TMPRSS2 transcriptional inhibition as a therapeutic strategy for COVID-19

Xinchen Wang, Ph.D.[1*], Ryan S. Dhindsa, Ph.D.[1,2], Gundula Povysil, M.D., Ph.D.[1], Anthony Zoghbi, M.D.[1,3,4], Joshua E. Motelow, M.D., Ph.D.[1,5], Joseph A. Hostyk, B.Sc.[1], Nicholas Nickols, M.D. Ph.D.[6,7], Matthew Rettig, M.D.[8,9], David B. Goldstein, Ph.D.[1,2*]

[1]Institute for Genomic Medicine, Columbia University Irving Medical Center, New York,
[2]Department of Genetics & Development, Columbia University Irving Medical Center, New York,
[3]Department of Psychiatry, Columbia University Irving Medical Center, New York
[4]New York State Psychiatric Institute, New York
[5]Division of Pediatric Critical Care, Department of Pediatrics, New York-Presbyterian Morgan Stanley Children's Hospital, Columbia University Irving Medical Center, New York
[6]Department of Radiation Oncology, University of California, Los Angeles, Los Angeles
[7]Department of Radiation Oncology, Veteran Affairs Greater Los Angeles Healthcare System, Los Angeles, California
[8]Division of Hematology and Oncology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles
[9]Division of Hematology and Oncology, VA Greater Los Angeles Healthcare System, Los Angeles, Los Angeles, California

*To whom correspondence should be addressed: xw2553@cumc.columbia.edu (X.W.), dg2875@cumc.columbia.edu (D.B.G.)

Abstract

There is an urgent need to identify effective therapies for COVID-19. The SARS-CoV-2 host factor protease *TMPRSS2* is required for viral entry and thus an attractive target for therapeutic intervention. In mouse, knockout of *tmprss2* led to protection against SARS-CoV-1 with no deleterious phenotypes, and in human populations genetic loss of *TMPRSS2* does not appear to be selected against. Here, we mined publicly available gene expression data to identify several compounds that down-regulate *TMPRSS2*. Recognizing the need for immediately available treatment options, we focused on FDA-approved drugs. We found 20 independent studies that implicate estrogenic and androgenic compounds as transcriptional modulators of *TMPRSS2*, suggesting these classes of drugs may be promising therapeutic candidates for clinical testing and observational studies of COVID-19. We also note that expression of *TMPRSS2* is highly variable and skewed in humans, with a minority of individuals having extremely high expression. Combined with literature showing that inhibition of *TMPRSS2* protease activity reduces SARS-CoV-2 viral entry in human cells, our results raise the hypothesis that modulation of *TMPRSS2* expression is a promising therapeutic avenue for COVID-19.

Introduction

The rapid international spread of the novel pathogenic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes the disease known as COVID-19, poses a global health emergency. As of April 5, 2020, there have been over 1,133,000 confirmed cases and 62,500 deaths worldwide[1]. The clinical presentation of COVID-19 ranges from mild respiratory symptoms to severe progressive pneumonia, multiorgan failure, and death[2]. Therapeutic interventions beyond supportive care in the literature have included oseltamivir, remdesivir, ganciclovir, α-interferon, hydroxychloroquine and lopinavir[2-7]. Lopinavir, a protease inhibitor, is the only drug with a completed clinical trial but failed to shorten time to improvement or viral shedding. Any effective intervention rapidly mobilized to the frontlines could profoundly impact resource allocation[8]. Effective treatments are therefore vital to handle the surge of COVID-19 infections.

SARS-CoV-2 host factors are attractive targets for therapeutic intervention. The SARS-CoV-2 spike (S) glycoprotein binds the angiotensin-converting enzyme 2 (*ACE2*), allowing the viral particle to enter host cells[9]. Viral entry into host cells also requires cleavage of the viral S protein by host proteases; this cleavage results in irreversible conformational changes to the S protein that allow the virus and host cell membranes to fuse[9]. S protein cleavage, called priming, can use the host serine protease *TMPRSS2* or the cysteine proteases cathepsin B or L (CatB/L)[10-14]. A recent single-cell RNA-sequencing study of human and non-human primate tissues revealed three major cell types that co-express *TMPRSS2* and *ACE2*: type II pneumocytes in the lung, absorptive enterocytes in the terminal ileum, and nasal goblet secretory cells[15].

Computational and *in vitro* screens are useful to identify compounds that either act directly against viral proteins, or that disrupt protein interactions between SARS-CoV-2 and host proteins required for its viral life cycle. Here we propose and develop a complementary approach seeking to identify transcriptional regulators of the host proteins most critical to viral entry and replication within host cells. Given the aggressiveness of this pandemic and the urgency of deploying effective treatments, our first efforts focus on the repurposing of existing drugs as an attractive alternative to novel compound discovery. We note, however, that this screening approach could also be applied to the discovery of new chemical entities with more desirable properties than already available approved medicines.

Several lines of evidence point to the host protease *TMPRSS2* that the virus employs to gain entry into lung epithelium as a promising target for pharmacologic targeting. Pharmacologic inhibition of *TMPRSS2* prevents SARS-CoV-2 entry into cultured human lung cells[11], and work in mouse also supports loss of *tmprss2* as being protective against SARS-CoV-1 (details in Supplementary Materials)[16]. Furthermore, loss-of-function of *TMPRSS2* is not strongly selected against in human populations (**Supplementary Fig. 1**)[17]. In contrast, for different reasons neither *ACE2* nor *CatB/L* appear as strong candidates for therapeutic intervention based on transcriptional inhibition (details in Supplementary Materials). For these reasons, we focus on transcriptional modulation of *TMPRSS2* as the highest priority, but have generated results for other critical host proteins using the same framework, in particular the other entry proteins and a broader set of proteins that interact with viral proteins[18]. The initial identification of these compounds highlights potential therapeutic targets and pathways that could be pursued by drug repurposing for the amelioratation of COVID-19 symptoms.

