Supplementary Materials for "TMPRSS2 transcriptional inhibition as a therapeutic strategy for COVID-19"

Evidence for TMPRSS2 as a candidate therapeutic target

1) *TMPRSS2* loss-of-function alleles are tolerated in the human population (gnomAD probability of loss of function intolerance, pLI = 0, **Supplementary Fig. 1**)[1]
2) *tmprss2* knockout in mouse is not lethal, and do not have any severe phenotypes[2]
3) *tmprss2* knockout mice infected with SARS-CoV-1 have less viral replication and loss of body weight compraed to wild-type mice[2].
4) Conventional inhibition of *TMPRSS2* with camostat mesilate can block viral entry by SARS-CoV-2[3].

Evidence for ACE2 and CatB/L as a therapeutic target

1) ACE2 is loss-of-function intolerant in human (pLI = 1).
2) While homozygous loss of ace2 in mouse prevents SARS-CoV-1 viral entry, partial loss is thought to contribute to the severity of lung pathologies following SARS-CoV-1 infection, indicating that down-regulation of *ACE2* in infected individuals may further harm the lungs of infected patients[4].
3) *CatB/L* (gene symbols *CTSB*, *CTSL*) are both loss-of-function tolerant in human populations (*CTSB*: pLI = 0; *CTSL*: pLI = 0.01), yet *Ctsl* loss in mouse appears to be deleterious. While *Ctsb* knockout mice are born normal, without gross abnormalities, and show resistance to induced pancreatitis[5,6], *Ctsl* knockout mice have hair-loss, skin defects, impaired T cell maturation, dilated cardiomyopathy, and high postnatal mortality[7]. Furthermore, *Ctsl* knockout mice have higher mortality after influenza A infection compared to wildtype mice, which might be due to defective immune responses caused by the *Ctsl* knockout[6]. In addition and perhaps most importantly, *Ctsb* and *Ctsl* are dispensible for viral entry and spread[3].

Identification of ACE2 transcriptional inhibitors

We searched for compounds that transcriptionally modulate *ACE2* expression, however we note that while *ACE2* is used as a host factor for viral entry, virus-induced

loss of *ACE2* expression is believed to exacerbate SARS-CoV symptoms. *ACE2* shows a very tissue-restricted expression pattern, reducing the number of experiments in which comparisons could be performed (**Supplementary Fig. 6**). Despite this limitation, we identified 9 comparison conditions that led to ACE2 down-regulation, and 24 that lead to *ACE2* up-regulation (**Supplementary Table 2**). Within these nine comparisons, we noticed that two 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**). A search in our database for other treatments involving BET inhibitors comparisons yielded two additional comparisons where *ACE2* differential expression was tested (in A375 and SET2 cells); however both comparisons were not statistically significant, likely due to low baseline expression of *ACE2* in those cell lines. Notably, the synthetic androgen R1881 was among compounds that up-regulate *ACE2*, in addition to *EGFR* inhibitors (gefitinib, erlotinib, WZ4002). *ACE2* expression was poorly imputed by the Connectivity Map (self-correlation = 0.38), and is poorly expressed in the three cell lines profiled by the drug transcriptome RNA-seq study.

Identification of CTSB and CTSL transcriptional inhibitors

Finally, we considered targets that could lead to transcriptional inhibition of cathepsin B and cathepsin L. 44 treatment conditions led to decreases in *CTSB* expression, and 74 led to creases in *CTSL* expression. Notably, cardiac glycosides used for heart failure such as proscillaridin and digoxin led to decreases in both *CTSB* and *CTSL* expression (~4-8-fold decreases in expression, **Supplementary Tables 3** and **4**). In the Connectivity Map, expression of *CTSL* was directly assayed by the L1000 array, and expression of *CTSB* was imputed to a higher accuracy than both *TMPRSS2* and *ACE2* (self-correlation = 0.82), suggesting Connectivity Map data could be used for compound identification for both CTSB and CTSL transcriptional inhibition (outputs available in **Supplementary Tables 5** and **6**). For both genes, the Connectivity Map analysis identified over 100 compounds that led to either significant up-regulation or down-regulation of *CTSB* or *CTSL*, and of note, treatment with the cardiac glycoside digoxin led to down-regulation of both cathepsin genes across a range of dosages.

