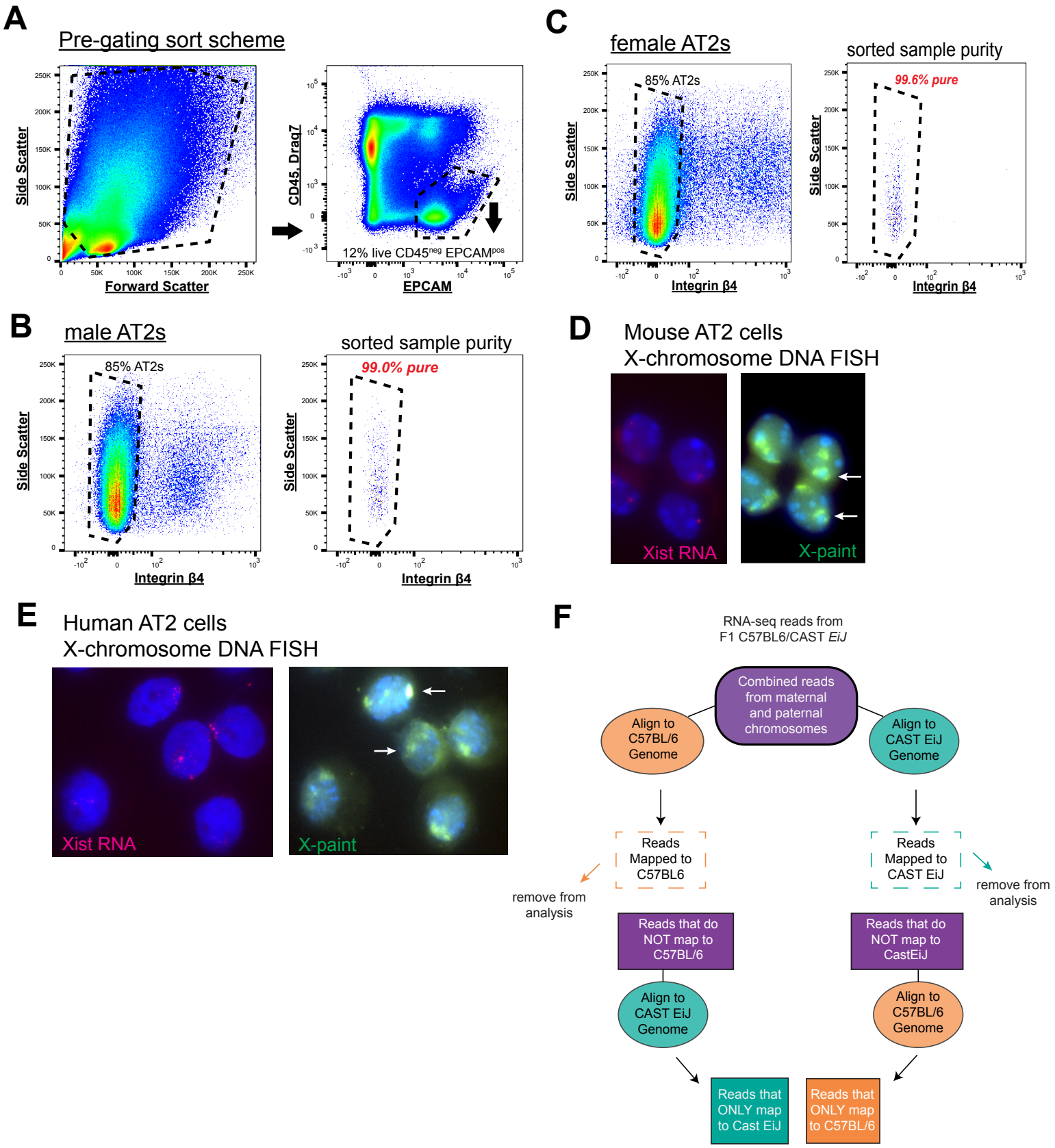
Stem Cell Reports, Volume *18*

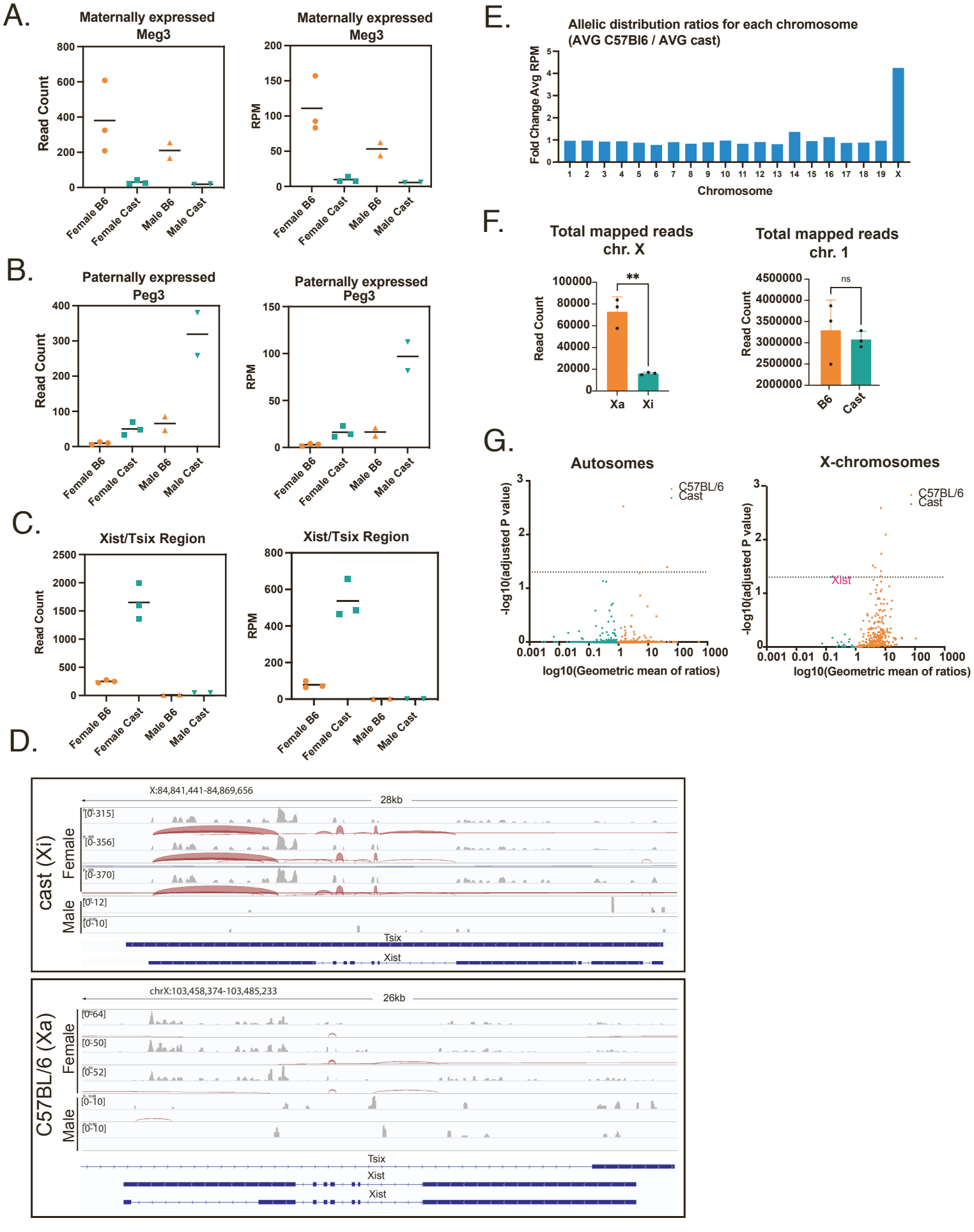
Supplemental Information

Unusual X chromosome inactivation maintenance in female alveolar type 2 cells is correlated with increased numbers of X-linked escape genes and sex-biased gene expression

