D862–D870 *Nucleic Acids Research, 2021, Vol. 49, Database issue* *Published online 19 November 2020*
doi: 10.1093/nar/gkaa1064

CSEA-DB: an omnibus for human complex trait and cell type associations

Yulin Dai[1,†], Ruifeng Hu[1,†], Astrid Marilyn Manuel[1], Andi Liu[2], Peilin Jia[1,*] and Zhongming Zhao[1,2,3,*]

[1]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA, [2]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA and [3]MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA

Received August 15, 2020; Revised October 18, 2020; Editorial Decision October 19, 2020; Accepted October 21, 2020

ABSTRACT

During the past decade, genome-wide association studies (GWAS) have identified many genetic variants with susceptibility to several thousands of complex diseases or traits. The genetic regulation of gene expression is highly tissue-specific and cell type-specific. Recently, single-cell technology has paved the way to dissect cellular heterogeneity in human tissues. Here, we present a reference database for GWAS trait-associated cell type-specificity, named Cell type-Specific Enrichment Analysis DataBase (CSEA-DB, available at https://bioinfo.uth.edu/CSEADB/). Specifically, we curated total of 5120 GWAS summary statistics data for a wide range of human traits and diseases followed by rigorous quality control. We further collected >900 000 cells from the leading consortia such as Human Cell Landscape, Human Cell Atlas, and extensive literature mining, including 752 tissue cell types from 71 adult and fetal tissues across 11 human organ systems. The tissues and cell types were annotated with Uberon and Cell Ontology. By applying our deTS algorithm, we conducted 10 250 480 times of trait-cell type associations, reporting a total of 598 (11.68%) GWAS traits with at least one significantly associated cell type. In summary, CSEA-DB could serve as a repository of association map for human complex traits and their underlying cell types, manually curated GWAS, and single-cell transcriptome resources.

INTRODUCTION

The past decade has witnessed great success in genome-wide association studies (GWAS) which reported thousands of genetic variants with statistical association with a few thousands of human complex diseases and traits. However, where and how these genetic factors manifest their impacts on the molecular changes remain mostly elusive. Previous studies have discovered that genetic variants tend to regulate the gene expression or function in specific tissues and cell types (1,2). The accurate assessment of disease-associated tissues or cell types becomes a critical step to understanding the etiology of these human complex diseases and traits (3,4). Recently, we successfully developed a t-statistics-based method 'decoding the tissue-specificity' (deTS) to measure the tissue-specific enrichment of 26 human complex diseases utilizing GWAS summary statistics and tissue gene expression profiling (5). Later, we expanded this method to assess the tissue-specific enrichment of ∼5000 collected GWAS over ∼70 tissues curated from the Genotype-Tissue Expression project (GTEx) (2) and Encyclopedia of DNA Elements project (ENCODE) (6). All the trait and tissue associations are stored in our Tissue-Specific Enrichment Analysis DataBase (TSEA-DB, https://bioinfo.uth.edu/TSEADB/) (7). However, due to the heterogeneity within the tissue, the bulk RNA-seq of tissue might not fully reflect the underlying biological basis.

Human bodies are composed of 11 major organ systems, ∼100 organs/tissues, and more than 100 unique cell types or thousands of sub-cell types (8). In recent years, large-scale single-cell transcriptome data have been generated by several pioneer studies (9–11) and multiple international consortia (e.g. Human Cell Atlas, Human Lung Cell Atlas, and Human Cell Landscape) (12–14). These studies aim to characterize the molecular features of the cell types in human

[*]To whom correspondence should be addressed. Tel: +1 713 500 3631; Email: zhongming.zhao@uth.tmc.edu
Correspondence may also be addressed to Peilin Jia. Tel: +1 713 500 3633; Email: peilin.jia@uth.tmc.edu
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

major tissues, which provides us rich resources to decode the cell type-specificity of human cell types and harness the genetic implications underlying human complex diseases and traits.

In this work, we aim to systematically explore the genetic signals of complex traits and diseases underlying human cell types. We conduct the following approaches. (i) We updated our GWAS summary statistics collection. (ii) We curated these large-scale single-cell transcriptome datasets of human tissues and calculate the *t*-statistics-based measurements to assess the cell type-specificity of genes within each tissue. (iii) We constructed a comprehensive association map of cell types and the human complex traits and diseases through conducting the cell type-specific enrichment analysis (CSEA) for thousands of traits we curated and maintained. (iv) We identified the trait-associated cell types, which will be good candidates to allocate the 'causal' or relevant cell types and shed light on the underlying mechanisms. (v) We highlighted those traits associated with the same cell type, which indicates the potential comorbidity and shared genetic basis (such as genes and pathways). (vi) All the curated data and associations have been managed and displayed in a user-friendly database to serve as the public repository of an omnibus map for human complex trait and cell type associations. (vii) Finally, we constructed a gene expression portal at cellular level to allow users to query and compare the relative abundance for genes of interest cross human tissues and cell types.

DATA COLLECTION, ANALYSIS AND APPLICATION

GWAS summary statistics collection and update

We adapted the ∼5000 GWAS summary statistics collected from the previous TSEA-DB (Tissue-Specific Enrichment Analysis DataBase) frozen by 19 June 2019. We further curated 260 GWAS traits from GWAS Catalog (15) and GRASP (grasp.nhlbi.nih.gov/) updated until 18 June 2020. Briefly, We collected the GWAS summary statistics from three major collections: the multi-trait collection (MTC) panel, which is a fixed collection of curations by previous studies (16–18); the UK Biobank (UKBB) panel, which deposited the UKBiobank 'GWAS round 2' results preprocessed by Neale's lab (http://www.nealelab.is/uk-biobank) on 1 August 2018 as the largest and most comprehensive resource of UKBB GWAS; the expanded trait collection (ETC), which is under recurrent curation from the new GWAS collected from GWAS Catalog and other resources. Both MTC and ETC panels were defined as the non-UKBB panel in our database.

Quality control of GWAS data

We adapted the same quality control (QC) strategies for the updated GWAS summary statistics in the ETC panel. Briefly, we only used GWAS conducted in European ancestry for this database. No trans-ethnic meta-analysis GWAS was included since no proper linkage disequilibrium information could be applied to them. We further filtered those GWAS with lambda <0.8 or >1.3 to exclude deflated and inflated studies.

Table 1. Summary of GWAS panel curation

Panel	Summary statistics	Number of TAG sets[a]
MTC[a]	432	1235
ETC[a]	316	1169
UKBB[a]	4372	11 370
Total	**5120**	**13 774**

[a] MTC: multi-trait collection panel, ETC: expanded trait collection panel, UKBB: UK Biobank collection panel, TAG: traits-associated-gene.
This table describes the distribution of 5120 GWAS summary statistics in three collection panels. The qualified GWAS traits-associated-gene (TAG) sets in each panel are listed in the third column.

Calculation of gene-based *P*-value and traits-associated-gene set

We updated our pipeline and applied a commonly used tool, Multi-marker Analysis of GenoMic Annotation (MAGMA v1.07) (19), to calculate the gene-level *P*-value. Specifically, we considered all SNPs in the gene body and 50 kb upstream and 35 kb downstream regions. We used the mean χ^2 statistic for these SNPs to obtain gene-based *P*-values, considering the effects of the gene length, SNP density, and local linkage disequilibrium (LD) structure. We used the 1000 Genome Project Phase 3 European population as the reference panel.

We further used a dynamic threshold for trait-associated-gene (TAG) sets given the different significance from each GWAS study. The gene-based *P*-value generated by MAGMA was stratified to groups by threshold $P < 0.05$, $P < 0.01$, $P < 0.001$, $P < 1 \times 10^{-4}$ and $P < 1 \times 10^{-5}$. We further limited the number of genes in each group into the range from 20 to 3000, aiming to avoid statistical significance biased by genes set size in TAG. For each GWAS study, we required at least one TAG set complying with the criteria for the number of genes. Overall, we obtained 432 in the MTC panel, 316 in ETC panel, and 4372 in the UKBB panel stored in our CSEA-DB (Table 1).

