periodic reanalysis of genomic data.

European Journal of Human Genetics; <https://doi.org/10.1038/s41431-026-02234-3>

INTRODUCTION

The UK government launched the 100,000 Genomes Project (100KGP) in 2013 [1]. 100KGP was a hybrid clinical-research initiative which used whole-genome sequencing (WGS) in a National Health Service (NHS) setting to study rare disease and cancer [2, 3]. In total, around 85,000 NHS patients had WGS [4]. Recruitment, phenotyping, sampling and feedback of results was delivered by 13 NHS Genomics Medicine Centres in England with related initiatives in Northern Ireland [5], Scotland [6] and Wales.

Responsibility for healthcare in Wales is devolved to the Welsh Government. This means the NHS in Wales is separate from the NHS in other parts of the UK. Wales has a single Medical Genetics Service (the All Wales Medical Genomics Service (AWMGS)) which provides clinics for inherited conditions throughout Wales. AWMGS has a single CPA-accredited NHS Genetics Laboratory

with expertise in sample preparation, genomic analysis, interpretation and reporting of genetic variation. Following the dissolution of the UK Genetic Testing Network in 2018 [7], AWMGS continued to access some gene panel tests still offered by other UK genetics laboratories, sent some panel tests to clinical laboratories in the European Union and employed clinical exome sequencing to investigate patients with brain malformations, early-onset epilepsy, intellectual disability (ID) and ectodermal dysplasia.

In 2016, a Genomic Medicine Centre (GMC) was established in Wales to contribute to the 100KGP. The GMC was a collaboration between AWMGS, Cardiff University, and Genomics England (GEL). The Wales GMC recruited participants for 100KGP between January and October 2018. The aims of 100KGP in Wales were aligned with those of the wider project: increasing

¹All Wales Medical Genomics Service, Wales Genomic Health Centre, Cardiff, UK. ²Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK. ³Department of Paediatric Neurology, University Hospital of Wales, Cardiff, UK. ⁴Big Data Institute, University of Oxford, Oxford, UK. ⁵Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ⁶School of Computation, Information and Technology, Technical University of Munich, Garching, Germany. ⁷Computational Health Center, Helmholtz Center Munich, Neuherberg, Germany. ⁸Institute of Human Genetics, School of Medicine, Technical University of Munich, Munich, Germany. ⁹School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK. ¹⁰Department of Ophthalmology, University Hospital of Wales, Cardiff, UK. ¹¹Wessex Genomics Laboratory Service (Salisbury), Salisbury District Hospital, Salisbury, UK. ¹²Present address: Northern Genetics Service, Institute of Genetic Medicine, International Centre for Life, Newcastle upon Tyne, UK. ¹³Present address: Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria. ¹⁴Present address: Department of Medical Biochemistry and Immunology, University Hospital of Wales, Cardiff, UK. ¹⁵Present address: Department of Research and Development, University Hospital of Wales, Cardiff, UK. ¹⁶These authors contributed equally: Jana Jezkova, Martin A. McClatchey. ✉email: fryae@cardiff.ac.uk

Received: 12 March 2026 Revised: 12 August 2026 Accepted: 3 September 2026

Published online: 28 September 2026

the rate of diagnosis for rare disease, studying the genetic basis of disease, integration of genomic medicine into routine NHS clinical pathways, providing training opportunities, and developing infrastructure and expertise for genomic medicine in Wales.

Here, we summarize the implementation and outcomes of the 100KGP in Wales. We describe the range of phenotypes in the recruited cohort, reported findings, and initial diagnostic yield. We estimate how often genomic findings would have been found by previous standard of care testing and the rate of incidental/additional findings in the 100KGP Wales cohort. We describe subsequent efforts to reanalyse the data and the resulting improvements in diagnostic yield that have been obtained.

SUBJECTS AND METHODS

The 100KGP in Wales operated within the regulatory approvals of the main 100KGP project and in a manner similar to other parts of the UK [6].

Participants

Recruitment of participants ran between January 2018 and October 2018. Patients were eligible if they had one of the rare phenotypes of presumed genetic origin as set out in the GEL Rare Disease List [8] and if previous NHS genetic testing had been negative. Proband was 85.7% White British, 3.2% any other white background, 5.2% Asian or Asian British, 4.5% mixed ethnicity, 0.6% Black or Black British-African and 0.6% other ethnicities. Nine families (5.8%) were known to be consanguineous, six of these were from Asian or Asian British backgrounds. Phenotypes of particular interest to the recruitment team included: ID, epilepsy, congenital malformations, inherited optic neuropathies, and retinal dystrophies.

Individuals were selected and referred by their clinician to the GMC recruitment team. Referring clinicians were either consultant clinical geneticists in AWMGS or an academic ophthalmologist (MV). Where possible, probands were co-recruited with parents and other affected family members. Parent-offspring trios were preferred but other family structures (singletons, parent-child duos, quads) were also eligible. The phenotype under investigation, associated Human Phenotype Ontology (HPO) terms [9], family structure and information on prior genetic testing were captured in GEL's OpenClinica database [10].

DNA sampling and genome sequencing

DNA was extracted from blood in the AWMGS Laboratory. DNA aliquots containing a minimum of 10 µg at 100 ng/µl with an A260/A280 ratio of 1.8–2.0 were sent to GEL for short-read genome sequencing on an Illumina sequencing platform as previously described [3]. Samples were submitted to the GEL biorepository between March 2018 and December 2018.

Clinical bioinformatic analysis

Paired-end 150 bp reads were mapped to GRCh38 for variant calling and analysed as per the GEL clinical bioinformatics pipeline for genome sequencing with panel-based gene analysis [1]. Small variants (single nucleotide variants (SNV) and indels <50 base pairs) were filtered based on frequency, segregation with the condition, and mode of inheritance. Variants were prioritised into three 'tiers'. Tier 1 (predicted protein-truncating or moderate-impact de novo variants e.g. missense) and Tier 2 (predicted protein-altering) were variants in genes with pre-existing evidence for involvement in the phenotype under investigation (based on a curated lists of 'green' genes from PanelApp [11]). Tier 3 variants were predicted protein-truncating or protein-altering in other genes. Tiered variant lists were returned to the Wales GMC between November 2018 and December 2019 using the GEL Interpretation Portal. Variants were also ranked by Exomiser score [12].

