

Long-read genome sequencing resolves a complex structural variant involving TBCD and exposes a gap in existing variant classification frameworks

Joseph Leslie

j.leslie@exeter.ac.uk

University of Exeter <https://orcid.org/0000-0003-0972-8818>

Alistair Pagnamenta

University of Exeter <https://orcid.org/0000-0001-7334-0602>

Laura Harrold

University of Exeter

Nishanka Ubeyratna

University of Exeter

Lucy Mallin

Exeter Genomics Laboratory

Olivia Wenger

New Leaf Center

Ethan Scott

New Leaf Center

Erik Puffenberger

Clinic for Special Children

Emma Cassidy

Salisbury District Hospital

Martina Owens

Exeter Genomics Laboratory

Emma Baple

University of Exeter <https://orcid.org/0000-0002-6637-3411>

Brief Communication

Keywords: intragenic, duplication, cismorphism, HiFi, PacBio, TBCD, Amish, founder effect

Posted Date: May 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9628820/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is no duality of interest

Abstract

Intragenic duplications can lead to loss-of-function by disrupting the native gene sequence, however, understanding the precise structural conformation is important for accurate predictions of functional outcomes. Here, we describe an Amish kindred with a positive history of Galloway-Mowat and pontocerebellar hypoplasia syndromes. The youngest child also harboured a homozygous intragenic exon 8–13 duplication involving *TBCD*, initially reported as likely-pathogenic. HiFi long-read genome sequencing and *cismorphism*-based phasing (spans of 24bp–9.8kb), demonstrated the variant to be a non-tandem duplication-inversion-duplication (DUP-INV-DUP) that leaves the native *TBCD* gene intact. Integration of long-read data with founder-population frequency strongly indicated that this DUP-INV-DUP was benign, however a formal classification using ACGS/ACMG guidelines for interpreting CNVs was not possible, as none of the specified evidence criteria are applicable. These results highlight the utility of long-read genome sequencing for fully characterising complex structural variants for improved pathogenicity assessment and expose a gap in existing SV/CNV interpretation frameworks.

Introduction

Chromosomal gains may cause disease through various mechanisms. Whole gene duplications may be pathogenic due to increased overall gene product (triplosensitivity), whereas intragenic duplications can lead to loss-of-function by disrupting the sequence of the native gene. Gene-level metrics are available for predicting which chromosomal gains may cause disease due to triplosensitivity¹. Conversely, to interpret the effect of intragenic duplications, the precise structural configuration must be determined. Short-read sequencing technologies have inherent limitations in resolving the precise structure of complex structural variants (SVs), and so intragenic duplications are frequently assumed to be in tandem, i.e. one copy inserted immediately after the other. However, it is recognised that ~ 20% of rare, intragenic duplications do not involve a direct-tandem configuration². In these scenarios, the original copy of the gene may be unaffected, and the partial gene copy is typically not expressed. Failure to resolve the arrangement of a complex SV may lead to incorrect variant classification and thus to misdiagnosis.

Long-read sequencing can resolve complex SVs through superior mapping, spanning complex arrangements, and improved phasing^{3,4}. As these technologies continue to be integrated into research and diagnostic settings, there are numerous examples where improved SV calling facilitated diagnoses that would have been missed with short-read sequencing alone⁵. However, an under-utilised benefit of long-read sequencing is the converse; resolution of non-tandem sub-genic SVs that leave the native gene intact, that may otherwise be reported as pathogenic. Here, we describe an individual misdiagnosed with *TBCD*-related encephalopathy due to identification of an intragenic duplication via exome sequencing. We demonstrate the utility of HiFi genome sequencing in comprehensively resolving the structure of the variant, which does not disrupt the *TBCD* gene, allowing for variant reclassification.

Patient and methods

Patient description

This individual was born to second-cousin Amish parents. At 3 weeks of age, he presented with poor feeding and weight gain, with irritability (head/back arching) after feeding. He had a normal head circumference. Family history was positive for two siblings (Fig. 1) with Galloway-Mowat syndrome (*WDR73*, MIM 251300) and one sibling with a dual diagnosis of pontocerebellar hypoplasia (*EXOSC3*, MIM 614678) and *TBCD*-related encephalopathy (MIM 617193), both of which are associated with irritability in the newborn period. Concern for inheritance of one of these conditions prompted genetic testing at a CLIA-accredited diagnostic laboratory. Detailed sequencing methodology and testing results are available as **Supplementary methods**.

