



Evolutionary history of type II transmembrane serine proteases involved in viral priming

Diego Forni¹ · Manuela Sironi¹ · Rachele Cagliani¹

Received: 11 October 2021 / Accepted: 15 January 2022 / Published online: 5 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Type II transmembrane serine proteases (TTSPs) are a family of trypsin-like membrane-anchored serine proteases that play key roles in the regulation of some crucial processes in physiological conditions, including cardiac function, digestion, cellular iron homeostasis, epidermal differentiation, and immune responses. However, some of them, in particular TTSPs expressed in the human airways, were identified as host factors that promote the proteolytic activation and spread of respiratory viruses such as influenza virus, human metapneumovirus, and coronaviruses, including SARS-CoV-2. Given their involvement in viral priming, we hypothesized that members of the TTSP family may represent targets of positive selection, possibly as the result of virus-driven pressure. Thus, we investigated the evolutionary history of sixteen TTSP genes in mammals. Evolutionary analyses indicate that most of the TTSP genes that have a verified role in viral proteolytic activation present signals of pervasive positive selection, suggesting that viral infections represent a selective pressure driving the evolution of these proteases. We also evaluated genetic diversity in human populations and we identified targets of balancing selection in *TMPRSS2* and *TMPRSS4*. This scenario may be the result of an ancestral and still ongoing host–pathogen arms race. Overall, our results provide evolutionary information about candidate functional sites and polymorphic positions in TTSP genes.

Introduction

In order to initiate a successful infection, viruses must deliver their genome to the host cell cytosol. To achieve this, enveloped viruses trigger a sequence of events that lead to the fusion of their membrane with that of the target host cell, either by direct fusion with the plasma membrane or by an internalization process into endosomes (Barrett and Dutch 2020; Harrison 2008). Usually, this process is mainly mediated by the interaction of specific viral attachment proteins expressed on the surface of the virion (envelope glycoproteins) with viral receptors. These latter are cell surface molecules that perform specific host cellular functions but are hijacked by viruses to assist infection. Multiple types of molecules, including proteins, carbohydrates, and lipids, can be used as viral receptors. However, proteins are considered to be more suitable receptors because of their stronger affinity and higher specificity for viral attachment, resulting in an

increase in the efficiency of viral entry and in an expansion of the viral host range and tropism (Zhang et al. 2019a, b; Maginnis 2018).

Viral receptors favor the recognition and attachment of the virus to the host cell, but the activation of the membrane fusion machinery often requires the involvement of additional host factors. In fact, most viral fusion proteins are synthesized as a precursor and require processing by host cell proteases to trigger the fusion of the viral lipid envelope with the cellular membrane. Several cellular proteases, including proprotein convertases (i.e., furin), trypsin, cathepsins, and type II transmembrane serine proteases (TTSPs), have been described to participate in the proteolytic activation of viral proteins. TTSPs are a family of trypsin-like membrane-anchored serine proteases that play a key role in the regulation of some crucial processes, including cardiac function, digestion, cellular iron homeostasis, epidermal differentiation, and immune responses (Bugge et al. 2009). Consequently, deregulation of these enzymes is an important feature of several pathological conditions (List et al. 2002; Du et al. 2008; Scott et al. 2001), including tumorigenesis (Choi et al. 2009).

✉ Rachele Cagliani
rachele.cagliani@lanostrafamiglia.it

¹ Scientific Institute IRCCS E. MEDEA, Bioinformatics,
23842 Bosisio Parini, Italy

TTSPs expressed in the human airways were identified as host factors that promote the proteolytic activation and spread of respiratory viruses such as influenza virus, human metapneumovirus, paramyxoviruses, human parainfluenza viruses, Sendai virus, orthoreoviruses, Group A rotaviruses and coronaviruses (Shirogane et al. 2008; Bertram et al. 2010, 2013; Shirato et al. 2013; Sakai et al. 2014; Sasaki et al. 2021; Nygaard et al. 2012), including SARS-CoV-2, the causative agent of the current COVID-19 pandemic (Zang et al. 2020; Hoffmann et al. 2020b).

For a successful infection, influenza viruses depend on the proteolytic activation of their entry-mediating surface protein, hemagglutinin (HA). HA may be activated by diverse host trypsin-like proteases *in vitro* or *in vivo*. In particular, influenza viruses (e.g. H1N1 and H7N9 strains) hijack transmembrane serine proteases (including TMPRSS2, TMPRSS4, TMPRSS13, TMPRSS11D, TMPRSS11E, and ST14) (Laporte and Naesens 2017), that cleave HA into subunits HA1 and HA2, a prerequisite for conformational changes that trigger fusion.

In coronaviruses, it is the spike (S) protein that promotes cell entry via membrane fusion (Shang et al. 2020; Li 2016). After receptor engagement, the coronavirus S protein is enzymatically processed by cellular proteases, mediating efficient membrane fusion with host cells (Bertram et al. 2013; Shirato et al. 2013; Hoffmann et al. 2020a, b; Matsuyama et al. 2020; Simmons et al. 2005).

Human coronaviruses can use a variety of cellular proteases to prime their S protein, including TMPRSS2 and other members of the TTSP family (e.g., TMPRSS4 and TMPRSS11) (Zang et al. 2020; Millet and Whittaker 2015). Likewise, coronaviruses can utilize different receptors to promote their entry into the host cells: both SARS-CoV and SARS-CoV-2 use angiotensin-converting enzyme 2 (ACE2), whereas MERS-CoV and HCoV-229E exploit DPP4 (dipeptidyl peptidase 4) and ANPEP (aminopeptidase N or CD13), respectively (Hoffmann et al. 2020b; Raj et al. 2013; Yeager et al. 1992; Forni et al. 2017).

Furthermore, previous studies have shown that SARS-CoV and other coronaviruses interact with multiple cell-surface molecules, such as CD209 (DC-SIGN), CLEC4M (L-SIGN/CD299), CLECG (LSEctin), and basigin (BSG or CD147) (Yang et al. 2004; Marzi et al. 2004; Gramberg et al. 2005; Vankadari and Wilce 2020; Wang et al. 2020a). Interestingly, many of these receptors are targets of positive selection (Cui et al. 2013; Bibiana et al. 2020; Guo et al. 2020; Damas et al. 2020; Ortiz et al. 2008; Enard et al. 2016; Sironi et al. 2015; Wang et al. 2020b).

In line with its role in the activation of viral glycoproteins, TMPRSS2 variants were associated with a higher risk of severe disease and increased susceptibility to human influenza (Cheng et al. 2015). Variants in the same gene they may also play a significant role in the interindividual

differences, particularly in the gender-related bias, of COVID-19 susceptibility and severity (Alshahawey et al. 2020). Moreover, TMPRSS2 polymorphisms may explain the higher mortality rate among Italians and the association to SARS-CoV-2 infection susceptibility in different populations (Asselta et al. 2020; Irham et al. 2020; Piva et al. 2021; Vargas-Alarcón et al. 2020; Hou et al. 2020).

Given their involvement in viral priming, we reasoned that members of the TTSP family may also represent targets of positive selection, possibly as the result of virus-driven pressure. To test this hypothesis, we systematically analyzed the evolutionary patterns of TTSP genes in a large phylogeny of mammals. We also evaluated their genetic diversity in human populations.

Material and methods

Sequences, alignments, and phylogenetic trees

Coding sequences for TTSPs were retrieved from the UCSC server (<http://genome.ucsc.edu/>). Mammalian homologs of human genes were included only if they represented 1-to-1 orthologs, as reported in the EnsemblCompara GeneTrees (Vilella et al. 2009). Sequences that carry stop codons or have a sequence coverage of less than 50% were excluded. The list of species analyzed for each gene is reported in Supplementary Table S1.

The RevTrans 2.0 utility was used to generate multiple sequence alignments (<http://www.cbs.dtu.dk/services/RevTrans/>, MAFFT v6.240 as an aligner) (Wernersson and Pedersen 2003). Phylogenetic trees were reconstructed using the phyML program with gamma-distributed rates, 4 substitution rate categories, and estimation of transition/transversion ratio and proportion of invariable sites (Guindon et al. 2009).

Evolutionary analysis in mammals

All alignments were screened for the presence of recombination using genetic algorithm recombination detection (GARD) (Kosakovsky Pond et al. 2006). GARD is a genetic algorithm implemented in the HYPHY suite, which uses phylogenetic incongruence among segments in the alignment to detect the best-fit number and location of recombination breakpoints.

To obtain an overview of the selective patterns of mammalian TTSPs coding genes, the average nonsynonymous substitution (dN)/synonymous substitution (dS) rate ratio was calculated using the single-likelihood ancestor counting (SLAC) method (Kosakovsky Pond and Frost 2005).

To detect positive selection, the codon-based codeml program implemented in the phylogenetic analysis by maximum likelihood (PAML) suite was applied (Yang 2007).

Specifically, a model (M8, positive selection model) that allows a class of sites to evolve with $dN/dS > 1$ was compared to two models (M7 and M8a, neutral models) that do not allow $dN/dS > 1$. To assess statistical significance, twice the difference of the likelihood ($\Delta \ln L$) for the models (M8a vs M8 and M7 vs M8) is compared to a χ^2 distribution (1 or 2 degrees of freedom for M8a vs M8 and M7 vs M8 comparisons, respectively). To assure reliability, two different codon substitution models were used (F3 \times 4 and F61).

Positively selected sites were identified using the following methods: (1) the Bayes empirical Bayes (BEB) analysis (with a cutoff of 0.90), which calculates the posterior probability that each codon is from the site class of positive selection (under model M8) (Anisimova et al. 2002); (2) fast unbiased Bayesian Approximation (FUBAR) (Murrell et al. 2013), an approximate hierarchical Bayesian method that generates an unconstrained distribution of selection parameters to estimate the posterior probability of positive diversifying selection at each site in a given alignment (with a cutoff ≥ 0.90); (3) the fixed effects likelihood (FEL) (Kosakovsky Pond and Frost 2005), a maximum-likelihood (ML) approach to infer dN/dS on a per-site basis, assuming that the selection pressure for each site is constant along the entire phylogeny (with a cutoff of ≥ 50). To be conservative, a site was considered under positive selection if it was detected by at least two methods. Protein structures were visualized with the PyMOL software (Schrödinger 2017), which was also used to create 3D protein figures.

GARD, FEL, FUBAR, and SLAC analyses were run locally through the HyPhy suite (Pond et al. 2005).

In order to identify specific branches with a proportion of sites evolving with $dN/dS > 1$, the adaptive Branch-Site Random Effects Likelihood method (aBS-REL) was used. This method applies sequential likelihood ratio tests to identify branches under positive selection without a priori knowledge about which lineages are of interest (Smith et al. 2015); branches identified using this approach were cross-validated using the branch-site likelihood ratio tests from PAML (models MA and MA1).

Population genetic analyses

Data from the 1000 Genomes Phase 1 Project (1000 Genomes Project Consortium et al. 2010) for individuals with East Asian (AS; Japanese plus Chinese) or European ancestry (CEU), and Yoruba subjects from Nigeria (YRI), were used to calculate nucleotide diversity and site frequency spectrum statistics for seven TTSP genes. Nucleotide diversity over the whole gene was measured as the average number of pairwise sequence nucleotide differences, π (Nei and Li 1979) and the expected per site heterozygosity, θ_W (Watterson 1975). The same statistics were also calculated in 5 kb sliding windows moving with a step of 250 bp. Three

site frequency spectrum (SFS) statistics (Tajima's D , DT) (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993) plus nucleotide diversity indicators were calculated for gene regions showing high level of diversity in the sliding-windows approach.

