COMMUNICATION

Low grade mosaicism in hereditary haemorrhagic telangiectasia identified by bidirectional whole genome sequencing reads through the 100,000 Genomes Project clinical diagnostic pipeline

Jessica M Clarke,[1,2] Mary Alikian,[1,2] Sihao Xiao,[2,3] Dalia Kasperaviciute,[4] Ellen Thomas,[1,4] Isobel Turbin,[1,3] Kike Olupona,[1] Elna Cifra,[1] Emanuel Curetean,[1] Teena Ferguson,[1] Julian Redhead,[1] The Genomics England Research Consortium,[4] Claire L Shovlin [1,2,3]

[1]West London Genomic Medicine Centre, Imperial College Healthcare NHS Trust, London, UK
[2]Genomics England Respiratory Clinical Interpretation Partnership (GeCIP), London, UK
[3]NHLI Cardiovascular Sciences, Imperial College London, London, UK
[4]Genomics England, London, UK

Correspondence to
Claire L Shovlin, Professor of Practice (Clinical and Molecular Medicine), NHLI Vascular Science, Imperial Centre for Translational and Experimental Medicine (ICTEM), Hammersmith Campus, London W12 0NN, UK;
c.shovlin@imperial.ac.uk

Check for updates

Whole-genome sequencing (WGS) has been championed within the UK National Health Service (NHS) and represents one of the approaches within the forthcoming UK National Genomic Test Directory to identify genetic variants that cause particular rare inherited diseases.[1]

Although diploid organisms such as man develop from a single cell, postzygotic somatic mutations occur and lead to mosaicism.[2] Sufficiently early generation of a disease-causing DNA sequence variant can result in a mosaic individual who is the first member of a family to be affected by an inherited disorder.[3 4] Their causative DNA sequence variant is likely to be present at less than the 50% expected for a heterozygote, and may be difficult to detect with confidence. Greater ability to detect mosaicism has been used to favour higher depth panel-based sequencing rather than WGS as, due to sequencing capacities, there are trade-offs between the number of target nucleotides sequenced and the average depth of reads at any given nucleotide.

It is increasingly recognised that hereditary haemorrhagic telangiectasia (HHT[5]) is a condition where the first affected member of the family may be a mosaic.[3 4] HHT is transmitted as an autosomal dominant trait via a single pathogenic DNA sequence variant, usually in *ENG*, *ACVRL1* or *SMAD4*.[5] There are hundreds of different pathogenic variants in these genes in different HHT families, all resulting in similar clinical features.[6] HHT is diagnosed clinically by the presence of at least three Curaçao Criteria: recurrent nosebleeds, mucocutaneous telangiectasia, visceral involvement such as pulmonary, hepatic, gastrointestinal or cerebral arteriovenous malformations (AVMs), and an affected first-degree relative.[5] A high index of suspicion for HHT is warranted, particularly in cases where more than one AVM is present or where multiple generations display these features. As we have shown, however, high proportions of referrals to mainstream services with pulmonary AVMs have no identified HHT causal genes on standard gene testing, even when there is clear evidence of HHT clinically.[7]

The proband was one of these[7] cases, a 54-year-old woman who had presented following major complications from pulmonary AVMs including a cerebral abscess and profound hypoxaemia. The pulmonary AVMs were treated by embolisation, and she was also advised to use antibiotic prophylaxis prior to future dental and surgical procedures.[5] On clinical assessment, she reported frequent nosebleeds and displayed characteristic HHT buccal mucosal telangiectasia. There was no antecedent family to suggest HHT across an extensive family history, although several individuals in the subsequent generations experienced nosebleeds. Fulfilling three Curaçao Criteria, the proband received a clinical diagnosis of 'Definite HHT'. Descendants with nosebleeds had two Curaçao Criteria and received a label of 'Possible HHT'. Using routine methodologies and processes including bidirectional Sanger sequencing, HHT gene testing for *ENG*, *ACVRL1* and *SMAD4* did not identify a genetic cause for disease in the proband. Due to the absence of a molecular test for the family, the condition could not be excluded in any descendants.

