Gene of the month: **TMPRSS2** (transmembrane serine protease 2)

Michelle Thunders 📩, Brett Delahunt 📩

**ABSTRACT**

Transmembrane serine protease 2 is encoded by the **TMPRSS2** gene. The gene is widely conserved and has two isoforms, both being autocatalytically activated from the inactive zymogen form. A fusion gene between the **TMPRSS2** gene and **ERG** (erythroblast-specific related gene), an oncogenic transcription factor, is the most common chromosomal aberration detected in prostate cancer, responsible for driving carcinogenesis. The other key role of **TMPRSS2** is in priming the viral spike protein which facilitates viral entry essential for viral infectivity. The protease activates a diverse range of viruses. Both SARS-CoV and SARS-CoV-2 (COVID-19) use angiotensin-converting enzyme 2 (ACE2) and **TMPRSS2** to facilitate entry to cells, but with SARS-CoV-2 human-to-human transmission is much higher than SARS-CoV. As **TMPRSS2** is expressed outside of the lung, and can therefore contribute to extrapulmonary spread of viruses, it warrants further exploration as a potential target for limiting viral spread and infectivity.

**INTRODUCTION**

The transmembrane serine protease 2 (**TMPRSS2**) is a tool used by a diverse range of viruses, including SARS-CoV-2, to enter human cells, and is also manipulated in tumourigenesis to drive androgen-responsive prostate cancer. The 70kDa serine protease is made as a precursor protein (zymogen) which undergoes autoproteolytic activation. In humans, **TMPRSS2** mutations have not been associated with any inherited pathologies, and similarly **Tmprss2** knockout mice are asymptomatic. This suggests functional redundancy involving one or more of the type II transmembrane serine protease family (**TTSP**). An alternative is that the protein could potentially have a highly specialised non-vital function, becoming apparent only in the context of stress, disease or another systemic problem. In terms of normal function, **TMPRSS2** has been associated with physiological and pathological processes such as digestion, tissue remodeling, blood coagulation, fertility, inflammatory responses, tumour cell invasion, apoptosis and pain. Expression of **TMPRSS2** is developmentally regulated and increases with ageing. There is strong expression in fetal brain, but little in adult brain. Expression is high in adult lung tissue being similar in males and females, and low in fetal lung tissue. **TMPRSS2** shows greater expression in bronchial epithelial cells when compared with surfactant producing type II alveolar cells and alveolar macrophages. There is no expression in type I alveolar cells that form the respiratory surface. Expression is also found in heart, liver and gastrointestinal tract. Recently, both **TMPRSS2** and angiotensin-converting enzyme 2 (ACE2) (the receptor for SARS-CoV-2) have been found to be expressed in human corneal epithelium suggesting ocular surface cells could also be potential viral entry points. **TMPRSS2** is also highly expressed in the epithelium of the human prostate gland and regulated by androgenic hormones. In the normal prostate, it contributes to proteolytic cascades that result in the activation of prostate-specific antigen, the protease with enzymatic reaction in seminal fluid analogous to fibrinolytic and blood coagulation. **TMPRSS2** was reported to regulate epithelial sodium channel activity in vitro, implying a possible role in epithelial sodium homeostasis within the prostate gland. The targeting of **TMPRSS2** to the apical plasma membrane and its presence in prostasomes suggests that it may also have a function in enhancing male reproductive capability.

**TMPRSS2 structure**

**TMPRSS2** is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, rat, chicken, zebrafish, *Caenorhabditis elegans* and frog. It is homologous to, but different from, the human enteropeptidase gene. Residing at 21q22.3, it has 15 exons and an open reading frame of 492 amino acids. It contains a type II transmembrane domain, an LDL receptor class A (LDLRA) domain (forms binding site for calcium), a scavenger receptor cysteine-rich (SRCR) domain (involved in binding to other cell surface or extracellular molecules) and a serine protease domain (S1 family that cleaves at arginine or lysine residues). Additionally, **TMPRSS2** harbours androgen-responsive elements in its 5’ UTR. Testosterone and dihydrotestosterone regulate transcription of the **TMPRSS2** gene through stimulation of the androgen receptor.