In order to obtain a statistical significance for these values, the same tests were performed for a control set of ~ 2000 genes, as previously described (Forni et al. 2014, 2013). Briefly, ~ 2000 genes were randomly selected, with the only limitations to be longer than 5000 bp and to have more than 80% human-outgroup (chimpanzee or orangutan genomes) aligning bases (Forni et al. 2014, 2013). This approach allowed us to generate a distribution of the above statistics and the calculation of empirical percentiles. We considered as statistically significant those values with a percentile rank higher than the 95th. For the gene region analysis, we selected 5 kb independent gene regions from the 2000 random genes and we used them to generate SFS distributions.

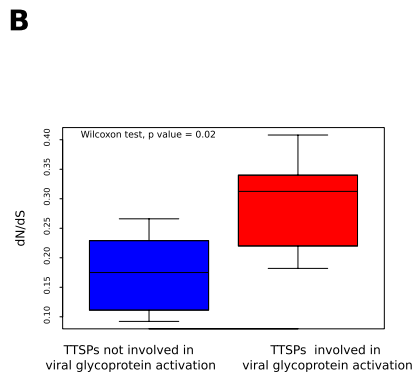
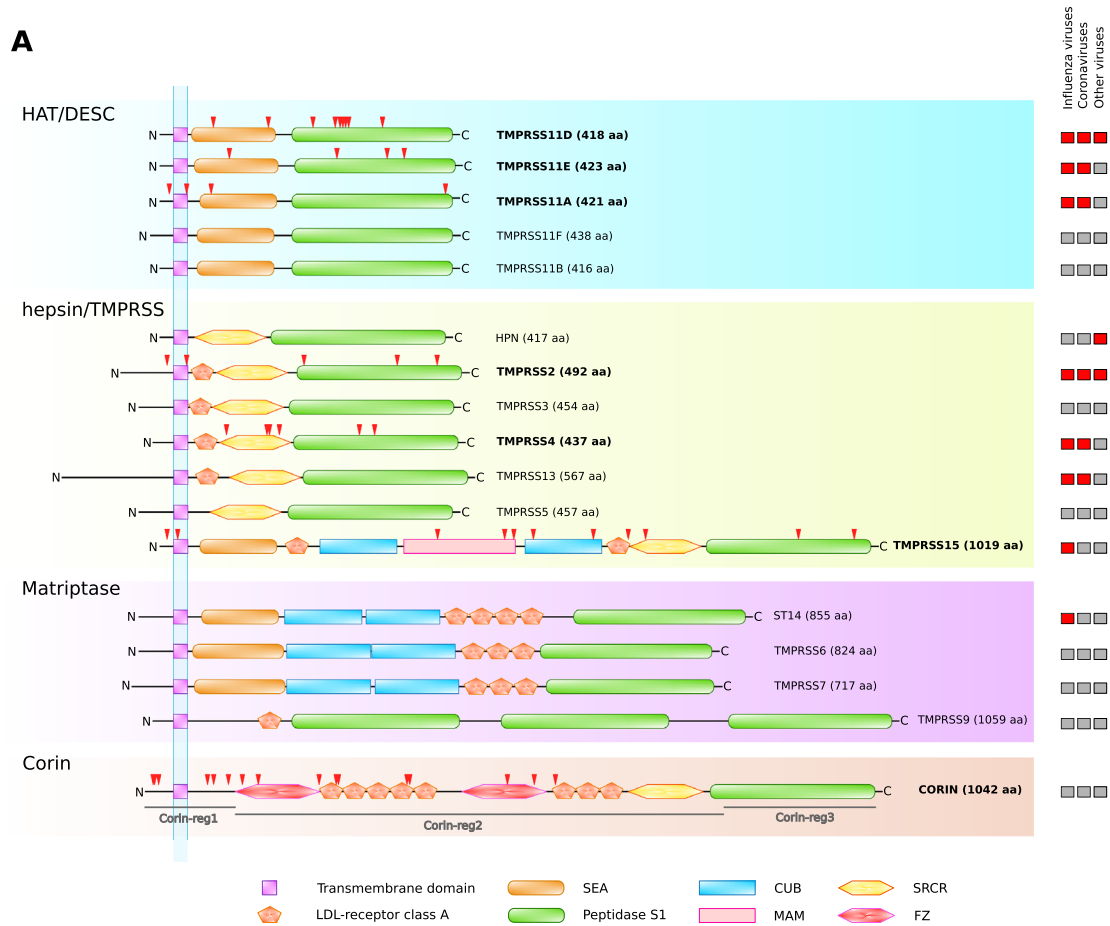
Orthologous genomic regions in chimpanzee, orangutan or macaque were retrieved using the liftOver tool (Kuhn et al. 2013).

Results and discussion

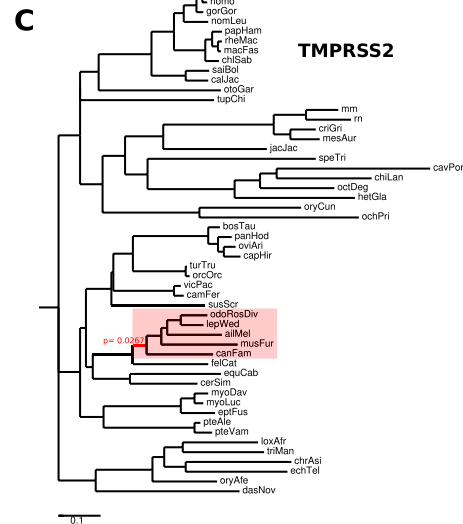
Evolutionary analysis of TTSPs in mammals

In humans, the TTSP family comprises 17 known members that share a number of common structural features (Fig. 1A). These include a short N-terminal cytoplasmic domain, a transmembrane domain, a variable stem region, and a C-terminal serine protease domain with a catalytic triad (Ser-His-Asp) (Böttcher-Friebertshäuser 2018). The stem region contains an assortment of modular structural domains that contribute to TTSP activation and interaction with substrates (Hooper et al. 2001; Antalis et al. 2010). Based on the composition of the extracellular protein portion and similarity in domain structure, chromosomal localization, and phylogenetic analysis of the serine protease domains, TTSPs are grouped in four subfamilies known as HAT/DESC, hepsin/TMPRSS (transmembrane protease/serine S1), matriptase, and corin (Böttcher-Friebertshäuser 2018; Szabo and Bugge 2008) (Fig. 1A).

TTSPs expressed in the human airways have been identified as host cell factors that may support the proteolytic activation and spread of respiratory viruses, including influenza viruses and coronaviruses. In particular, in vitro studies have demonstrated the involvement of 8 TTSP proteases (TMPRSS11D, TMPRSS11E, TMPRSS11A, TMPRSS2, TMPRSS4, TMPRSS13, TMPRSS15, and ST14) in the activation of HA and of spike proteins (Zang et al. 2020; Hoffmann et al. 2020b; Böttcher-Friebertshäuser 2018; Hayashi et al. 2018). Other viruses also require TTSPs (at



Gene	dN/dS
TMPRSS11D	0.408
TMPRSS11E	0.317
TMPRSS11A	0.331
TMPRSS11F	0.211
HPN	0.092
TMPRSS2	0.251
TMPRSS3	0.099
TMPRSS4	0.308
TMPRSS5	0.266
TMPRSS13	0.182
TMPRSS15	0.349
ST14	0.189
TMPRSS6	0.124
TMPRSS7	0.158
TMPRSS9	0.192
CORIN	0.306



least in vitro) for entry into the host cell. Indeed mammalian orthoreovirus, paramyxoviruses, human metapneumovirus, and human parainfluenza viruses require proteolytic activation through TMPRSS2 and/or TMPRSS11D at the viral

entry step (Nygaard et al. 2012; Böttcher-Friebertshäuser 2018). More recently, Sasaki and colleagues demonstrated that TMPRSS2 and TMPRSS11D also mediate proteolytic activation and trypsin-independent infection in group

Fig. 1 HYPERLINK "sps:id::fig1lllocator::gr1llmediaobject::0" Evolutionary Analysis of TTSPs in Mammals. **A** Schematic representation of Type II Transmembrane Serine Proteases (TTSPs). TTSPs are grouped into four subfamilies based on domain composition and phylogenetic analyses (Szabo and Bugge, 2008). The name and the localization of the domains in the protein sequences are taken from the Uniprot database (<https://www.uniprot.org/>). LDL-receptor class A: Low-density lipoprotein receptor domain class A; SRCR: Scavenger receptor Cys-rich; SEA: sea urchin sperm protein, enterokinase, agrin; CUB: C1r, C1s, uEGF, and bone morphogenetic protein; MAM: domain in meprin, A5, receptor protein tyrosine phosphatase mu; SRCR: Scavenger receptor Cys-rich; FZ: Frizzled. Red arrow heads indicate the position of positively selected sites (see Table 1). The CORIN regions obtained on the basis of the recombination breakpoints are also reported. Viruses target of proteolytic activation by TTSPs are also reported (see Supplementary Table S2). **B**. Comparison of dN/dS values. Boxplot comparison of average dN/dS values calculated for TTSP genes. *p* values from the Wilcoxon rank sum test and average dN/dS values for each gene are reported. TTSPs involved (red) or not involved (blue) in the activation of viral glycoproteins are indicated. **C**. Branch-site analysis of positive selection. aBS-REL analysis for *TMPRSS2* in mammals identified one branch showing evidence of positive selection (pink box). Branch lengths are scaled to the expected number of substitutions per nucleotide. The *p* value (red) from aBS-REL is reported. See Supplementary Table 1 for the full name of the species reported in the tree. See also Figure S1

A rotavirus (Sasaki et al. 2021) (Supplementary Table S2, Fig. 1A).

To comprehensively analyze the selective pressure acting on TTSPs and to obtain an estimate of the extent of constraint acting on these genes, we retrieved coding sequence information for all available mammalian species in the UCSC database (Supplementary Table 1). In particular, we retrieved sequences for at least 46 species for 16 TTSP genes. We excluded *TMPRSS11B* from the analysis, as the existence of a human pseudogene (*TMPRSS11BNL*) could affect evolutionary inference. Concerning *TMPRSS13*, we excluded from the analysis the region encoding the first 89 amino acids, as it showed limited similarity across species.

Because recombination can generate false positive inferences of positive selection (Sironi et al. 2015; Anisimova et al. 2003), all alignments were screened for the presence of recombination using genetic algorithm recombination detection (GARD) (Kosakovsky Pond et al. 2006). Evidence of recombination was detected only for the *CORIN* gene. The *CORIN* coding alignment was thus split on the basis of the recombination breakpoints in three sub-regions (*CORIN-reg1-3*) and then used for the analyses.

For coding genes, the strength of selection can be quantified by comparing the rate of non-synonymous nucleotide substitutions per non-synonymous site (dN) with that of synonymous substitution per synonymous site (dS) (Hughes and Nei 1989). We thus calculated the average dN/dS ratio (also referred to as ω) using the single-likelihood ancestor counting (SLAC) method (Kosakovsky Pond and Frost 2005). dN/dS values greater than 1 imply positive (diversifying)

selection, whereas ratios lower than 1 indicate purifying selection (functional constraint). The expected dN/dS under selective neutrality is 1. In the 16 TTSPs, the average dN/dS ratio varied from 0.092 to 0.408 (Fig. 1B). Thus, as is the case of most mammalian genes (Sironi et al. 2015), purifying selection is the major force acting on these proteases. However, comparison of dN/dS values between TTSPs not involved (as of present knowledge) in the activation of viral glycoproteins and those with a verified role in this process indicated that the latter show generally faster evolutionary rates based on overall dN/dS (Wilcoxon test, *p* value = 0.02) (Fig. 1B). This is consistent with the idea that positive selection may act on specific sites in proteins that are otherwise selectively constrained. To test this hypothesis in TTSP genes, we applied the “site models” implemented in the *codeml* tool from the PAML suite (see methods). Using likelihood ratio tests (LRTs), *codeml* compares models of gene evolution that allow (model M8) or disallow (models M8a and M7) a class of codons to evolve with dN/dS > 1. Specifically, M8 represents the positive selection model that is tested against the neutral M8a and M7 models.

