Transcriptional profiling of 1.3 million brain cells with the Chromium Single Cell Gene Expression Solution

Introduction

High-throughput single cell RNA sequencing (scRNA-seq) technologies measure genome-wide transcriptome data at single cell resolution and have revolutionized our ability to study complex biological systems. However, most scRNA-seq technologies can only process hundreds to thousands of cells in a single experiment—far fewer than the number of cells in most tissues and organs. Our understanding of complex biological systems will be fueled by technologies that can process hundreds-of-thousands to millions of cells at a time.

10x Genomics recently described a fully-integrated, droplet-based approach, the Chromium Single Cell Gene Expression Solution (1), which enables 3' mRNA digital profiling of hundreds to millions of cells. To demonstrate the scalability of the 10x platform and software, the Chromium Single Cell Gene Expression Solution was used in a study of the mammalian brain, one of the most complex biological systems in nature. Specifically, ~1.3 million brain cells from two embryonic mice were profiled. The single experiment successfully detected major neuronal and non-neuronal cells types and uncovered many substructures within each major subtype, including rare interneurons, without the need for cell enrichment procedures.

Highlights

The Chromium Single Cell Gene Expression Solution enables characterization of large and complex biological tissues at single cell resolution

Cell Ranger allows genome-wide transcriptome analysis of ~1.3 million brain cells with computational efficiency

At a modest sequencing depth (~18,500 reads/cell), major neuronal subtypes were detected in expected proportions, and rare interneurons were uncovered without the use of enrichment strategies

Loupe Cell Browser is an intuitive, easy-to-use desktop application that allows novices and seasoned bioinformaticians alike to visualize and analyze single cell RNA-seq data

Methods

Cell encapsulation, library preparation, and sequencing

Combined cortex, hippocampus, and ventricular zones, dissected from 2 embryonic day 18 mice, were purchased from BrainBits (www.brainbits.com, C57EHCV). Samples were dissected and shipped on the same day (overnight on dry ice), and stored at 40°C until being prepared for scRNA-seq.

Brain tissues were prepared following our demonstrated protocol for dissociation of mouse embryonic neural tissue (support.10xgenomics.com/single-cell/sample-prep). Details of the Chromium Single Cell Gene Expression workflow were described previously (10xGenomics.com/single-cell). Barcoded cDNA library preparation was performed with reagents from Single Cell 3' v2 kits (Product Code:120237). 133 libraries were generated with 17 chips, with each library aiming to recover ~10,000 cells.

Libraries were sequenced on 11 HiSeq 4000 flowcells using paired-end sequencing with 26 bp (R1), 8 bp (i7) and 98 bp (R2) run configuration, resulting in an average depth of ~18,500 reads per cell.

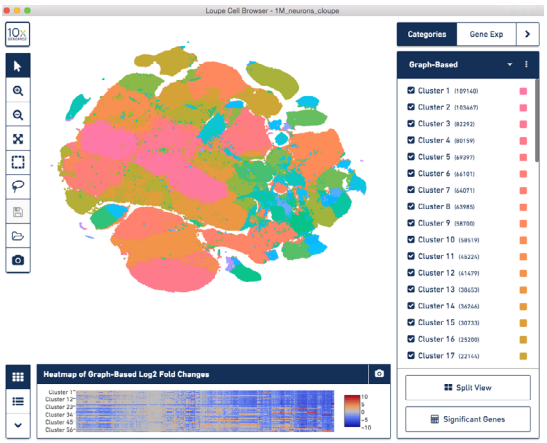

Figure 1. Screenshot of a Loupe Cell Browser window showing 1.3 million brain cells. Key functionalities of the Loupe Cell Browser include: dynamic, interactive visualization of tSNE scatter plots, generation of custom subgroups by manual selection, identification of differentially expressed genes between groups, formation of gene expression heatmaps, and export of differential expression tables and plots.

Easy-to-use computational analysis and data visualization tools

The Cell Ranger software suite was used to perform demultiplexing, barcode processing, transcript counting and clustering analysis. First, each sequencing library was processed individually. Then, the output from all 133 libraries was aggregated in Cell Ranger to produce an aggregated gene-barcode matrix. Finally, principal component and clustering analyses were performed on the aggregate data using Cell Ranger secondary analysis tools. Alignment, aggregation, and transcript counting were performed on the full dataset; clustering analyses were performed on a randomly selected subset of 20,000 cells.

