Single Cell RNA-Seq Analysis of Pancreatic Cells

Group Van Gogh

Data Curator: Elysha Sameth

Programmer: Monil Gandhi

Analyst: Andrew Gjelsteen

Biologist: Lindsay Wang

INTRODUCTION

Gene expression studies comparing tissues in bulk offer the possibility of studying gene-regulatory differences across a large number of normal and pathological samples. However, this often masks the biological signal because of variability in cell type proportions. While there have been significant advances in characterizing the transcriptomes of individual cells using in situ and RNA sequencing (RNA-seq) approaches, there remains a difficulty in obtaining tissues from donors and developing a system that captures a sufficient number of cells. The single cell RNA-seq (scRNA-seq) inDrop method combats this issue, providing a systematic approach for capturing thousands of cells without pre-sorting using high-throughput droplet microfluidics that barcode the RNA from individual cells. By using single-cell transcriptomics, substructures within cell populations that can classify cell type by function and marker gene expression can be performed.

Although the inDrop approach has been shown to address the problem of needing pre-sorting when using high-throughput sequencing, the fact that the barcoding of the RNA from individual cells presents some difficulty when processing data. Namely, this requires the processing of the sequencing libraries into UMI count matrices, which allows for variation in approach.

In order to investigate gene expression differences between diabetics and healthy individuals, Baron et al[1] used inDrop data from an independent set of pancreatic islets. Cells were divided into clusters that matched previously characterized cell types. With these clustered cell data, Baron et al were able to study the human and mouse pancreas and to describe its constituent cell types at a higher resolution than that provided by immuno-histochemical

characterization and transcriptional analyses of bulk samples or cell type-enriched samples. In doing so, novel expressions of transcription factors, signaling receptors, and medically relevant genes were detected, subpopulations and heterogeneity within pancreatic cell types were identified, and bulk gene expression samples were deconvolved using single-cell data.

Our study aims to reproduce the results of Baron et al by examining different cell types of sequencing data from the pancreas of a 51 year old female donor[1]. Barcode extraction was done for the most relevant barcodes and UMIs from the parsed data. Gene expression levels for each of the cells was calculated and the resulting data was clustered using K-nearest neighbor (KNN) algorithm. Marker genes and cell types were determined and used to reason associated function.

DATA

Within Baron et al, human pancreatic islets from four cadaveric donors and islets from five ICR and C57BL/6 mice strains were obtained. Using the inDrop microfluidics system, ~10,000 human and ~2,000 mouse pancreatic cells were isolated, sequenced, and processed for transcriptomic analysis (Figure 1). Encapsulation of the cells and reverse transcription reaction, as well as library preparation, was carried out according to the protocol outlined by Klein et al[2]. Paired end sequencing was performed on Illumina Hiseq 2500, with reads lacking known cell barcodes, adaptor sequence, or beginning of the poly-T tail in read 1 removed. Reads were then trimmed using *Trimmomatic* 0.32[3] and split into individual cells based on their match against both cell barcode sequences, with errors of up to two nucleotides mismatch corrected. Finally, read 2 was mapped to the reference transcriptome using *bowtie2* 1.1.1[4] and expressions were quantified for clustering.

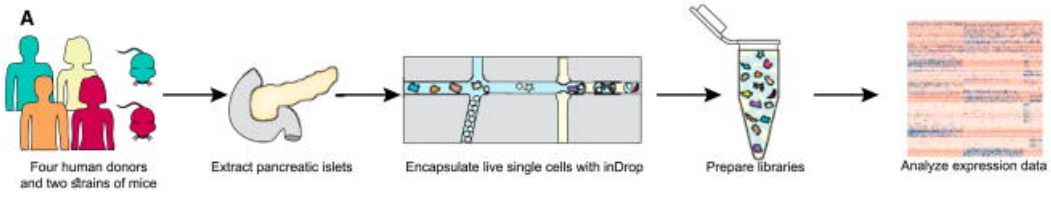


Figure 1. Single-cell RNA-seq was carried out on human and mouse pancreatic islets using the inDrop microfluidics system to generate data that allow for quantification of transcript abundance across cells and genes (from Figure 1A of Baron et al).

For our study, only SRR files from the 51 year old female human donor were processed and analyzed. To identify the SRA accession number, the metadata was downloaded from the

GEO accession number GSE84133, which contains the thirteen samples (sequencing libraries) and four human individuals. From the sample information, it was determined that the short read archive (SRA) accession number for the donor was SRP07832 and included the runs SRR3879604, SRR3879605, and SRR3879606 that were used for further analysis. Raw read 1 barcodes for these runs could not be matched to the InDrops barcode scheme due to noise in the protocol, therefore barcodes were pre-processed and padded to be 19 bases with 6 UMI bases when a valid barcode pattern was identified in the read. These pre-processed read 1 data, as well as read 2, for the donor was then further processed and analyzed in downstream analyses.

Run	AvgSpotLen	Bases	BioSample	Sample Name	BMI
SRR3879604	99	56122043408	SAMN05379703	GSM2230758	21.1
SRR3879605	99	38870435402	SAMN05379703	GSM2230758	21.1
SRR3879606	99	36467999606	SAMN05379703	GSM2230758	21.1

Table 1. Run information for the 51 year old female donor from GEO accession number GSE84133 that will be used in our study. Read 1 data from the runs SRR3879604, SRR3879605, and SRR3879606 will be processed

METHODS

Processing Single Cell Sequencing Reads and Choosing Barcodes

To determine the number of cells actually sequenced and whitelist informative barcodes, the number of reads per distinct barcode was calculated to eliminate reads with infrequent barcodes from consideration. Barcodes from each run were extracted using *awk* and *sed*, and the counts of each distinct barcode determined using *sort* and *uniq -c*. To identify the first *n* barcodes to filter, four methods were performed and compared (Table 2). Based on these results, merging runs, dropping barcodes that are not of length 19, and further filtering by the inflection point of the cumulative read counts[5] (Method 2) was determined to be the most appropriate and was used for further processing.

