Abstract. Severe acute respiratory syndrome (SARS) coronavirus-2 (SARS-CoV2) is the cause of a new disease (COVID-19) which has evolved into a pandemic during the first half of 2020. Older age, male sex and certain underlying diseases, including cancer, appear to significantly increase the risk for severe COVID-19. SARS-CoV-2 infection of host cells is facilitated by the angiotensin-converting enzyme 2 (ACE-2), and by transmembrane protease serine 2 (TMPRSS2) and other host cell proteases such as cathepsin L (CTSL). With the exception of ACE-2, a systematic analysis of these two other SARS-CoV2 infection mediators in malignancies is lacking. Here, we analysed genetic alteration, RNA expression, and DNA methylation of TMPRSS2 and CTSL across a wide spectrum of tumors and controls. TMPRSS2 was overexpressed in cervical squamous cell carcinoma and endocervical adenocarcinoma, colon adenocarcinoma, prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), uterine corpus endometrial carcinoma and uterine carcinosarcoma, with PRAD and READ exhibiting the highest expression of all cancers. CTSL was upregulated in lymphoid neoplasm diffuse large B-cell lymphoma, oesophageal carcinoma, glioblastoma multiforme, head and neck squamous cell carcinoma, lower grade glioma, pancreatic adenocarcinoma, skin cutaneous melanoma, stomach adenocarcinoma, and thymoma. Hypo-methylation of both genes was evident in most cases where they have been highly upregulated. We have expanded on our observations by including data relating to mutations and copy number alterations at pan-cancer level. The novel hypotheses that are stemming out of these data need to be further investigated and validated in large clinical studies.

Introduction

Severe acute respiratory syndrome (SARS) coronavirus-2 (SARS-CoV2) is the cause of COVID-19, which was first identified at the end of 2019 and has evolved into a pandemic during the following months (1). Based on increasing data, older age and male sex predispose to severe COVID-19, whilst a number of underlying diseases/conditions are also directly related with significantly higher risk for adverse clinical outcomes from COVID-19 (1). The latter include diabetes, hypertension, obesity, immunosuppression, asthma, and chronic obstructive pulmonary disease (1).

Entry of SARS-CoV2 into its host cells is facilitated by its spike proteins which bind to the angiotensin-converting enzyme 2 (ACE-2) (2). Moreover, the spike viral proteins are primed by the transmembrane protease serine 2 (TMPRSS2) (2). As such, SARS-CoV2 infection of host cells is
Figure 1. SARS-CoV2 attaches to the ACE-2 receptor and it can be taken into the host cell in endosomes where it can be cleaved and activated by CTSL proteases. Moreover, TMPRSS2 is co-expressed with ACE-2 on the cell membrane and it can prime the spike viral proteins, thus activating the fusion of the virus with the membrane lipid layer and facilitating the uptake into the host cell. ACE-2, angiotensin-converting enzyme 2; CTSL, cathepsin L; TMPRSS2, transmembrane protease serine 2.

Figure 2. RNA expression aberration of CTSL in tumors. (A) the gene expression profile across all tumor samples and paired normal tissues, each dot represents expression of samples. (B) The gene expression profile across all tumor samples and paired normal tissues. The height of bar represents the median expression of certain tumor type or normal tissue. CTSL, cathepsin L.

Figure 3. RNA expression aberration of TMPRSS2 in tumors. (A) the gene expression profile across all tumor samples and paired normal tissues, each dot represents expression of samples. (B) The gene expression profile across all tumor samples and paired normal tissues. The height of bar represents the median expression of certain tumor type or normal tissue. TMPRSS2, transmembrane protease serine 2.
facilitated by the cleavage of the spike proteins by TMPRSS2 and host cell proteases, such as cathepsin L (CTSL) (Fig. 1) (3).

Cancer has been identified as a risk factor for severe COVID-19, as many patients can be immunocompromised (4). In a very recent robust systematic analysis, Chai et al curated a pan-cancer analysis of ACE-2 detailing the expression and mutations across a wide spectrum of tumors (5). Following the same approach with this elegant study, here we expanded on these excellent observations by curating data for the other two mediators of SARS-CoV-2 infection, namely CTSL and TMPRSS2.

Materials and methods

Bioinformatic analysis. TMPRSS2 and CTSL were validated in The Cancer Genome Atlas (TCGA), GEPIA (http://gepia.cancer-pku.cn/), UALCAN (http://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl). The pan-cancer cohort of TCGA was downloaded through eBioPortal (https://www.ebiportal.org/). The datasets used for the two genes were: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma;
Results and Discussion

In TCGA datasets for all cancers, CTSL was upregulated in DLBC, ESCA, GBM, HNSC, LGG, PADD, SKCM, STAD and THYM, while it has lower expression than normal in COAD, LAML and READ (Fig. 2A and B). On the other hand, TMPRSS2 was upregulated in CESC, COAD, KICH, PRAD, READ, UCEC and UCS, with PRAD and READ exhibiting the highest expression of all cancers (Fig. 3A and B).