Outputs from Cell Ranger (.cloupe files) were visualized with Loupe Cell Browser (Figure 1), our desktop application for Windows and MacOS that allows you to easily visualize and analyze Chromium Single Cell 3' data. The browser's main window displays data as a two dimensional tSNE projection, while intuitive tools allow users to highlight significant genes, identify cell types, and explore substructure within cell clusters.

Computational analysis of 1.3 million cells – an engineering feat!

Cell Ranger is our open-source software solution for analyzing single cell 3' RNA-seq data produced by the 10x Chromium Platform. It is easy to install and scales from workstation to cluster to cloud. Computing resources for the analysis of 1.3 million brain cells are summarized below.

Demultiplexing, alignment, and transcript counting:

- ~20 CPU hours
- ~12 GB memory
- FASTQs of 1.3 million cells = 3.6 TB

Aggregate/reanalyze data from 133 libraries:

- ~700 CPU hours
- ~175 GB memory
- matrix files of 1.3 million cells = 3.9 GB

Principal component analysis and clustering:

- ~350 CPU hours
- ~300 GB memory
- clustering output files = 773 MB

t-SNE Analysis

t-distributed stochastic neighbor embedding (t-SNE) is a machine learning algorithm that embeds high-dimensional data into a space of two or three dimensions, which can then be displayed in a scatter plot (2). The method is commonly used for visualizing scRNA-seq datasets.

Results for 20,000 cells analysis

Cellular viability was maintained throughout disassociation and library preparation protocols

Fewer than 0.5% of cells show >5% mitochondrial (mito) transcripts (UMI counts), suggesting that the majority of the cells are healthy (Figure 2a). At a sequencing depth of ~18.5k reads per cell, a median of ~1,870 genes and ~5,000 transcripts were detected per cell (Figure 2b and c).

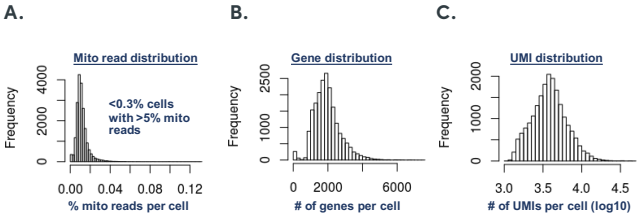

Figure 2. Quality analysis of 20,000 cells.

Comprehensive cell type identification

The proportion of neuronal and non-neuronal cells can be estimated based on the expression of marker genes detected in different clusters (Figure 3). The population comprises ~75% neurons, which further sub-cluster into ~60% excitatory neurons, ~15% inhibitory neurons and ~25% glial cells, including ~12% astrocytes and ~9% oligodendrocytes.

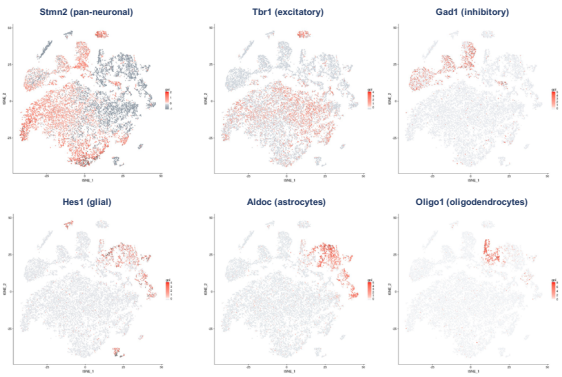

Figure 3. Gene expression profiles overlaid on tSNE projections for six selected brain cell marker genes.

Unbiased discovery recapitulates known neuronal cell populations

Seven major classes of cells were identified based on the gene markers that are enriched in each class: excitatory and inhibitory neurons, astrocytes, mural cells, oligodendrocytes, microglia, and ependymal cells (Figure 4a). The brain cells were also classified based on their best match to the average expression profile of reference transcriptomes defined in a recent study from the Allen Brain Institute (3) (4b) and observed a consistent cell-type identification. In both cases, clustering patterns suggest that there is extensive substructure within each main classification.

