OXFORD

ADEIP: an integrated platform of age-dependent expression and immune profiles across human tissues

Xuan Liu[†], Wenbo Chen[†], Yu Fang[†], Siqi Yang, Liuping Chang, Xingyu Chen, Haidong Ye, Xinyu Tang, Shan Zhong, Wen Zhang, Zhiqiang Dong, Leng Han and Chunjiang He (iD)

Corresponding authors: Chunjiang He, College of Biomedicine and Health, Huazhong Agricultural University, Wuhan 430070, China.
E-mail: che@whu.edu.cn; Leng Han, Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, Texas A&M University, Houston, TX 77030, USA. E-mail: leng.han@tamu.edu; Zhiqiang Dong, College of Biomedicine and Health, Huazhong Agricultural University, Wuhan 430070, China. E-mail: dongz@mail.hzau.edu.cn
[†]These authors contributed equally to this work.

Abstract

Gene expression and immune status in human tissues are changed with aging. There is a need to develop a comprehensive platform to explore the dynamics of age-related gene expression and immune profiles across tissues in genome-wide studies. Here, we collected RNA-Seq datasets from GTEx project, containing 16 704 samples from 30 major tissues in six age groups ranging from 20 to 79 years old. Dynamic gene expression along with aging were depicted and gene set enrichment analysis was performed among those age groups. Genes from 34 known immune function categories and immune cell compositions were investigated and compared among different age groups. Finally, we integrated all the results and developed a platform named ADEIP (http://gb.whu.edu.cn/ADEIP or http://geneyun.net/ADEIP), integrating the age-dependent gene expression and immune profiles across tissues. To demonstrate the usage of ADEIP, we applied two datasets: severe acute respiratory syndrome coronavirus 2 and human mesenchymal stem cells-assoicated genes. We also included the expression and immune dynamics of these genes in the platform. Collectively, ADEIP is a powerful platform for studying age-related immune regulation in organogenesis and other infectious or genetic diseases.

Xuan Liu is a PhD student at College of Informatics, Huazhong Agricultural University.

Wenbo Chen is a research assistant at College of Biomedicine and Health, Huazhong Agricultural University.

Yu Fang is a postdoc at College of Biomedicine and Health, Huazhong Agricultural University.

Siqi Yang is a master student at School of Basic Medical Science, Wuhan University.

Liuping Chang is a master student at College of Biomedicine and Health, Huazhong Agricultural University.

Xingyu Chen is a master student at College of Biomedicine and Health, Huazhong Agricultural University.

Haidong Ye is a master student at College of Biomedicine and Health, Huazhong Agricultural University.

Xinyu Tang is an undergraduate student at School of Basic Medical Science, Wuhan University.

Shan Zhong is a professor at School of Basic Medical Science, Wuhan University.

Wen Zhang is a professor at College of Informatics, Huazhong Agricultural University. He is an expert in bioinformatics and database construction.

Zhiqiang Dong is a professor at College of Biomedicine and Health, Huazhong Agricultural University.

Leng Han is an Associate Professor at Center for Epigenetics and Disease Prevention, Institute of Biosciences and Technology, Texas A&M University He is an expert in bioinformatics, especially for immune data mining.

Chunjiang He is a professor at College of Biomedicine and Health, Huazhong Agricultural University. He is an expert in bioinformatics and immunoinformatics.

Submitted: 26 May 2021; **Received (in revised form):** 21 June 2021

Introduction

Age has been reported to play critical roles in gene expression in organogenesis and diseases and causes the diverse responses to the virus infection or efficacy of drugs, bringing challenges to clinical therapy [1, 2]. Identification of genes associated with age and defining the function of these genes have important significance in the study of age-related diseases [3]. Previous work based on high-throughput transcriptome sequencing datasets has revealed the association of gene expression and tissue specificity, quantitative trait locus (QTLs) and sex [4, 5]. Some studies have analyzed age-related expression profiles based on a small number of samples in specific conditions or diseases but not across the life span of individuals [3, 6–8]. In addition, the innate immune cell functions, such as cell migration and pattern recognition receptors signaling, are impaired in aged individuals [9]. Aging is known to be associated with the lower frequency of naive T cells [10]. In addition, in the pandemic of coronavirus disease 2019 (COVID-19), different organs and ages exhibited different responses to infections and treatment [11–13]. Therefore, there is a need to develop a comprehensive platform to explore the age-dependent immune alterations across human tissues.

