Defects in tubulin beta 2A class IIa (TUBB2A) are associated with a range of complex cerebral cortex dysplasias. Defects in tubulin beta 2A class IIa (TUBB2A) are associated with a range of complex cerebral cortex dysplasias. Defects in tubulin beta 2A class IIa (TUBB2A) are associated with a range of complex cerebral cortex dysplasias. However, several studies reporting NM_001069.3:c.743C>T p.(Ala248Val) as a recurrent pathogenic mutation, it is listed in ClinVar with conflicting interpretations. To resolve these inconsistencies, we scanned data from the 100,000 Genomes Project1 (100KGP) and identified 58 individuals where p.(Ala248Val) had been called. Read alignment analysis suggested that the variant was genuine in 5/58 individuals, all of whom had a primary neurodevelopmental phenotype. In the remaining cases which spanned non-specific disease phenotypes, low allelic ratios (1%–19%) suggest recurrent mismapping artefacts. Alpha and beta tubulins form heterodimers that polymerise to form microtubules, dynamic components of the cytoskeleton that play an important role in cell division, migration and intracellular transport. Variants in several tubulin genes are associated with a variety of cortical brain malformation phenotypes, including lissencephaly, polymicrogyria, microlissencephaly and simplified gyration, collectively termed 'tubulopathies'. A recently described tubulopathy involving TUBB2A (MIM #615763) has been associated with brain phenotypes ranging from a normal cortex to extensive dysgyria. One particular TUBB2A variant, p.(Ala248Val), has been reported in several studies, in most cases arising de novo. Additional unpublished clinical cases also report a de novo origin (www.ncbi.nlm.nih.gov/clinvar/variation/127101).

Multiple occurrences of the same de novo mutation in patients with overlapping phenotypes would typically provide strong evidence supporting pathogenicity. However, on closer inspection, p.(Ala248Val) becomes harder to interpret, particularly when applying the American College of Medical Genetics and Genomics (ACMG) population allele frequency (AF) criteria PM2/BS1. The AF in gnomAD v2.1.1 is 8/237 044 in exomes and 79/15 882 in genomes, an unexpected skew for a low penetrance allele. This mismapping hypothesis is also supported by the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

Segmental duplications are known to result in reads with low mapping quality on short-read sequencing, and this can cause mismapping artefacts. Indeed, several regions share similarity with TUBB2A. Although the highest identity is with TUBB2B, other beta tubulin genes (TUBB3/TUBB4A/TUBB6) and a pseudogene (TUBB2BP1) share >90% identity with TUBB2A exon 4 (online supplemental table S1). Notably, TUBB2BP1 contains the analogous base to p.(Ala248Val) in TUBB2A, and this 'cismorphism' is in a region relatively depleted for other cismorphisms (figure 1). Thus, we speculate that mismapping of reads from TUBB2BP1 may result in p.(Ala248Val) being called in TUBB2A as an artefact and thus the apparently high AF in gnomAD.

Searching data from 78 195 individuals sequenced as part of the 100KGP (online supplemental material, Methods section) uncovered 58 subjects apparently heterozygous for p.(Ala248Val). On reviewing read alignment statistics, two distinct clusters were seen. In 5/58 individuals, the p.(Ala248Val) variant appeared being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.

The p.(Ala248Val) variant fails quality control filters in the gnomAD genome datasets and is only visible when the ‘filtered variants’ checkbox is selected. In contrast, it is a PASS variant in the exome subset of gnomAD v2.1.1. This inconsistent AF data likely explains the conflicting interpretations in ClinVar—currently one benign, one likely benign, two likely pathogenic and two pathogenic assessments. This degree of conflict is unusual, as diagnostic laboratories apply ACMG guidelines conservatively and typically report variants as being of uncertain significance when doubt arises.
three nearby TUBB2A-TUBB2BP1 cismorphisms, which can be observed in the same reads (figure 2B).

All five patients with apparently ‘genuine’ variants had neuro-developmental presentations involving intellectual disability. Three patients were reported to have seizures (one with electroencephalogram showing hypsarrhythmia); three had hypoplasia of the corpus callosum; and three had asymmetric ventricules; the findings were not atypical of the clinical tubulinopathy spectrum (online supplemental table S2 and figure S1). In four of five of these cases, genome sequencing had been performed as parent–child trios, and in these, the variant was confirmed to have arisen de novo. The other 53 individuals spanned several disease areas and included unaffected family members, as well as germline samples from patients with cancer (online supplemental table S3).