Method	Filter Description and Whitelist	Software
1	Inflection point of the cumulative read counts per cellular barcode of each individual run. Whitelist was determined based on the inflection point and then merged with one unique barcode per line.	R[6]
2	Inflection point of the cumulative read counts per cellular barcode of the runs merged	R[6]

	with barcodes without a length of 25 dropped. Whitelist was determined based on the inflection point.	
3	Inflection point of the cumulative read counts per cellular barcode of each individual run. Outputted whitelists were merged with one unique barcode per line.	UMI-tools[7]
4	Inflection point of the cumulative read counts per cellular barcode of the runs merged.	UMI-tools[7]

Table 2. Summary of the method and software used to determine an appropriate filter for cell-containing barcode, as well as the processing of the whitelisted barcodes.

To quantify the reads, a cell-by-gene (UMI) count matrix was generated using *salmon alevin*[8]. A *whitelist.txt* of one distinct barcode per line that passed the filters and the *fastq* files were inputted into *salmon alevin* with the parameters *--end 5 --barcodeLength 19 --umiLength 6* based on the custom barcode and UMI lengths. To enable *salmon* to collapse from the transcript to the gene level, the *--tgMap* option was given, which takes in a transcript to gene map file of each transcript present in the reference to the corresponding gene. This file was created by downloading the current human reference (GRCh38.p13) transcript sequences from Gencode[9] and mapping the transcript ID (ENSTXXX…) to gene (ENSGXXX…). Furthermore, the transcript sequences were used to build the index of the reference transcriptome using *salmon index* and was inputted using the *-i* argument. After running the command as a job on the Shared Computing Cluster[10] (SCC) with the library type ISR (inward, stranded, reverse strand), a UMI matrix of the number of reads in a cell originating from the corresponding gene, as well as a read mapping summary statistics, were outputted and analyzed in downstream analysis.

Processing the UMI Counts Matrix

The UMI count matrix was imported with the help of *tximport*[17] and consisted of 27648 features across 13471 samples. The next step was to convert the UMI mtix into a Seurat object which was accomplished by the *Seurat*[12] package. Following this step, Ensemble gene identifiers were converted to the Gene symbols using *EnsDb.Hsapiens.v79*[13]. From the matrix, the genes detected in greater than 3 cells and cells with at least 200 detected genes were filtered, resulting in 26060 features across 12277 samples. Figure 2 shows the number of genes per cell, the number of molecular counts associated with genes per cell, and the percent of reads that map to the mitochondrial genome. According to the distribution in Figure 2, the cells with less that 200 unique gene counts and greater than 4000 are not included. The cutoff thresholds are based on the criteria for not including the group of cells with very few genes and the group of cells

with high number of genes, which is based on the plot distribution in Figure 2. Cells with unique gene counts greater than 4000 or less than 200 and also for the mitochondrial genes of higher than 5% were excluded from the dataset. The filtering step therefore takes into consideration nFeature and percent.mt from Figure 2, in the overall filtering of the dataset for quality control.

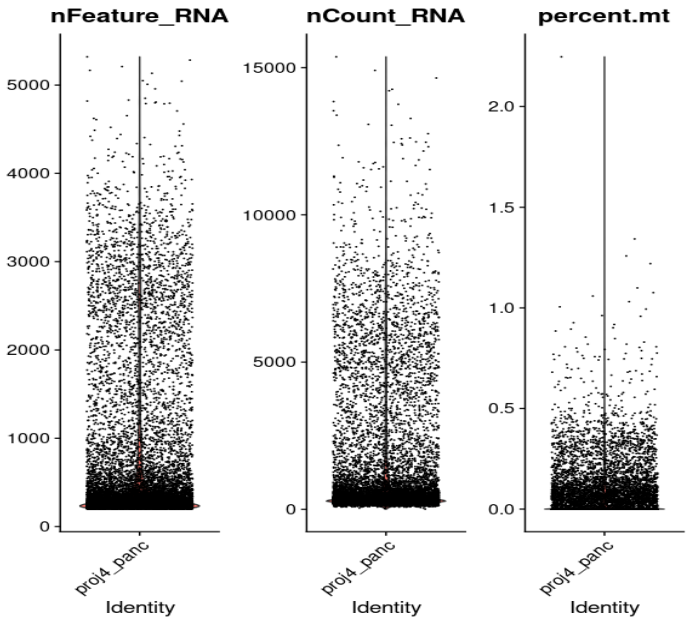


Figure 2. Violin plot of 1) number of genes per cell, 2) number of molecular counts associated with genes per cell, and 3) percent of reads that map to the mitochondrial genome. The plot gives information regarding the metadata and the filtering threshold to consider for the number of genes and the percent mitochondrial content