To explore expression and immune profiles in tissues along with aging, we developed an integrated platform named ADEIP (age-dependent expression and immune profile) based on all available datasets of The Genotype-Tissue Expression (GTEx) project (v8) [14]. ADEIP includes gene expression data from >16 000 samples from 30 major human tissues, from donors within the age range of 20–80. Differential expressed genes and enriched pathways were inspected among age groups. Genes belonging to classic immune categories were also compared. Immune cell composition was inferred and compared among age groups. To demonstrate the usage of ADEIP, we demonstrated two examples by using two gene sets including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and human mesenchymal stem cells (hMSC)-associated genes. To our knowledge, ADEIP is the first integrated platform displaying a large-scale view of age-related expression and immune dynamics. ADEIP provides a powerful platform to investigate aging-related expressions and immune trajectory across tissues, which can contribute to understand the mechanisms of infectious or genetic diseases associated with ages, allowing the development of therapy procedures accordingly.

Materials and methods

Datasets collection

The quantitative data included raw counts and TPM (Transcripts Per Kilobase of exon model per million mapped reads) of 54 592 genes in 30 human tissues were downloaded from the GTEx project (v8, https://gtexportal.org/) [14]. According to GTEx documents, the sequencing reads were aligned to the human reference genome GRCH38/hg38 using STAR v2.5.3a [15]. Gene-level expression quantification was based on the GENCODE v26 gene annotation, and read counts and TPM values were produced with RNA-SeQC v1.1.9. Currently, ADEIP includes 16 704 samples derived from 948 donors, which are from different age groups. The age groups were 20–29, 30–39, 40–49, 50–59, 60–69 and 70–79.

Differential analysis of aging-associated genes

DESeq2 [16] was used to perform differential expression analysis between any two groups in six age groups. At least two replicates were needed in at least one of the two compared groups. Genes with a fold-change >2 and P-value <0.05 (correction by Benjamini and Hochberg method) were treated as significant differential expressed genes.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was implemented using GSEA [17] command line tool (https://www.gsea-msigdb.org). A total of 7481 biological process terms of Gene Ontology [18] (http://geneontology.org/) were acquired from molecular signatures database (MSigDB) [19] (https://www.gsea-msigdb.org/) and used as the reference gene sets. Age groups with less than three samples in one tissue were discarded in the analysis according to GSEA instructions.

Classification of immune function categories

A total of 34 known immune function categories including gene annotation were collected from four public databases: ImmPort [20] (https://www.immport.org/), InnateDB [21] (https://www.innatedb.ca/), HisgAtlas [22] (http://biokb.ncpsb.org/HisgAtlas/) and Gene Ontology [18] (GO:0002376, http://geneontology.org/). Gene official symbols and Ensemble ID are used to match the associated genes in the platform.

Definition of subcellular location

The subcellular localization annotation of genes was collected from Hum-mPLoc [23] (v3.0) (http://www.csbio.sjtu.edu.cn/bioinf/Hum-mPLoc3/) and Euk-mPLoc [24] (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi/). Genes annotated in extracellular localization were excluded and 12 types of subcellular localization including 17 645 genes were retained.

Inference of immune cell composition

CIBERSORT [25] (https://cibersort.stanford.edu/) was used to infer the composition of immune cells from RNA-Seq quantitative data. CIBERSORT R source code was downloaded from the original site, and R package Immunedeconv [26] was used to invoke CIBERSORT to calculate the composition of immune cells in each tissue based on gene TPM. A total of 22 immune cell types were inferred. The two-tailed Mann–Whitney U-test was used to compare the relative ratios of cell types between samples through the Python package SciPy [27].