To assess statistical significance, twice the difference of the likelihood ($\Delta\ln L$) for the models (M7 vs M8 and M8a vs M8) was compared to a χ^2 distribution (2 degrees of freedom for the M7 vs M8 comparison, 1 degree of freedom for M8a vs M8). To be conservative and to obtain robust results we applied two different codon frequency models (F3 × 4 and F61) and we called a gene as a target of positive selection only if it was detected by both comparisons (M7 versus M8, M8a versus M8) with both models.

Applying these criteria, we found seven TTSP genes (*TMPRSS11D*, *TMPRSS11E*, *TMPRSS11A*, *TMPRSS15*, *TMPRSS2*, *TMPRSS4*, *CORIN-reg1* and *CORIN-reg2*) to be targets of positive selection in mammals (Table 1). To identify specific sites targeted by positive selection, we again applied a conservative strategy and we called a site as positively selected only if it was detected by at least two of the following methods: BEB analysis from the M8 model, FUBAR and FEL methods (Kosakovsky Pond and Frost 2005; Murrell et al. 2013). Positively selected sites were identified in each of the seven genes and, as evident in Fig. 1A, they were not clustered, but rather distributed along the protein sequences. The only exception was *TMPRSS11D*, where a preponderance of selected sites was observed in the protease S1 domain.

We next extended our analysis to explore possible variations in selective pressure across lineages. In particular, we aimed to assess whether specific internal branches in the mammalian phylogenies evolved under episodic positive selection. Because we did not want to make any a priori assumption about which lineages were more likely to have experienced adaptive evolution, we applied the branch site-random effects likelihood (aBS-REL) method (Smith et al.

Table 1 Likelihood ratio test statistics for models of variable selective pressure among sites (F3×4 and F61 codon frequency model) for membrane-associated type II transmembrane serine protease genes

Gene/LRT model	n. of species	F3×4		F61		Positively selected sites ^c
		−2ΔlnL ^a	p value ^b	−2ΔlnL ^a	p value ^b	
<i>TMPRSS11D</i>	57			0.408		
M8a vs M8		19.9	2.81 × 10⁻¹⁰	29.29	1.95 × 10⁻¹⁴	75T, 156E, 215S, 248F, 254R, 259L, 260I, 265 K, 314Q
M7 vs M8		24.57	2.13 × 10⁻¹¹	35.03	6.12 × 10⁻¹⁶	
<i>TMPRSS11E</i>	57					
M8a vs M8		15.62	2.28 × 10⁻⁸	17.68	2.74 × 10⁻⁹	98E, 252K, 324Y, 347A
M7 vs M8		27.33	1.35 × 10⁻¹²	27.12	1.67 × 10⁻¹²	
<i>TMPRSS11A</i>	54					
M8a vs M8		2.95	0.0151	3.73	0.0063	14N, 39V, 62I, 413N
M7 vs M8		9.37	8.52 × 10⁻⁵	8.42	0.00022	
<i>TMPRSS11F</i>	56					
M8a vs M8		0	1	0	1	
M7 vs M8		0	1	1.39	0.25	
<i>HPN</i>	57					
M8a vs M8		0	1	0	1	
M7 vs M8		5.19	0.00557	2.77	0.063	
<i>TMPRSS2</i>	52					
M8a vs M8		2.12	0.0395	4.36	0.0031	103C, 140G, 301P, 433V, 489I
M7 vs M8		9.6	6.67 × 10⁻⁵	11.63	8.9 × 10⁻⁶	
<i>TMPRSS3</i>	54					
M8a vs M8		0	1	0.16	0.57	
M7 vs M8		10.62	2.44 × 10 ⁻⁵	5.53	0.004	
<i>TMPRSS4</i>	56					
M8a vs M8		9.26	1.68 × 10⁻⁵	8.56	3.51 × 10⁻⁵	103A, 161D, 166E, 179S, 298F, 319T
M7 vs M8		12.71	3.02 × 10⁻⁶	13.74	1.08 × 10⁻⁶	
<i>TMPRSS13</i>						
M8a vs M8		1.33	0.1	-0.18	1	
M7 vs M8		6.92	0.000988	3	0.05	
<i>TMPRSS5</i>	47					
M8a vs M8		1.17	0.13	0.03	0.81	
M7 vs M8		2.76	0.063	0.09	0.91	
<i>TMPRSS15</i>	57					
M8a vs M8		33.74	2.13 × 10⁻¹⁶	39.03	9.99 × 10⁻¹⁹	12H, 26A, 389A, 486I, 499Y, 545S, 621 V, 669E, 701 T, 918 V, 997L
M7 vs M8		59.31	1.75 × 10⁻²⁶	61.51	1.93 × 10⁻²⁷	
<i>ST14</i>	58					
M8a vs M8		0.12	0.62	9.68	1.08 × 10 ⁻⁵	
M7 vs M8		13.17	1.91 × 10 ⁻⁶	17.05	3.94 × 10 ⁻⁸	
<i>TMPRSS6</i>	57					
M8a vs M8		0	1	1.15	0.13	
M7 vs M8		11.86	7.07 × 10 ⁻⁶	17.12	3.67 × 10 ⁻⁸	
<i>TMPRSS7</i>	56					
M8a vs M8		0.12	0.62	0.82	0.2	
M7 vs M8		7.79	0.000414	9.01	0.00012	
<i>TMPRSS9</i>	56					
M8a vs M8		0.546	0.3	6.58	0.0003	
M7 vs M8		6.666	0.00127	11.66	8.63 × 10 ⁻⁶	
<i>CORIN-reg1</i>	57					

Table 1 (continued)

Gene/LRT model	n. of species	F3×4		F61		Positively selected sites ^c
		− 2ΔlnL ^a	<i>p</i> value ^b	− 2ΔlnL ^a	<i>p</i> value ^b	
M8a vs M8		8.96	0.0028	14.86	5.9 × 10^{−4}	12R, 13C, 21P, 90E, 99 T, 118 V
M7 vs M8		23.82	1.06 × 10^{−6}	37.28	8.03 × 10^{−9}	
<i>CORIN-reg2</i>	57					
M8a vs M8		6.98	0.0082	13.08	0.0014	138A, 163V, 252V, 274N, 278A, 375V, 380H, 520 V,
M7 vs M8		21.84	2.96 × 10^{−6}	37.52	7.12 × 10^{−9}	558S, 588R
<i>CORIN-reg3</i>	57					
M8a vs M8		0.12	0.729	12.64	0.0018	
M7 vs M8		0.92	0.337	13.4	0.0012	

Note: − 2ΔlnL values and *p* values of genes target of positive selection are shown in bold

^aTwice the difference of likelihood for the two models compared

^b*p* value of rejecting the neutral models (M8a and M7) in favor of the positive selection model (M8)

^cPositively selected sites detected by at least two methods among BEB, FEL, and FUBAR

2015) to the seven positively selected TTSPs. For three genes, aBS-REL identified the presence of a few positively selected branches with borderline *p* values (Fig. 1C and Supplementary Fig. S1). In particular, evidence of selection was detected for the branch leading to *Eptesicus fuscus* and the *Myotis* clade for *TMPRSS11D*, to the myomorpha suborder of *Rodentia* for *TMPRSS11A*, and to some lineages of the *carnivora* order for *TMPRSS2*. However, these results were not confirmed using the *codeml* branch-site LRT models (Zhang et al. 2005), suggesting that positive selection in TTSP genes is mainly pervasive rather than episodic.

As mentioned above, an involvement of TTSP genes in the activation and spread of influenza viruses and/or coronaviruses was verified for 8 TTSP genes (Böttcher-Friebertshäuser 2018; Hayashi et al. 2018). Notably, six of the seven positively selected genes (*TMPRSS11D*, *TMPRSS11E*, *TMPRSS11A*, *TMPRSS15*, *TMPRSS2*, and *TMPRSS4*) are involved in the processing of viral glycoproteins, strongly suggesting that viral infections could represent the selective pressure driving the evolution of these TTSPs. Conversely, proteases with no known role in viral cell entry displayed no positive selection signals. The only exception was the *CORIN* gene, in which positive selection signals were identified in two of the 3 analyzed sub-regions.

Corin is a cardiac membrane serine protease and plays a key role in maintaining cardiac function and regulating blood pressure (Yan et al. 2000). Corin deficiency causes spontaneous hypertension and cardiac hypertrophy in mice (Chan et al. 2005), and variants in the human *CORIN* gene can alter corin protein conformation, inhibit corin zymogen activation, and contribute to hypertension and cardiovascular diseases (Zhou and Wu 2014; Peng et al. 2015).

The N-terminal cytoplasmic tail and the transmembrane domain are unnecessary for corin expression and

catalytic activity (Knappe et al. 2003). However truncation of the cytoplasmic tail can create a new N-terminus that inhibits corin trafficking in the Golgi (He et al. 2020). Interestingly, three out of the 16 positively selected sites localized to the N-cytoplasmic tail. In particular, positively selected site Cys13 is polymorphic in human populations (rs2289433, p.Cys13Tyr) and it has been associated with serum corin levels (Zhang et al. 2019a, b).

Taking into consideration the physiological functions of this protein and that, up to date, the involvement of corin in viral activation and spread has not been reported, the signal of positive selection is most likely driven by non-viral factors.

For the other six positively selected genes, their role in the processing of viral glycoproteins has been widely described (Böttcher-Friebertshäuser 2018; Hayashi et al. 2018; Zmora et al. 2018; Kishimoto et al. 2021; Zang et al. 2020), supporting the hypothesis that viral infections represented the main selective force in the evolutionary history of these genes. For instance, in swine and mouse, *TMPRSS2* can activate HA with monobasic cleavage site, indicating that homologous proteases are involved in HA cleavage in different host species (Tarnow et al. 2014; Peitsch et al. 2014).

In addition to the canonical activator *TMPRSS2*, other TTSPs have been reported to function as entry cofactors of different viral fusion proteins, at least in vitro. For example, it was recently reported that *TMPRSS4* works synergistically with *TMPRSS2* to promote SARS-CoV-2 infection of small intestinal enterocytes (Zhang et al. 2020), and three members of the HAT/DESC subfamily (*TMPRSS11D*, *TMPRSS11A*, and *TMPRSS11E*) are able to activate the SARS-CoV, SARS-CoV-2, and MERS-CoV spike proteins,

as well as influenza HA (Böttcher-Friebertshäuser 2018; Hayashi et al. 2018; Zmora et al. 2018; Kishimoto et al. 2021; Zhang et al. 2020; Bertram et al. 2011).

The redundancy of function of the different proteases may appear to be in contrast with the presence of positive selection signals in the TTSP genes. However, we wish to mention that several lines of evidence suggest that this might not be the case. First, not all TTSPs have the same tissue/cell expression and not all of them may have the same cleavage efficiency. Second, protease activation may occur through different mechanisms in cells, as recently suggested by Liu and colleagues based on proteomic interaction data (Liu et al. 2021). Third, most data suggesting redundancy of TTSP function in terms of viral protein cleavage have been obtained in vitro and do not necessarily translate to the situation in vivo or in natural infections. Finally, since more proteases are able to cleave proteins of different viruses and the same viral protein can be cleaved by more proteases, different infectious agents may have exerted distinct selection pressures on the same gene, and this could explain, at least in part, the presence of selected sites distributed in different domains.

