


Article

# scRNASeqDB: A Database for RNA-Seq Based Gene Expression Profiles in Human Single Cells

Yuan Cao <sup>1,†</sup>, Junjie Zhu <sup>2,†</sup>, Peilin Jia <sup>1</sup> and Zhongming Zhao <sup>1,3,4,\*</sup> 

<sup>1</sup> Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; yuan.cao@uth.tmc.edu (Y.C.); Peilin.Jia@uth.tmc.edu (P.J.)

<sup>2</sup> Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA; jjzhu@stanford.edu

<sup>3</sup> Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>4</sup> Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

\* Correspondence: zhongming.zhao@uth.tmc.edu; Tel.: +1-713-500-3631

† These authors contributed equally to this work.

Received: 13 November 2017; Accepted: 30 November 2017; Published: 5 December 2017

**Abstract:** Single-cell RNA sequencing (scRNA-Seq) is rapidly becoming a powerful tool for high-throughput transcriptomic analysis of cell states and dynamics at the single cell level. Both the number and quality of scRNA-Seq datasets have dramatically increased recently. A database that can comprehensively collect, curate, and compare expression features of scRNA-Seq data in humans has not yet been built. Here, we present scRNASeqDB, a database that includes almost all the currently available human single cell transcriptome datasets ( $n = 38$ ) covering 200 human cell lines or cell types and 13,440 samples. Our online web interface allows users to rank the expression profiles of the genes of interest across different cell types. It also provides tools to query and visualize data, including Gene Ontology and pathway annotations for differentially expressed genes between cell types or groups. The scRNASeqDB is a useful resource for single cell transcriptional studies. This database is publicly available at <https://bioinfo.uth.edu/scrnaseqdb/>.

**Keywords:** single cell; RNA sequencing; database; expression profile; cell type; differential expression

## 1. Introduction

Single-cell RNA sequencing (scRNA-Seq) is rapidly evolving as a powerful tool for high-throughput transcriptomic analysis of cell states and dynamics [1]. So far, most gene expression studies in the literature have been based on the averaged expression information from bulk tissue [2]. scRNA-Seq can accurately measure RNA expression characteristics at the single cell level in order to explore cell phenotype, function, genetic alterations in various conditions, and transcriptomic heterogeneity. While it is a new technology, both the quality and quantity of scRNA-Seq data have dramatically increased during the past a few years [3–9]. Such an overwhelming volume of public scRNA-Seq data requires effective quality control, platform and analysis assessment, data mining and integration, and information management. Therefore, there is strong need for a centralized web resource that curates and provides features such as single-cell gene expression profiles among various cell types and studies. Such a resource will be beneficial to the broad biological and biomedical research communities.

So far, there have been three reported web resources for single cell transcriptome data, but all of them were developed for mouse data [10–12]. Du et al. [10] developed a web resource which only included one single-cell gene expression dataset that was involved in the development of mouse lung tissue. Similarly, Nestorowa et al. [11] built a web interface using only one single-cell transcriptome dataset for hematopoietic stem and progenitor cells (HPSC), providing insights into the differentiation

of blood stem cell in mice. Finally, Biase et al. [12] developed an online genome browser for mouse scRNA-Seq data. Recently, the single-cell portal for brain research was launched by the Broad Institute of MIT and Harvard ([https://portals.broadinstitute.org/single\\_cell](https://portals.broadinstitute.org/single_cell)). This portal was developed to facilitate sharing scientific results and disseminate datasets resulting from the National Institutes of Health (NIH) Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative, which contains mostly brain related scRNA-Seq datasets in humans and mice. To date, it includes 39 studies covering 234,551 cells. However, there are only 15 shared datasets that are freely accessible, and they contain no more than 20 human cell lines or cell types. Although there are some databases that offer abundant transcriptomic information for one or multiple cell lines or cell types, an in-depth investigation of transcriptional features and data comparison of the transcriptomes across different studies in humans are still lacking [13,14].

To the best of our knowledge, there is no dedicated database for the comprehensive curation and annotation of human scRNA-Seq data yet. Here, we developed human single-cell RNA-Seq database (scRNASeqDB), a database that collects and curates publicly available single cell gene expression datasets for humans. Through metadata annotation and unified data processing procedures, it covers 200 human cell groups and 13,440 samples (as of 1 September 2016). These data and the follow-up analysis results are available through a user-friendly web interface. First, scRNASeqDB allows ranking gene expression across different datasets; that is, the user can analyze a specific gene across different cell lines and cell types. Second, based on the transcriptional features of each cell group in one dataset, we developed a visualization interface that allows the flexible display and comparison of user-specified gene(s). Third, scRNASeqDB provides users with the gene rank list for a specific cell group. This database is publicly available at <https://bioinfo.uth.edu/scrnaseqdb/>. It helps researchers within the fields of biology and medicine to facilitate gene expression studies in human single cells.

