# Multiomics Evaluation of Gastrointestinal and Other Clinical Characteristics of COVID-19

Mulong Du,[1,2,3] Guoshuai Cai,[4] Feng Chen,[1,2] David C. Christiani,[3,5] Zhengdong Zhang,[6,7] and Meilin Wang[6,7]

[1]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China; [2]China International Cooperation Center for Environment and Human Health, Nanjing Medical University, Nanjing, China; [3]Departments of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; [4]Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina; [5]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; [6]Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China; and [7]Department of Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

Since December 2019, coronavirus disease 2019 (COVID-19) outbreak caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has produced a worldwide panic. Beyond the principal human-to-human transmission method by droplet and contact, there is still limited knowledge about possible alternate transmission methods to guide clinical care. Recent clinical studies have observed digestive symptoms in patients with COVID-19,[1] possibly because of the enrichment and infection of SARS-CoV-2 in the gastrointestinal tract, mediated by virus receptor of angiotensin converting enzyme 2 (ACE2),[2] which suggests the potential for a fecal-oral route of SARS-CoV-2 transmission.[3,4] However, there is still a large gap in the biological knowledge of COVID-19. In this study, via a bulk-to-cell strategy focusing on ACE2, we performed an integrated omics analysis at the genome, transcriptome, and proteome levels in bulk tissues and single cells across species to decipher the potential routes for SARS-CoV-2 infection in depth.

## Methods

Clinical and epidemiologic data of patients with COVID-19 were collected from a continually updated resource.[5] The transcriptome and proteome derived from bulk tissues and cells were accessed from multiple databases. A phenome-wide association study data set was supplied for genetic analysis on the ACE2 pathway. *P* values were calculated from *t* test and gene set analysis. More details are shown in Supplementary Methods.

## Results

### Clinical Symptoms of Coronavirus Disease 2019 at Diagnosis

We constructed a user-friendly interface for the visualization of clinical symptoms of COVID-19 (https://mulongdu.shinyapps.io/map_covid/; Supplementary Figure 1). Fever was the most common symptom at onset of illness (>70%). Notably, 5.13% and 3.34% of patients had recorded digestive symptoms from Hubei and the outside (Supplementary Table 1), respectively; 1.67% of asymptomatic carriers were recorded with positive SARS-CoV-2.

### ACE2 Expression Pattern in Bulk Tissues

As shown in Figure 1, *ACE2* was widely expressed across tissues. *ACE2* was considered intestine specific because its expression was enriched more than 4-fold in the intestinal tract (Figure 1A) compared with other tissues. The protein detection supported the activity of ACE2 in digestive, excretory, and reproductive organs (Figure 1A). A similar expression pattern could be found in extra data sets (Figure 1B–D). However, *ACE2* expression was mild in the lung.

### ACE2 Expression Pattern in Cells

We further performed cell-specific analysis to decipher the expression pattern of *ACE2* in target organs. *ACE2* messenger RNA (mRNA) in the intestinal cell lines was significantly higher than in the lung cell lines ($P = 2.88 \times 10^{-6}$) (Figure 1E) but did not differ significantly between the kidney and lung cell lines. Moreover, the mean values of ACE2 protein in both the intestinal (0.31) and kidney (0.61) cell lines were higher than those in the lung (–0.26), although there were no significant differences among the cells (Figure 1F).

Most current article

BRIEF COMMUNICATIONS
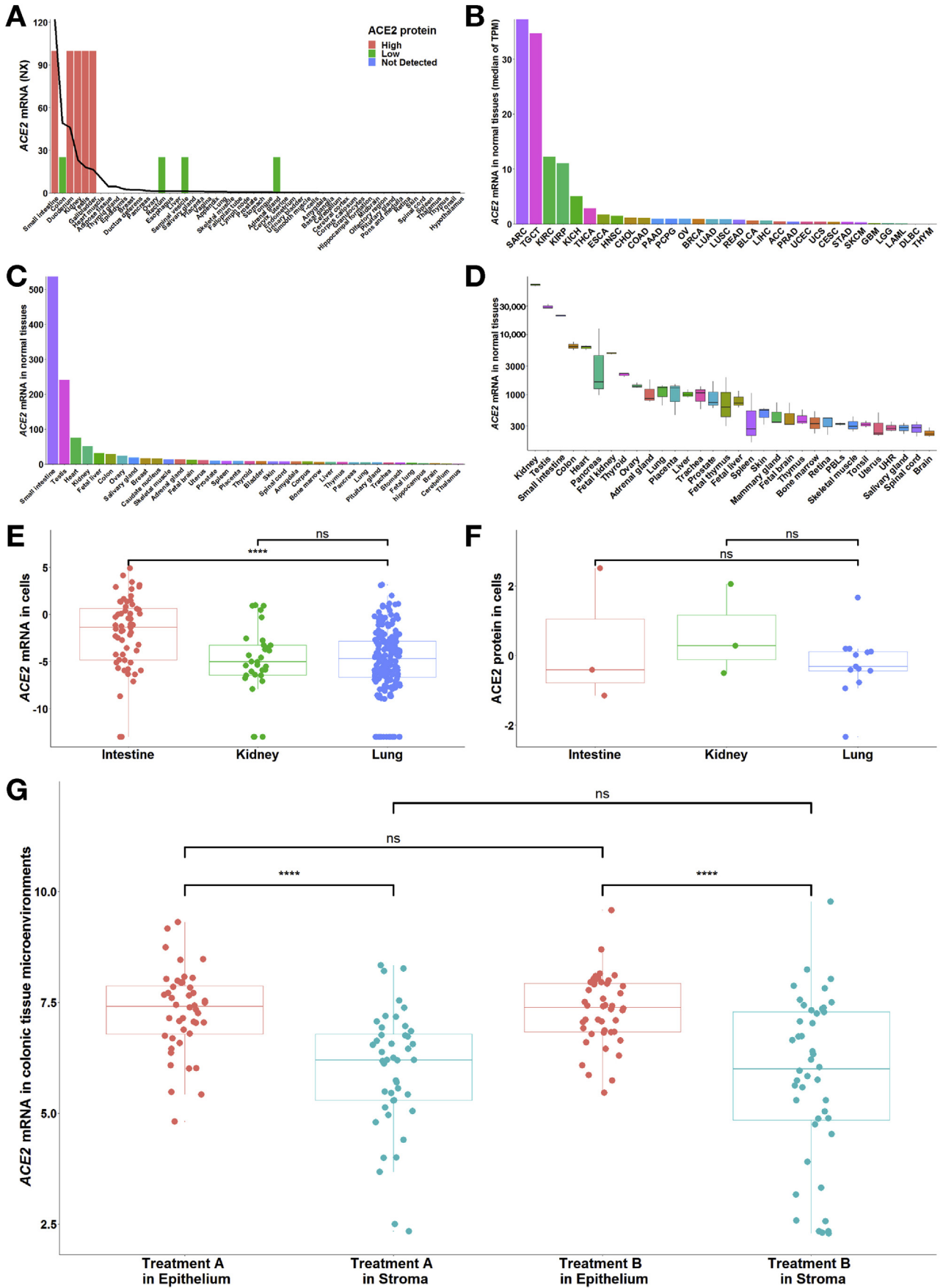
In colonic tissue microenvironments, *ACE2* exhibited significantly higher expression in epithelia than in stroma ($P = 4.60 \times 10^{-8}$ in treatment A; $P = 8.81 \times 10^{-7}$ in treatment B) (Figure 1*G*); however, intriguingly, there was no significant difference in *ACE2* mRNA between aspirin intervention and placebo (Figure 1*G*).