Correlation analysis of gene expression and cell composition

Pearson correlation between gene expression and relative ratios of cell types was performed by R package psych (https://cran.r-project.org). Only those genes or cell types detected (with TPM or ratio >0) in >75% of samples were retained in the analysis. Adjusted P-value <0.05 corrected by Benjamini and Hochberg method was used to infer the significance.
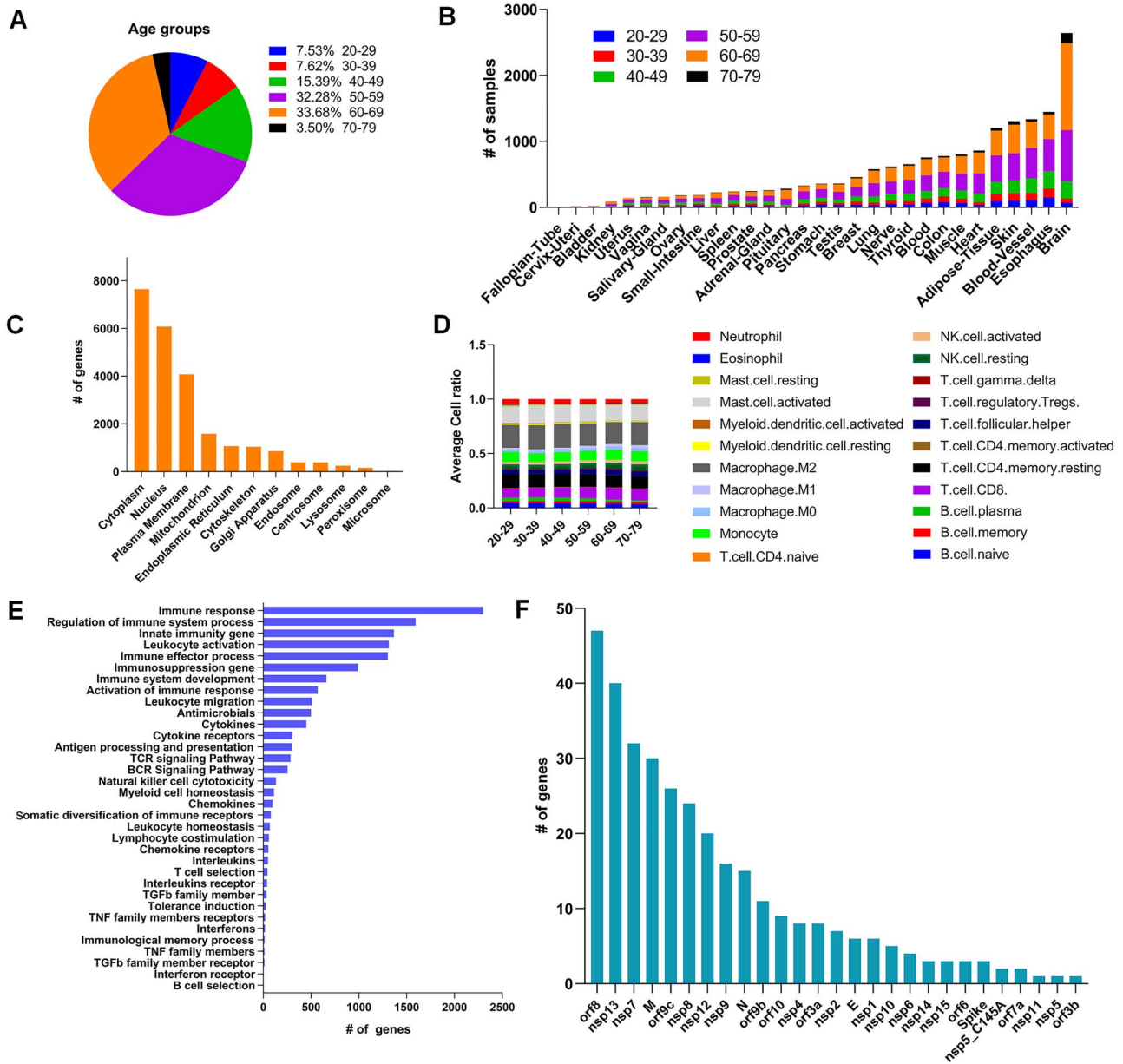
Figure 1. Overview and content of ADEIP. (**A**) Number of samples distributed in six age groups. (**B**) Number of samples distributed in 30 tissues and six age groups. (**C**) Distribution of genes in 12 subcellular locations. (**D**) Average ratio of 22 cell types in six age groups. (**E**) Distribution of genes in 34 immunological categories collected in the analysis. (**F**) Distribution of genes binding to 27 antigens of SARS-CoV-2.

Collection of gene set example

Two gene sets were collected to demonstrate the usage of our platform. A list of human genes interacting with SARS-CoV-2 proteins was collected from previous research [28], which revealed these genes are potentially involved in the invasion of SARS-CoV-2. Target genes corresponding to drugs used in COVID-19 treatment were also collected. Furthermore, human multipotent mesenchymal stem cells (hMSC)-associated genes were extracted from previous work [29], which is identified by comparing gene expression from multipotent mesenchymal stem cells of elderly osteoporosis patients (hMSC-OP), elderly wounds of non-osteoporotic (hMSC-old) and middle-aged healthy wounds (hMSC-C). Target genes corresponding to the drugs used in osteoporosis treatment were also collected [30, 31].

Statistical analysis

Differential expression analysis among age groups was performed by two-sample *t*-test (*$P < 0.05$,**$P < 0.01$,***$P < 0.001$). Images were plotted by R packages (ggplot2, pheatmap, igraph and corplot) and Graphpad Prism.

