

Bioinformatic analysis of reproductive system**Bioinformatic analysis reveals that the reproductive system
is potentially at risk from SARS-CoV-2**

**Jiawei Zhang[#], Yuqi Wu[#], Rui Wang[#], Keshi Lu, Menjiang Tu, Huan Guo,
Weijie Xie, Zizhen Qin, Shulin Li, Pei Zhu*, Xiangwei Wang***

Department of Urology & Carson International Cancer Center, Shenzhen University
General Hospital & Shenzhen University Clinical Medical Academy Center,
Shenzhen University, Shenzhen 518000, PR China

*** Correspondence:**

Xiangwei Wang

winn0324@szu.edu.cn

Pei Zhu

zhup158@163.com

[#] These are joint first authors.

Running title: Bioinformatic analysis of reproductive system

Keywords: SARS-CoV-2, ACE2, reproductive system, single-cell RNA-seq

Abstract

An outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) occurred in China towards the end of 2019, and has spread rapidly ever since. Previous studies showed that some virus could affect the reproductive system and cause long-term complications. Recent studies exploring the source of SARS-CoV-2 using genomic sequencing have revealed that SARS-CoV-2 enters the host cells via the angiotensin-converting enzyme II (ACE2), the receptor that recognizes SARS-CoV. To investigate the expression of ACE2 and to explore the potential risk of infection in the reproductive system, we performed a thorough bioinformatic analysis on data from public databases involving RNA expression, protein expression, and single-cell RNA expression studies. The analyzed data showed high levels of ACE2 mRNA and protein expression in the testis and spermatids and equal levels of ACE2 expression in the uterus and lung. Comprehensive single-cell analysis identified

Bioinformatic analysis of reproductive system

ACE2 expression in the lung, testis, spermatids, and uterus. In conclusion, this study revealed the potential risk associated with the SARS-CoV-2 infection in the reproductive system and predicted that long-term complications might have a significant impact on the prevention and management of COVID-19, the disease caused upon infection with SARS-CoV-2.

1 Introduction

An outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1]—a newly identified coronavirus named by the International Committee on Taxonomy of Viruses—occurred in Wuhan, China and has rapidly spread to other countries since the end of 2019. Whole-genome sequencing analysis revealed that SARS-CoV-2 belongs to the same subgenus as the severe acute respiratory syndrome virus (SARS-CoV) and the Middle East respiratory syndrome virus (MERS-CoV) [2]. The RNA sequence of SARS-CoV-2 shows the greatest similarity with that of the coronavirus isolated from *Rhinolophus* bat species [3, 4], which indicates that bats might possibly be the primary source or an intermediate host of SARS-CoV-2 that causes COVID-19, the resulting disease, as named by WHO [5]. The main clinical manifestations of COVID-19 are fever, cough, dyspnea, and bilateral lung infiltration. Although most of the currently reported cases are not severe, approximately 20% of the patients are diagnosed with critical illness characterized by respiratory failure, septic shock, or other organ failure, all of which require intensive care [6-8].

Previous studies [2-4] have revealed that both SARS-CoV and SARS-CoV-2 express a similar spike glycoprotein, which can strongly bind to angiotensin-converting enzyme II (ACE2), the virus-specific receptor expressed on the cell surface, and induce the primary steps governing infection in human cells. The multiplication of SARS-CoV-2 within the host causes an immune response. Similar to SARS and MERS, respiratory droplets and direct contact are the confirmed routes of transmission for SARS-CoV-2 because of the high expression of ACE2 on type II alveolar cells (AT2) of the lung. A recent report by Zhong et al. [6] revealed that gastrointestinal secretions and urine samples are positive for SARS-CoV-2, which indicates other potential routes of transmission. Therefore, investigation of the expression of ACE2 in human cells might provide relevant information regarding the susceptibility to SARS-CoV-2 and might provide evidence regarding other routes of SARS-CoV-2 transmission.

Some viruses, such as the mumps virus [9], may affect the testes and other reproductive organs thus causing long-term complications, such as orchitis and infertility; however, it is not known whether SARS-CoV-2 can directly invade or indirectly damage the reproductive organs. Herein, we investigated the expression of ACE2 in various organs in addition to the human lung, and compared ACE2 expression levels in the reproductive system with that in the lung using single-cell

Bioinformatic analysis of reproductive system

transcriptome-based data available on public databases. These findings highlight the need for clinicians to pay attention to SARS-CoV-2-associated long-term complications in the reproductive system, and provide new data that may significantly impact the prevention and management of COVID-19.