Copy-number variant (CNV) calling was introduced to the GEL pipeline in January 2019, but was not retrospectively applied to closed cases [13]. The Canvas software [14] was used to find CNVs >10 kb with a call quality score >10. Mode of inheritance was not considered. Tier A CNVs overlapped with a green gene in a panel applied to the patient. Heterozygous CNVs covering a recessive gene were tiered. Tier null CNVs were those that did not meet the criteria for tier A reporting. Manta software [15] was also used to call CNVs and structural variants (SV) but this data was not used for clinical calling.

Variant interpretation

AWMGS clinical scientists performed variant interpretation and classification of tiered variants as per American College of Medical Genetics and Genomics (ACMG) [16] and UK ACGS 2018 guidelines [17]. They analysed all Tier 1 and 2 variants and the top five Tier 3 variants in known human disease genes ranked by Exomiser.

Variant review included a clinical discussion at multidisciplinary team (MDT) meetings, when needed. All variant classifications were recorded by laboratories but only ACMG/ACGS Class 5 (pathogenic), Class 4 (likely pathogenic) and 'hot' Class 3 (uncertain significance) variants were included in diagnostic reports. Here, 'hot' meant a variant of uncertain significance (VUS) that narrowly missed a 'likely pathogenic' classification and additional evidence might lead to it being upgraded. In contrast, a 'cold' VUS was considered unlikely to be disease causing and not reported. Reported variants were validated by Sanger sequencing using retained clinical DNA samples and formally reported back to the referring consultant clinician for return to participants.

The 2025 NHS diagnostic pipeline used for the reanalysis is described in the Supplementary Methods. In a separate analysis AbSplice score was applied to 100KGP data to find SNVs near splice sites related to developmentally regulated splicing events [18]. AbSplice score predicts the likelihood that a variant causes abnormal splicing in a given tissue [19]. A full description of this analysis is given in the Supplementary Methods.

RNA extraction and quality control

The RNA methods used to analyse an *ENPP1* variant have previously been described [20]. *NCAPD2*, *COQ4* and *GATAD2B* variants were analysed by the Wessex Genomics Laboratory Service (Salisbury) using published methods [21]. In brief, RNA was extracted from peripheral blood, converted to cDNA and targeted PCR performed. PCR products were visualized by agarose gel electrophoresis and subsequently analysed by Sanger sequencing. *COQ4* and *THOC2* variants were investigated using whole transcriptome RNA-seq on fibroblast RNA performed in-house in AWMGS. A detailed description of the RNA-seq methods used are given in the Supplementary Methods.

RESULTS

Proband demographics

In total, 154 probands and 284 family members were recruited. Among probands, 104 (67.5%) were under 18 years at recruitment, 85 (55%) were male and 69 (45%) were female. The majority of probands (98, 63.6%) were in the Neurology and Neurodevelopmental Condition category, 66 of these (42.9% of all probands) had ID. The remainder were distributed across other categories (Fig. 1A, Supplementary Table 1). Fifteen probands (9.7%) were recruited under more than one condition category. One hundred and forty-five (94%) probands were recruited with family members: 119 (77%) parent-child trios, 13 (8.4%) parent-child duos, 9 (5.8%) singletons, 5 (3.2%) parent-child quads and 8 (5.2%) other family structures (Fig. 1B, Supplementary Table 2). Twenty-seven relatives (9.5% of the co-recruited family members) were considered to have the same rare condition as the proband.

Variant assessment

One or more Tier 1 and/or Tier 2 variants (SNV and indels) were found in 92 of 154 families (59.7%); one family (3 affected siblings and their unaffected mother) had no tiered variants returned; all other families had three or more Tier 3 variants returned (average 81.4 per family). The numbers of tiered variants for different family structures are shown in Fig. 1C. Consistent with the findings of Scottish 100KGP report [6], the number of tiered variants returned for interpretation reduced with increasing family size. Tiered CNV were only returned for 45/154 families (29.2%). On average 1.2 Tier A (high priority) and 67.2 Tier Null (low priority) CNVs were reported per family. Twenty-two variants were discussed in formal MDT meetings (20 were reported), 27 other variants were discussed by email between the laboratory scientists and clinicians (23 subsequently reported).

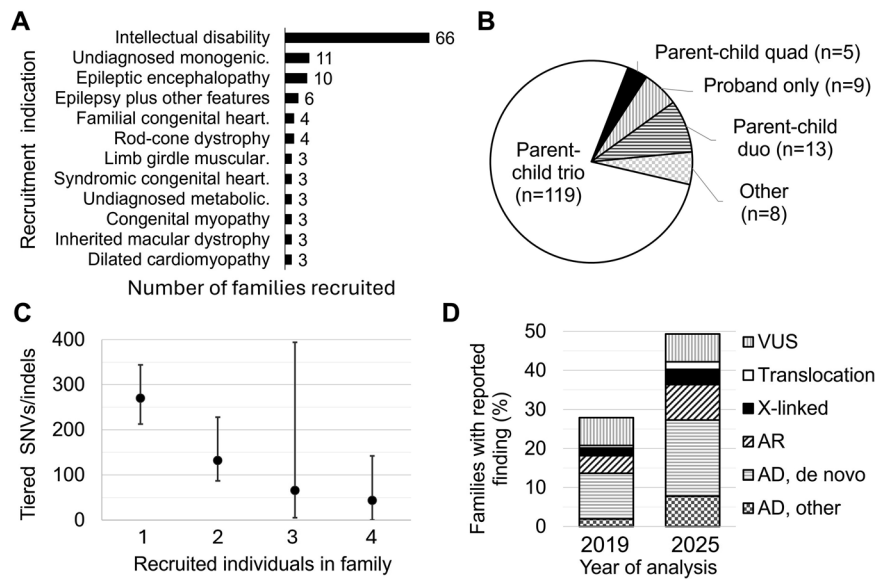


Fig. 1 Recruitment and results from the 100,000 Genomes Project in Wales. **A** Top 12 recruitment disease categories. Categories are as those set out in the Genomics England Rare Disease List (v1.7, June 2017) [8]. Where a proband was entered with conditions in more than one category, the proband is included within counts for all relevant categories. Abbreviated categories: Undiagnosed monogenic (disorder seen in a specialist genetics clinic); Limb girdle muscular (dystrophy); Syndromic congenital heart (disease) and Undiagnosed metabolic (disorders). **B** Family structures among recruited individuals. **C** Average numbers of tiered variants for different family structures (range as error bars). **D** Rates of diagnostic findings following the initial analysis of returned variants (2019) and currently (2025). AD autosomal dominant, AR autosomal recessive, VUS variant of uncertain significance.

