cmgh
CELLULAR AND MOLECULAR GASTROENTEROLOGY AND HEPATOLOGY

RESEARCH LETTER

Identification of Transcription Factors Regulating SARS-CoV-2 Entry Genes in the Intestine

Gastrointestinal symptoms of coronavirus disease 2019 (COVID-19), including diarrhea, nausea, and vomiting, are more common than previously thought. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) likely causes these symptoms by infecting the epithelial cells lining the gastrointestinal tract,[1] with angiotensin I converting enzyme 2 (ACE2) functioning as the viral receptor[2] and transmembrane serine protease 2 (TMPRSS2) functioning in viral spike protein priming.[3] Du et al[4] recently reported that ACE2 not only is expressed in lung alveolar type II (AT2) cells but also highly expressed in absorptive enterocytes. However, the regulatory mechanisms and transcription factors driving expression of *Ace2* and *Tmprss2* remains unclear. Using epigenomic approaches and mouse genetic models, we identify 4 key transcriptional regulators (caudal-type homeobox 2 (CDX2), hepatocyte nuclear factor 4 (HNF4), Smad family member 4 (SMAD4), or GATA binding proteins) that bind to the loci of these genes, alter chromatin looping, shape epigenetic modifications, and, ultimately, show a dramatic impact on *Ace2* and *Tmprss2* gene expression upon transcription factor knockout.

We began by investigating the expression of COVID-19–related host genes throughout the body. Chromatin accessibility and RNA transcript levels show a tissue-specific expression pattern for *Ace2* and *Tmprss2*, with greatest expression observed in intestine, kidney, and lung tissues (Figure 1*A* and *B* and Supplementary Figure 1). *Ace2* and *Tmprss2* are expressed more robustly in isolated intestinal epithelium compared with the remaining subepithelium (Figure 1*C*). Temporally, transcript levels of *Ace2* and *Tmprss2* increase during embryonic development (Figure 1*D* and *F*). In the adult tissue, *Ace2* transcripts are increased in the villus compared with the crypt (Figure 1*E* and *F*). We next focused on the epithelial cell populations expressing the SARS-CoV-2–related genes. single cell RNA sequencing analysis defined cell populations within crypt epithelium that express markers of stem, progenitor, and differentiating epithelial cells (Supplementary Figure 2*A*). Cells expressing canonic goblet, Paneth, tuft, enteroendocrine, or enterocyte lineage markers were each identified, as expected (Supplementary Figure 2*B–F*, respectively). Other transcripts are expressed more broadly throughout the epithelium (Supplementary Figure 2*G*). We found *Ace2* expression to be enriched in cells co-expressing mature enterocyte markers, whereas *Tmprss2* was expressed more broadly throughout the epithelium (Supplementary Figure 2*H*).

We next examined how intestinal transcription factor regulatory networks impact the expression of genes important for SARS-CoV-2 infection. CDX2 is required for specification of the intestine during embryonic development,[5] and in adult life is required for intestinal maturation and proper enterocyte function.[6] HNF4 factors are required for maturation of the embryonic intestine,[7] and work in conjunction with SMAD4 to promote expression of adult enterocyte genes.[8] GATA family transcription factors are important for intestinal regionalization.[9] Chromatin immunoprecipitation sequencing of these transcription factors shows multiple binding regions at the loci of genes involved in CoV-2 infection (Figure 2*A*). Notably, the binding regions occur at locations of accessible chromatin (indicated by assay for transposase-accessible chromatin (ATAC)-seq in pink in Figure 2*A*), consistent with transcriptional enhancer functions. Active chromatin modifications were observed at these loci, and, importantly, chromatin modifications at the *Ace2* locus showed dependence on knockout of CDX or HNF factors in the epithelium using the *Villin-Cre*^{ERT2} driver (indicated by asterisks in Figure 2*B*). Aside from *Ace2*, we also observed increased levels of these active chromatin modifications at the *Tmprss2* locus upon knockout of HNF4 or CDX transcription factors (Figure 2*B*). Altered chromatin structure at the *Ace2* and *Tmprss2* loci upon CDX and HNF4 loss would predict a corresponding change in transcript levels in these knockout models. Indeed, *Ace2* was downregulated significantly upon loss of CDX2, HNF4 factors, SMAD4, or GATA factors (Figure 2*C–F*). Although *Ace2* levels were diminished in these knockout models, *Tmprss2* showed an increase in transcript levels (Figure 2*C–F*). Finally, consistent with their dynamic expression changes in response to HNF4 knockout, we also observed dynamic chromatin looping events at these loci, as measured by Hi-C chromatin immunoprecipitation (HiChIP).[10] Although the *Ace2* locus showed fewer chromatin looping events upon HNF4 knockout, the *Tmprss2* locus showed increased contacts between presumed enhancers and the *Tmprss2* transcriptional start site (Figure 2*G*).