2 Materials and methods

2.1 Data sources

The expression profiles of ACE2 in normal adult tissue, including testis, kidney, fallopian tubes, kidney cortex, seminal vesicles, vagina, ovary, bladder, lung, prostate, placenta, tonsil, epididymis, vas deferens, renal pelvis, endometrium, uterus, and endocervix were obtained from the Expression Atlas database (<https://www.ebi.ac.uk/gxa/home>). Immunohistochemistry images of the lung, testis, seminal vesicle, prostate, fallopian tube, and bladder were acquired from The Human Protein Atlas (<https://www.proteinatlas.org/>). Single-cell RNA-seq (scRNA-seq) data for the lung and uterus were obtained from samples GSM4008628 and GSM4008666 (series number GSE134355) deposited in the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/>) database. Testis data were from series number GSE124263 and spermatid data from series number GSE109037.

2.2 Quality control

Low-quality samples with fewer than 1 000 or greater than 20 000 expressed genes were removed. We set the percentage of Unique Molecular Identifiers(UMIs) mapped to mitochondrial or ribosomal genes at <10%.

2.3 Dimension reduction, cell clustering, and gene expression analysis

The single-cell data expression matrix was processed using the R package, Seurat (version 3.1). First, we used “LogNormalize” to normalize single-cell gene expression data. Then, we identified the highly variable genes using the “FindVariableGenes” function, with the “vst” method. Principle component analysis was performed on the single-cell expression matrix with significant highly variable genes using the “RunPCA” function. We then used the t-Distributed Stochastic Neighbor Embedding (t-SNE) function to obtain cell scatter plots. Gene expression markers of identity classes were identified using the “FindMarkers” function (min.pct = 0.25, logfc.threshold = 0.25). Feature plots and violin plots were generated using Seurat to show the imputed gene expression levels [10-13].

3 Results

3.1 Expression of ACE2 in different tissues based on RNA sequencing

Bioinformatic analysis of reproductive system

The body map of median *ACE2* expression in normal samples was obtained from GEPIA (<http://gepia2.cancer-pku.cn/#general>, Fig. 1A). These data were derived from the RNA-seq datasets of the TCGA and GTEx databases and displayed as $\log_2(\text{transcripts per million [TMP]} + 1)$ after homogenization. In conjunction with the heatmap (Fig. 1B) obtained from the Expression Atlas database, the results showed that in addition to its reported expression in the respiratory (lung $\log_2[\text{TMP} + 1]$, 0.89), digestive (esophagus $\log_2[\text{TMP} + 1]$, 1.43; stomach $\log_2[\text{TMP} + 1]$, 0.4; colon $\log_2[\text{TMP} + 1]$, 1.04), and urinary (kidney $\log_2[\text{TMP} + 1]$, 3.73; bladder $\log_2[\text{TMP} + 1]$, 0.67) systems, *ACE2* was also expressed in male and female reproductive systems (prostate $\log_2[\text{TMP} + 1]$, 0.5; ovary $\log_2[\text{TMP} + 1]$, 0.94; ectocervix $\log_2[\text{TMP} + 1]$, 0.41). Moreover, *ACE2* expression was higher in testis ($\log_2[\text{TMP} + 1]$: 5.16) than that in other tissues. Therefore, the reproductive system, especially the testis, should be considered as a potentially high risk in the context of SARS-CoV-2 infection.

3.2 Expression of *ACE2* in different tissues based on immunocytochemistry

Although the mRNA expression reflects the protein content to a certain extent, it is often not directly proportional the level of protein expression, as different processes regulate mRNA translation in different tissues [14]. Therefore, we obtained immunohistochemical profiles for *ACE2* expression in the lung, testis, seminal vesicle, prostate, fallopian tubes, and bladder from The Human Protein Atlas public database.

We only observed *ACE2* expression in the cytoplasm and membranes of macrophages, while almost no *ACE2* expression was observed in pneumocytes (Fig. 2A). In contrast, high levels of *ACE2* expression were observed in seminiferous duct cells, including the sperm cells and Leydig cells, with quantities greater than 75% (Fig. 2B). *ACE2* protein was expressed at moderate levels in the glandular cells of the seminal vesicles, although its quantity was less than 25% (Fig. 2C). Similarly, we also observed moderate *ACE2* expression in the fallopian tubes (Fig. 2E). *ACE2* expression was also observed in the prostate (Fig. 2D) and bladder (Fig. 2F), although at relatively low levels. Consequently, we speculated that SARS-CoV-2 may infect testis, and sperm and that there is some risk of sexual transmission.

