

Opinion

Chronic Pancreatitis: The True Pathogenic Culprit within the *SPINK1* N34S-Containing Haplotype Is No Longer at Large

Na Pu^{1,2}, Emmanuelle Masson^{1,3}, David N. Cooper⁴ , Emmanuelle Génin¹ , Claude Férec^{1,3} and Jian-Min Chen^{1,*} 

¹ Inserm, Univ Brest, EFS, UMR 1078, GGB, F-29200 Brest, France; punayeah@163.com (N.P.); emmanuelle.masson@univ-brest.fr (E.M.); emmanuelle.genin@inserm.fr (E.G.); claude.ferec@univ-brest.fr (C.F.)

² Department of Critical Care Medicine, Research Institute of General Surgery, Jinling Hospital, Medical School of Nanjing University, Nanjing 210016, China

³ Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, F-29200 Brest, France

⁴ Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK; CooperDN@cardiff.ac.uk

* Correspondence: Jian-Min.Chen@univ-brest.fr

Abstract: A diverse range of loss-of-function variants in the *SPINK1* gene (encoding pancreatic secretory trypsin inhibitor) has been identified in patients with chronic pancreatitis (CP). The haplotype harboring the *SPINK1* c.101A>G (p.Asn34Ser or N34S) variant (rs17107315:T>C) is one of the most important heritable risk factors for CP as a consequence of its relatively high prevalence worldwide (population allele frequency \approx 1%) and its considerable effect size (odds ratio \approx 11). The causal variant responsible for this haplotype has been intensively investigated over the past two decades. The different hypotheses tested addressed whether the N34S missense variant has a direct impact on enzyme structure and function, whether c.101A>G could affect pre-mRNA splicing or mRNA stability, and whether another variant in linkage disequilibrium with c.101A>G might be responsible for the observed association with CP. Having reviewed the currently available genetic and experimental data, we conclude that c.-4141G>T (rs142703147:C>A), which disrupts a PTF1L-binding site within an evolutionarily conserved HNF1A-PTF1L *cis*-regulatory module located \sim 4 kb upstream of the *SPINK1* promoter, can be designated as the causal variant beyond reasonable doubt. This case illustrates the difficulties inherent in determining the identity of the causal variant underlying an initially identified disease association.

Keywords: chronic pancreatitis; enhancer; linkage disequilibrium; regulatory variant; *SPINK1* gene



Citation: Pu, N.; Masson, E.; Cooper, D.N.; Génin, E.; Férec, C.; Chen, J.-M. Chronic Pancreatitis: The True Pathogenic Culprit within the *SPINK1* N34S-Containing Haplotype Is No Longer at Large. *Genes* **2021**, *12*, 1683. <https://doi.org/10.3390/genes12111683>

Academic Editor: Rodolfo Iuliano

Received: 11 October 2021

Accepted: 23 October 2021

Published: 23 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Chronic pancreatitis (CP) is a complex disease that can be caused by genetic and/or environmental factors ([1] and references therein). Genetic studies over the past 25 years have served to identify a trypsin-dependent pathway that is central to our understanding of the etiology of CP [2]. The recognition of this pathway emerged from the identification and characterization of gain-of-function variants in *PRSS1* (encoding cationic trypsinogen; MIM# 276000) and loss-of-function variants in *SPINK1* (encoding pancreatic secretory trypsin inhibitor; MIM# 167790) and *CTRC* (encoding chymotrypsin C (MIM# 601405), which specifically degrades all human trypsinogen/trypsin isoforms) in patients with the disease ([3–5] and references therein) (Figure 1).

A diverse range of variants in the *SPINK1* gene has been reported in the literature [3,4]. Whilst the pathogenic relevance of large deletions, splice-site variants and nonsense variants is usually self-evident, that of other types of lesion, particularly missense variants, often has to be determined by in-depth functional analysis. In this regard, a missense variant in the *SPINK1* gene, p.Asn34Ser or simply N34S (c.101A>G; rs17107315:T>C) [6], represents one of the most important CP-associated heritable risk factors owing to its relatively high prevalence (an allele frequency of 0.009028 according to gnomAD v2.1.1 [7])

and its considerable effect size (odds ratio (OR) = 10.90; 95% confidence interval 7.56–15.72) in accordance with a recent meta-analysis [8]) worldwide. However, whether or not the N34S missense substitution per se represents the underlying pathogenic variant that predisposes to CP has been the subject of continued research interest over the past two decades. The clarification of this issue has not only considerable biological interest but also potential diagnostic and therapeutic value.