Broad transcriptional inhibitors of host proteins that interact with SARS-CoV-2 viral proteins

We next considered a larger set of 332 host proteins that may be required for viral infection based on their protein-protein interactions with SARS-CoV-2 viral proteins[8]. These proteins were recently identified by affinity-purification mass spectrometry of SARS-CoV-2 proteins expressed in human cells. Using the Connectivity Map, we sought to identify compounds that could down-regulate these host proteins. The Connectivity Map uses a "connectivity score" to assess each tested compound's ability to reverse a query signature. This score ranges from -100 to +100, with a score of -100 indicating complete reversal. We used three different subsets of the 332 host genes. First, we only considered the 33 genes included in the 978 landmark genes directly profiled in the L1000 assay (**Supplementary Fig. 7A**). In the second query, we added in an additional 33 genes that are well-imputed (**Supplementary Fig. 7B**). In the third query, we included the top 150 genes by fold-change in affinity-purification from the protein-protein-interaction map, regardless of imputation quality (**Supplementary Fig. 7C**). In total, we identified 12 compounds that achieved a Connectivity Score stronger than the recommended cutoff of -90 across all three queries (**Supplementary Table 7**). Because these compounds target more proteins than only those required for viral entry, these types of candidates may be efficacious in more broadly limiting viral entry and replication.

References for Supplementary Information

1. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
2. Iwata-Yoshikawa, N. *et al.* TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol.* **93,** (2019).
3. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 1–19 (2020). doi:10.1016/j.cell.2020.02.052
4. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. *Nat Med* **11,** 875–879 (2005).
5. Halangk, W. *et al.* Role of cathepsin B in intracellular trypsinogen activation and the onset of acute pancreatitis. *J. Clin. Invest.* **106,** 773–781 (2000).
6. Xu, X., Greenland, J. R., Gotts, J. E., Matthay, M. A. & Caughey, G. H. Cathepsin L Helps to Defend Mice from Infection with Influenza A. *PLoS ONE* **11,** e0164501 (2016).
7. Petermann, I. *et al.* Lysosomal, cytoskeletal, and metabolic alterations in cardiomyopathy of cathepsin L knockout mice. *FASEB j.* **20,** 1266–1268 (2006).
8. Gordon, D. E. *et al.* A SARS-CoV-2-Human Protein-Protein Interaction Map

Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv.org*
doi:10.1101/2020.03.22.002386

Supplementary Fig. 1: Human constraint scores for host proteins involved in SARS-CoV-2 viral entry

TMPRSS2

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	143.6	132	Z = 0.76 o/e = 0.92 (0.8 - 1.06)
Missense	323.2	303	Z = 0.4 o/e = 0.94 (0.85 - 1.03)
pLoF	30.9	20	pLI = 0 o/e = 0.65 (0.46 - 0.94)

ACE2

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	101	84	Z = 1.33 o/e = 0.83 (0.7 - 1)
Missense	281.8	223	Z = 1.25 o/e = 0.79 (0.71 - 0.88)
pLoF	31	3	pLI = 1 o/e = 0.1 (0.04 - 0.25)

Cathepsin B (CTSB)

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	82.6	168	Z = -7.39 o/e = 2.03 (1.73 - 1.99)
Missense	207.9	344	Z = -3.35 o/e = 1.65 (1.51 - 1.81)
pLoF	19.2	17	pLI = 0 o/e = 0.89 (0.61 - 1.33)