The proband was recruited to the 100,000 Genomes Project.[8] Following WGS of DNA extracted from a second sample of peripheral blood and sequence alignment to Genome build GRCh38 and ENST00000373203, the recruiting Genomic Medicine Centre (GMC) was informed that she had a Tier 2 variant in *ENG*, the gene most commonly responsible for pulmonary AVM-associated HHT. The synonymous variant *ENG* c.1134G>A, p.(Ala378=), rs1329127701, has a Genome Aggregation Database (gnomAD) frequency of 0.00003, substitutes an A for the final G of *ENG* exon 8, and is predicted by five splicing programmes to reduce splice site efficiency by a mean of 33% (figure 1). The variant is listed as pathogenic on the HHT Mutation Database,[6] and ClinVar.[9] At the multidisciplinary team meeting, it was concluded that the variant could explain the patient's phenotype, and DNA was sent for Sanger sequencing validation.
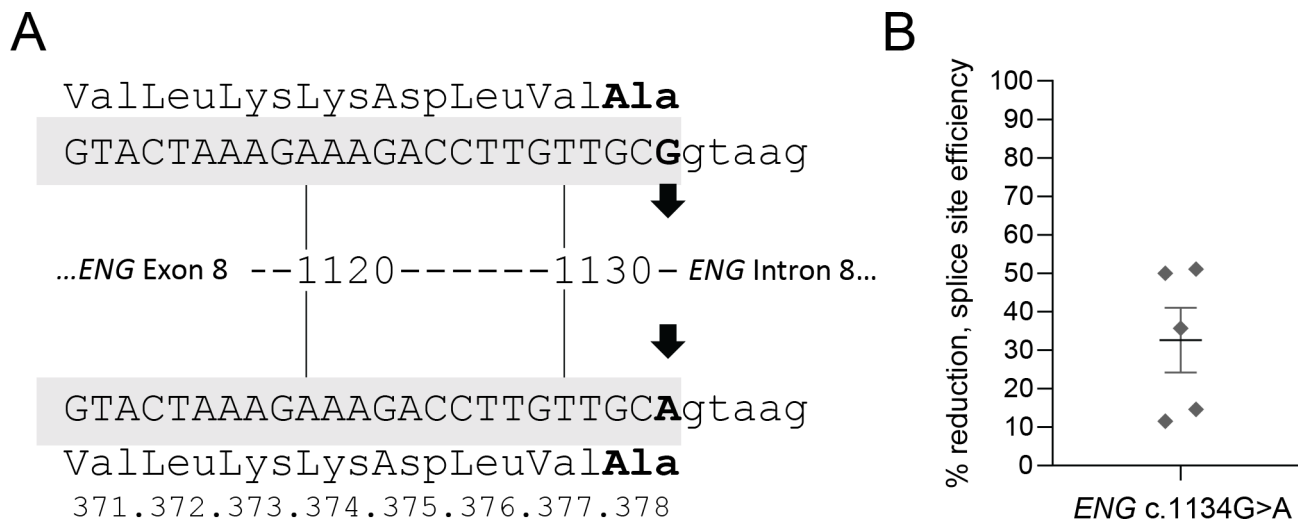
OPEN ACCESS

BMJ

Figure 1 Variant consequences: (A) schematic of variant *ENG* (NM_001114753) c.1134G>A, p.(Ala378=) and (B) splice site predictions: the symbols indicate from highest to lowest, the percentage (%) reduction in splice site efficiency calculated by SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer and HumanSplicingFinder. All five exceeded the 10% reduction considered significant, and three exceeded 30% reduction. The variant therefore meets sufficient American College of MedicalGenetics and Genomics (ACHG) / Association for Molecular Pathology (AMP) criteria[10] to be classified as 'likely pathogenic': 1 strong criterion (Ps4: Genome Aggregation Database frequency 0.00003), plus ≥2 supporting criteria (PP3: predicted by five splicing programmes to reduce splice site efficiency; PP4: highly specific patient phenotype; and PP5: listed as pathogenic on two reputable sources[7 9]).