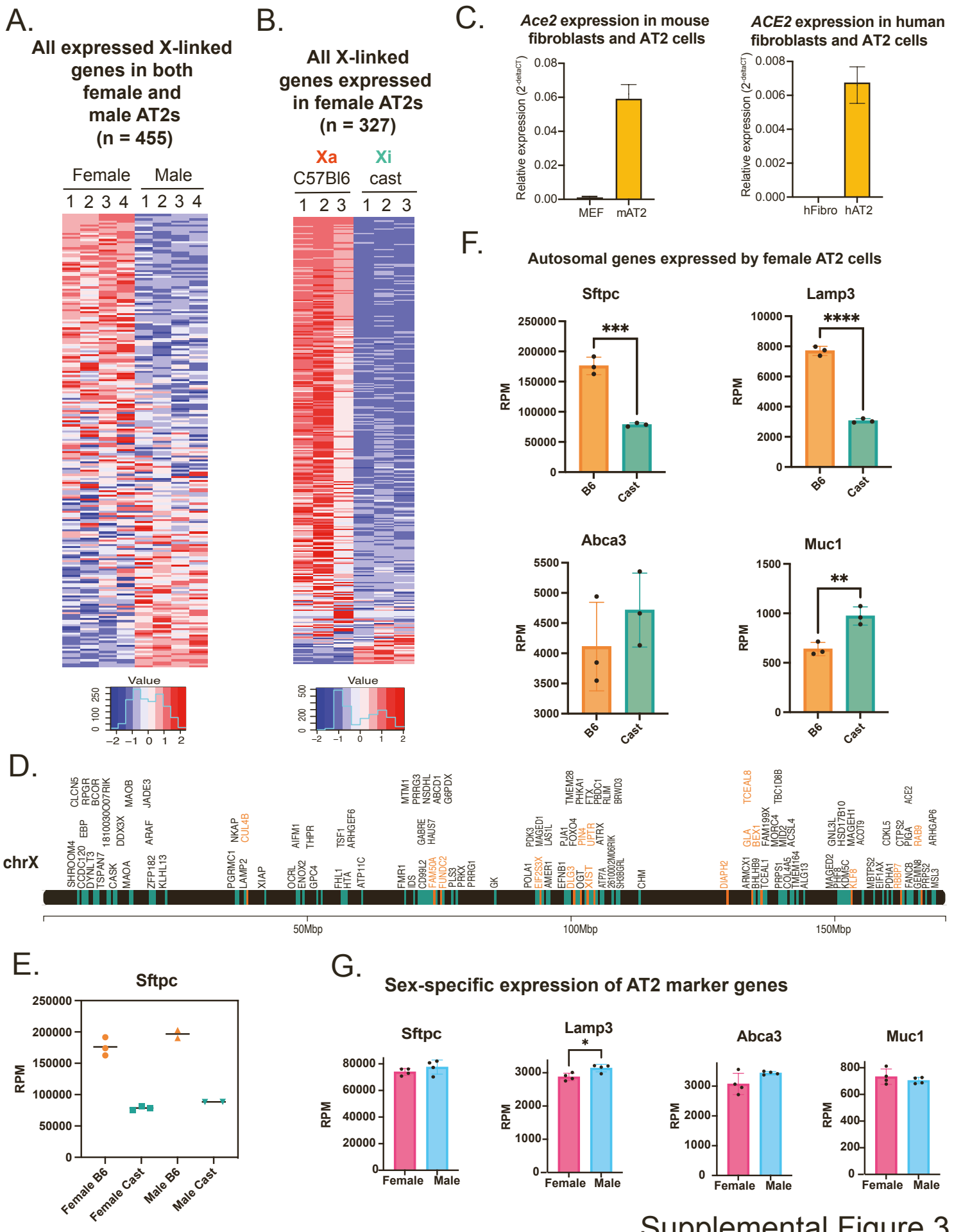
Isabel Sierra, Sarah Pyfrom, Aaron Weiner, Gan Zhao, Amanda Driscoll, Xiang Yu, Brian D. Gregory, Andrew E. Vaughan, and Montserrat C. Anguera