Subsequently, we carried out single-cell analysis to dissect the *ACE2* expression pattern. *ACE2* was enriched specifically in the enterocytes of mice small intestine epithelia (Supplementary Figure 2*A* and *B*), which was consistent with the findings in humans.[6] *ACE2* was highly concentrated in epithelia at the renal proximal tubule in both humans and mice (Supplementary Figure 2*C–E*).

### Genetic Effect of the Angiotensin Converting Enzyme 2 Pathway on Clinical Symptoms

We obtained 7 phenotypes related to the intestinal tract and kidney deposited in the phenome-wide association study data set and extracted genome-wide association study summary statistics with more than 3000 single-nucleotide polymorphisms assigned to 33 genes of the *ACE2* pathway for the gene set analysis (Supplementary Table 2 and Supplementary Figure 3). In the pathway-based level, we found no significant genetic association of the *ACE2* pathway with 7 phenotypes but a modest performance of the prediction model in nephrotic syndrome (area under the receiver operating characteristic curve, 0.607) (Supplementary Table 3). Partitioned to the gene-based level, the significant joint effect of each gene extended across different phenotypes (Supplementary Table 3). In the disorders of the digestive and excretory systems, *TGFB1* was associated with colorectal cancer, *ACE* and *MAPK3* with nephrotic syndrome, and *KLK1* and *KNG1* with urolithiasis. Similarly, in the blood test parameters, *ACE2*, along with *ENPEP*, exhibited a significant association only with Alb.

## Discussion

On the basis of prior evidence suggesting that ACE2 mediated the entry of SARS-CoV-2 into cells,[2] we previously found that *ACE2* expression was significantly higher in smokers than in nonsmokers, especially in distinct lung cell types.[7] This was corroborated by the clinical observation that the patients with severe cases of COVID-19 were more likely to have a smoking history (22.1%) than those with nonsevere COVID-19 (13.1%).[1] In this study, *ACE2* was confirmed to be enriched in the epithelia of the intestinal tract; therefore, a mutual interaction potentially occurred such that SARS-CoV-2 disrupted ACE2 activity, infected the intestinal epithelium by its cytotoxicity, and shed into feces, resulting in gastrointestinal manifestations and/or positive