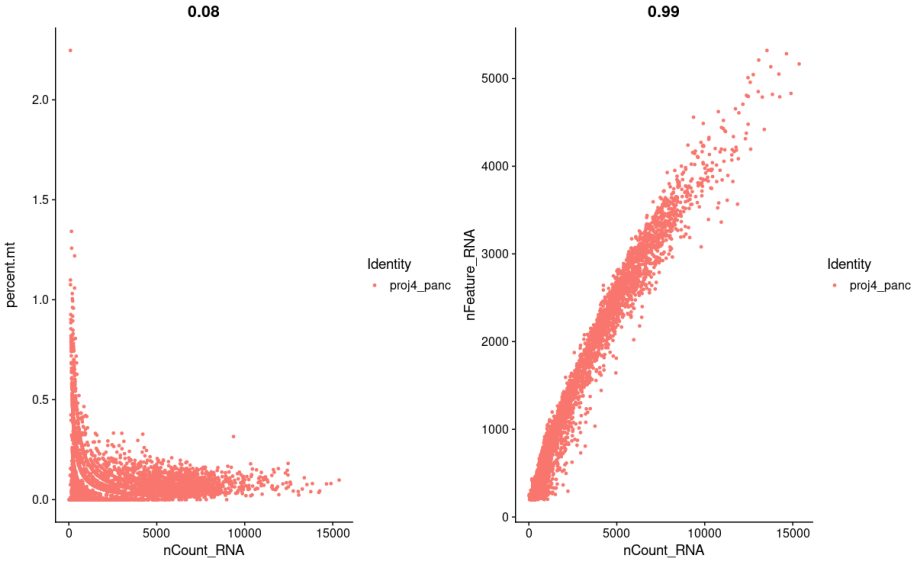


Figure 3. Scatter Plot for correlation, 1) Correlation between percent.mt and nCount_RNA, 2) Correlation between nFeature_RNA and nCount_RNA

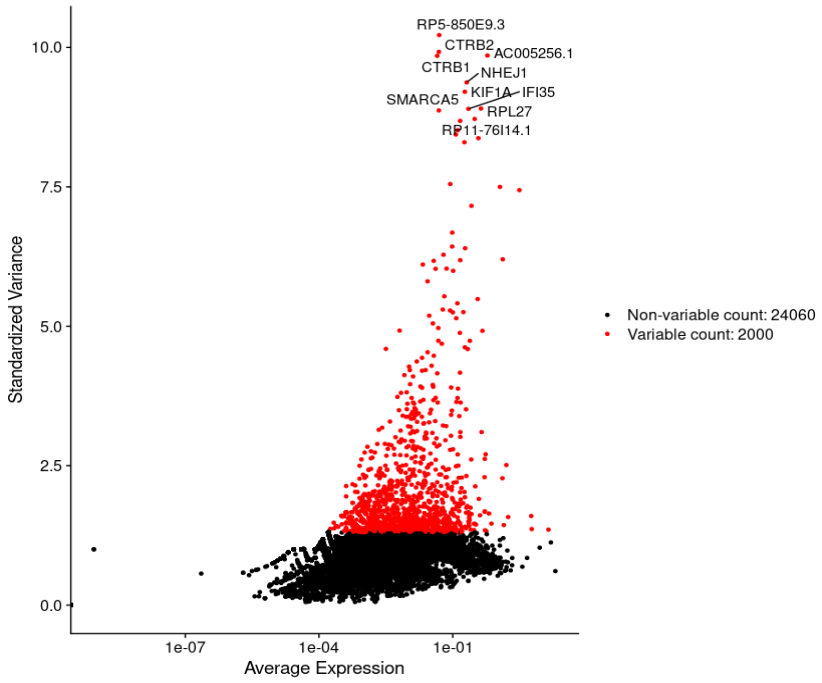


Figure 4. Scatter plot showing highly variable genes in red and naming the top 10 genes.

After filtering of low quality cells, the data was normalized using a method called "LogNormalize", to normalize the gene expression data for each cell across all samples and then multiplied by a scaling factor of 1000. Normalization helped in selecting the genes with high variation. The top 10% highly variable genes were therefore selected for further analysis. Figure 4, identifies 2000 as the most variable genes.

Cluster Marker Genes

With the normalized gene set, the data was clustered into cell types. The data clustered into 14 different cell types, using Seurat's FindAllMarkers() function, with corresponding log2 Fold-Change and p-value statistics included. In order to identify the type of pancreatic cell that each cluster represented, the top 5 significant genes, as measured by $\log_2$FC and p-value, were searched for. The top marker genes were visualized across all clusters in a Violin Plot using R's VlnPlot() and a feature plot using R's FeaturePlot() commands. A non-linear dimensional reduction plot was rendered using Seurat's RunUMAP() function. These plots are displayed in the Appendix (Appendix 4 and Appendix 5).

Marker Gene Analysis

Using adjusted p value < 0.01, significant cluster marker genes were filtered to prepare for further analysis. Then gene set enrichment analysis was performed on the marker genes for each cluster, using Enrichr[14]. The cell type for each cluster is determined from Descartes Cell Types and Tissue 2021 database in Enrichr, and the top 3 enrichment pathways are determined from KEGG 2019 Human database in Enrichr.

RESULTS

Read Processing and Barcode Selection

When merging all reads from the donor, there were 2914895 distinct barcodes that had to be filtered to remove potential bias from low-quality cells, and determine barcodes corresponding to a cell-containing droplet. Invalid barcodes according to the pre-processing padding were removed, resulting in 2833359 barcodes. Finally, a cumulative distribution graph (Figure 5) showed that the inflection point is approximately 100, thus the first 100 barcodes were removed and the remaining barcodes used as input for *salmon alevin*[8].