3.3 Cell type-specific *ACE2* expression in the reproductive system

To further explore the expression of *ACE2* in different cell types of different organs in the reproductive system, we analyzed previously published scRNA-seq data. We analyzed scRNA-seq data from the lung to construct a reference dataset. A cluster of lung AT2 cells was identified based on the co-expression of *ACE2*, *ABCA3*, and *MUC1*, using the “Human_cell_markers” data obtained from PanglaoDB (<https://panglaodb.se/>) as reference (Fig. 3). Consistent with the findings of Han et al.,

Bioinformatic analysis of reproductive system

the proportion of *ACE2*-positive AT2 cells was found to be approximately 1% of total AT2 cells[15].

ACE2 was highly expressed in the somatic cells and germ cells of the testes. *THY1* was used to identify the somatic cells and *MAGEA4* was used to identify the germ cells in the testes (Fig. 4). Spermatid scRNA-seq data showed that germ cells from spermatids contained *ACE2*-positive cells and a cluster of sperm cells was identified based on the co-expression of *ENG* and *ITGB1* (Fig. 5). These results further implied that organs of the male reproductive system, especially the testes and spermatids, may be at a high risk of infection with SARS-CoV-2, and can serve as host cells for the virus.

In the uterus, the expression of *ACE2* was observed in the epithelial cells (Fig. 6). *ACE2* expression was relatively lower in the uterus than that in the testes and spermatids, but was close to that observed in the lungs. Therefore, the uterus is also at a risk of being infected by the virus.

4 Discussion

This study showed the potential vulnerability of the reproductive system including testis, spermatids, and uterus to infection with SARS-CoV-2, by analyzing the mRNA and protein expression data for the receptor, *ACE2*, from public databases. We further investigated the mRNA expression of *ACE2* at the single-cell level in the lung tissue and in three different reproduction system tissues, based on data from the GEO database. Previous studies [4, 16] have confirmed the relationship between the differential expression of *ACE2* in human systems and clinical symptoms reported for each system, such as dyspnea in the respiratory system, diarrhea in the digestive system, and acute kidney injury in the urinary system. Lastly, Zhong et al. reported the presence of SARS-CoV-2 in the gastrointestinal tract, saliva, and urine by polymerase chain reaction [6]. The expression of *ACE2* in testes and spermatids indicates that SARS-CoV-2 may infect the male reproductive system. Orchitis and infertility are considered to be additional complications associated with multiple types of viruses, such as the mumps virus, human immunodeficiency virus, and Zika virus [9]. Xu et al. suggested that SARS resulted in orchitis after analyzing pathological changes in testes from six patients who died of SARS [17]. In addition, Zhao et al. reported the presence of SARS-CoV in testicular epithelial cells and Leydig cells in autopsies of two patients who died of SARS [18]. Therefore, we speculate that SARS-CoV-2 may also exist in the testis or even in the seminal fluid, which may not only cause orchitis and infertility, but may also increase the possibility of sexual transmission. Enhanced protection methods may be recommended during digital rectal examination and prostatic massage of patients with suspected or documented SARS-CoV-2 infection. Orchitis and infertility may also be considered as long-term complications of COVID-19.

Bioinformatic analysis of reproductive system

ACE2 expression has previously been identified in the placenta and its levels increase in the renal tissue during pregnancy [16, 19]. In our study, we also showed relatively high expression of ACE2 in the uterus. Additional attention may need to be paid to the investigation of the susceptibility to SARS-CoV infection during pregnancy. Bharadwaj et al. found that ACE2 deficiency inhibits gestational weight gain and restricts fetal growth in pregnant mice [20]. High incidence of spontaneous miscarriage, preterm delivery, and intrauterine growth restriction has been reported in pregnant women presented with SARS in Hong Kong and Guangzhou, China [21, 22]. As SARS-CoV and SARS-CoV-2 have the same target receptor, ACE2, COVID-19 can possibly affect fetal health. A recent retrospective review of medical records by Chen et al. identified two cases of fetal distress among the nine monitored pregnancies, but did not identify any evidence of intrauterine infection [23]. However, a larger number of patients with pregnancies spread across different trimesters are needed to confirm this finding.

Our study has some limitations. Although we constructed a single-cell map of reproductive organs, the total number of cells included in our analysis was low, which may cause analysis bias during the determination of ACE2 expression. The reasons for this low number of cells may be the strict criteria used in the screening period and the limited number of datasets uploaded in the GEO database. However, the results of next-generation sequencing and ACE2 immunohistochemistry data were included as additional lines of evidence. Furthermore, no clinical specimens have been analyzed due to safety concerns and because of the unavailability of patients with SARS-CoV-2 infection. Further laboratory testing is needed to verify the existence of SARS-CoV-2 in the seminal fluid, placenta, and in other reproductive organs.