## 2. Materials and Methods

### 2.1. Data Collection and Metadata Annotations

We searched the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database [15] for gene expression profiling experiments using the following keywords: scRNAseq, single-cell RNA-Seq, and single-cell transcriptome. Such strategies have been successfully utilized in our previous work and in a wide range of databases developed by our group [16–18] and by others [19–22]. A detailed description is provided in Table S1. Next, we carefully reviewed the resultant papers and datasets. A total of 38 datasets for human single-cell RNA-Seq analysis was obtained from this systematic search (Table S2). The number of samples ranged between 7 and 3162. The number of scRNA-Seq accessible experiments in humans has increased dramatically since 2012, with 2016 alone contributing 5348 samples. These datasets included 200 human cell groups, which are related to reproduction, immune system, the brain and nervous system, cancer, and stem cells, among others (Table S3). The metadata of GEO datasets were downloaded and imported into MySQL using the R package GEOquery [23] and RMySQL, respectively. Metadata of cell types and experiment conditions were manually curated according to the characteristics of each dataset or description in the original publications. For RNA-Seq experiments, the gene expression matrices were also retrieved from the GEO and converted to Transcripts Per Million (TPM) or read count format by using our in-house R scripts (available upon request). For cells in datasets where the fragments per kilobase of exon per million reads mapped (FPKM) were available, we computed the TPM for gene  $i$ , according to:

$$\text{TPM}_i = \left( \frac{\text{FPKM}_i}{\sum_j \text{FPKM}_j} \right) \times 10^6 \quad (1)$$

For datasets that provided expression data with unique molecular identifier (UMI) values instead of FPKM values, we similarly applied the same normalization as we did for TPM calculation for each cell. This conversion enables the units to be consistent for dataset-to-dataset comparison.

In addition, the sample accession number for each sample was added, which uniquely identifies the exact experiment in the NCBI GEO dataset [24].

## 2.2. Rank of Genes

RankProd is a non-parametric analysis tool that employs the rank of the expression value of genes to prioritize those genes [25]. In scRNASeqDB, RankProd was adopted to compare gene expression level across different scRNA-Seq datasets. Each gene was assigned an order in the specific cell lines or cell types. A low gene order means it is highly expressed in the cell line or cell type of the dataset in examination. In our database, we used a horizontal bar plot to represent a gene's rank in each cell line or cell type. Specifically, we defined the rank of a gene in a particular cell line or cell type as follows:

$$\frac{\text{total gene number} - \text{gene order}}{\text{total gene number}} \times 100 \quad (2)$$

For each gene, its rank in all datasets was presented using a bar plot, providing the users with an overview and direct comparison of the gene's expression rank across the different datasets.

## 2.3. Differential Gene Expression

For particular studies that contained more than one cell line or experimental condition, such as those that included pancreatic islet acinar versus pancreatic islet duct [26], or those that included iPSC 409B2 (41 days) versus iPSC 409B2 (65 days), we conducted differentially expressed (DE) gene analysis between cell groups for each dataset. BPSC, representing Beta-Poisson model for Single-Cell RNA-seq data analyses, [27] was employed to analyze the DE genes. BPSC is an analysis tool based on the beta-Poisson model for single-cell gene expression data. It addresses practical and realistic issues such as non-integer expression values or low expression values. Utilizing BPSC, we analyzed DE genes in all 38 datasets by parallel computing in a server with 16 CPUs and 256 GB RAM. The total running time was 25.7 days. The genes were identified as significantly differentially expressed at a false discovery rate (FDR) <0.05 [28]. Finally, we deposited these DE genes into our database and made them available to users at the bottom section in 'Dataset View'.

## 2.4. Web Interface

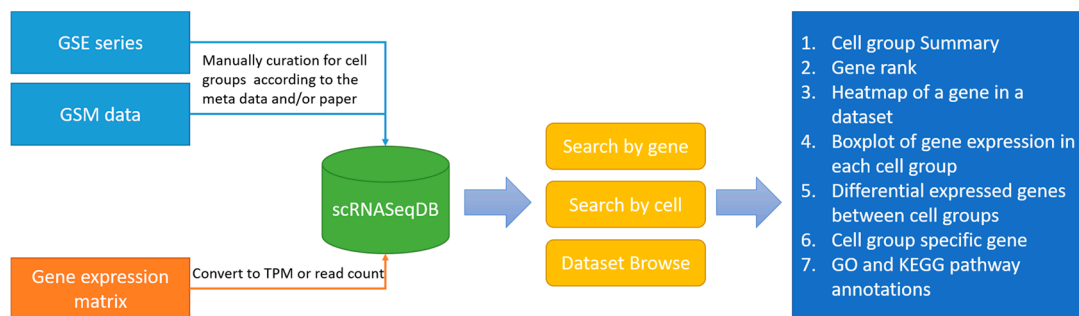
Currently, scRNASeqDB release 1.0 has collected 13,440 samples belonging to 200 human cell groups from 38 datasets. The web interface of scRNASeqDB was implemented in PHP and JavaScript using the Yii framework, which enables users to search across the database easily without requiring much computer expertise. Interactive heatmap and box plots were constructed dynamically to display gene expression for individual cells and cell groups in one dataset using the HighCharts component. Users can export the charts as images in PDF, PNG, JPG, or SVG format for record or further analysis.