Results

Literature-wide screen for transcriptional inhibitors of SARS-CoV-2 host factors reveals drug repurposing candidates

To identify compounds that transcriptionally inhibit host factors required for SARS-CoV-2 viral entry, we performed a literature-wide screen of RNA-seq datasets in the NCBI Sequence Read Archive (SRA) that incorporated keywords relating to drug treatments. Of 252,877 human RNA-seq datasets in the SRA that were uniformly mapped by the Skymap project[19], we identified 29,550 samples in 1,222 studies that involved a drug treatment (Methods). Within each study, we manually assigned samples as case or control for each comparison group based on sample descriptions and literature reviews, yielding 3,089 distinct case-control comparisons. We used these comparisons to create a database of differentially expressed genes under various drug treatments.
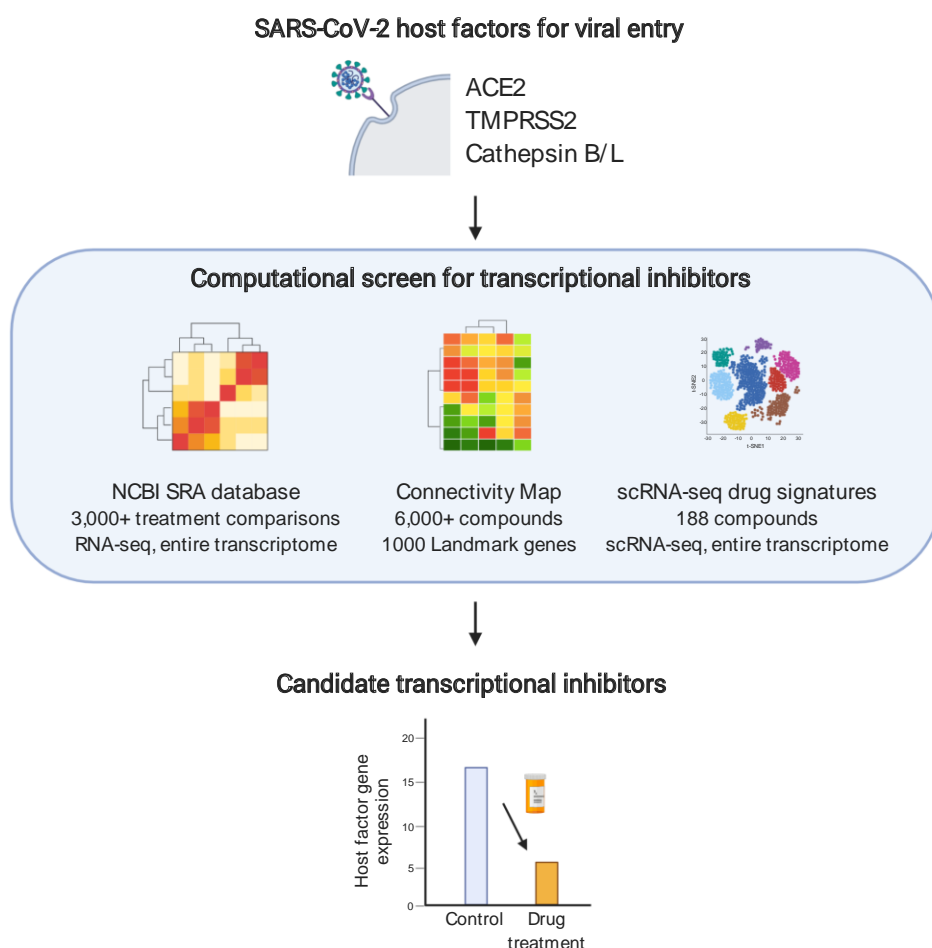

Fig. 1: Overview of repurposing approach to iden tify transcriptional inhibitors of SARS-CoV -2 host factors

Identification of TMPRSS2 transcriptional inhibitors

Next, we queried this database to identify drug treatments that led to significant differential expression of SARS-CoV-2 host factor genes. We first focused on *TMPRSS2*, given its promise as a candidate for therapeutic intervention. At a Bonferroni-corrected p-value of 0.05 (raw p < 2.06 x 10^{-5}), we identified 32 treatment conditions that led to significant down-regulation of *TMPRSS2* in human cell lines, and 76 conditions that led to up-regulation. While the drugs that down-regulate *TMRPSS2* may be useful as anti-viral targets, the drugs that up-regulate are also important to recognize because they may exacerbate viral infection. Notably, 12 of 32 drug treatments that significantly down-regulated TMPRSS2 (seven independent studies) and 24 of 76 treatments that led to TMPRSS2 up-regulation (15 independent studies) involved estrogens, androgens, or agonists or antagonists or their receptors. These results are consistent with studies showing that TMPRSS2 is regulated by androgens[20]. Specifically, treatment with estradiol (longer than 3 hours, see **Supplementary Fig. 2** for time-course data), genistein (a phytoestrogen that modulates ERα and ERβ), and MDV3100/enzalutamide (an androgen receptor antagonist commonly used in advanced prostate cancer) led to statistically significant down-regulation of *TMPRSS2* between 1.6-fold and 14-fold, depending on experimental conditions, cell lines and choice of controls (example in **Fig. 2**, data in **Supplementary Table 1**). Accordingly, we also replicated the enzalutamide signal in the LAPC4 prostate cancer cell line. (**Fig. 2**). Conversely, *TMPRSS2* expression increased between 1.4-fold and 20-fold following treatment with androgens (e.g. testosterone, any duration of treatment, see **Supplementary Fig. 2**), synthetic androgens (R1881/Metribolone) or *short-term* exposure to estradiol (under 3 hours, see **Supplementary Fig. 2**). In addition to estrogen and androgen-related compounds, other treatments that decrease *TMPRSS2* expression include: dual TGFβ and EGF treatment (347-fold decrease in expression in HeLa cells, single study only), and chaetocin (non-specific histone lysine methyltransferase inhibitor, ~4-fold decrease).

To identify more compounds that modulate *TMPRSS2* expression, we considered data from the Connectivity Map[21], which includes an unbiased screen of compounds in multiple cell types followed by expression profiling of 978 landmark genes (L1000 platform) and statistical imputation for the rest of the transcriptome. However, *TMPRSS2* was not well-imputed by the Connectivity Map (self-correlation = 0.56 between RNA-seq expression and imputed expression). Consistent with a recent preprint on the lack of reproducibility for L1000-based gene imputation values[22], we did not observe a significant difference in *TMPRSS2* expression following estradiol treatment in the breast or prostate cancer cells used in the RNA-seq studies

described above. Interestingly, the prostate cancer cell line used in these studies, PC3, lacks androgen receptor (AR) expression and has only minimal estrogen receptor (ER) expression, so the absence of a *TMPRSS2* signal is not unexpected.