As stated above, positively selected sites in TTSPs genes are not clustered, but rather distributed along the protein sequences. However, all TTSPs involved in the processing of viral glycoproteins have at least one selected site in the peptidase domain (Fig. 1A). Although such sites are not located near the active site (catalytic triad) of the protease domain (Fig. 2), we cannot exclude that they may have a role in modulating the proteolytic activity of these proteins. Likewise, we cannot rule out the possibility that the sites identified in the protease domains, as well as in other protein regions, are involved in direct interactions with the viral target proteins (e.g., spike protein in coronavirus, HA in influenza viruses, VP4 protein in group A rotaviruses). It is currently unknown whether proteases are able to establish direct interactions with their substrates.

In summary, evolutionary analyses indicate that most of the TTSP genes that have a verified role in proteolytic activation and spread of influenza viruses and/or coronaviruses present pervasive positive selection signals. This evidence strongly suggests that viral infections represent a major selective pressure driving the evolution of these type II transmembrane serine proteases in mammals. Nonetheless, we cannot exclude that the selection signals are due, at least partially, to other selective pressures associated with the physiological functions of these proteins. Future studies are required to clarify this aspect and to evaluate the functional relevance and selective advantages associated with these genetic changes in TTSP proteins during evolution.

Population genetic analysis in humans

Infectious diseases have exerted strong selective pressures throughout human history (Haldane 1932; Fumagalli et al. 2011). Indeed, a number of human loci have been shown to evolve in response to such pressures; the best documented example is the major histocompatibility complex (MHC) (Bernatchez and Landry. 2003), which is highly polymorphic in several species as a consequence of balancing selection. This is a process whereby genetic variability is maintained in populations due to some selective pressure. Molecular signals of balancing selection have also been described for several genes involved in immune defense, including toll-like receptors (Fumagalli et al. 2009; Ferrer-Admetlla et al. 2008), various cytokine genes (Fumagalli et al. 2009; Ferrer-Admetlla et al. 2008), antimicrobial peptides (AMPs) (Cagliani et al. 2008; Hollox and Armour. 2008), and *TRIM5* (Cagliani et al. 2010).

Thus, we addressed the role of natural selection in the shaping of genetic diversity at the seven positively selected TTSP genes in human populations. To this aim, we calculated θ_W (an estimate of the expected per site heterozygosity) (Watterson 1975) and π (the average number of pairwise sequence nucleotide differences between haplotypes) (Nei and Li 1979) in individuals with European ancestry (CEU), in Yoruba subjects from Nigeria (YRI), and in East Asian subjects (AS; Japanese plus Chinese). We then compared the values with the distribution obtained from 2000 randomly selected genes (see Methods for details).

As reported in Supplementary Table S3, only *TMPRSS2* and *TMPRSS4* showed evidence of high nucleotide diversity in all populations, although the rank of θ_W and π did not reach the 95th percentile in all of them.

High nucleotide diversity might be suggestive of balancing selection. Yet, as a result of mutations and recombination, balancing selection signatures tend to extend over relatively short genomic intervals (Wiuf et al. 2004; Charlesworth. 2006). Thus, in order to map the region(s) responsible for high nucleotide diversity in *TMPRSS2* and *TMPRSS4*, we calculated θ_W and π in 5 kb sliding-windows moving along these genes. In order to analyze potential regulatory regions, we extended the analysis to 20 kb upstream and downstream each gene. Overall, the same trend of θ_W and π was evident in all populations for both genes (Figs. 3, 4). In *TMPRSS2*, sliding-windows analysis highlighted 3 regions with high nucleotide diversity in at least two populations (Fig. 3 and Table 2): one located in the 3' region of *TMPRSS2* (TMPRSS2-reg1) and falling in the *MX1* gene, one in *TMPRSS2* intron 8 (TMPRSS2-reg2), and one covering the promoter region (TMPRSS2-reg3). In the case of *TMPRSS4*, only two intragenic regions were identified (Fig. 4 and Table 2): one covering the 2 kb surrounding

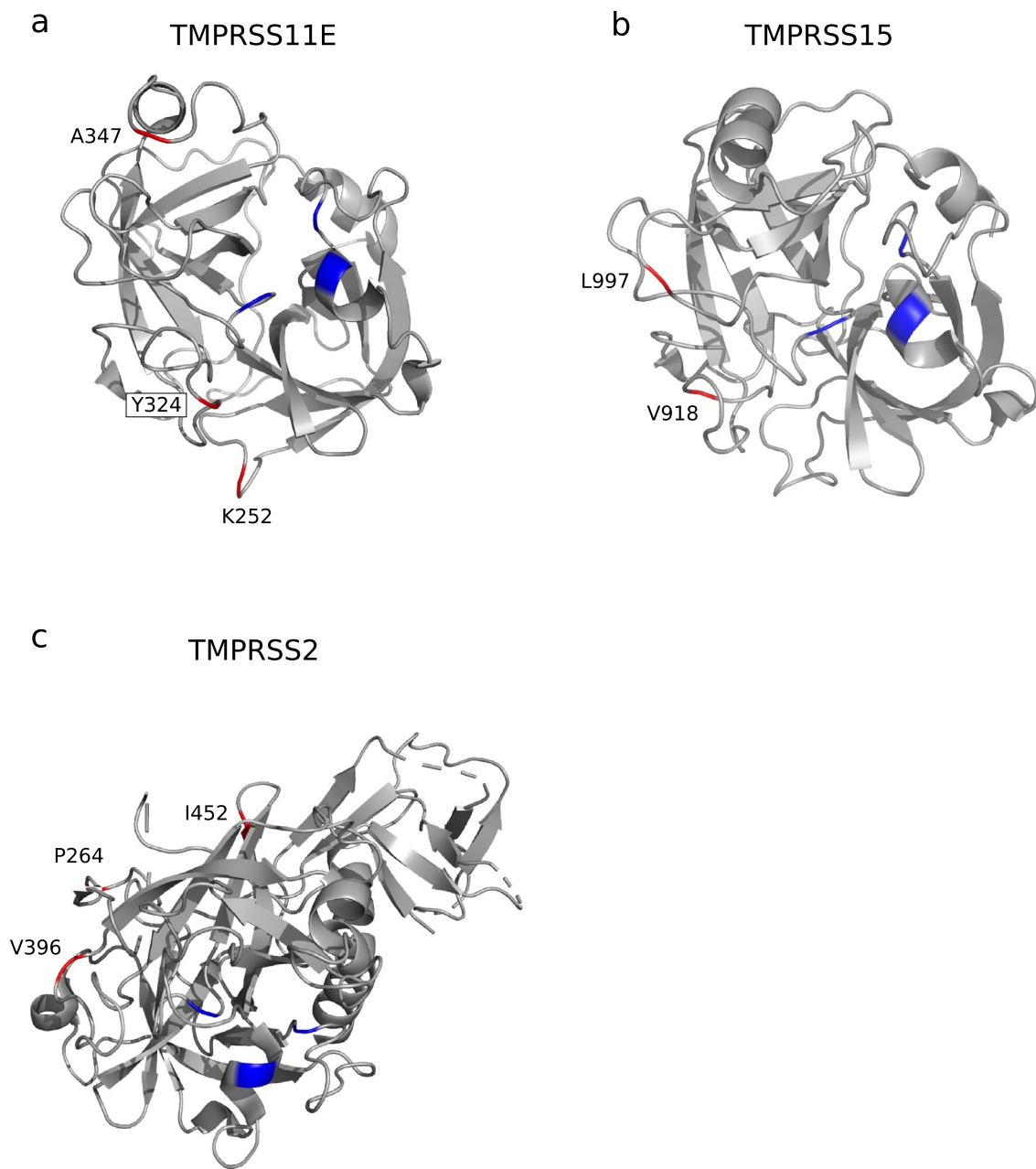


Fig. 2 Positive selection acting on TTSP serine protease domains. Cartoon 3D representation of the serine protease domain for TMPRSS11E (a), TMPRSS15 (b), and TMPRSS2 (c). Positively selected sites are colored in red and refer to the human protein; the catalytic triad (Ser-His-Asp) is shown in blue. TMPRSS11E,

TMPRSS15, and TMPRSS2 serine protease domain structures are derived from the Protein Data Bank (PDB) ids 2OQ5, 4DGJ, and 7MEQ, respectively. For all other TTSP domains with positively selected sites (see Fig. 1A and Table 1), no 3D structures were available in the PDB database

exon 2 (TMPRSS4-reg1) and one covering the entire 3'UTR (TMPRSS4-reg2).

An effect of balancing selection is a distortion of the site frequency spectrum (SFS) toward intermediate frequency alleles. Deviation in the SFS is identified by neutrality tests, including Tajima's D (DT) (Tajima 1989) and Fu and Li's D^* and F^* (Fu and Li. 1993). DT tests the departure

from neutrality by comparing θ_w and π , and positive values indicate an excess of intermediate frequency variants. Fu and Li's F^* and D^* are also based on SNP frequency spectra and differ from DT in that they also take into account whether mutations occur in external or internal branches of a genealogy. For TMPRSS2-reg1 and TMPRSS2-reg2, tests indicated departure from neutrality with significantly