A disulphide bond links the catalytic domain and the membrane bound portion of the peptide; the protease domain is, however, released into the extracellular space. The N terminal is in the cytoplasm. Next to this is a hydrophobic transmembrane domain, a stem region and a protease domain containing a catalytic amino acid triad essential for proteolytic activity comprising histidine, aspartic acid and serine (figure 1). The intracellular domain may interact with cytoskeletal components and signalling molecules and could be important for the correct intracellular trafficking of the peptide. The stem region (the LDLRA and SRCR domains) might mediate protein–protein interactions and binding. The catalytic domain cleaves cell-membrane receptors, growth factors, cytokines and components of
Figure 1  Schematic of TMPRSS2 492 amino acid peptide (isoform2). TM is the transmembrane domain, LDLRA is LDL receptor class A domain, SRCR is the scavenger receptor cysteine rich domain. H, D and S are the catalytic amino acid triad histidine, aspartic acid and serine, essential for proteolytic activity.

the extracellular matrix. The 32 kDa serine protease domain undergoes autolysis, secretion into the epithelia and interacts with cell surface proteins, the extracellular matrix and proteins of neighbouring cells. The SARS-CoV binding site is formed by the insertion of 26 amino acid residues from the protease domain and the remainder of the protein. This permits the protease domain to undergo conformational change necessary to transition into its active state. The alternative isoforms suggest that there could be a membrane bound and circulating form. There is a single N-terminal fragment for isoform 2, while two fragments are detected for isoform 1, indicating a possible cleavage specificity and also possible altered intracellular localisation. The majority of mature protease are membrane bound after autolysis, but there are some present in the extracellular matrix. Isoform 1 co-localises with viral haemagglutinin cleaving and activating the viral protein; it also activates the SARS-CoV spike protein for cathepsin L-independent entry into target cells.

Recently, there has been interest in identification of unique, but prevalent, variants in TMPRSS2 as potential contributing factors influencing susceptibility and severity to SARS-CoV-2. It is possible that such TMPRSS2 variants can affect the efficiency of the protease. Hou et al identified six germline variants in the TMPRSS2 coding region which were also identified as somatic mutations in The Cancer Genome Atlas (TCGA) and Catalogue Of Somatic Mutations In Cancer (COSMIC) databases; these are pVal160Met, pGly181Arg, pArg240Cys, pGly259Ser, pPro335Leu and pGly432Ala. All populations carry the pVal160Met variant with high frequency (rs12329760). The pAsp435Tyr was only found in European populations. Other previously reported polymorphisms of interest have been in the non-coding sequence, including SNP rs2070788 in the upstream regulatory region. This variant has been associated with increased TMPRSS2 expression and higher risk for severe illness with the 2009 pandemic H1N1 influenza. A logical target for treatment could therefore be a TMPRSS2 inhibitor. Experiments using camostat mesylate (a clinically proven serine protease inhibitor) showed significant infection reduction, though not complete cessation of infection, via blocking SARS-CoV-2 infection of lung cells. SNP rs8134378 has been reported to reduce binding and activation by the androgen receptor. TMPRSS2 as a protease

TMPRSS2 is unusual among the type II TTSP family regarding its formation of multiple complexes. The majority of the TMPRSS2 released from complexes is in the latent (zymogen) form, which would not be expected to form stable complexes with endogenous protease inhibitors. The TMPRSS2 complexes are, therefore, apparently not a product of conventional mechanisms governing the zymogen activation and inhibition of active serine proteases. Furthermore, TMPRSS2 has 22 cysteine residues, although the serine protease domain in TMPRSS2 contains eight cysteine residues in conserved positions, the pairing counterpart cysteine for Cys140 is absent. This unpaired cysteine in the serine protease domain may be the structural element, which facilitates the formation of disulfide-linked complexes. It will be of great interest to determine whether the extensive formation of TMPRSS2 complexes represents a novel mechanism by which the protease is regulated and to explore the functional significance of complex formation.

