

INSIGHTS

Familial thrombocytopenia: The long and short of it

Lauren Murphy¹ and Adam J. Mead¹

In this issue, Wahlster, Verboon, and colleagues (2021. *J. Exp. Med.* <https://doi.org/10.1084/jem.20210444>) describe a multigenerational family with inherited thrombocytopenia where the causal variant was not identified using conventional genome sequencing approaches. Long-read sequencing and RNA sequencing revealed a complex structural variant, causing overexpression of a pathogenic gain-of-function *WAC-ANKRD26* fusion transcript.

Inherited thrombocytopenia 2 (THC2) is an autosomal dominant disorder characterized by ankyrin repeat domain 26 (*ANKRD26*) mutations, which cause aberrant *ANKRD26* overexpression during megakaryocyte differentiation with consequent impaired platelet production and a predisposition to myeloid malignancies (Noris et al., 2011; Pippucci et al., 2011). While THC2 is rare, distinguishing this condition from other noninherited causes of thrombocytopenia such as myelodysplasia (Kewan et al., 2020) or immune thrombocytopenia is crucial. THC2 is typically caused by single nucleotide variants (SNVs) in the 5' untranslated region of the *ANKRD26* gene. These mutations result in loss of runt-related transcription factor 1 (*RUNX1*) and friend leukemia integration 1 transcription factor (*FLI1*) binding (Bluteau et al., 2014), with consequent failure to repress *ANKRD26* expression during hematopoietic differentiation (see figure).

In this issue, Wahlster, Verboon, and colleagues describe a multigenerational family with inherited thrombocytopenia with a THC2-like phenotype (Wahlster et al., 2021). Based on this clinical phenotype, targeted Sanger sequencing of the *ANKRD26* gene was performed, but no pathogenic variants were identified. Furthermore, no pathogenic variants were identified through targeted sequencing of additional genes known to be mutated in familial thrombocytopenia (*RUNX1*, *GATA1*,

and *MPL*). A number of cases were therefore taken forward for whole-exome sequencing (WES) and/or whole-genome sequencing (WGS), which also failed to detect putative causal variants. However, reanalysis of the sequencing data revealed increased coverage of specific portions of the *ANKRD26* gene, raising the possibility of an underlying complex structural variant (SV).

While short-read WES and WGS is a well-established tool to identify disease-associated SNVs in rare diseases and is now in routine clinical use (Turro et al., 2020), the difficulty in identifying SVs, partly due to their large size, often renders short-read mapping incapable of fully resolving the entire variant (Amarasinghe et al., 2020). Therefore, the authors employed long-read sequencing of a trio within the larger pedigree to enable accurate SV assembly (Amarasinghe et al., 2020). Through combinatorial data analysis, they were able to identify and characterize a duplication spanning exons 10–20 of *ANKRD26*, which was part of a larger, complex paired-duplication inversion SV. Short-read WGS and targeted genotyping by a customized PCR confirmed segregation of this SV with all affected family members across multiple generations and absence of the SV in unaffected family members. Together, these data provide compelling evidence that this SV was pathogenic, causing the thrombocytopenia



Insights from Murphy and Mead.