**Figure 1.** Expression pattern of *ACE2* across tissues and cells in the human body. (*A*) At the mRNA level, in combined data from Human Protein Atlas, Genotype Tissue Expression, and Functional Annotation of The Mammalian Genome, the top 5 tissues included those from 3 digestive organs (intestinal tract), along with the kidney and testis, indicating the intestine specificity of *ACE2* (with normalized expression values of 122, 49.1 and 46 in small intestine, colon, and duodenum, respectively). The line represents the normalized expression value of *ACE2* across all tissues. At the protein level, in samples from Human Protein Atlas, 9 tissues were stained with ACE2, among which 5 tissues were highly stained, including tissues from 2 digestive organs (duodenum and small intestine), and 4 tissues were minimally stained, including tissues from digestive organs (colon and rectum). Furthermore, similar to *ACE2* mRNA, ACE2 protein was enriched in the kidney and testis. The bars indicate the immunohistochemical results for ACE2 protein. Antibody staining was reported as not detected, low, medium, or high, and the score was based on the staining intensity and fraction of stained cells. (*B–D*) Among the normal tissues with *ACE2* mRNA expression from (*B*) The Cancer Genome Atlas (TCGA) and (*C*, *D*) Gene Expression Omnibus, the top tissues also included those from the intestine, kidney, and testis. We excluded SARC in TCGA because it contained only 2 samples of normal tissues. (*C*) GSE2361-supplied gene expression profiles in 36 normal human tissues with only 1 sample per tissue. (*D*) GSE7905-collected gene expression profiles from 31 normal tissues and a Universal Human Reference RNA, which used 3 replicates per tissue for a total of 96 samples. (*E*) At the mRNA level, *ACE2* expression in the intestinal cell lines (n = 60; –2.08 ± 4.02) was higher than that in the lung cell lines (n = 194; –5.00 ± 3.51; ****$P = 2.88 \times 10^{-6}$), but there was no significant difference between the kidney (n = 32; –4.92 ± 3.59) and lung cell lines. (*F*) At the protein level, the differences in ACE2 expression among the intestine (n = 3; 0.31 ± 1.94), kidney (n = 3; 0.61 ± 1.31) and lung (n = 13; –0.26 ± 0.89) cell lines were not statistically significant. Values are reported as mean ± standard deviation. Both mRNA and protein expression levels were adjusted by the global normalization. (*G*) GSE71571-provided gene expression profiles in the epithelial and stromal cells of normal colon in 44 healthy individuals recruited for an aspirin intervention trial. Briefly, participants were randomly assigned to the order in which they received treatment A or treatment B, respectively. (The type of treatment included active [aspirin] and placebo but was masked. The study authors must be contacted directly for unblinding). A paired *t* test was used for the comparison of *ACE2* mRNA between epithelial and stromal cells (****$P = 4.60 \times 10^{-8}$ in treatment A; ****$P = 8.81 \times 10^{-7}$ in treatment B). An unpaired *t* test was applied for the comparison of *ACE2* mRNA between treatment A and B. More details are provided in the Supplementary Methods. ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; DLBC, Lymphoid neoplasm diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and Neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower-grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MESO, mesothelioma; ns, no significance; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THY, thymoma; TPM, transcripts per million; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; UHR, Universal Human Reference RNA; UVM, uveal melanoma.

SARS-CoV-2 in stool.[4,8] Considering the physiologic renewal of intestinal epithelia every 4–5 days, our results warn that more attention must be given to the possibility of fecal-oral transmission of SARS-CoV-2, especially by asymptomatic carriers.

The renal proximal tubule enriched in *ACE2* indicated that viral shedding in the urine was feasible; however, no evidence supported the actual detection of SARS-CoV-2 in the urine.[4] Nevertheless, renal impairment was common in patients with severe COVID-19,[1] which could be supported by the potential that SARS-CoV-2 damaged renal tubular cells and induced the disruption of the ACE2 pathway referring to *ACE2*, *ACE*, *ENPEP*, *TGFB1*, *THOP1*, *MAS1*, and *NLN* involved in kidney dysfunction.

In summary, *ACE2* enriched in the intestinal tract and kidney—more specifically, in the epithelium—could mediate the entry of SARS-CoV-2 into cells to accumulate and cause cytotoxicity but does not respond to nonsteroidal anti-inflammatory drugs. It is reasonable to emphasize the monitoring of digestive and excretory system complications in patients with COVID-19 and the possibility of SARS-CoV-2 transmission via the fecal-oral route by individuals with suspected infection and asymptomatic carriers (Supplementary Figure 4).

## Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2020.03.045.

### References

1. Guan WJ, et al. N Engl J Med 2020;382:1708–1720.
2. Hoffmann M, et al. Cell 2020;181:271–280.
3. Gu J, et al. Gastroenterology 2020;158:1518–1519.
4. Xiao F, et al. Gastroenterology 2020;158:1831–1833.
5. Xu B, et al. Lancet Infect Dis 2020;20:534.
6. Zhang H, et al. bioRxiv 2020:2020.01.30.927806. Available at: https://www.biorxiv.org/content/10.1101/2020.01.30.927806v1. Accessed May 22, 2020.
7. Cai G, et al. Am J Respir Crit Care Med 2020.
8. Xu Y, et al. Nat Med 2020;26:502–505.

**Correspondence**
Address correspondence to: Zhengdong Zhang, MD, PhD, Department of Environmental Genomics, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Jiangning District, Nanjing 211166, China. e-mail: drzdzhang@njmu.edu.cn; fax: +86 25 86868499; or Meilin Wang, MD, PhD, Department of Environmental Genomics, School of Public Health, Nanjing Medical University, 101 Longmian Avenue, Jiangning District, Nanjing 211166, China. e-mail: mwang@njmu.edu.cn; fax: +86 25 86868499.

**BRIEF COMMUNICATIONS**

# Supplementary Methods

## Clinical and Epidemiologic Information

The clinical and epidemiologic data of patients with COVID-19 were collected from a continually updated resource (https://tinyurl.com/s6gsq5y),[1] including 26,357 cases worldwide. The analyzed data set was downloaded on Feb 23, 2020, which enrolled 14,729 confirmed cases from Hubei, Wuhan province, and 11,628 cases from the outside of Hubei around the world. The visualization of symptoms of COVID-19 cases worldwide was built using *plotly* and *shiny* packages in the R library (R Core Team, Vienna, Austria).

