NGS mismapping confounds~~Resolving~~ the clinical~~conflicting~~ interpretation~~s~~ of the *PRSS1* p.Ala16Val (c.47C>T) variant in chronic pancreatitis

Emmanuelle Génin,[1,2] David N. Cooper,[3] Emmanuelle Masson,[1,2] Claude Férec,[1,2] Jian-Min Chen,[1]

[1]Inserm, Univ Brest, EFS, UMR 1078, GGB, F-29200 Brest, France

[2]Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, F-29200 Brest, France

[3]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom

Word count (main text plus Figure and Table legends): 600

Abbreviations: CP, chronic pancreatitis; NGS, next generation sequencing

Keywords: chronic pancreatitis; next generation sequencing; gnomAD; paralog; *PRSS1* p.Ala16Val variant

Correspondence to Dr Jian-Min Chen, INSERM UMR1078 – EFS – UBO, 22 avenue Camille Desmoulins, 29238 BREST, France; Tel: +33-2-98018174; Fax: +33-2-98016474; e-mail: jian-min.chen@univ-brest.fr

We read with interest the recent publication by Weiss and colleagues, which addressed the pitfalls of using next-generation sequencing (NGS) to diagnose *PRSS1* variants in chronic pancreatitis (CP).[1] Specifically, having failed to authenticate an NGS-identified "*PRSS1* c.47C>T (p.Ala16Val) variant" by Sanger sequencing the *PRSS1* gene in the supposed carrier, they postulated that this artefact could have arisen from sequence reads emanating from one of *PRSS1*' highly homologous and closely linked (7q34) pseudogenes, *PRSS3P2*. Herein, we address another NGS-related pitfall that contributes to confusion in relation to the clinical interpretation of *PRSS1* p.Ala16Val.

p.Ala16Val is the third most commonly detected rare *PRSS1* variant in CP; its putative pathological involvement is supported by its ability to increase trypsinogen autoactivation.[2] ClinVar, however, ascribes to it conflicting interpretations of pathogenicity (i.e., likely benign (1); pathogenic (3) and uncertain significance (2)).[3] The main reason for this appears to be its relatively high allele frequency (i.e., 0.006607) in all gnomAD v2.1.1 populations.[4] Since no p.Ala16Val homozygotes were present in gnomAD, its carrier frequency would be 0.0132, which would be ~30 times the estimated prevalence of CP (i.e., 30-50/100,000).

To resolve this conundrum, we first surveyed the p.Ala16Val variant data in gnomAD v2.1.1. In all four examples of p.Ala16Val heterozygotes whose BAM files were available, a mismapping artefact mimicking gene conversion[5] can be assumed, bearing in mind the following four considerations. First, the p.Ala16Val variant always occurs in association with another three variants in close proximity (i.e., the G-T-T-T track in Figure 1A). Second, there is a "donor" sequence for these *cis*-linked variants in another closely linked *PRSS1* pseudogene, *TRY7* (trypsinogen D; https://www.ncbi.nlm.nih.gov/gene/?term=try7) (Figure 1B). Third, allelic ratios of p.Ala16Val in these examples are all <25%, significantly lower

than the 50% that would be expected for a genuine heterozygous variant. Lastly, despite an apparent allele frequency of 0.006607, no p.Ala16Val homozygotes are present in gnomAD v2.1.1. We also surveyed the p.Ala16Val variant data in the French Exome (FrEx) Project database,[6] obtaining quite similar findings: all "p.Ala16Val" carriers (n = 159; all heterozygotes) among the 525 sequenced French individuals carry the aforementioned G-T-T-T track; additionally, the allelic ratios of p.Ala16Val in these supposed heterozygous carriers are unbalanced with a maximum value of 25% (Supplementary Figure S1).