# 3. Results

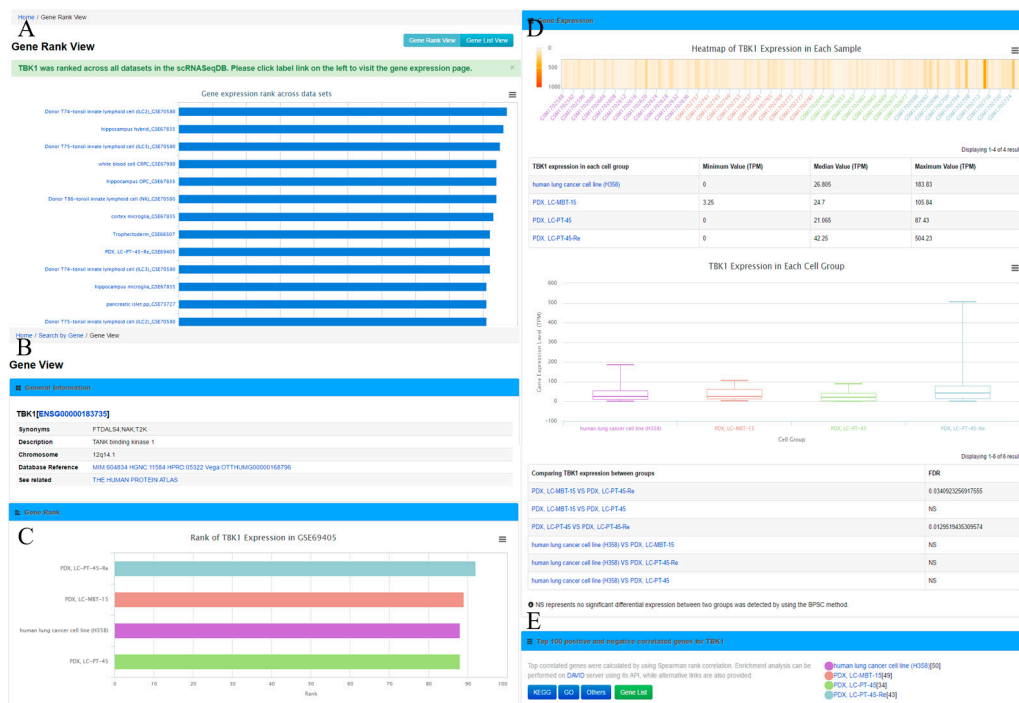
## 3.1. Database Description and Summary of Features

The main purpose of scRNASeqDB is to facilitate the analysis and visualization of gene expression profiles across various human single cells based on the public datasets. The flowchart of data collection and database construction is provided in Figure 1. We have collected 38 single-cell RNA-seq datasets of humans covering 200 cell groups in the scRNASeqDB. The homepage of the web resource provides three ways to search the database. First, the user can search the gene of interest using gene symbols or gene Ensembl IDs. A successful search will lead the user to the 'Gene Rank List' webpage (Figure 2A), which displays the gene rank across different cell lines or types in this database. The 'Gene View' page provides detailed information of the gene expression within the dataset (Figure 2B–E). Second, the user can search the cell group of interest by inputting the name of a cell line or type. As an outcome, it will generate a 'Cell View' webpage with rich information. Finally, the user can browse the datasets by

choosing a dataset ID; the system will generate a ‘Dataset View’ webpage. The user can also hit the tags of some popular genes in the ‘Gene Cloud’ section to start a quick gene search.



**Figure 1.** Flowchart of data collection and database construction. GSE: accession number of series; GSM: accession number for samples; TPM: Transcripts Per Million.



**Figure 2.** Gene Rank View and Gene View. (A) Gene rank list of an example gene *TBK1* across all datasets; (B) Display general information of *TBK1* in ‘Gene View’ page; (C) Show the gene rank of *TBK1* in each cell group of GSE69405; (D) The section of gene expression displayed by heatmap and box plots. The minimum value, median value, and maximum value of the gene expression and the FDR (false discovery rate) of the gene expression comparison between groups are also displayed; (E) The top 100 positively and negatively correlated genes and annotation links are provided.