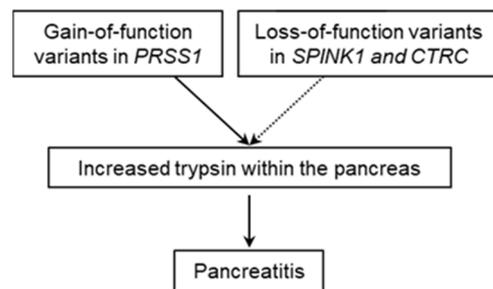


Figure 1. Schematic illustration of the trypsin-dependent pathway for chronic pancreatitis stemming from the identification of genetic variants in the *PRSS1*, *SPINK1*, and *CTRC* genes.

Witt and colleagues initially postulated that N34S, which is located near the reactive site of *SPINK1* (Lys41-Ile42), might impair the enzyme's inhibitory activity on prematurely activated trypsin within the pancreas, thereby conferring predisposition to CP [6]. However, functional analyses of the wild-type and N34S mutant *SPINK1* enzymes expressed in *Saccharomyces cerevisiae* [9], Chinese hamster ovary cells [10], and human embryonic kidney 293T (HEK293T) cells [11] failed to identify any measurable effect of the N34S variant on the expression, secretion or trypsin inhibitory activity of *SPINK1*. An alternative hypothesis that either c.101A>G or one of its four *cis*-linked intronic variants may affect pre-mRNA splicing or mRNA stability [12] also failed to garner support from experiments employing cell culture-based minigene [13] and full-length gene [14,15] splicing assays. Herein, it should be emphasized that in our full-length gene splicing assay [14,15], it is the *SPINK1* genomic sequence extending from the translation initiation codon to the translation termination codon that was inserted into the expression vector. In other words, a possible effect of c.101A>G and its four *cis*-linked intronic variants on pre-mRNA splicing and/or mRNA stability was simultaneously analyzed in the gene's natural genomic sequence context as far as the entire coding and intronic sequences are concerned. Employing both qualitative and quantitative reverse transcription-PCR (RT-PCR) analyses, we did not observe any measurable effect of the c.101A>G or its four *cis*-linked intronic variants on pre-mRNA splicing and/or mRNA stability [14,15]. Moreover, RT-PCR analysis of total RNA prepared from pancreatic tissue of two c.101A>G (N34S) homozygotes yielded only wild-type transcripts [16]. However, expression levels of *SPINK1* mRNA were not subjected to quantitative analysis in this latter study. Additionally, no significant single-tissue expression quantitative trait loci (eQTLs) for *SPINK1* in human pancreatic tissue are listed in the Genotype-Tissue Expression (GTEx) database [17].

These negative findings stimulated us to embark upon a new hypothesis-driven approach that sought a causal regulatory variant in linkage disequilibrium (LD) with c.101A>G (N34S). Employing an LD threshold of $r^2 \geq 0.40$ and querying the 1000 Genomes Project Phase 1 data in the context of the European population by means of HaploReg v4.1, we identified a total of 25 single nucleotide polymorphisms in strong LD (r^2 values ranging from 0.87 to 1) with c.101A>G in the region spanning 20 kb 3' of *SPINK1* to 18 kb 5' of *SPINK1* (Figure 2).