Note Too many missense variants; missense z score < -5

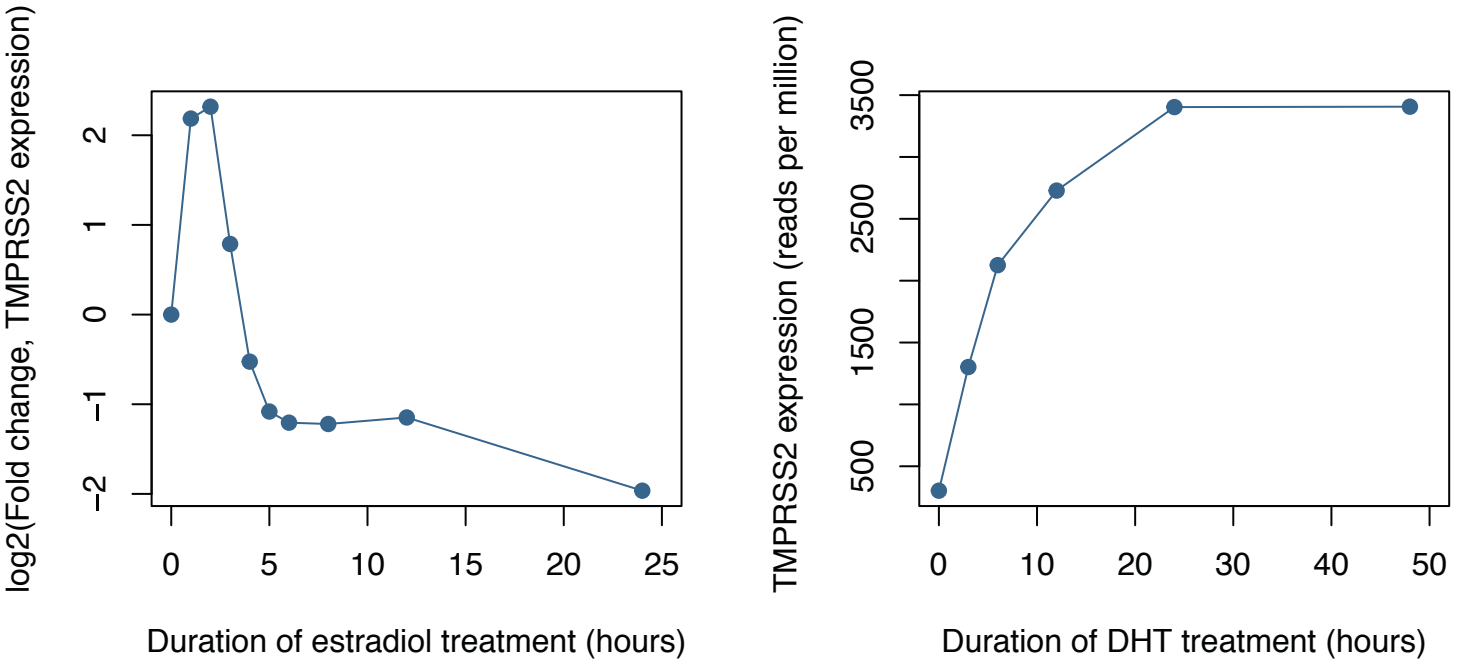
Note Too many or too few synonymous variants; synonymous z score < -5 or synonymous z score > 5

Cathepsin L (CTSL)

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	64.2	59	Z = 0.51 o/e = 0.92 (0.74 - 1.14)
Missense	180.7	167	Z = 0.36 o/e = 0.92 (0.81 - 1.05)
pLoF	15.2	6	pLI = 0.01 o/e = 0.4 (0.22 - 0.78)