Results

Exome sequencing identifies a sub-genic duplication involving *TBCD*

Read depth analysis identified the familial chromosomal gain involving exons 8–13 of *TBCD* (NM_005993.5). Results indicated a total of 4 copies, suggesting a homozygous duplication (Fig. 1). As two direct tandem copies of exons 8–13 would be predicted to result in a frameshift [NM_005993.5:r.772_1318dup, p.(Val440GlyfsTer5)], the variant was reported as likely pathogenic (LP) and the individual received a diagnosis of autosomal recessive *TBCD*-related encephalopathy. Two caveats were noted on the report. Firstly, the 4 copies could also have been made up from a triplication *in-trans* with a normal allele and so establishing phase was recommended. Secondly, it is possible that the variant may not have involved a direct-tandem duplication, in which case the original copy of *TBCD* is likely unaffected.

TBCD's 39 exons lie within 17q25.3. Biallelic pathogenic variants in *TBCD* cause a progressive early-onset encephalopathy (MIM 617193)^{6–8}. The typical presentation comprises severely delayed psychomotor development apparent in the first months of life, neurodevelopmental delay, seizures, progressive microcephaly and hypotonia or spasticity. However, on further examination of the child at 5 months of age, he had normal growth parameters including head circumference and was meeting all developmental milestones. Neurological and systemic evaluations were unremarkable. No parental concerns were raised and there was no suspicion of seizure activity. Given this, we interrogated genome/exome sequencing data from 400 Amish and Mennonite individuals and identified two additional homozygotes. The first was a 3-year-old male who presented with moderate developmental delay, mild hypotonia, and febrile seizures. He had normal growth parameters, was non-dysmorphic, and had not received a genetic diagnosis. The second was an unrelated healthy mother.

Given the discordance with the expected disease trajectory and the identification of at least one homozygote without features of *TBCD*-related disorder, we recruited the individual to our research

programme.

HiFi sequencing reveals a complex SV involving TBCD with three possible structures

We performed PacBio HiFi genome sequencing to delineate the precise structure of the *TBCD* variant. Seven autosomal ROHs of > 5.0 Mb (**Figure S1**) were identified. A focussed analysis of 89,572 high-quality SNVs on chromosome 17 detected three regions of homozygosity, with *TBCD* lying in the most distal of these (Fig. 2A). This 1.70 Mb ROH region on 17q25.3 was consistent with the 4 copies being due to a homozygous duplication lying within a region that is identical-by-descent.

The duplication involving *TBCD* was resolved to 42,069bp (NC_000017.11:g.82,792,115_82,834,184 [CN = 4], NM_005993.5:c.772–5642_1318 + 19250dup). However, breakpoint analysis indicated that this was interlinked with a second, more distal duplication of 221,713bp. The distance between the two duplications was 149,025 bp. Although the larger of the two duplications was called using HiFiCNV (v1.0.0), coordinates were imprecise (Fig. 2B). SV calling, performed using pbsv v2.9.0 (<https://github.com/PacificBiosciences/pbsv>) detected reciprocal breakend (BND) calls (chr17:82,792,115 – 82,983,209 and chr17:82,834,184 – 83,204,922, GRCh38) that coincided with increased read coverage and confirmed the two duplications to be linked in a complex duplication-inversion-duplication (DUP-INV-DUP) structure^{9,10}. The BND calls and the copy number status could be explained by three possible structures (Fig. 2C-E), with only two of these resulting in *TBCD* disruption (Fig. 3A).

Haplotype phasing using cismorphisms reveals intact TBCD gene

The mean HiFi read length obtained was 8,189bp (N50 = 7,798bp), so we were unable to span the 42 kb segment and capture both breakpoints in a single read. However, 14 *cis*-morphic SNVs within the duplication (**Table S1**) meant that haplotype assignments from HiPhase¹¹ could be used to infer that the two breakpoints lay on either side of the same (more distal) exon 8–13 duplicated segment (**Figure S2**). The maximum jump between informative SNVs was 9,788 bp and the results of algorithmic phasing were confirmed manually in IGV (Fig. 3B). Given these results point to the structure where the native *TBCD* gene is intact, the child does not show clinical signs of *TBCD*-associated disorder, and we had identified other unaffected homozygous individuals, our assessment of this complex SV was to downgrade from LP to LB (data was shared with the reporting laboratory, ClinVar:SCV007541313). However, a formal classification using the ACGS/ACMG CNV interpretation guidelines was not possible as none of the evidence criteria listed in the guidelines are applicable^{12,13}.