Of the five patients where the variant was suspected to be genuine, three were white; one was Pakistani; and for one, ethnicity data were unavailable. Of the remaining 53 individuals, 34% were African/Caribbean; 30% were Asian; 13% were white; and for 23%, ethnicity data were not available. The increased prevalence of likely artefactual variant calls in individuals of African ethnicity mirrors the pattern seen in gnomAD. This may reflect TUBB2BP1 polymorphisms or additional tracts of common paralogous sequence in that population.

On a technical note, where Sanger sequencing is used for validation, primer design is critically important. In the original study by Cushion et al, a low allelic fraction was observed in the electropherogram. Rather than reflecting mosaicism, this was likely due to coamplification of TUBB2B (figure 1). We propose an alternative reverse primer (online supplemental table S4) that increases specificity towards TUBB2A and also demonstrate that poor primer design can lead to erroneous validation of NGS artefacts (online supplemental figure S2). Where similar methods are used, we recommend filtering p.(Ala248Val) variant calls at an allelic fraction of >20% and requiring >2 reads on both strands.

For one case, retrospective analysis of exome sequencing validated p.(Ala248Val) but further emphasised the impact of read lengths on mapping quality (online supplemental figure S3). Applying a similar analytical strategy on TUBB2B identified two patients from 100KGP with cortical brain malformations harbouring the corresponding p.(Ala248Val) variant (online supplemental material), with a similar clustering pattern observed (online supplemental figures S4, S5).

Our cautionary tale highlights the difficulty in distinguishing bona fide gene-conversion events from mapping artefacts using short-read data. It is anticipated that increased uptake of long-read sequencing technologies will be beneficial to help fully resolve repetitive loci such as this. The value of plotting
read-alignment statistics across a large cohort of individuals analysed using a uniform pipeline (eg, 100KGP) is also highlighted. It is likely that similar approaches may be useful for other genes where conversion events represent an important mutational mechanism.

**Author affiliations**

1. Wellcome Centre for Human Genetics, Oxford University, Oxford, UK
2. NIHR Biomedical Research Centre, Oxford, UK
3. East Midlands Regional Molecular Genetics Service, Nottingham University Hospitals NHS Trust, Nottingham, UK
4. Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK
5. Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK
6. Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, Nottingham, UK
7. North West Thames Regional Genetics Service, Northwick Park Hospital, Harrow, UK
8. Department of Pediatrics, Amsterdam University Medical Centres, Duivendrecht, Noord-Holland, Netherlands
9. West Midlands Regional Clinical Genetics Service and Birmingham Health Partners, Birmingham Women’s and Children’s Hospitals NHS Foundation Trust, Birmingham, UK

**Acknowledgements**

We thank Daniel Martin for performing Sanger validation in one of the families and the anonymous reviewer for suggesting we extend our analysis to include TUBB2A. We also thank the DDD study (www.ddduk.org) for sharing read alignment data for the overlapping case.

**Collaborators**

Genomics England Research Consortium: J C Ambrose; P Arumugam; E L Bale; M Bleday; F Boardman-Pretty; J M Boissiere; C R Boustred; H Brittain; M J Caulfield; G C Chan; C E H Craig; L C Daugherty; A de Burca; A Devereau; G Elgar; R E Foulger; T Fowler; P Furio-Tari; J M Hackett; D Halai; A Hamblin; S Henderson; J E Holman; T J P Hubbard; K Ibañez; R Jackson; L J Jones; D Kasprzyk; M Kayici; A Kousathanas; L Lahstejn; K Lawson; S E A Leigh; U L Seong; F J Lopez; F Maleady-Crowe; J Mason; E M McDonagh; L Moustianas; M Mueller; N Murugaeus; A C Need; C A Odhams; C Patch; M B Pereira; D Perez-Gil; D Polychronopoulos; J Pullinger; T Rahim; A Renderon; P Riesgo-Ferreiro; T Rogers; M Ryden; K Savage; K Sawant; R H Scott; A Siddiqi; A Sieghart; D Smedly; K R Smith; S C Smith; A Sosinsky; W Spooner; H E Stevens; A Stuckey; R Sultana; E A Thomas; R Thompson; C Tregidgo; A Tucci; E Walsh; S A Watters; M J Welland; E Williams; K Witkowska; S M Wood; M Zarowiecki.