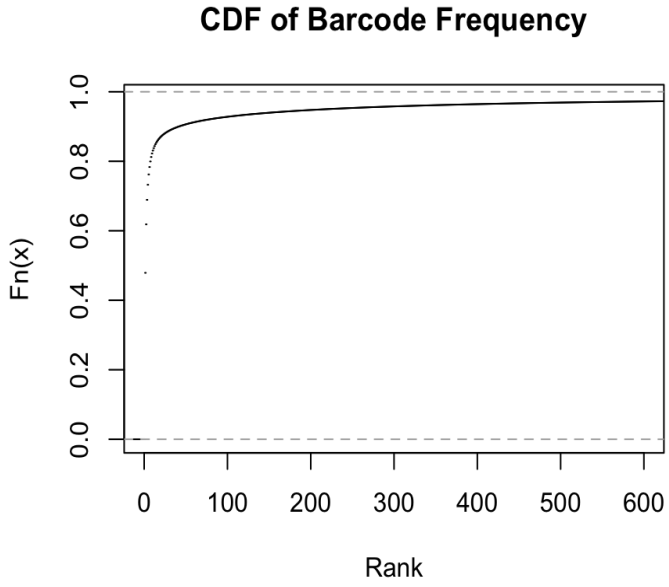


Figure 5. Cumulative distribution plot of the barcode frequencies of the SRR3879604, SRR3879605, and SRR3879606 *fastq* run files from donor SRP07832. Pre-processing of barcodes included removing those that do not have 19 bases. The x-axis is the n^{th} barcode ordered by frequency, with the top 600 visualized for clarity. The y-axis is the probability that the variable takes a value less than or equal to x.

The results from *salmon* (Table 2) show that 12.6852% of the cellular barcodes matched a low number of reads (noisy) and were dropped. These low-quality libraries could have originated from experimental factors such as damaged or stressed cells, or failures in library

preparation due to sequencing or PCR errors. By dropping these barcodes and the ~81,500 barcodes that were deemed invalid, we retained 2637113 barcodes that represent independent cells. Furthermore, the proportion of uniquely mapped reads out of the 1324837961 input reads (mapping rate) was found to be 42.7501%. This low mapping rate may be due to the *k*-mer size when building the index, with smaller values of *k* potentially improving sensitivity. However, since the authors did not state the mapping rate, these results may be expected. Given these results, Method 2 was shown to be successful as the Method 1 inflection point (knee) was approximately 100 for each read, which would contain barcodes of invalid lengths (Appendix 1). Meanwhile, Method 3 dropped 56.3396% of cellular barcodes due to noise (Appendix 2), and Method 4 99.9949% reads (Appendix 3). As a result, despite the low mapping rate, this method was successful in filtering noisy cellular barcodes, demultiplexing the samples, mapping to transcriptome, and quantifying reads.

Statistic	**Value**	**Statistic**	**Value**
Barcode Processing		**Quantifying**	
Transcripts Found	233613	Total Reads in Experiment	1324837961
Noisy cellular barcodes (dropped)	12.6852%	Reads with N (dropped)	60564
Total Unique Barcodes	4251176	UMIs after Deduplicating	28217780
Used Barcodes except Whitelist	24029	Mean UMIs per cell	10
Final Number CBs	2637113	Mean Genes per cell	7
Mapping Rate: 42.7501%			

Table 2. Results from *salmon alevin* using a whitelist of barcodes above the inflection point of the cumulative read counts per cellular barcode of the runs merged with barcodes without a length of 25 dropped.

Marker Genes

Linear transformation scaling was applied as a preprocessing step prior to applying PCA dimensionality technique. Scaling shifts the expression of each gene so that the mean and variance across cells are 0 and 1 respectively. After scaling, PCA was applied to the scaled data and an elbow plot was constructed to visualize the PC based on the percentage of variance explained by each PC. Figure 6 shows a decrease in the variance around PC9 - PC10, which indicates that the first 10 PC's capture the majority of the variance reported in the data.

Clustering algorithm KNN was applied to identify cell clusters. Seurat's FindNeighbours function helped with the clustering based on KNN technique. FindClusters function helped to cluster the cells, based on the modularity optimization technique called Louvain algorithm. Resolution for FindClusters function was chosen to be 1.5. The clustering based on the first 10PC dimensions, resulted in 10 clusters shown in Figure 7.

