SOCIETY for NEUROSCIENCE

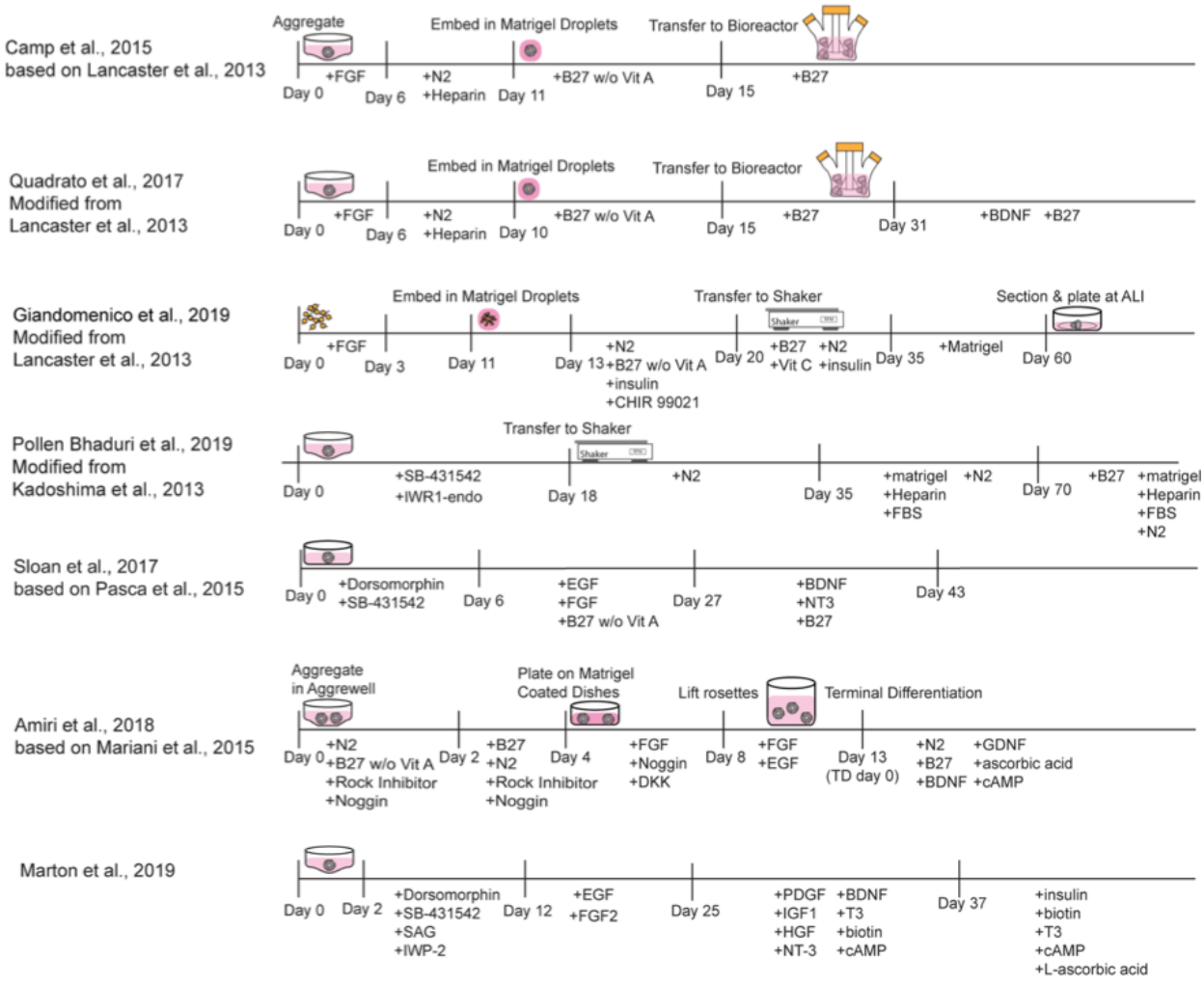
1121 14th Street NW
Suite 1010
Washington, DC 20005

Phone (202) 962-4000
Fax (202) 962-4941
Web SfN.org

Module 4 Resource List: Protocols and Analysis Guide

The resources in the protocols and analysis guide below were selected by Madeline Lancaster, Aparna Bhaduri, and Madeline Andrews, faculty from Module 4 of Stem Cells and Reprogramming Methods for Neuroscience: An SfN Training Series.