We also considered single-cell RNA-sequencing (scRNA-seq) data from a recent study that measured transcriptomic signatures following treatment with 188 compounds in three cancer cell lines (MCF7 breast cancer, K562 leukemia, A549 lung adenocarcinoma)[23]. *TMPRSS2* expression was only present in MCF7 cells. In MCF7 cells, we identified two compounds that led to statistically significant increase in *TMPRSS2* expression: JQ-1, a BET bromodomain inhibitor (q-value = 1.58×10^{-24}, normalized effect size = 0.19), and fulvestrant, an estrogen receptor antagonist (q-value = 3.96×10^{-11}, normalized effect size = 0.14). Interestingly, JQ1 is known to inhibit ER expression in MCF7, so the effects of JQ1 on *TMPRSS2* expression may be mediated through the ER[24]. Together, these results identify existing drug compounds that can potentially be repurposed to transcriptionally inhibit *TMPRSS2* expression, and suggest that the activation of estrogen pathways or inhibition of androgen pathways can be a promising modality for clinical intervention in SARS-CoV-2 infection.
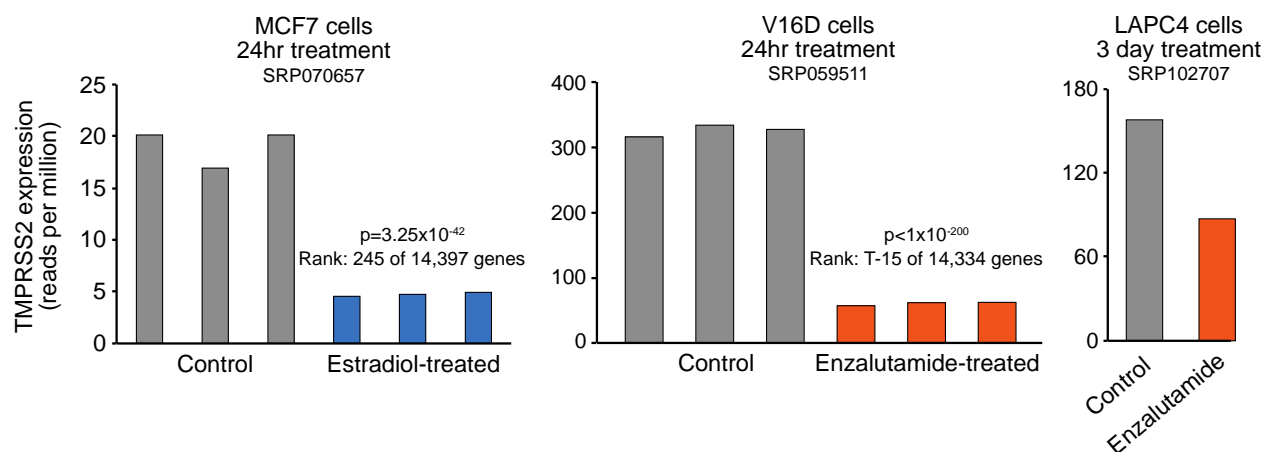

Fig. 2. **Example effects of transcriptional inhibitory effects of estradiol and enzalutamide treatment on TMPRSS2 expression.** Data from RNA-seq studies on MCF7 cells (breast cancer line), V16D cells (prostate cancer line) and LAPC4 cells (prostate cancer line, no known TMPRSS2 fusion). Bars correspond to distinct biological replicates. Difference in control expression of TMPRSS2 corresponds to differences in baseline TMPRSS2 expression in prostate & breast tissue (see Supplementary Fig. 3). p-values calculated using DESeq2, comparing entire transcrip- tome between treatment and control. Rank represents rank of TMPRSS2 differential expression p-value compared to all other genes in transcriptome. SRP value corresponds to accession number of corresponding study in the NCBI SRA.

Identification of transcriptional regulators of other host proteins involved in viral entry and replication

In addition to *TMPRSS2*, we also searched for and found strong inhibitors for transcriptional regulators of other key host proteins involved in SARS-CoV-2 viral entry and replication, including *ACE2*, *CatB/L*, and a recently published set of 332 host proteins that may be required for viral infection based on their protein-protein interactions with SARS-CoV-2 viral proteins[18] (Supplementary Materials). As an example, treatment with the BET bromodomain inhibitors, JQ-1 and CPI-203, led to reduction of *ACE2* expression. Notably, the CPI-203 treatment comparison was performed in bronchial epithelial cells, and treatment led to ~6-fold *ACE2* down-regulation in cells from both healthy patients and cystic fibrosis patients (38th strongest change out of 15,701 genes tested upon CPI-203 treatment, within-study FDR = 2.75×10^{-20}, **Supplementary Fig. 3**)[25].

Relevance in lung tissue of estradiol-based transcriptional inhibition of TMPRSS2

One key limitation of the drug repurposing analysis presented above is that all experiments were performed in breast cancer or prostate cancer cell lines, leading to concerns about applicability and relevance to lung tissue. We next asked whether treatment with estradiol or androgen receptor antagonists could potentially transcriptionally inhibit *TMPRSS2* expression in human lung tissue. First, we note that androgen treatment was shown to significantly increase *TMPRSS2* expression in A549 lung adenocarcinoma cells[20], and also that the androgen receptor which is known to regulate *TMPRSS2* is expressed in human lung[26]. To test whether treatment with anti-androgenic or pro-estrogenic compounds would also down-regulate expression of *TMPRSS2* in human lungs, we leveraged *TMPRSS2*'s extremely high variability in gene expression across individuals (**Supplementary Fig. 4**). Specifically, we hypothesized that if *TMPRSS2* expression in human lung tissue were regulated by estrogen and androgens, then its expression in lung would vary in tandem with known estrogen and androgen response genes (**Fig. 3A**). To test this hypothesis, we quantified the correlation in gene expression between *TMPRSS2* and 12,276 other genes expressed in lung across all 427 individuals with RNA-seq data (TPM > 5, Spearman correlation, GTEx v7 data). Notably, using Gene Set Enrichment Analysis (GSEA) on the gene list ranked by correlation with *TMPRSS2* expression[27], we observed that the hallmark early and late estrogen response gene sets, as well as the hallmark androgen response gene set are three of the top four sets most enriched in genes with high positive correlation with *TMPRSS2* expression in lung (**Fig. 3B,C**).