Platform implementation

The web application of ADEIP was developed using MySQL (v.5.6, https://www.mysql.com) and ThinkPHP (v5.0, https://github.com/top-think). The interface of the website was designed and implemented using Bootstrap (https://getbootstrap.com), Echarts (http://echarts.apache.org), and Highcharts (https://www.highcharts.com.cn) were used to generate the images. The website has been tested in several popular web browsers,
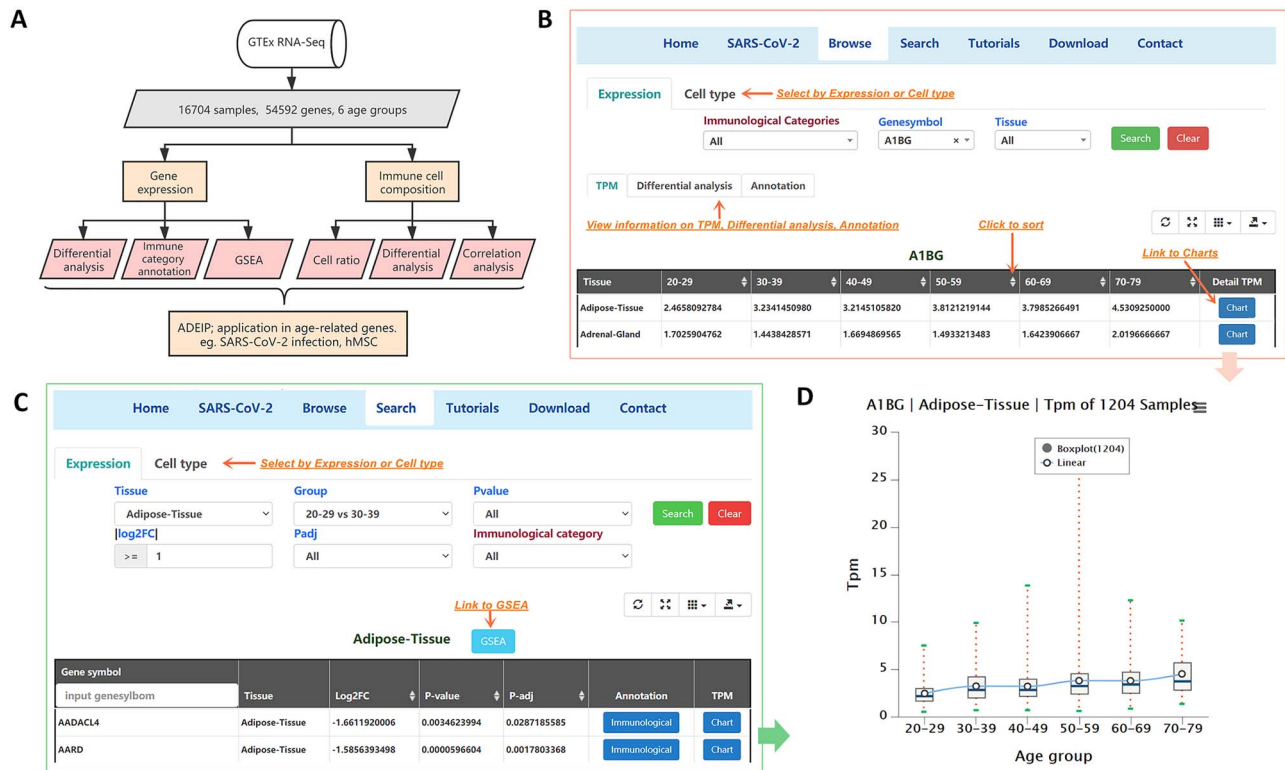
Figure 2. Interface and exhibition of ADEIP. (**A**) Pipeline of ADEIP contruction. All the datasets were collected and subsequent analysis including differential expression analysis, immune categories annotation, GSEA, cell ratio inference and correlation analysis were performed. All the results were included and integrated in the database. (**B**) Browse interface. Two panels including expression and cell type were established and available for query. (**C**) Search interface. User can select different parameters or thresholds to filter the results. (**D**) Chart of boxplot illustrates the detailed expression of genes in six age groups.

including Google Chrome and Firefox. The entire content of the ADEIP is freely available and can be downloaded from the website.

Results

Platform content

ADEIP utilized RNA-Seq datasets from GTEx project, including 16 704 samples of 30 major tissues (adipose, adrenal, bladder, blood, blood vessel, brain, breast, cervix uteri, colon, esophagus, fallopian tube, heart, kidney, liver, lung, muscle, nerve, ovary, pancreas, pituitary, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, uterus, vagina) from 948 human donors in six age groups: 20–29, 30–39, 40–49, 50–59, 60–69 and 70–79 (Figure 1A). Each tissue has a different number of samples (Figure 1B). The RNA-Seq reads are aligned to the human genome and annotated into 54 592 genes. Differential expression analysis was performed between any two of six age groups and significantly differential results (Fold-change >2 and adjusted *P*-value <0.05) were selected and imported into ADEIP. GSEA was also performed between age groups and pathways with the absolute value of normalized enrichment score (NES) >1 and the *P*-value <0.05 were considered as significant.