Compliance with Ethical Guidelines

This manuscript does not involve a research protocol requiring approval by the relevant institutional review board or ethics committee.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

Xiangwei Wang and Pei Zhu contributed to the conception of the study. Jiawei Zhang, Yuqi Wu and Rui Wang contributed significantly to analysis and manuscript preparation; Keshi Lu, Menjiang Tu and Huan Guo, performed the data analyses and

Bioinformatic analysis of reproductive system

wrote the manuscript; Weijie Xie, Zizhen Qin and Shulin Li helped perform the analysis with constructive discussions.

Funding

This study was supported by the Science and Technology Innovation Commission Foundation of Shenzhen (grant no. JCYJ20180305124827261).

Acknowledgments

The authors would like to thank Editage for editing the manuscript prior to submission.

References

1. Gorbalenya AE. Severe acute respiratory syndrome-related coronavirus – The species and its viruses, a statement of the Coronavirus Study Group. bioRxiv (2020). doi: 10.1101/2020.02.07.937862
2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature (2020). doi: 10.1038/s41586-020-2012-7
3. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv (2020). doi: 10.1101/2020.01.22.914952
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet (2020). 10.1016/S0140-6736(20)30251-8
5. WHO: WHO Director-General's remarks at the media briefing on 2019-nCoV. <https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the->

Bioinformatic analysis of reproductive system

media-briefing-on-2019-ncov-on-11-february-2020.

6. Guan W-H, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. Clinical characteristics of 2019 novel coronavirus infection in China. medRxiv (2020). doi: 10.1101/2020.02.06.20020974
7. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020) 395:497-506. doi: 10.1016/S0140-6736(20)30183-5
8. Yang Y, Lu Q, Liu M, Wang Y, Zhang A, Jalali N, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. medRxiv (2020). doi: 10.1101/2020.02.10.20021675
9. Liu W, Han R, Wu H, Han D. Viral threat to male fertility. *Andrologia* (2018) 50(11): e13140.
10. McDavid A, Finak G, Chattopadhyay PK, Dominguez M, Lamoreaux L, Ma SS, et al. Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* (2013) 29(4): 461-7.
11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550.
12. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* (2014) 32:381-6.
13. Andrew McDavid GF, Masanao Yajima. MAST: Model-based Analysis of

Bioinformatic analysis of reproductive system

- Single Cell Transcriptomics. R package version 1.2.1.
<https://github.com/RGLab/MAST/> 2017.
14. Guo Y, Xiao P, Lei S, Deng F, Xiao GG, Liu Y, et al. How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim Biophys Sin (Shanghai)* (2008) 40:426-36.
 15. Xin ZKC, Jiawei Z, Peiyi H, Jie H, Zeguang H. The single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to Wuhan 2019-nCoV infection. *Front Med* (2020).
doi: 10.1007/s11684-020-0754-0
 16. Hamming I, Cooper ME, Haagmans BL, Hooper NM, Korstanje R, Osterhaus ADME, et al. The emerging role of ACE2 in physiology and disease. *J Pathol* (2007) 212:1-11.
 17. Xu J, Qi L, Chi X, Yang J, Wei X, Gong E, et al. Orchitis: a complication of severe acute respiratory syndrome (SARS). *Biol Reprod* (2006) 74:410-6.
 18. Zhao J, Zhou G, Sun Y. SARS coronavirus could cause multi-organ infection. *Med J Chinese People's Liberation Army* (2003) 28:697-8.
 19. Brosnihan KB, Neves LAA, Joyner J, Averill DB, Chappell MC, Sarao R, et al. Enhanced renal immunocytochemical expression of ANG-(1-7) and ACE2 during pregnancy. *Hypertension* (2003) 42: 749-53.
 20. Bharadwaj MS, Strawn WB, Groban L, Yamaleyeva LM, Chappell MC, Horta C, et al. Angiotensin-converting enzyme 2 deficiency is associated with

Bioinformatic analysis of reproductive system

impaired gestational weight gain and fetal growth restriction. Hypertension (2011) 58:852-8.