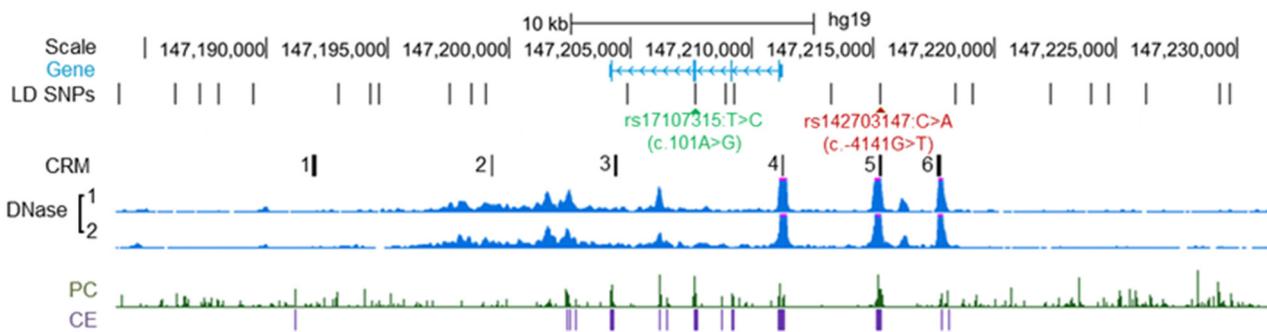


Figure 2. Illustration of the *SPINK1* locus, the locations of the *SPINK1* N34S (c.101A>G; rs17107315:T>C) and c.-4141G>T (rs142703147C>A) variants, and the spatial coincidence of c.-4141G>T with regulatory features. In the “Gene” panel, the four exons of the *SPINK1* gene are denoted by vertical lines, with the arrows indicating the direction of transcription. LD SNPs, single nucleotide polymorphisms in strong linkage disequilibrium with *SPINK1* c.101A>G. It should be noted that only 24 of the 25 LD SNPs are shown in the Figure, the variant that was not shown, rs138251740A>G (located further downstream of chr5:147,230,000), is neither located within a putative PTF1L–HNF1A CRM nor within a region showing strong evolutionary conservation and high chromatin accessibility. CRM, *cis*-regulatory modules harboring binding sites for transcription factors HNF1A and PTF1L. DNase, DNase I-accessible DNA regions in the pancreatic tissues of two donors as determined by DNase-seq. PC and CE, Placental Mammal Conservation by PhastCons (PC) and Placental Mammal Conserved Elements (CE) were obtained from the UCSC Genome Browser. The Figure was adapted from our work [18] with permission (Copyright 2021 Wiley Periodicals LLC).

Of the 25 LD variants only one, rs142703147:C>A (c.-4141G>T relative to the A of the translation initiation codon of *SPINK1* demarcated as +1 [19]), was found to be both located within an evolutionarily conserved region and one of the three most accessible chromatin regions in pancreatic tissue. The three most accessible chromatin regions each harbor a putative HNF1A–PTF1L *cis*-regulatory module (CRM) (a CRM, usually 100–1000 base-pairs in length, contains several transcription factor binding sites, and provides the structural basis for coordinating the action of the corresponding transcription factors required for the temporal and spatial expression of neighboring genes [20]). Both HNF1A and PTF1L are basic components of the exocrine pancreas-specific transcriptional network for digestive function and pancreatic acinar cell homeostasis [21–23]. Importantly, c.-4141G>T occurs within one of these CRMs and is predicted to disrupt the corresponding PTF1L-binding site. Co-transfection transactivation experiments [18] have demonstrated that this variant leads to reduced gene expression. Contemporaneously, Kereszturi and Sahin-Tóth reported that two pancreatic cancer cell lines heterozygous for the *SPINK1* N34S haplotype exhibited reduced expression of the variant allele and suggested that c.-4141G>T might be a candidate causal variant [24].

Taken together, the findings from the aforementioned studies, and particularly the two from 2017 [18,24], suggested that c.-4141G>T was likely to be the causal variant underlying the association of the *SPINK1* N34S haplotype with CP. This notwithstanding, three recent studies have claimed that the pathogenic mechanism underlying the *SPINK1* N34S haplotype is still unknown and have therefore continued to search for a possible direct effect of the N34S missense variant on *SPINK1* structure and function. Two of these studies involved purely hypothetical simulation or modeling [25,26]. The other explored the possible impact of physiological stress factors (e.g., altered ion concentrations, temperature shifts and environmental pH) on the structure and trypsin inhibitory function of *SPINK1* (and its N34S variant) using biophysical methods, but did not obtain any significant findings [27]. Herein, it is pertinent to mention a recent study from the Sahin-Tóth laboratory [28] that pointed out two serious shortcomings of the two early N34S-*SPINK1* binding studies [9,11]. First, the methodology used was semiquantitative at best. Second, the use of either nonsulfated recombinant human cationic trypsin, bovine trypsin or a commercial human trypsin of unspecified nature in the binding assays could have impacted the functional relevance of the binding assays given that human trypsins are

invariably sulfated on Tyr154 [29,30]. Using authentic sulfated human trypsins and more robust experimental conditions, Sahin-Tóth and colleagues provided conclusive evidence that N34S has no impact on trypsin inhibition [28].