Finally, we tested whether the transcriptionally repressive direction of effect for TMPRSS2 modulation by estradiol and androgen receptor antagonists might be the same in the human lung as it is in cell lines. To do so, we considered temporal gene expression data from MCF7 breast cancer cells and LNCaP prostate cancer cells in response to estradiol and dihydrotestosterone (DHT, androgen) treatment (SRA accession #: SRP070657 and SRP059762)[28]. Notably, we observed that in human post-mortem lungs, genes with expression patterns that most strongly correlate with *TMPRSS2* also behave similarly to *TMPRSS2* in response to estradiol and DHT treatment in MCF7 and LNCaP cells, respectively (**Fig. 3D,** middle panels). Conversely, genes with expression patterns most inversely correlated with *TMPRSS2* in human lungs behave inversely in response to estradiol and DHT treatment (**Fig. 3D,** right panels). These results are consistent with a prior study that demonstrated both androgen-induced transcription of *TMPRSS2*, and androgen receptor binding to enhancer elements within the *TMPRSS2* gene, in human lung adenocarcinoma cells[20]. Collectively, these results indicate that expression of *TMPRSS2* in the human lung changes alongside known estrogen and androgen response genes, and with the same direction of effect. This suggests that the expression of *TMPRSS2* in the lung can be repressed by treatment with estrogens or with androgen receptor antagonists, and that expression can be activated by androgens. Furthermore, these data suggest that within human populations, *TMPRSS2* expression in the lung is modulated by estrogens and androgens.

(next page) Fig. 3. Relevance of estrogen-based TMPRSS transcriptional inhibition in lung. (A) Overview of approach to test whether estrogen and androgen-response genes correlate with TMPRSS2 expression in lung (n=427). (B) Top five enriched gene sets reveals expression of estrogen and androgen response genes in lung are highly correlated with expression of TMPRSS2. (C) Gene set enrichment plots for estrogen and androgen response sets. (D) Genes with most correlated expression to TMPRSS2 in lung also have the same estradiol and DHT-dependent gene expression patterns in MCF7 and LNCaP cells, respectively. Genes with least correlated expression to TMPRSS2 in lung have estradiol and DHT-dependent expression patterns different from TMPRSS2. Shared areas correspond to 95% CI.
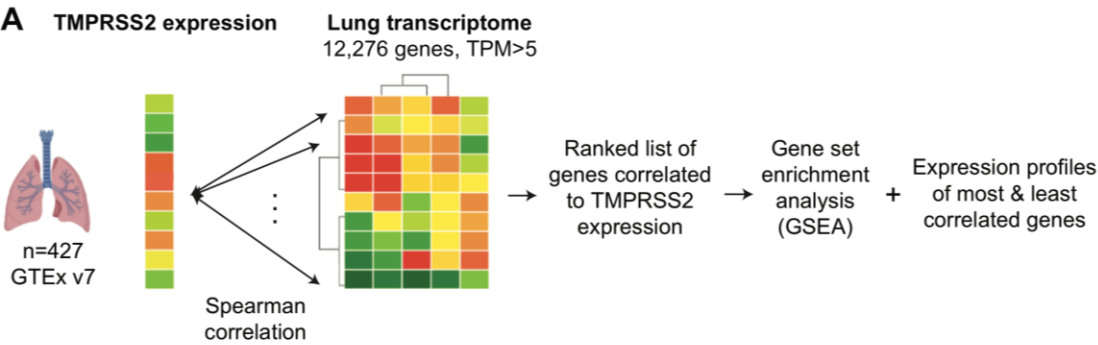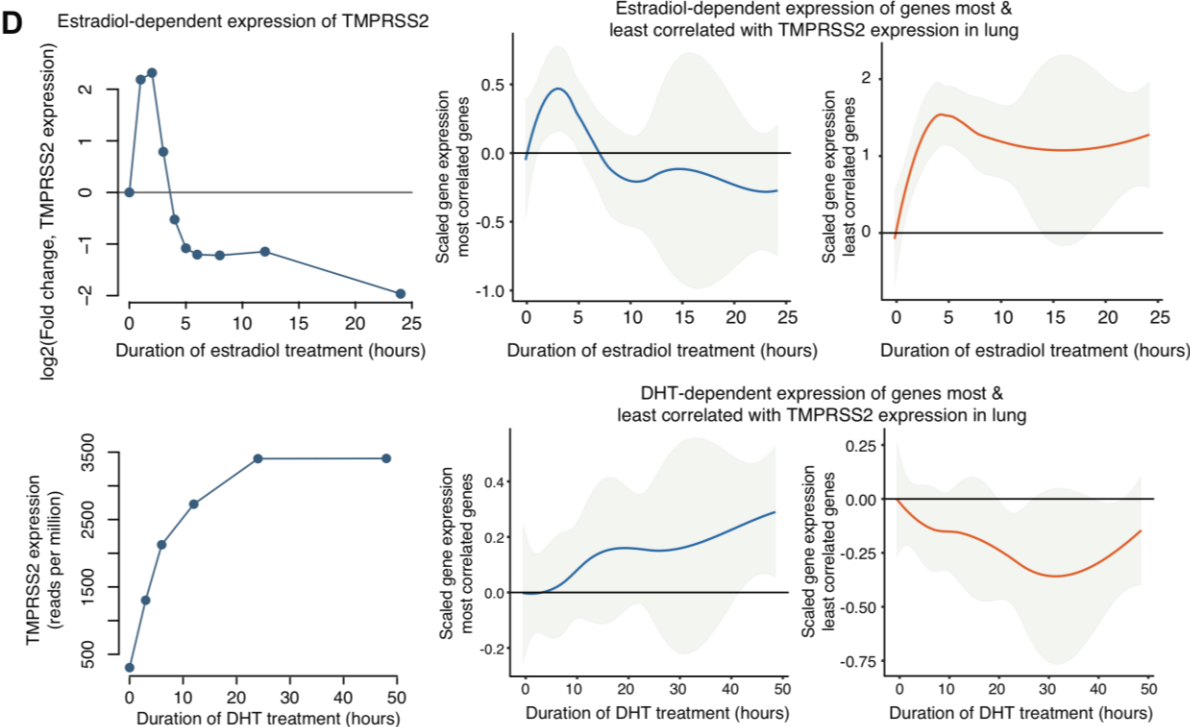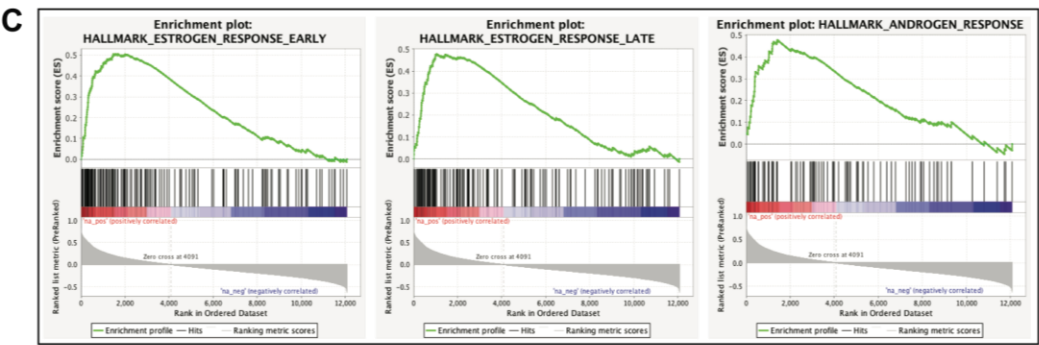
High variability in *TMPRSS2* expression within human populations

