Using Embryoid Bodies to Understand Human Biology and Evolution

Kenneth Barr[1], Katie Rhodes[1,2], Bryan Pavlovic[3] and Yoav Gilad[1,2]

1. Department of Genetic Medicine, University of Chicago 2. Department of Human Genetics, University of Chicago
3. University of California San Francisco, Department of Neurology

Abstract

The vast majority of genetic variants that are associated with disease lie within non-coding DNA and are thought to affect gene regulation. This has inspired large-scale efforts to identify variants that affect gene expression levels (eQTLs) in a wide range of adult tissues. However, most disease-associated SNPs – though they are located in putatively regulatory regions – have not been found to be eQTLs. One reason for this could be that we have not examined gene regulation in the cell types or states most relevant for disease. Many tissues and cell types are inaccessible due to practical or ethical constraints, and it is often impossible to obtain multiple samples from the same individuals. Thus, the pace of genetic discovery is fundamentally limited by access to relevant human tissues.

The discovery that mature human cells can be transformed into stem cells was an important step toward solving this problem. Induced pluripotent stem cells (iPSCs) provide a renewable source of human tissue that can, in theory, develop into any cell type. In practice, however, it can take years to discover how to produce any single tissue from iPSCs. We have identified an opportunity to generate and study many, or even most, human cell types simultaneously. When grown in the proper conditions, stem cells form spontaneously differentiating organoids known as embryoid bodies (EBs). Cells within EBs differentiate asynchronously into cell types originating from all three germ layers, including pluripotent, intermediate, and mature cell types. By applying single-cell RNA-sequencing (scRNA-seq) to cells within EBs, we can jointly identify eQTLs across a multitude of cell types, all within a controlled genetic environment. The use of EBs will also allow us to observe cellular transitions and regulatory events that are not evident in static cell culture.

EBs also afford opportunities to study evolution and human-specific novelty. Since 2011, the National Institute of Health has placed a moratorium on new research on chimpanzees. iPSCs generated from chimpanzee are the only possible way to study diverse cell living chimpanzee cell types. We are generating EBs from both human and chimpanzee iPSCs in order to understand evolutionary changes in gene expression in diverse mature and developmental cell types.

Using RCC Midway Resources

All computation is performed on RCC midway2. We use snakemake to manage computational pipelines. We align individual datasets to the genome using the alignment software STAR. Because we often pool individuals with different genotypes, we find the genetic identity of each individual cell using the computational tool Demuxlet. Similarly, in evolutionary comparisons, we pool cells from both humans and chimpanzees. We use in-house tools written in python to compare genome alignments and assign each individual cell to the proper species. We then analyze the resulting single cell datasets with packages for the R statistical programming language. These packages include Seurat, Monocle, and URD.

Single cell datasets are large and sparse and require novel computational tools. The human genome contains over 40,000 genes and we are generating data for tens to hundreds of thousands of individual cells. These large matrices are often too large to fit into memory. We use hdf5 formats to analyze data in pieces in a memory-efficient manner.

EBs Generation, Cell Type Identification, and Trajectory Inference

We formed EBs in micropatterned Aggrewell dishes at a density of 2000 iPSCs per well. We grew the EBs for 3 weeks, feeding 2mL of E6 media every other day. On day 21, we dissociated the cells with a 30 minute incubation in TryLE and submitted them for single-cell isolation and sequencing on the 10x genomics Chromium playform. We sequenced 5000 cells on 1 lane of a HiSeq4000. Data analysis was performed with Seurat. We identified cell types by marker gene expression. We identified trajectories with URD.__