This study provides an overview of how the genes known to facilitate SARS-CoV-2 infection are spatially and temporally expressed and transcriptionally regulated. Transcription factors function in complex and collaborative networks to promote proper cell function. It is interesting that 4 key intestinal transcription factors, CDX2, HNF4 factors, SMAD4, or GATA factors, all activate *Ace2* in the intestine. In addition, these transcription factors also might work as suppressors for *Tmprss2*, because *Tmprss2* is increased upon loss of these key intestinal transcription factors. These findings could help understand variable susceptibility to COVID-19 within the population owing to variable utilization of these transcription factors or their binding sites, which might cause variable expression of *Ace2* or *Tmprss2*. These regulatory mechanisms ultimately could lead to potential avenues for altering host gene expression to reduce the susceptibility or severity of SARS-CoV-2 infection.

L. CHEN[1,2]
A. MARISHTA[1]
C. E. ELLISON[1]
M. P. VERZI[1,2,3]
[1]Department of Genetics, Human Genetics Institute of New Jersey, Rutgers University, Piscataway, New Jersey
[2]Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey
[3]Rutgers Center for Lipid Research, New Brunswick, New Jersey

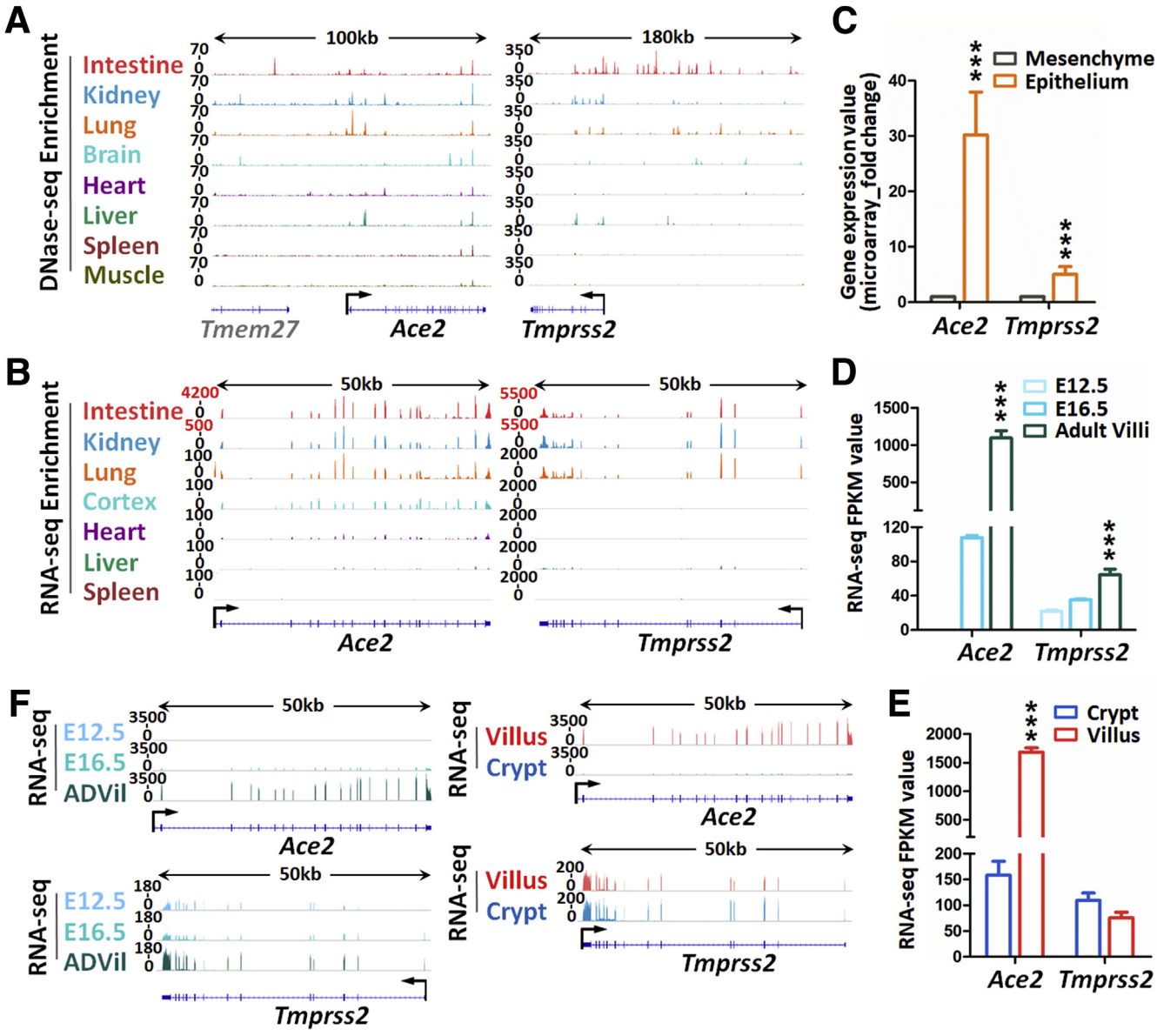

Figure 1. *Ace2* and *Tmprss2* are highly expressed in the intestinal epithelium. (*A*) DNase sequencing (DNase-seq) (GSE51336) enrichment at gene loci of *Ace2* and *Tmprss2* among different tissues. (*B*) RNA sequencing (RNA-seq) (GSE36025) tracks of *Ace2* and *Tmprss2* in different tissues. (*C*) Microarray data of mesenchymal and epithelial cells of the developing intestine (GSE6383). Data are collected from microarray probes showing representative expression fold change. (*D*) RNA sequencing of intestinal epithelium (GSE115541) shows that *Ace2* and *Tmprss2* increasingly are expressed across developmental time, from embryonic day (E)12.5 to adult. (*E*) In the adult tissue, *Ace2* is expressed more robustly in the villus vs the crypt cells (GSE53545, GSE70766, and GSE102171). Data are presented as means ± SEM. ***$P < .001$. (*F*) integrative genomics viewer tracks of RNA sequencing. ADVil, adult villi; FPKM, fragments per kilobase of transcript per million mapped reads.