Discussion

As well as being harder to detect than their deletion counterparts, duplications are more challenging to interpret¹⁴. As demonstrated here, the extra DNA copy can be inserted further along the chromosome, leaving it unlikely to cause gene disruption unless it has been inserted into another gene. This necessitates accurate resolution of the duplication's structure to interpret pathogenicity. The interlinking

of two duplicated segments further complicates interpretation, and we required long-read data to resolve the structure.

Despite determining the precise configuration of the *TBCD* SV, formal classification of benign under current ACGS/ACMG guidelines for interpreting CNVs was not possible¹⁵. Intragenic duplications involving coding exons are interpreted similar to SNVs, which supported the original LP classification based on presumed tandem duplication and predicted nonsense-mediated decay (PVS1_strong). After long-read sequencing demonstrated that the native *TBCD* gene is intact, we reassessed the variant under CNV guidance. However, current frameworks lack positive benign criteria to apply as evidence for fully resolved non-disruptive structural variation.

Benign non-tandem intragenic duplications have been observed before. Notably, several studies have identified such duplications in the *DMD* gene that have been aberrantly associated with Duchenne's muscular dystrophy^{16,17}. Accurately establishing the precise structure of these variants becomes increasingly important in the context of carrier screening and newborn screening, where the absence of a compatible phenotype complicates interpretation¹⁸. Interestingly, we identified several carriers of a similar SV in *TBCD* in UK Biobank (**Figure S3, S4, Note S1**), and thus our findings provide generalizable lessons for SV interpretation.

Due to the founder effect, many geographically isolated or endogamous populations are enriched for particular sets of genetic variants that are rare globally but comparatively common in local communities¹⁹. Many founder variants are already known in the Amish and Mennonite populations, and newly described variants can be introduced alongside other founder variants into screening efforts²⁰. Conversely, founder variants that are difficult to interpret have high potential to be falsely assessed and reported as pathogenic, which may cause unnecessary distress to families who are led to believe that they/their child will develop a severe disorder. Compiling a reciprocal list of cryptic founder variants that are benign but have high potential for generating false assertions is an equally important endeavour.

In summary, using long-read sequencing and *cismorphism* based phasing, we deciphered the structure of a complex SV involving *TBCD*. Reclassification of this founder variant will improve the accuracy of genetic diagnostics in Amish populations. This work underscores the critical importance of characterising the breakpoints of intragenic duplications to help improve variant interpretation, and highlights gaps in current CNV interpretation guidelines for classifying non-disruptive structural variants as benign.

Declarations

Author Contribution Statement

ATP, LH, and JSL led the data analysis and manuscript drafting. All authors (ATP, LH, NU, LM, OKW, EMS, MO, EJC, EGP, ELB and JSL) reviewed the manuscript and contributed essential intellectual input. NU

performed wet lab sequencing protocols. LM performed bioinformatic processing of sequencing data. OKW, EMS and ELB provided and interpreted clinical data. JSL and ATP oversaw the project.

Ethical Approval

This study adhered to the principles of the Declaration of Helsinki and was approved by the Akron Children's Hospital institutional review board (IRB) (#986876) with all research participants or their legal guardians providing written informed consent for study participation and the publication of clinical details. Affected individuals were identified through their clinical care provider.

Competing Interests

The authors declare no competing interests.

Acknowledgements

First and foremost, we are grateful to the family for their involvement in this study. We thank Tim Hall for sharing code to assess CNVs in the UK Biobank Research Analysis Platform. This research was funded by the NHS Rare and Inherited Disease Genomic Network of Excellence and the National Institute for Health and Care Research (NIHR) Exeter Biomedical Research Centre (BRC). JSL is funded through an NIHR Exeter BRC Translational Fellowship. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. UK Biobank data access was obtained under the approved UKB project 'Understanding the role of rare and common genetic variation in human phenotypes', application number 103356.

Data Availability Statement

Data that support the findings of this study are available on request from the corresponding author with reasonable patient privacy restrictions.