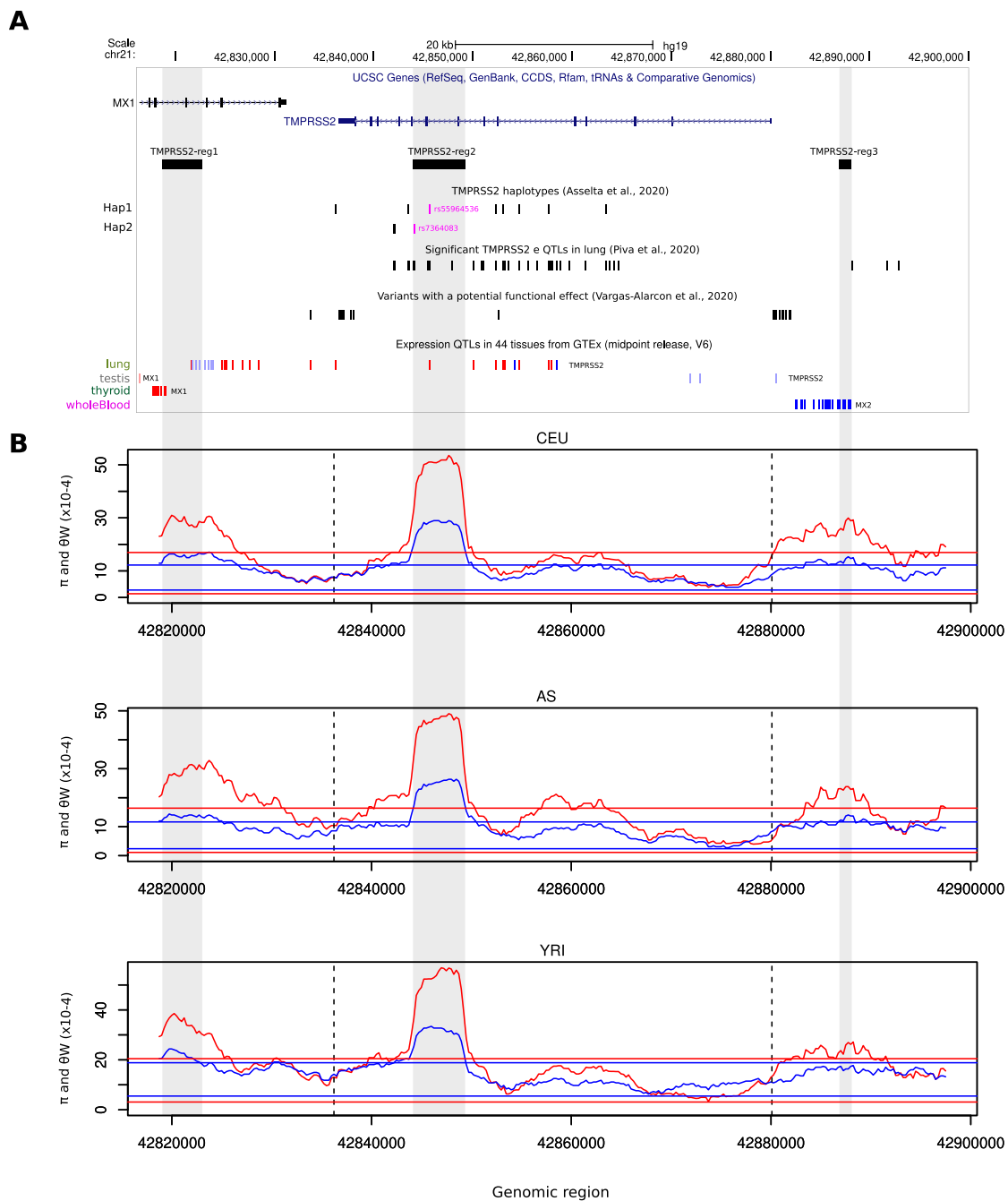


Fig. 3 Human population genetic analysis of *TMRSS2*. **A** Screenshot from the UCSC genome browser (<http://genome.ucsc.edu/>, GRCh37/hg19). The panel shows the analyzed genomic region (chr21:42816232–42900085), polymorphic variants associated with COVID-19 and with a potential functional effects (Asselta et al. 2020; Piva et al. 2021; Vargas-Alarcón et al. 2020), as well as a track with genetic variants likely affecting proximal gene expression in human tissues from the genotype-tissue expression (GTEx, V6 data release). The eQTL item color indicates the effect size attributed to

the eQTL: red, high positive; light red, moderate positive; light blue, moderate negative; blue, high negative. **B** Nucleotide diversity analysis in human populations. Data from 1000 genomes Project were used to calculate θ_W and π in sliding windows of 5 kb moving with a step of 250 bp. Horizontal lines represent the 95th percentiles in the distribution of θ_W (blue) and π (red). Vertical dashed lines indicate the start and the end of *TMRSS2*. Regions with θ_W and π greater than the 95th percentile in at least two populations are shaded in gray

positive values for most statistics in all populations (Table 2). Conversely, for *TMRSS2*-reg3, neutrality tests showed more significant values in the CEU population.

With respect to *TMRSS4*, significantly positive values for most statistics were observed only in the region covering the 3'-UTR (Table 2). Overall, these analyses indicate that

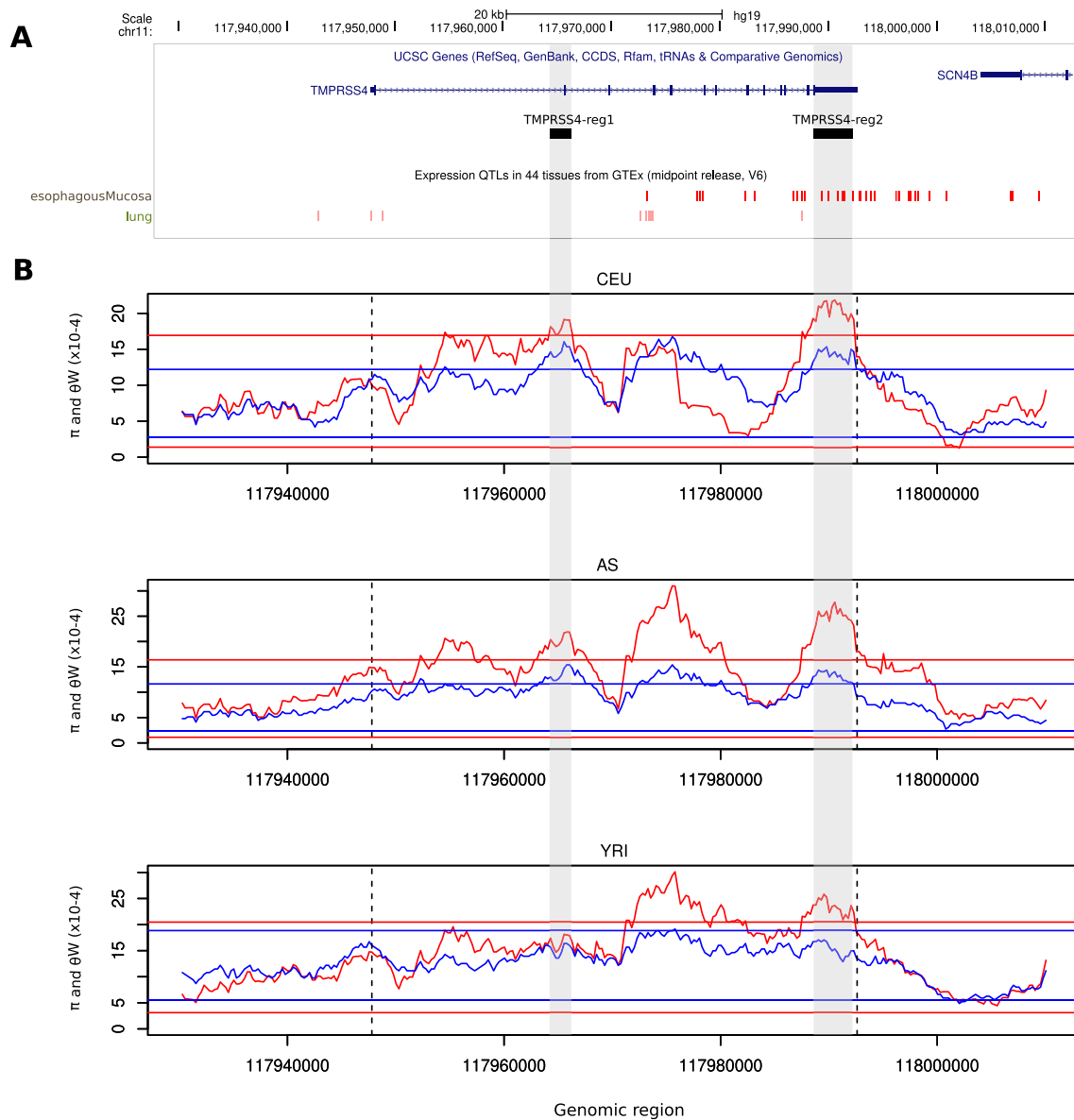


Fig. 4 Human population genetic analysis of *TMPRSS4*. **A** Screenshot from the UCSC genome browser (<http://genome.ucsc.edu/>, GRCh37/hg19). The panel shows the analyzed genomic region (chr11:117927793–118012605) and a track with genetic variants likely affecting proximal gene expression in human tissues from the genotype-tissue expression (GTEx, V6 data release). The eQTL item color indicates the effect size attributed to the eQTL: red, high positive; light red, moderate positive; light blue, moderate negative; blue,

high negative. **B** Nucleotide diversity analysis in human populations. Data from 1000 genomes Project were used to calculate θ_W and π in sliding windows of 5 kb moving with a step of 250 bp. Horizontal lines represent the 95th percentiles in the distribution of θ_W (blue) and π (red). Vertical dashed lines indicate the start and the end of *TMPRSS4*. Regions with θ_W and π greater than the 95th percentile in at least two populations are shaded in gray

both *TMPRSS2* and *TMPRSS4* may be targets of ongoing balancing selection in human populations.

The action of natural selection obviously implies the presence of a functional variant with an effect on fitness and a selective pressure to act on it. Recently, variants in *TMPRSS2* with a potential functional role and/or associated to viral infection susceptibility have been identified

(Asselta et al. 2020; Irham et al. 2020; Piva et al. 2021; Vargas-Alarcón et al. 2020). Specifically, two variants in *TMPRSS2*-reg2 (rs55964536 and rs7364083) define two distinct haplotypes with different frequencies among populations and were predicted to modulate *TMPRSS2* expression (Asselta et al. 2020) (Fig. 3). The rs55964536 polymorphism was reported as an eQTL in lung tissue (genotype-tissue

Table 2 Nucleotide diversity and neutrality tests for the *TMPRSS2* and *TMPRSS4* gene regions

Gene (chromosomal position)	Population	S^a	$\pi (\times 10^{-4})$		$\theta_w (\times 10^{-4})$		Tajima's D		Fu and Li's D^*		Fu and Li's F^*	
			Value	Rank ^b	Value	Rank ^b	Value	Rank ^b	Value	Rank ^b	Value	Rank ^b
<i>TMPRSS2-reg1</i> (chr21:42,818,732–42,822,732)	CEU	38	29.96	99	16.57	99	2.38	97	1.5	97	2.25	99
	YRI	58	38.95	99	25.24	99	1.65	99	0.98	80	1.53	97
	AS	33	29.12	99	14.12	99	3.05	99	2.06	99	2.98	99
<i>TMPRSS2-reg2</i> (chr21:42,843,982–42,849,232)	CEU	86	50.54	99	28.57	99	2.39	98	1.55	97	2.32	99
	YRI	99	54.68	99	32.82	99	2.08	99	1.88	99	2.37	99
	AS	78	47.47	99	25.43	99	2.66	98	1.23	93	2.25	99
<i>TMPRSS2-reg3</i> (chr21:42,886,982–42,888,232)	CEU	9	27.71	99	12.56	96	2.79	99	1.28	93	2.17	99
	YRI	10	25.3	98	13.93	83	1.94	99	-0.19	36	0.69	81
	AS	10	28.52	99	13.7	98	2.54	97	0.56	77	1.52	93
<i>TMPRSS4-reg1</i> (chr11:117,964,293–117,966,293)	CEU	21	20.03	98	18.31	99	0.26	56	-0.15	48	0.01	51
	YRI	21	17.41	92	18.28	95	-0.13	67	0.81	74	0.53	75
	AS	21	24.05	99	17.97	99	0.91	73	0.3	69	0.66	74
<i>TMPRSS4-reg2</i> (chr11:117,988,543–117,992,043)	CEU	31	23.79	99	15.45	99	1.55	89	1.65	98	1.93	97
	YRI	31	25.76	98	15.42	89	1.93	99	1.65	97	2.12	99
	AS	29	27.32	99	14.18	99	2.62	98	0.39	73	1.58	94

Significant values are in bold

^aNumber of segregating sites

^bPercentile rank relative to a distribution of 5-kb regions from 2000 random genes

expression, GTEx) (GTEx Consortium et al. 2017) and the haplotype to which rs55964536 belongs is functionally linked to another eQTL variant (rs8134378) located at a known androgen-responsive enhancer for *TMPRSS2* (Asselta et al. 2020). Based on these observations, Asselta and colleagues suggested that upregulation of *TMPRSS2* expression level through androgens could explain the sex bias in COVID-19 severity. On the other hand, rs7364083 belongs to the second haplotype. This latter is also defined by a variant (rs2070788) associated with increased gene expression, a higher risk of severe disease with the 2009 H1N1 pandemic influenza, and increased susceptibility to human H7N9 flu (Cheng et al. 2015).

In *TMPRSS2-reg2*, another variant (rs2298661) in association with both COVID-19 disease susceptibility and with increased risk of worse disease outcomes was identified through a trans-ancestry meta-analysis (Horowitz et al. 2021). The rs2298661 polymorphism is in tight linkage disequilibrium ($LD R^2 > 0.8$) with a non-synonymous variant (rs12329760, p.V160M). V160M falls in an exonic splicing enhancer site (Srp40) and is also predicted to decrease protein stability in structural models, with a plausible role in viral entry (Vishnubhotla et al. 2020).