DATA COLLECTION

Human organ system tissue single-cell transcriptome data

We conducted a deep literature-mining for human tissue-single-cell transcriptome data. We downloaded the datasets from three major sources, Human Cell Landscape (http://bis.zju.edu.cn/HCL/) (14), Single Cell Expression Atlas (https://www.ebi.ac.uk/gxa/sc/home), and extensive literature curation (9–13,20–22) (Supplementary file S1). We only collected the healthy tissue single-cell transcriptome data with detailed cell type annotation by original works. Firstly, we collected and curated the transcriptome matrix by tissue. We excluded those genes expressed in <30 cells in each tissue. Considering the statistical power, we further filtered those cell types with the number of cells no smaller than 30 in each tissue. Overall, we curated 71 tissue samples (55 unique tissues) and 752 tissue cell types (TCs) in adult and fetal tissues. The total number of genes in different tissues range from 3427 in adult ascending colon to 21 758 in placenta decidua.

The cell numbers of each cell type in each study were also recorded in the database as the information for cell type distribution in each tissue. The total number of cells in each tissue panel range from 995 in pancreas to 94 257 in spleen (12,20). We also provided the resource information and download link for each study for users to download (Supplementary file S1)

Tissue-Cell type (TC) structure

We used a hierarchical structure to store the cell type in transcriptome data. Under the assumption that the cell types work collaboratively within each tissue, their tissue context should serve as another layer of information. We generated a unique tissue-cell type (TC) id for each of the cell types identified from the focal tissue single-cell data. This structure is displayed on the front page and the Browse function.

UNIFORM PROCESSING PIPELINE

We constructed a standardized pipeline to preprocess the single-cell transcriptome datasets from different resources.

Read count normalization

Read count matrix was obtained from each single-cell data resource. Pre-normalized data were also reversed to the read count matrix. Then, the matrix was normalized to counts per millions mapped reads. G_i is gene read count or Unique Molecular Identifier (UMI); N_i is the total mapped read count in each cell; 10^6 is the scale factor. The CPM matrix was further added by 1 and subsequently logistic transformed by 2 to eliminate the effect of extreme values.

$$\mathrm{CPM_i} = \frac{\mathrm{G_i}}{\frac{N_\mathrm{i}}{10^6}}$$

TISSUE AND CELL TYPE ONTOLOGY ANNOTATION

Tissue anatomy ontology

We collected 55 unique human tissues from 11 distinct organ systems of the human body according to the anatomy. We further annotated them with their id in the Uberon system (23), which is an integrated cross-species anatomical ontology system. For each study, we recorded two Uberon ids (author-inferred tissue id and CSEA-DB annotated tissue id) for each tissue. If the study provides the tissue Uberon id, we kept the information as the author-inferred tissue and set as NA for tissue without author annotation. For CSEA-DB annotated tissue, we annotated its highest level Uberon system node. This CSEA-DB annotated tissue Uberon id would be used in the later on cell type annotation to determine whether the cell type is tissue-specific. The Uberon ontology obo file was downloaded from http://purl.obolibrary.org/obo/uberon.obo (accessed on 6 July 2020). We used the 'is_a' relationship to obtain all the descendants for each of the CSEA-DB annotated tissue id as the 'tissue Uberon id set' (Figure 1B).

Cell type ontology

Another important feature of our CSEA-DB is all the cell type information adapted from the single-cell transcriptome data were further curated by CSEA-DB according to the 'inferred cell type name' from the original study. Specifically, we manually queried the 'tissue name' + 'inferred cell type name' in Cell Ontology (https://www.ebi.ac.uk/ols/ontologies/cl) (24). We recorded the most relevant cell type name and id returned from the Cell Ontology. Thus, each unique tissue cell type stored in our database have 'inferred cell type name', 'Cell Ontology id', and 'Cell Ontology full name'. All the queries at the Cell Ontology website were conducted by 27 July 2020. We also downloaded the Cell Ontology obo file from https://github.com/obophenotype/cell-ontology (accessed on July 6, 2020). This file includes all the Cell Ontology id information and some annotated with Uberon id information through 'part_of' relationship. Firstly, we obtained all the Cell Ontology ancestors id through recursive 'is_a' relationship as a union for one focal cell type. Next, we searched all the possible 'part_of' relationships with annotation of the Uberon id (Figure 1B). Thus, we obtained a 'cell type Uberon id set' for all the focal Cell Ontology id in our database.

Definition of tissue-specific cell types

Due to the wide-spreading connective tissue cell, including fixed cells (fibrocytes and adipocytes) and 'wandering cells' such as leukocytes (25,26), they might not have the cell type-specificity or directly related to the tissue function. Thus, we utilized an ontology-based method to check whether these tissue-cell types (TCs) were annotated as one cell type belongs to the corresponding tissue ontology. Specifically, for each of our 752 TCs, we overlap its Uberon id set described in the previous '**Cell type ontology**' session with its 'tissue Uberon id set'. If there is an overlapping of these two Uberon id sets, we would annotate this focal TC to tissue-specific or non-tissue-specific cell type, respectively (Figure 1B).

Application of cell type-specific enrichment analysis

We modified our previously developed 'tissue-specific enrichment analysis' (5) and applied it to explore the cell type-specificity within each tissue. Briefly, we used the $\log_2$ (CPM + 1) normalized single-cell transcriptome matrix to calculate the cell type-specific expression within the cell types (number of cells $\geq$ 30) in each tissue. The *t*-statistics (t_{ij}) for the coefficient of lm(y_i~x_j) is calculated, where y_i is a vector of the normalized expression of *i* gene; x_j is a design matrix indicating the cells either in or outside of the *j* cell type; lm is the linear model regression. Thus, t_{ij} represents the *t*-statistics of i gene in *j* cell type. Then we defined the top 5% *t*-statistic score gene in focal cell type as the cell type-specific genes. Lastly, we conducted a fisher exact test whether the TAG set from each trait is overrepresented with the cell type-specific genes, where *P*-value indicates the significance of this CSEA.

DESCRIPTION OF THE WEBSITE AND TOOLS

Overview of CSEA-DB

The front page of CSEA-DB includes an overview of human 11 organ systems. We used a hierarchical structure to

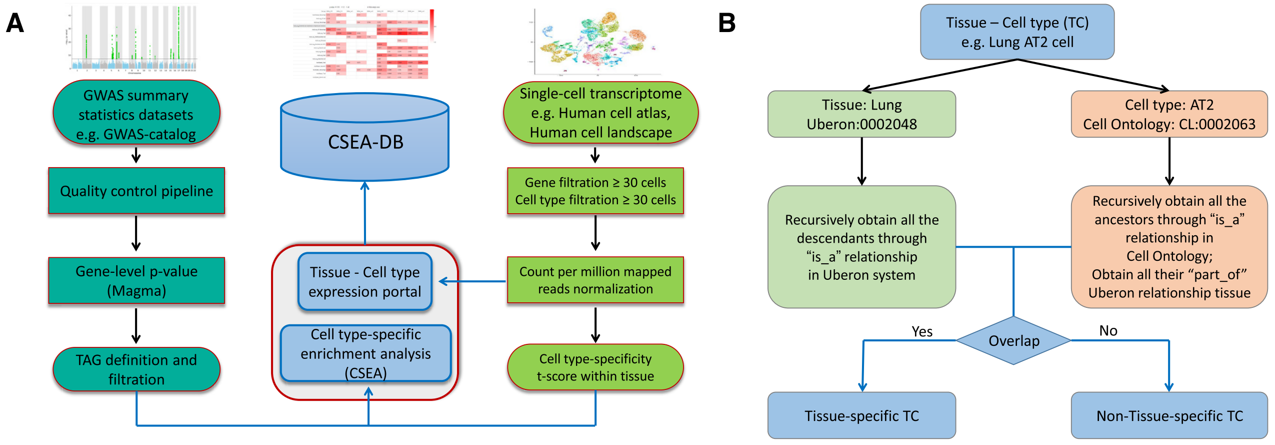

A
GWAS summary statistics datasets e.g. GWAS-catalog
Quality control pipeline
Gene-level p-value (Magma)
TAG definition and filtration
CSEA-DB
Tissue - Cell type expression portal
Cell type-specific enrichment analysis (CSEA)
Single-cell transcriptome e.g. Human cell atlas, Human cell landscape
Gene filtration ≥ 30 cells
Cell type filtration ≥ 30 cells
Count per million mapped reads normalization
Cell type-specificity t-score within tissue
B
Tissue – Cell type (TC) e.g. Lung AT2 cell
Tissue: Lung Uberon:0002048
Cell type: AT2 Cell Ontology: CL:0002063
Recursively obtain all the descendants through "is_a" relationship in Uberon system
Recursively obtain all the ancestors through "is_a" relationship in Cell Ontology; Obtain all their "part_of" Uberon relationship tissue
Overlap
Yes
No
Tissue-specific TC
Non-Tissue-specific TC