References

1. Collins RL, Glessner JT, Porcu E, Lepamets M, Brandon R, Lauricella C *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* 2022; 185: 3041–3055.e25.
2. Newman S, Hermetz KE, Weckselblatt B, Rudd MK. Next-Generation Sequencing of Duplication CNVs Reveals that Most Are Tandem and Some Create Fusion Genes at Breakpoints. *The American Journal of Human Genetics* 2015; 96: 208–220.
3. Yuan Y, Zhou X, Deng J, Zhu Q, Peng Z, Chen L *et al.* Case report: Long-read sequencing identified a novel 14.9-kb deletion of the α -globin gene locus in a family with α -thalassemia in China. *Front Genet* 2023; 14: 1156071.

4. Paivandy A, Lenner F, Eisfeldt J, Jonson T, Ehrencrona H, Lindstrand A et al. Flexible and rapid validation of structural variation using adaptive sampling. *Eur J Hum Genet* 2026. doi:10.1038/s41431-026-02039-4.
5. Miller DE, Sulovari A, Wang T, Loucks H, Hoekzema K, Munson KM et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet* 2021; 108: 1436–1449.
6. Miyake N, Fukai R, Ohba C, Chihara T, Miura M, Shimizu H et al. Biallelic TBCD Mutations Cause Early-Onset Neurodegenerative Encephalopathy. *The American Journal of Human Genetics* 2016; 99: 950–961.
7. Flex E, Niceta M, Cecchetti S, Thiffault I, Au MG, Capuano A et al. Biallelic Mutations in TBCD, Encoding the Tubulin Folding Cofactor D, Perturb Microtubule Dynamics and Cause Early-Onset Encephalopathy. *The American Journal of Human Genetics* 2016; 99: 962–973.
8. Ocampo-Chih C, Dennis H, Lall N, Pham N, Liang B, Verma S et al. PEBAT, an Intriguing Neurodegenerative Tubulinopathy Caused by a Novel Homozygous Variant in TBCD: A Case Series and Literature Review. *Pediatr Neurol* 2023; 139: 59–64.
9. Jung H, Yang T-P, Walker S, Danecek P, Garcia-Salinas OI, Neville MDC et al. Complex de novo structural variants are an underestimated cause of rare disorders. *Nat Commun* 2025; 16: 9528.
10. 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: 1235–1244.
11. Holt JM, Saunders CT, Rowell WJ, Kronenberg Z, Wenger AM, Eberle M. HiPhase: jointly phasing small, structural, and tandem repeat variants from HiFi sequencing. *Bioinformatics* 2024; 40. doi:10.1093/bioinformatics/btae042.
12. Durkie M, Cassidy E-J, Berry I, Owens M, Turnbull C, Scott RH et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2024. 2024<http://gnomad.broadinstitute.org/>.
13. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A et al. 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* 2020; 22: 245–257.
14. Retterer K, Scuffins J, Schmidt D, Lewis R, Pineda-Alvarez D, Stafford A et al. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genetics in Medicine* 2015; 17: 623–629.
15. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A et al. 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* 2020; 22: 245–257.
16. Ding Q, Balan J, Vidal-Folch N, Pickart AM, Sun G, Walsh JR et al. Rethinking the pathogenicity of intragenic DMD duplications detected by carrier screening: High prevalence of nontandem duplications revealed by long-read sequencing. *Genetics in Medicine* 2025; 27: 101539.

17. Seiwert EU, Shi X, Cayan A, Crawford S, Husami A, Tian C et al. Reclassification of intragenic DMD gene duplications by optical genome mapping resolves uncertainty and improves clinical management. *Neuromuscular Disorders* 2026; 60: 106335.
18. Horton R, Wright CF, Firth H V, Turnbull C, Lachmann R, Houlston RS et al. Challenges of using whole genome sequencing in population newborn screening. *BMJ* 2024; 384: e077060.
19. Salter CG, Cai Y, Lo B, Helman G, Taylor H, McCartney A et al. Biallelic *PI4KA* variants cause neurological, intestinal and immunological disease. *Brain* 2021; 144: 3597–3610.
20. Crowgey EL, Washburn MC, Kolb EA, Puffenberger EG. Development of a Novel Next-Generation Sequencing Assay for Carrier Screening in Old Order Amish and Mennonite Populations of Pennsylvania. *The Journal of Molecular Diagnostics* 2019; 21: 687–694.