Recent epidemiological data indicate that COVID-19 may cause severe illness in up to 30% of infected individuals with a case fatality-rate exceeding 20% in high risk populations[2,29]. Sex-specific infection and severity rates vary, with some data pointing to equal infection rates between men and women, and other data showing higher rates of infection and critical illness in men[2,30-32]. Given that *TMPRSS2* expression is modulated by the sex hormones estradiol and androgens, and given evidence that *tmprss2* loss of function in mouse could be protective against severity of SARS-CoV-1[16], we next asked whether demographic differences in expression of *TMPRSS2*, as well as *ACE2*, could explain trends observed in COVID-19 epidemiological data.

The most striking feature of *TMPRSS2* and *ACE2* expression data is how variable their gene expression is amongst individuals. Using post-mortem gene expression data from human lungs (n=427, age range 20-80, GTEx Consortium v7)[26], we noticed that both *TMPRSS2* and *ACE2* have highly variable expression amongst individuals (**Supplementary Fig. 4**). This variability is not due to differences between age groups or sex, but is present within each demographic group. As an example, we considered a set of 156 men aged 40-59. In this group, *TMPRSS2* expression in the lung varied by two orders of magnitude from 2.3 TPM (transcripts per million) to 249.5 TPM (median 44.3 +/- 32.0 TPM). Further, while median *ACE2* expression is much lower than *TMPRSS2* (median TPM 1.1 in same individuals), we observed a similarly broad spread in expression spanning two orders of magnitude, with *ACE2* expression ranging from 0.17 TPM to 14.5 TPM in the same demographic group. Given the high spread in the expression of *TMPRSS2* and *ACE2*, as well as literature evidence that loss of *TMPRSS2* is associated with severity of infection[16], we hypothesized that the high variance in expression of *TMPRSS2* and *ACE2* could be associated with the wide range in COVID-19 severity observed in the human population. To test this, we quantified the skewness in expression of the four SARS-CoV-2 host factors, and noticed that all four genes have highly skewed expression patterns (skewness between 1.6 and 5.6), where the majority of individuals tested have low expression of each gene, but a small subset of individuals have very high expression. Notably, the expression of the four viral entry host factor genes is more skewed than the median gene expressed in lung (69th percentile or above for all genes, **Fig. 4A**). *ACE2* in particular has one of the most skewed expression patterns in the entire transcriptome (97th percentile), and even among genes with similar expression levels (94th percentile among genes with median TPM between 0.5 and 3, **Fig. 4B**). These results are consistent with a model where a small subset of

individuals have high expression of SARS-CoV-2 host factors and are at particularly high risk for being infected.
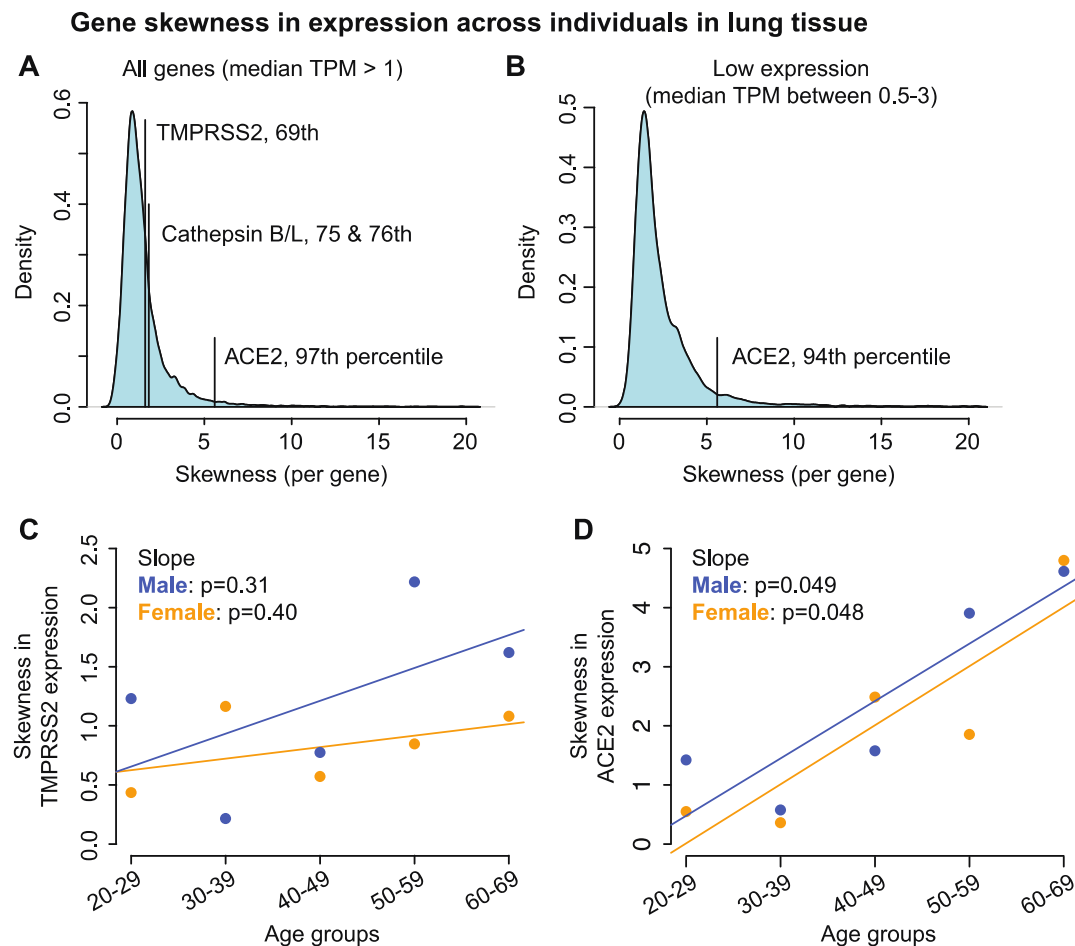