As in the original NHS gene test however, the variant was not initially confirmed by Sanger sequencing, and the case was referred to the Respiratory Genomics England Clinical Interpretation Partnership.

Careful review of traces on Genomics England's Integrative Genomics Viewer indicated that the variant nucleotide was not present at the 50:50 distribution expected for an autosomal dominant disorder (figure 2A). The variant was identified on forward and reverse reads, suggesting it was unlikely to be a sequencing artefact (figure 2B). Subsequent Sanger sequencing identified the signal from the variant nucleotide at a level that could be easily disregarded as background noise in Sanger-based testing without prior knowledge (figure 2C).

The 100,000 Genomes Project operates using a second alignment and filtering pipeline for research interrogations. The same weighting of signal:noise escape might be viewed less favourably if applied genome wide than for clinical diagnostic sequencing where a single rare variant is sought in one of the RefSeq genes. We therefore reviewed the individual's variant call format (vcf) file in the Genomics England Research Environment where raw sequencing data had been separately analysed by Illumina Issac for sequence alignment and Starling for small variant calling.[8] The variant was not identified in these more stringent pipelines.

We concluded that *ENG* c.1134G>A, p.(Ala378=) was confirmed in the proband and represents a further case of mosaicism in the first clinically affected member of an HHT family.[3 4] This case also allowed us to conclude that the set of automated pipelines in place through the 100,000 Genomes Project for clinical diagnostic purposes are able to identify low-grade mosaicism. The clinical diagnostic algorithms used Platypus, which employs an allele bias filter for variants based on expectation under heterozygous segregation in a diploid organism. However, Platypus rejects variants only if the fraction of variant reads is less than 0.5 *and* the p value under a binomial model is less than 0.001. In other words, the fraction of variant reads that trigger this filter depends on the total coverage, and as we have shown, for a read depth of 35, a fractional read of 0.2 will escape the filter, increasing the likelihood of detecting pathogenic variants present in mosaics. We suggest such potential mosaic calls could be highlighted to GMCs to facilitate the necessary rigorous Sanger sequence inspections.

On a wider note, there are two important general implications. The first is the message that mosaicism should be kept in mind in cases of inherited diseases such as HHT where no pathogenic variant is identified; other methods to look for mosaicism, for example, examination of an oral mucosa swab, or tissue targeted methods may be considered. Second, the case highlights that traditional Sanger sequencing is unreliable when it comes to mosaic cases. Even with only modest coverage (35X), WGS with a robust bioinformatics pipeline clearly identified the pathogenic variant. Recognising that the cost of next generation sequencing is continuously going down, and throughput is continually going up, we would recommend next-generation sequencing for all clinical testing.