Figures

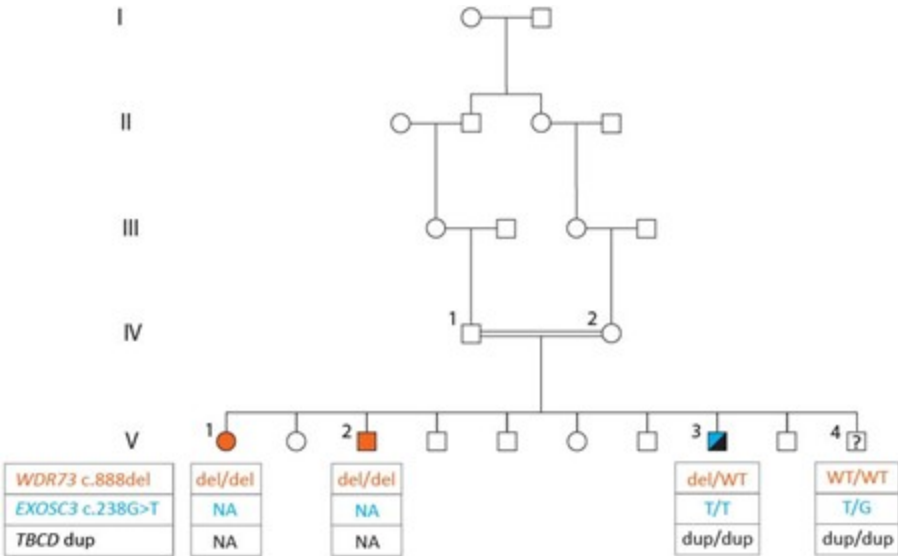


Figure 1

Positive family history for three recessive disorders informs exome sequencing revealing a pathogenic biallelic sub-genic duplication in *TBCD*. Clinical suspicion of genetic disorders in V:1, V:2 and V:3 led to exome sequencing performed by Invitae. V:1 and V:2 both received diagnoses of Galloway-Mowat syndrome due to a homozygous pathogenic Amish founder variant in *WDR73* [NM_032856.5:c.888del, p.(Phe296LeufsTer26)] (affected status is indicated by orange shapes). V:3 is homozygous for both a pathogenic Amish founder variant in *EXOSC3* [NM_016042.4:c.238G>T p.(Val80Phe)], and a likely pathogenic sub-genic duplication involving exons 8-13 of *TBCD*, which if in tandem, predicts a frameshift [NM_005993.5:r.772_1318dup, p.(Val440GlyfsTer5)]. They therefore received a dual diagnosis of pontocerebellar hypoplasia type 1B (indicated in blue) and *TBCD*-related encephalopathy (indicated in black). Suspicion of a genetic disorder (noted as ?) in the proband of this study (V:4) combined with the extensive positive family history of genetic disease informed exome sequencing by Invitae, which revealed the same homozygous sub-genic duplication impacting *TBCD* as present in V:3. Segregation of

the three variants in all individuals who underwent exome sequencing is noted at the bottom of the pedigree. Del = deletion, WT = wildtype, dup = duplication, NA = individuals not tested for this variant.