Overview of Selected Organoid Generation Protocols



SOCIETY for NEUROSCIENCE
1121 14th Street NW
Suite 1010
Washington, DC 20005
Phone (202) 962-4000
Fax (202) 962-4941
Web SfN.org

Links to Useful Step-By-Step Protocols and Protocol Papers

Generation of Cerebral Organoids from Human Pluripotent Stem Cells
The original Lancaster et. al. protocol.

Lancaster Lab Resources Page
This site is continually being updated with improved protocols and help with organoid analysis.

Feeder-Free, Xeno-Free Generation of Cortical Spheroids From Human Pluripotent Stem Cells
Generation and Assembly of Human Brain Region-Specific Three-Dimensional Cultures
Two protocols for cortical spheroids.

Detailed Analysis Guide Including Annotated Code Examples

Below are the methods behind the data analysis presented in the Module 4 video, "Using Organoids and Single Cell RNA Sequencing Approaches to Study Human Cortical Development."

10X Capture and Sequencing

Single-cell capture was performed following the 10X v2 Chromium manufacturer's instructions. In each case, 10,000 cells were targeted for capture and 12 cycles of amplification for each the cDNA amplification and library amplification were performed. Libraries were sequenced as per manufacturer recommendation on a NovaSeq S2 flow cell.

Clustering

Clustering was performed as previously described. Prior to clustering, batch correction was performed in the spirit of Peng, et al Cell 2019. Briefly, each set of cells within a batch were normalized to the highest expressing gene, making the range of expression from 0 to 1. These values were multiplied by the average number of counts within the batch. These normalized datasets were piped into Seurat v2 (Butler, et al Nature Biotechnology 2018), where cells with less than 500 genes per cell or greater than 10% of reads aligning to mitochondrial genes being discarded. Normalized counts matrices were $\log_2$ transformed, and variable genes were calculated using default Seurat parameters. Data was scaled in the space of these variables, and batch was regressed out. Principal component analysis was performed using FastPCA, and significant PCs were identified using the formula outlined in Shekhar et al Cell 2016. In the space of these significant PCs, the k=10 nearest neighbors were identified as per the RANN R package. The distances between these neighbors was weighted by their Jaccard distance, and louvain clustering was performed

1121 14th Street NW
Suite 1010
Washington, DC 20005
Phone (202) 962-4000
Fax (202) 962-4941
Web SfN.org

using the igraph R package. If any clusters contained only 1 cell, the process was repeated with k=11 and up until no clusters contained only 1 cell. Cluster markers and tSNE plots were generated with Seurat package default parameters.

Cell Type Annotations

Primary cell type annotations of clusters were performed by comparison to previously annotated cell types (primarily Nowakowski et al Science 2017), and when a repository of substantial matching was not available, a combination of literature-based annotation of layer or maturation stage identity was used. When a cluster was substantially enriched based upon an age or an areal metadata property, this empirical observation was used to inform the annotation. Organoid cell types were first annotated by their similarity to primary cell clusters (using correlation analysis described below), if the correspondence was at or above 0.4 and only one primary cell type had such a high correspondence, the primary cell type was applied to the organoid cluster. If the correspondence was between 0.2 – 0.4 and included only one similarity, that cell type was used to identify the organoid cell type unless there was an obvious discrepancy in top marker gene expression between the two clusters. If no correlation was above 0.2, literature annotations or unknown identities were assigned. If an organoid cluster correlated equally well (within 10%) of multiple primary subtypes of the same or similar cell type, "pan" identity was assigned.

