ESHG

COMMENT

The absence of *CFHR3* and *CFHR1* genes from the T2T-CHM13 assembly can limit the molecular diagnosis of complement-related diseases

Abderaouf Hamza[1], Carine El-Sissy[2], Nadhir Yousfi[3], Paula Vieira Martins[2], Cédric Rafat[4,5], Julien Masliah-Planchon[1], Véronique Frémeaux-Bacchi[2,6] and Laurent Mesnard[3,4,5,7]✉

© The Author(s), under exclusive licence to European Society of Human Genetics 2023

European Journal of Human Genetics (2023) 31:730–732; https://doi.org/10.1038/s41431-023-01350-8

The Telomere-to-Telomere consortium (T2T) work, published in *Science* on April 1, 2022, offers the first complete sequence of a human genome [1]. Using the T2T-CHM13 assembly as a reference genome should improve genetic variation analysis [2] and thus help in the molecular diagnosis of genetic diseases. Compared to GRCh38, it adds 200 million base pairs of sequences containing 1956 gene predictions but removes sequences containing 263 genes. Among those are *CFHR3* and *CFHR1*, which, along with the *CFH*, *CFHR2*, *CFHR4*, and *CFHR5* genes, constitute the CFHR-Factor H gene cluster in 1q32 (Fig. 1A).

The CFHR-Factor H gene cluster is medically relevant. It is associated with complement disorders like atypical hemolytic and uremic syndrome (aHUS) [3, 4]. Structural variants (SV) genes leading to the formation of gene fusions between *CFH* and its paralogs could be the cause of up to 5% of aHUS cases. In one such genomic rearrangement (*CFH*::*CFHR1*) the C-terminal SCR region of the CFH protein is replaced by the C-terminal portion of CFHR1 [5–7], leading to the partial loss of SCR function [8]. C3 glomerulopathy, another human kidney disease that shares complement over-activation as a common pathophysiological pathway with aHUS, has also been connected to sequence variations of the CFHR-Factor H gene cluster [9, 10].

The molecular diagnosis of these diseases is complicated by the highly repetitive nature of the CFHR-Factor H gene cluster locus sequence. On the one hand, short-read NGS can detect various types of rearrangement, but their detection can be impaired by read mapping issues caused by the numerous segmental duplications at this locus; on the other hand, multiplex ligation-dependent probe amplification (MLPA) can only detect copy number variants [4]. The performance of SV detection with long-read sequencing methods and their increasing accessibility make them a promising alternative. As such, long-read sequencing paired with an improved reference could revolutionize the diagnosis of aHUS and other CFH-related diseases, but the impact of the absence of relevant genes from T2T-CHM13 needs to be evaluated.

To assess how using T2T-CHM13 as a reference genome could affect the study of the CFHR-Factor H gene cluster (and ultimately the diagnosis of aHUS), we used targeted nanopore sequencing with adaptive sampling [11] on DNA samples obtained from three normal controls and a patient with aHUS. In this patient, MLPA and functional analysis demonstrated the causal variant to be a paradigmatic *CFH*::*CFHR1* recombination event but the exact breakpoints of the SV were still to be determined [3]. To resolve the exact breakpoint coordinates, we targeted the CFHR-Factor H gene cluster region with adaptive sampling using the GRCh38.p14 assembly (chr1:196598161-197040250). Reads were then mapped against both GRCh38.p14 (Fig. 1B) and T2T-CHM13 v.2 assemblies (Fig. 2).

Aberrant alignment patterns, indicating major discrepancies between the samples' genomes and the T2T genome, were observed in the aHUS sample and in controls 1 and 2 but not in control 3 (Fig. 2). Owing to the repetitive nature of the region and despite the long read lengths obtained (N50 of 18,423, 19,803, 25,354, and 21,423 bp for the aSHU patients and controls 1, 2, and 3, respectively) those aberrant patterns could not be resolved into interpretable SVs, either by SV detection tools or manual review. Importantly, evidence of the presence of the *CFH*::*CFHR1* hybrid gene visible on the GRCh38 alignment (Fig. 1B) was lost when using T2T-CHM13 as a reference genome (Fig. 2). This demonstrates that using the T2T as a reference genome can obscure clinically relevant variations in the CFHR-Factor H gene cluster, with major implications for the diagnosis of aHUS and other CFH-related diseases.