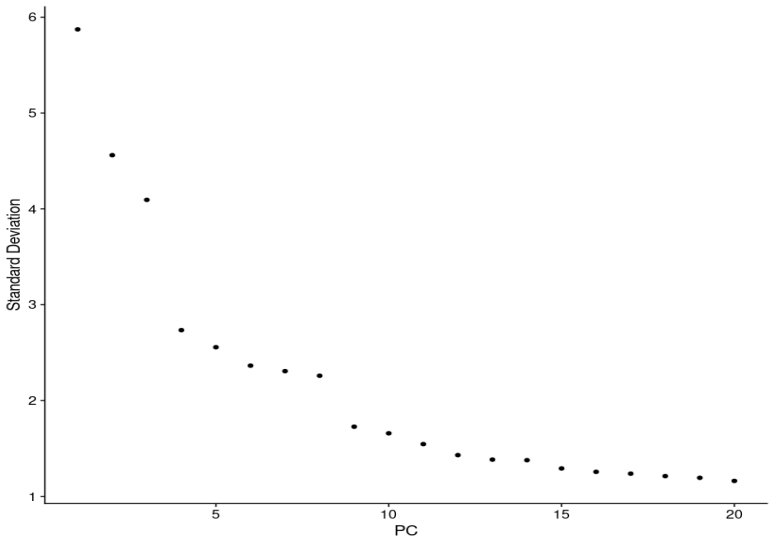


Figure 6. Elbow plot for PC vs percentage of variance. Elbow Effect neat PC10

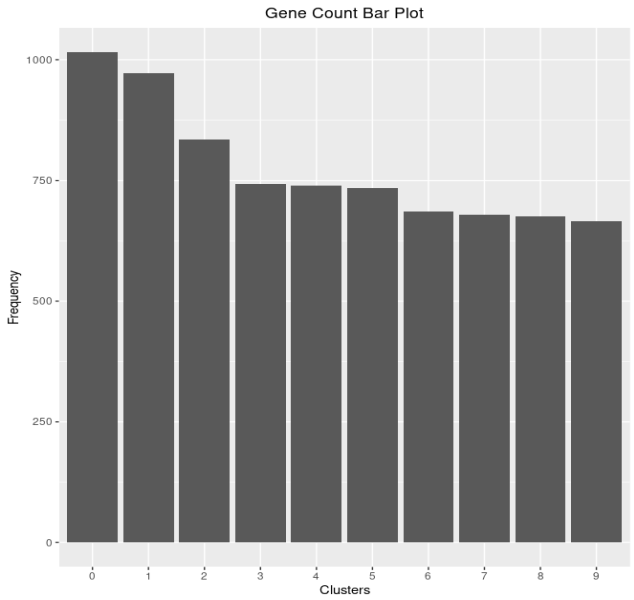


Figure 7. Histogram showing the 10 identified clusters with the frequency for cells in each

Marker Gene Enrichment Analysis

Cluster	# of Marker Genes	Enriched Terms
0 Islet endocrine cells	30	- Neuroactive ligand-receptor interaction - Lysosome - Focal adhesion
1 Islet endocrine cells	35	- Type I diabetes mellitus - Maturity onset diabetes of the young - Sulfur metabolism
2 Pancreas Beta Cells*	21	- ECM-receptor interaction - Protein digestion and absorption - Amoebiasis
3 Acinar cells	66	- Ribosome - Protein digestion and absorption - Pancreatic secretion
4 Islet endocrine cells	29	- Type I diabetes mellitus - Antigen processing and presentation - Complement and coagulation cascades
5 Ductal cells	66	- Shigellosis - Viral myocarditis - Salmonella infection
6 Islet endocrine cells	1070	- Ribosome - Protein export - Oxidative phosphorylation
7 Erythroblasts	146	- Fatty acid biosynthesis - Bladder cancer - Thiamine metabolism
8 Islet endocrine cells	1474	- Ribosome - Oxidative phosphorylation - Parkinson disease
9 Stromal cells	252	- Ribosome - ECM-receptor interaction - Focal adhesion
10 Ductal cells	1932	- Ribosome - Huntington disease - Bacterial invasion of epithelial cells
11 Acinar cells	1117	- Ribosome - Protein processing in endoplasmic reticulum - Non-alcoholic fatty liver disease (NAFLD)
12	94	- Lysosome

Myeloid cells		- Shigellosis - Pathogenic Escherichia coli infection

Table 3. The enrichment pathways for each cluster marker gene. *The cell type for cluster 2 is obtained from MSigDB 2020 Hallmark database, because no pancreas related term was found in Descartes Cell Types and Tissue 2021 database.

DISCUSSION

From our analysis, we identified 6 Islet endocrine cells, 2 types of blood cells (myeloid cells, Erythroblasts), 2 types of acinar cells, 2 types of ductal cells, and 1 type of stromal cell. Based on the database we are using, we were unable to identify the specific cell type. Overall, we have identified fewer cell types compared to the original article. However, the cell species generally correspond to the results reported in the original article. One of the reasons for the misclassification and incomplete result could be due to our custom identification of barcodes. Since we do not have access to the original barcode, valid barcodes were customized and padded manually.

Furthermore, the pathways enriched for each gene cluster generally correspond with the findings reported in the original article or can be supported by other research articles. For instance, for cluster 3 islet endocrine cells, the enrichment terms are “Type I diabetes mellitus” and “Antigen processing and presentation” which correspond to the pathology and mechanism of Type I diabetes. Also, for cluster 8 islet endocrine cells, one of the enrichment pathways associated pancreas secretion with Parkinson's disease which is supported by Knudsen et al.’s research[15]. The enrichment terms for cluster 9 suggested that stromal cells participate in ECM (extracellular matrix) receptor interactions, which agreed with the description of transcriptome of stellate cell in the original article. As discussed in Weniger’s article[16], ECM promotes cell survival, proliferation angiogenesis, and metastasis, which could explain the presence of erythroblasts in the sample.

CONCLUSION

In this paper, we tried to replicate the results obtained from Baron et al. using samples from a 51-year-old female. We mapped the readings based on customized barcodes, and eliminated the low quality readings. Then, using the Stuart package in R we were able to cluster

the cells based on gene expression and identify the marker gene for each cluster. In our efforts to reconstruct the analysis in the paper, we have identified 13 cell types. Despite the minor inconsistency with the original article, the cell types we identified generally consist with the original article. Also, the enrichment analysis also yielded the similar biological functions as in the original article. We believe that better result reproduction can be gained when using more samples to perform the analysis. Overall, the study provided a novel method to characterize transcriptome profiling using single cell RNA-Seq and use the result to study disease pathology.

REFERENCES

[1] Baron, Maayan, Adrian Veres, Samuel L. Wolock, Aubrey L. Faust, Renaud Gaujoux, Amedeo Vetere, Jennifer Hyoje Ryu, et al. 2016. “A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-Cell Population Structure.” Cell Systems 3 (4): 346–60.e4. PMID: 27667365

[2] Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161:1187–1201.

[3] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–2120.

[4] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–359.

[5] Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May 21;161(5):1202-1214. doi: 10.1016/j.cell.2015.05.002. PMID: 26000488; PMCID: PMC4481139.

[6] “The R Project for Statistical Computing.” R, www.R-project.org/.

[7] Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017 Mar;27(3):491-499. doi: 10.1101/gr.209601.116. Epub 2017 Jan 18. PMID: 28100584; PMCID: PMC5340976.