Figure 1. Workflow for Cell type-Specific Enrichment Analysis DataBase (CSEA-DB). (**A**) Workflow for CSEA-DB and statistics; Three figures (Manhattan plot, heatmap, and t-SNE plot) illustrate the features of the GWAS summary statistics dataset, Cell Type-Specific Enrichment Analysis Database (CSEA-DB), and single-cell transcriptome dataset, respectively. We process the datasets and construct the database based on this workflow. (**B**) Workflow for annotating tissue-cell types and identifying its tissue-specificity. For each tissue-cell type maintained in our database, we manually curate its ontology information in Uberon system and Cell Ontology and identify its tissue-specificity through this workflow.

store the tissue single-cell data information. The navigation bar has four featured functions 'Browse', 'Search', 'Multi-trait' comparison, and 'scExpression', Users can navigate the whole database through these four featured functions. The 'Browse' function stores all the trait and cell type maps split by trait panels (MTC, ETC and UKBB) as well as the trait-associated TCs map and TC-associated trait map. The search function supports the fuzzy search for trait name, tissue, and cell type of interest. And the return page includes all the possible results related to the keywords.

Trait-associated TCs page

In the current CSEA-DB, we conducted the CSEA for 13 774 TAG sets (size $\geq$20 and $\leq$ 3000) from 5120 GWAS data sets over 752 TCs (Table 1). We found 99.97% of TAG sets (13 770 /13 774) have at least one cell type associated with a nominal P-value <0.05. After Bonferroni correction for 13 774 TAG sets and 752 TCs ($\sim$5 $\times$ 10^{-9}), we still observed 9.92% of TAG sets (1367/13 774) having at least one cell type. At the GWAS traits level, we observed 100% (5120/5120) and 11.68% (598 /5120) of GWAS have at least one cell type with significant associations before and after multiple-testing correction. Specifically for the non-UKBB panel, we observed 100% (748 /748) and 20.86% (156 /748) of GWAS identified with the significant association before and after multiple-testing correction (Figure 2A, left). Thus, we identified that non-UKBB GWAS traits tend to have a higher proportion of Bonferroni corrected significance than the UKBB GWAS traits. Figure 2B shows an overview for one specific GWAS trait and its basic information, including trait name, case and control number, reference, and TAG sets information. The CSEA results for the trait could be displayed in all TCs or one tissue and its corresponding TCs.

TC-associated traits page

Among 752 TC, 100% TCs has at least one nominal significant association with one GWAS trait, while 61.17% TCs (460 /752) have at least one Bonferroni corrected significance with at least one GWAS phenotype (Figure 2A, right). We did not observe the tissue-specific cell types (60.51%, 118/195) have a different proportion of Bonferroni corrected significance (comparing to nominal significance) from the non-tissue-specific cell types (61.40%, 342/557). We aimed to explore the TC-associated traits both tissue-wisely and TC-wisely. As shown in Figure 2C, the TCs and traits are connected by edges with P-value of association smaller than a certain threshold, while TCs are connected with their corresponding tissue. Users can filter the P-value threshold if the network is too sparse or too dense. Below the threshold, we also provide the information of cell types and their numbers within this tissue. Users can click the hyperlink to look into the specific TC-associated traits. Interestingly, we identified some of the cell types from the same tissue that might share strong associations with the same trait, indicating that these cell types might work together to contribute to the corresponding GWAS phenotype. As shown in Figure 2D, one specific TC-associated traits page only contains the association of one TC. This page is similar to the association page for TCs or traits as shown in Figure 2C.

Bridging to TSEA-DB

In addition to our CSEA result, we provide a hyperlink to TSEA-DB for each trait to allow users to browse the traits-associated tissues (TATs). TSEA-DB provides a comprehensive overview of each trait by displaying the Manhattan plot for GWAS summary statistics and gene-level P-values. Moreover, users could compare both results of TSEA and CSEA in shared tissues, providing deeper biological insight of genetic signals at cellular resolution as