**Contributors**

ATP and JCT conceived the work. VR, ATP and RLH performed data analysis. EG provided bioinformatics support and, along with MAMcC, gave critical comments. JRS, MS, AG, J-McC, DO and AEF recruited patients and reviewed clinical information. VR, ATP, MAMcC and JCT drafted the manuscript, which was revised and approved by all authors.

**Funding**

The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre Programme and the Wellcome Trust (203141/Z/16/Z). This research was also made possible through access to the data and findings generated by the 100,000 Genomes Project. The 100,000 Genomes Project is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The 100,000 Genomes Project is funded by the NIHR and National Health Service (NHS) England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The 100,000 Genomes Project uses data provided by patients and collected by the NHS as part of their care and support.

**Competing interests**

None declared.

**Patient consent for publication**

Not required.

**Ethics approval**

HRA Committee East of England, Cambridge South (REC: 14/EE/1112).

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Supplemental material**

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access**

This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

**ORCID iDs**

Alistair T Pagnamenta http://orcid.org/0000-0001-7334-0602
Andrew E Fry http://orcid.org/0000-0001-9778-6924
REFERENCES

1. Cushion TD, Pociorkowski AR, Pilz DT, Mullins JGL, Seltzer LE, Marion RW, Tuttle E, Ghoneim D, Christian SL, Chung S-K, Rees MJ, Dobyns WB. De novo mutations in the beta-tubulin gene TUBB2A cause simplified gyral patterning and infantile-onset epilepsy. Am J Hum Genet 2014;94:634–41.

2. Brock S, Vanderhasselt T, Vermaning S, Keymolen K, Régal L, Romaniello R, Wieczorek D, Storm TM, Schaeferhoff K, Hetz U, Kuechler A, Krägeloh-Mann I, Haack TB, Kasteleijn J, Schot R, Mancini GMS, Webster R, Mohammad S, Leventer RJ, Mirzaa G, Dobyns WB, Bahi-Buisson N, Meuwissen M, Jansen AC, Stouffs K. Defining the phenotypical spectrum associated with variants in TUBB2A. J Med Genet 2021;58:33–40.

3. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, Halai D, Baple E, Craig C, Hamblin A, Henderson S, Patch C, O'Neill A, Devereau A, Smith K, Martin AR, Sosinsky A, McDonagh EM, Sultana R, Mueller M, Smedley D, Toms A, Dinh L, Fowler T, Bale M, Hubbard T, Rendon A, Hill S, Caulfield MJ, Genomes P, 100000 Genomes Project. The 100000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ 2018;361:k1687.

4. Bahi-Buisson N, Poirier K, Fourniol F, Saillour Y, Valence S, Lebrun N, Huly M, Bianco CF, Boedraert N, Elie C, Lascelles K, Souville I, Beldjord C, Chelly J. LIS-Tubulinopathies Consortium. The wide spectrum of tubulinopathies: what are the key features for the diagnosis? Brain 2014;137:1676–700.

5. Romaniello R, Arrigoni F, Fry AE, Bassi MT, Rees MJ, Borgatti R, Pilz DT, Cushion TD. Tubulin genes and malformations of cortical development. Eur J Med Genet 2016;61:744–54.

6. Cai S, Li J, Wu Y, Jiang Y. De novo mutations of TUBB2A cause infantile-onset epilepsy and developmental delay. J Hum Genet 2020;65:601–8.

7. Retterer K, Jussola J, Cho MT, Vitazka P, Millan F, Gibellini F, Bertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bii R, Suchy S, Friedman B, Tahiliani I, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S. Clinical application of whole-exome sequencing across clinical indications. Genet Med 2016;18:696–704.

8. Rodan LH, El Achkar CM, Berry GT, Poduri A, Prabhu SP, Yang E, Anselm I. De novo TUBB2A variant presenting with anterior temporal Pachygyria. Child Neurol 2017;32:127–31.

9. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee A, ACMG Laboratory Quality Assurance Committee. 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;17:405–23.

10. Ebbert MTW, Jensen TD, Jensen-West K, Sens JP, Reddy JS, Ridge PG, Kauwe JSK, Belzil V, Pregent L, Carraquillo MM, Keene D, Larson E, Crane P, Asmann YW, Ertekin-Taner N, Younkin SG, Ross OA, Rademakers R, Petrucelli L, Fryer JD. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol 2019;20:97.