Fig. 4. Skewness in gene expression of SARS-CoV-2 host co-factors. (A) Skewness in gene expression across 427 lung samples for TMPRSS2, Cathepsin B/L and ACE2. Numbers after gene names correspond to percentiles of skewness compared to entire transcriptome. (B) ACE2 has high skewness even when compared to genes matched for low expression values. (C,D) Skewness in expression of TMPRSS2 (panel C) and ACE2 as a function of age (panel D). p-values calculated from linear regression of skewness against age.

In addition, we considered sex and age-specific gene expression patterns for the four host factor genes. Using the GTEx dataset, we did not observe strong sex-specific differences in the expression of any of the four genes (**Supplementary Figs. 4** and **5**), however a recent preprint reported that *TMPRSS2* has significantly higher expression in males specifically when considering gene expression data in bronchial epithelial cells from 170 individuals[33]. We also noticed that the skewness of *ACE2* expression within demographic groups increased

significantly with age (p=0.048 in females, p=0.049 in males, **Fig. 4D**). A more modest but not statistically significant increase in skewness was observed for *TMPRSS2* expression across age (**Fig. 4C**).

Discussion

Here, we have used a computational approach that leverages publicly available transcriptomic data investigating how a large number of compounds, including hundreds of FDA-approved drugs, perturb gene expression in multiple cell types. These analyses have identified a number of FDA-approved drugs that have been shown to reliably down-regulate the host proteins that are critical to the entry of SARS-CoV-2 into host cells. Amongst these, the estrogen-related compounds estradiol, genistein, and the androgen receptor antagonist enzalutamide are shown to downregulate *TMPRSS2*, which is required for SARS-CoV-2 spike protein priming and viral entry, and appear to be the most promising repurposing candidates for further tests in cellular models of SARS-CoV-2. Convergence of multiple compounds on the same biological pathway suggests that other commonly prescribed drugs acting on the same pathway but not within our database may also be promising repurposing candidates to transcriptionally inhibit *TMPRSS2*.

Fortuitously, a wealth of FDA approved drugs block AR signaling. These include the GnRH analogs that reduce pituitary release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), thereby potently suppressing testosterone production, and anti-androgens that interfere with binding of androgens to the AR. These drugs are predominantly used in the treatment of prostate cancer, have well known and generally tolerable side effect profiles, and exhibit reversibility of their biologic effects. Additionally, 5-alpha reductase inhibitors, such as dutasteride and finasteride, inhibit the peripheral conversion of testosterone to the more potent dihydrotestosterone, and may offer an alternative, less intense effect on *TMPRSS2* expression. Consistent with these data, a mouse study of SARS-CoV infection showed that in female mice, ovariectomy or treatment with estrogen receptor antagonists (both expected to increase lung *Tmprss2* expression) resulted in increased mortality to SARS-CoV-1 infection[34]. Of the host factors involved in SARS-CoV-2 entry, *TMPRSS2* appears the most promising candidate for transcriptional inhibition because: (i) *TMPRSS2* expression levels have been shown to be associated with severity in mouse models, (ii) population genetic considerations suggest that transcriptional inhibition of *TMPRSS2* may not be harmful, and (iii) *TMPRSS2* is essential for SARS-CoV-2 entry to host cells, unlike cathepsins B and L. Finally, we emphasize that the core hypotheses underlying the transcriptional inhibition approach are testable with observational studies to assess whether widely-used drugs affecting estrogen and androgen signaling correlate with COVID-19 severity. In support of this hypothesis, a recent cross-sectional study of female COVID-19 patients in Wuhan, China found an inverse correlation between serum

estradiol levels and COVID-19 severity, leading the authors to hypothesize that estradiol supplements may be protective against COVID-19[35].

While it is already known that sex hormones can regulate *TMRPSS2* expression[20], this work shows that estrogen-related compounds and androgen receptor antagonists appear to be the most securely identified down-regulators of TMPRSS2 expression amongst FDA approved drugs and other widely tested compounds. It is therefore a high priority to evaluate how estrogen-related compounds perform both in *in vitro* viral entry assays and in symptom amelioration in patients. While lower priority, the regulators of *ACE2* and *CatB/L* should also be evaluated in *in vitro* viral entry assays, and it seems possible that *ACE2* down-regulation could theoretically play a role in prophylactic prevention of infection, but not reduction in symptom severity, if such transcriptional inhibition were shown to be safe.

Expression variability of *ACE2* and *TMPRSS2*

Of particular interest is the wide variability of *TMPRSS2* expression levels in the human population, suggesting a possible explanation for much of the variability amongst people in the severity of disease. While we do not observe a strong correlation between increased *TMPRSS2* or *ACE2* expression and age in adults, we do observe a significant increase in skewness of *ACE2* expression with age and non-significant increase in skewness of *TMPRSS2*, which raises a possible connection between higher expression of these host factors and the marked increase in mortality rate in the older population. Further, the dramatic variation in expression of both *TMPRSS2* and *ACE2* also suggests a credible hypothesis for variation of vulnerability within age groups. We note that this hypothesis can be readily tested by evaluating *TMPRSS2* expression levels in patients with a range of severities, including comparisons of all symptomatic patients with the general populations. We consider these assessments a high priority, and suggest consideration of whether these evaluations can be performed in the same RNA samples used for SARS-CoV-2 diagnostics, given that scRNA-seq studies have shown *TMPRSS2* and *ACE2* expression in nasal goblet cells.