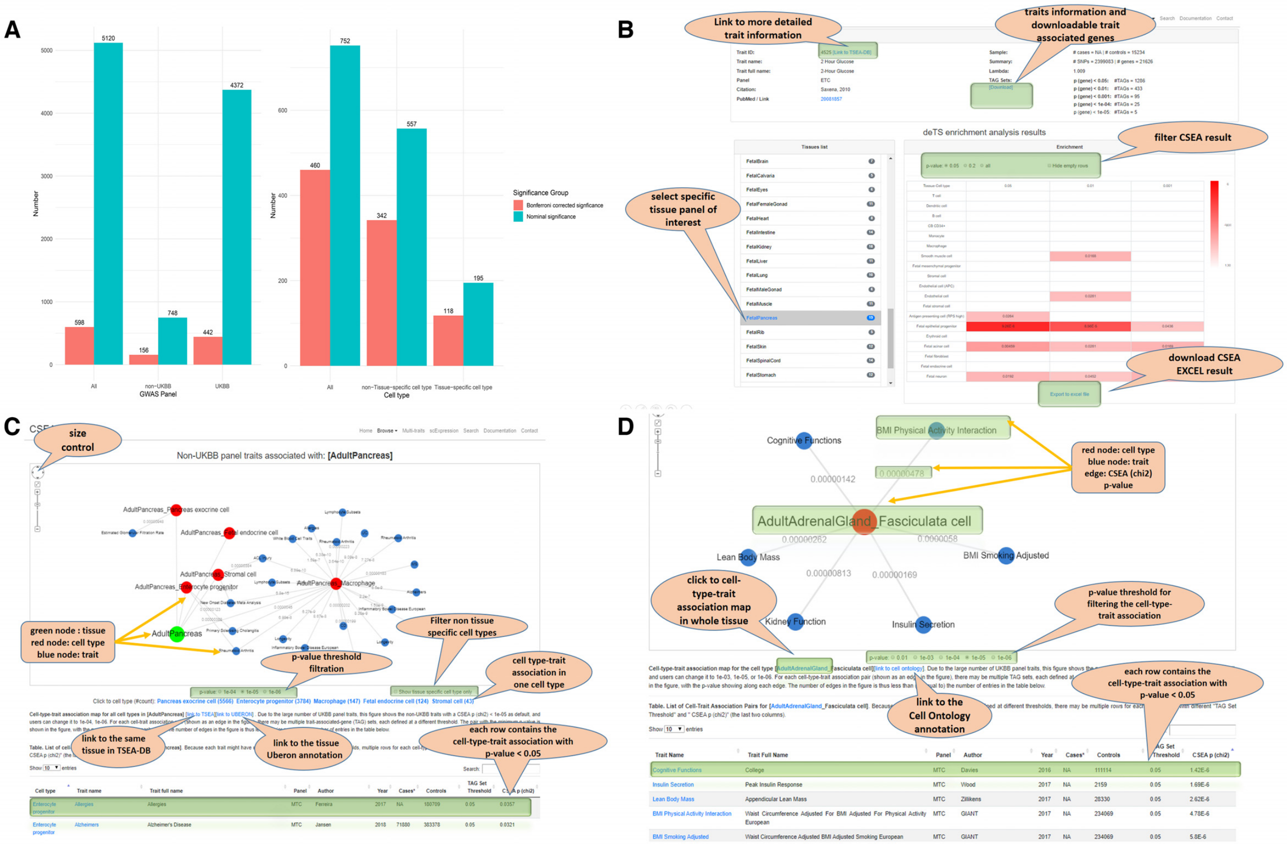

A
5120
598
156
748
442
4372
All
non-UKBB
UKBB
GWAS Panel
Number
752
460
557
342
118
195
All
non-Tissue-specific cell type
Tissue-specific cell type
Cell type
Significance Group
Bonferroni corrected significance
Nominal significance
B
Link to more detailed trait information
traits information and downloadable trait associated genes
deTS enrichment analysis results
filter CSEA result
select specific tissue panel of interest
download CSEA EXCEL result
C
size control
Non-UKBB panel traits associated with: [AdultPancreas]
green node : tissue
red node: cell type
blue node: trait
Filter non tissue specific cell types
p-value threshold filtration
cell type-trait association in one cell type
link to the same tissue in TSEA-DB
link to the tissue Uberon annotation
each row contains the cell-type-trait association with p-value < 0.05
D
Cognitive Functions
BMI Physical Activity Interaction
AdultAdrenalGland_Fasciculata cell
Lean Body Mass
BMI Smoking Adjusted
Kidney Function
Insulin Secretion
red node: cell type
blue node: trait
edge: CSEA (chi2) p-value
click to cell-type-trait association map in whole tissue
p-value threshold for filtering the cell-type-trait association
link to the Cell Ontology annotation
each row contains the cell-type-trait association with p-value < 0.05

Figure 2. Statistics of trait and tissue-cell type association map and vignettes for database functions. (**A**) Trait and tissue-cell type map statistics; left: number of trait and tissue-cell type associations with at least one nominal or Bonferroni corrected significant association in at least one of the 752 tissue-cell types (TCs); y-axis is the number of unique traits; x-axis is the three categories, all GWAS traits, UK Biobank (UKBB) panel traits, and non-UKBB panel traits; right: number of trait and tissue-cell type associations with at least one nominal or Bonferroni corrected significant association in at least one of 5,120 GWAS traits; y-axis is the number of unique TCs; x-axis is three categories, all TCs, non-tissue-specific TCs, and tissue-specific TCs. (**B**) Cell type-specific enrichment analysis (CSEA) result from browsing the GWAS trait page. Top rows show trait information, including trait name, links to related sources, study summary, and trait-associated-gene (TAG) sets used for CSEA. The heatmap table in the bottom shows the enriched results of the trait in different cell types. Users can check the results in different tissues. The heatmap table is interactive and allows users to set different *P*-value thresholds for displaying the enriched results. The empty row which means there is no enriched result at the selected *P*-value level can be hidden by checking the option box. If the user selects the option 'all', all the enriched result values will be displayed in the heatmap table, only values that are lower than 0.2 will be marked with color proportions. (**C**) Network view of tissue-cell types-associated traits. Green node denotes a tissue, red nodes are cell types in this focal tissue, and blue nodes are associated traits with the cell types. The edges are labeled with *P*-values. The network view can be modified by setting different *P*-value cutoffs or hide non-specific cell types. A short note shows the description of the network data source. At the bottom, the data table shows a full list of cell type-trait association pairs for the tissue. Because each trait might have multiple TAG sets defined at different thresholds, multiple rows for each cell type-trait pair with different 'TAG Set Threshold' and 'CSEA p (chi2)' (the last two columns). (**D**) Network view of one specific cell type associated traits. This page is similar to the page in (C).

well as rescuing those diluted signals in the bulk tissue expression.

scExpression page

The scExpression page provides normalized gene expression information at both tissue and cell type level. Users can submit one gene symbol name of interest and the scExpression page will return an overview (barplot) of the average CPM normalized gene expression within each of the 68 tissues with UMI-based data. The three SMART-seq2 datasets based on the full-length RNA-seq method are not listed, as their normalized expression profiles are not comparable with other UMI-based data (27,28). Once the users further click the tissue bar of interest, it will generate barplot for the average CPM normalized expression of each cell type in that tissue. This page is a useful tool for researchers who aim to check the gene expression across human tissues and cell types. Overall, we provide a one-stationary curation for the average normalized gene expression of tissue and cell type.

Multi-trait comparison function

As we demonstrated in the trait-associated cell type, containing >5120 traits information and their CSEA result across over 752 TCs tissue-wisely or together. Since there are multiple studies for the same or similar trait, the comparison of these traits could help to identify the consistent enriched cell types shared by the same trait from different studies. Moreover, multiple tissues or TCs might be related to complex traits and diseases. Thus, we provide this multi-trait comparison function to explore the shared or unique cell types across multiple traits and multiple TCs of interest.

To this end, this multi-trait comparison function could explore the associations between multiple traits and multiple TCs of interests simultaneously. We used 'Ashma' as an example in Figure 3. Asthma is a condition that leads to the inflammation in the airways and the bronchial tubes that carry air into the lung, which makes patients difficult to breathe (29). We selected two Asthma studies with 10 TAG sets and all TCs in three tissues, including two disease-relevant tissue trachea and lung, and one disease irrelevant tissue adipose through the multi-trait comparison function.

We identified the same cell types might act differently in disease-relevant and irrelevant tissues. For instance, we found macrophage has most significant association with $P > 0.005$ in adipose, $P = 6.98 \times 10^{-5}$ in lung, and $P = 2.86 \times 10^{-5}$ in trachea. We also identified endothelial cell (APC) has relatively high risk in both lung ($P = 2.29 \times 10^{-4}$) and trachea ($P = 1.37 \times 10^{-5}$), although T cell in lung has the most significant association ($P = 7.71 \times 10^{-7}$). Nevertheless, we also find the lung B cell has the association of $P = 1.05 \times 10^{-4}$, while the plasma B cell has no association with $P < 0.05$ in any of 10 TAG sets. Overall, this systematic characterization of the TCs underlying genetic signals could help to dissect the potential mechanism of asthma.

TSEA and CSEA comparison

To understand the relationship between TSEA and CSEA, we compare their results for one autoimmune disease (Crohn's Disease) in TSEA-DB and CSEA-DB (Figure 4A and B) (4). Specifically, we identified five associated tissues (adipose visceral, lung, small intestine, whole blood, and spleen) that have at least one TAG with $P < 0.01$ in the TSEA-DB. (Figure 4A). The top associated tissues are whole blood ($P = 3.08 \times 10^{-10}$), spleen ($P = 5.27 \times 10^{-7}$), lung ($P = 6.59 \times 10^{-7}$), small intestine ($P = 0.001$), and adipose visceral ($P = 0.008$). We selected the CSEA result from the corresponding tissues (AdultAdipose, AdultDuodenum, AdultIleum, AdultJeJunum, AdultLung, AdultPeripheralBlood and AdultSpleen) in CSEA-DB (Figure 4B). Since the cell type-specificity is calculated within each tissue, their significance is comparable within each focal tissue. We identified that dendritic cell (7.81×10^{-9}), M2 macrophage (2.01×10^{-8}) and M2 macrophage (1.57×10^{-6}) are the top significant cell types in lung. Interestingly, we found dendritic cell (3.20×10^{-6}), macrophage (5.47×10^{-6}), and endothelial cell (APC) (3.18×10^{-6}) are the most enriched cell types in three small intestine tissues (disease-relevant tissues). Moreover, endothelial cells (APC) in non-small intestine tissues all have $P > 0.001$, suggesting that the endothelial cell (APC) might work differently in small intestine tissues and contribute to the Crohn's disease along with the dendritic cell, and macrophage. Besides, the proportions of 'causal' cell types within each tissue could also be an important indicator to assess their effect. Lastly, the genetic factors could only explain a small proportion of the disease. The microenvironment that tissues are exposed to might have contributions to the disease pathogenesis.

Documentation page

We built a documentation page to briefly describe the data collection, preprocessing, and analysis (https://bioinfo.uth.edu/CSEADB/document.php). We also provide vignettes for website functions, including Browse, Search, Multi-trait comparison and scExpression functions.