## Bulk Tissue Resources

The consensus normalized expression value of *ACE2* mRNA was extracted from the combined transcriptomics data sets by Human Protein Atlas,[2] Genotype Tissue Expression,[3] and Functional Annotation of The Mammalian Genome,[4] which enrolled 55 tissue types and 6 blood cell types. Briefly, the Human Protein Atlas provided transcriptome and proteome profiles across approximately 40 kinds of cells, tissues, and organs in the human body and grouped genes into categories describing the tissue specificity. The Genotype Tissue Expression project collected samples from 54 nondiseased tissue sites from nearly 1000 individuals. The Functional Annotation of The Mammalian Genome project supplied transcriptome data for every major human organ and more than 200 cancer cell lines.

The Cancer Genome Atlas and Gene Expression Omnibus also supplied transcriptome profiles derived from bulk tissues. We extracted *ACE2* mRNA in normal tissues from both The Cancer Genome Atlas, spanning 33 cancer types deposited in the Gene Expression Profiling Interactive Analysis platform,[5] and the Gene Expression Omnibus, including major organs (GSE2361,[6] GSE7905,[7] and GSE71571[8]). We further grouped cells into the subtypes intestine, kidney, and lung according to the cell progenitor using the Cancer Cell Line Encyclopedia data set and performed differential expression analysis for ACE2 mRNA and protein across each subgroup using the *t* test. The Cancer Cell Line Encyclopedia project provided omics data, including transcriptome[9] and proteome,[10] covering 1100 cell lines.

## Single-Cell RNA Sequencing Analysis

We used the Single Cell Expression Atlas (https://www.ebi.ac.uk/gxa/sc/home) and Kidney Interactive Transcriptomics (http://humphreyslab.com/SingleCell/) to visualize the expression pattern of *ACE2* mRNA in each cell cluster of intestinal tract and kidney. Briefly, single-cell RNA sequencing of intestinal and kidney tissues were carried out using the InDrop, DropSeq, or 10× Chromium platforms, and the distribution of *ACE2* was shown in the cell subgroups derived from both humans (including 4524 nuclei in complete kidney tissues[11] and 4259 nuclei in kidney epithelia[12]) and mice (including 1522 cells in small intestinal epithelia[13] and 11,395 cells and nuclei in kidney epithelia[14]).

## Selection of Angiotensin Converting Enzyme 2 Pathways and Genetic Variants

The key pathways containing *ACE2* were accessed from the Kyoto Encyclopedia of Genes and Genomes (Renin-angiotensin system; hsa04614) and PathCards (ACE Inhibitor Pathway, Pharmacodynamics SuperPath). Considering that TMPRSS2 could cooperate with ACE2 to enable the virus to enter the host cells,[15] we ultimately included 33 genes in the *ACE2* network for further genetic analysis (Supplementary Table 2), among which *ACE2*, *ATP6AP2*, and *AGTR2* were located on chromosome X. The ACE2 pathway network was constructed using STRING (https://string-db.org/) (Supplementary Figure 3). Genomic information for the candidate genes was extracted from the National Center for Biotechnology Information assembly by GRCh 37 (Supplementary Table 2). The 1000 Genomes Projects (phase 3) provided the genetic information for each gene, along with the following quality control criteria for selecting genetic variants underlying east Asian ancestry (CHB: Han Chinese in Beijing, China; JPT: Japanese in Tokyo, Japan): minor allele frequency >0.01, *P* value for Hardy-Weinberg equilibrium $>1.0 \times 10^{-6}$, and genotyping call rate >95%.

## Analysis of Phenome-wide Association Studies

We initially used the largest Asian phenome-wide association study database, the Japanese Encyclopedia of Genetic Associations by Riken,[16,17] to evaluate the genetic effect of the ACE2 pathway on the clinical symptoms of COVID-19 and the disease process that occurs in *ACE2*-enriched tissues. More than 200 traits or diseases with related genome-wide association studies were deposited in the Japanese Encyclopedia of Genetic Associations by Riken, of which the summary statistics were calculated by logistic or linear regression analysis for the association between each single nucleotide polymorphism and phenotypes. When mapping the phenotypes related to the intestinal tract and kidney deposited in the phenome-wide association study data set, we used colorectal cancer, nephrotic syndrome, and urolithiasis for candidate digestive and excretory system diseases. In addition, we considered the blood test parameters, including the albumin/globulin ratio and the Albumin, C-reactive protein, and total protein concentrations, as reflections of patients' health conditions in terms of the clinical symptoms of COVID-19.[18]

To avoid unexplainable and weak effects of a single genetic variant on a phenotype, we thus performed gene set analysis using Multi-marker Analysis of genomic Annotation (MAGMA),[19] which aggregated genetic variants into candidate gene or pathway sets, to evaluate and predict the phenotypes susceptibility. We also used SummaryAUC to evaluate the performance of polygenic risk prediction models in binary outcomes underlying summary statistics.[20]

## Ethics Approval

All public databases involving human participants were approved by the ethics committees of original studies, and this study was approved by the institutional review board of Nanjing Medical University (NJMUIRB -2020-020).