Supplemental Figure 1

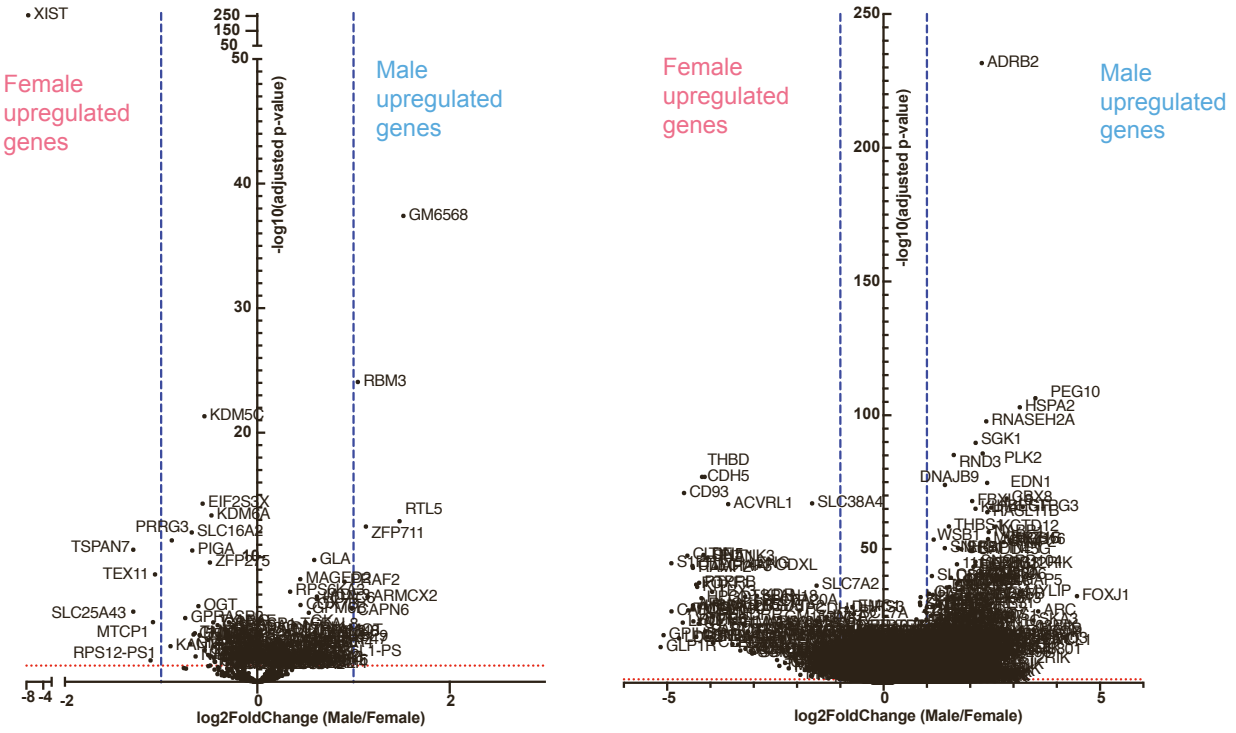


Supplemental Figure 2

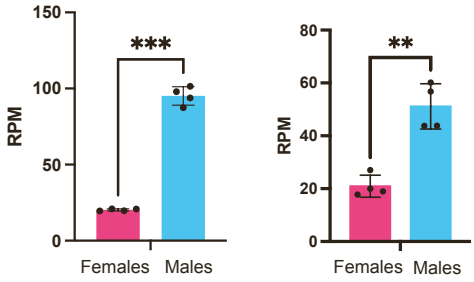


Supplemental Figure 3

A.



B.



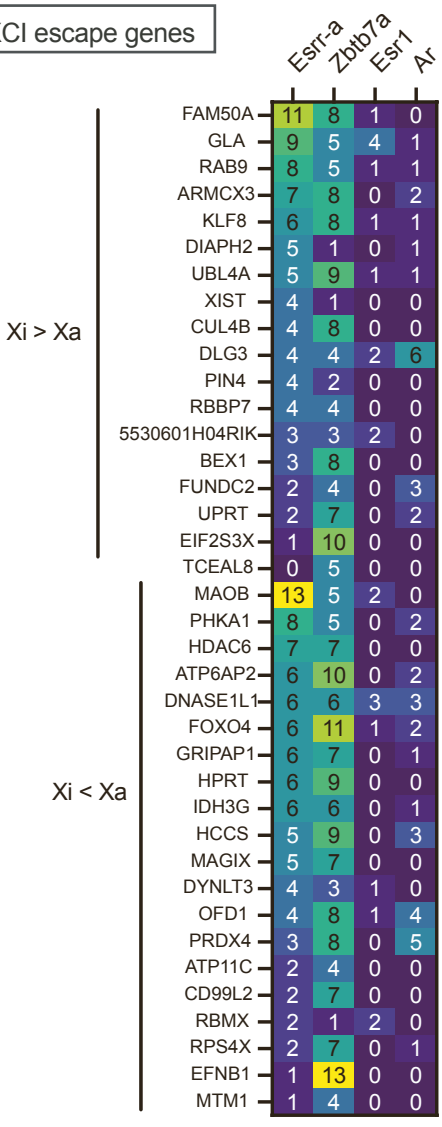
C.