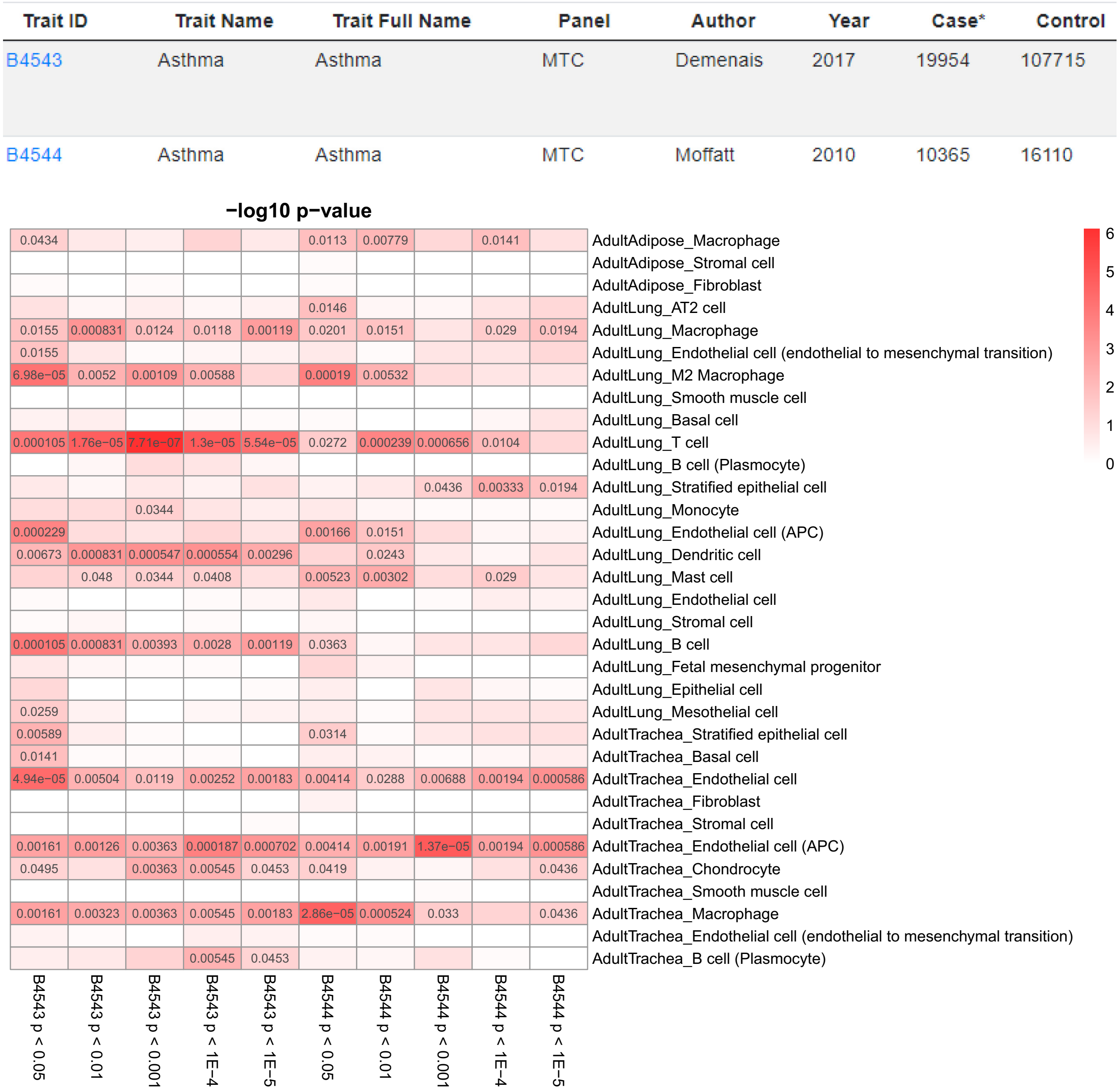
DATABASE DESIGN AND UPDATES

The CSEA-DB web interfaces were constructed with standard HTML and Bootstrap 4 libraries (http://getbootstrap.com/). The data were processed using R and python scripts. The processed and annotated data and summary statistics were stored in MySQL. PHP was the main language used for implementing the functions of CSEA-DB, such as, browsing, searching and data exporting. The interactive and dynamic web pages were implemented through several JavaScript libraries (CytoscapJS, zTReeJS, HighchartJS) and Ajax strategies. Our database could be easily expanded with the newly updated data through highly efficient scripts. In the previous work TSEA-DB (7), we have built a standardized workflow to select, preprocess, and conduct quality control on the GWAS summary statistics. In this work, we used that pipeline and updated 171 qualified GWAS summary statistics since the last update on 19 June 2019. We will update GWAS summary statistics annually. In this work, we built a pipeline to process the single-cell tissue transcriptome data, including data quality control, ontology annotation and cell type-specific expression panel construction. Due to rapid advances in single-cell genomics technology, we will seasonally update our single-cell dataset collection to include the newly available human tissues and related cell types. The CSEA could be implemented using modified scripts based on our deTS package (5). All the results in this work are based on the release on 13 August 2020.

CONCLUDING REMARKS AND FUTURE DEVELOPMENT

In previous TSEA-DB (https://bioinfo.uth.edu/TSEADB/), we have successfully decoded the diseases relevant tissues, most of which aligned with the current knowledge. In this Cell type-Specific Enrichment Analysis DataBase (CSEA-DB, https://bioinfo.uth.edu/CSEADB/), we mine the various data deeper into the trait-associated tissue cell types. We updated and reprocessed the long-termly maintained GWAS summary statistics with the new pipeline. We collected, annotated, and processed 71 single-cell transcriptome studies covering 55 unique adult and fetal tissues in 11 human organ systems.

Interestingly, we observed a broad expression of immune cells among human tissues. As shown in Figure 4C, we found human macrophage, T cell, dendritic cell existed in 74.1%, 63.8%, 63.8% out of 58 tissue from HCL, respectively. This discovery supports the widespread connective tissue cells such as macrophage, T cell, and dendritic cell in human tissues. Moreover, we summed up the cellular proportion of top five lymphatic cell types within each human tissues and identified that ascending colon (85.3%), adipose (82.3%), spleen (81.0%), peripheral blood (70.9%),



Trait ID	Trait Name	Trait Full Name	Panel	Author	Year	Case*	Control
B4543	Asthma	Asthma	MTC	Demenais	2017	19954	107715
B4544	Asthma	Asthma	MTC	Moffatt	2010	10365	16110