[8] Srivastava, A., Malik, L., Smith, T. et al. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol 20, 65 (2019). https://doi.org/10.1186/s13059-019-1670-y

[9] Adam Frankish, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M Mudge, Cristina Sisu, James Wright, Joel Armstrong, If Barnes, Andrew

Berry, Alexandra Bignell, Silvia Carbonell Sala, Jacqueline Chrast, Fiona Cunningham, Tomás Di Domenico, Sarah Donaldson, Ian T Fiddes, Carlos García Girón, Jose Manuel Gonzalez, Tiago Grego, Matthew Hardy, Thibaut Hourlier, Toby Hunt, Osagie G Izuogu, Julien Lagarde, Fergal J Martin, Laura Martínez, Shamika Mohanan, Paul Muir, Fabio C P Navarro, Anne Parker, Baikang Pei, Fernando Pozo, Magali Ruffier, Bianca M Schmitt, Eloise Stapleton, Marie-Marthe Suner, Irina Sycheva, Barbara Uszczynska-Ratajczak, Jinuri Xu, Andrew Yates, Daniel Zerbino, Yan Zhang, Bronwen Aken, Jyoti S Choudhary, Mark Gerstein, Roderic Guigó, Tim J P Hubbard, Manolis Kellis, Benedict Paten, Alexandre Reymond, Michael L Tress, Paul Flicek, GENCODE reference annotation for the human and mouse genomes, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D766–D773, https://doi.org/10.1093/nar/gky955

[10] “SCC Quick Start Guide.” BU TechWeb RSS, www.bu.edu/tech/support/research/system-usage/scc-quickstart/.

[11] Soneson C, Love MI, Robinson MD (2015). “Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.” F1000Research, 4. doi: 10.12688/f1000research.7563.1.

[12] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck III, Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zagar, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P. Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar B. Fleming, Bertrand Yeung, Angela J. Rogers, Juliana M. McElrath, Catherine A. Blish, Raphael Gottardo, Peter Smibert, Rahul Satija bioRxiv 2020.10.12.335331; doi: https://doi.org/10.1101/2020.10.12.335331

[13] Rainer J (2017). EnsDb.Hsapiens.v79: Ensembl based annotation package. R package version 2.99.0.

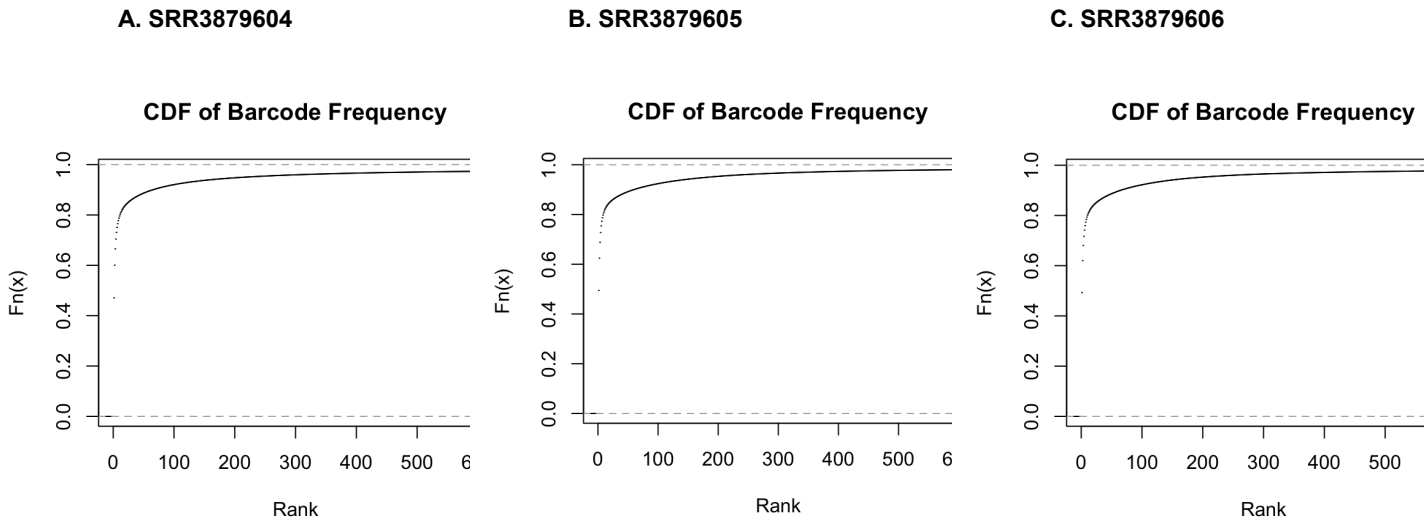
[14] Chen, E.Y., Tan, C.M., Kou, Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013). https://doi.org/10.1186/1471-2105-14-128

[15] Knudsen K, Hartmann B, Fedorova TD, et al. Pancreatic Polypeptide in Parkinson's Disease: A Potential Marker of Parasympathetic Denervation. J Parkinsons Dis. 2017;7(4):645-652. doi:10.3233/JPD-171189