21. Wong SF, Chow KM, Leung TN, Ng WF, Ng TK, Shek CC, et al. Pregnancy and perinatal outcomes of women with severe acute respiratory syndrome. Am J Obstet Gynecol (2004) 191:292-7.
22. Wang Y, Zhang R, Zhang J. Effects of SARS during pregnancy on the pregnant outcome. Chinese J Perinat Med (2004) 7:155-8.
23. Chen H, Guo J, Wang C, Luo F, Yu X, Zhang W, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. Lancet (2020). doi: 10.1016/S0140-6736(20)30360-3

Bioinformatic analysis of reproductive system

Figure Legends

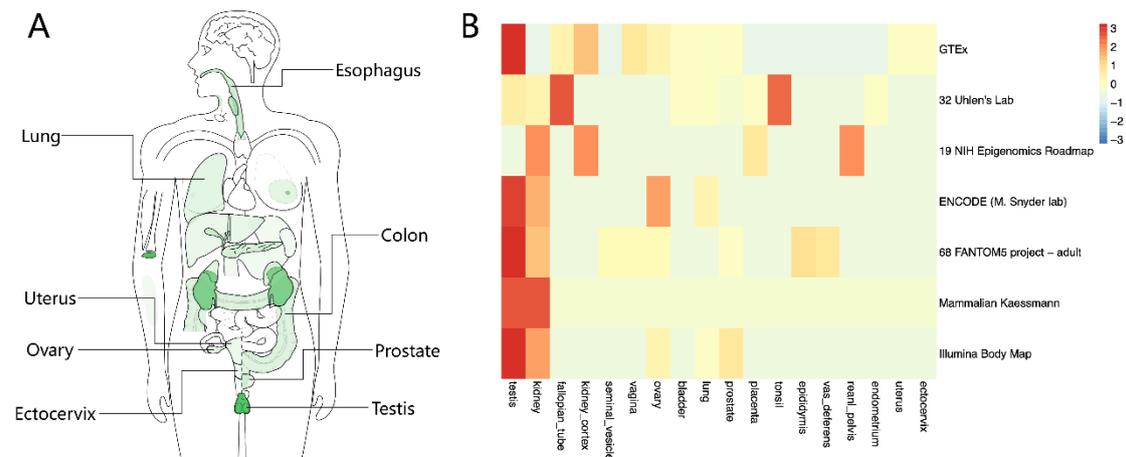


Figure 1. *ACE2* expression analysis based on RNA-seq data deposited in public databases

(A) The bodymap of median *ACE2* expression, calculated as $\log_2(\text{TMP} + 1)$, in different organs was obtained from the GEPIA database (<http://gepia2.cancer-pku.cn/#general>). (B) The heatmap of differential *ACE2* expression in organs of the respiratory system, urinary system, and reproductive system was acquired from the Expression Atlas (<https://www.ebi.ac.uk/gxa/home>), and included data from the GTEx, 32 Uhlen's Lab, 19 NIH Epigenomics Roadmap, ENCODE (M. Snyder lab), 68 FANTOM5 project - adult, Mammalian Kaessmann, and Illumina Body Map databases.

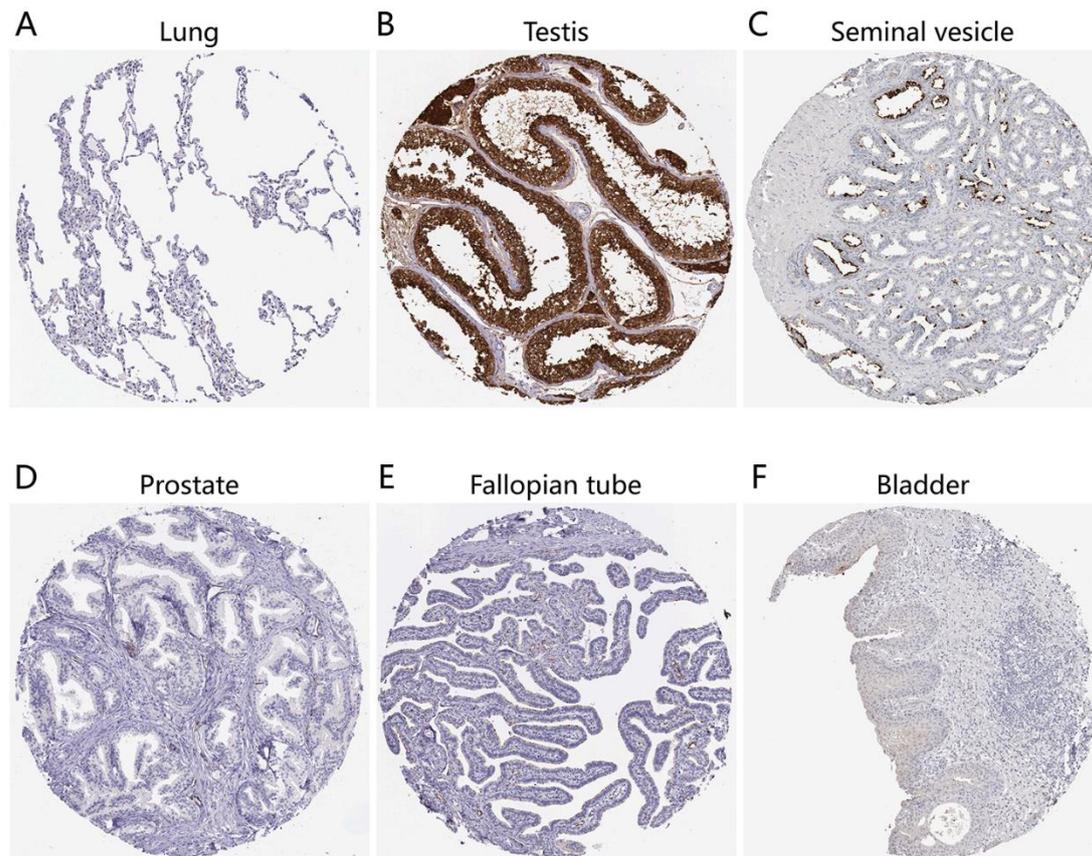
Bioinformatic analysis of reproductive system