Taken together, these evidences suggest that the balancing selection signal in *TMPRSS2-reg2* could be secondary to the presence of variants with a potential functional effect, most likely on gene expression regulation. Overall, these variants with different allele frequencies in different populations could play an important role in viral entry,

helping to explain the heterogeneity observed in different ethnic groups with regard to incidence, severity, progression, and therapeutic response to COVID-19 (Alshahawey et al. 2020).

As reported in Fig. 3, other eQTL variants fall in *TMPRSS2-reg1* and *-reg3*. In particular, a cluster of eQTLs modulating expression levels of *TMPRSS2* and *MX1* in lung and thyroid falls in *TMPRSS2-reg1*, whereas whole-blood *MX2* eQTLs are present in *TMPRSS2-reg3*. *MX1* and *MX2* are involved in antiviral response. In particular, *MX1* can restrict a wide variety of viruses, including orthomyxoviruses (e.g., influenza and Thogoto), bunyaviruses (e.g., La Crosse and rift valley viruses), and hepatitis B virus (Verhelst et al. 2013). Conversely, *MX2* is known to counteract retroviruses (Kane et al. 2013), and to exert a weak effect against vesicular stomatitis virus infection (Liu et al. 2012). Moreover, *MX1* expression is higher in COVID-19 patients compared to non-infected subjects and tends to decrease with age (Bizzotto et al. 2020). Variants in *TMPRSS2-reg1* might thus contribute to modulate both *TMPRSS2* and *MX1* expression, acting on the ability to infect and to trigger an effective antiviral response.

As stated above, *TMPRSS2-reg1* falls within the *MX1* gene, hence deviation from neutrality with significantly positive values for most SFS statistics in human populations may be attributable to a specific selection signal for the *MX1* gene rather than *TMPRSS2*. However, within this region, there are eQTL variants for *TMPRSS2*; these variants may act to determine a different expression of the protease and,

consequently, a different ability to activate the viral proteins that are *TMPRSS2* protease targets.

With respect to *TMPRSS4*, information on its regulation is more limited. However, eQTLs are present in *TMPRSS4-reg2*. Specifically, different eQTLs for the esophageal mucosa map to this genomic region.

In summary, high nucleotide diversity and a shift in the SFS toward intermediate-frequency alleles suggest that the natural selection signals in these *TMPRSS2* and *TMPRSS4* regions are driven by the presence of polymorphic variants with a functional role that could contribute to infection with and spread of viruses, and/or to the modulation of disease clinical course.

Conclusions

Type II transmembrane serine proteases (TTSPs) are involved in the proteolytic activation of different viruses. In particular, the cleavage of surface glycoproteins by some proteases has been reported for influenza viruses and coronaviruses, but also for other respiratory viruses. These observations led us to hypothesize that the members of this family may be targets of virus-driven selective pressure. It is known that natural selection events acted on the receptors used by viruses to promote their entry into the cell during the early stages of infection. From this point of view, coronavirus receptors, such as ACE2, DPP4, and ANPEP have been extensively studied. However, until now, systematic evolutionary studies had never been carried out on TTSP gene family.

In this study, the application of specific molecular evolution tests showed that 7 TTSP genes evolved adaptively in mammals. Six of these are involved in the proteolytic activation of different viral proteins, suggesting that host–pathogen arms races drive the evolution of these genes.

Since more proteases are able to prime proteins of different viruses, we may speculate that they play an important role in determining viral tropism and disease progression, explaining, at least partially, the apparent redundancy of the function of these proteases. Regarding the positively selected sites, we observed that all the TTSP genes have at least one site in the protease domain, that these sites are not in close proximity to the catalytic domain and that in general the selected sites are distributed along the gene. Taken together, these results suggest that different infectious agents may have exerted distinct selection pressures on the same gene and that sites targeted by positive selection in the protease and in other domains could participate in the formation of bonds or in stabilizing the reciprocal position of the host and viral proteins.

Of particular interest, especially at the present time, is *TMPRSS2*. This protease represents the canonical activator

of SARS-CoV-2, although its role in priming has been confirmed for many other respiratory viruses.

Notably, *TMPRSS2* presents pervasive positive selection signals in the mammalian phylogeny and could be target of ongoing balancing selection in human populations. In the latter case, the signals of natural selection could be driven by the presence of polymorphic variants with a functional role that may contribute to the infection with and the spread of viruses, and/or to the modulation of the clinical course of the disease.

As the world continues to wrestle with the COVID-19 pandemic, research on the virus aims to develop therapies to control the health and socio-economic emergency.

Different vaccine platforms are being used for vaccine development (Kyriakidis et al. 2021), most of them aiming to induce neutralizing antibodies against the spike protein (S). Some of these are already used in vaccination campaigns worldwide, others are close to receiving approval or authorization for large-scale vaccinations. However, the emergence of new lineages of SARS-CoV-2 carrying variants in the spike protein raises concerns about a reduction in the efficacy of the vaccines currently in use. Furthermore, vaccination during the pandemic could exert selective pressures that might favor the expansion of vaccine-resistant SARS-CoV-2 lineages. In this context, it is important to combine vaccinations with alternative therapies aimed at impairing the virus capacity to spread through the host cells. Thus, several studies have identified serine proteases as potential drug targets for the treatment of SARS-CoV-2 infection (Baughn et al. 2020; Qiao et al. 2020; Tang et al. 2020; Zarubin et al. 2020; Deng et al. 2021).

TMPRSS2 is a well-characterized TTSP and represents one of the essential host factors in viral activation and cell entry of influenza viruses, such as pandemic influenza H1N1, and different coronaviruses, including SARS-CoV-2. *TMPRSS2* is dispensable for development and homeostasis in mice (Kim et al. 2006). Thus, this protease is potentially one of the most promising drug targets for COVID-19 therapy. Therapeutics targeting *TMPRSS2* expression and/or activity may be beneficial not only for coronavirus-infected patients but also for those infected with influenza viruses. Indeed, promising data have been obtained from the use of *TMPRSS2* inhibitors such as camostat and nafamostat mesylate, which seem to have good efficacy against coronaviruses (Hoffmann et al. 2020b, c). The design of selective inhibitors will be facilitated by a better understanding of functional genetic diversity, both in humans and in other mammals that are used as model organisms for drug development.

Data reported in this study indicate that, in analogy to other genes with a role in viral infections, TTSPs have evolved adaptively in mammals (Sironi et al. 2015). In particular a continuum of selective pressure acting on *TMPRSS2*

and *TMPRSS4* was observed as these genes also represent selection targets in human populations. This scenario may be the result of an ancestral and still ongoing host–pathogen arms race.

Overall, our results provide evolutionary information about candidate functional sites that may shape infection outcomes across mammals and within humans.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00439-022-02435-y>.

Author contributions Conceptualization, RC and MS; formal analysis, RC, DF, and MS; investigation, RC; visualization, RC and DF; writing—original draft, RC, MS and DF; supervision, RC and MS. All authors reviewed the manuscript.

Funding This work was supported by the Italian Ministry of Health (“Ricerca Corrente 2021” to MS).

Declarations

Conflict of interest The authors declare no competing interests.

References

- 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073. <https://doi.org/10.1038/nature09534>
- Alshahawey M, Raslan M, Sabri N (2020) Sex-mediated effects of ACE2 and TMPRSS2 on the incidence and severity of COVID-19; the need for genetic implementation. *Curr Res Transl Med* 68:149–150 (S2452-3186(20)30054-4)
- Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol Biol Evol* 19:950–958
- Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164:1229–1236
- Antalis TM, Buzza MS, Hodge KM, Hooper JD, Netzel-Arnett S (2010) The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment. *Biochem J* 428:325–346. <https://doi.org/10.1042/BJ20100046>
- Asselta R, Paraboschi EM, Mantovani A, Duga S (2020) ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging (albania NY)* 12:10087–10098. <https://doi.org/10.18632/aging.103415>
- Barrett CT, Dutch RE (2020) Viral membrane fusion and the trans-membrane domain. *Viruses* 12:693. <https://doi.org/10.3390/v12070693>
- Baughn LB, Sharma N, Elhaik E, Sekulic A, Bryce AH, Fonseca R (2020) Targeting TMPRSS2 in SARS-CoV-2 infection. *Mayo Clin Proc* 95:1989–1999 (S0025-6196(20)30626-1)
- Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol* 16:363–377. <https://doi.org/10.1046/j.1420-9101.2003.00531.x>
- Bertram S, Glowacka I, Blazejewska P, Soilleux E, Allen P, Danisch S, Steffen I, Choi SY, Park Y, Schneider H, Schughart K, Pöhlmann S (2010) TMPRSS2 and TMPRSS4 facilitate trypsin-independent spread of influenza virus in Caco-2 cells. *J Virol* 84:10016–10025. <https://doi.org/10.1128/JVI.00239-10>
- Bertram S, Glowacka I, Müller MA, Lavender H, Gnirss K, Nehlmeier I, Niemeyer D, He Y, Simmons G, Drosten C, Soilleux EJ, Jahn O, Steffen I, Pöhlmann S (2011) Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. *J Virol* 85:13363–13372. <https://doi.org/10.1128/JVI.05300-11>
- Bertram S, Dijkman R, Habjan M, Heurich A, Gierer S, Glowacka I, Welsch K, Winkler M, Schneider H, Hofmann-Winkler H, Thiel V, Pöhlmann S (2013) TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. *J Virol* 87:6150–6160. <https://doi.org/10.1128/JVI.03372-12>
- Bibiana SOF, Vargas-Pinilla P, Amorim CEG, Sortica VA, Bortolini MC (2020) ACE2 diversity in placental mammals reveals the evolutionary strategy of SARS-CoV-2. *Genet Mol Biol* 43:e20200104. <https://doi.org/10.1590/1678-4685-GMB-2020-0104>
- Bizzotto J, Sanchis P, Abbate M, Lage-Vickers S, Lavignolle R, Toro A, Olszewicki S, Sabater A, Cascardo F, Vazquez E, Cotignola J, Gueron G (2020) SARS-CoV-2 infection boosts MX1 antiviral effector in COVID-19 patients. *iScience* 23:101585. <https://doi.org/10.1016/j.isci.2020.101585>
- Böttcher-Friebertshäuser E (2018) Membrane-anchored serine proteases: host cell factors in proteolytic activation of viral glycoproteins. In: Böttcher-Friebertshäuser E, Garten W, Klenk HD (eds) *Activation of Viruses by Host Proteases*. Springer, Switzerland, pp 153–203
- Bugge TH, Antalis TM, Wu Q (2009) Type II transmembrane serine proteases. *J Biol Chem* 284:23177–23181. <https://doi.org/10.1074/jbc.R109.021006>
- Cagliani R, Fumagalli M, Riva S, Pozzoli U, Comi GP, Menozzi G, Bresolin N, Sironi M (2008) The signature of long-standing balancing selection at the human defensin beta-1 promoter. *Genome Biol* 9:R143. <https://doi.org/10.1186/gb-2008-9-9-r143>
- Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, Pozzoli U, Bonaglia MC, Bresolin N, Clerici M, Sironi M (2010) Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Hum Genet*. <https://doi.org/10.1007/s00439-010-0884-6>
- Chan JC, Knudson O, Wu F, Morser J, Dole WP, Wu Q (2005) Hypertension in mice lacking the proatrial natriuretic peptide convertase corin. *Proc Natl Acad Sci USA* 102:785–790. <https://doi.org/10.1073/pnas.0407234102>
- Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* 2:e64. <https://doi.org/10.1371/journal.pgen.0020064>
- Cheng Z, Zhou J, To KK, Chu H, Li C, Wang D, Yang D, Zheng S, Hao K, Bossé Y, Obeidat M, Brandsma CA, Song YQ, Chen Y, Zheng BJ, Li L, Yuen KY (2015) Identification of TMPRSS2 as a susceptibility gene for severe 2009 pandemic A(H1N1) influenza and A(H7N9) influenza. *J Infect Dis* 212:1214–1221. <https://doi.org/10.1093/infdis/jiv246>
- Choi SY, Bertram S, Glowacka I, Park YW, Pöhlmann S (2009) Type II transmembrane serine proteases in cancer and viral infections. *Trends Mol Med* 15:303–312. <https://doi.org/10.1016/j.molmed.2009.05.003>
- Cui J, Eden JS, Holmes EC, Wang LF (2013) Adaptive evolution of bat dipeptidyl peptidase 4 (dpp4): implications for the origin and emergence of Middle East respiratory syndrome coronavirus. *Virol J* 10:304–422X-10–304. <https://doi.org/10.1186/1743-422X-10-304>
- Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, Hiller M, Koepfli KP, Pfenning AR, Zhao H, Genereux DP, Swofford R, Pollard KS, Ryder OA, Nweeia MT, Lindblad-Toh