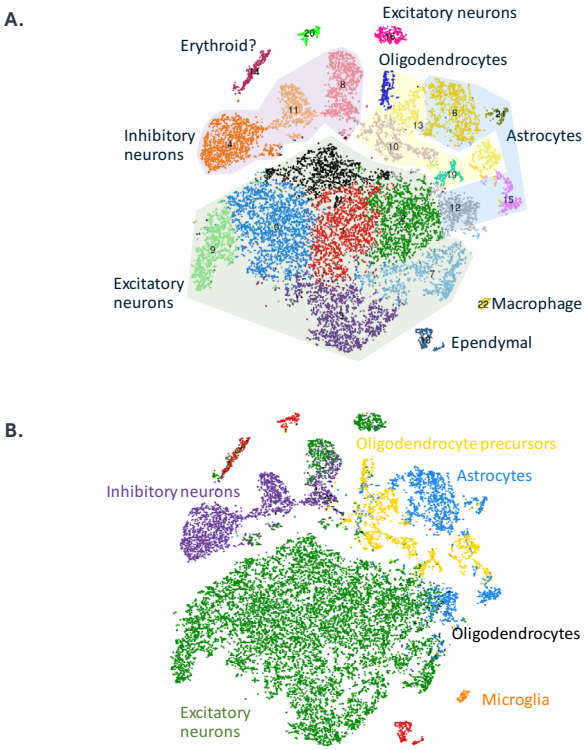

Figure 4. t-SNE projections of 20,000 brain cells where each cell is represented as a dot. (A) Shading highlights major clusters identified in the dataset, and color clustering within these shaded regions indicates additional complexity within each major class. (B) Cells were colored based on their best match to the average expression profile of reference transcriptomes (Tasic et al., 2016).

Software and Data Sharing

10x Genomics is committed to data sharing and open-source software in the hopes that the research community will apply their engineering, analytical and biological expertise to the vast dataset from this study. Here, 20,000 out of 1.3 million cells and only a few of the known marker genes are the focus of analysis, making this dataset ripe for further discovery. With the Loupe Cell Browser (https://support.10xgenomics.com/single-cell/software/downloads/latest, Figure 1), users can intuitively explore 3' gene expression data generated by the Cell Ranger pipeline. The raw FASTQ data is also available for developers who wish to use their own tools or develop new ones.

Software, data sets, and data sharing forums

- Cell Ranger Pipeline: https://support.10xgenomics.com/single-cell/software/pipelines/latest/what-is-cell-ranger
- Loupe Cell Browser: https://support.10xgenomics.com/single-cell/software/downloads/latest
- 1.3 million brain cell dataset: https://support.10xgenomics.com/single-cell/datasets/1M_neurons

Conclusion

10x Genomics has produced the largest brain scRNA-seq dataset to date. With only 1.5% of cells being explored in detail, the whole-genome transcriptome data for 1.3 million murine brain cells is open to the research community for method development and biological discovery.

References

1. G Zheng et al., Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat Commun*. 8, 14049 (2017).
2. L van der Maaten, G Hinton, Visualizing Data using t-SNE. *J Mach Learn Res*. 9, 2579–2605 (2008).
3. B Tasic et al., Adult Mouse Cortical Cell Taxonomy Revealed by Single Cell Transcriptomics. *Nat Neurosci*. 19, 335–46 (2016).

Resources from 10x Genomics

We are dedicated to helping you get the most out of your 10x Genomics system by offering multiple helpful resources:

10x University

Immerse yourself in 10x University, a comprehensive step-by-step learning and training environment containing video tutorials and trainings.
10xgenomics.com/10x-university

10x Genomics library

Easy access to our complete library of product literature, customer publications, application notes, protocols, and many other useful resources.
10xgenomics.com/resources/library

Support

Visit the support site for documentation, software, and datasets that will help you get the most out of your 10x Genomics products.
support.10xgenomics.com

10x Genomics compatible products

Access our list of key partner products that have been certified compatible to work with our various solutions.
10xgenomics.com/compatible-products

Solutions and products

Along with our suite of complete solutions, we offer an ever-growing catalogue of services to help you find the answers to your research questions.
10xgenomics.com/solutions

10x Blog

Keep up to date with the 10x Genomics Blog, where you'll find everything from tips and tricks to the latest 10x news.
10xgenomics.com/blog

Contact us

10x Genomics
6230 Stoneridge Mall Road
Pleasanton, CA 94588-3260
+1 925 401 7300
+1 800 709 1208

info@10xgenomics.com **10xgenomics.com**

For more locations in US, EU and Asia visit:
10xgenomics.com/company/#locations

© 2020 10x Genomics, Inc. FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.
LIT000015 Rev B Application Note Transcriptional profiling of 1.3 million brain cells with the Chromium Single Cell Gene Expression Solution