Subsequently, the expression of these two genes from GEPIA were correlated with the methylation status using the Ualcan database. For most of the cancers there was a strong correlation of the gene expression with its promoter methylation status in agreement with the findings for ACE-2 (5). CTSL in COAD and READ has a higher beta value; hence, lower expression of the gene, while in STAD, the lower methylation of the promoter, leads to higher
gene expression (Fig. 4A and B). Five cancers were found to have lower methylation than normal for the CTSL, and seven cancers for the TMPRSS2. In LUSC and BRCA, the very high methylation rate of the TMPRSS2 gene leads to a very low expression level (Fig. 5A and B), while the very low methylation rate in PRAD and READ lead to very high numbers of gene transcripts.

Moreover, using the Human Protein Atlas analysis, of the consensus transcript expression levels that combines the Human Protein Atlas (HPA), Genotype-Tissue Expression; (GTEx) and the Functional Annotation of Mammalian Genomes 5 (FANTOM5) datasets, the expression of these two genes in organ/tissue samples were examined. CTSL is highly expressed in the placenta, adipose, appendix and lung tissue (Fig. 6). TMPRSS2 is highly expressed in prostate [indeed, TMPRSS2 and PCA are utilized as biomarkers for prostate cancer (6)], stomach, colon and small intestine tissues (Fig. 7).

Furthermore, using the cBioportal pan-cancer panel, the region and the types of mutations were identified which these two genes have in all the examined cancer types (Figs. 8A and 9A). Most of the CTSL mutations are lying on the peptidase region and are mostly found in CESC, ESCA, Mature B-cell Neoplasms, Melanoma and COAD (Fig. 8B). Of note, in most of the cancers the majority of the patients had deletions and partly some gains and amplifications (Fig. 8C). TMPRSS2 mutations were lying across the whole gene region and mostly consist of gene fusions (TMPRSS2-ERG) in prostate adenocarcinoma (Fig. 9B and C).
Finally, examination of the overall survival (OS) of all cancers, showed that CTSL low expression in KIRC (P=0.0001) has poor prediction for the patients, while low expression of CTSL in LUSC (P=0.0077) had better prognosis. High expression of TMPRSS2 in BRCA, SARC and UM had poor prediction for the patients, while gene expression was not statistically relevant for the patients of the other types of cancers (Figs. S1-S4).

Of note, neither of the two proteins were differentially regulated in LUAD; a comorbidity of severe COVID-19 contrary to ACE-2 (7). TMPRSS2 is significantly upregulated in prostate cancer, where it harbors gene fusion events, as documented in this study. While preparing this report, another interesting study was published which corroborated the expression of TMPRSS2 in the prostate. In this study, the authors have put forward the question of whether the TMPRSS2 increased expression in prostate is involved with the documented sexual dimorphism of COVID-19 (8). This is an exciting hypothesis that needs to be studied further given that the clinical data indicate that male sex is among the risk factors for adverse COVID-19 related outcomes.

In our analysis we also demonstrate that the pancreas is riddled with deep deletions for TMPRSS2 where ACE-2 is co-expressed. Interestingly, Liu et al (9) have recently shown that ACE-2 expression in the pancreas may lead to pancreatic damage following SARS-CoV-2 infection. For CTSL, higher expression was noted in the human placenta where ACE-2 is also co-expressed (10). This finding warrants further investigation into the role of these viral entry mediators in placentation, vertical transmission of SARS-CoV-2 to the fetus, as well as

---

Figure 9. (A) TMPRSS2 mutations were distributed across the TMPRSS2 gene (cBioPortal) (12). Mutation diagram circles are colored with respect to the corresponding mutation types. In case of different mutation types at a single position, color of the circle is determined with respect to the most frequent mutation type. Mutation types and corresponding color codes are as follows: green: Missense mutations; black: Truncating mutations. (B) The relevance of genetic disorders and TMPRSS2 expression were as shown, in prostate cancer there is a 30% fusion of the gene and in the stomach adenocarcinoma 8% deep deletion. (C) DNA copy variation were not statistically relevant to RNA TMPRSS2 expression and as shown, many pancreatic cancer patients had deep deletions. CTSL, cathepsin L; TMPRSS2, transmembrane protease serine 2.
their potential involvement in maternal and neonatal complications (11). These hypotheses remain to be investigated in large clinical studies.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

All data generated or analysed during this study are included in this published article (and its supplementary information files).

Authors’ contributions

PK generated the data, and produced the figures; VA and KC contributed to critical revision of the article; HSR, DAS contributed to the writing of the manuscript and final edits; IK and EK contributed equally to the conception of the work and data analysis and interpretation. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

DAS is the Editor-in-Chief for the journal, but had no personal involvement in the reviewing process, or any influence in terms of adjudicating on the final decision, for this article. The other authors declare that they have no competing interests.

References


This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.