The ‘Gene View’ page displays the expression profile of a gene. It consists of four sections. (1) The general information section includes gene symbol, synonyms, Ensembl gene ID, description, and chromosome locations, among others. To further enable the exploration of gene information, this section provides web links of the gene to other online resources, such as OMIM [29], Ensembl [30], HPRD [31], and Vega [32]; (2) In the gene rank section, a bar plot shows the specific gene rank across different cell types/groups in this study; (3) In the gene expression section, an interactive heatmap is available to display gene expression across individual cells in one dataset. A box plot and a table summarizing the minimum, median, and maximum value of the gene expression in each cell group

from each dataset is also provided. If there are more than one cell group in a dataset, a comparison of query gene expression among cell groups will also be listed to indicate significantly differentially expressed genes; (4) It also shows the top 100 positively and negatively correlated genes across cells in a dataset for the query gene. Pathway annotations from Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) [33] for these genes are provided as well. In addition, the user can obtain the top 100 positively and negatively correlated genes by clicking the 'Gene List' button for further analysis.

In the 'Cell View' webpage (Figure 3), the user can explore the ranked genes in the query cell group and the relationship between two genes in this specific cell group. In the first section, a summary of the cells and experiments are provided based on the metadata of the GEO. In the second section, a table is used to present the information of cell samples in the cell group.

Home / Search by Cell / Cell View

## A Cell View

**Cell Information**

**Donor T74-tonsil innate lymphoid cell (NK)**

**Overall Design** Sequencing libraries were prepared from FACS sorted individual ILCs with the Smart-Seq2 protocol (Picelli et al. Nature Methods 2013)

**Summary** Single cell RNA-sequencing of human tonsil Innate lymphoid cells (ILCs) from three independent tonsil donors.

**Dataset view** GSE70580

**PMID** 26878113

## B Samples in Donor T74-tonsil innate lymphoid cell (NK)

Displaying 1-10 of 25 results.

Series	Sample	Instrument	Organism	Title	Cell Source
GSE70580	GSM1810511	Illumina HiSeq 2000	Homo sapiens	T74_P1_B4_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810513	Illumina HiSeq 2000	Homo sapiens	T74_P1_B9_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810520	Illumina HiSeq 2000	Homo sapiens	T74_P1_E6_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810521	Illumina HiSeq 2000	Homo sapiens	T74_P1_E7_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810528	Illumina HiSeq 2000	Homo sapiens	T74_P1_G4_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810529	Illumina HiSeq 2000	Homo sapiens	T74_P1_G9_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810531	Illumina HiSeq 2000	Homo sapiens	T74_P2_B4_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810532	Illumina HiSeq 2000	Homo sapiens	T74_P2_B9_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810543	Illumina HiSeq 2000	Homo sapiens	T74_P2_D11_NK	Donor T74-tonsil innate lymphoid cell (NK)
GSE70580	GSM1810544	Illumina HiSeq 2000	Homo sapiens	T74_P2_D2_NK	Donor T74-tonsil innate lymphoid cell (NK)

« < 1 2 3 > »

## C Gene rank in Donor T74-tonsil innate lymphoid cell (NK)

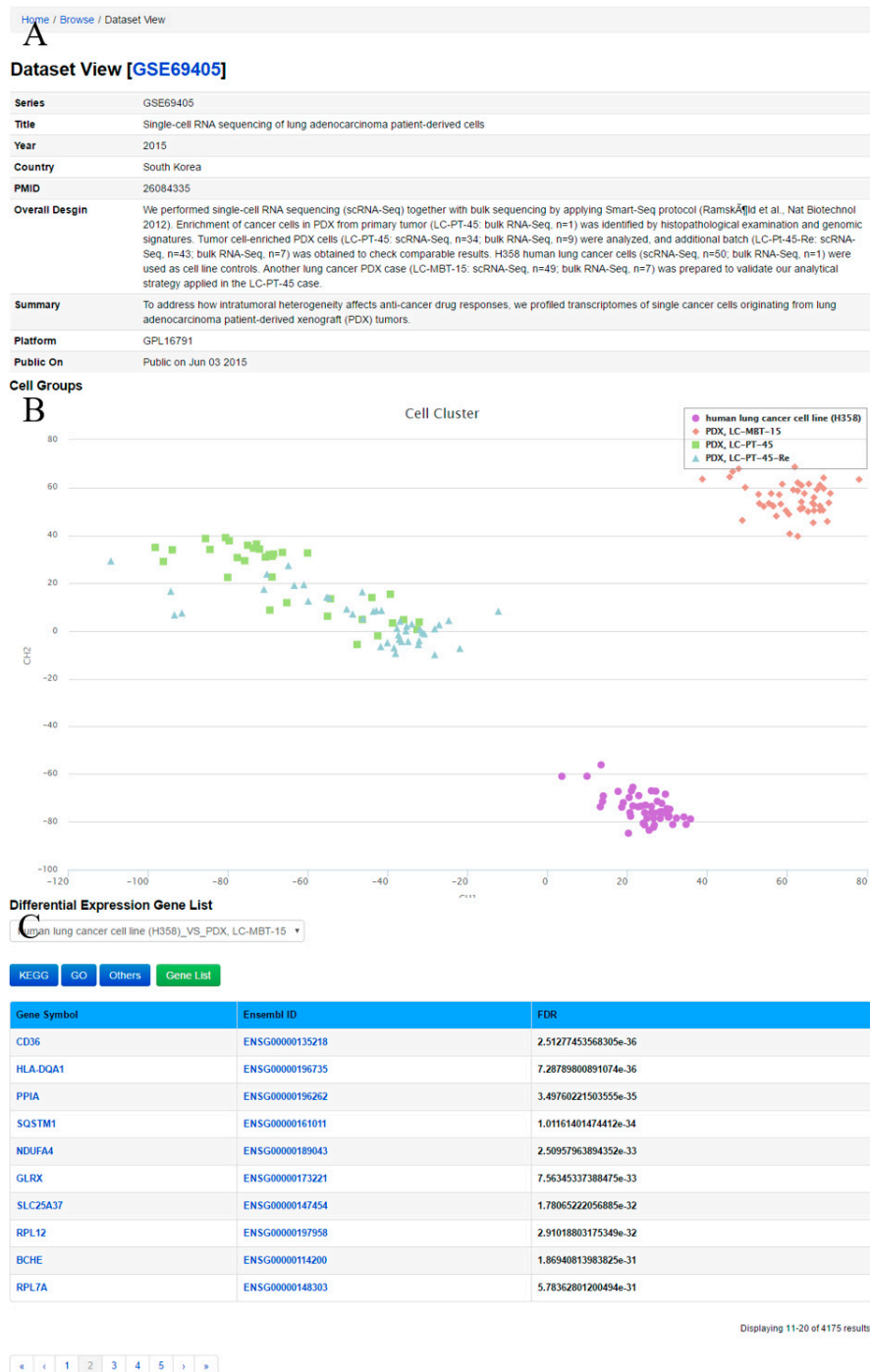
Displaying 1-10 of 31933 results.