A total of 17 645 genes with 12 subcellular locations were extracted and included (Figure 1C). A total of 22 types of immune cells in all 16 704 samples were predicted. The proportion of each immune cell in each sample belonging to each age group was presented (Figure 1D). In addition, differential ratios of those immune cells in different age groups were performed. The correlation between the proportion of immune cells and gene

expression was also inspected. A total of 4636 genes attributed to 34 immune-related functional categories were classified and established in the database (Figure 1E).

We also collected two examples to illustrate the aging-related expression dynamics. A total of 333 genes binding to 27 antigens of SARS-CoV-2 (Figure 1F) and 44 target genes corresponding to 55 drugs used in COVID-19 treatment were included (Supplementary Tables S1 and S2). Another 3915 hMSC-associated genes and nine target genes corresponding to 27 drugs used in osteoporosis treatment were also included (Supplementary Tables S3 and S4).

Platfom access

All these contents are included and integrated into an integrated platform named ADEIP (Figure 2A). We developed a user-friendly interface to allow user browse, search and download. Users can browse the ADEIP database through various options (Figure 2B). The 'Expression' option allows users to browse and filter the gene information. Users can also select 'TPM', 'Differential analysis' and 'Annotation' panels for more results. The 'TPM' panel includes expression level (the average TPM) in each of the six age groups and 30 tissues, details of the expression curve can be viewed by clicking the 'chart' button (Figure 2D). 'Differential analysis' panel exhibits the differential expressed genes among six age groups. 'Annotation' panel presents the gene annotations including Ensembl ID, chromosome type, gene type, subcellular location, immunesource function category, etc.

Furthermore, users can visualize the relative ratios, differential results among age groups and correlated genes of immune
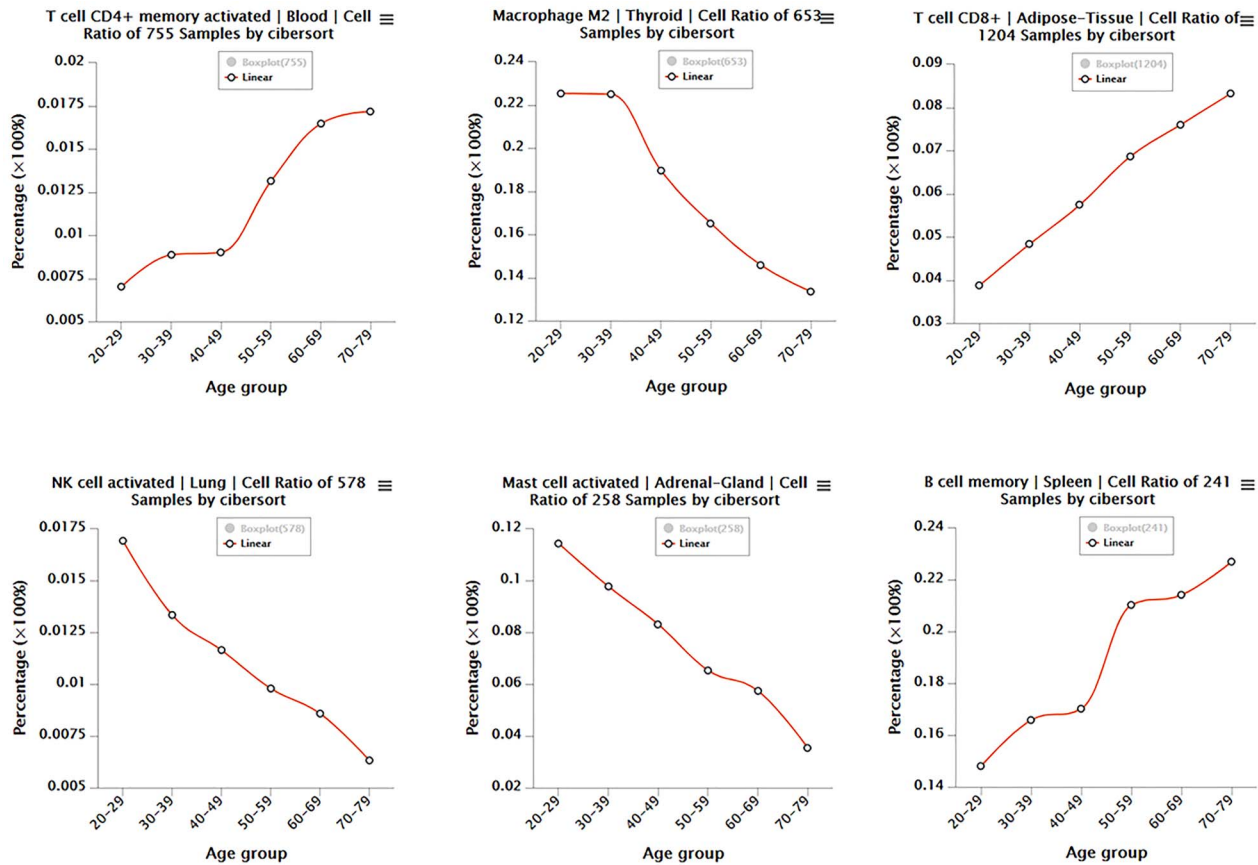
Figure 3. Exhibition of cell ratios in six age groups in ADEIP. Ratios of six cell types including CD4+ T cell, Macrophage M2, CD8+ T cell, activated NK cell, activated mast cell and memory B cell in blood, thyroid, adipose, lung, adrenal gland and spleen separately, in six age groups.