(A)

Symptoms of COVID-19
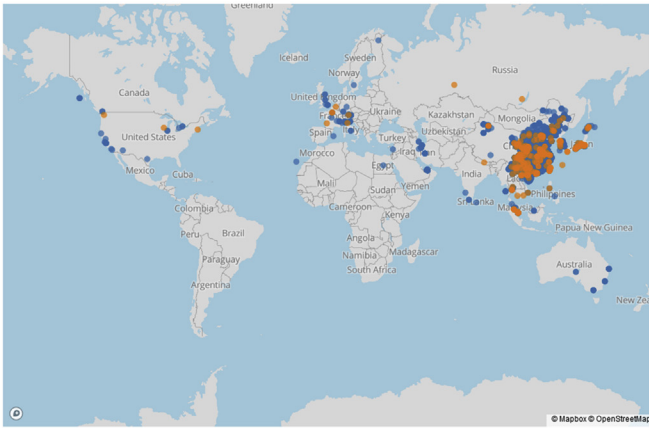


(B)

Digestive symptoms of COVID-19



● no record  ● record
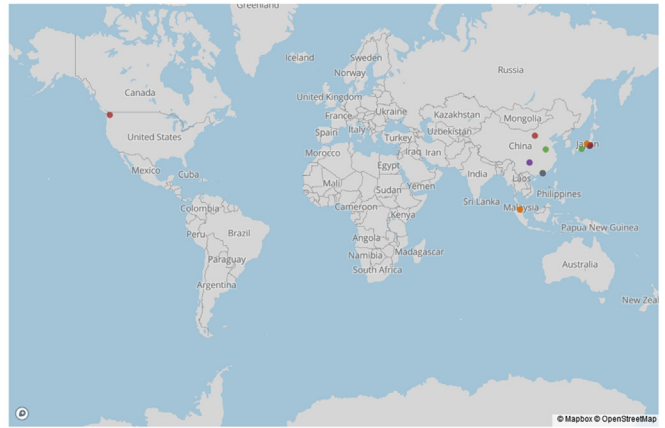
● Anorexia  ● Asymptomatic  ● Diarrhea  ● Nausea  ● Vomiting/emesis