Diagnostic yield

Following the initial analysis of returned variants in 2018/19, pathogenic or likely pathogenic variants consistent with the phenotype were found in 32/154 families (20.8%) (Fig. 1D). In addition, 11 families (7.1%) had 'hot' VUS that were returned to participants. For the remaining 111 families (72.1%) no genetic cause for the phenotype was identified. The initial findings included four Tier A CNVs, the rest were small variants. Standard of care (SOC) in Wales in 2018 was a clinical exome/Mendeliome (TruSight One [TSO] Expanded sequencing panel). We found 83.7% (36/43) of reported findings were in target regions of the TSO panel (annotated in Supplementary Table 3). This may overestimate the yield of SOC testing as sequencing coverage was not modelled in detail. Variants likely to be missed by SOC testing included non-coding variants, SV and variants in genes not included in the TSO panel.

Secondary findings

At recruitment, participants were asked if they wanted secondary findings (also known as additional findings) fed back. GEL looked for genomic findings known to cause serious conditions for which there was good evidence that knowing about them could influence healthcare. The conditions were based on the 2013 ACMG secondary findings gene list but did not include genes associated with inherited cardiac conditions [22, 23]. Secondary findings were returned for 7 out of 422 (1.7%) 100KGP Wales participants who opted in to receiving health-related additional findings. This rate is similar to the rate of additional findings returned for rare disease probands (1.58%) and relatives (1.7%) in other parts of 100KGP [24]. Two individuals had an *LDLR* variant that was classified as a 'cold' class 3 VUS and not fed back. Pathogenic variants in *BRCA2*, *MSH6* and *RET* were returned to participants (Supplementary Table 4).

Ongoing review and reanalysis

In the 7 years following the initiation of 100KGP Wales, data from 100KGP Wales has contributed to at least 10 research publications [20, 25–33]. In addition to findings identified by researchers using

the GEL Research Environment ($n = 18$), there have been ongoing local efforts to review and reclassify variants ($n = 18$). The 36 variants included 3 new VUS and 3 previous VUS that were upgraded to pathogenic or likely pathogenic. At present (2025), pathogenic or likely pathogenic variants that explained the phenotype have been found in 65/154 families (42.2%), an uplift from the original diagnostic yield (32/154) of 103% (Fig. 1D). In addition, 11 families (7.1%) have a VUS reported. Variants returned to participants are listed in Supplementary Table 3. Several approaches contributed to the increased diagnostic yield. These included new literature and discovery of new disease genes ($n = 12$), phenotype-directed review of tiered variants ($n = 7$), detection of SV ($n = 4$), improved in silico splicing prediction ($n = 3$), RNA analysis ($n = 5$), breakpoint analysis of known translocations ($n = 2$), improved indel calling ($n = 3$), additional reports in clinical variant databases ($n = 2$), testing of additional relatives ($n = 1$) and episignature analysis ($n = 1$). Some new diagnoses were based on evidence from two or more approaches. For the remaining 78 families (50.6%) no genetic cause for the phenotype has been identified to date (Fig. 1D).

We systematically reanalysed 100KGP Wales data using a modern diagnostic pipeline developed by AWMGS for WGS interpretation. Re-analysis included review of SNV, CNV, and short tandem repeats (STR) in the Variant call format (VCF) files. An average of 58 variants per proband were identified for further evaluation. Of the 144 families with VCF files available via the GEL Interpretation Portal, 70 had a reported variant (pathogenic, likely pathogenic and VUS) from previous analysis. In total, 60/70 (85.7%) reported variants were detected by the modern diagnostic pipeline. Key differences between the modern analysis pipeline and the initial 100KGP analysis included access to contemporary variant databases such as ClinVar ($n = 12$ new variants identified), annotation with SpliceAI ($n = 4$), and use of Manta/Canvas data ($n = 2$). Variants found by the modern analysis pipeline are annotated in Supplementary Table 3. Four diagnostic variants (in *ARCNI1*, *KMT2A*, *UPF3B* and *CACNA1F*) were in the VCF file but filtered out due to low quality metrics or read depth, two variants were filtered due to being inherited from parents labelled as

unaffected (*NPRL3*, *CIC*), one variant was mosaic (*SCN2A*, allele fraction 18.8% (9/48 reads)), and one variant was filtered due to gnomAD frequency (*TUBB2A* variant in a region of high homology) [28]. A homozygous SV in *ENPP1* and the t(6;13) translocation were not detected in the SV VCF supplied by GEL. Components of the *ENPP1* SV could be found in the VCF but full assembly required manual inspection and curation [20].

RNA analysis

RNA analysis was used to clarify the clinical significance of variants in *COQ4* (individual P1), *NCAPD2* (individual P2), *GATAD2B* (individual P3), *THOC2* and *ENPP1*. Detailed case reports for P1, P2 and P3 are provided in the supplementary material. The *ENPP1* results are published elsewhere [20]. A manuscript describing the *THOC2* results is being prepared.

RNA analysis was key to the diagnosis of *COQ4*-associated coenzyme Q₁₀ (CoQ₁₀) deficiency in participant P1. This individual is a 16-year-old male with early-onset intractable epilepsy, severe developmental delay, dystonia, lactic acidosis, and abnormal muscle respiratory chain enzymes studies. However, muscle CoQ₁₀ level was normal. Review of individual P1's WGS data found compound heterozygous *COQ4* missense variants: (Tier 2, Exomiser rank 3) NM_016035.5, c.205A>G, p.(Met69Val) and c.211G>A, p.(Ala71Thr). These variants were close to an exon edge and adjacent to known pathogenic variants [34].