Rank order	Gene Symbol	Ensembl ID
1	B2M	ENSG00000166710
2	TMSB4X	ENSG00000205542
3	J01415.4	
4	HLA-C	ENSG00000204525
5	MT-CO2	ENSG00000198712
6	MALAT1	ENSG00000251562
7	MT-CO1	ENSG00000198804
8	RPL3P4	ENSG00000232573
9	MT-ND4L	ENSG00000212907
10	MT-CYB	ENSG00000198727

« < 1 2 3 4 5 > »

**Figure 3.** Cell View, displaying tonsil innate lymphoid cell (NK) information as an example. (A) Summary information; (B) Metadata of cell samples; (C) The top 200 upregulated genes and correlation matrices for these genes for the specific cell type or group. Note that multiple pages will be displayed when the sample or gene list is long.

Through ‘Dataset View’ (Figure 4), the user can obtain a comprehensive description and cell groups in the query dataset. The method Clustering through Imputation and Dimensionality Reduction (CIDR) [34] is used to construct cell clusters based on the datasets. Meanwhile, lists of differentially expressed genes between two cell groups in the datasets as well as GO and KEGG pathway annotations for differentially expressed genes are displayed.



**Figure 4.** Dataset View, displaying dataset GSE69405 as an example. (A) Metadata information for the dataset; (B) Cell groups and cell cluster are presented; (C) A list of differentially expressed genes between different groups is provided.

### 3.2. Example

Many genes demonstrate a tissue-specific expression profile [24,35]. For example, Leucine-rich-alpha-2-glycoprotein1 (LRG1) is encoded by an oncogene that was recently found to be vital to the progression of human cancer [36]. *LRG1* mRNA level is upregulated in most hepatocellular carcinoma (HCC) cell lines [36]. Searching *LRG1* expression in scRNASeqDB confirmed similar expression in single cells, with the highest expression rank (top 4% rank) observed in liver cancer cells (Figure S1). This example indicates that scRNASeqDB can be used as a tool to identify tissue-selective genes in a single cell.

## 4. Discussion

scRNASeqDB is a user-friendly database that timely collects and curates single cell gene expression profiles in human cells. The database currently includes 38 datasets covering the gene expression of 13,440 single cells from 200 cell groups. It provides various features such as gene expression in different cell types, expression patterns at the pathway levels, and tools for the visualization and exploration of gene expression in single cells. In order to provide investigators with a tool to accurately determine under which condition(s) and in which cell line(s) and cell type(s) the expression of a gene of interest is altered, we developed analysis tools for the study of the dynamics of gene expression regulation across cell lines and types. The online results can be easily saved in different file formats for other purposes. Furthermore, scRNASeqDB was designed to facilitate the identification of specific gene markers for each cell subpopulation by using differential gene analysis.