Sex-specific Gene Ontology Analyses

Female-specific upregulation Category	Count	Log10(P)
autophagy	43	-8.05
protein modification by small protein conjugation or removal	85	-8.04
positive regulation of catabolic process	64	-6.46
phosphatidylinositol biosynthetic process	21	-6.17
Herpes simplex virus 1 infection	56	-5.43

Male-specific Category	Count	Log10(P)
Nonsense Mediated Decay independent of the Exon Junction Complex	43	-33.54
mRNA processing	51	-9.02
Parkinson disease	36	-8.72
Cholesterol biosynthesis	8	-7.16
nucleobase-containing compound biosynthetic process	56	-6.93

D.



Supplemental Figure 4

Supplementary Figure 1. AT2 cell isolation for imaging and allele-specific transcriptional profiling. A. FACS sorting scheme for AT2 isolation. Total lung digest is pre-gated on forward and side scatter and live (DRAQ7^{neg}), CD45^{neg}, EPCAM^{pos}. B-C. Within the gates shown in A, male (B) and female (C) AT2s were further defined by gating on integrin β4^{neg}. Sort purity was extremely high, >99% EPCAM^{pos} integrin β4^{neg} for both male and female sorts. D-E. DNA FISH for the X chromosomes in mouse (D) and (E) human AT2 cells. The same field is shown for Xist RNA FISH (left) and X-paint (right). White arrows indicated the Xi, based on overlapping Xist RNA signal. (F). Schematic diagram showing the overview for the alignment to *mus musculus* and *mus castaneus* genomes for the allele-specific RNAseq pipeline.

Supplementary Figure 2. Allele-specific RNAseq profiling of female and male AT2 cells. (A) Allele-specific RNAseq analyses of imprinted genes: *Meg3*, *Peg3*, and *Xist.* Read counts and normalized RPM values for each gene (A) *Meg3* (chr.12), (B) *Peg3* (Chr.7), and (C) *Xist/Tsix* region (chr.X) are shown. The transcript levels for each allele in female and male AT2 cells are shown. (D) IGV coverage (grey) and splice junction (red) tracks showing allele-specific alignment of unique reads that map to the *Xist* and *Tsix* region in female and male F1 mus x cast mice. Note that these reads are not strand-specific, thus we cannot distinguish whether the reads from C57BL/6 (Xa) are Tsix or Xist transcripts. (E) Allelic distribution of mapped reads for all chromosomes in female AT2 cells. The fold change of average RPM values (shown as the ratio of C57Bl6/cast RPM) for all autosomes and X chromosome. (F) The allelic distribution for total mapped reads for each allele, for the X chromosome (Xa or Xi) and chromosome 1

(C57BL/6 or cast). Statistical significance determined using unpaired t-test, ** $p < 0.005$; ns, not significant. (G) The geometric mean of gene expression ratios for allelic expression in female F1 mus x cast mice, for each autosomal gene (left) and X-linked gene (right). C57BL/6 ratios in orange; cast ratios in teal. P values were calculated using ratio paired t-tests and corrected for multiple comparisons using the Holm-Sidak method (alpha = 0.05).