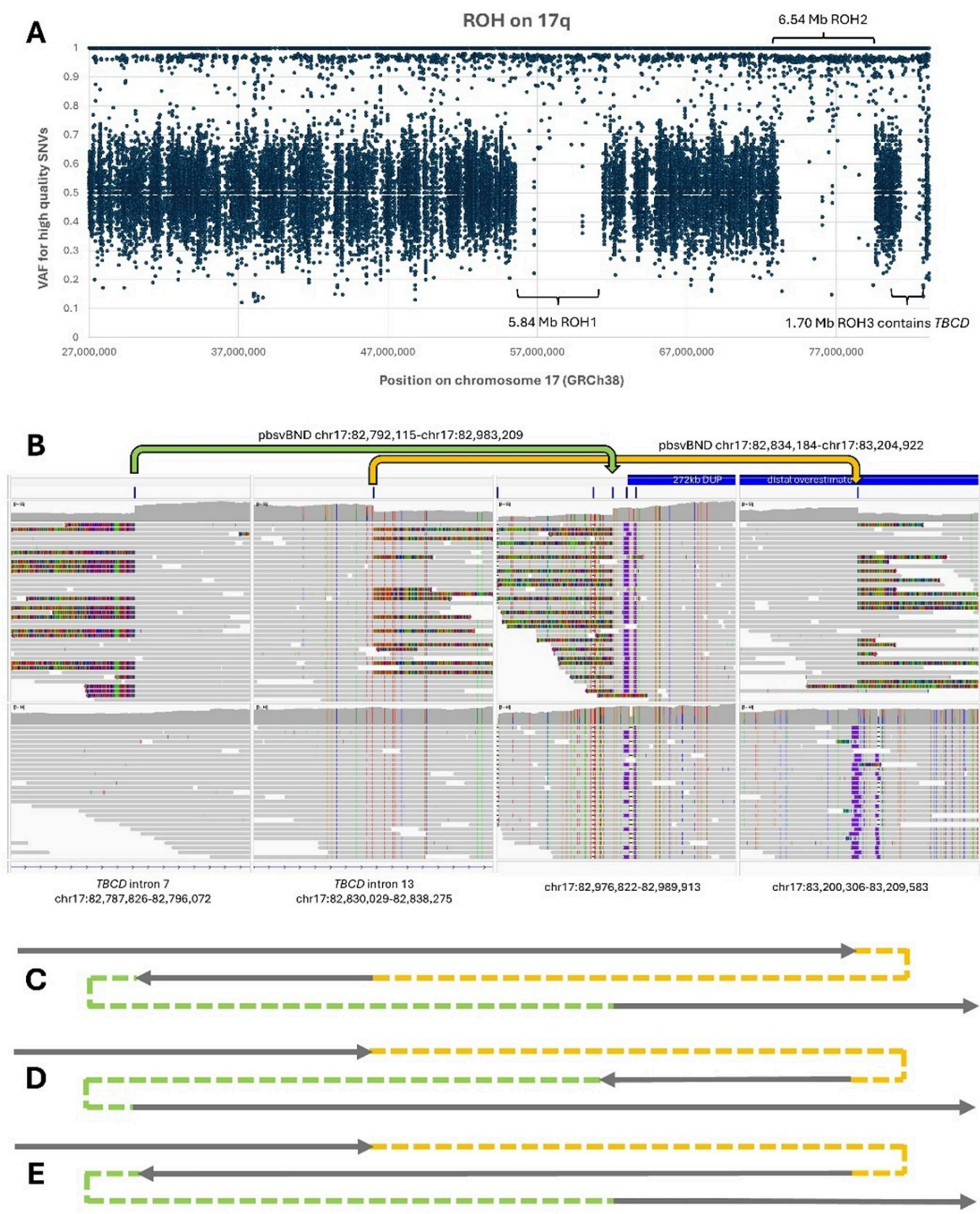


Figure 2

PacBio HiFi genome sequencing identifies regions of homozygosity on 17q that overlaps a non-tandem *TBCD* duplication. A) Analysis of SNVs detected identified 3 regions of homozygosity on 17q. A set of

high confidence variants was generated by filtering the small variant vcf (called by DeepVariant v1.6.0) for SNVs that were i) flagged as PASS and ii) for which the sum of the allelic depths was 20-50x. SNVs thought to define the ROH regions were then checked in IGV. The GRCh38 coordinates of the three ROH regions are 17:55,521,126-61,361,697, 17:73,087,812-79,627,302 (ROH2) and 17:81,317,159-83,017,161 (ROH3). The proximal end of ROH2 was a SNP at 19x but identified on visual analysis. The third ROH region contains the *TBCD* duplication. B) HiFi read alignments shown in IGV that support two interlinked duplications on 17q25.3. The positions of BND calls are highlighted. The distal duplication, although called by HiFiCNV is overestimated at the distal end and underestimated at the proximal end. C-E) The two interlinked duplications are inverted with respect to one another and there are 3 possible structures that could explain the data as shown in the subway plots. Dotted lines represent breakpoints and are colour-coded to match the respective BND calls shown in panel B.