A common task in many single-cell studies is to detect differentially expressed genes between cell populations [37]. To the best of our knowledge, there are some methods that are designed specifically for scRNA-Seq data, such as scDD (a statistical approach for identifying differential distributions in single-cell RNA-seq experiments) [38], D3E (discrete distributional differential expression) [39], MAST (Model-based Analysis of Single-cell Transcriptomics) [40], SCDE (a set of statistical methods for analyzing single-cell RNA-seq data) [41], and BPSC [27]. In Jaakkola's method evaluation [37], the authors reported that ROTS (reproducibility-optimized test statistic) [42] had the best performance after they compared SCDE, MAST, DESeq, Limma, and ROTS using three benchmark datasets. In our other study, we compared the performance across several tools in scRNA-Seq differential expression analysis. Although scDD showed better performance in some datasets, it took much more time when the sample size is not small (unpublished data). BPSC shows comparable performance with higher sensitivity, specificity, and reproducibility, as well as being less time-consuming. Therefore, we employed BPSC for DE analysis in our database.

With the rapid evolution of single-cell sequencing technologies, we expect an even faster generation of scRNA-Seq data in the next a few years. We plan to update the database by following our in-house pipeline. Updated keywords representing scRNA-Seq studies will be incorporated to include newly arising datasets. Our pipeline will also be continuously updated to keep with the development of new data formats or data types. In addition, we provide function to allow users to submit their own single-cell RNA-seq data. The user can conduct this task by submitting the request for review to the web administrator through our web interface. Upon approval, the dataset along with related information will be imported to this database. Such data, along with the data we retrieve from public domains like GEO, will be analyzed in our future database release. Second, gene expression dataset from relevant bulk tissue will be integrated for the assessment of expression features and biases between scRNA-Seq and traditional RNA-Seq. Third, a Gene Set Enrichment Analysis (GSEA) module will be integrated into scRNASeqDB to allow the user to perform advanced gene feature analysis based on the scRNA-Seq data. In addition to these future expansions, our database will provide an overview of the features of gene expression in various single cells, providing useful and general information on expression in different cell lines and tissues. Thus, when data eventually increase to beyond our maintenance capabilities and become sufficient for a general resource of scRNA-Seq for representative

cell types and tissues, we believe that we will have accomplished our endeavor for such purposes and conclude our development of this database.

## 5. Conclusions

In summary, we have created scRNASeqDB to allow users to search, analyze, compare, and visualize the gene expression from most available scRNA-Seq data so that they can seek deeper insights into gene expression across human single cells. scRNASeqDB will be regularly maintained to include future scRNA-Seq datasets, both from the user's direct submission to the scRNASeqDB and from the annotation of scRNA-Seq data from public resources such as GEO. More functions will be added to our database's web interface as well. So far, this is the first single-cell RNASeq database for humans.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2073-4425/8/12/368/s1](http://www.mdpi.com/2073-4425/8/12/368/s1). Figure S1: *LRG1* gene expression rank across scRNA-Seq datasets. The figure was generated by searching *LRG1* from scRNASeqDB and saved in one of the available formats (PDF, PNG, JPG, or SVG). Table S1: Summary of the search results in the PubMed database by different search strategies (by 1 July 2016). Table S2: Summary of datasets in scRNASeqDB. Table S3: Summary of cell lines or types in scRNASeqDB.

**Acknowledgments:** This work was partially supported by National Institutes of Health grants (R01LM012806 and R01LM011177), China Scholarship Council, National Natural Science Foundation of China (81572620), and the Shandong Provincial Natural Science Foundation of China (ZR2015HM003). We thank the IT support from the School of Biomedical Informatics, the University of Texas Health Science Center at Houston, and Guangchun Han for his help of some data processing.

**Author Contributions:** Y.C., Z.Z. and P.J. conceived and designed the experiments; Y.C. and J.Z. performed the data process, annotation, and database/web development; Y.C. and J.Z. analyzed the data; Y.C., Z.Z., J.Z. and P.J. wrote the manuscript. All authors read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Linnarsson, S.; Teichmann, S.A. Single-cell genomics: Coming of age. *Genome Biol.* **2016**, *17*, 97. [[CrossRef](#)] [[PubMed](#)]
2. Shalek, A.K.; Satija, R.; Adiconis, X.; Gertner, R.S.; Gaublomme, J.T.; Raychowdhury, R.; Schwartz, S.; Yosef, N.; Malboeuf, C.; Lu, D.; et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **2013**, *498*, 236–240. [[CrossRef](#)] [[PubMed](#)]
3. Dixit, A.; Parnas, O.; Li, B.; Chen, J.; Fulco, C.P.; Jerby-Arnon, L.; Marjanovic, N.D.; Dionne, D.; Burks, T.; Raychowdhury, R.; et al. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **2016**, *167*, 1853.e17–1866.e17. [[CrossRef](#)] [[PubMed](#)]
4. Achim, K.; Pettit, J.-B.; Saraiva, L.R.; Gavriouchkina, D.; Larsson, T.; Arendt, D.; Marioni, J.C. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **2015**, *33*, 503–509. [[CrossRef](#)] [[PubMed](#)]
5. Yu, Y.; Tsang, J.C.H.; Wang, C.; Clare, S.; Wang, J.; Chen, X.; Brandt, C.; Kane, L.; Campos, L.S.; Lu, L.; et al. Single-cell RNA-seq identifies a PD-1(hi) ILC progenitor and defines its development pathway. *Nature* **2016**, *539*, 102–106. [[CrossRef](#)] [[PubMed](#)]
6. Tirosh, I.; Venteicher, A.S.; Hebert, C.; Escalante, L.E.; Patel, A.P.; Yizhak, K.; Fisher, J.M.; Rodman, C.; Mount, C.; Filbin, M.G.; et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* **2016**, *539*, 309–313. [[CrossRef](#)] [[PubMed](#)]
7. Stubbington, M.J.T.; Lönnberg, T.; Proserpio, V.; Clare, S.; Speak, A.O.; Dougan, G.; Teichmann, S.A. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **2016**, *13*, 329–332. [[CrossRef](#)] [[PubMed](#)]
8. Macosko, E.Z.; Basu, A.; Satija, R.; Nemeshe, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A.R.; Kamitaki, N.; Martersteck, E.M.; et al. highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **2015**, *161*, 1202–1214. [[CrossRef](#)] [[PubMed](#)]