Figure 2. ACE2 expression analysis based on data hosted on immunocytochemistry databases

Immunohistochemical profiles were obtained from The Human Protein Atlas database (<https://www.proteinatlas.org/>). (A) ACE2 staining in normal lung tissue collected from a 75-year-old woman (patient id: 496); (B) testis tissue was isolated from a 57-year-old man (patient id: 1180); (C) seminal vesicle tissue was extracted from a 51-year-old man (patient id: 1761); (D) prostate tissue was collected from a 68-year-old woman (patient id: 613); (E) fallopian tube tissue was isolated from a 41-year-old woman (patient id: 1019); and (F) bladder tissue was extracted from a 46-year-old man (patient id: 3184).

Bioinformatic analysis of reproductive system

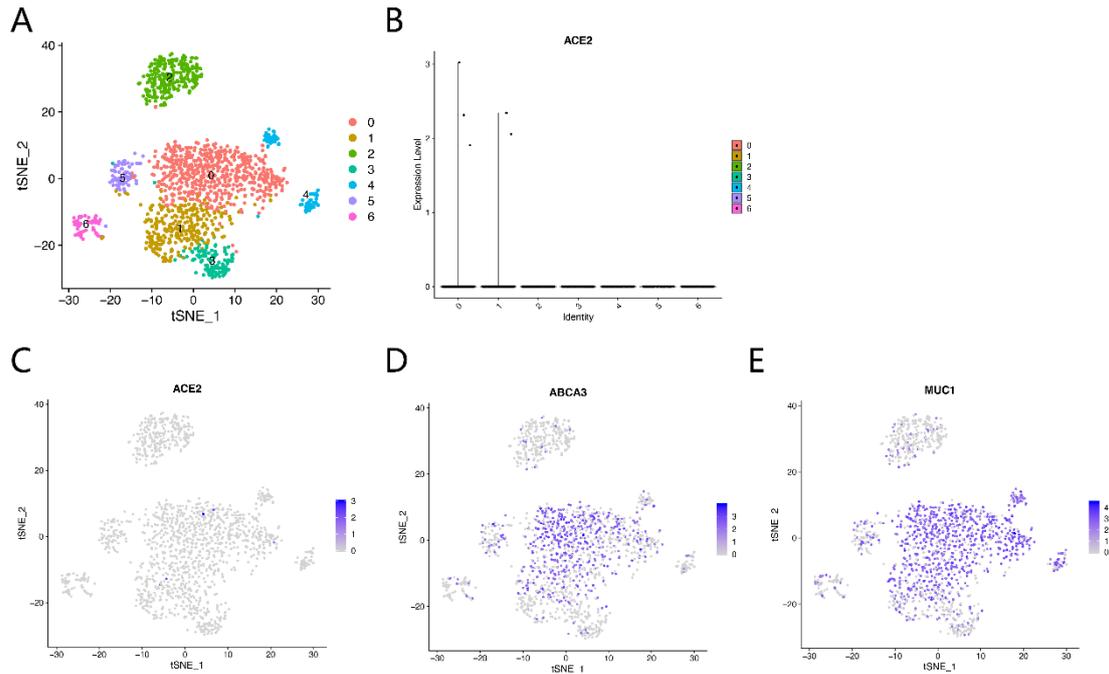


Figure 3. Data analysis of lung cell scRNA-seq

(A) Cells were categorized into six clusters. (B) The violin plot shows the distribution of *ACE2* expression in different cell clusters. (C) The t-SNE feature plot shows that the cluster of cells with *ACE2* expression also expresses the canonical markers of AT2 cells, (D) *ABCA3* and (E) *MUC1*.