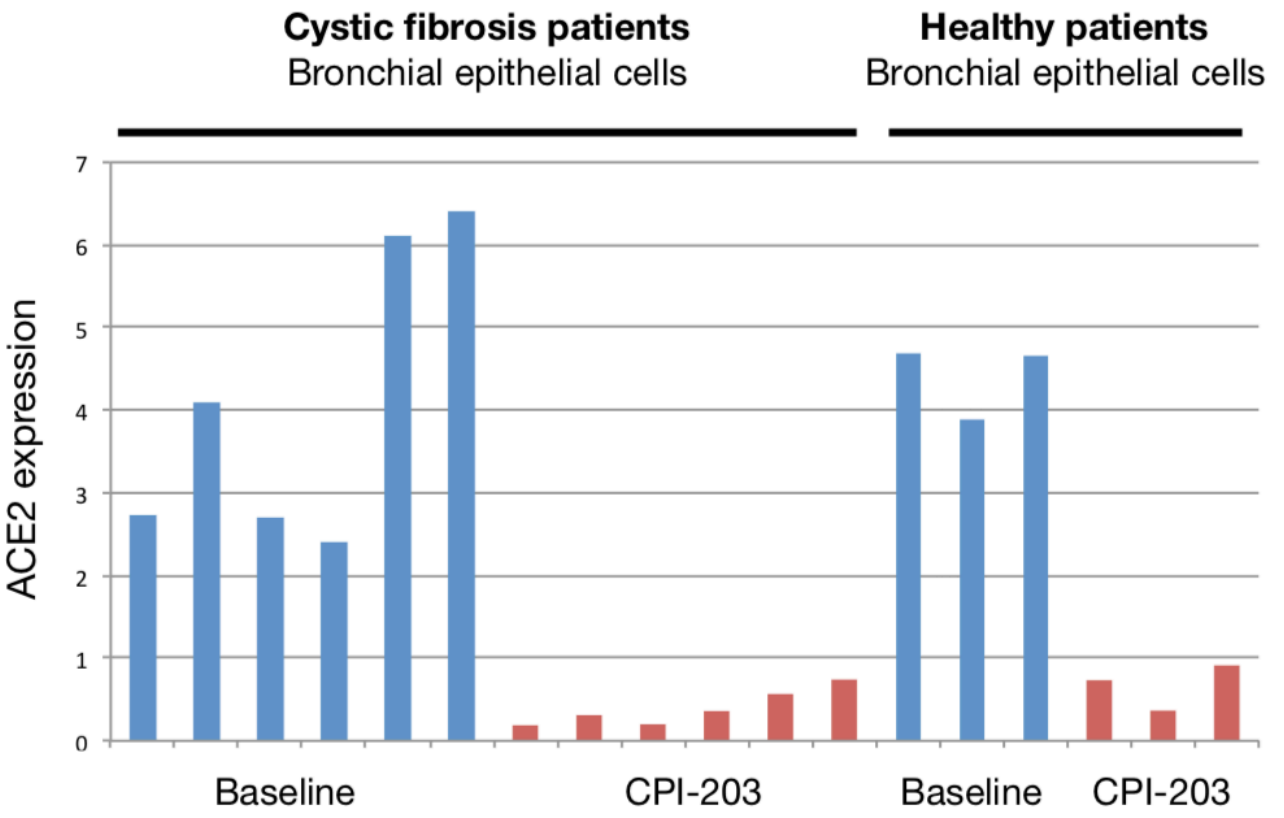
Supplementary Fig. 1: Human population-based constraint scores for TMPRSS2, ACE2, Cathepsin B and Cathepsin L. Data taken from the gnomAD browser (v2.1.1). Of the four host proteins, only ACE2 is strongly intolerant of loss-of-function mutations

Supplementary Fig. 2: Time-dependent effects of estradiol and DHT on TMPRSS2 expression