cells by selecting the 'Cell type' option. The percentage of cell ratio in six age groups in each tissue is presented in 'Cell ratio' panel. By clicking 'chart', users can view the distribution curves (Figure 3). The results of differential cell ratio among six age groups are presented in the 'differential analysis' panel. 'Correlation analysis' panel includes the potential genes correlated with the cell types.

Users can query the details of gene expression in the platform by selecting multiple thresholds including tissue type, age groups, P-values/adjusted P-value, log2 fold-change or immune function categories through the 'Expressoin' panel in the search page (Figure 2C). Gene symbol can also be searched in the detailed tables. Gene set enrichment results among those age groups in each tissue can be viewed by clicking the 'GSEA' button (Figure 2D). Details of immune cell composition can also be searched by selecting tissue type, age groups and P-value through the 'Cell type' panel. Users can also select cell type or gene symbol in the detailed tables.

All the original and result files including sample information, gene expression, cell composition in all samples, and lists of associated genes and drugs of example gene sets in ADEIP can be downloaded through the download page. Gene annotation used in ADEIP can also be downloaded through the download page.

Example illustration

Two gene sets were used as examples to illustrate the usage of ADEIP. Users can view the detailed information of SARS-CoV-2 or hMSC-associated genes through the example

page (Figure 4A). Clicking 'gene symbol' links to the browse page to establish the detailed information. For example, the expression of SARS-CoV-2-associated gene CEP250 in adipose is increased along with aging. Reversely, the expression of another associated gene HDAC2 in testis is decreased along with aging (Figure 4B). The 'drug' tab includes the detailed information of SARS-CoV-2 or hMSC-associated drugs and target genes. The statistics of samples with differential expression among six age groups can be viewed by clicking 'chart'.