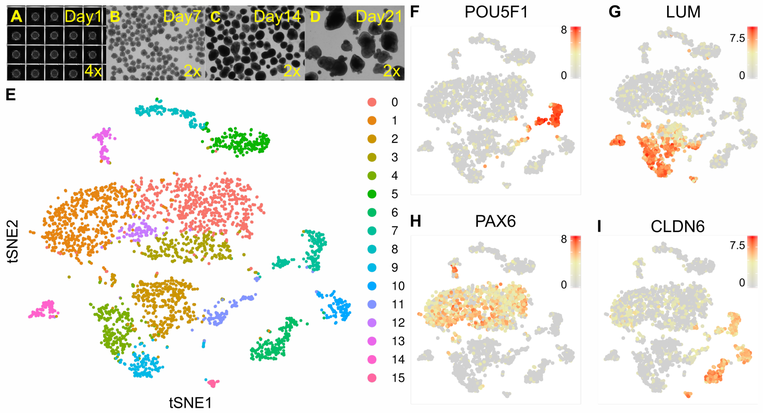

Figure 1: (A) Brightfield microscopy of iPSC aggregates in AggreWellTM dishes after 24 hours. Imaged at 4x. (B-D) Brightfield microscopy of EBs at 7, 14, and 21 days. Imaged at 2x. (E) tSNE of transcriptomes of 21 day EBs. Seurat identified 16 clusters using resolution = 1. (F-I) tSNE colored by the number of Unique Molecular Identifiers (UMIs) detected for pluripotency marker POU5F1, mesoderm marker LUM, ectoderm marker PAX6, and endoderm marker CLDN6. Number of UMIs is indicated with a color scale bar.

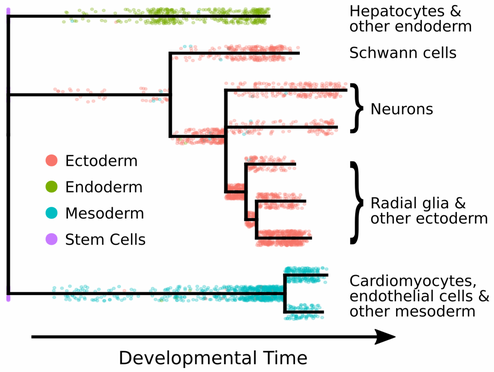

Figure 2. Dendogram of embryoid body development from scRNA-seq on day 21. Each point represents the transcriptome of a single cell. The cells are ordered on the x-axis according to developmental time (inferred with URD). Cells are colored by the inferred germ layer. Branches are labeled by cell types that we have identified through marker gene expression.

Comparing Delelopmental Trajectories of Humans and Chimpanzees

We formed EBs from 6 humans and 3 chimps on Agrrewell dishes as previously described. After 21 days, we dissociated EBs with a 30 minute incubation in Accumax. We pooled cells from all individuals at equal proportions and sequenced the transcriptomes of 50,000 cells on the 10x Chromium platform.

We alighed reads to both human and chimpanzee genomes using STAR solo. We then assigned each cell to the proper species by comparing alignments. We identified discordant reads, which map better in one species than the other. We identified a cell as belonging to a species if 80% or more discordant reads mapped to that species, and there were at least 10 discordant reads. After species assignment, we assigned cells to each individual using Demuxlet.

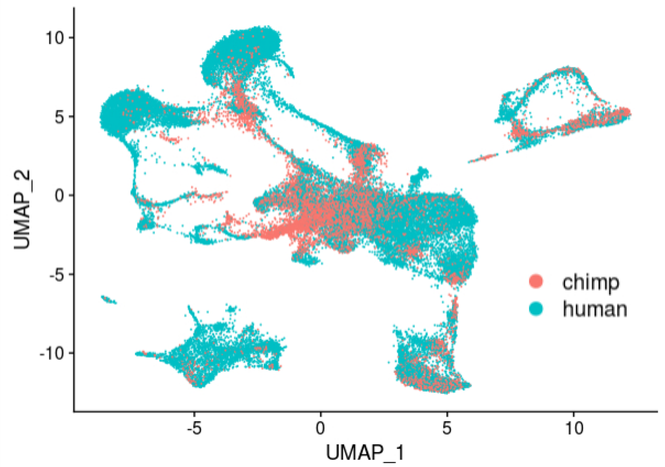

Figure 3: UMAP of 21 day human and chimp EBs. We generated the UMAP using Seurat with default parameters. Humans and chimpanzee cells are generally well dispersed, indicating that human and chimpanzee EBs are composed of the same cell types.

Future directions

We are sequencing 21 day human and chimp EBs at greater depth, and generating 2 additional biological replicates. From this data we will identify differentially expressed genes in humans and chimpanzees. Additionally, we will sequence EBs from a genetic reference panel of Yoruba iPSCs. From this panel we will identify eQTLs.