TMPRSS2 and viral entry

Viruses found to use this protein for cell entry include influenza virus and the human coronaviruses HCoV-229E, MERS-CoV, SARS-CoV and SARS-CoV-2 (COVID-19 virus). CoV are extracellular particles in an envelope with protruding spike proteins. In order for them to infect a cell, the viral spike protein needs to engage with the cell surface protease, TMPRSS2. The protease engages in a complex with ACE2, the co-receptor for SARS-CoV. TMPRSS2 cleaves ACE2 at Arginine 697–716, which enhances viral uptake. Variants in ACE2, resulting in loss of the TMPRSS2 cleavage site, could also contribute to milder symptoms and reduced infectivity. The SARS-CoV spike protein is cleaved at Arginine 700–716, which allows protease activity to be initiated. This cleavage site is conserved in SARS-CoV-2, which therefore requires the presence of TMPRSS2 to enter cells.
to bind to the host cell receptor. Following binding, there needs to be proteolytic cleavage of the bound spike protein for fusion between viral and host cell membranes to occur. Exposure of the fusion peptide allows viral entry (S mediated cell–cell fusion). This process involves the conformational change of the spike protein until the viral mRNA genome is deposited into the host cell cytoplasm. These conformational changes and fusion require cellular proteases, the availability of which is a rate-limiting step in CoV entry. There are several proteases crucial to coronavirus viral entry and they are found at different subcellular locations. In particular, the type II transmembrane serine proteases are anchored into the plasma membranes. If cell surface proteases are present, this enables the spike proteins to be cleaved and for viral fusion to occur near the plasma membrane. If cell surface proteases are absent, then late viral entry can occur through endocytosis triggered by endosomal proteases, such as cathepsin L, as in the case of MERS-CoV.31 The recent development of the Viral Integrated Structural Evolution Dynamic Database (VISTEDD) should help further elucidate the viral proteome dynamics and their interaction with host cell receptors.23

Respiratory viruses with monobasic cleavage sites typically have an expression profile confined to the aerodigestive tract. Viral replication is limited to these organs as the cleavage site is cleaved by few cellular proteases, restricting viral growth to these limited areas. Viruses with multibasic cleavage sites are cleaved by several common cellular proteases. These viruses, therefore, have the potential to grow systemically in the host, being activated by ubiquitously expressed proprotein convertases such as furin. This facilitates systemic spread of the disease on a massive scale. Highly pathogenic avian influenza viruses have multibasic cleavage sites cut by ubiquitously expressed proteases.28 It has been suggested that the origin of SARS-CoV-2 is zoonotic in nature27; however, all CoV in bats and pangolins have monobasic cleavage sites. How a multibasic cleavage site occurred in SARS-CoV-2 is an interesting question, particularly in the context of potential viral adaptation in response to overexpression of protease inhibitors. Interestingly, a study of Tmprss2 knockout mice found that influenza strain H3N2 was avirulent but became lethal after 10 passages in mice.28 The passaged virus carried an N-glycosylation mutation at the base of the haemagglutinin stalk region. Loss of this glycan could alter accessibility to the cleavage loop and provide access to alternative host proteases. In cell culture experiments, overexpression of Tmprss2 resulted in haemagglutinin cleavage and multicycle replication of influenza B virus.29

SARS-CoV-2 S protein has two functional domains S1 and S2; S1 is the receptor binding domain and S2 contains functional elements involved in membrane fusion. Multiple cleavage sites are present; one of these is at the S1/S2 boundary and another within S2. The S1/S2 cleavage site contains multiple arginine residues (multibasic) previously unknown in animal coronaviruses.30 Hofmann et al31 report that cleavage of the spike protein at the S1/S2 site is at an exposed multibasic loop, mediated by furin. This cleavage is required for viral entry into human lung cells and is also a critical step in cell–cell fusion. The formation of multinucleated giant cells is therefore S-protein dependent, being enhanced by Tmprss2. Giant cell formation also has implications for virus cell–cell spread, anoikis/alternative accelerated cell death and evasion of antibody neutralisation.32 Experiments deleting the multibasic motif resulted in a spike protein that could not induce giant cell formation, whereas adding another arginine to the S1/S2 site increased the numbers of giant cells. This suggests that augmented cell–cell spread and pathogenicity could be attained in viral variants with optimised S1/S2 sites.31 It is likely that a combination of drugs targeting both furin and Tmprss2 is required for effective viral control. MI-1851 is a potent synthetic furin inhibitor, its use in conjunction with a drug such as camostat mesylate is currently being explored for treatment of SARS-CoV-2.33

Experiments in transfected cells34 have shown that cleavage of SARS S is frequently incomplete, leaving uncleaved/cleavage intermediates in the cellular supernatant, which may interfere with antibody-mediated neutralisation. The pseudotypes generated in the presence of Tmprss2 retain infectivity and are protected against neutralisation by antibodies. This protection is due to the presence of shed SARS S in the virion preparation, where the S protein fragments function as antibody decoys. Removal of shed SARS S from these preparations removed resistance to neutralisation, adding shed SARS S provided neutralisation resistance. The conserved receptor-binding domain in SARS-CoV-2, as well as the S protein could be important targets for neutralising antibodies, as cleavage of this fragment as an antibody decay function is a possibility.