Finally, it should be noted that the c.-4141G>T functional enhancer variant is in perfect LD with c.101A>G (N34S) in the French population ($r^2 = 1$; 548 chronic patients and 562 controls genotyped) as well as in the European population of the 1000 Genomes Project [31] but not in the Chinese population ($r^2 = 0.80$; 1104 patients and 1196 controls genotyped) or the Indian population ($r^2 = 0.59$; 347 patients and 264 controls genotyped) [18]. Conditional analyses performed at the time suggested that both variants influenced disease risk [18]. With hindsight, this latter observation should be interpreted with a considerable degree of caution. Our main concern lies with the data obtained from the Indian population for which the lowest r^2 was observed. Only in the Indian population was c.101A>G (N34S) found to have a higher odds ratio (OR) for CP than c.-4141G>T (15.12 versus 14.82). By contrast, in the Chinese population, c.-4141G>T had a higher OR than c.101A>G (6.13 versus 5.47) whilst the two variants had an equal OR in the French population where they are in complete LD. It should also be pointed out that (i) of the three populations studied, the Indian population had the smallest sample size [18] and (ii) the Indian population is well known to have a complex population substructure [32]. The constellation of these concerns suggests that the counterintuitive result obtained from the Indian population probably arose as a consequence of sampling error and/or population stratification. However, an alternative possibility could conceivably be the presence of an additional CP-predisposing variant in the Indian population that is in stronger LD with c.101A>G than with c.-4141G>T.

In summary, based upon the currently available genetic and experimental data, we conclude that c.-4141G>T represents the true culprit underlying the association of the *SPINK1* N34S haplotype with CP. In other words, the true culprit underlying the *SPINK1* N34S haplotype-disease association should no longer be thought to be at large [33]. This notwithstanding, one may still argue that further experiments are needed to unambiguously confirm causality. For example, discarded biopsy or autopsy tissue from N34S heterozygotes could, in principle, be analyzed with a view to showing that the N34S-containing allele is indeed associated with reduced *SPINK1* mRNA expression in vivo. Alternatively, given the availability of 2 cell lines heterozygous for both the c.101A>G (N34S) and c.-4141G>T variants, CRISPR/cas9 technology could be employed to generate isogenic cell lines carrying only the c.101A>G (N34S) variant or the c.-4141G>T variant. Subsequent comparison of *SPINK1* mRNA expression levels in these cell lines could then serve to strengthen the evidence for causality. Indeed, it is always better to provide additional experimental evidence if resources are available. However, in this particular case, the evidence supporting our contention that the c.-4141G>T variant is the true pathogenic variant is, we believe, beyond reasonable doubt.

The conclusion that c.-4141G>T is the true pathogenic culprit within the *SPINK1* N34S-containing haplotype has one immediate clinical implication. Genetic correction of the c.-4141G>T variant in the corresponding carriers could be explored as a patient-tailored therapy.

Author Contributions: N.P. and J.-M.C. drafted the manuscript. E.M., D.N.C., E.G. and C.F. revised the manuscript with important intellectual input. All authors have read and agreed to the published version of the manuscript.

Funding: This work received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: N.P. received a one year visiting student scholarship from the China Scholarship Council, the Ministry of Education of the People's Republic of China (no. 202006190267).

Conflicts of Interest: The authors are unaware of any conflict of interest.