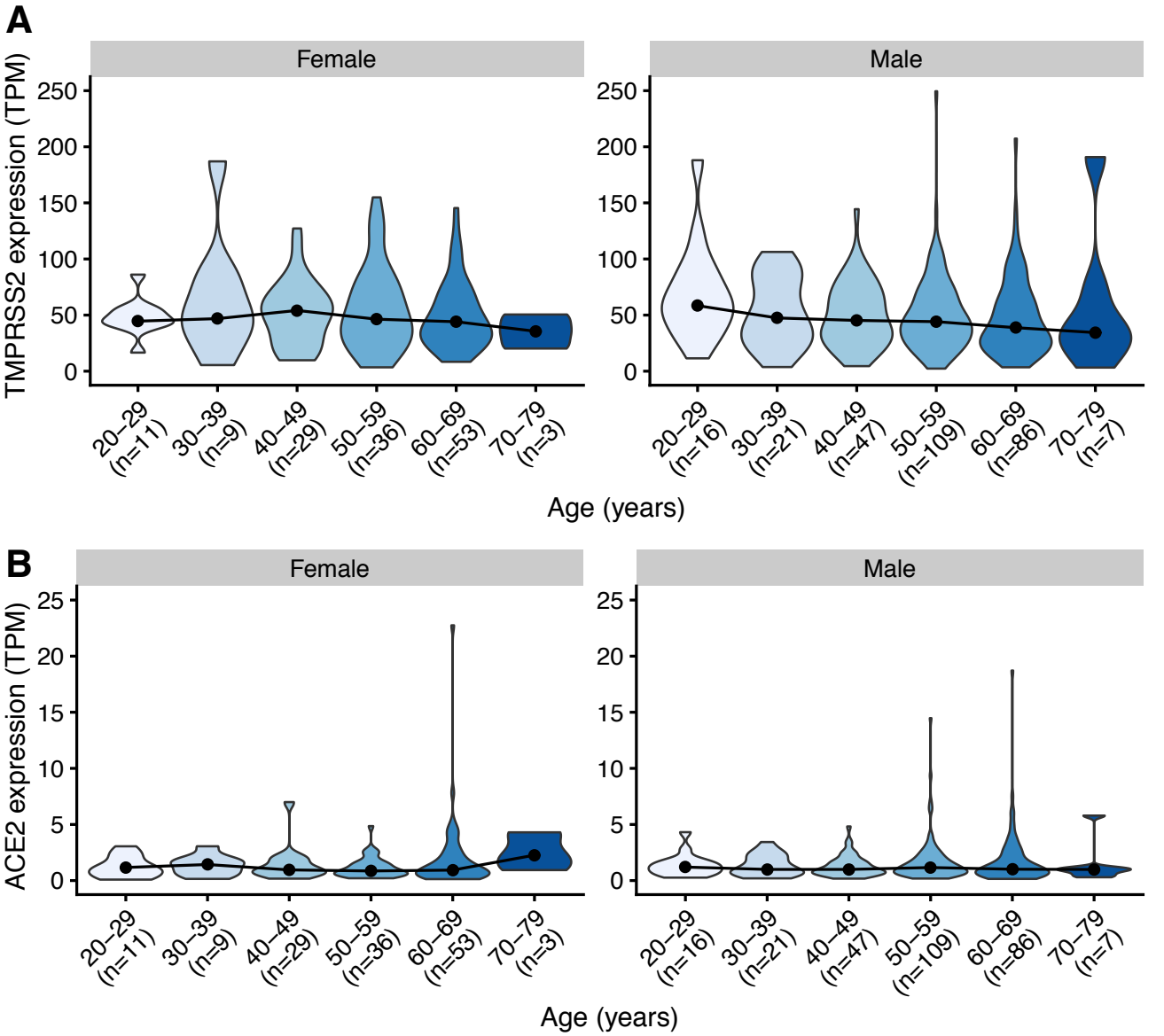


Supplementary Fig. 2: Time-dependent effects of estradiol and DHT on TMPRSS2 expression. Estradiol data from Baran-Gale et al. (SRP070657), DHT data from GEO accession GSE70150.

Supplementary Fig. 3: CPI-203 treatment in bronchial epithelial cells reduces ACE2 expression