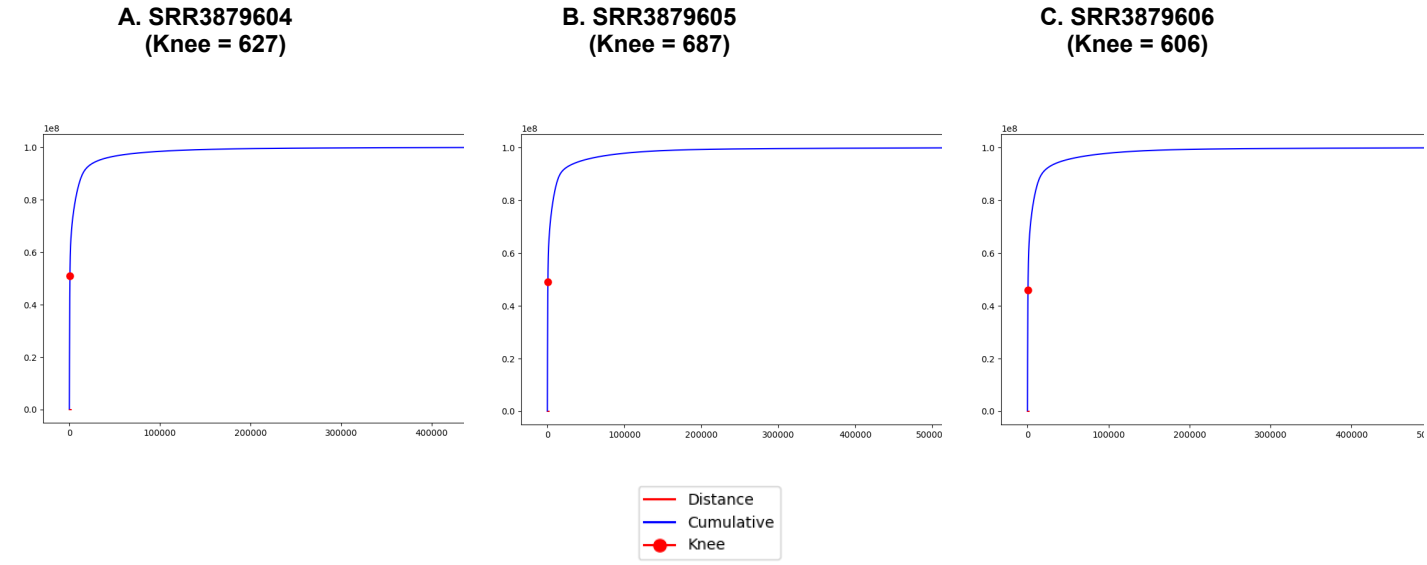
[16] Weniger M, Honselmann KC, Liss AS. The Extracellular Matrix and Pancreatic Cancer: A Complex Relationship. Cancers (Basel). 2018;10(9):316. Published 2018 Sep 6. doi:10.3390/cancers10090316

[17] *tximport*. Bioconductor. (n.d.). https://bioconductor.org/packages/release/bioc/html/tximport.html.