9. Klein, A.M.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D.A.; Kirschner, M.W. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **2015**, *161*, 1187–1201. [[CrossRef](#)] [[PubMed](#)]
10. Du, Y.; Guo, M.; Whitsett, J.A.; Xu, Y. “LungGENS”: A web-based tool for mapping single-cell gene expression in the developing lung. *Thorax* **2015**, *70*, 1092–1094. [[CrossRef](#)] [[PubMed](#)]
11. Nestorowa, S.; Hamey, F.K.; Pijuan Sala, B.; Diamanti, E.; Shepherd, M.; Laurenti, E.; Wilson, N.K.; Kent, D.G.; Göttgens, B. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **2016**, *128*, e20–e31. [[CrossRef](#)] [[PubMed](#)]
12. Biase, F.H.; Cao, X.; Zhong, S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res.* **2014**, *24*, 1787–1796. [[CrossRef](#)] [[PubMed](#)]
13. Sheng, X.; Wu, J.; Sun, Q.; Li, X.; Xian, F.; Sun, M.; Fang, W.; Chen, M.; Yu, J.; Xiao, J. MTD: A mammalian transcriptomic database to explore gene expression and regulation. *Brief. Bioinform.* **2017**, *18*, 28–36. [[CrossRef](#)] [[PubMed](#)]
14. Ye, F.; Huang, W.; Guo, G. Studying hematopoiesis using single-cell technologies. *J. Hematol. Oncol.* **2017**, *10*, 27. [[CrossRef](#)] [[PubMed](#)]
15. Clough, E.; Barrett, T. The gene expression omnibus database. *Stat. Genom. Methods Protoc.* **2016**, *1418*, 93–110. [[CrossRef](#)]
16. Kim, P.; Zhao, J.; Lu, P.; Zhao, Z. mutLBSgeneDB: Mutated ligand binding site gene DataBase. *Nucleic Acids Res.* **2017**, *45*, D256–D263. [[CrossRef](#)] [[PubMed](#)]
17. Wang, J.; Liu, T.; Zhao, B.; Lu, Q.; Wang, Z.; Cao, Y.; Li, W. sRNATarBase 3.0: An updated database for sRNA-target interactions in bacteria. *Nucleic Acids Res.* **2016**, *44*, D248–D253. [[CrossRef](#)] [[PubMed](#)]
18. Jia, P.; Han, G.; Zhao, J.; Lu, P.; Zhao, Z. SZGR 2.0: A one-stop shop of schizophrenia candidate genes. *Nucleic Acids Res.* **2017**, *45*, D915–D924. [[CrossRef](#)] [[PubMed](#)]
19. Li, J.-R.; Sun, C.-H.; Li, W.; Chao, R.-F.; Huang, C.-C.; Zhou, X.J.; Liu, C.-C. Cancer RNA-Seq Nexus: A database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res.* **2016**, *44*, D944–D951. [[CrossRef](#)] [[PubMed](#)]
20. Mei, S.; Qin, Q.; Wu, Q.; Sun, H.; Zheng, R.; Zang, C.; Zhu, M.; Wu, J.; Shi, X.; Taing, L.; et al. Cistrome Data Browser: A data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.* **2017**, *45*, D658–D662. [[CrossRef](#)] [[PubMed](#)]
21. Holtman, I.R.; Noback, M.; Bijlsma, M.; Duong, K.N.; van der Geest, M.A.; Ketelaars, P.T.; Brouwer, N.; Vainchtein, I.D.; Eggen, B.J.L.; Boddeke, H.W.G.M. Glia Open Access Database (GOAD): A comprehensive gene expression encyclopedia of glia cells in health and disease. *GLIA* **2015**, *63*, 1495–1506. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, Q.; Yang, B.; Chen, X.; Xu, J.; Mei, C.; Mao, Z. Renal Gene Expression Database (RGED): A relational database of gene expression profiles in kidney disease. *J. Biol. Database Curation* **2014**, *2014*. [[CrossRef](#)] [[PubMed](#)]
23. Davis, S.; Meltzer, P.S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **2007**, *23*, 1846–1847. [[CrossRef](#)] [[PubMed](#)]
24. Shaul, Y.D.; Yuan, B.; Thiru, P.; Nutter-Upham, A.; McCallum, S.; Lanzkron, C.; Bell, G.W.; Sabatini, D.M. MERAV: A tool for comparing gene expression across human tissues and cell types. *Nucleic Acids Res.* **2016**, *44*, D560–D566. [[CrossRef](#)] [[PubMed](#)]
25. Hong, F.; Breitling, R.; McEntee, C.W.; Wittner, B.S.; Nemhauser, J.L.; Chory, J. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **2006**, *22*, 2825–2827. [[CrossRef](#)] [[PubMed](#)]
26. Li, J.; Klughammer, J.; Farlik, M.; Penz, T.; Spittler, A.; Barbieux, C.; Berishvili, E.; Bock, C.; Kubicek, S. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* **2016**, *17*, 178–187. [[CrossRef](#)] [[PubMed](#)]
27. Vu, T.N.; Wills, Q.F.; Kalari, K.R.; Niu, N.; Wang, L.; Rantalainen, M.; Pawitan, Y. Beta-Poisson model for single-cell RNA-seq data analyses. *Bioinformatics* **2016**, *32*, 2128–2135. [[CrossRef](#)] [[PubMed](#)]
28. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, *57*, 289–300.
29. Amberger, J.; Bocchini, C.A.; Scott, A.F.; Hamosh, A. McKusick’s Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.* **2009**, *37*, D793–D796. [[CrossRef](#)] [[PubMed](#)]