CFHR-Factor H gene cluster is highly polymorphic [12]. In particular, *CFHR3-CFHR1* deletions are enriched in individuals of African and East Asian descent (0.36 and 0.47 allele frequency, respectively [12]) and an estimated 5% of the general population presents a homozygous deletion [13]. The CHM13hTERT cell line (as well as control 3 described here) could be a carrier of this minor allele, explaining the absence of *CFHR3* and *CFHR1* in the T2T-CHM13 assembly. As no one individual can serve as a

[1]Department of Genetics, Institut Curie, PSL Research University, Paris, France. [2]Department of Biological Immunology, Hôpital Européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France. [3]Unité Mixte de Recherche S1155, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France. [4]Service de Soins Intensifs Néphrologiques et Rein Aigu (SINRA), French Intensive Renal Network, Hôpital Tenon, Assistance Publique-Hôpitaux de Paris, Paris, France. [5]Faculté de Médecine, Sorbonne Université, Paris, France. [6]Unité Mixte de Recherche S1138, Institut National de la Santé et de la Recherche Médicale (INSERM), Centre de Recherche des Cordeliers, Paris, France. [7]Institut des Sciences du Calcul et des Données, Sorbonne Université, Paris, France. ✉email: laurent.mesnard@aphp.fr

Received: 27 September 2022 Revised: 14 March 2023 Accepted: 20 March 2023
Published online: 10 April 2023

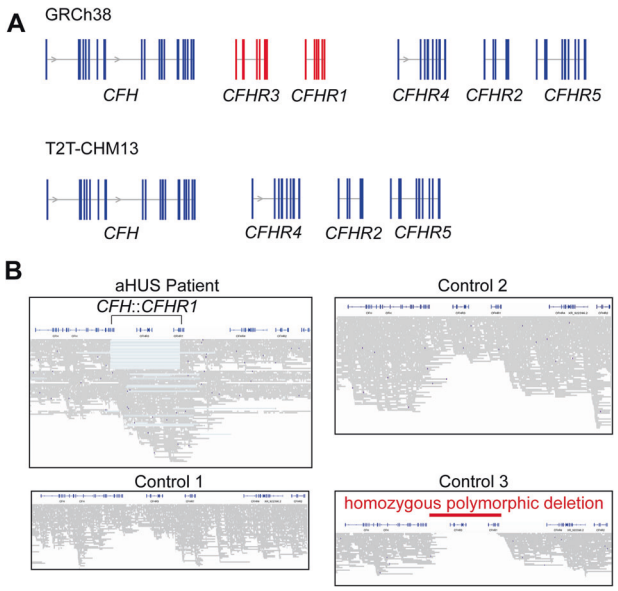

Fig. 1 Long-read sequencing and mapping against the GRCh38 reference genome reveal a structural variant in the CFHR-Factor H cluster region, allowing the precise molecular diagnosis of aHUS. **A** MANE-Select transcripts [15] in the CFHR-Factor H gene cluster region in GRCh38 (top) and T2T-CMH13 (bottom). The CFHR3 and CFHR1 genes (red) are absent from T2T-CMH13. Blue and red boxes represent exons and gray lines introns. **B** Using GRCh38 as a reference genome, split reads mapping both to the CFH and CFHR1 loci are indicative of the presence of a structural variant responsible for the CFH::CFHR1 gene fusion, enabling the molecular diagnosis of aHUS. Control 1 and control 2 show no such alignment pattern; the absence of reads aligning on the CFHR3 and CFHR1 in control 3 suggests that this individual is a carrier of a polymorphic CFHR3-CFHR deletion [12]. Note that all three alignment patterns are easily interpretable; the aHUS patient data have been aligned with specific parameters to better visualize SV breakpoints with split reads. On this Integrative Genomics Viewer (IGV) visualization, blue lines and boxes represent gene introns and exons, respectively, and gray rectangles represent aligned reads. Thin horizontal light blue lines connecting two portions of aligned reads indicate discontinuous alignment ("split reads").

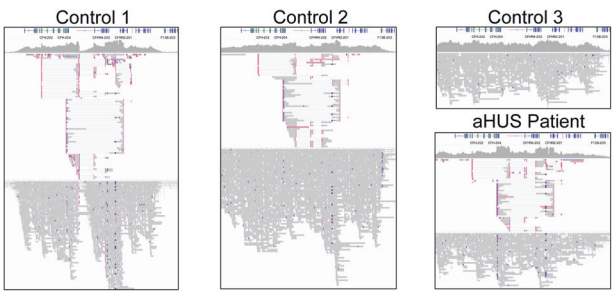

Fig. 2 The T2T-CMH13 reference genome generates aberrant patterns of alignment in the CFHR-Factor H cluster region, hindering the molecular diagnosis of aHUS. Alignment of control 1, control 2, and the aHUS patient's long-read sequencing data shows split reads mapped to the T2T-CMH13 CFHR-Factor H region that cannot be resolved into structural variants, obscuring any relevant variation, and making the region extremely challenging to analyze. In particular, the CFH::CFHR1 gene fusion found in the aHUS patient's sequencing data aligned against GRCh38 (Fig. 1) is not discernible anymore. Alignment data of control 3 show no such aberrant pattern, suggesting that the subject carries the same haplotype as the CHM13hTERT cell line used to generate T2T-CMH13. On this IGV visualization, colored lines and boxes represent gene introns and exons, respectively, and gray rectangles represent aligned reads.

perfect reference for all humankind, this is an intrinsic limit of the current paradigm of genomics analysis (resequencing and mapping against a linear reference sequence) that pangenome-based methods aim to address in the future [14]. In the meantime, scientists and clinicians working on complement-related diseases should be aware that using the T2T-CHM13 assembly as a reference genome can impair the detection of alterations involving CFHR-Factor H cluster genes. More broadly, even though this new assembly is undoubtedly overall a net upgrade from the previous reference genome, one should carefully and systematically evaluate how the differences between reference genomes can impact their specific object of study.