**Supplementary Figure 1.** The global distribution of COVID-19 cases. The data set was downloaded on February 23, 2020 (https://tinyurl.com/s6gsq5y). (*A*) On that date, a total of 26,357 cases were distributed around the world, among which 398 cases (39 cases derived from Hubei and 359 from outside Hubei) recorded with complete clinical symptoms at the onset of illness are marked in orange, and the cases without recorded clinical symptoms are shown in blue. (*B*) Patients with symptoms in digestive system or asymptomatic carriers are marked in varying colors. A Web site location provided a user-friendly interface for visualization of clinical symptoms (https://mulongdu.shinyapps.io/map_covid/).
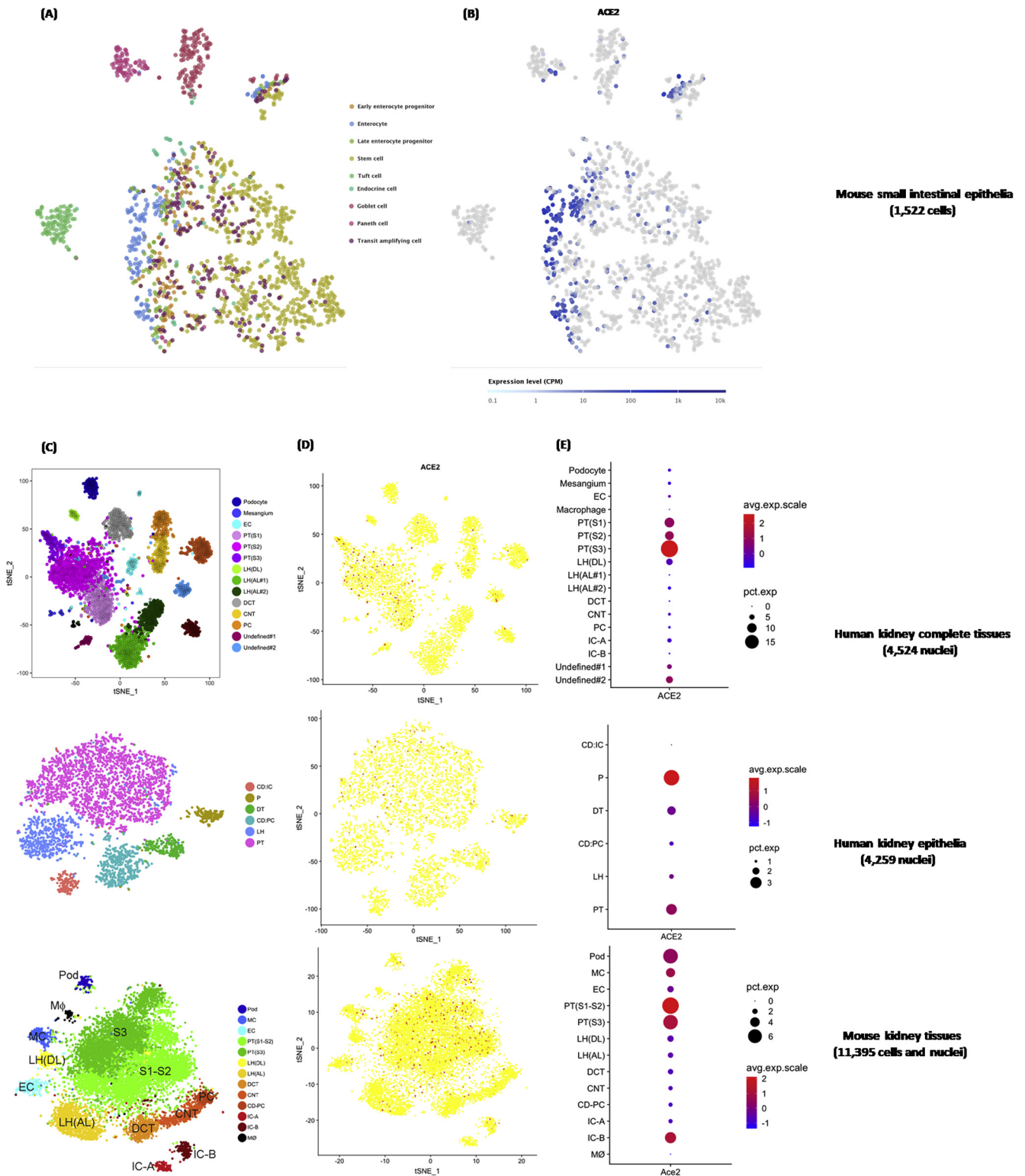
**Supplementary Figure 2.** Single-cell RNA sequencing data showing the *ACE2* expression pattern in the intestinal tract and kidney. (*A*) tSNE plot of small intestinal epithelial cell subgroups. Cells were partitioned into 9 groups: early enterocyte progenitor, enterocyte, late enterocyte progenitor, stem cell, tuft cell, endocrine cell, goblet cell, Paneth cell, and transit amplifying cell. (*B*) Cells expressing *ACE2* expression are colored blue, indicating enrichment of *ACE2* in intestinal enterocytes. (*C*) tSNE plot of cell subgroups for each kidney data set. Cells from human and mouse kidney tissues were grouped by kidney anatomy: ascending limb (AL), collecting duct–principal cell (CD-PC), connecting tubule (CNT), distal convoluted tubule (DCT/DT), descending limb (DL), endothelial cell (EC), intercalated cell (IC), loop of Henle (LH), mesangial cell (MC), macrophage (MΦ), podocyte (Pod/P), and proximal tubule (PT). (*D*) Cells expressing *ACE2* are colored red, indicating enrichment of *ACE2* in the proximal tubule of the kidney. (*E*) The abundance of *ACE2* expression across each cell subgroup in human and mouse kidneys. avg, average; exp, expression; CPM, counts per million; pct, percent; tSNE, t-distributed stochastic neighbor embedding.
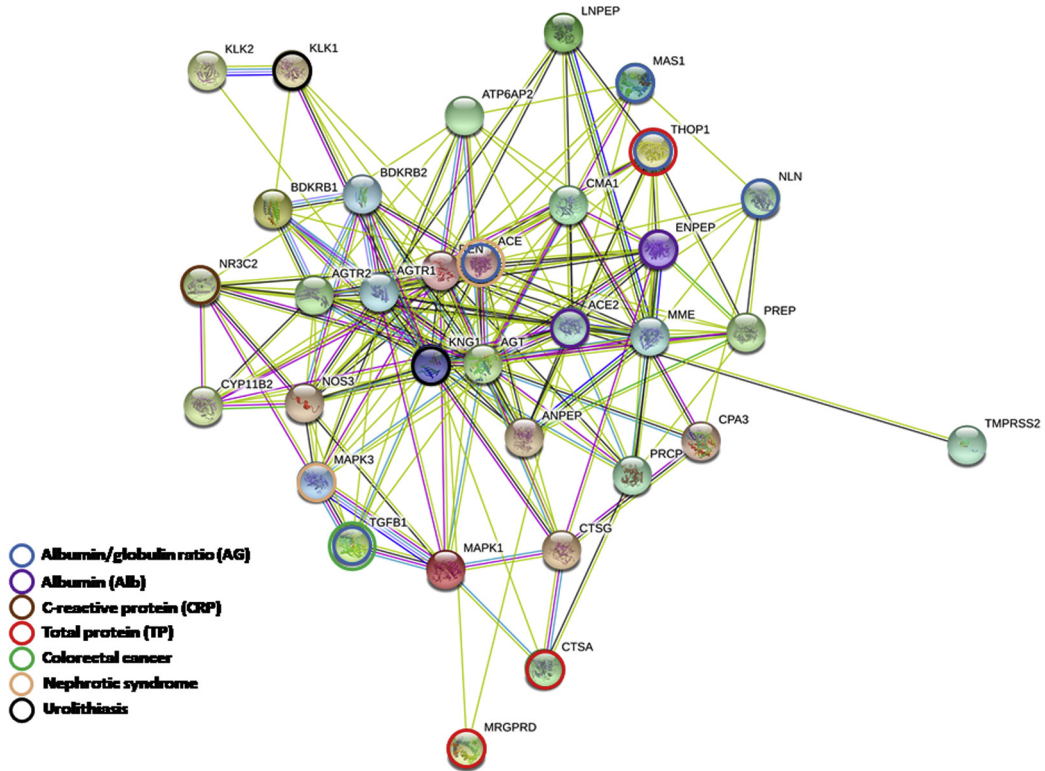
**Supplementary Figure 3.** Network of genes in the ACE2 pathway. The network was constructed by STRING (https://string-db.org/) with the default parameters.
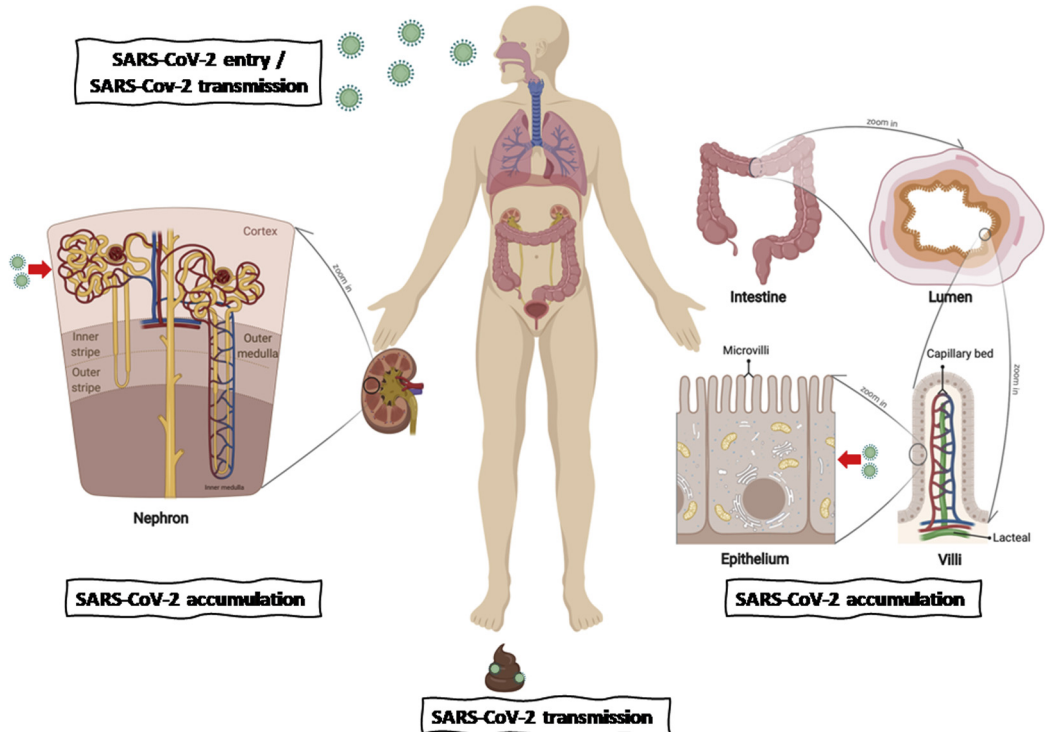