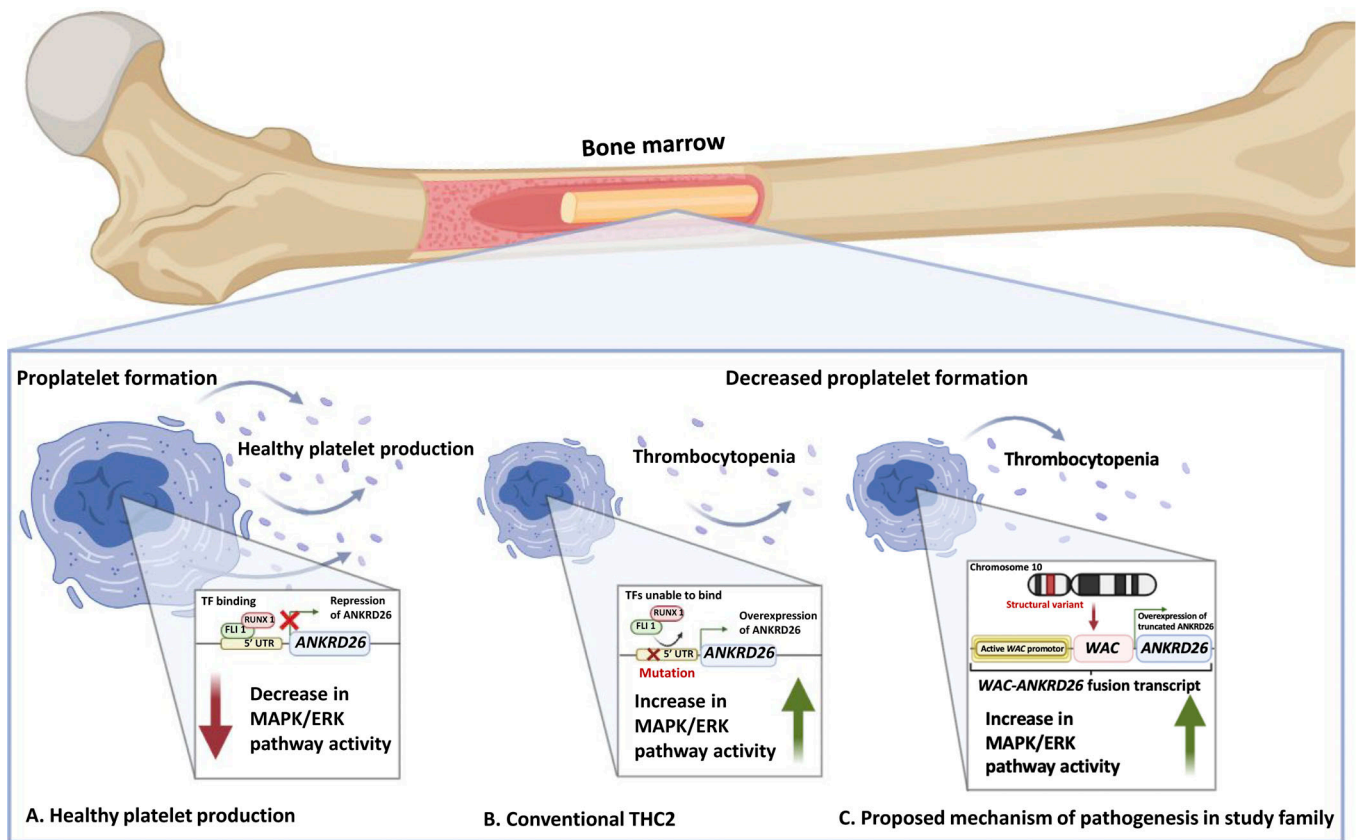
phenotype with high penetrance, although with considerable heterogeneity of thrombocytopenia between affected family members.

Wahlster et al. (2021) next focused on understanding the molecular pathogenesis of the SV. In contrast to previous THC2 cases, sequencing analysis of this SV identified no pathogenic variants in the proximal promoter region of *ANKRD26*. The SV only led to structural changes in two genes, *ANKRD26* and *WAC*, although the *WAC* gene retained an intact open reading frame. The altered adjacent positioning of the genes resulted in a potential fusion transcript between exon 1 of *WAC* and exons 10–34 of *ANKRD26*. Notably, the *WAC* gene is constitutively expressed throughout the hematopoietic system, suggesting that the SV might lead to aberrant overexpression of an N-terminally truncated form of *ANKRD26*. To elucidate this further, the authors performed RNA sequencing of peripheral blood mononuclear cells from three affected individuals and healthy controls. This analysis

Medical Research Council (MRC) Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, National Institute for Health Research Biomedical Research Centre, University of Oxford, Oxford, UK.

Adam J. Mead: adam.mead@imm.ox.ac.uk.

© 2021 Murphy and Mead. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).