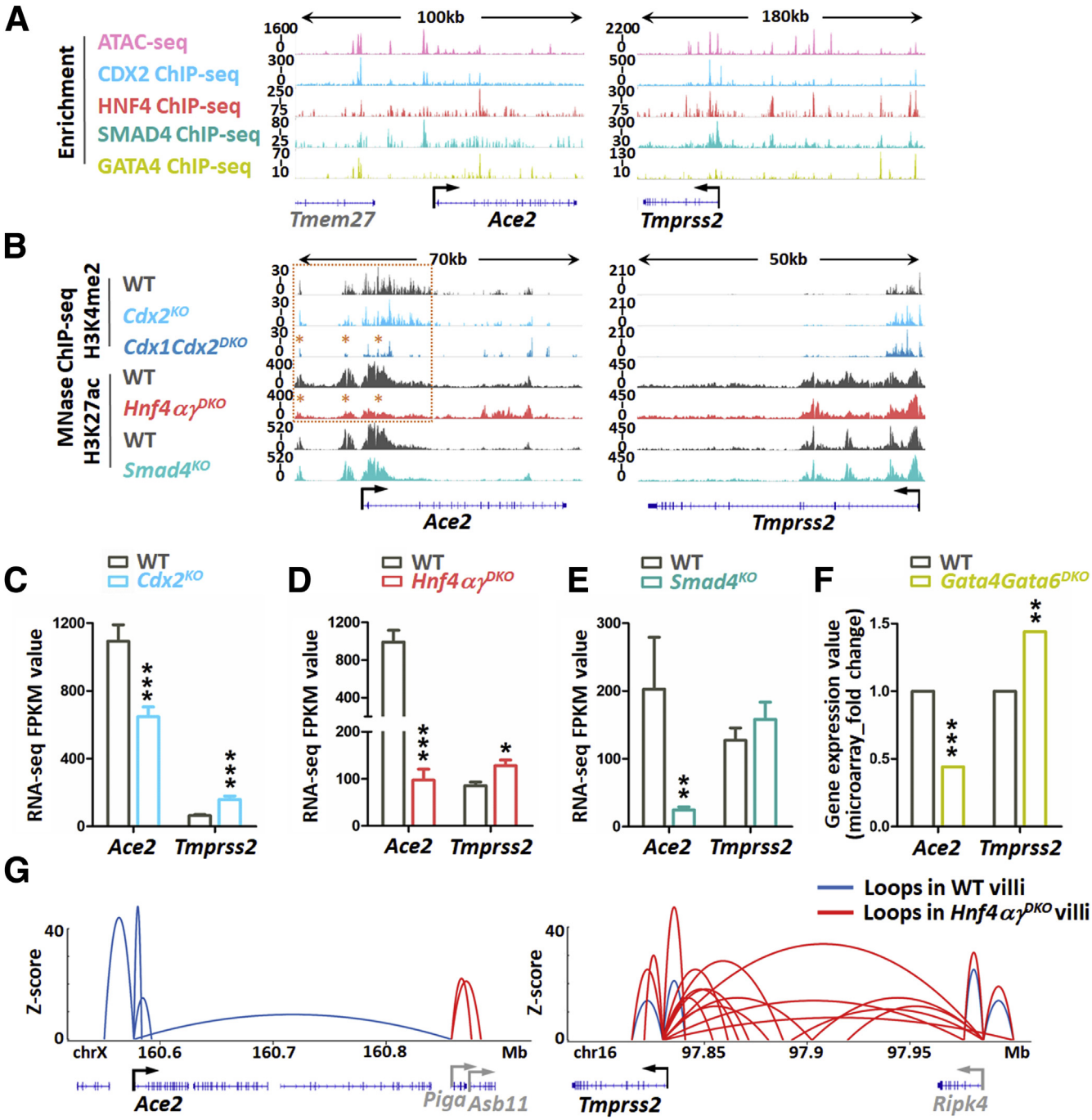

Figure 2. HNF4, CDX2, SMAD4, and GATA transcription factors are key regulators of *Ace2* and *Tmprss2* expression in the intestinal epithelium. (*A*) Chromatin accessibility (ATAC-seq) and transcription factor binding (chromatin immunoprecipitation sequencing [ChIP-seq]) at COVID-19–related gene loci. (*B*) Histone post-translational modifications associated with enhancers and promoters are reduced upon knockout of either CDX or HNF4 at the *Ace2* locus (indicated by asterisks), and, conversely, are unchanged or increase at the *Tmprss2* locus. (*C–F*) Transcriptome analysis to measure corresponding gene expression changes of COVID-19–related host genes upon knockout of these transcription factors (*$P < .05$, **$P < .01$, and ***$P < .001$). (*G*) H3K4me3-HiChIP assays to measure 3-dimensional enhancer–promoter looping show multiple contacts between COVID-related gene promoters and nearby regulatory elements. Loops are visualized by Sushi, and shown with $q \leq 0.0001$ and counts ≥ 8 (2 combined biological replicates per condition). ATAC-seq, assay for transposase-accessible chromatin with high throughput sequencing; FPKM, fragments per kilobase of transcript per million mapped reads; H3K4me3-HiChIP, Hi-C chromatin immunoprecipitation of H3K4me3; MNase, micrococcal nuclease; WT, wild-type.

Address correspondence to e-mail: lchen@dls.rutgers.edu or verzi@biology.rutgers.edu.

References

1. Lamers MM, et al. Science 2020;369:50–54.
2. Yan R, et al. Science 2020;367:1444–1448.
3. Hoffmann M, et al. Cell 2020;181:271–280 e8.
4. Du M, et al. Gastroenterology 2020;158:2298–2301 e7.
5. Gao N, et al. Dev Cell 2009;16:588–599.
6. Verzi MP, et al. Mol Cell Biol 2011;31:2026–2039.