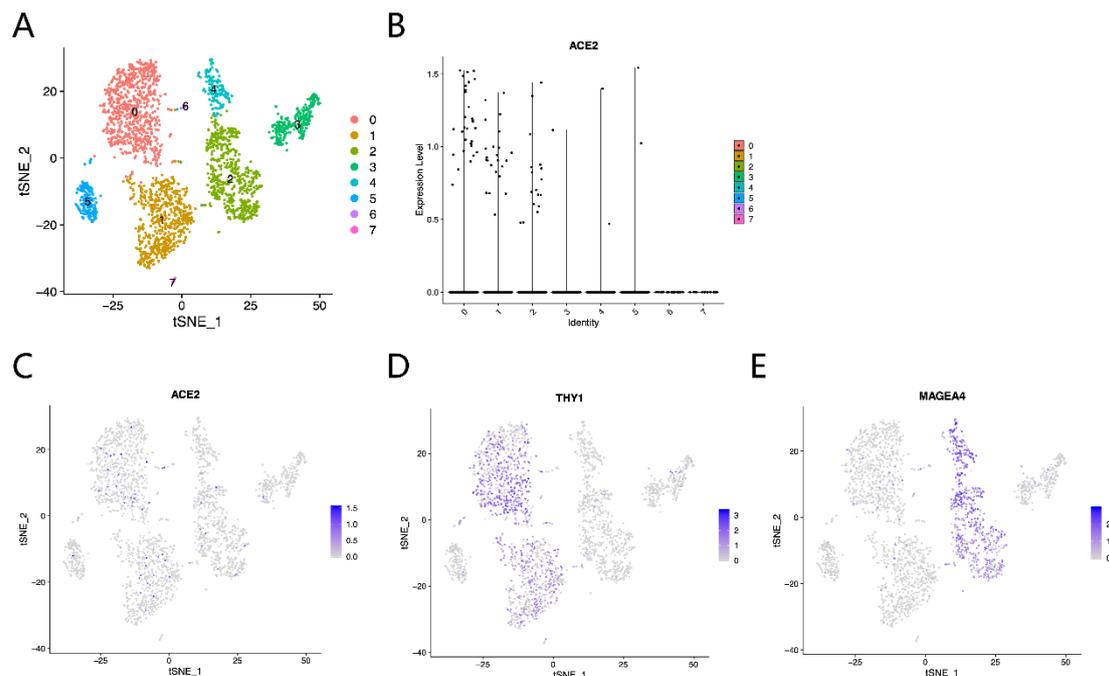


Figure 4. Data analysis of testis cell scRNA-seq

Bioinformatic analysis of reproductive system

(A) Cells were categorized into seven clusters. (B) The violin plot shows the distribution of *ACE2* expression in different cell clusters. (C) The t-SNE feature plot shows that the cluster of cells with *ACE2* expression also expresses the canonical markers of the somatic and germ cells of the testes, (D) *THY1* and (E) *MAGEA4*.