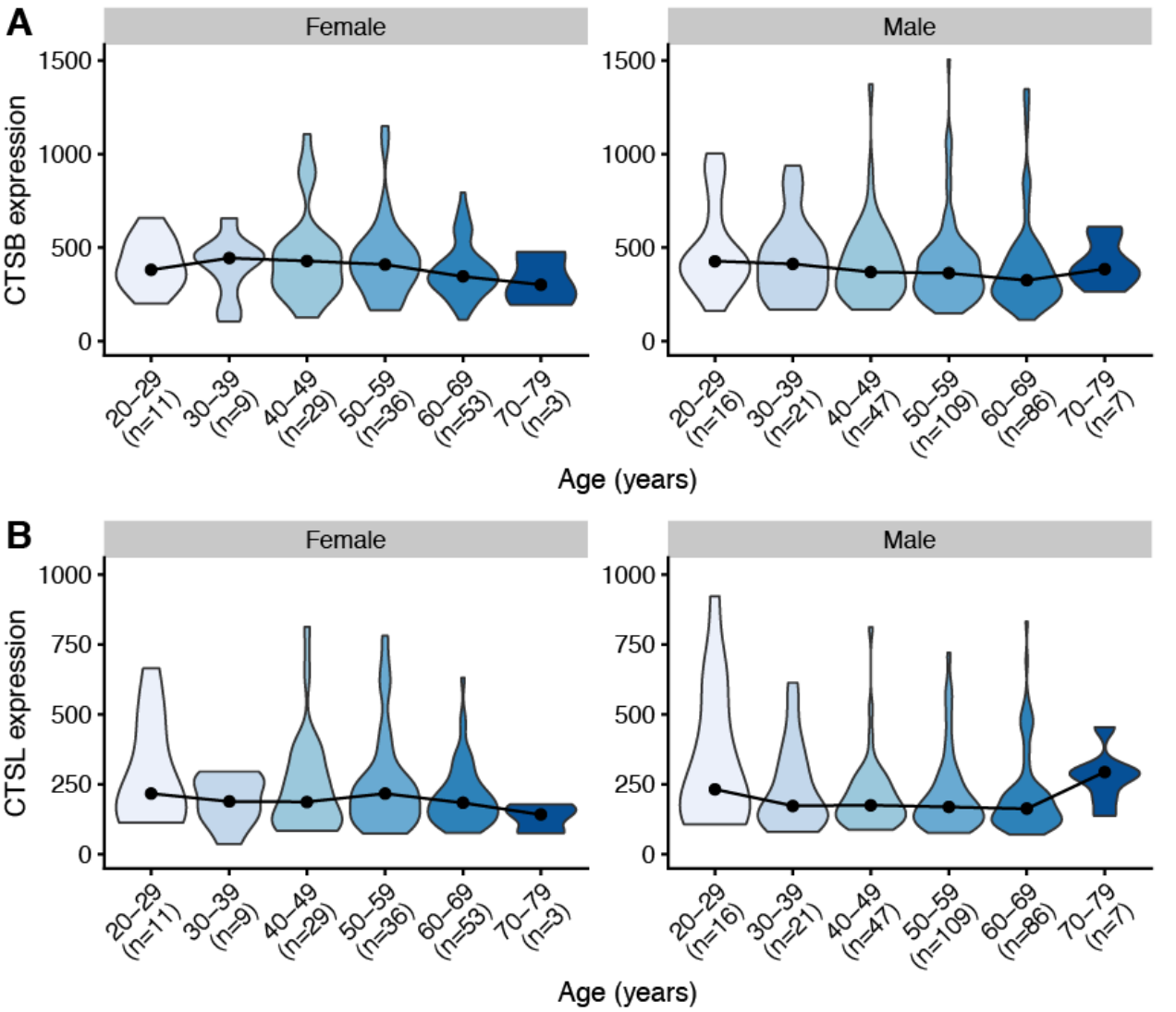


Supplementary Fig. 4: High variability in TMPRSS2 and ACE2 expression within human population