7. Chen L, et al. Development 2019;146:dev179432.
8. Chen L, et al. Nat Genet 2019;51:777–785.
9. Thompson CA, et al. Cell Mol Gastroenterol Hepatol 2017;3:422–446.
10. Mumbach MR, et al. Nat Methods 2016;13:919–922.

Abbreviations used in this letter: **CDX, caudal-type homeobox 2; COVID-19, coronavirus disease 2019; HNF4, hepatocyte nuclear factor 4; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SMAD4, Smad family member 4.**

2352-345X
https://doi.org/10.1016/j.jcmgh.2020.08.005

Received June 17, 2020. Accepted August 13, 2020.

CRediT Authorship Contributions

Lei Chen, (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Investigation: Lead; Methodology: Lead; Visualization: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Argit Marishta, (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Methodology: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting)

Christopher E. Ellison, (Data curation: Supporting; Formal analysis: Supporting; Investigation: Supporting; Methodology: Supporting)

Michael Paul Verzi, (Conceptualization: Lead; Funding acquisition: Lead; Project administration: Lead; Supervision: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)

Conflicts of interest

The authors disclose no conflicts.

Funding

This research was funded by National Institutes of Health grants R01CA190558 (M.P.V.) and R01GM130698 (C.E.E.). Also supported by the Intestinal Stem Cell Consortium from the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under grant number U01 DK103141 (M.P.V.). Supported by the New Jersey Commission on Cancer Research grant DFHS18PPC051 (L.C.). Support also was received from the Sequencing Facility of the Rutgers Cancer Institute of New Jersey (P30CA072720). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.