30. Aken, B.L.; Ayling, S.; Barrell, D.; Clarke, L.; Curwen, V.; Fairley, S.; Fernandez Banet, J.; Billis, K.; García Girón, C.; Hourlier, T.; et al. The Ensembl gene annotation system. *J. Biol. Database Curation* **2016**, *2016*, baw093. [[CrossRef](#)] [[PubMed](#)]
31. Goel, R.; Harsha, H.C.; Pandey, A.; Prasad, T.S.K. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. *Mol. Biosyst.* **2012**, *8*, 453–463. [[CrossRef](#)] [[PubMed](#)]
32. Wilming, L.G.; Gilbert, J.G.R.; Howe, K.; Trevanion, S.; Hubbard, T.; Harrow, J.L. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* **2008**, *36*, D753–D760. [[CrossRef](#)] [[PubMed](#)]
33. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
34. Peijie, L.; Michael, T.; Joshua, W.K.H. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-Seq data. *Genome Biol.* **2017**, *18*, 59.
35. Kim, P.; Park, A.; Han, G.; Sun, H.; Jia, P.; Zhao, Z. TissGDB: Tissue specific Gene DataBase in cancer. *Nucleic Acids Res.* **2017**. [[CrossRef](#)] [[PubMed](#)]
36. Wang, C.-H.; Li, M.; Liu, L.-L.; Zhou, R.-Y.; Fu, J.; Zhang, C.Z.; Yun, J.-P. LRG1 expression indicates unfavorable clinical outcome in hepatocellular carcinoma. *Oncotarget* **2015**, *6*, 42118–42129. [[CrossRef](#)] [[PubMed](#)]
37. Jaakkola, M.K.; Seyednasrollah, F.; Mehmood, A.; Elo, L.L. Comparison of methods to detect differentially expressed genes between single-cell populations. *Brief. Bioinform.* **2017**, *18*, 735–743. [[CrossRef](#)] [[PubMed](#)]
38. Korthauer, K.D.; Chu, L.-F.; Newton, M.A.; Li, Y.; Thomson, J.; Stewart, R.; Kendziorski, C. A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol.* **2016**, *17*, 222. [[CrossRef](#)] [[PubMed](#)]
39. Delmans, M.; Hemberg, M. Discrete distributional differential expression (D3E)—A tool for gene expression analysis of single-cell RNA-seq data. *BMC Bioinform.* **2016**, *17*, 110. [[CrossRef](#)] [[PubMed](#)]
40. Finak, G.; McDavid, A.; Yajima, M.; Deng, J.; Gersuk, V.; Shalek, A.K.; Slichter, C.K.; Miller, H.W.; McElrath, M.J.; Prlic, M.; et al. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **2015**, *16*, 278. [[CrossRef](#)] [[PubMed](#)]
41. Kharchenko, P.V.; Silberstein, L.; Scadden, D.T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **2014**, *11*, 740–742. [[CrossRef](#)] [[PubMed](#)]
42. Seyednasrollah, F.; Rantanen, K.; Jaakkola, P.; Elo, L.L. ROTS: Reproducible RNA-seq biomarker detector-prognostic markers for clear cell renal cell cancer. *Nucleic Acids Res.* **2016**, *44*, e1. [[CrossRef](#)] [[PubMed](#)]