Genomics England Research Consortium Ambrose J. C. 1, Arumugam P.1, Baple E. L. 1, Bleda M. 1, Boardman-Pretty F. 1,2, Boissiere J. M. 1, Boustred C. R. 1, Brittain H.1, Caulfield M. J.1,2, Chan G. C. 1, Craig C. E. H. 1, Daugherty L. C. 1, de Burca A. 1, Devereau, A. 1, Elgar G. 1,2, Foulger R. E. 1, Fowler T. 1, Furió-Tarí P. 1, Hackett J. M. 1, Halai D. 1, Hamblin A.1, Henderson S.1,2, Holman J. E. 1, Hubbard T. J. P. 1, Ibáñez K.1,2, Jackson R. 1, Jones L. J. 1,2, Kasperaviciute D. 1,2, Kayikci M. 1, Lahnstein L. 1, Lawson K. 1, Leigh S. E. A. 1, Leong I. U. S. 1, Lopez F. J. 1, Maleady-Crowe F. 1, Mason J. 1, McDonagh E. M. 1,2, Moutsianas L. 1,2, Mueller M. 1,2, Murugaesu N. 1, Need A. C. 1,2, Odhams C. A. 1, Patch C. 1,2, Perez-Gil D. 1, Polychronopoulos D. 1, Pullinger J. 1, Rahim T. 1, Rendon A. 1, Riesgo-Ferreiro P.1, Rogers T. 1, Ryten M. 1, Savage K. 1, Sawant K. 1, Scott R. H. 1, Siddiq A. 1, Sieghart A. 1, Smedley D. 1,2, Smith K. R. 1,2, Sosinsky A. 1,2, Spooner W. 1, Stevens H. E. 1, Stuckey A. 1, Sultana R. 1, Thomas E. R. A. 1,2, Thompson S. R. 1, Tregidgo C. 1, Tucci A. 1,2, Walsh E. 1, Watters, S. A. 1, Welland M. J. 1, Williams E. 1, Witkowska K. 1,2, Wood S. M. 1,2, Zarowiecki M. 1. 1, Genomics England, London, UK 2. William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.
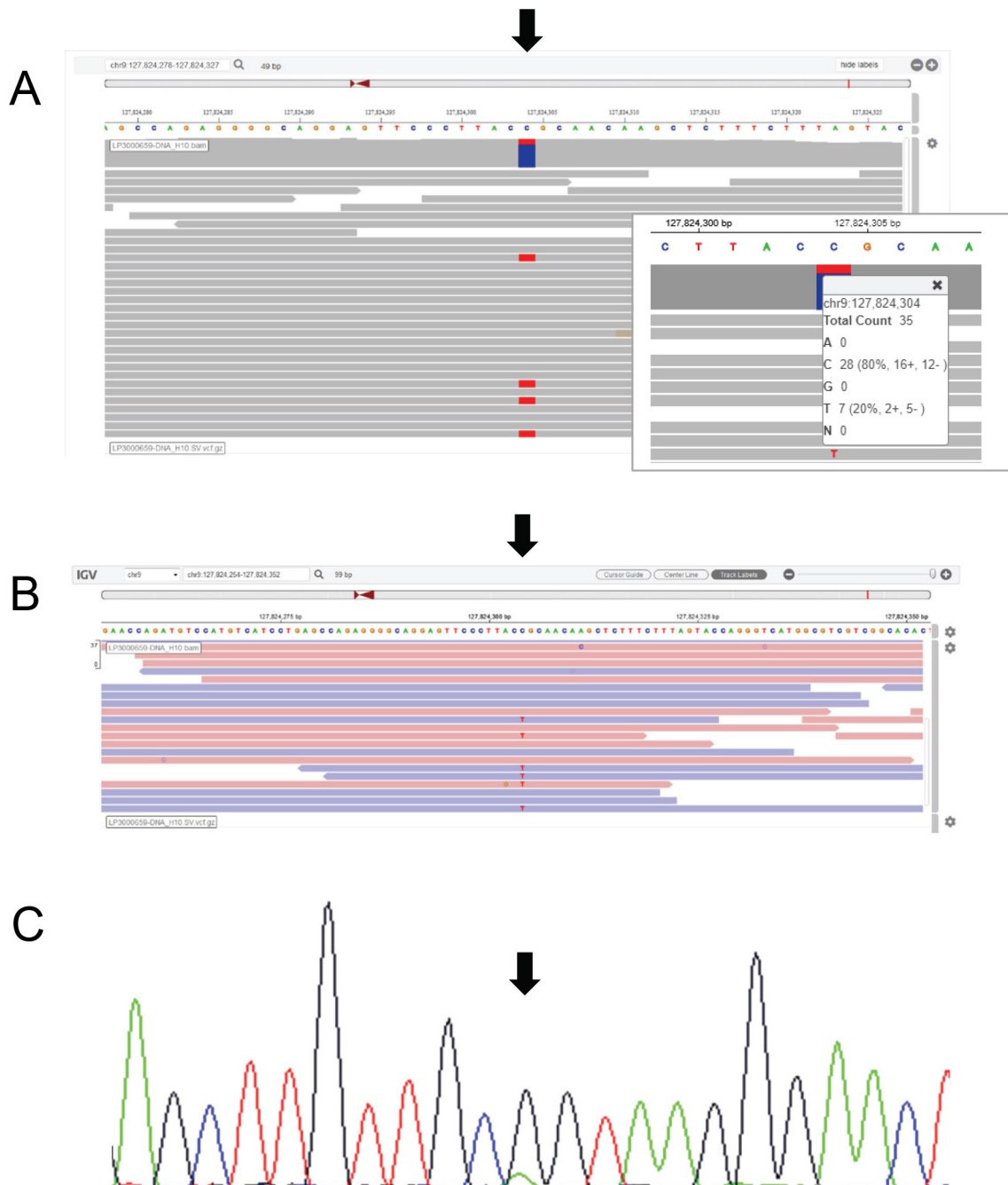
Figure 2 Variant detection: (A) of the 35 reads at Chr9:127 824 304 (arrowed), 28 (80%) were wild type and 7 (20%) were the variant sequence; (B) the 18 forward strand reads, 16 (88.9%) were wild type, 2 (11.1%) were variant. Of the 17 reverse strand reads, 12 (70.6%) were wild type and 5 (29.4%) were variant. (C) The clean Sanger sequencing trace at the locus demonstrating wild-type (black) and variant (green) sequences. The wild-type peak was quantified as 793 RFUs, compared with the variant peak of 121 RFUs representing 15.3% of the total. RFU, relative fluorescence unit.