Considerations for hormone-based trials in COVID-19

The computational review of hundreds of FDA-approved compounds for transcriptional inhibitors of *TMPRSS2* allows us to select the best starting points focused on maximizing transcriptional inhibition. Our computational analysis led to our hypothesis that robustly reducing AR transcriptional output will suppress expression of TMPRSS2 in target cells, interfere with

SARS-CoV-2 entry and priming, thus reducing severity or duration of COVID-19. Notably, the AR is the direct target of dozens of approved drugs. A trial in COVID-19 could therefore directly enter in Phase II. The next step for application for COVID-19 is to select among the numerous existing therapeutics. The ideal choice would offer rapid and reproducible suppression of AR transcriptional activity with tolerable side effects. Many AR antagonists, including enzalutamide, are highly potent AR inhibitors but take several weeks to achieve steady state levels – far too long for use against COVID-19. GnRH agonists, such as leuprolide and goserelin, potently suppress androgen synthesis. However, these agents cause a transient but inevitable surge in testosterone (potentially worsening COVID-19) followed by a period of up to 2-3 weeks to achieve nadir levels of serum testosterone. In contrast, the GnRH antagonists, such as degarelix, elicit a rapid and reproducible suppression of circulating testosterone to castrate levels within 2-3 days without an initial surge and a median 88% reduction in serum testosterone at 24 hours[36].

The next step towards clinical translation is to select the optimal patient population. Although data continue to evolve, many patients with COVID-19 exhibit a mild disease course and therapeutic intervention may be riskier than supportive measures, although a mild form of hormonal modulation could be considered. The end stage of COVID-19, characterized by respiratory and multi-organ failure, may be driven by a hyperinflammatory response rather than active viral replication. Thus, we contend that the optimal patient population would be those who manifest symptomatic disease that warrants hospitalization for supportive care, but are not yet exhibiting signs characteristic of acute respiratory distress syndrome (ARDS). Finally, it may be prudent to evaluate androgen suppression first in male COVID-19 patients, as suppression of androgen in females *reduces* endogenous estrogenic activity and thereby could increase expression of *TMPRSS2* and exacerbate COVID-19.

Given that transcriptional suppression of *TMPRSS2* operates independently of all other COVID-19 treatments currently being tested or considered, a combination therapy approach using anti-androgens in tandem with antivirals such as remdesivir or competitive inhibitors of viral entry proteins would also be an obvious priority for evaluation. It is worth noting that a small molecule inhibitor of *TMPRSS2*, camostat mesilate, has recently entered trials for COVID-19 (NCT04321096). In addition, testing combination therapies and comparing different hormone-based treatments in each sex is also a high priority for pre-clinical models.

In summary, we have developed a therapeutic strategy designed to complement existing antiviral strategies focused on inhibition of viral proteins and strategies focused on small molecule disruption of host protein interactions with the virus. Our strategy seeks to transcriptionally inhibit the key proteins that SARS-CoV-2 relies upon, and has identified immediate opportunities for therapeutic intervention using approved estrogen-related compounds or anti-androgens that modulate expression of *TMPRSS2*, critical to viral entry. Depending on the degree of *TMPRSS2* transcriptional inhibition needed for a protective effect, we hypothesize these transcriptional inhibitors can be used either alone or in combination with other direct inhibitors. Further, this framework can be expanded to identify the most effective down regulators of both viral entry proteins and proteins critical to the life cycle of the virus within cells.

Supplementary Tables

Supplementary Table 1: Transcriptional modulators of TMPRSS2 by RNA-seq using NCBI SRA

Supplementary Table 2: Transcriptional modulators of ACE2 by RNA-seq using NCBI SRA

Supplementary Table 3: Transcriptional modulators of CTSB by RNA-seq using NCBI SRA

Supplementary Table 4: Transcriptional modulators of CTSL by RNA-seq using NCBI SRA

Supplementary Table 5: Transcriptional modulators of CTSB by Connectivity Map

Supplementary Table 6: Transcriptional modulators of CTSL by Connectivity Map

Supplementary Table 7: Transcriptional inhibitors of host proteins with protein-protein interactions to SARS-CoV-2 viral proteins

Materials and Methods

Computational screening for transcriptional inhibitors

We downloaded uniformly processed RNA-seq datasets from 3,764,506 high-throughput sequencing samples from SkyMap in raw count format[19]. We subsetted these for RNA-seq datasets performed on human samples, yielding 252,877 samples. Potential studies incorporating drug treatments were identified by searching for studies incorporating the term "treatment" or "nM", yielding 29,550 unique samples (designated by SRS ID) in 1,222 unique studies (designated by SRP ID). Samples within each study were manually assigned to case or control status, matching for other conditions (e.g. cell line used). To ensure consistency, all manual assignments were performed by a single Ph.D-level researcher with expertise in wet-lab experimental design using sample annotations present in the SRA, as well as reading relevant papers when case-control assignments are ambiguous or insufficient annotations were provided. Samples were allowed to be assigned to multiple comparisons (e.g. same samples used as controls for different drug perturbations), and samples with similar conditions (e.g. similar dosages or treatment times) were grouped to increase statistical power in some cases. In these cases, non-grouped sample comparisons were also kept as a separate comparison entry. In total, the 1,222 studies yielded 3,089 case control comparisons. Within each comparison, differentially expressed genes were identified using the DESeq2 R package on raw

count data, and for the current analysis, we focused specifically on a subset of comparisons with at least one biological replicate. scRNA-seq drug treatment data was downloaded from Srivatsan *et al.*[23].

TMPRSS2 gene expression analyses

TMPRSS2 gene expression across human individuals were downloaded from the GTEx project (version 7). Correlation between TMPRSS2 and other genes expressed in lung (n=427) was performed in R by calculating the Spearman correlation between TMPRSS2 expression and other genes, restricting to only 12,276 genes with median expression greater than 5 transcripts per million. Skewness of all expression patterns was calculated using the "skewness" function in the e1071 package in R.

Gene Set Enrichment Analysis was run using GSEA (v4.0.3) using the pre-ranked function, with values for each gene from -1 to 1 taken from their Spearman correlation with TMPRSS2 expression in human lung samples.