Both variants are listed in ClinVar (VCV001345226.9, conflicting classifications, and VCV003342649.1, likely pathogenic). RNA-seq data from P1's skin fibroblasts was compared with GTEx fibroblast count data and a set of in-house controls. After adjustment for multiple testing only two genes were significant expression outliers genome-wide ($P < 0.05$): *COQ4* and *C9ORF129* (the latter looked artefactual). *COQ4* RNA was significantly reduced in P1 (fold change 0.69, $z = -7.46$, adjusted $P = 5.72 \times 10^{-7}$). The *COQ4* transcripts in P1 also had a higher proportion of a truncated isoform (25% compared to ~1% on controls, Supplementary Fig. 1). Both alleles were present in the RNA transcripts (Supplementary Fig. 2). A qualitative reduction of *COQ4* RNA was also confirmed by targeted PCR (Supplementary Fig. 3). The *COQ4* variants were both classified as likely pathogenic.

Individual P2 is a 14-year-old male with congenital microcephaly (-8 SD), moderate ID, short stature, bilateral hearing loss and previous primitive neuroectodermal tumour. Review of individual P2's WGS data found compound heterozygous variants in *NCAPD2* (Tier 3, Exomiser rank 3). The variants were a canonical splice site variant (NM_014865.4, c.1954+1G>A) and a splice region variant (c.3019G>A, p.(Gly1007Ser)). SpliceAI scores were 1.00 for c.1954+1G>A (predicted donor loss 1 bp 5' at the adjacent splice site) and 0.11 for c.3019G>A (predicted donor loss at that nucleotide). The c.3019G>A variant had a higher Pangolin score (0.5) with the predicted donor loss at the same position. Targeted analysis of *NCAPD2* RNA from blood demonstrated both variants resulted in aberrant transcripts (Supplementary Figs. 4, 5). The c.1954+1G>A variant caused an in-frame single exon deletion, r.1715_1954del, p.(Gly572_Thr651del). The c.3019G>A variant resulted in an out-of-frame single exon deletion, r.2908_3019del, p.(Asn970AlafsTer40). The transcript containing r.2908_3019del appeared to be present at low levels as an apparently full-length transcript containing the r.3019g>a variant was also visible. Both variants were classified as likely pathogenic.

Individual P3 is a 16-year-old female with global developmental delay, ID, hypotonia and absence seizures. The *GATAD2B* variant was detected through analysis of AbSplice scores for SNVs near splice sites related to developmentally regulated splicing events [18]. The variant (NM_020699.4: c.1217-399C>T) was de novo and predicted to create a pseudoexon in intron 6 of *GATAD2B* with a splice acceptor site 99 bp 5' of the variant and a donor site 2 bp 5' of the variant (SpliceAI scores 0.74 and 0.50, respectively). Targeted analysis of *GATAD2B* RNA from the proband's blood

confirmed the presence of aberrant transcripts containing the pseudoexon (Supplementary Figs. 6, 7). The variant resulted in an out-of-frame insertion: r.1216_1217ins1217-498_1217-401, p.(Lys407TrpfsTer28) and was classified as likely pathogenic.

DISCUSSION

We have described the execution and outcomes of the 100KGP in Wales. In total, 154 families, predominantly parent-child trios, were recruited and underwent WGS. Participants recruited to 100KGP Wales came from a wide range of phenotype categories and had been subject to previous genetic investigation without reaching a diagnosis. The initial diagnostic yield of 20.8% was slightly lower than the rate reported for the 100KGP pilot (25%) [3]. This likely reflects differences in recruitment and analysis. For example, 100KGP Wales recruited more children (67.5% versus 26% in the pilot), a higher proportion of ID probands (42.9% versus 6% in the pilot) and more parent-offspring trios (77% versus 36.5% in the pilot). A detailed comparison of demographic and disease category differences between 100KGP Wales and the pilot is given in Supplementary Table 1. In addition, around 14% of the diagnostic findings in the pilot came from research beyond the standard pipeline – something that was not part of initial analysis in Wales.

In the 7 years since recruitment ended, WGS data generated by 100KGP Wales has enabled the identification of novel disease genes and expanded gene-phenotype associations, resulting in new diagnoses for affected individuals [20, 25–33]. The diagnostic yield of 100KGP Wales has now doubled (42.2%). Key reasons for the improved yield include iterative reanalysis of the data, clinical reassessment of patients/families, new disease associations, and additional information from new diagnostic techniques such as RNA analysis.

We found a modern NHS WGS analysis pipeline was able to detect 85.7% of diagnostic variants identified by the research community using 100KGP Wales data. The main reasons for the NHS pipeline missing diagnostic variants were variants being inherited from a parent annotated as unaffected, and issues with the quality or read depth of the variant in the VCF file. The VCF files available on the Clinical Interpretation Portal were generated in 2018 and unprocessed raw data file types (e.g. FASTQ files) were not available. Reprocessing the original FASTQ files with updated software may improve the quality of variant calls. Inherited dominant disease alleles remain a challenge for interpretation. It is laborious to distinguish these from the large number of neutral inherited heterozygous variants. Furthermore, variants are likely to be de-prioritized if they are inherited from a parent thought to be unaffected (e.g. due to non-penetrance, milder expression or lack of documentation) or if a variant is not shared by affected relatives due to genetic heterogeneity (as may have happened to the family with no tiered variants in our cohort).

Following completion of recruitment to 100KGP Wales, AWMGS developed its own genomic sequencing capacity, leading to the first NHS-commissioned trio rapid WGS service in the UK [35]. Local sequencing solved many of the challenges of repatriating genomic data; however it means Welsh genomic data does not flow to the UK National Genomics Research Library (NGRL). No equivalent to the NGRL currently exists in Wales which limits research access to NHS genomic data. Neither routine reanalysis of genomic data nor RNA-seq are currently NHS funded in Wales.

Our findings align with the growing evidence that periodic reanalysis of genomic data from unsolved cases can substantially improve diagnostic rates in rare disease cohorts. Multiple studies have demonstrated significant increases in diagnostic rates following reanalysis, with yields ranging from 2.3% to 22% [36–38]. Dai et al. conducted a meta-analysis showing that reanalysis had an overall diagnostic yield of 10% [39]. Application of clinical expertise can also improve reanalysis yield. Involvement of clinical geneticists in data interpretation has been shown to

Table 1. Clinical and genetic findings in patients with NCAPD2-related microcephaly.