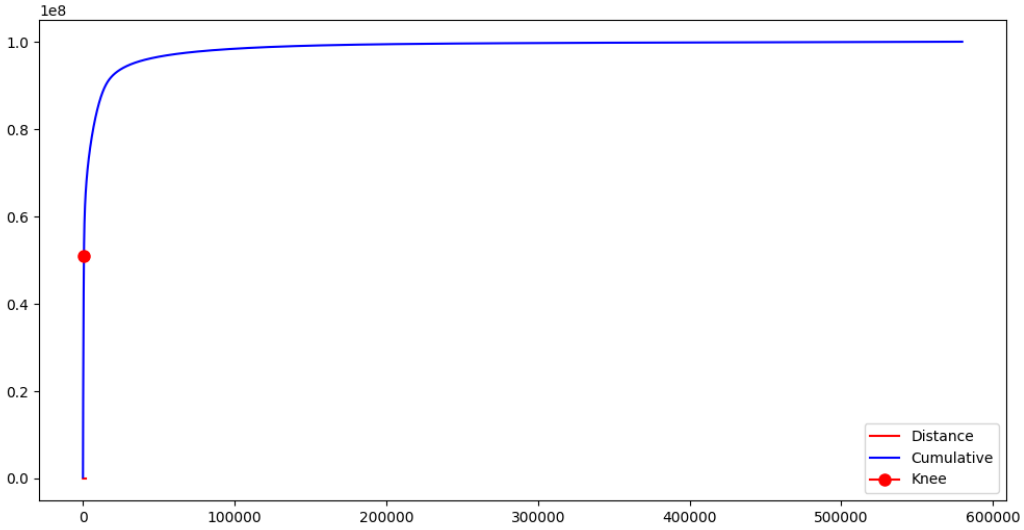
APPENDIX



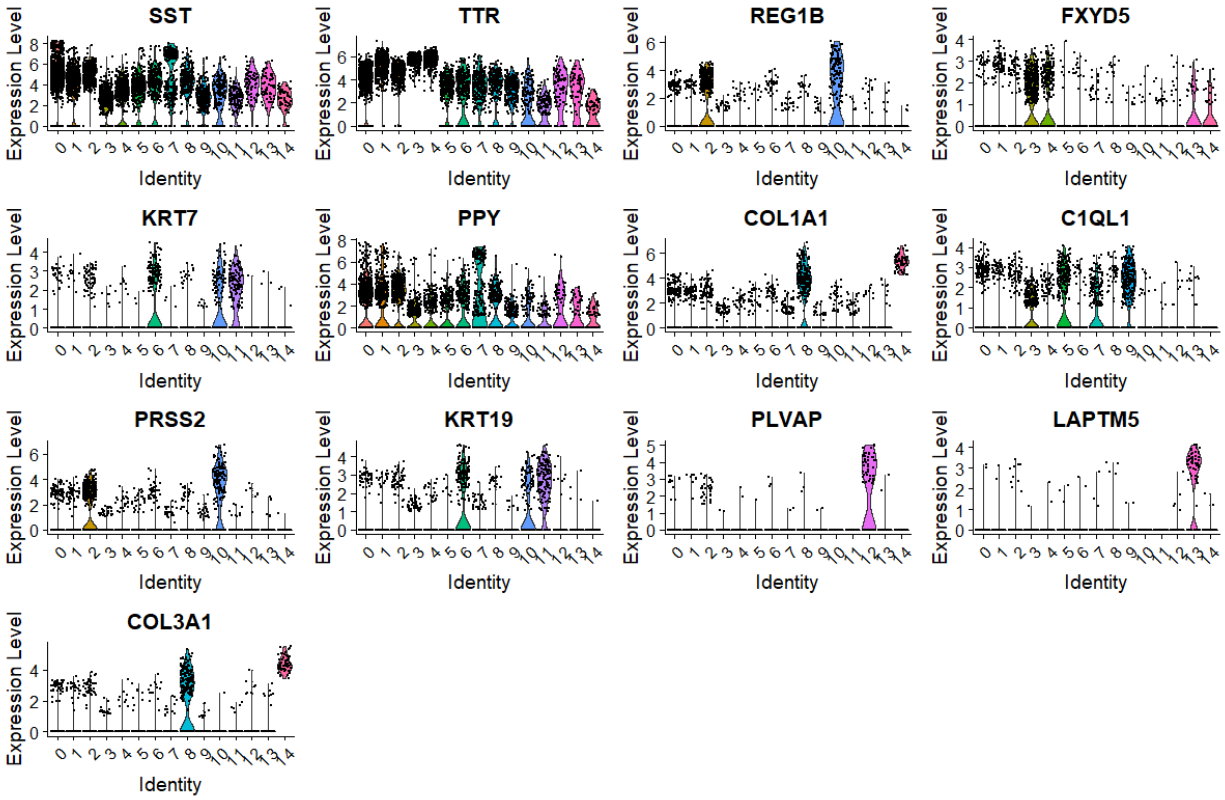
Appendix 1. R density plot of the cumulative read counts per cellular barcode for SRR3879604, SRR3879605, and SRR3879606 *fastq* run files from donor SRP07832 (Method 1). The inflection point (knee) is approximately 100 for each read, which is too low and would contain invalid lengthed barcodes.



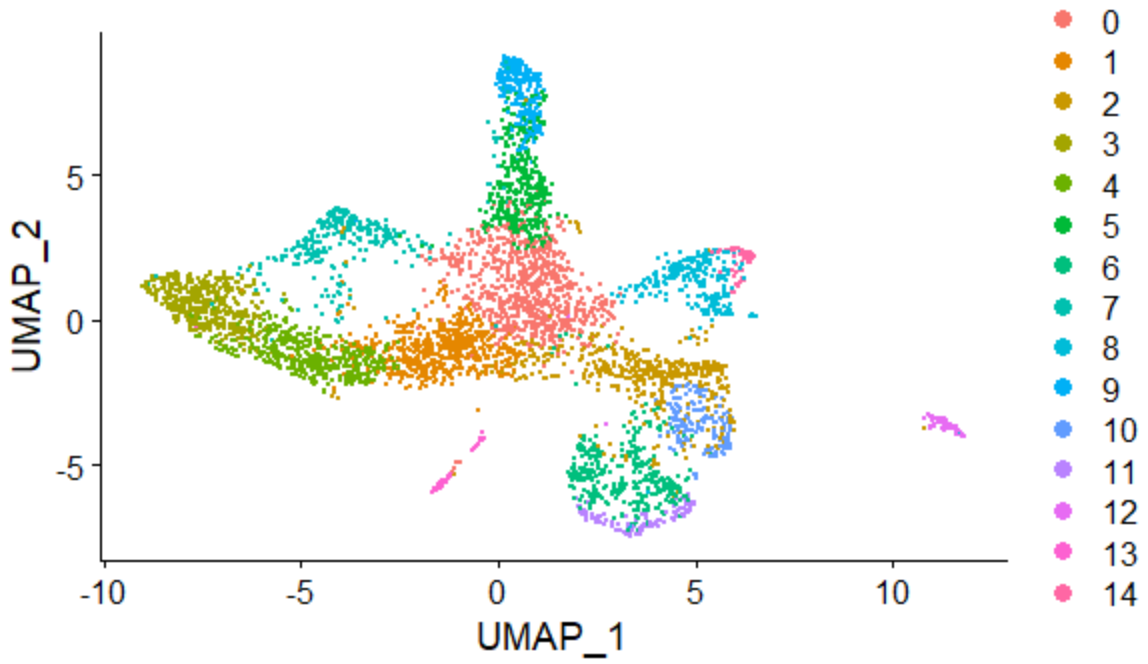
Appendix 2. UMI-tools density plot of the cumulative read counts per cellular barcode for SRR3879604, SRR3879605, and SRR3879606 *fastq* run files from donor SRP07832 (Method 3). The red dot (knee) represents the number of cell barcodes which should be accepted. After performing *salmon alevin*, 56.3396% reads were dropped.



Appendix 3. UMI-tools density plot of the cumulative read counts per cellular barcode for SRR3879604, SRR3879605, and SRR3879606 *fastq* run files from donor SRP07832 merged (Method 4). The red dot (knee) represents the number of cell barcodes which should be accepted. After performing *salmon alevin*, 99.9949% reads were dropped.



Appendix 4. Violin plot showing one of the differentially-expressed genes across each cluster. Note that the cluster count is different from the prior analysis because due to an error in Seurat's ReadRDS function being unable to read a .rda file, another group's data was borrowed for just this portion of analysis. These results show an enrichment of a couple of genes in a single primary cluster, such as PLVAP and LAPTM5 in clusters 12 and 13, respectively.



Appendix 5: Non-linear dimensional reduction plot of pancreatic gene clusters. Note that there are more clusters in this image due to last-minute borrowing of data from another group as Seurat's ReadRDS returned an error for our .rda file. The origin of the data remains the same, so the clustered data show similar findings to what we found, though they were clustered to a higher degree. Figure generated using R's Seurat DimPlot() function.