Schematic demonstrating normal and aberrant *ANKRD26* expression in healthy individuals and those with TCH2. **(A)** The binding of transcription factors (TF) RUNX1 and FLI1 to the promoter region of *ANKRD26* results in silencing of *ANKRD26* and reduced MAPK-ERK signaling leading to normal megakaryocyte maturation, proplatelet formation, and platelet release. **(B)** A mutation in the 5' untranslated region (UTR) of *ANKRD26* prevents binding of RUNX1 and FLI1 and the repression of the gene. This leads to hyperactivation of the MAPK-ERK pathway, resulting in impaired proplatelet formation in THC2 individuals. **(C)** A complex structural variant results in a WAC-*ANKRD26* fusion transcript encoding a functional truncated form of *ANKRD26*. Overexpression of this gene, driven by the active WAC promoter, likely leads to increased MAPK activation mirroring the phenotype seen in conventional THC2. Adapted from [Bluteau et al., 2014](#) and created with BioRender.com (2021).

revealed transcripts spanning *WAC* exon 1 and *ANKRD26* exon 10 that were present exclusively in the thrombocytopenic individuals. Furthermore, *ANKRD26* was markedly overexpressed in affected individuals, and this was limited to exons contained within the *WAC-ANKRD26* fusion transcript, consolidating previous observations and confirming the SV was responsible for the generation of a partial fusion transcript and the subsequent overexpression of a truncated form of *ANKRD26*.

After confirming the ability of the SV to induce overexpression of a region of the *ANKRD26* gene, [Wahlster et al. \(2021\)](#) explored various isoforms of *WAC* and *ANKRD26* and evaluated the ability of the transcripts to translate stable protein. This was complex due to the presence of multiple isoforms of both the *WAC* and *ANKRD26* genes. Analysis of the RNA sequencing data

supported that the most likely *ANKRD26* isoform encompassed a skipped exon, enabling translation of a truncated form of *ANKRD26* starting from a methionine in exon 11. They compared an in-frame *WAC-ANKRD26* fusion with the full-length *ANKRD26* and a truncated “exon 11+” methionine initiating *ANKRD26* that starts at exon 11 and lacks the preceding ankyrin repeats. The absence of protein expression in HEK 293T cells transfected with *WAC-ANKRD26* fusion cDNA indicated the full fusion transcript was incapable of producing a stable protein. As the full-length and exon 11+ truncated *ANKRD26* transfected cells were able to achieve protein translation, the authors took these forward for further functional studies. cDNAs encoding the full and exon 11+ *ANKRD26* transcripts were delivered into human CD34⁺ hematopoietic stem and progenitor cells (HSPCs)

via lentiviral transduction, and increased expression of *ANKRD26* was confirmed. Intriguingly, upon starvation and restimulation with thrombopoietin, [Wahlster et al. \(2021\)](#) demonstrated, via flow cytometry, a significant increase in ERK phosphorylation correlating with the overexpression of the full-length and truncated *ANKRD26* gene in HSPCs. These data confirm that the truncated *ANKRD26* transcript retains its function, and when overexpressed, driven by the *WAC* gene promoter, this is likely to lead to the increased MAPK activation in megakaryocytes with the resulting thrombocytopenia phenotype seen in THC2 (see figure, panel C).

Clearly, these findings might be informative for other rare families with a THC2-like phenotype but unidentified causative gene; it remains to be seen whether additional families are identified in due course

with similar SVs affecting this locus, or whether this family carries a unique pathogenic SV. In conventional THC2 cases, where SNVs within the regulatory region of the *ANKRD26* gene are the primary cause of disease, these point mutations correlate with a predisposition to myeloid malignancies with incomplete penetrance (Bluteau et al., 2014; Noris et al., 2011). Whether the overexpression of a truncated form of *ANKRD26* will lead to a similar susceptibility to hematologic malignancy remains unknown. It is noteworthy that the truncated exon 11+ *ANKRD26* was more weakly activating than full-length protein, suggesting that other regions of the protein encoded by exons 1–10 might also have a functional role. It is also intriguing that the thrombocytopenia was highly heterogeneous between affected family members, with a suggestion of an age-dependent variation, which is of interest in relation to age-associated alterations in hematopoiesis and platelet production (Grover et al., 2016).