Discussion

Many diseases, including idiopathic pulmonary fibrosis, cardiovascular disease, type-2 diabetes, primary osteoporosis and COVID-19, exhibited the association of age and raised challenges for clinical treatment and prognosis [29, 32–36]. Thus, genome-wide investigation of gene expression change along with aging could benefit therapy. In this work, we developed ADEIP, an integrated platform to visualize the age-dependent expression and immune profiles across human tissues. Through ADEIP, we observed that the gene expression in different tissues was changed along with aging. Differential expresseion and pathway analysis revealed different functional enrichment among age groups. Furthermore, an inspection of the subcellular location of genes could contribute to decipher how aging influences gene expression inside cells. Enrichment of known immune functional categories also revealed the age-dependent patterns, which could help to understand the immune status dynamics in different ages.
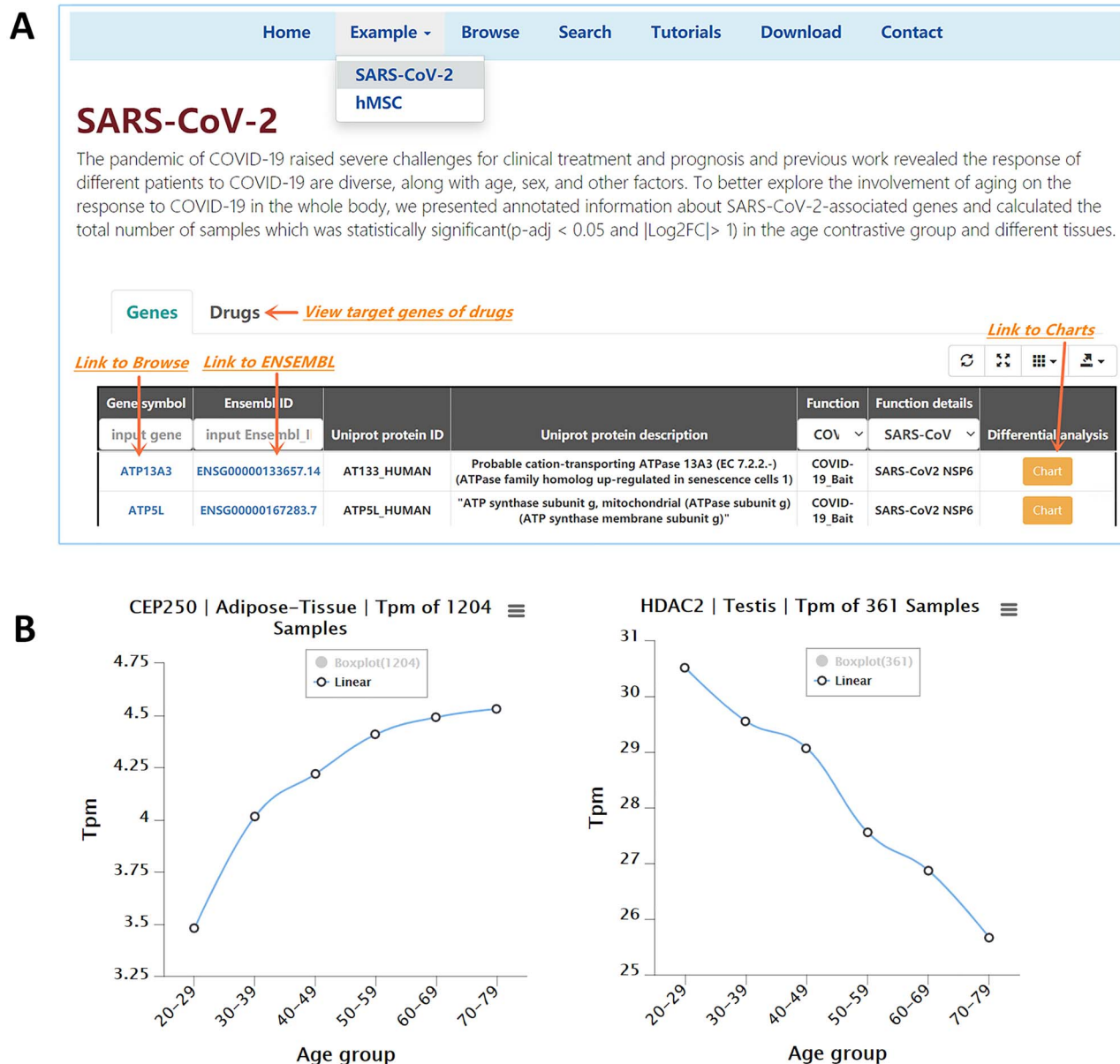
Figure 4. Application of ADEIP in SARS-CoV-2-associated genes. (**A**) Platform collection of SARS-CoV-2-associated genes. (**B**) Expression distribution of two SARS-CoV-2-associted genes CEP250 and HDAC2 in adipose and testis separately, in six age groups.

Based on the popular deconvolution method, CIBERSORT [25], which was been developed to estimate the immune cell abundance from gene expression profiles, we explored the correlation between cell composition and aging globally. We observed various specific immune cells established different abundance in different ages such as neutrophils, CD8+ T cells and other subtypes, revealing the aging phenotype was involved in the development of immune cells. We also explored the correlation between the proportion of immune cells in each tissue and the level of gene expression. Results exhibited many genes are significantly correlated with the proportion of immune cells in different tissues and age groups, which could be used as potential new cell markers. By integrating these results into ADEIP, researchers can easily compare the different immune profiles along with aging.

Overall, we present the integrated platform, ADEIP, to inspect the age-dependent expression and immune profiles in tissues, which can be applied to various diseases such as cancer or infectious diseases, and decipher the different responses and outcomes of therapy among ages. As functions of more genes are revealed in specific diseases, ADEIP will become a useful platform to study aging and diseases.

> **Key Points**
> - A comprehensive platform of age-dependent gene and immune profiles based on large-scale analysis.
> - Characterize the involvement of age on expression and immune regulation across tissues.
> - Explore the effects of age on SARS-CoV-2-associated genes and treatment.

Supplementary Data

Supplementary data are available online at *Briefings in Bioinformatics*.