	Martin et al. [55]	Masih et al. [58]	Lin et al. [59]	Lin et al. [59]	Reuter et al. [60]	Reuter et al. [60]	This report
Ethnic origin	Indian	Indian (consanguineous)	Chinese	Chinese	Iraqi (consanguineous)	Iraqi (consanguineous)	White British
NCAPD2 genotype (NM_014865.4)	Homozygous c.4120+2T>C	Homozygous c.4120+2T>C	Homozygous c.3477+2T>C	Homozygous c.3477+2T>C	Homozygous c.23T>C	Homozygous c.23T>C	Compound heterozygous c.1954+1G>A/ c.3019G>A
Protein	p.(Asp1374GlyfsTer29)	p.(Asp1374GlyfsTer29)	p.(Gly1160ValfsTer16)	p.(Gly1160ValfsTer16)	p.(Phe8Ser)	p.(Phe8Ser)	p.(Gly572_Thr651del), p.(Asn970AlafsTer40)
Age at report	3 y	1 y 2 m	13 y	Terminated at 34 w gestation	n/k	n/k	12.5 y
Sex	Male	Female	Female	Male	Male	Female	Male
OFC	34.8 cm (−11.9 SD)	29 cm (−13.07 SD)	45 cm (<−3 SD)	Progressive microcephaly	Microcephaly	Microcephaly	45 cm (−6.3 SD)
Height	74 cm (−5.8 SD)	62 cm (−4.55 SD)	151.2 cm (>−1 SD)	n/k	n/k	n/k	136 cm (−2.0 SD)
Weight	7 kg (−7.25 SD)	5 kg (−5.33 SD)	38 kg (>−1 SD)	n/k	n/k	n/k	35.4 kg (−0.8 SD)
Birth weight	1 kg (−4.88 SD at 37w)	1300 g	3.1 kg (>−1 SD at term)	n/k	Low	Low	2.98 kg (−1.2 SD)
Clinical findings	Severe ID, autistic-like behaviours, no speech, low-birth weight, and short stature	Severe ID, prominent nose, short neck, delayed teeth eruption, hyperactivity, short stature, speech delay	Mild ID (IQ 59), sloping forehead, poor learning performance, emotional instability and hyperactivity.	Sibling of previous patient. Microcephaly detected by abdominal ultrasound and intra utero MRI after 24 weeks of pregnancy	Mild ID, low-birth weight, short stature. 1 st cousin of next patient. Also share homozygous ENO2 missense.	Mild ID, low-birth weight, short stature. 1 st cousin of previous patient. Also share homozygous ENO2 missense.	Moderate ID, severe speech and language delay, autism spectrum disorder, ADHD, challenging behaviour, mixed hearing loss
MRI brain	Published images suggestive of gyral simplification	Microcephaly; thinned out splenium of corpus callosum and dorsal cyst communicating with third ventricle	Reduced cerebral cortex size with simplified gyri	Foetal MRI shown in original report	n/k	n/k	Normal in infancy, supratentorial primitive neuroectodermal tumour at 2.5 y
RNA analysis	Retention of intron 30. Small amounts of wild-type transcript still detectable.	Not done	Not done, retention of intron 26 speculated. SpliceAI predicts an in-frame deletion r.3391_3477del p.(Val1131_Lys1159del)	Not done, retention of intron 26 speculated.	Not done	Not done	c.1954+1G>A causes in-frame single exon deletion; c.3019G>A causes out-of-frame single exon deletion. Apparently full-length transcript containing r.3019g>a also visible

w weeks, m months, y years, ID intellectual disability, SD standard deviation, n/k not known.

lead to a 15.5% increase in diagnoses [40]. Similarly, phenotypic evaluation and clinical insights led to an upgraded classification of several tiered variants in the 100KGP Wales cohort (7/36 (19%) of the newly reported variants or upgraded VUS in 2025).

We recognise the limitations of this study. The 100KGP Wales cohort size of 154 families is relatively small, which may limit the generalizability of the findings. In addition, while our study demonstrates the value of reanalysis, it did not determine the optimal frequency of reanalyses or the cost-effectiveness of this approach in a resource-limited healthcare setting. Despite the benefits of reanalysis, implementing systematic reanalysis in clinical practice faces several challenges. Workforce capacity has been identified as a major barrier, along with ethical, legal and economic considerations [41, 42]. To address these issues, various implementation strategies have been proposed, including the development of automated reanalysis pipelines and increased use of artificial intelligence [43, 44].

We found RNA analysis particularly useful for the investigation of tiered variants classified as VUS. RNA sequencing was critical to establishing the diagnosis of COQ4-associated CoQ₁₀ deficiency in participant P1. CoQ₁₀ (or ubiquinone) is an essential component of the electron transfer chain, where it shuttles electrons from complex I or II to complex III [45]. The spectrum of COQ4-associated disease includes patients with neonatal- or infantile-onset severe neurological impairment, often accompanied by cardiomyopathy and respiratory distress [46, 47] and patients with later-onset ataxia and spasticity [48, 49]. Individual P1 was a diagnostic challenge because although his phenotype was consistent with a mitochondrial disorder, his muscle CoQ₁₀ level was normal and blood CoQ₁₀ level was only at the lower end of the reference range. Respiratory chain enzymes studies showed P1 had borderline low complex II + III and low complex IV activity in muscle. Low complex II + III activity is common in COQ4-associated CoQ₁₀ deficiency [46, 50]. Low complex I + III activity is also reported in COQ4 patients [46, 47] but this was not measured in P1. Reduced complex IV activity is not considered a typical feature but reduced activity (30–50% of mean control values) has been reported in some COQ4 patients [46].

Biochemical studies in COQ4 patient muscle and fibroblasts show CoQ₁₀ is often severely reduced (<10% mean control value) [46]. However, some COQ4 patients have CoQ₁₀ levels similar to controls [51]. The functional severity of mutations may partly explain this. Patients with later onset or long-term survival (like individual P1) have only relative reduction (40–70% of mean control CoQ₁₀ values) [46, 49] or levels at the lower end of the reference range [52]. Timing of CoQ₁₀ testing may also influence the result. P1's muscle biopsy and blood CoQ₁₀ test were performed when he was clinically stable possibly missing a transient reduction in CoQ₁₀. P1's CoQ₁₀ production may not be sufficient during metabolic stress (e.g. intercurrent illness or prolonged seizures) explaining why he had higher plasma lactate during acute illness. A decrease in plasma CoQ₁₀ concentration was noted in a previous COQ4 patient whose disease worsened due to rotavirus infection [53].