Correlation Analysis

Correlation analysis was generated in the space of marker genes. For each cluster, a marker specificity score was generated for each gene. This score equaled the 'enrichment' – $\log_2$(fold change) of the marker compared to other clusters – and the 'specificity' – the percent of the relevant cluster expressing the marker divided by the percent of other clusters expressing the marker. A matrix of all markers across all clusters was created for each individual dataset; if a marker was not expressed at all in a certain cluster, it was marked as 0. If a value was divided by 0 to calculate the score, the score was placed as a dummy score at 1500. Matrices between comparisons were correlated in the space of overlapping marker gene space using Pearson's correlations.

Additional Details for Correlation Analysis

Example snippet of marker genes from Seurat Marker Gene Analysis:

- The first column includes gene names, but they are often numbered if the gene is seen as a marker for multiple clusters, so we use the last column. The p-value and adjusted p-value calculation details can be found in the Seurat documentation. The avg_logFC indicates the $\log_2$(fold change) of the expression value of the gene in this cluster versus all other clusters. pct.1 represents the

1121 14th Street NW
Suite 1010
Washington, DC 20005
Phone (202) 962-4000
Fax (202) 962-4941
Web SfN.org

percent of cells in this cluster expressing the gene while pct.2 represents the percent of cells in the remaining clusters expressing the gene.

- Example calculation of gene score using $\log_2$(fold change) and the specificity of the gene
- This results in a gene score for every gene in every cluster, as seen below
- From these scores, we can make a matrix for every cluster and every marker gene. We do this using a simple perl script on the text file, but this can be done in a number of alternative ways as well. The resulting matrix will have n genes in the column and k clusters in the rows.
- To correlate this matrix to another dataset or itself to look for similarity, simply correlate the transposed matrix to another object. To correlate organoids to primary we used the following commands:

```
markergenes <- intersect(colnames(organoids), colnames(primary))
correlated <- cor(t(organoids[,genes]), t(primary[,genes]))
write.table(correlated, "organoid_vsprimary_correlation.txt", sep= "\t", quote=F,
col.names=NA)
```

- The resulting file can be visualized in Morpheus from the Broad Institute: https://software.broadinstitute.org/morpheus/

Linear Mixed Models

VariancePartition was used for linear mixed model analysis. Analysis was performed in a randomized subset of 50,000 genes in the space of expressed genes across the meta data properties. Age was used as a continuous variable and all other variables were assigned as discrete.

WGCNA and Maturation Analysis

WGCNA networks were calculated as previously described in 10,000 randomly chosen primary radial glia cells. These networks were applied to the remaining primary and organoid cells using the applyModule function. Pseudoage was calculated by taking networks that correlated highly to age in the 10,000 cell subset and combining their genes into a single gene set. PCA was performed in this gene space in the full space of radial glia and the loading of the first principle component dictated the pseudoage. Please also see this WGCNA Tutorial.

Implementations of WGCNA in single-cell RNA sequencing analysis:

Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex
Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution

SOCIETY for NEUROSCIENCE

1121 14th Street NW
Suite 1010
Washington, DC 20005

Phone (202) 962-4000
Fax (202) 962-4941
Web SfN.org

Area Signatures

Area signatures were obtained by performing pairwise differential expression between each of the seven cortical areas and the six remaining areas. Differential expression across all of the areas was combined, with a count of how many times a gene was differentially expressed in an area from each of the pairwise comparison. While combining the lists, the enrichment and specificity were averaged across all six analyses and multiplied by the number of times the gene appeared as a marker for an area of interest. This value, the "area specificity score" was compared across all areas. For any genes that were considered markers of multiple areas, the area with the highest area specificity score was allocated the gene as a marker, thus making all area markers unique to one area alone. This is how some areas have a higher percentage of cells assigned to another area other than their area or origin, and enables cleaner comparison of areal pattern emergence. Each set of area marker genes were designated as a network, and the correlation of each cell to this area was calculated by applyModules and calculating a module eigengene. After assignment, in order to normalize unequal module eigengene distributions, within a dataset the module eigengenes were normalized by area and the assigned area for a cell was the area for which that cell had the highest module eigengene.