Acknowledgment

Funding

References

1. Pera A, Campos C, Lopez N, *et al*. Immunosenescence: implications for response to infection and vaccination in older people. *Maturitas* 2015; **82**: 50–5.
2. Akbar AN, Henson SM, Lanna A. Senescence of T lymphocytes: implications for enhancing human immunity. *Trends Immunol* 2016; **37**: 866–76.
3. Zeng L, Yang J, Peng S, *et al*. Transcriptome analysis reveals the difference between "healthy" and "common" aging and their connection with age-related diseases. *Aging Cell* 2020; **19**: e13121.
4. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, *et al*. The impact of sex on gene expression across human tissues. *Science* 2020; **369**: eaba3066.
5. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020; **369**: 1318–30.
6. Lissner MM, Thomas BJ, Wee K, *et al*. Age-related gene expression differences in monocytes from human neonates, young adults, and older adults. *PLoS One* 2015; **10**: e0132061.
7. Jia K, Cui C, Gao Y, *et al*. An analysis of aging-related genes derived from the Genotype-Tissue Expression project (GTEx). *Cell Death Discov* 2018; **4**: 26.
8. Wang K, Wu D, Zhang H, *et al*. Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases. *Sci Rep* 2018; **8**: 10929.
9. Shaw AC, Goldstein DR, Montgomery RR. Age-dependent dysregulation of innate immunity. *Nat Rev Immunol* 2013; **13**: 875–87.
10. Pinti M, Appay V, Campisi J, *et al*. Aging of the immune system: focus on inflammation and vaccination. *Eur J Immunol* 2016; **46**: 2286–301.
11. Arunachalam PS, Wimmers F, Mok CKP, *et al*. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* 2020; **369**: 1210–20.
12. Chen YM, Zheng Y, Yu Y, *et al*. Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *EMBO J* 2020; **39**: e105896.
13. Blanco-Melo D, Nilsson-Payant BE, Liu WC, *et al*. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 2020; **181**: 1036–45.e9.
14. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45**: 580–5.
15. Dobin A, Davis CA, Schlesinger F, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**: 15–21.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.
17. Subramanian A, Tamayo P, Mootha VK, *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**: 15545–50.
18. Mi H, Muruganujan A, Ebert D, *et al*. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 2019; **47**: D419–26.
19. Liberzon A, Birger C, Thorvaldsdottir H, *et al*. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems* 2015; **1**: 417–25.
20. Bhattacharya S, Dunn P, Thomas CG, *et al*. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Scientific Data* 2018; **5**: 180015.
21. Breuer K, Foroushani AK, Laird MR, *et al*. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res* 2013; **41**: D1228–33.
22. Liu Y, He M, Wang D, *et al*. HisgAtlas 1.0: a human immuno-suppression gene database. *Database (Oxford)* 2017; **2017**: bax094.
23. Zhou H, Yang Y, Shen HB. Hum-mPLoc 3.0: prediction enhancement of human protein subcellular localization through modeling the hidden correlations of gene ontology and functional domain features. *Bioinformatics* 2017; **33**: 843–53.
24. Chou KC, Shen HB. Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. *J Proteome Res* 2007; **6**: 1728–34.
25. Newman AM, Liu CL, Green MR, *et al*. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015; **12**: 453–7.
26. Sturm G, Finotello F, List M. Immunedeconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data. *Methods Mol Biol* 2020; **2120**: 223–32.
27. Virtanen P, Gommers R, Oliphant TE, *et al*. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020; **17**: 261–72.
28. Gordon DE, Jang GM, Bouhaddou M, *et al*. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, Nature 2020; **583**: 459–468.
29. Benisch P, Schilling T, Klein-Hitpass L, *et al*. The transcriptional profile of mesenchymal stem cell populations in primary osteoporosis is distinct and shows overexpression of osteogenic inhibitors. *Plos One* 2012; **7**: e45142.
30. Zhu Z, Li J, Ruan G, *et al*. Investigational drugs for the treatment of osteoarthritis, an update on recent developments. *Expert Opin Investig Drugs* 2018; **27**: 881–900.

31. Wang K, Xu J, Hunter DJ, *et al*. Investigational drugs for the treatment of osteoarthritis. *Expert Opin Investig Drugs* 2015; **24**: 1539–56.

32. Mora AL, Rojas M, Pardo A, *et al*. Emerging therapies for idiopathic pulmonary fibrosis, a progressive age-related disease. *Nat Rev Drug Discov* 2017; **16**: 755–72.

33. Tarry-Adkins JL, Ozanne SE. Nutrition in early life and age-associated diseases. *Ageing Res Rev* 2017; **39**: 96–105.

34. Zhang J, Litvinova M, Liang Y, *et al*. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* 2020; **368**: 1481–6.

35. Qian J, Zhao L, Ye RZ, *et al*. Age-dependent gender differences in COVID-19 in Mainland China: comparative study. *Clin Infect Dis* 2020; **71**: 2488–94.

36. Han YN, Feng ZW, Sun LN, *et al*. A comparative-descriptive analysis of clinical characteristics in 2019-coronavirus-infected children and adults. *J Med Virol* 2020; **92**: 1596–602.