RNA analysis also helped interpret the *NCAPD2* variants identified in participant P2. *NCAPD2* encodes non-SMC condensin I complex subunit D2 a component of the condensin complex which plays a key role in chromosome condensation [54]. Mutations in *NCAPD2* are associated with autosomal recessive primary microcephaly type 21 (OMIM 617983). *NCAPD2* patients are extremely rare. Table 1 summarizes previously reported patients with biallelic *NCAPD2* variants. P2 shares several features with previously reported individuals including severe microcephaly, short stature, severe ID, ASD, emotional instability, and hyperactivity. Interestingly, splice site variants have been common in the *NCAPD2* patients reported to date. The 'leaky' nature of the splice variants may allow some full-length protein to be produced.

This may indicate that complete loss of *NCAPD2* expression may not be compatible with survival.

Mutations in genes encoding centrosomal proteins (e.g. *ASPM* or *MCPH1*) cause the mitotic spindle to fail to align leading to oblique cell cleavage events that result in premature asymmetric cell division reducing the neural stem cell pool. In contrast, condensin mutations introduce chromatin bridges but do not alter spindle orientation [55]. Consistent with this, P2 did not have premature chromosome condensation when tested. The chromatin bridges cause chromosome segregation errors leading to micronucleus formation and increased aneuploidy in daughter cells. This is expected to reduce cell proliferation and survival, depleting the neuronal number generated during neurogenesis resulting in microcephaly [55]. An increased rate of aneuploidy in neural progenitors might have contributed to the brain PNET in individual P2. To date, no other *NCAPD2* patients or their heterozygous relatives have been reported to have cancer. However, a patient with *NCAPD3* variants (the paralog of *NCAPD2* encoding the corresponding subunit in the condensin II complex) developed a highly malignant anaplastic medulloblastoma at 8 years of age [55]. *NCAPD2* has recently been recognised as an oncogene which is expressed in a wide range of cancer types. However, the mechanism for this role may be different as clinical stage and poor prognosis were associated with over-expression of *NCAPD2* rather than deficiency [56, 57].

In conclusion, we have described the implementation and outcomes of the 100KGP in Wales. Our results demonstrate that genome sequencing, coupled with systematic reanalysis, significantly improves diagnostic yield in rare diseases. The 103% increase in diagnostic rate over seven years highlights the dynamic nature of genomic diagnosis and the need for healthcare systems to implement protocols for periodic reanalysis of genomic data. Challenges remain in determining the optimal frequency and methodology for reanalysis. Our findings demonstrate the value of integrating new techniques, including RNA-seq and improved SV detection, into routine genomic diagnostics. Furthermore, our experience highlights the value of embedding genomic research alongside routine clinical care, facilitating new discoveries while improving diagnostic rates for patients.

DATA AVAILABILITY

Data from the NGRl used in this research are available within the secure Genomics England Research Environment. Access to NGRl data is restricted to adhere to consent requirements and protect participant privacy. Data used in this research include: Variants returned to 100KGP Wales participants (Supplementary Tables 3, 4). 19 variants identified by analysis of AbSplice scores (Supplementary Table 5). Access to NGRl data is provided to approved researchers who are members of the Genomics England Research Network, subject to institutional access agreements and research project approval under participant-led governance. For more information on data access, visit: <https://www.genomicsengland.co.uk/research>.

REFERENCES

- Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes project: bringing whole genome sequencing to the NHS. *BMJ*. 2018;361:k1687.
- Ballard LM, Horton RH, Dheensa S, Fenwick A, Lucassen AM. Exploring broad consent in the context of the 100,000 genomes project: a mixed methods study. *Eur J Hum Genet*. 2020;28:732–41.
- 100,000 Genomes Project Pilot Investigators, Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, et al. 100,000 genomes pilot on rare-disease diagnosis in health care - preliminary report. *N Engl J Med*. 2021;385:1868–80.
- 100,000 Genomes Project. Genomics England. 2025. <https://www.genomicsengland.co.uk/initiatives/100000-genomes-project>.
- Kerr K, McKenna C, Heggarty S, Baillie C, McMullan J, Crowe A, et al. A formative study of the implementation of whole genome sequencing in Northern Ireland. *Genes*. 2022;13:1104.
- Hocking LJ, Andrews C, Armstrong C, Ansari M, Baty D, Berg J, et al. Genome sequencing with gene panel-based analysis for rare inherited conditions in a