Comparisons of temporal gene expression for genes with high and low correlation were performed by taking the top and bottom 100 genes by Spearman's correlation (corresponding to correlation > 0.627 and < -0.5032). Estradiol data taken from SRP070657, and DHT data taken from SRP059762. As we consider the expression patterns of many genes, all with different baseline expression, we scaled the expression of every gene by dividing the expression at each time point by the expression at t=0. Plots were generated in ggplot2 using geom_smooth (loess).

Connectivity Map analyses

Connectivity Map L1000 data were downloaded from the NCBI GEO database (GSE70138 and GSE92742). We extracted imputed expression data for TMPRSS2 and ACE2, and for each treatment compound, we performed a non-parametric Mann-Whitney U test between control DMSO-treated samples and drug-treated samples. As Connectivity Map data tests drug treatments across a wide range of concentrations, we performed comparisons using the following dosage groups: (i) any dosage, (ii) below 0.5uM, (iii) 0.2uM to 1uM, (iv) 0.5uM to 2uM, (v) below 1uM. Multiple testing correction was performed using the Bonferroni-Hochberg correction.

To identify broad transcriptional inhibitor of SARS-CoV-2 viral protein interacting partners, we queried the Connectivity Map (clue.io) to identify compounds most likely to down-regulate host proteins required for SARS-CoV-2 pathogenesis as determined via a protein-protein interaction analysis[18]. The Connectivity Map considers up to 150 genes per query. We used three different signatures as input. In the first, we included the 33 out of the 332 host proteins that are directly quantified by the L1000 platform (i.e. "landmark genes"). The second signature included these landmark genes and 33 other well-imputed (self-correlation >0.8) genes. In the third signature, we considered the top 150 genes from the protein-protein interaction map, ranked by fold change. We considered the "summary" Connectivity Score for each compound, which represents the average Connectivity Score achieved in each of the nine core CMap cell lines.

References

1.   World Health Organization. (2020). Novel Coronavirus (2019-nCoV): Situation Reports. Geneva: WHO; 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

2.   Guan W-J, Ni Z-Y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020;

3.   Wu C, Chen X, Cai Y, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med 2020;

4.   Cao B, Wang Y, Wen D, et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. N Engl J Med 2020;

5.   Holshue ML, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 2020;382(10):929–36.

6.   Xu Y, Li X, Zhu B, et al. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. Nat Med 2020;:1–9.

7.   Mitjà O, Clotet B. Correspondence. The Lancet Global Health 2020;:1–2.

8.   Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA 2020;

9.   Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science 2020;:eabb2762.

10.   Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. Proc Natl Acad Sci USA 2005;102(33):11876–81.

11.   Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 2020;1–19.

12.   Glowacka I, Bertram S, Muller MA, et al. Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. J Infect Dis 2011;190(1):4122–34.

13.   Simmons G, Zmora P, Gierer S, Heurich A, Pöhlmann S. Antiviral Research. Antiviral Research 2013;100(3):605–14.

14.   Matsuyama S, Nagata N, Shirato K, Kawase M, Takeda M, Taguchi F. Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. J Virol 2010;84(24):12658–64.

15.   Ziegler, C. *et al.* SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues. *CellPress Sneak Peek.* Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3555145

16.   Iwata-Yoshikawa N, Okamura T, Shimizu Y, Hasegawa H, Takeda M, Nagata N. TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. J Virol 2019;93(6).

17.   Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285–91.

18.   Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. bioRxiv.org preprint, doi: 10.1101/2020.03.22.002386

19.   Tsui B, Dow M, Skola D, Carter H. Extracting allelic read counts from 250,000 human sequencing runs in Sequence Read Archive. Pac Symp Biocomput 2019;24:196–207.

20.   Mikkonen L, Pihlajamaa P, Sahu B, Zhang F-P, Jänne OA. Androgen receptor and androgen-dependent gene expression in lung. Molecular and Cellular Endocrinology 2010;317(1-2):14–24.

21.   Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 2017;171(6):1437–1452.e17.

22.   Lim N, Pavlidis P. Evaluation of Connectivity Map shows limited reproducibility in drug repositioning. bioRxiv.org preprint, doi: 10.1101/845693

23.   Srivatsan SR, McFaline-Figueroa JL, Ramani V, et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science 2020;367(6473):45–51.

24.   Sengupta S, Biarnes MC, Jordan VC. Cyclin dependent kinase-9 mediated transcriptional de-regulation of cMYC as a critical determinant of endocrine-therapy resistance in breast cancers. Breast Cancer Res Treat 2014;143(1):113–24.

25.   Chen K, Campfield BT, Wenzel SE, et al. Antiinflammatory effects of bromodomain and extraterminal domain inhibition in cystic fibrosis lung inflammation. JCI Insight 2016;1(11):L257.

26.   GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348(6235):648–60.

27.   Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102(43):15545–50.

28.   Baran-Gale J, Purvis JE, Sethupathy P. An integrative transcriptomics approach identifies miR-503 as a candidate master regulator of the estrogen response in MCF-7 breast cancer cells. RNA 2016;22(10):1592–603.

29.   Livingston E, Bucher K. Coronavirus Disease 2019 (COVID-19) in Italy. JAMA 2020;

30.   Cai H. Correspondence. The Lancet Respiratory 2020:1–1. doi:10.1016/S2213-2600(20)30117-X

31.   Yang, X, Yu, Y, Xu, J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet 2020. doi:10.1016/S2213-2600(20)30079-5

32.   Zhang J-J, Dong X, Cao Y-Y, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020;395(10223):507.

33.   Asselta R, Paraboschi EM, Mantovani A, Duga S. ACE2and TMPRSS2variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. medRxiv.org preprint, doi:10.1101/2020.03.30.20047878

34.   Channappanavar R, Fett C, Mack M, Eyck Ten PP, Meyerholz DK, Perlman S. Sex-Based Differences in Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Infection. J Immunol 2017;198(10):4046–53.

35.   Ding T, Zhang J, Wang T, et al. A Multi-hospital Study in Wuhan, China：Protective Effects of Non-menopause and Female Hormones on SARS-CoV-2 infection. medRxiv.org preprint, doi: 10.1101/2020.03.26.20043943

36.   Klotz L, Boccon-Gibod L, Shore ND, et al. The efficacy and safety of degarelix: a 12-month, comparative, randomized, open-label, parallel-group phase III study in patients with prostate cancer. BJU Int 2008;102(11):1531–8.