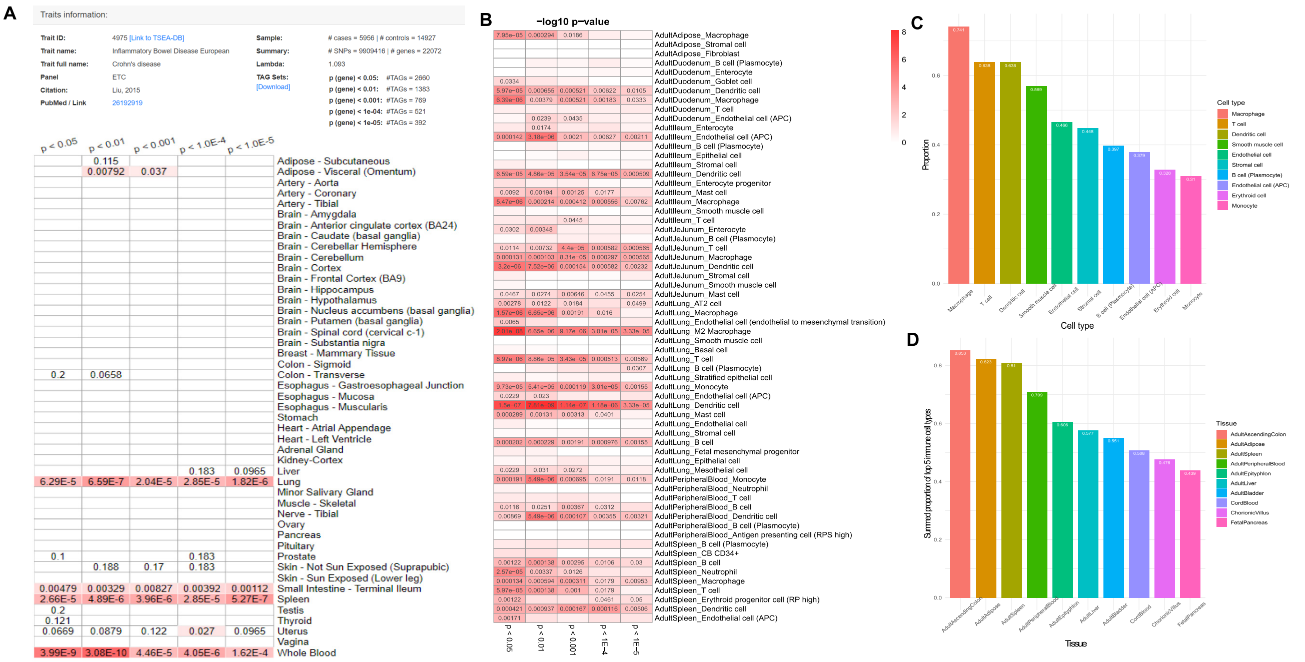
−log10 p−value

B4543 p < 0.05	B4543 p < 0.01	B4543 p < 0.001	B4543 p < 1E−4	B4543 p < 1E−5	B4544 p < 0.05	B4544 p < 0.01	B4544 p < 0.001	B4544 p < 1E−4	B4544 p < 1E−5	
0.0434					0.0113	0.00779		0.0141		AdultAdipose_Macrophage
										AdultAdipose_Stromal cell
										AdultAdipose_Fibroblast
					0.0146					AdultLung_AT2 cell
0.0155	0.000831	0.0124	0.0118	0.00119	0.0201	0.0151		0.029	0.0194	AdultLung_Macrophage
0.0155										AdultLung_Endothelial cell (endothelial to mesenchymal transition)
6.98e−05	0.0052	0.00109	0.00588		0.00019	0.00532				AdultLung_M2 Macrophage
										AdultLung_Smooth muscle cell
										AdultLung_Basal cell
0.000105	1.76e−05	7.71e−07	1.3e−05	5.54e−05	0.0272	0.000239	0.000656	0.0104		AdultLung_T cell
										AdultLung_B cell (Plasmocyte)
							0.0436	0.00333	0.0194	AdultLung_Stratified epithelial cell
		0.0344								AdultLung_Monocyte
0.000229					0.00166	0.0151				AdultLung_Endothelial cell (APC)
0.00673	0.000831	0.000547	0.000554	0.00296		0.0243				AdultLung_Dendritic cell
	0.048	0.0344	0.0408		0.00523	0.00302		0.029		AdultLung_Mast cell
										AdultLung_Endothelial cell
										AdultLung_Stromal cell
0.000105	0.000831	0.00393	0.0028	0.00119	0.0363					AdultLung_B cell
										AdultLung_Fetal mesenchymal progenitor
										AdultLung_Epithelial cell
0.0259										AdultLung_Mesothelial cell
0.00589					0.0314					AdultTrachea_Stratified epithelial cell
0.0141										AdultTrachea_Basal cell
4.94e−05	0.00504	0.0119	0.00252	0.00183	0.00414	0.0288	0.00688	0.00194	0.000586	AdultTrachea_Endothelial cell
										AdultTrachea_Fibroblast
										AdultTrachea_Stromal cell
0.00161	0.00126	0.00363	0.000187	0.000702	0.00414	0.00191	1.37e−05	0.00194	0.000586	AdultTrachea_Endothelial cell (APC)
0.0495		0.00363	0.00545	0.0453	0.0419				0.0436	AdultTrachea_Chondrocyte
										AdultTrachea_Smooth muscle cell
0.00161	0.00323	0.00363	0.00545	0.00183	2.86e−05	0.000524	0.033		0.0436	AdultTrachea_Macrophage
										AdultTrachea_Endothelial cell (endothelial to mesenchymal transition)
			0.00545	0.0453						AdultTrachea_B cell (Plasmocyte)

Figure 3. Multi-trait multi-cell types comparison using two asthma studies in three human tissues (Adult Adipose, Adult Lung and Adult Trachea). An example of our Multi-trait function by comparing the asthma GWAS from two datasets in three tissues (33 tissue-cell types). Our 'Multi-trait' function allows users to compare at most 10 data sets in all the tissue-types in our database at one time. The top two rows demonstrate the study information about these traits and the heatmap below is the cell type-specific enrichment analysis (CSEA) for the GWAS with multiple trait-associated-gene (TAG) sets. The color is proportional to the –log(P) value of the associations. The cells with $P < 0.05$ are filled with the corresponding P-values and the tables with $P \geq 0.05$ are left blank.

and epityphlon (appendix) (60.6%) are the top five human tissues that contain the largest proportion of these five lymphatic cells (macrophage, T cell, dendritic cell, B cell (plasmocyte) and monocyte) (Figure 4D). Considering the widespreading lymphatic cells in human tissues, we did a systematic curation for cell types identified in the transcriptome data with those connective cells in each tissue. However, the 'tissue-specific' and 'non-tissue-specific' cell types are not a rigorous biological definition. We provided this cell type filtration option to eliminate the potential 'noise' from the non-tissue-specific connective cells such as fibrocytes, adipocytes, and leukocytes cells, highlighting the associations of 'tissue-specific' cell types. Overall, CSEA-DB could provide systematic potential insights into the biological mechanisms of human complex diseases at cellular resolution.

In the future, CSEA-DB aims to update the database in the following three directions. (i) Current CSEA-DB only curates the GWAS with samples of European Ancestry. With the recent advent of GWAS in other popu-



Traits information:

Trait ID:	4975 [Link to TSEA-DB]	**Sample:**	# cases = 5956 | # controls = 14927
Trait name:	Inflammatory Bowel Disease European	**Summary:**	# SNPs = 9909416 | # genes = 22072
Trait full name:	Crohn's disease	**Lambda:**	1.093
Panel	ETC	**TAG Sets:** [Download]	**p (gene) < 0.05:** #TAGs = 2660 **p (gene) < 0.01:** #TAGs = 1383 **p (gene) < 0.001:** #TAGs = 769 **p (gene) < 1e-04:** #TAGs = 521 **p (gene) < 1e-05:** #TAGs = 392
Citation:	Liu, 2015		
PubMed / Link	26192919		

	p < 0.05	p < 0.01	p < 0.001	p < 1.0E-4	p < 1.0E-5
Adipose - Subcutaneous		0.115			
Adipose - Visceral (Omentum)		0.00792	0.037		
Artery - Aorta					
Artery - Coronary					
Artery - Tibial					
Brain - Amygdala					
Brain - Anterior cingulate cortex (BA24)					
Brain - Caudate (basal ganglia)					
Brain - Cerebellar Hemisphere					
Brain - Cerebellum					
Brain - Cortex					
Brain - Frontal Cortex (BA9)					
Brain - Hippocampus					
Brain - Hypothalamus					
Brain - Nucleus accumbens (basal ganglia)					
Brain - Putamen (basal ganglia)					
Brain - Spinal cord (cervical c-1)					
Brain - Substantia nigra					
Breast - Mammary Tissue					
Colon - Sigmoid					
Colon - Transverse	0.2	0.0658			
Esophagus - Gastroesophageal Junction					
Esophagus - Mucosa					
Esophagus - Muscularis					
Stomach					
Heart - Atrial Appendage					
Heart - Left Ventricle					
Adrenal Gland					
Kidney-Cortex					
Liver				0.183	0.0965
Lung	6.29E-5	6.59E-7	2.04E-5	2.85E-5	1.82E-6
Minor Salivary Gland					
Muscle - Skeletal					
Nerve - Tibial					
Ovary					
Pancreas					
Pituitary					
Prostate	0.1			0.183	
Skin - Not Sun Exposed (Suprapubic)		0.188	0.17	0.183	
Skin - Sun Exposed (Lower leg)					
Small Intestine - Terminal Ileum	0.00479	0.00329	0.00827	0.00392	0.00112
Spleen	2.66E-5	4.89E-6	3.96E-6	2.85E-5	5.27E-7
Testis	0.2				
Thyroid	0.121				
Uterus	0.0669	0.0879	0.122	0.027	0.0965
Vagina					
Whole Blood	3.99E-9	3.08E-10	4.46E-5	4.05E-6	1.62E-4