References

1. Ru, N.; Xu, X.N.; Cao, Y.; Zhu, J.H.; Hu, L.H.; Wu, S.Y.; Qian, Y.Y.; Pan, J.; Zou, W.B.; Li, Z.S.; et al. The impacts of genetic and environmental factors on the progression of chronic pancreatitis. *Clin. Gastroenterol. Hepatol.* **2021**. online ahead of print. [[CrossRef](#)] [[PubMed](#)]
2. Hegyi, E.; Sahin-Tóth, M. Genetic risk in chronic pancreatitis: The trypsin-dependent pathway. *Dig. Dis. Sci.* **2017**, *62*, 1692–1701. [[CrossRef](#)]
3. Genetic Risk Factors in Chronic Pancreatitis. Available online: <http://www.pancreasgenetics.org/index.php> (accessed on 27 September 2021).
4. Girodon, E.; Rebours, V.; Chen, J.M.; Pagin, A.; Lévy, P.; Férec, C.; Bienvenu, T. Clinical interpretation of *SPINK1* and *CTRC* variants in pancreatitis. *Pancreatology* **2020**, *20*, 1354–1367. [[CrossRef](#)]
5. Girodon, E.; Rebours, V.; Chen, J.M.; Pagin, A.; Lévy, P.; Férec, C.; Bienvenu, T. Clinical interpretation of *PRSS1* variants in patients with pancreatitis. *Clin. Res. Hepatol. Gastroenterol.* **2021**, *45*, 101497. [[CrossRef](#)] [[PubMed](#)]
6. Witt, H.; Luck, W.; Hennies, H.C.; Classen, M.; Kage, A.; Lass, U.; Landt, O.; Becker, M. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat. Genet.* **2000**, *25*, 213–216. [[CrossRef](#)]
7. GnomAD (Genome Aggregation Database). Available online: <https://gnomad.broadinstitute.org/> (accessed on 27 September 2021).
8. Chen, J.M.; Herzig, A.F.; Genin, E.; Masson, E.; Cooper, D.N.; Férec, C. Scale and scope of gene-alcohol interactions in chronic pancreatitis: A systematic review. *Genes* **2021**, *12*, 471. [[CrossRef](#)]
9. Kuwata, K.; Hirota, M.; Shimizu, H.; Nakae, M.; Nishihara, S.; Takimoto, A.; Mitsushima, K.; Kikuchi, N.; Endo, K.; Inoue, M.; et al. Functional analysis of recombinant pancreatic secretory trypsin inhibitor protein with amino-acid substitution. *J. Gastroenterol.* **2002**, *37*, 928–934. [[CrossRef](#)]
10. Boulling, A.; le Maréchal, C.; Trouvé, P.; Raguénès, O.; Chen, J.M.; Férec, C. Functional analysis of pancreatitis-associated missense mutations in the pancreatic secretory trypsin inhibitor (*SPINK1*) gene. *Eur. J. Hum. Genet.* **2007**, *15*, 936–942. [[CrossRef](#)] [[PubMed](#)]
11. Király, O.; Wartmann, T.; Sahin-Tóth, M. Missense mutations in pancreatic secretory trypsin inhibitor (*SPINK1*) cause intracellular retention and degradation. *Gut* **2007**, *56*, 1433–1438. [[CrossRef](#)]
12. Chen, J.M.; Mercier, B.; Audrézet, M.P.; Raguénès, O.; Quéré, I.; Férec, C. Mutations of the pancreatic secretory trypsin inhibitor (*PSTI*) gene in idiopathic chronic pancreatitis. *Gastroenterology* **2001**, *120*, 1061–1064. [[CrossRef](#)]
13. Kereszturi, E.; Király, O.; Sahin-Tóth, M. Minigene analysis of intronic variants in common *SPINK1* haplotypes associated with chronic pancreatitis. *Gut* **2009**, *58*, 545–549. [[CrossRef](#)]
14. Boulling, A.; Chen, J.M.; Callebaut, I.; Férec, C. Is the *SPINK1* p.Asn34ser missense mutation *per se* the true culprit within its associated haplotype? *WebmedCentral Gene* **2012**, *3*, WMC003084.
15. Wu, H.; Boulling, A.; Cooper, D.N.; Li, Z.S.; Liao, Z.; Férec, C.; Chen, J.M. Analysis of the impact of known *SPINK1* missense variants on pre-mRNA splicing and/or mRNA stability in a full-length gene assay. *Genes* **2017**, *8*, 263. [[CrossRef](#)] [[PubMed](#)]
16. Masamune, A.; Kume, K.; Takagi, Y.; Kikuta, K.; Satoh, K.; Satoh, A.; Shimosegawa, T. N34s mutation in the *SPINK1* gene is not associated with alternative splicing. *Pancreas* **2007**, *34*, 423–428. [[CrossRef](#)]
17. The Genotype-Tissue Expression (GTEx) Database. Available online: <https://gtexportal.org/home/> (accessed on 7 October 2021).
18. Boulling, A.; Masson, E.; Zou, W.B.; Paliwal, S.; Wu, H.; Issarapu, P.; Bhaskar, S.; Genin, E.; Cooper, D.N.; Li, Z.S.; et al. Identification of a functional enhancer variant within the chronic pancreatitis-associated *SPINK1* c.101A>G (p.Asn34ser)-containing haplotype. *Hum. Mutat.* **2017**, *38*, 1014–1024. [[CrossRef](#)]
19. den Dunnen, J.T.; Dalgleish, R.; Maglott, D.R.; Hart, R.K.; Greenblatt, M.S.; McGowan-Jordan, J.; Roux, A.F.; Smith, T.; Antonarakis, S.E.; Taschner, P.E. HGVS recommendations for the description of sequence variants: 2016 update. *Hum. Mutat.* **2016**, *37*, 564–569. [[CrossRef](#)]
20. Lelli, K.M.; Slattery, M.; Mann, R.S. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.* **2012**, *46*, 43–68. [[CrossRef](#)] [[PubMed](#)]
21. Masui, T.; Swift, G.H.; Hale, M.A.; Meredith, D.M.; Johnson, J.E.; Macdonald, R.J. Transcriptional autoregulation controls pancreatic Ptf1a expression during development and adulthood. *Mol. Cell. Biol.* **2008**, *28*, 5458–5468. [[CrossRef](#)]
22. Holmstrom, S.R.; Deering, T.; Swift, G.H.; Poelwijk, F.J.; Mangelsdorf, D.J.; Kliewer, S.A.; MacDonald, R.J. LRH-1 and PTF1-L coregulate an exocrine pancreas-specific transcriptional network for digestive function. *Genes Dev.* **2011**, *25*, 1674–1679. [[CrossRef](#)]
23. Molero, X.; Vaquero, E.C.; Flandez, M.; Gonzalez, A.M.; Ortiz, M.A.; Cibrian-Uhalte, E.; Servitja, J.M.; Merlos, A.; Juanpere, N.; Massumi, M.; et al. Gene expression dynamics after murine pancreatitis unveils novel roles for Hnf1 α in acinar cell homeostasis. *Gut* **2012**, *61*, 1187–1196. [[CrossRef](#)] [[PubMed](#)]
24. Kereszturi, É.; Sahin-Tóth, M. Pancreatic cancer cell lines heterozygous for the *SPINK1* p.N34s haplotype exhibit diminished expression of the variant allele. *Pancreas* **2017**, *46*, e54–e55. [[CrossRef](#)] [[PubMed](#)]
25. Sun, Z.; Kolossváry, I.; Kozakov, D.; Sahin-Tóth, M.; Vajda, S. The N34S mutation of *SPINK1* may impact the kinetics of trypsinogen activation to cause early trypsin release in the pancreas. *bioRxiv* **2020**. [[CrossRef](#)]
26. Kulke, M.; Nagel, F.; Schulig, L.; Geist, N.; Gabor, M.; Mayerle, J.; Lerch, M.M.; Link, A.; Delcea, M. A hypothesized mechanism for chronic pancreatitis caused by the N34S mutation of serine protease inhibitor Kazal-type 1 based on conformational studies. *J. Inflamm. Res.* **2021**, *14*, 2111–2119. [[CrossRef](#)] [[PubMed](#)]