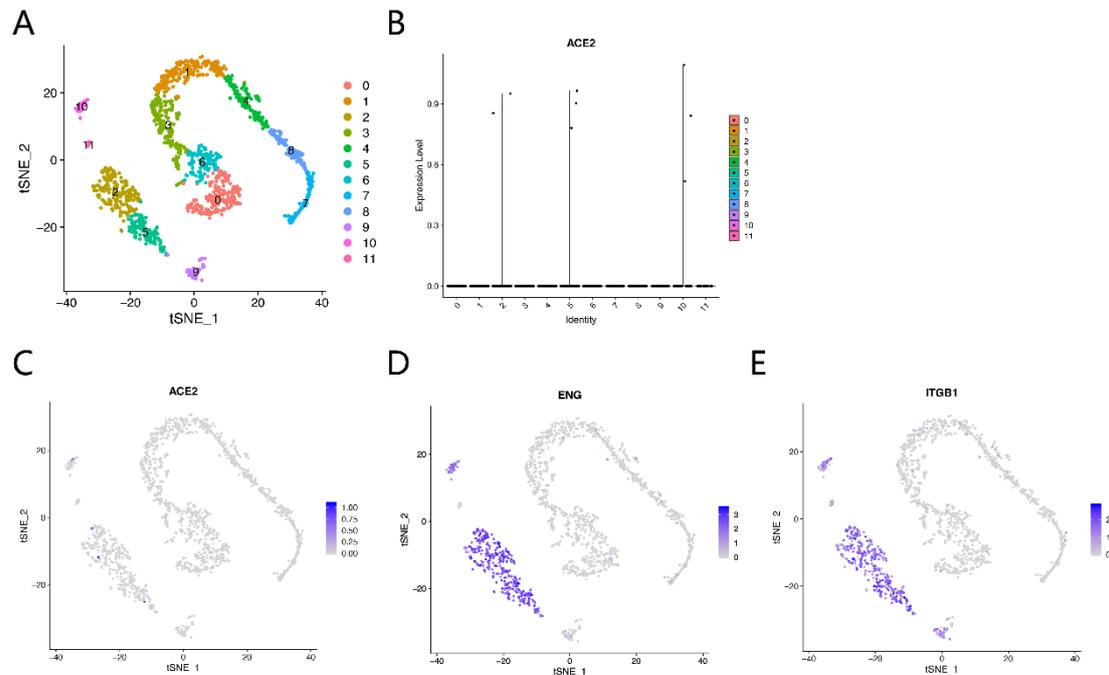


Figure 5. Data analysis of spermatid scRNA-seq

(A) Cells were categorized into eleven clusters. (B) The violin plot shows the distribution of *ACE2* expression in different cell clusters. (C) The t-SNE feature plot shows that the cluster of cells with *ACE2* expression also expresses the canonical markers of germ cells, (D) *ENG* and (E) *ITGB1*.

Bioinformatic analysis of reproductive system

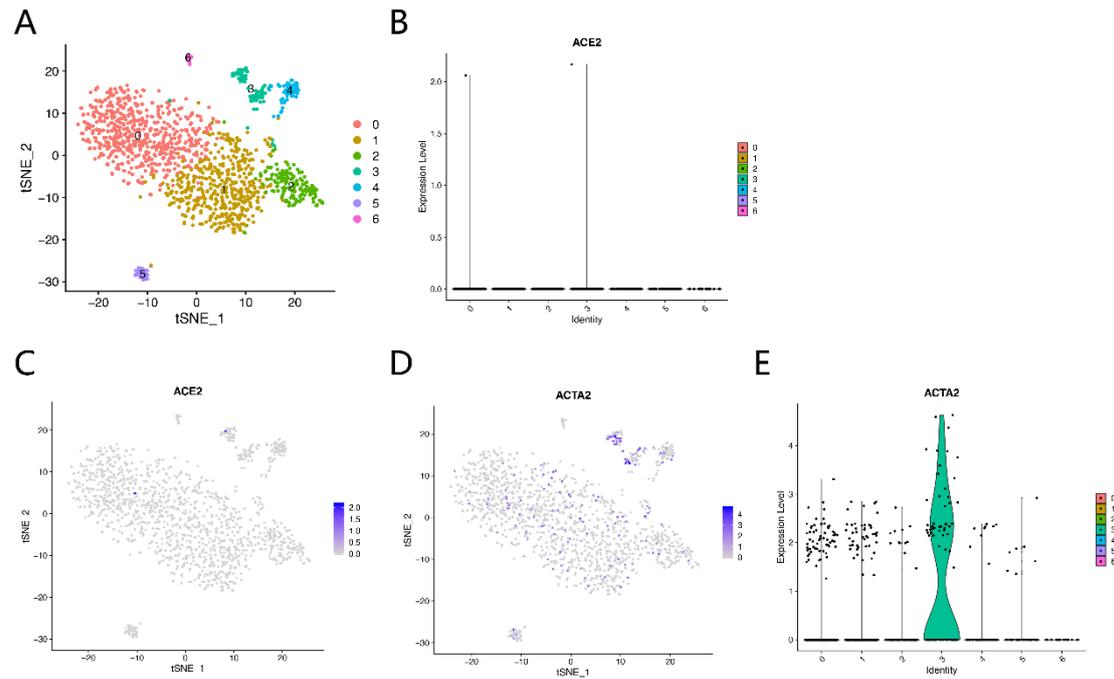


Figure 6. Data analysis of uterus cell scRNA-seq

(A) Cells were categorized into six clusters. (B) The violin plot shows the *ACE2* expression distribution of different cell clusters. (C, D, and E) The t-SNE feature plots and violin plot show that the cluster of cells with *ACE2* expression also expresses the canonical marker of epithelial cells, *ACTA2*.