- K, Teeling EC, Karlsson EK, Lewin HA (2020) Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc Natl Acad Sci USA* 117:22311–22322. <https://doi.org/10.1073/pnas.2010146117>
- Deng Q, ur Rasool R, Russell RM, Natesan R, Asangani IA (2021) Targeting androgen regulation of TMPRSS2 and ACE2 as a therapeutic strategy to combat COVID-19. *Iscience* 24:102254
- Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, Khovananth K, Mudd S, Mann N, Moresco EM, Beutler E, Beutler B (2008) The serine protease TMPRSS6 is required to sense iron deficiency. *Science* 320:1088–1092. <https://doi.org/10.1126/science.1157121>
- Enard D, Cai L, Gwennap C, Petrov DA (2016) Viruses are a dominant driver of protein adaptation in mammals. *Elife*. <https://doi.org/10.7554/eLife.12469.e12469>
- Ferrer-Admetlla A, Bosch E, Sikora M, Marques-Bonet T, Ramirez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, Casals F (2008) Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol* 181:1315–1322
- Forni D, Cagliani R, Pozzoli U, Colleoni M, Riva S, Biasin M, Filippi G, De Gioia L, Gnudi F, Comi GP, Bresolin N, Clerici M, Sironi M (2013) A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles. *Immunity* 38:1129–1141. <https://doi.org/10.1016/j.immuni.2013.04.008>
- Forni D, Cagliani R, Tresoldi C, Pozzoli U, De Gioia L, Filippi G, Riva S, Menozzi G, Colleoni M, Biasin M, Lo Caputo S, Mazzotta F, Comi GP, Bresolin N, Clerici M, Sironi M (2014) An evolutionary analysis of antigen processing and presentation across different timescales reveals pervasive selection. *PLoS Genet* 10:e1004189. <https://doi.org/10.1371/journal.pgen.1004189>
- Forni D, Cagliani R, Clerici M, Sironi M (2017) Molecular evolution of human coronavirus genomes. *Trends Microbiol* 25:35–48 (S0966-842X(16)30133-0)
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Riva S, Clerici M, Bresolin N, Sironi M (2009) Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med* 206:1395–1408. <https://doi.org/10.1084/jem.20082779>
- Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet* 7:e1002355. <https://doi.org/10.1371/journal.pgen.1002355>
- Gramberg T, Hofmann H, Möller P, Lalor PF, Marzi A, Geier M, Krumbiegel M, Winkler T, Kirchhoff F, Adams DH, Becker S, Münch J, Pöhlmann S (2005) LSECtin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology* 340:224–236 (S0042-6822(05)00367-3)
- GTEC Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEC (eGTEC) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration & Visualization—EBI, Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts: Laboratory, Data Analysis & Coordinating Center (LDACC): NIH program management: Biospecimen collection: Pathology: eQTL manuscript working group: Battle A, Brown CD, Engelhardt BE, Montgomery SB (2017) Genetic effects on gene expression across human tissues. *Nature* 550:204–213. <https://doi.org/10.1038/nature24277>
- Guindon S, Delsuc F, Dufayard JF, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol* 537:113–137. https://doi.org/10.1007/978-1-59745-251-9_6
- Guo H, Hu BJ, Yang XL, Zeng LP, Li B, Ouyang S, Shi ZL (2020) Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. *J Virol* 94:e00902–e920. <https://doi.org/10.1128/JVI.00902-20> (Print 2020 Sep 29)
- Haldane JBS (1932) The causes of evolution. Harper & Row, New York
- Harrison SC (2008) Viral membrane fusion. *Nat Struct Mol Biol* 15:690–698. <https://doi.org/10.1038/nsmb.1456>
- Hayashi H, Kubo Y, Izumida M, Takahashi E, Kido H, Sato K, Yamaya M, Nishimura H, Nakayama K, Matsuyama T (2018) Enterokinase enhances influenza A virus infection by activating trypsinogen in human cell lines. *Front Cell Infect Microbiol* 8:91. <https://doi.org/10.3389/fcimb.2018.00091>
- He M, Zhang Y, Li H, Liu M, Dong N, Wu Q (2020) A common CORIN variant in hypertension reduces corin intracellular trafficking by exposing an inhibitory N-terminus. *Biochem Biophys Res Commun* 530:35–41
- Hoffmann M, Kleine-Weber H, Pöhlmann S (2020a) A Multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 78:779–784.e5 (S1097-2765(20)30264-1)
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S (2020b) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181:271–280.e8 (S0092-8674(20)30229-4)
- Hoffmann M, Schroeder S, Kleine-Weber H, Müller MA, Drosten C, Pöhlmann S (2020c) Nafamostat mesylate blocks activation of SARS-CoV-2: new treatment option for COVID-19. *Antimicrob Agents Chemother* 64:e00754–e820. <https://doi.org/10.1128/AAC.00754-20> (Print 2020 May 21)
- Hollox EJ, Armour JA (2008) Directional and balancing selection in human beta-defensins. *BMC Evol Biol* 8:113. <https://doi.org/10.1186/1471-2148-8-113>
- Hooper JD, Clements JA, Quigley JP, Antalis TM (2001) Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *J Biol Chem* 276:857–860 (R000020200)
- Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, Justice AE, Banerjee N, Coignet MV, Yadav A, Leader JB, Marcketta A, Park DS, Lanche R, Maxwell E, Knight SC, Bai X, Guturu H, Sun D, Baltzell A, Kury FSP, Backman JD, Girschick AR, O'Dushlaine C, McCurdy SR, Partha R, Mansfield AJ, Turissini DA, Li AH, Zhang M, Mbatchou J, Watanabe K, Gurski L, McCarthy SE, Kang HM, Dobbyn L, Stahil E, Verma A, Sirugo G, Ritchie MD, Jones M, Balasubramanian S, Siminovich K, Salerno WJ, Shuldiner AR, Rader DJ, Mirshahi T, Locke AE, Marchini J, Overton JD, Carey DJ, Habegger L, Cantor MN, Rand KA, Hong EL, Reid JG, Ball CA, Baras A, Abecasis GR, Ferreira MA (2021) Common genetic variants identify targets for COVID-19 and individuals at high risk of severe disease. medRxiv. <https://doi.org/10.1101/2020.12.14.20248176>
- Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, Sharifi N, Erzurum S, Eng C, Cheng F (2020) New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2

- polymorphism analysis. *BMC Med* 18:216-020-01673-z. <https://doi.org/10.1186/s12916-020-01673-z>
- Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 86:958–962
- Irham LM, Chou WH, Calkins MJ, Adikusuma W, Hsieh SL, Chang WC (2020) Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. *Biochem Biophys Res Commun* 529:263–269 (**S0006-291X(20)31116-5**)
- Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, Wilson SJ, Schoggins JW, Rice CM, Yamashita M, Hatzioannou T, Bieniasz PD (2013) MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* 502:563–566. <https://doi.org/10.1038/nature12653>
- Kim TS, Heinlein C, Hackman RC, Nelson PS (2006) Phenotypic analysis of mice lacking the Tmprss2-encoded protease. *Mol Cell Biol* 26:965–975 (**26/3/965**)
- Kishimoto M, Uemura K, Sanaki T, Sato A, Hall WW, Kariwa H, Orba Y, Sawa H, Sasaki M (2021) TMPRSS11D and TMPRSS13 activate the SARS-CoV-2 spike protein. *Viruses* 13:384. <https://doi.org/10.3390/v13030384>.doi:384
- Knappe S, Wu F, Masikat MR, Morser J, Wu Q (2003) Functional analysis of the transmembrane domain and activation cleavage of human corin: design and characterization of a soluble corin. *J Biol Chem* 278:52363–52370 (**S0021-9258(20)75241-5**)
- Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208–1222 (**msi105**)
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23:1891–1901. <https://doi.org/10.1093/molbev/msl051>
- Kuhn RM, Haussler D, Kent WJ (2013) The UCSC genome browser and associated tools. *Brief Bioinform* 14:144–161. <https://doi.org/10.1093/bib/bbs038>
- Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO (2021) SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ Vacc* 6:28-021-00292-w. <https://doi.org/10.1038/s41541-021-00292-w>
- Laporte M, Naesens L (2017) Airway proteases: an emerging drug target for influenza and other respiratory virus infections. *Curr Opin Virol* 24:16–24 (**S1879-6257(16)30210-3**)
- Li F (2016) Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* 3:237–261. <https://doi.org/10.1146/annurev-virology-110615-042301>
- List K, Haudenschild CC, Szabo R, Chen W, Wahl SM, Swaim W, Engelholm LH, Behrendt N, Bugge TH (2002) Matriptase/MT-SP1 is required for postnatal survival, epidermal barrier function, hair follicle development, and thymic homeostasis. *Oncogene* 21:3765–3779. <https://doi.org/10.1038/sj.onc.1205502>
- Liu SY, Sanchez DJ, Aliyari R, Lu S, Cheng G (2012) Systematic identification of type I and type II interferon-induced antiviral factors. *Proc Natl Acad Sci USA* 109:4239–4244. <https://doi.org/10.1073/pnas.1114981109>; [10.1073/pnas.1114981109](https://doi.org/10.1073/pnas.1114981109)
- Liu X, Huuskonen S, Laitinen T, Redchuk T, Bogacheva M, Salokas K, Pöhner I, Öhman T, Tonduru AK, Hassinen A, Gawryski L, Keskitalo S, Vartiainen MK, Pietiäinen V, Poso A, Varjosalo M (2021) SARS-CoV-2-host proteome interactions for antiviral drug discovery. *Mol Syst Biol* 17:e10396. <https://doi.org/10.15252/msb.202110396>
- Maginnis MS (2018) Virus–receptor interactions: the key to cellular invasion. *J Mol Biol* 430:2590–2611 (**S0022-2836(18)30630-2**)
- Marzi A, Gramberg T, Simmons G, Möller P, Rennekamp AJ, Krumbiegel M, Geier M, Eisemann J, Turza N, Saunier B, Steinkasserer A, Becker S, Bates P, Hofmann H, Pöhlmann S (2004) DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *J Virol* 78:12090–12095 (**78/21/12090**)
- Matsuyama S, Nao N, Shirato K, Kawase M, Saito S, Takayama I, Nagata N, Sekizuka T, Katoh H, Kato F, Sakata M, Tahara M, Kutsuna S, Ohmagari N, Kuroda M, Suzuki T, Kageyama T, Takeda M (2020) Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc Natl Acad Sci USA* 117:7001–7003. <https://doi.org/10.1073/pnas.2002589117>
- Millet JK, Whittaker GR (2015) Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res* 202:120–134 (**S0168-1702(14)00496-1**)
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30:1196–1205. <https://doi.org/10.1093/molbev/mst030>
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Nygaard RM, Golden JW, Schiff LA (2012) Impact of host proteases on reovirus infection in the respiratory tract. *J Virol* 86:1238–1243. <https://doi.org/10.1128/JVI.06429-11>
- Ortiz M, Kaessmann H, Zhang K, Bashirova A, Carrington M, Quintana-Murci L, Telenti A (2008) The evolutionary history of the CD209 (DC-SIGN) family in humans and non-human primates. *Genes Immun* 9:483–492. <https://doi.org/10.1038/gene.2008.40>
- Peitsch C, Klenk HD, Garten W, Böttcher-Friebertshäuser E (2014) Activation of influenza A viruses by host proteases from swine airway epithelium. *J Virol* 88:282–291. <https://doi.org/10.1128/JVI.01635-13>
- Peng H, Zhang Q, Cai X, Liu Y, Ding J, Tian H, Chao X, Shen H, Jiang L, Jin J, Zhang Y (2015) Association between high serum soluble corin and hypertension: a cross-sectional study in a general population of China. *Am J Hypertens* 28:1141–1149. <https://doi.org/10.1093/ajh/hpv002>
- Piva F, Sabanovic B, Cecati M, Giulietti M (2021) Expression and co-expression analyses of TMPRSS2, a key element in COVID-19. *Eur J Clin Microbiol Infect Dis* 40:451–455. <https://doi.org/10.1007/s10096-020-04089-y>
- Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679 (**bti079**)
- Qiao Y, Wang XM, Mannan R, Pitchiaya S, Zhang Y, Wotring JW, Xiao L, Robinson DR, Wu YM, Tien JC, Cao X, Simko SA, Apel IJ, Bawa P, Kregel S, Narayanan SP, Raskind G, Ellison SJ, Parolia A, Zelenka-Wang S, McMurry L, Su F, Wang R, Cheng Y, Deleka AD, Mei Z, Pretto CD, Wang S, Mehra R, Sexton JZ, Chinnaiyan AM (2020) Targeting transcriptional regulation of SARS-CoV-2 entry factors ACE2 and TMPRSS2. *Proc Natl Acad Sci USA* 118:e2021450118. <https://doi.org/10.1073/pnas.2021450118>
- Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Dijkman R, Muth D, Demmers JA, Zaki A, Fouchier RA, Thiel V, Drosten C, Rottier PJ, Osterhaus AD, Bosch BJ, Haagmans BL (2013) Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495:251–254. <https://doi.org/10.1038/nature12005>
- Sakai K, Ami Y, Tahara M, Kubota T, Anraku M, Abe M, Nakajima N, Sekizuka T, Shirato K, Suzuki Y, Ainai A, Nakatsu Y, Kanou K, Nakamura K, Suzuki T, Komase K, Nobusawa E, Maenaka K, Kuroda M, Hasegawa H, Kawaoka Y, Tashiro M, Takeda M (2014) The host protease TMPRSS2 plays a major role in vivo replication of emerging H7N9 and seasonal influenza viruses. *J Virol* 88:5608–5616. <https://doi.org/10.1128/JVI.03677-13>