	p < 0.05	p < 0.01	p < 0.001	p < 1E−4	p < 1E−5
AdultAdipose_Macrophage	7.95e−05	0.000294	0.0186		
AdultAdipose_Stromal cell					
AdultAdipose_Fibroblast					
AdultDuodenum_B cell (Plasmocyte)					
AdultDuodenum_Enterocyte					
AdultDuodenum_Goblet cell	0.0334				
AdultDuodenum_Dendritic cell	5.97e−05	0.000655	0.000521	0.00622	0.0105
AdultDuodenum_Macrophage	6.39e−06	0.00379	0.000521	0.00183	0.0333
AdultDuodenum_T cell					
AdultDuodenum_Endothelial cell (APC)		0.0239	0.0435		
AdultIleum_Enterocyte		0.0174			
AdultIleum_Endothelial cell (APC)	0.000142	3.18e−06	0.0021	0.00627	0.00211
AdultIleum_B cell (Plasmocyte)					
AdultIleum_Epithelial cell					
AdultIleum_Stromal cell					
AdultIleum_Dendritic cell	6.59e−05	4.86e−05	3.54e−05	6.75e−05	0.000509
AdultIleum_Enterocyte progenitor					
AdultIleum_Mast cell	0.0092	0.00194	0.00125	0.0177	
AdultIleum_Macrophage	5.47e−06	0.000214	0.000412	0.000556	0.00762
AdultIleum_Smooth muscle cell					
AdultIleum_T cell			0.0445		
AdultJeJunum_Enterocyte	0.0302	0.00348			
AdultJeJunum_B cell (Plasmocyte)					
AdultJeJunum_T cell	0.0114	0.00732	4.4e−05	0.000582	0.000565
AdultJeJunum_Macrophage	0.000131	0.000103	8.31e−05	0.000297	0.000565
AdultJeJunum_Dendritic cell	3.2e−06	7.52e−06	0.000154	0.000582	0.00232
AdultJeJunum_Stromal cell					
AdultJeJunum_Smooth muscle cell					
AdultJeJunum_Mast cell	0.0467	0.0274	0.00646	0.0455	0.0254
AdultLung_AT2 cell	0.00278	0.0122	0.0184		0.0499
AdultLung_Macrophage	1.57e−06	6.65e−06	0.00191	0.016	
AdultLung_Endothelial cell (endothelial to mesenchymal transition)	0.0065				
AdultLung_M2 Macrophage	2.01e−08	6.65e−06	9.17e−06	3.01e−05	3.33e−05
AdultLung_Smooth muscle cell					
AdultLung_Basal cell					
AdultLung_T cell	8.97e−06	8.86e−05	3.43e−05	0.000513	0.00569
AdultLung_B cell (Plasmocyte)					0.0307
AdultLung_Stratified epithelial cell					
AdultLung_Monocyte	9.73e−05	5.41e−05	0.000119	3.01e−05	0.00155
AdultLung_Endothelial cell (APC)	0.0229	0.023			
AdultLung_Dendritic cell	1.5e−07	7.81e−09	1.14e−07	1.18e−06	3.33e−05
AdultLung_Mast cell	0.000289	0.00131	0.00313	0.0401	
AdultLung_Endothelial cell					
AdultLung_Stromal cell					
AdultLung_B cell	0.000202	0.000229	0.00191	0.000976	0.00155
AdultLung_Fetal mesenchymal progenitor					
AdultLung_Epithelial cell					
AdultLung_Mesothelial cell	0.0229	0.031	0.0272		
AdultPeripheralBlood_Monocyte	0.000191	5.49e−06	0.000695	0.0191	0.0118
AdultPeripheralBlood_Neutrophil					
AdultPeripheralBlood_T cell					
AdultPeripheralBlood_B cell	0.0116	0.0251	0.00367	0.0312	
AdultPeripheralBlood_Dendritic cell	0.00869	5.49e−06	0.000107	0.00355	0.00321
AdultPeripheralBlood_B cell (Plasmocyte)					
AdultPeripheralBlood_Antigen presenting cell (RPS high)					
AdultSpleen_B cell (Plasmocyte)					
AdultSpleen_CB CD34+					
AdultSpleen_B cell	0.00122	0.000138	0.00295	0.0106	0.03
AdultSpleen_Neutrophil	2.57e−05	0.00337	0.0126		
AdultSpleen_Macrophage	0.000134	0.000594	0.000311	0.0179	0.00953
AdultSpleen_T cell	5.97e−05	0.000138	0.001	0.0179	
AdultSpleen_Erythroid progenitor cell (RP high)	0.00122			0.0461	0.05
AdultSpleen_Dendritic cell	0.000421	0.000937	0.000167	0.000116	0.00506
AdultSpleen_Endothelial cell (APC)	0.00171				

Figure 4. Comparison between TSEA-DB and CSEA-DB for Crohn's disease, and cell type distribution among tissues in CSEA-DB. (**A**) Tissue-specific enrichment analysis (TSEA) results for one Crohn's disease GWAS across 47 tissues in Genotype-Tissue Expression (GTEx). Top rows are the basic information about the Crohn's disease GWAS, the heatmap below demonstrates the TSEA results of this Crohn's disease GWAS with multiple trait-associated-gene (TAG) sets from the Tissue-Specific Enrichment Analysis DataBase (TSEA-DB). (**B**) Cell type-specific enrichment analysis (CSEA) results of Crohn's disease GWAS in tissue-cell type level with multiple TAG sets from CSEA-DB. (**C**) Barplot for top 10 frequently identified cell types in CSEA-DB curated tissue transcriptome datasets.). (**D**) Barplot for top 10 tissues with the largest summed proportions of the five lymphatic cells (macrophage, T cell, dendritic cell, B cell (plasmocyte), and monocyte).

lations (e.g. African and East Asian population), we will integrate the more population panel into our GWAS curation as well as the annual update referring to GWAS-catalog. Moreover, the chromatin interaction information for chromosome 3D data has also been integrated to better interpret the effect of variants in long-term chromosomal interactions (30,31). We will actively update our pipeline to better assess the gene-level *P*-value from GWAS summary statistics. (ii) We will continuously collect and curate the emerging single-cell transcriptome data quarterly. We will have a more comprehensive collection of human single-cell transcriptome data along with the advance of ongoing projects like Human cell atlas and single-cell expression atlas with more data and more accurate annotation. Since more single-cell platforms adapted the UMI-based method (32), we expect to integrate more single-cell transcriptome datasets (e.g. developmental and temporal-spatial data) into our database. (iii) In this study, we observed that some specific complex human disease is significantly associated with multiple cell types in one specific tissue, suggesting these cell types from this tissue might all contribute to the etiology of diseases. As shown in Figure 3, we found that both immune cells and epithelial cells (AT2) are all extremely enriched in human adult lung, indicating their co-occurrence might be associated with the underlying mechanism. Nevertheless, different cells within certain microenvironments will communicate with each other and coordinate to transduce signals (such as immune response in immune cells) (33). Deciphering the genetic risks underlying such intercellular interactions will further help us to understand the etiology of human complex traits and diseases (34).

In summary, we constructed an omnibus map for over 5120 human GWAS phenotypes and 752 human tissue cell types. We identified many tissue-specific cell types that play crucial roles; and such results align with previous discoveries in TSEA-DB. Meanwhile, some widespread cell types might play different roles in different tissues and contribute to the disease pathogenesis contextually. Moreover, tissue cell type could be related to multiple GWAS phenotypes and multiple cell types carrying genetic risks within one tissue might communicate and work collaboratively. These discoveries could provide new insights into understanding the mechanism of human complex traits and diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Xiaoqian Jiang for providing computational resource support in our data process. We would like to thank all the members of Bioinformatics and Systems Medicine Laboratory (BSML) for valuable discussion. All the tissue illustration was created with BioRender.com.