- publicly funded healthcare system: implications for future testing. *Eur J Hum Genet.* 2023;31:231–8.
7. UK Genetic Testing Network (UKGTN) - series closed 2018. NHS England Digital. 2025. <https://digital.nhs.uk/data-and-information/publications/statistical/uk-gtn>.
 8. Genomics England. The 100,000 Genomes Project – Current Rare Disease List v 1.7. 2017.
 9. Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* 2021;49:D1207–D1217.
 10. Introduction to OpenClinica A Genomics England Quick Guide. 2017. <https://www.genomicsengland.co.uk/assets/forms/Quickguide-Open-Clinica-registration-consent-and-data-submission.pdf> (accessed 22 July 2026).
 11. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet.* 2019;51:1560–5.
 12. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc.* 2015;10:2004–15.
 13. Hyder Z, Calpena E, Pei Y, Toozee RS, Brittain H, Twigg SRF, et al. Evaluating the performance of a clinical genome sequencing program for diagnosis of rare genetic disease, seen through the lens of craniosynostosis. *Genet Med.* 2021;23:2360–8.
 14. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics.* 2016;32:2375–7.
 15. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics.* 2016;32:1220–2.
 16. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–24.
 17. Ellard S, Baple EL, Owens M, Eccles DM, Turnbull C, Abbs S, et al. ACGS Best Practice Guidelines for Variant Classification 2018. 2018. https://www.acgs.uk.com/media/10793/uk_practice_guidelines_for_variant_classification_2018_v10.pdf (accessed 7 Aug 2025).
 18. Wagner N, Neverov AM, Martin-Geary AC, Londhe S, Schröder C, Yépez VA, et al. Aberrant splicing prediction during human organ development. 2025. <https://doi.org/10.1101/2025.07.16.665183> (accessed 14 Nov 2025).
 19. Wagner N, Çelik MH, Hölzlwimmer FR, Mertes C, Prokisch H, Yépez VA, et al. Aberrant splicing prediction across human tissues. *Nat Genet.* 2023;55:861–70.
 20. Moore AR, Yu J, Pei Y, Cheng EWY, Taylor Tavares AL, Walker WT, et al. Use of genome sequencing to hunt for cryptic second-hit variants: analysis of 31 cases recruited to the 100 000 Genomes Project. *J Med Genet.* 2023;60:1235–44.
 21. Wai HA, Lord J, Lyon M, Gunning A, Kelly H, Cibin P, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med.* 2020;22:1005–14.
 22. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 2013;15:565–74.
 23. Nolan J, Buchanan J, Taylor J, Almeida J, Bedenham T, Blair E, et al. Secondary (additional) findings from the 100,000 Genomes Project: Disease manifestation, health care outcomes, and costs of disclosure. *Genet Med.* 2024;26:101051.
 24. Smith EJ, Hill M, Chitty LS, Morris S. Costs and cost-effectiveness of returning secondary findings from genomic sequencing based on the return of additional findings in the 100,000 Genomes Project. *Genet Med.* 2025;27:101479.
 25. Pagnamenta AT, Yu J, Walker S, Noble AJ, Lord J, Dutta P, et al. The impact of inversions across 33,924 families with rare disease from a national genome sequencing project. *Am J Hum Genet.* 2024;111:1140–64.
 26. McClatchey MA, du Toit ZD, Vaughan R, Whatley SD, Martins S, Hegde S, et al. Focal segmental glomerulosclerosis and mild intellectual disability in a patient with a novel de novo truncating TRIM8 mutation. *Eur J Med Genet.* 2020;63:103972.
 27. Fry AE, Marra C, Derrick AV, Pickrell WO, Higgins AT, Te Water Naude J, et al. Missense variants in the N-terminal domain of the A isoform of FHF2/FGF13 cause an X-linked developmental and epileptic encephalopathy. *Am J Hum Genet.* 2021;108:176–85.
 28. Ragoussis V, Pagnamenta AT, Haines RL, Giacopuzzi E, McClatchey MA, Sampson JR, et al. Using data from the 100,000 Genomes Project to resolve conflicting interpretations of a recurrent TUBB2A mutation. *J Med Genet.* 2022;59:366–9.
 29. Parry DA, Martin C-A, Greene P, Marsh JA, Genomics England Research Consortium, Blyth M, et al. Heterozygous lamin B1 and lamin B2 variants cause primary microcephaly and define a novel laminopathy. *Genet Med.* 2021;23:408–14.
 30. Senum SR, Li YSM, Benson KA, Joli G, Olinger E, Lavu S, et al. Monoallelic IFT140 pathogenic variants are an important cause of the autosomal dominant polycystic kidney-spectrum phenotype. *Am J Hum Genet.* 2022;109:136–56.
 31. Klöckner C, Sticht H, Zacher P, Popp B, Babcock HE, Bakker DP, et al. De novo variants in SNAP25 cause an early-onset developmental and epileptic encephalopathy. *Genet Med.* 2021;23:653–60.
 32. Rabin R, Radmanesh A, Glass IA, Dobyns WB, Aldinger KA, Shieh JT, et al. Genotype-phenotype correlation at codon 1740 of SETD2. *Am J Med Genet A.* 2020;182:2037–48.
 33. Lee S, Menzies L, Hay E, Ochoa E, Docquier F, Rodger F, et al. Epigenotype-genotype-phenotype correlations in SETD1A and SETD2 chromatin disorders. *Hum Mol Genet.* 2023;32:3123–34.
 34. Sondheimer N, Hewson S, Cameron JM, Somers GR, Broadbent JD, Ziosi M, et al. Novel recessive mutations in COQ4 cause severe infantile cardiomyopathy and encephalopathy associated with CoQ10 deficiency. *Mol Genet Metab Rep.* 2017;12:23–27.
 35. Sloper E, Jezkova J, Thomas J, Dawson K, Halstead J, Gardner J, et al. Wales Infants' and children's Genome Service (WINGS): providing rapid genetic diagnoses for unwell children. *Arch Dis Child.* 2024;109:409–13.
 36. Bullich G, Matalonga L, Pujadas M, Papakonstantinou A, Piscia D, Tonda R, et al. Systematic collaborative reanalysis of genomic data improves diagnostic yield in neurologic rare diseases. *J Mol Diagn.* 2022;24:529–42.
 37. Halfmeyer I, Bartolomaeus T, Popp B, Radtke M, Helms T, Hentschel J, et al. Approach to cohort-wide re-analysis of exome data in 1000 individuals with neurodevelopmental disorders. *Genes.* 2022;14:30.
 38. Schobers G, Schieving JH, Yntema HG, Pennings M, Pfundt R, Derks R, et al. Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. *Genome Med.* 2022;14:66.
 39. Dai P, Honda A, Ewans L, McLaughran J, Burnett L, Law M, et al. Recommendations for next generation sequencing data reanalysis of unsolved cases with suspected Mendelian disorders: a systematic review and meta-analysis. *Genet Med.* 2022;24:1618–29.
 40. Basel-Salmon L, Orenstein N, Markus-Bustani K, Ruhrman-Shahar N, Kilim Y, Magal N, et al. Improved diagnostics by exome sequencing following raw data reevaluation by clinical geneticists involved in the medical care of the individuals tested. *Genet Med.* 2019;21:1443–51.
 41. Best S, Fehlberg Z, Richards C, Quinn MCJ, Lunke S, Spurdle AB, et al. Reanalysis of genomic data in rare disease: current practice and attitudes among Australian clinical and laboratory genetics services. *Eur J Hum Genet.* 2024;32:1428–35.
 42. van der Geest MA, Maeckelberghe ELM, van Gijn ME, Lucassen AM, Swertz MA, van Langen IM, et al. Systematic reanalysis of genomic data by diagnostic laboratories: a scoping review of ethical, economic, legal and (psycho)social implications. *Eur J Hum Genet.* 2024;32:489–97.
 43. Abdelwahab O, Torkamaneh D. Artificial intelligence in variant calling: a review. *Front Bioinform.* 2025;5:1574359.
 44. Welland MJ, Ahlquist KD, De Fazio P, Austin-Tse C, Pais L, Wedd L, et al. Automated reanalysis of genomic data for rare disease diagnostics at scale. *Nat Med.* 2026. <https://doi.org/10.1038/s41591-026-04477-5>.
 45. Alcázar-Fabra M, Navas P, Brea-Calvo G. Coenzyme Q biosynthesis and its role in the respiratory chain structure. *Biochim Biophys Acta.* 2016;1857:1073–8.
 46. Brea-Calvo G, Haack TB, Karall D, Ohtake A, Invernizzi F, Carozzo R, et al. COQ4 mutations cause a broad spectrum of mitochondrial disorders associated with CoQ10 deficiency. *Am J Hum Genet.* 2015;96:309–17.
 47. Laugwitz L, Seibt A, Herebian D, Peralta S, Kienle I, Buchert R, et al. Human COQ4 deficiency: delineating the clinical, metabolic and neuroimaging phenotypes. *J Med Genet.* 2022;59:878–87.
 48. Bosch AM, Kamsteeg E-J, Rodenburg RJ, van Deutekom AW, Buis DR, Engelen M, et al. Coenzyme Q10 deficiency due to a COQ4 gene defect causes childhood-onset spinocerebellar ataxia and stroke-like episodes. *Mol Genet Metab Rep.* 2018;17:19–21.
 49. Cordts I, Semmler L, Prasuhn J, Seibt A, Herebian D, Navaratnarajah T, et al. Bi-allelic COQ4 variants cause adult-onset ataxia-spasticity spectrum disease. *Mov Disord.* 2022;37:2147–53.
 50. Yu MH-C, Tsang MH-Y, Lai S, Ho MS-P, Tse DML, Willis B, et al. Primary coenzyme Q10 deficiency-7: expanded phenotypic spectrum and a founder mutation in southern Chinese. *NPJ Genom Med.* 2019;4:18.
 51. Mero S, Salvati L, Leuzzi V, Rubegni A, Calderan C, Nardecchia F, et al. New pathogenic variants in COQ4 cause ataxia and neurodevelopmental disorder without detectable CoQ10 deficiency in muscle or skin fibroblasts. *J Neurol.* 2021;268:3381–9.
 52. Caglayan AO, Gumus H, Sandford E, Kubisiak TL, Ma Q, Ozel AB, et al. COQ4 mutation leads to childhood-onset ataxia improved by CoQ10 administration. *Cerebellum.* 2019;18:665–9.
 53. Pan P, Zhou N, Sun Y, Chen Z, Han J, Zhou W. The Spectrum of clinical manifestations in newborns with the COQ4 mutation: case series and literature review. *Front Pediatr.* 2024;12:1410133.
 54. Hirano T. Condensin-based chromosome organization from bacteria to vertebrates. *Cell.* 2016;164:847–57.