27. Buchholz, I.; Nagel, F.; Klein, A.; Wagh, P.R.; Mahajan, U.M.; Greinacher, A.; Lerch, M.M.; Mayerle, J.; Delcea, M. The impact of physiological stress conditions on protein structure and trypsin inhibition of serine protease inhibitor Kazal type 1 (SPINK1) and its N34S variant. *Biochim. Biophys. Acta Proteins Proteom.* **2020**, *1868*, 140281. [[CrossRef](#)]
28. Szabó, A.; Toldi, V.; Gazda, L.D.; Demcsak, A.; Tozser, J.; Sahin-Tóth, M. Defective binding of SPINK1 variants is an uncommon mechanism for impaired trypsin inhibition in chronic pancreatitis. *J. Biol. Chem.* **2021**, *296*, 100343. [[CrossRef](#)]
29. Sahin-Tóth, M.; Kukor, Z.; Nemoda, Z. Human cationic trypsinogen is sulfated on Tyr154. *FEBS J.* **2006**, *273*, 5044–5050. [[CrossRef](#)] [[PubMed](#)]
30. Szabó, A.; Salameh, M.A.; Ludwig, M.; Radisky, E.S.; Sahin-Tóth, M. Tyrosine sulfation of human trypsin steers S2' subsite selectivity towards basic amino acids. *PLoS ONE* **2014**, *9*, e102063. [[CrossRef](#)] [[PubMed](#)]
31. LDlink. Available online: <https://ldlink.nci.nih.gov/?tab=home> (accessed on 21 October 2021).
32. Sengupta, D.; Choudhury, A.; Basu, A.; Ramsay, M. Population stratification and underrepresentation of Indian subcontinent genetic diversity in the 1000 genomes project dataset. *Genome Biol. Evol.* **2016**, *8*, 3460–3470. [[CrossRef](#)]
33. Chen, J.M.; Férec, C. The true culprit within the *SPINK1* p.N34S-containing haplotype is still at large. *Gut* **2009**, *58*, 478–480. [[CrossRef](#)]