Supplementary Figure 3. Sex-specific and allele-specific profiling of X-linked and AT2 marker genes. (A) Expression of all X-linked genes (n=455 genes) in female and male AT2 cells. Heatmap showing z-score normalized expression the X-linked genes in female (n = 4) and male (n = 4) AT2 samples. Genes were considered 'expressed' if RPM > 1 across all 4 female samples. (B) Allele-specific expression of X-linked genes in female AT2 cells. Heatmap showing allele-specific expression from the Xa (C57BL/6) or the Xi (cast) for all expressed X-linked genes containing a SNP (n=327 genes). (C) Ace2/ACE2 expression in fibroblasts and AT2 cells from mice and humans using qPCR. (D) Location of AT2-specific XCI escape genes across the X-chromosome. Green locations indicate genes that have higher expression from the Xa; orange locations indicate genes that have higher expression from the Xi. (E) Allele-specific expression of *Sftpc* in female and male AT2 cells. (F) Allele-specific expression of known AT2 marker genes *Sftpc, Lamp3, Abca3,* and *Muc1* in female AT2 cells. Statistical significance determined using unpaired t-test, *** $p < 0.0005$; **** $p < 0.00005$. (G) Sex-specific expression of *Sftpc, Lamp3, Abca3, Muc1* in female and male AT2 cells. Statistical significance determined using unpaired t-test, * $p < 0.005$.

Supplementary Figure 4. (A) Volcano plot of X-linked (left) and autosomal (right) genes upregulated in female and male AT2 cells. Blue dashed lines indicate the cutoff for log2 fold change of expression (x-axis) and the y-axis is the adjusted p-value for significance of differential expression. (B) Sex-specific expression (diploid mapped reads) for two autosomal genes (*Adrb2, Gnas*) upregulated in male AT2 cells. Statistical significance determined using student's t-test. (C) Gene ontology analyses for pathways upregulated in female (top) or male (bottom) AT2 cells. Top 5 GO categories are shown; complete lists are shown in **Supplemental Table 3**. (D) Numbers of binding sites for hormone (Ar, Esr1) and nuclear receptor (Esrr-alpha and Zbtb7a) transcription factors in promotor regions (2kb of TSS) of two groups of XCI escape genes expressed in AT2 cells: Xi > Xa denotes genes where there is more expression from the Xi; Xi < Xa denotes genes where there is more expression from Xa.

Supplemental Table 1: List of XCI escape genes in female AT2s that escape according to the binominal confidence interval calculation. The allelic ratio (Xi/Xa) is shown for each XCI escape gene.

Supplemental Table 2: All expressed genes (genome-wide) for male and female replicates (n=4), showing total read count (first tab) and RPM values (second tab) for each replicate sample. Gene ID, gene name, and chromosome are also shown.

Supplemental Tables 3a, 3b, 3c: Gene Ontology (GO) Analyses for genes upregulated in female AT2s (Table 3a), upregulated in male AT2s (Table 3b), and XCI escape genes (Table 3c). Top 20 clusters with their representative enriched terms (one per cluster). "Count" is the number of genes in the user-provided lists with membership in the given ontology term. "%" is the percentage of all of the user-provided genes that are found in the given ontology term (only input genes with at least one ontology term annotation are included in the calculation). "Log10(P)" is the p-value in log base 10. "Log10(q)" is the multi-test adjusted p-value in log base 10.

Supplemental Procedures

Mice

$Xist^{2lox/2lox}$ animals (129Sv/Jae) were a gift of R. Jaenisch (Csankovszki *et al.*, 1999), and live animals were kindly transferred to our colony from J.T.Lee. $Xist^{2lox/2lox}$ mice were backcrossed to C57Bl6 background for 10 generations, and then mated to B-actin Cre Recombinase mice (strain 033984; Jackson labs). F1 mus/cast mice (Xa $^{Xist\ KO}$ Xi) were generated by mating $Xist^{+/KO}$ females with wild-type *Mus castaneus* (*Cast*) males. F1 mus/cast $Xist^{+/KO}$ females from this mating always inactivate the paternal *cast* X chromosome. Mice were screened by PCR primers for Cre-Recombinase and $Xist^{2lox}$ alleles using previously published primer sets(Yildirim *et al.*, 2013). Animal experiments were approved by the University of Pennsylvania Institutional Animal Care and Use Committee (IACUC). Euthanasia via carbon dioxide was used for animal sacrifice prior to lung isolation.

Immunofluorescence (IF)

For IF analyses, slide locations were recorded for Xist RNA FISH images first, and then slides were blocked for 30 min in blocking buffer (PBS with 0.2% Tween-20 and 5% BSA) and then incubated for 2 hours at room temperature with respective primary antibodies (1:100 dilution): H3K27me3 (39155, Active Motif); H2AK119-ubiquitin (8240, Cell Signaling). Slides were incubated with the appropriate FITC conjugated secondary antibody, washed, then imaged at the corresponding locations using a fluorescence microscope.