The authors also make a strong case that this novel SV is significant as an exemplar more broadly of how SVs can be missed by conventional sequencing strategies (Sedlazeck et al., 2018). Fully phased long-read sequence data increase the yield of SV detection by as much as fivefold in comparison with short-read WGS data (Huddleston et al., 2017). Although SNVs are the most common type of disease-causing variant in the human genome, it is now apparent that an increasing number of diseases are caused by genetic variants much larger than single base-

pair substitutions or small (<50 bp) indels (Eichler, 2019). SVs result in significant rearrangement of the genome and have numerous implications, including alteration of gene expression through amplification, deletion, or disruption of noncoding genome regulatory elements (Chiang et al., 2017), as previously reported in familial thrombocythemia (Saliba et al., 2015). The generation of fusion transcripts is a well-known cause of disrupted gene expression in human cancer (Li et al., 2020) and may emerge over the coming years as an under-recognized cause of rare inherited conditions. In persons or families with a rare disease of unknown cause, where short-read WGS and other conventional approaches have failed to identify a causative variant, long-read technologies look set to have an important role in the coming years. Defining the optimal role for these technologies in diagnostic pipelines and appropriate approaches for the analysis and integration of different data modalities remains a challenge. However, the development and dissemination of standardized methods and comprehensive reference databases will be an important step (Amarasinghe et al., 2020; Huddleston et al., 2017; Sedlazeck et al., 2018). Moreover, beyond identification of SVs, the study by Wahlster et al. (2021) highlights the need for complementary functional assays to pin down the mechanism by which SVs cause a phenotype. While such extensive experimentation cannot form part of routine clinical diagnostics, for novel SVs, such studies are essential before assigning causality.

In summary, the study by Wahlster et al. (2021) not only provides new insights into *ANKRD26* and its role in familial thrombocytopenia but also highlights the importance of considering SVs, including gene fusions, as an alternative mechanism driving pathogenesis in rare congenital disorders.

Acknowledgments

A.J. Mead is supported by a Cancer Research UK Senior Cancer Research Fellowship (C42639/A26988).

References

- Amarasinghe, S.L., et al. 2020. *Genome Biol.* <https://doi.org/10.1186/s13059-020-1935-5>
- Bluteau, D., et al. 2014. *J. Clin. Invest.* <https://doi.org/10.1172/JCI71861>
- Chiang, C., et al. 2017. *Nat. Genet.* <https://doi.org/10.1038/ng.3834>
- Eichler, E.E. 2019. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMra1809315>
- Grover, A., et al. 2016. *Nat. Commun.* <https://doi.org/10.1038/ncomms11075>
- Huddleston, J., et al. 2017. *Genome Res.* <https://doi.org/10.1101/gr.214007.116>
- Kewan, T., et al. 2020. *J. Investig. Med. High Impact Case Rep.* <https://doi.org/10.1177/2324709620938941>
- Li, Y., et al. 2020. *Nature.* <https://doi.org/10.1038/s41586-019-1913-9>
- Noris, P., et al. 2011. *Blood.* <https://doi.org/10.1182/blood-2011-02-336537>
- Pippucci, T., et al. 2011. *Am. J. Hum. Genet.* <https://doi.org/10.1016/j.ajhg.2010.12.006>
- Saliba, J., et al. 2015. *Nat. Genet.* <https://doi.org/10.1038/ng.3380>
- Sedlazeck, F.J., et al. 2018. *Nat. Methods.* <https://doi.org/10.1038/s41592-018-0001-7>
- Turro, E., et al. 2020. *Nature.* <https://doi.org/10.1038/s41586-020-2434-2>
- Wahlster, L., et al. 2021. *J. Exp. Med.* <https://doi.org/10.1084/jem.20210444>