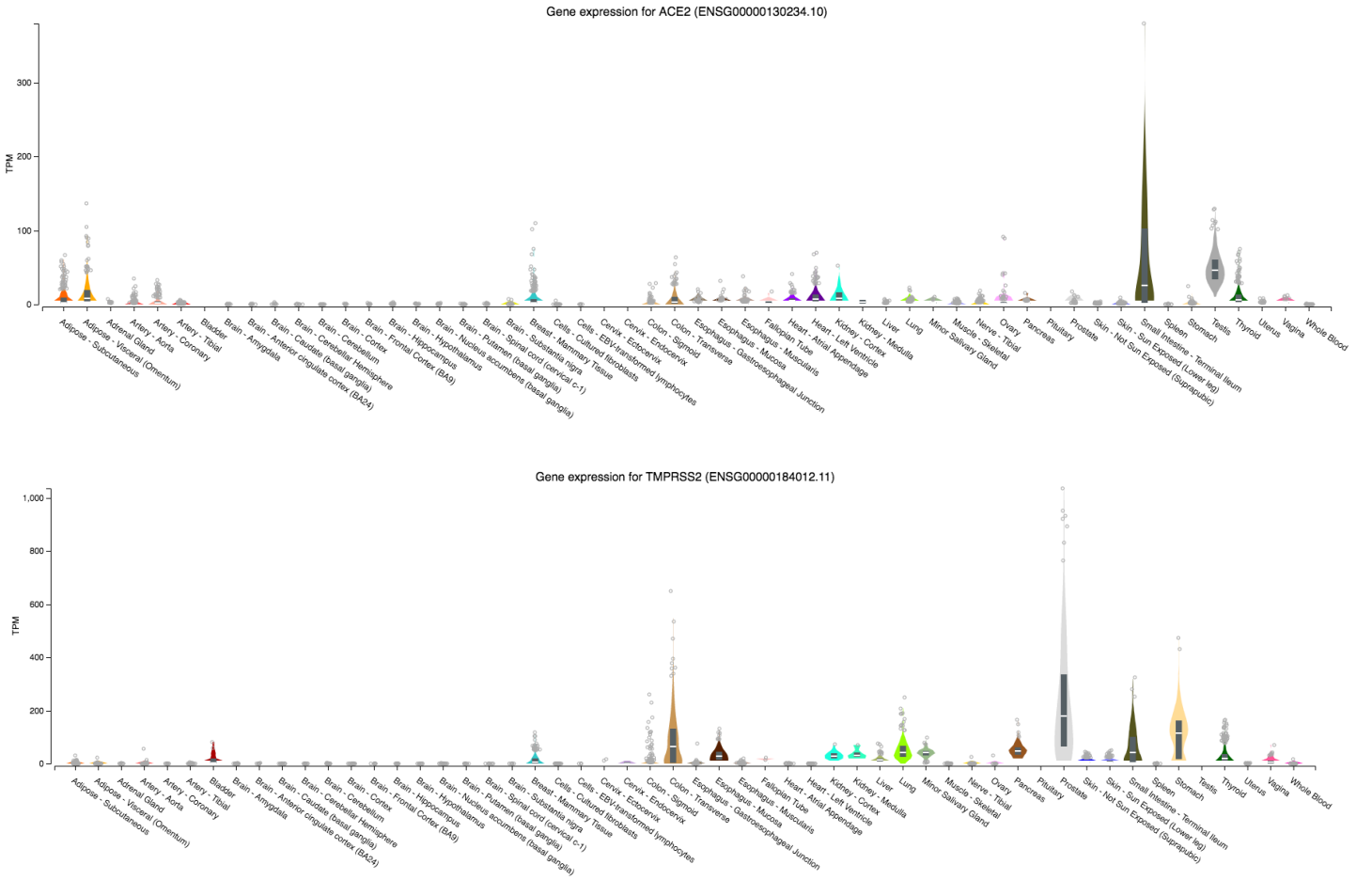


Supplementary Fig. 4: High variability in TMPRSS2 and ACE2 expression within human population. Data from GTEx consortium (v7), samples split by sex and age group. Expression values represented in transcripts per million (TPM)

Supplementary Fig. 5: Expression of CTSB and CTSL in lung across demographic groups