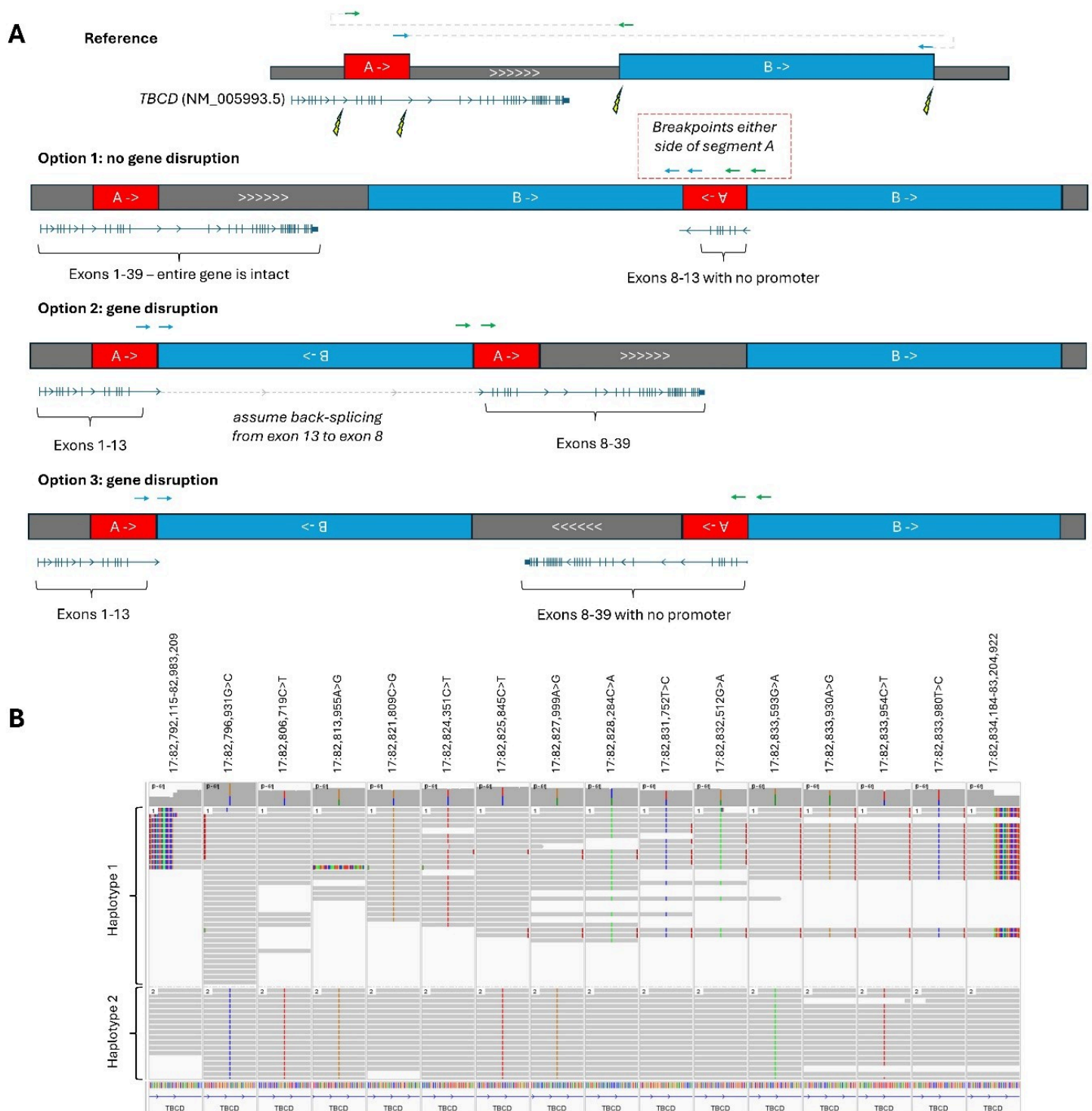


Figure 3

Resolving ambiguous structure using *cis*-morphism analysis aids variant interpretation. A) Schematic diagram showing how the split reads can be explained by three possible structures of which Options 2 and 3 are predicted to disrupt gene function. Option 1, where the inverted partial copy of *TBCD* is inserted more distally, would not disrupt the original copy of *TBCD*. Assuming efficient back-splicing from exon 13 to exon 8, Option 2 would predict the same consequence as if the SV had been a simple direct-tandem duplication, i.e. a frameshift [NM_005993.5:r.772_1318dup, p.(Val440GlyfsTer5)]. B) Split-

window IGV graphic showing read alignments which support the 14 *cis*-morphisms (**Table 1**) used for algorithmic phasing and that helped distinguish the proximal and distal copies of the 42 kb duplication. As the breakpoints involve the same duplicated segment, the structure “Option 1” shown in panel 2A (and the subway plot in **Figure 1C**) appears to represent the correct SV configuration.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TBCDSupplementarymaterials050526.docx](#)