**Supplementary Figure 4.** Schematic of the bulk-to-cell strategy for evaluating SARS-CoV-2 infection, accumulation, and transmission across hosts. The diagram was constructed with BioRender (https://biorender.com/).

**Supplementary Table 1.** Clinical Characteristics of Patients Infected With SARS-CoV-2

| Characteristics | Patients in Hubei | | Patients outside of Hubei | |
| --- | --- | --- | --- | --- |
| | n = 39 | % | n = 359 | % |
| Age, y | | | | |
| <18 | 0 | 0.00 | 12 | 3.34 |
| 18–29 | 0 | 0.00 | 44 | 12.26 |
| 30–39 | 1 | 2.56 | 73 | 20.33 |
| 40–49 | 2 | 5.13 | 73 | 20.33 |
| 50–59 | 4 | 10.26 | 65 | 18.11 |
| 60–69 | 11 | 28.21 | 56 | 15.60 |
| 70–79 | 9 | 23.08 | 14 | 3.90 |
| ≥80 | 11 | 28.21 | 5 | 1.39 |
| Sex | | | | |
| Male | 26 | 66.67 | 201 | 55.99 |
| Female | 13 | 33.33 | 148 | 41.23 |
| Countries | | | | |
| Belgium | | | 1 | 0.28 |
| Cambodia | | | 1 | 0.28 |
| China | | | 256 | 71.31 |
| France | | | 3 | 0.84 |
| Germany | | | 1 | 0.28 |
| Italy | | | 1 | 0.28 |
| Japan | | | 52 | 14.48 |
| Malaysia | | | 7 | 1.95 |
| Nepal | | | 1 | 0.28 |
| Philippines | | | 1 | 0.28 |
| Russia | | | 2 | 0.56 |
| Singapore | | | 9 | 2.51 |
| South Korea | | | 9 | 2.51 |
| Thailand | | | 5 | 1.39 |
| United States | | | 3 | 0.84 |
| Vietnam | | | 7 | 1.95 |
| Symptoms | | | | |
| Fever | 31 | 79.49 | 266 | 74.09 |
| Cough | 21 | 53.85 | 128 | 35.65 |
| Fatigue | 5 | 12.82 | 16 | 4.46 |
| Digestive (diarrhea, nausea, vomiting/emesis, anorexia) | 2 | 5.13 | 12 | 3.34 |
| Asymptomatic | 0 | 0.00 | 6[a] | 1.67 |

[a]3 patients were from Japan, and 3 were from Malaysia.