Read alignment for allele-specific RNA sequencing

500,000 to 1 million AT2 cells were sorted from male and female F1 mus x cast mice, and placed in Trizol. Total RNA isolations were performed following manufacturers protocol. We used the TruSeq Stranded mRNA Library Prep (20020594, Illumina), and libraries were checked for quality using TapeStation, pooled, and subsequently run on the Illumina NextSeq 500 using 75bp single-end reads. Total number of reads per sample varied between 30-80 million. For allele-specific mapping, RNA-seq reads were first aligned to the Cast genome using STAR (v2.6.0a) with default parameters, except for the outSAMunmapped flag, which was set to Within KeepPairs to allow for unmapped reads to be extracted from alignment output. Unmapped reads were extracted using samtools and converted to Fastq format using bamToFastq. These reads were then aligned to the C57Bl/6 genome using STAR with the outFilterMultimapNmax flag set to 1 to filter out reads that mapped to more than one locus. With the output from this second alignment, HTSeq-count (v0.10.0) was used to count allele-specific reads mapping to genes in the

C57Bl/6 genome. To quantify read counts from the Cast genome, the same strategy was employed, except reads were first mapped to the C57Bl/6 genome and unmapped reads were then mapped to the Cast genome.

Hormone receptor binding-site analysis

Bed files (MA0007.3, MA0141.3, MA0592.3, MA0112.3, MA0258.2, MA0750.2) containing select predicted transcription factors binding sites (TFBS) in the mm10 genome were downloaded from http://expdata.cmmt.ubc.ca/JASPAR/downloads/UCSC_tracks/2022/mm10/. Bedfiles were merged and intersected with a 4kb region around the transcriptional start site (TSS) of mm10 genes using the Bedtools Suite (Quinlan and Hall, 2010).

Supplemental References

Camelo, A., Dunmore, R., Sleeman, M.A., and Clarke, D.L. (2014). The epithelium in idiopathic pulmonary fibrosis: breaking the barrier. Front Pharmacol *4*, 173. 10.3389/fphar.2013.00173.
Chapman, H.A., Li, X., Alexander, J.P., Brumwell, A., Lorizio, W., Tan, K., Sonnenberg, A., Wei, Y., and Vu, T.H. (2011). Integrin alpha6beta4 identifies an adult distal lung epithelial population with regenerative potential in mice. J Clin Invest *121*, 2855-2862. 10.1172/JCI57673.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.
Evans, M.J., Cabral, L.J., Stephens, R.J., and Freeman, G. (1975). Transformation of alveolar type 2 cells to type 1 cells following exposure to NO2. Exp Mol Pathol *22*, 142-150. 10.1016/0014-4800(75)90059-3.
Huang, X., Pearce, R., and Zhang, Y. (2020). De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. Aging (Albany NY) *12*, 11263-11276. 10.18632/aging.103416.
Leong-Poi, H., et al. (2005b). Angiotensin-converting enzyme 2 protects from severe acute lung failure. Nature *436*, 112-116. 10.1038/nature03712.
Katsura, H., Sontake, V., Tata, A., Kobayashi, Y., Edwards, C.E., Heaton, B.E., Konkimalla, A., Asakura, T., Mikami, Y., Fritch, E.J., et al. (2020). Human Lung Stem

Cell-Based Alveolospheres Provide Insights into SARS-CoV-2-Mediated Interferon Responses and Pneumocyte Dysfunction. Cell Stem Cell *27*, 890-904 e898. 10.1016/j.stem.2020.10.005.
Lau, E.S., McNeill, J.N., Paniagua, S.M., Liu, E.E., Wang, J.K., Bassett, I.V., Selvaggi, C.A., Lubitz, S.A., Foulkes, A.S., and Ho, J.E. (2021). Sex differences in inflammatory markers in patients hospitalized with COVID-19 infection: Insights from the MGH COVID-19 patient registry. PLOS ONE *16*, e0250774. 10.1371/journal.pone.0250774.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.
Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun *10*, 1523. 10.1038/s41467-019-09234-6.