Tmprss2 and prostate cancer

In prostate cancer cells, Tmprss2 is strongly upregulated in response to androgens. Tmprss2 is expressed on the luminal side of prostate epithelium and this is greater in cancer than in non-cancerous tissue. Fusion of the S’UTR of the Tmprss2 gene and ERG (erythroblast-specific-related gene), an oncogenic transcription factor, is the most common chromosomal aberration in prostate cancer and explains the overexpression of the ERG proto-oncogene seen in malignant cells.34 ERG has been estimated to be overexpressed in 40%-50% of primary prostate cancers, secondary to the androgen-dependent genomic rearrangement on 21q22.2–22.3 producing the fusion. The mutation occurs through chromosomal translocation or intergenic deletion, with both genes on the same arm of chromosome 21, and results in overexpression of chimeric mRNA of ERG. The prevalence of the fusion varies by ethnicity; being highest in European, then Asian and lowest in African (49%, 27%, and 25%, respectively). This suggests alternative mechanisms might explain the origin of prostate cancer in different populations.35

Tmprss2-ERG fusions occur early in prostate carcinogenesis. The coding sequence of Tmprss2 is not involved in the gene fusion. Consequently, there is no resultant recombinant protein for the Tmprss2-ETS gene fusion and the promoterless copy of Tmprss2 is silenced. This results in reduced expression of Tmprss2 mRNA in those patients with prostate cancer with the gene fusion. Expression of Tmprss2 is regulated by androgen11 being highly expressed in normal and neoplastic prostate tissue.36 Tmprss2 may contribute to prostate carcinogenesis, not only by the increased expression, but also through aberrant subcellular localisation attributable to the loss of epithelial polarity in the transformed cells. This latter abnormality may allow the protease to inappropriately gain access to and/or activate some cancer-promoting substrates, which its normal subcellular localisation would preclude under normal physiological conditions.

There is impairment of apoptosis in Tmprss2-ERG positive cancer cells,33 possibly due to disruption of the intracellular death domain or decoy receptors. Men with higher genetically determined transcriptional activity of the androgen receptor have higher risk of Tmprss2-ERG fusion-positive prostate cancer but not fusion-negative prostate cancer. Tumours with the fusion also have higher insulin/insulin-like growth factor signalling and so may modify how hormonal factors, such as obesity, influence risk of metastasis. An androgen-regulated gene can promote prostate cancer tumour growth and metastasis via matrix activation and degradation of ECM laminin beta 1 and nidogen-1. This means
that it is implicated in the acceleration of prostate cancer progression up to an aggressive growth phase. 38

CONCLUSION
Arguably, the most interesting genes are those whereby a knockout of function appears to have no deleterious effect. Their persistence throughout evolution is suggestive that they are not redundant but vital to the complexity and adaptability of human homeostatic function. TMPRSS2 encodes an exceptional protease. A further understanding of how this is used by viruses is vitally important for the development of antiviral treatments and the understanding of viral spread for current and future pandemics.

Take home messages

► TMPRSS2 encodes a widely expressed protease found anchored into cell membranes.
► TMPRSS2 is used by a range of viruses, including SARS-CoV-2 (COVID-19), to enter host cells.
► Understanding how TMPRSS2 contributes to viral infectivity is important in the development of effective antiviral treatments and to reduce viral spread in this, and future, pandemics.

Handling editor Tahir S Pillay

Contributors MT and BD facilitated the conception and design of the project, drafted the manuscript and revised it critically for intellectual content.

Funding
The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

This article is made freely available for use in accordance with BMJ's website handling editor. This article is made freely available for use in accordance with BMJ's website

ORCID IDs
Michelle Thunders http://orcid.org/0000-0001-6722-7072
Brett Delahunt http://orcid.org/0000-0002-5398-0300

RECOMMENDATIONS

Conflict of interest None declared.