- Sasaki M, Itakura Y, Kishimoto M, Tabata K, Uemura K, Ito N, Sugiyama M, Wastika CE, Orba Y, Sawa H (2021) Host serine proteases TMPRSS2 and TMPRSS11D mediate proteolytic activation and trypsin-independent infection in group A rotaviruses. *J Virol* 95:e00398–e421. <https://doi.org/10.1128/JVI.00398-21>
- Schrödinger L (2017) The PyMOL molecular graphics system, version 2.0 Schrödinger, LLC. Google Scholar there is no corresponding record for this reference
- Scott HS, Kudoh J, Wattenhofer M, Shibuya K, Berry A, Chrast R, Guipponi M, Wang J, Kawasaki K, Asakawa S, Minoshima S, Younus F, Mehdi SQ, Radhakrishna U, Papasavvas MP, Gehrig C, Rossier C, Korostishevsky M, Gal A, Shimizu N, Bonnet-Tamir B, Antonarakis SE (2001) Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. *Nat Genet* 27:59–63. <https://doi.org/10.1038/83768>
- Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F (2020) Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci USA* 117:11727–11734. <https://doi.org/10.1073/pnas.2003138117>
- Shirato K, Kawase M, Matsuyama S (2013) Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* 87:12552–12561. <https://doi.org/10.1128/JVI.01890-13>
- Shirogane Y, Takeda M, Iwasaki M, Ishiguro N, Takeuchi H, Nakatsu Y, Tahara M, Kikuta H, Yanagi Y (2008) Efficient multiplication of human metapneumovirus in Vero cells expressing the transmembrane serine protease TMPRSS2. *J Virol* 82:8942–8946. <https://doi.org/10.1128/JVI.00676-08>
- Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA* 102:11876–11881 (0505577102)
- Sironi M, Cagliani R, Forni D, Clerici M (2015) Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nat Rev Genet* 16:224–236. <https://doi.org/10.1038/nrg3905>
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32:1342–1353. <https://doi.org/10.1093/molbev/msv022>
- Szabo R, Bugge TH (2008) Type II transmembrane serine proteases in development and disease. *Int J Biochem Cell Biol* 40:1297–1316. <https://doi.org/10.1016/j.biocel.2007.11.013>
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S (2020) Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antivir Res* 178:104792 (S0166-3542(20)30206-0)
- Tarnow C, Engels G, Arendt A, Schwalm F, Sediri H, Preuss A, Nelson PS, Garten W, Klenk HD, Gabriel G, Böttcher-Friebertshäuser E (2014) TMPRSS2 is a host factor that is essential for pneumotropism and pathogenicity of H7N9 influenza A virus in mice. *J Virol* 88:4744–4751. <https://doi.org/10.1128/JVI.03799-13>
- Vankadari N, Wilce JA (2020) Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerg Microbes Infect* 9:601–604. <https://doi.org/10.1080/22221751.2020.1739565>
- Vargas-Alarcón G, Posadas-Sánchez R, Ramírez-Bello J (2020) Variability in genes related to SARS-CoV-2 entry into host cells (ACE2, TMPRSS2, TMPRSS11A, ELANE, and CTSL) and its potential use in association studies. *Life Sci* 260:118313 (S0024-3205(20)31065-1)
- Verhelst J, Hulpiau P, Saelens X (2013) Mx proteins: antiviral gatekeepers that restrain the uninvited. *Microbiol Mol Biol Rev* 77:551–566. <https://doi.org/10.1128/MMBR.00024-13>
- Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 19:327–335. <https://doi.org/10.1101/gr.073585.107;10.1101/gr.073585.107>
- Vishnubhotla R, Vankadari N, Ketavarapu V, Amanchy R, Avanthi S, Bale G, Reddy DN, Sasikala M (2020) Genetic variants in TMPRSS2 and Structure of SARS-CoV-2 spike glycoprotein and TMPRSS2 complex. *BioRx*. <https://doi.org/10.1101/2020.06.30.179663>
- Wang K, Chen W, Zhang Z, Deng Y, Lian JQ, Du P, Wei D, Zhang Y, Sun XX, Gong L, Yang X, He L, Zhang L, Yang Z, Geng JJ, Chen R, Zhang H, Wang B, Zhu YM, Nan G, Jiang JL, Li L, Wu J, Lin P, Huang W, Xie L, Zheng ZH, Zhang K, Miao JL, Cui HY, Huang M, Zhang J, Fu L, Yang XM, Zhao Z, Sun S, Gu H, Wang Z, Wang CF, Lu Y, Liu YY, Wang QY, Bian H, Zhu P, Chen ZN (2020a) CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. *Signal Transduct Target Ther* 5:283. <https://doi.org/10.1038/s41392-020-00426-x>
- Wang W, Zhao H, Han GZ (2020b) Host–virus arms races drive elevated adaptive evolution in viral receptors. *J Virol* 94:e00684–e720. <https://doi.org/10.1128/JVI.00684-20> (Print 2020 Jul 30)
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276
- Wernersson R, Pedersen AG (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res* 31:3537–3539
- Wiuf C, Zhao K, Innan H, Nordborg M (2004) The probability and chromosomal extent of trans-specific polymorphism. *Genetics* 168:2363–2372. <https://doi.org/10.1534/genetics.104.029488>
- Yan W, Wu F, Morser J, Wu Q (2000) Corin, a transmembrane cardiac serine protease, acts as a pro-atrial natriuretic peptide-converting enzyme. *Proc Natl Acad Sci USA* 97:8525–8529 (150149097)
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang ZY, Huang Y, Ganesh L, Leung K, Kong WP, Schwartz O, Subbarao K, Nabel GJ (2004) pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J Virol* 78:5642–5650 (78/11/5642)
- Yeager CL, Ashmun RA, Williams RK, Cardellicchio CB, Shapiro LH, Look AT, Holmes KV (1992) Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* 357:420–422 (BF357420a0)
- Zang R, Gomez Castro MF, McCune BT, Zeng Q, Rothlauf PW, Sonnek NM, Liu Z, Brulois KF, Wang X, Greenberg HB, Diamond MS, Ciorba MA, Whelan SPJ, Ding S (2020) TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci Immunol* 5:eabc582. <https://doi.org/10.1126/sciimmunol.abc3582>
- Zarubin A, Stepanov V, Markov A, Kolesnikov N, Marusin A, Khitrinskaya I, Swarovskaya M, Litvinov S, Ekomasova N, Dzhaubermzov M, Maksimova N, Sukhomyasova A, Shtygasheva O, Khuznutdinova E, Radzhabov M, Kharkov V (2020) Structural variability, expression profile, and pharmacogenetic properties of TMPRSS2 gene as a potential target for COVID-19 therapy. *Genes (Basel)* 12:19. <https://doi.org/10.3390/genes12010019>
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479. <https://doi.org/10.1093/molbev/msi237>
- Zhang H, Mo X, Qian Q, Zhou Z, Zhu Z, HuangFu X, Xu T, Wang A, Guo Z, Lei S, Zhang Y (2019a) Associations between potentially

- functional CORIN SNPs and serum corin levels in the Chinese Han population. *BMC Genet* 20:99-019-0802-4. <https://doi.org/10.1186/s12863-019-0802-4>
- Zhang Z, Zhu Z, Chen W, Cai Z, Xu B, Tan Z, Wu A, Ge X, Guo X, Tan Z, Xia Z, Zhu H, Jiang T, Peng Y (2019b) Cell membrane proteins with high N-glycosylation, high expression and multiple interaction partners are preferred by mammalian viruses as receptors. *Bioinformatics* 35:723–728. <https://doi.org/10.1093/bioinformatics/bty694>
- Zhang C, Zhang Y, Zhang S, Wang Z, Sun S, Liu M, Chen Y, Dong N, Wu Q (2020) Intracellular autoactivation of TMPRSS11A, an airway epithelial transmembrane serine protease. *J Biol Chem* 295:12686–12696. <https://doi.org/10.1074/jbc.RA120.014525>
- Zhou Y, Wu Q (2014) Corin in natriuretic peptide processing and hypertension. *Curr Hypertens Rep* 16:415-013-0412-7. <https://doi.org/10.1007/s11906-013-0415-7>
- Zmora P, Hoffmann M, Kollmus H, Moldenhauer AS, Danov O, Braun A, Winkler M, Schughart K, Pöhlmann S (2018) TMPRSS11A activates the influenza A virus hemagglutinin and the MERS coronavirus spike protein and is insensitive against blockade by HAI-1. *J Biol Chem* 293:13863–13873. <https://doi.org/10.1074/jbc.RA118.001273>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.