55. Martin C-A, Murray JE, Carroll P, Leitch A, Mackenzie KJ, Halachev M, et al. Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. *Genes Dev.* 2016;30:2158–72.
56. Feng L, Yang Y, Lin Z, Cui M, Jin A, Cui A. NCPAD2 is a favorable predictor of prognostic and immunotherapeutic biomarker for multiple cancer types including lung cancer. *Genes Environ.* 2024;46:2.
57. Dong X, Liu T, Li Z, Zhai Y. Non-SMC condensin I complex subunit D2 (NCPAD2) reveals its prognostic and immunologic features in human cancers. *Aging.* 2023;15:7237–57.
58. Masih S, Moirangthem A, Shambhavi A, Rai A, Mandal K, Saxena D, et al. Deciphering the molecular landscape of microcephaly in 87 Indian families by exome sequencing. *Eur J Med Genet.* 2022;65:104520.
59. Lin Y, Zeng C, Lu Z, Lin R, Liu L. A novel homozygous splice-site variant of NCPAD2 gene identified in two siblings with primary microcephaly: The second case report. *Clin Genet.* 2019;96:98–101.
60. Reuter MS, Tawamie H, Buchert R, Hosny Gebril O, Froukh T, Thiel C, et al. Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA Psychiatry.* 2017;74:293–9.

ACKNOWLEDGEMENTS

We would like to thank the patients, families, clinicians and laboratory staff who contributed to this work. We gratefully acknowledge the participants of the NGRL, whose contributions made this research possible. Secure access to the NGRL under project IDs 354 and 1404 was provided by Genomics England, which delivers the NGRL in partnership with NHS England, and is wholly owned by the UK Department of Health and Social Care. The NGRL contains participants' health data collected by the NHS as part of their care, along with samples and data from their participation in research, for which fully informed consent has been obtained. This includes genomic and clinical data provided through the NHS Genomic Medicine Service, as well as data obtained through research studies, including the 100,000 Genomes Project and the Generation Study, both of which are delivered in partnership with the NHS, and from other research cohorts involving external collaborators.

AUTHOR CONTRIBUTIONS

Conceptualization: JRS, AEF and HW; Funding acquisition: JRS; Project administration: IE and RV; Investigation: RV, IE, AK, OM, VV, RI, JFG, AA, IT, VJ, MTR, FHS, AJC, CP, MVR, JtWN, MV, SM, JRS and AEF; Formal analysis: SW, JJ, MAM, AEF, DJB, ACM-G, NW and JG; Software: JH, KA and PG; Data curation: SS; Supervision: JRS, AEF, and SM; Writing – original draft: AEF, JJ, MAM and HW; Writing – review & editing: All authors.

FUNDING

The 100KGP in Wales was support by grant MC_PC_16035 from the Medical Research Council. MAM and AEF were supported by the Wales Gene Park. Wales Gene Park was funded by Health and Care Research Wales. JJ was supported by a Research Building Capacity Collaboration Wales Postdoctoral Fellowship. This study was supported by

the Deutsche Forschungsgemeinschaft (DFG) through the IT Infrastructure for Computational Molecular Medicine (#461264291). NW and JG were supported by the DFG via the project NFDI 1/1 "GHGA - German Human Genome-Phenome Archive" (#441914366), the European Research Council (ERC) (EPIC; 101118521) and the European Union (EU). JG was supported by the European Rare Diseases Research Alliance (ERDERA). ERDERA has received funding from the EU's Horizon Europe research and innovation programme under grant agreement N°101156595. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the EU or the ERC Executive Agency. Neither the EU nor the granting authority can be held responsible for them.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL

The 100,000 Genomes Project was approved by the East of England – Cambridge South Ethics Committee (14/EE/1112). The parents of individuals P1, P2 and P3 provided written informed consent for publication of their case reports.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-026-02234-3>.

Correspondence and requests for materials should be addressed to Andrew E. Fry.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026