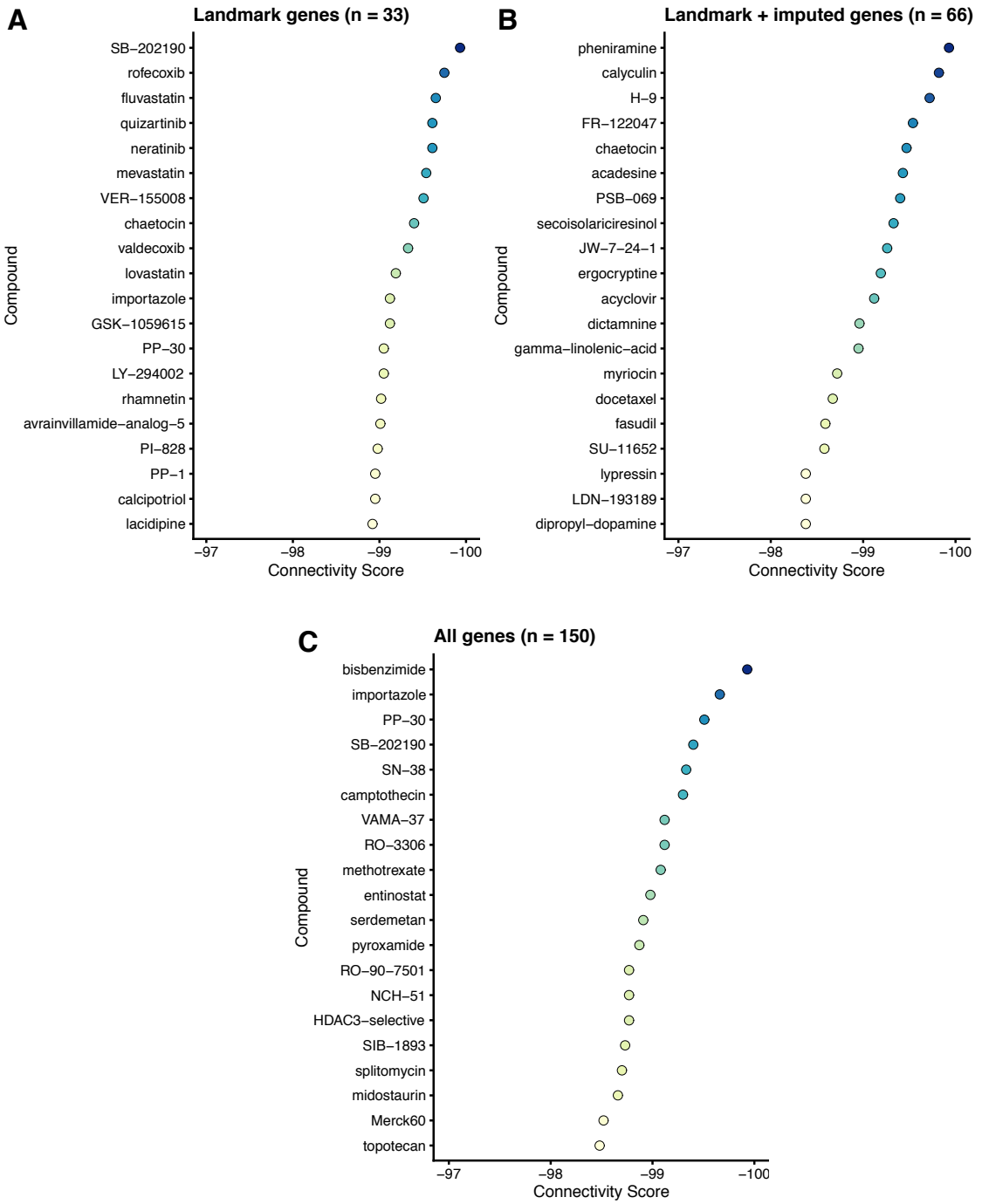


Supplementary Fig. 6: ACE2 and TMPRSS2 have restricted expression patterns across human tissues



Supplementary Fig. 6: ACE2 and TMPRSS2 have restricted expression patterns across human tissues. Data taken from the GTEx project (v8).

Supplementary Fig. 7: Top hits from Connectivity Map for compounds that down-regulate host proteins that interact with SARS-CoV-2 proteins



Supplementary Fig. 7: Top hits from Connectivity Map for compounds that down-regulate host proteins that interact with SARS-CoV-2 proteins. Top compounds shown for three comparisons: landmark genes (panel A, directly assayed by L1000 array), landmark + top imputed genes (panel B), and all genes (top 150 by fold-change). Gene number per analysis listed in plot titles, strongest down-regulator compounds have scores closest to -100.