DATA AVAILABILITY

The FASTQ files of reads aligning on the CFH-gene cluster region (GRCh38 coordinates: chr1:196598161-197040250) are available on the Sequence Read Archive under the accession number PRJNA916871. The complete sequencing data files are available from the corresponding author upon reasonable request.

REFERENCES

1. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376:44–53.
2. Aganezov S, Yan SM, Soto DC, Kirsche M, Zarate S, Avdeyev P, et al. A complete reference genome improves analysis of human genetic variation. Science. 2022;376:eabl3533.
3. Heinen S, Sanchez-Corral P, Jackson MS, Strain L, Goodship JA, Kemp EJ, et al. De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. Hum Mutat. 2006;27:292–3.
4. Detection of genetic rearrangements in the regulators of complement activation RCA cluster by high-throughput sequencing and MLPA | SpringerLink [Internet]. [cited 2021 Oct 23]. Available from: https://link.springer.com/protocol/10.1007/978-1-0716-1016-9_16.
5. Abarrategui-Garrido C, Martínez-Barricarte R, López-Trascasa M, de Córdoba SR, Sánchez-Corral P. Characterization of complement factor H-related (CFHR) proteins in plasma reveals novel genetic variations of CFHR1 associated with atypical hemolytic uremic syndrome. Blood. 2009;114:4261–71.
6. Francis NJ, McNicholas B, Awan A, Waldron M, Reddan D, Sadlier D, et al. A novel hybrid CFH/CFHR3 gene generated by a microhomology-mediated deletion in familial atypical hemolytic uremic syndrome. Blood. 2012;119:591–601.
7. Zipfel PF, Edey M, Heinen S, Józsi M, Richter H, Misselwitz J, et al. Deletion of complement factor H-related genes CFHR1 and CFHR3 is associated with atypical hemolytic uremic syndrome. PLoS Genet. 2007;3:e41.
8. Schwarz C, Brehon A, Mousseaux C, Luque Y, Senet P, Mariani P, et al. Ockham's razor defeated: about two atypical cases of hemolytic uremic syndrome. BMC Nephrol. 2020;21:269.
9. Smith RJH, Appel GB, Blom AM, Cook HT, D'Agati VD, Fakhouri F, et al. C3 glomerulopathy—understanding a rare complement-driven renal disease. Nat Rev Nephrol. 2019;15:129–43.
10. Fremeaux-Bacchi V, Fakhouri F, Garnier A, Bienaimé F, Dragon-Durey MA, Ngo S, et al. Genetics and outcome of atypical hemolytic uremic syndrome: a nationwide French series comparing children and adults. Clin J Am Soc Nephrol. 2013;8:554–62.
11. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nat Biotechnol. 2021;39:442–50.
12. Cantsilieris S, Nelson BJ, Huddleston J, Baker C, Harshman L, Penewit K, et al. Recurrent structural variation, clustered sites of selection, and disease risk for the complement factor H (*CFH*) gene family. Proc Natl Acad Sci USA 2018;115:E4433–42. Available from: https://pnas.org/doi/full/10.1073/pnas.1717600115.
13. Moore I, Strain L, Pappworth I, Kavanagh D, Barlow PN, Herbert AP, et al. Association of factor H autoantibodies with deletions of CFHR1, CFHR3, CFHR4, and with mutations in CFH, CFI, CD46, and C3 in patients with atypical hemolytic uremic syndrome. Blood. 2010;115:379–87.
14. Wang T, Antonacci-Fulton L, Howe K, Lawson HA, Lucas JK, Phillippy AM, et al. The Human Pangenome Project: a global resource to map genomic diversity. Nature. 2022;604:437–46.
15. Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604:310–5.

AUTHOR CONTRIBUTIONS

Conceptualization: LM. Methodology: AH, LM. Investigation: AH. Visualization: AH. Funding acquisition: not applicable. Project administration: not applicable. Supervision: LM, VF-B. Writing—original draft: AH, LM. Writing—review and editing: AH, LM, CE-S, PVM, CR, JM-P, VF-B.

FUNDING

No specific funding was received for the study.

COMPETING INTERESTS

VF-B manages a genetic testing facility that uses MLPA in the diagnosis of aHUS. The other authors declare no competing interests.

ETHICS APPROVAL

The study was performed in accordance with the ethical standards of the Declaration of Helsinki. Written informed consent was obtained regarding de-identified clinical and personal patient data collection, analysis and publication. The study has been approved by an Institutional Review Board (Direction de la Recherche Clinique et de l'Innovation (APHP220461)) and the Ethic board of Sorbonne Université (CER-2022-009).

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Laurent Mesnard.

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.