**Supplementary Table 2.** Information on 33 Genes in the ACE2 Pathway

| Gene name | Gene ID | Chromosome | Position | Strand |
|-----------|---------|------------|----------|--------|
| *REN* | 5972 | 1 | 204123944-204135465 | − |
| *AGT* | 183 | 1 | 230838269-230850336 | − |
| *AGTR1* | 185 | 3 | 148415658-148460790 | + |
| *CPA3* | 1359 | 3 | 148583043-148614874 | + |
| *MME* | 4311 | 3 | 154797436-154901518 | + |
| *KNG1* | 3827 | 3 | 186435098-186462199 | + |
| *ENPEP* | 2028 | 4 | 111397229-111484493 | + |
| *NR3C2* | 4306 | 4 | 148999915-149365850 | − |
| *NLN* | 57486 | 5 | 65018023-65125111 | + |
| *LNPEP* | 4012 | 5 | 96271346-96365115 | + |
| *PREP* | 5550 | 6 | 105725442-105850999 | − |
| *MAS1* | 4142 | 6 | 160320218-160329339 | + |
| *NOS3* | 4846 | 7 | 150688144-150711687 | + |
| *CYP11B2* | 1585 | 8 | 143991975-143999259 | − |
| *MRGPRD* | 116512 | 11 | 68747490-68748455 | − |
| *PRCP* | 5547 | 11 | 82535409-82612733 | − |
| *CMA1* | 1215 | 14 | 24974712-24977471 | − |
| *CTSG* | 1511 | 14 | 25042724-25045466 | − |
| *BDKRB2* | 624 | 14 | 96671016-96710666 | + |
| *BDKRB1* | 623 | 14 | 96721641-96735304 | + |
| *ANPEP* | 290 | 15 | 90328126-90358119 | − |
| *MAPK3* | 5595 | 16 | 30125426-30134630 | − |
| *ACE* | 1636 | 17 | 61554422-61575741 | + |
| *THOP1* | 7064 | 19 | 2785464-2813599 | + |
| *TGFB1* | 7040 | 19 | 41836812-41859831 | − |
| *KLK1* | 3816 | 19 | 51322402-51327043 | − |
| *KLK2* | 3817 | 19 | 51376689-51383823 | + |
| *CTSA* | 5476 | 20 | 44519591-44527459 | + |
| *TMPRSS2* | 7113 | 21 | 42836236-42880085 | − |
| *MAPK1* | 5594 | 22 | 22113946-22221970 | − |
| *ACE2* | 59272 | X | 15579156-15620192 | − |
| *ATP6AP2* | 10159 | X | 40440141-40465889 | + |
| *AGTR2* | 186 | X | 115301958-115306225 | + |

ID, identification.

**Supplementary Table 3.** Gene Set Analysis to Evaluate the Genetic Effect of the ACE2 Pathway on the 7 Phenotypes

| Sources | Phenotypes | Pathway name | Number of genes/SNPs | Sample size, N | $P_{\text{pathway based}}$ | AUC (variance) |
|---|---|---|---|---|---|---|
| Disease | Colorectal cancer | ACE2 pathway | 33 | 202,807 | .920 | 0.542 ($1.21 \times 10^{-5}$) |
| | Nephrotic syndrome | ACE2 pathway | 33 | 212,453 | .933 | 0.607 ($8.24 \times 10^{-5}$) |
| | Urolithiasis | ACE2 pathway | 33 | 212,453 | .743 | 0.537 ($1.29 \times 10^{-5}$) |
| Blood test | Albumin | ACE2 pathway | 33 | 102,223 | .306 | NA |
| | Albumin/globulin ratio | ACE2 pathway | 33 | 98,626 | .349 | |
| | C-reactive protein | ACE2 pathway | 33 | 75,391 | .365 | |
| | Total protein | ACE2 pathway | 33 | 113,509 | .541 | |
| | | Gene name | Number of SNPs | | | |
| Disease | Colorectal cancer | TGFB1 | 23 | 202,807 | .028 | NA |
| | Nephrotic syndrome | ACE | 37 | 212,453 | .012 | |
| | Nephrotic syndrome | MAPK3 | 3 | 212,453 | .012 | |
| | Urolithiasis | KLK1 | 13 | 212,453 | .001 | |
| | Urolithiasis | KNG1 | 146 | 212,453 | .002 | |
| Blood test | Albumin | ENPEP | 130 | 102,223 | .020 | NA |
| | Albumin | ACE2 | 18 | 102,223 | .029 | |
| | Albumin/globulin ratio | ACE | 33 | 98,626 | .012 | |
| | Albumin/globulin ratio | TGFB1 | 23 | 98,626 | .013 | |
| | Albumin/globulin ratio | THOP1 | 36 | 98,626 | .020 | |
| | Albumin/globulin ratio | MAS1 | 13 | 98,626 | .032 | |
| | Albumin/globulin ratio | NLN | 253 | 98,626 | .037 | |
| | C-reactive protein | NR3C2 | 624 | 75,391 | .021 | |
| | Total protein | MRGPRD | 1 | 113,509 | .008 | |
| | Total protein | CTSA | 16 | 113,509 | .012 | |
| | Total protein | THOP1 | 36 | 113,509 | .012 | |

AUC, area under the receiver operating characteristic curve (calculated by SummaryAUC); NA, not available; SNP, single nucleotide polymorphism.

## Supplementary References

1. Xu B, et al. Lancet Infect Dis 2020;20:534.
2. Uhlen M, et al. Science 2015;347:1260419.
3. Nat Genet 2013;45:580–585.
4. Yu NY, et al. Nucleic Acids Res 2015;43:6787–6798.
5. Tang Z, et al. Nucleic Acids Res 2017;45:W98–W102.
6. Ge X, et al. Genomics 2005;86:127–141.
7. Dezso Z, et al. BMC Biol 2008;6:49.
8. Thomas SS, et al. Genom Data 2015;6:154–158.
9. Barretina J, et al. Nature 2012;483:603–607.
10. Nusinow DP, et al. Cell 2020;180:387–402.
11. Wu H, et al. Cell Stem Cell 2018;23:869–881.
12. Wu H, et al. J Am Soc Nephrol 2018;29:2069–2080.
13. Haber AL, et al. Nature 2017;551:333–339.
14. Wu H, et al. J Am Soc Nephrol 2019;30:23–32.
15. Hoffmann M, et al. Cell 2020;181:271–280.
16. Kanai M, et al. Nat Genet 2018;50:390–400.
17. Ishigaki K, et al. bioRxiv 2019:795948.
18. Chen N, et al. Lancet 2020.
19. de Leeuw CA, et al. PLoS Comput Biol 2015;11: e1004219.
20. Song L, et al. Bioinformatics 2019;35:4038–4044.