Finally, we performed a meta-analysis of studies that (i) analyzed p.Ala16Val in both patients and controls and (ii) detected the variant at least once by means of Sanger sequencing or other conventional mutation screening methods using *PRSS1*-specific primers. The meta-analysis comprised a manual survey of all *PRSS1*-related publications (until December 2018) collated in reference 2, complemented by a keyword search (*PRSS1* and "chronic pancreatitis") in PubMed. In cases of overlapping studies from the same laboratory, the latest was used for analysis. Three eligible studies were identified;[7-9] taken together, p.Ala16Val was detected in 18 (1.8%; all heterozygotes) of 983 CP patients but was absent from 2288 controls (odds ratio = infinity; $p < 0.001$) (Table 1).

Table 1. *PRSS1* p.Ala16Val variant in CP patients and controls derived from meta-analysis of eligible genetic studies

Study	Patients +/n	Controls +/n
Chen et al.[7]	2/221	0/400
Rosendahl et al.[8]	14/660	0/1758
Schubert et al.[9]	2/102	0/130
Combined	18/983	0/2288

In summary, we conclude that the *bona fide* *PRSS1* p.Ala16Val variant is (i) extremely rare in the general population and (ii) of genuine pathological significance. This resolves the conflicting interpretations of *PRSS1* p.Ala16Val in CP and emphasizes the need for careful use of gnomAD data in variant assessment.[10]

Contributors EG performed the bioinformatics analysis and revised the paper. DNC, EM and CF contributed to data interpretation and revised the paper. JMC conceived the study, performed the meta-analysis and drafted the paper. All authors approved the final manuscript.

Patients and public involvement Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Funding This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), France.

Competing interests None declared

REFERENCES

1. Weiss FU, Laemmerhirt F, Lerch MM. Next generation sequencing pitfalls in diagnosing trypsinogen (*PRSS1*) mutations in chronic pancreatitis. *Gut* 2020 Sep 28. doi: 10.1136/gutjnl-2020-322864.
2. Genetic Risk Factors in Chronic Pancreatitis. http://www.pancreasgenetics.org/. Accessed 12 April 2021.
3. NM_002769.5(PRSS1):c.47C>T (p.Ala16Val). https://www.ncbi.nlm.nih.gov/clinvar/variation/38363/. Accessed 12 April 2021.

4. gnomAD. https://gnomad.broadinstitute.org/. Accessed 12 April 2012.
5. Chen JM, Férec C. Gene conversion-like missense mutations in the human cationic trypsinogen gene and insights into the molecular evolution of the human trypsinogen family. *Mol Genet Metab* 2000;71:463-9.
6. The French Exome (FrEx) Project. http://lysine.univ-brest.fr/FrExAC/. Accessed 12 April 2021.
7. Chen JM, Piepoli Bis A, Le Bodic L, *et al*. Mutational screening of the cationic trypsinogen gene in a large cohort of subjects with idiopathic chronic pancreatitis. *Clin Genet* 2001;59:189-93.
8. Rosendahl J, Landt O, Bernadova J, *et al*. *CFTR*, *SPINK1*, *CTRC* and *PRSS1* variants in chronic pancreatitis: is the role of mutated *CFTR* overestimated? *Gut* 2013;62:582-92.
9. Schubert S, Traub F, Brakensiek K, *et al*. *CFTR*, *SPINK1*, *PRSS1*, and *CTRC* mutations are not associated with pancreatic cancer in German patients. *Pancreas* 2014;43:1078-82.
10. Girodon E, Rebours V, Chen JM, *et al*. Clinical interpretation of *PRSS1* variants in patients with pancreatitis. *Clin Res Hepatol Gastroenterol* 2021;45:101497.

FIGURE LEGEND

Figure 1. *PRSS1* p.Ala16Val (c.47C>T) variant as a mismapping artefact mimicking gene conversion in gnomAD v.2.1.1. (**A**) Partial view of genome sequencing data from a "heterozygous *PRSS1* p.Ala16Val" carrier in gnomAD v2.1.1. (**B**) Sequence alignment of *PRSS1* and two of its pseudogenes, *PRSS3P2* and *TRY7*. Note the presence of a "donor" sequence for the "p.Ala16Val variant and its three *cis*-linked variants" in *TRY7*.