ORCID iD
Claire L Shovlin http://orcid.org/0000-0001-9007-5775

REFERENCES

1 NHS England. 2019/2020 National Genomic Test Directory, 2019. Available: https://www.england.nhs.uk/publication/national-genomic-test-directories/ [Accessed 26 Feb 2020].

2 Martincorena I. Somatic mutation and clonal expansions in human tissues. *Genome Med* 2019;11:35.

3 Tørring PM, Kjeldsen AD, Ousager LB, Brusgaard K. ENG mutational mosaicism in a family with hereditary hemorrhagic telangiectasia. *Mol Genet Genomic Med* 2018;6:121–5.

4 McDonald J, Wooderchak-Donahue WL, Henderson K, Paul E, Morris A, Bayrak-Toydemir P. Tissue-specific mosaicism in hereditary hemorrhagic telangiectasia: implications for genetic testing in families. *Am J Med Genet A* 2018;176:1618–21.

5 Shovlin CL, Buscarini E, Kjeldsen AD, Mager HJ, Sabba C, Droege F, Geisthoff U, Ugolini S, Dupuis-Girod S. European Reference Network For Rare Vascular Diseases (VASCERN) Outcome Measures For Hereditary Haemorrhagic Telangiectasia (HHT). *Orphanet J Rare Dis* 2018;13:136.

6 University of Utah. HHT Mutation Database, hosted by the ARUP Laboratories. Available: http://www.arup.utah.edu/database/HHT/ [Accessed 26 Feb 2020].

7 Turbin IG, Brownlow S, Shovlin CL. Directing HHT patients to the 100,000 genomes project. *Angiogenesis* 2018;21.

8 Genomics England. The 100,000 Genomes Project. Available: https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/ [Accessed 26 Feb 2020].

9 National Center for Biotechnology Information. ClinVar NM_000118.3(ENG):c.1134G>A (p.Ala378=). Available: https://www.ncbi.nlm.nih.gov/clinvar/variation/458328/ [Accessed 26 Feb 2020].

10 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. 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.