FUNDING

National Institutes of Health grant [R01LM012806] (in part); Cancer Genomics Core funded by the Cancer Prevention and Research Institute of Texas (CPRIT) [RP180734]; Data Science and Informatics Core for Cancer Research by the Cancer Prevention and Research Institute of Texas [RP170668]; the funder had no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. Funding for open access charge: [R03DE027711 to Jia].

Conflict of interest statement. None declared.

REFERENCES

1. Finucane,H.K., Reshef,Y.A., Anttila,V., Slowikowski,K., Gusev,A., Byrnes,A., Gazal,S., Loh,P.R., Lareau,C., Shoresh,N. *et al.* (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.*, **50**, 621–629.
2. GTEx Consortium; Laboratory, Data Analysis and Coordinating Center (LDACC)-Analysis Working Group; Statistical Methods groups-Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
3. Dai,Y., Hu,R., Pei,G., Zhang,H., Zhao,Z. and Jia,P. (2020) Diverse types of genomic evidence converge on alcohol use disorder risk genes. *J. Med. Genet.*, **57**, 733–743.
4. Dai,Y., Pei,G., Zhao,Z. and Jia,P. (2019) A convergent study of genetic variants associated with crohn's disease: evidence from GWAS, gene expression, methylation, eQTL and TWAS. *Front. Genet.*, **10**, 318.
5. Pei,G., Dai,Y., Zhao,Z. and Jia,P. (2019) deTS: tissue-specific enrichment analysis to decode tissue specificity. *Bioinformatics*, **35**, 3842–3845.
6. Davis,C.A., Hitz,B.C., Sloan,C.A., Chan,E.T., Davidson,J.M., Gabdank,I., Hilton,J.A., Jain,K., Baymuradov,U.K., Narayanan,A.K. *et al.* (2018) The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.*, **46**, D794–D801.
7. Jia,P., Dai,Y., Hu,R., Pei,G., Manuel,A.M. and Zhao,Z. (2020) TSEA-DB: a trait-tissue association map for human complex traits and diseases. *Nucleic Acids Res.*, **48**, D1022–D1030.
8. Tortora,G.J. and Derrickson,B.H. (2017) In: *Introduction to the Human Body*. John Wiley & Sons.
9. Lake,B.B., Chen,S., Sos,B.C., Fan,J., Kaeser,G.E., Yung,Y.C., Duong,T.E., Gao,D., Chun,J., Kharchenko,P.V. *et al.* (2018) Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.*, **36**, 70–80.
10. MacParland,S.A., Liu,J.C., Ma,X.Z., Innes,B.T., Bartczak,A.M., Gage,B.K., Manuel,J., Khuu,N., Echeverri,J., Linares,I. *et al.* (2018) Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.*, **9**, 4383.
11. Voigt,A.P., Mulfaul,K., Mullin,N.K., Flamme-Wiese,M.J., Giacalone,J.C., Stone,E.M., Tucker,B.A., Scheetz,T.E. and Mullins,R.F. (2019) Single-cell transcriptomics of the human retinal pigment epithelium and choroid in health and macular degeneration. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 24100–24107.
12. Madissoon,E., Wilbrey-Clark,A., Miragaia,R.J., Saeb-Parsy,K., Mahbubani,K.T., Georgakopoulos,N., Harding,P., Polanski,K., Huang,N., Nowicki-Osuch,K. *et al.* (2019) scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.*, **21**, 1.
13. Vieira Braga,F.A., Kar,G., Berg,M., Carpaij,O.A., Polanski,K., Simon,L.M., Brouwer,S., Gomes,T., Hesse,L., Jiang,J. *et al.* (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.*, **25**, 1153–1163.
14. Han,X., Zhou,Z., Fei,L., Sun,H., Wang,R., Chen,Y., Chen,H., Wang,J., Tang,H., Ge,W. *et al.* (2020) Construction of a human cell landscape at single-cell level. *Nature*, **581**, 303–309.
15. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
16. Liu,B., Gloudemans,M.J., Rao,A.S., Ingelsson,E. and Montgomery,S.B. (2019) Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.*, **51**, 768–769.
17. Pickrell,J.K., Berisa,T., Liu,J.Z., Segurel,L., Tung,J.Y. and Hinds,D.A. (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.*, **48**, 709–717.
18. Pei,G., Sun,H., Dai,Y., Liu,X., Zhao,Z. and Jia,P. (2019) Investigation of multi-trait associations using pathway-based analysis of GWAS summary statistics. *BMC Genomics*, **20**, 79.
19. de Leeuw,C.A., Mooij,J.M., Heskes,T. and Posthuma,D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.
20. Segerstolpe,A., Palasantza,A., Eliasson,P., Andersson,E.M., Andreasson,A.C., Sun,X., Picelli,S., Sabirsh,A., Clausen,M., Bjursell,M.K. *et al.* (2016) Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.*, **24**, 593–607.
21. Vento-Tormo,R., Efremova,M., Botting,R.A., Turco,M.Y., Vento-Tormo,M., Meyer,K.B., Park,J.E., Stephenson,E., Polanski,K., Goncalves,A. *et al.* (2018) Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature*, **563**, 347–353.
22. Enge,M., Arda,H.E., Mignardi,M., Beausang,J., Bottino,R., Kim,S.K. and Quake,S.R. (2017) Single-Cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell*, **171**, 321–330.
23. Mungall,C.J., Torniai,C., Gkoutos,G.V., Lewis,S.E. and Haendel,M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.*, **13**, R5.
24. Diehl,A.D., Meehan,T.F., Bradford,Y.M., Brush,M.H., Dahdul,W.M., Dougall,D.S., He,Y., Osumi-Sutherland,D., Ruttenberg,A., Sarntivijai,S. *et al.* (2016) The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics*, **7**, 44.
25. Alitalo,K. (2011) The lymphatic vasculature in disease. *Nat. Med.*, **17**, 1371–1380.
26. Halper,J. and Kjaer,M. (2014) Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. *Adv. Exp. Med. Biol.*, **802**, 31–47.
27. Picelli,S., Faridani,O.R., Bjorklund,A.K., Winberg,G., Sagasser,S. and Sandberg,R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.*, **9**, 171–181.
28. Vieth,B., Parekh,S., Ziegenhain,C., Enard,W. and Hellmann,I. (2019) A systematic evaluation of single cell RNA-seq analysis pipelines. *Nat. Commun.*, **10**, 4667.
29. Kudo,M., Ishigatsubo,Y. and Aoki,I. (2013) Pathology of asthma. *Front. Microbiol.*, **4**, 263.
30. Sey,N.Y.A., Hu,B., Mah,W., Fauni,H., McAfee,J.C., Rajarajan,P., Brennand,K.J., Akbarian,S. and Won,H. (2020) A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.*, **23**, 583–593.
31. Wang,Q., Chen,R., Cheng,F., Wei,Q., Ji,Y., Yang,H., Zhong,X., Tao,R., Wen,Z., Sutcliffe,J.S. *et al.* (2019) A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nat. Neurosci.*, **22**, 691–699.
32. Hagemann-Jensen,M., Ziegenhain,C., Chen,P., Ramskold,D., Hendriks,G.J., Larsson,A.J.M., Faridani,O.R. and Sandberg,R. (2020) Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.*, **38**, 708–714.
33. Cohen,M., Giladi,A., Gorki,A.D., Solodkin,D.G., Zada,M., Hladik,A., Miklosi,A., Salame,T.M., Halpern,K.B., David,E. *et al.* (2018) Lung Single-Cell signaling interaction map reveals basophil role in macrophage imprinting. *Cell*, **175**, 1031–1044.
34. Efremova,M., Vento-Tormo,M., Teichmann,S.A. and Vento-Tormo,R. (2020) CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.*, **15**, 1484–1506.