PLANT SCIENCE

Defining the developmental program leading to meiosis in maize

Brad Nelms* and Virginia Walbot*

Geneiotic Breeders

A person who combines the fields genetics and plant breeding to study meiosis.

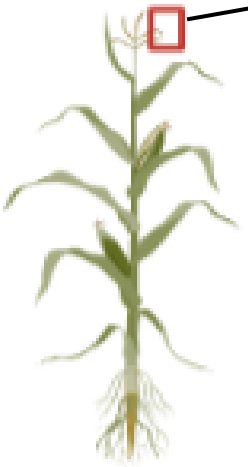
Tanner Cook
Yu-Ru Chen
Juan Panelo
Sean McLaughlin
Chiteri Kevin

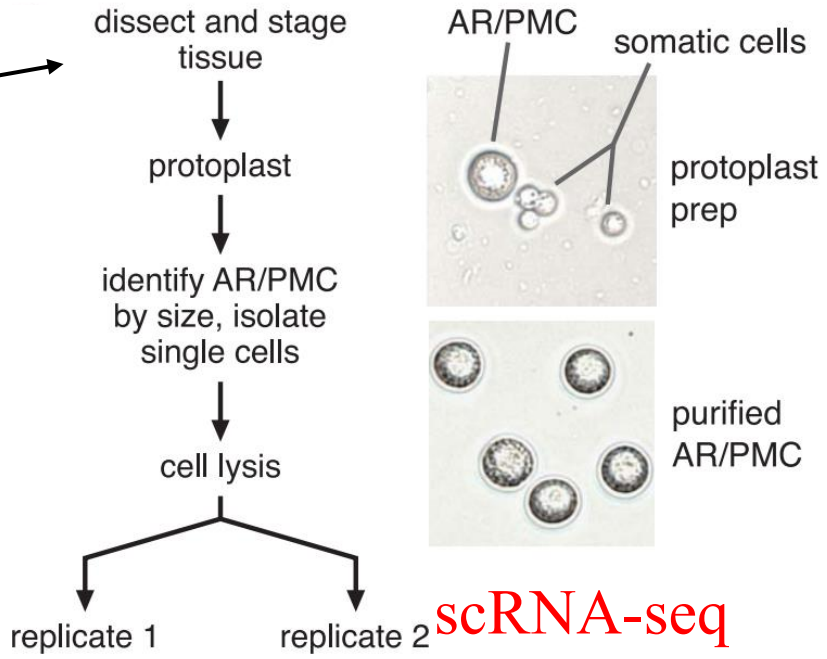
BCB546x
4/30/2021

Objectives

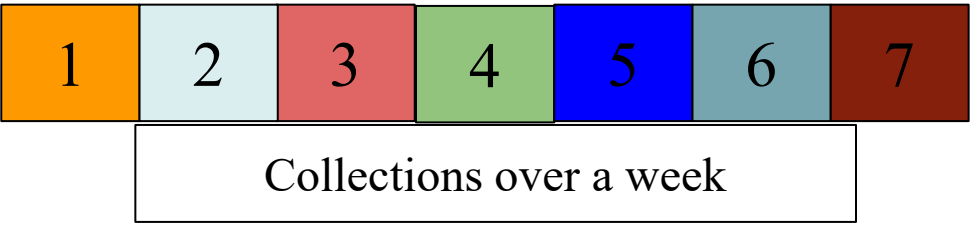
- Reconstruct the developmental program of maize male meiosis by single cell RNA-seq
- Identify cell differentiation/morphology
- Identify genes and their expression profiles

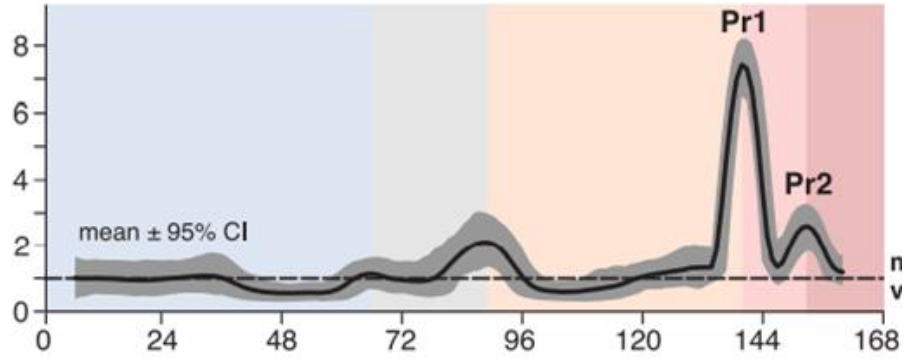
Background





scRNA-seq





Created in BioRender.com

Modified from Nelms et al., 2019

Workflow

gene	Cell 1 replicate 1	Cell 1 replicate 2
Zm00001d027231	7	5
Zm00001d027248	0	0

Dataset

31184 genes, 216 cells.

Filter cells out that did not pass quality control

~ 12,500 genes

144 cells passed QC with 1,2 reps

Analyses comparing technical replicates

Heatmap

128 cells that passed QC with 2 reps

Normalize, log transform, add pseudo count 11, filtering by counts/gene

PCA

Pseudotime

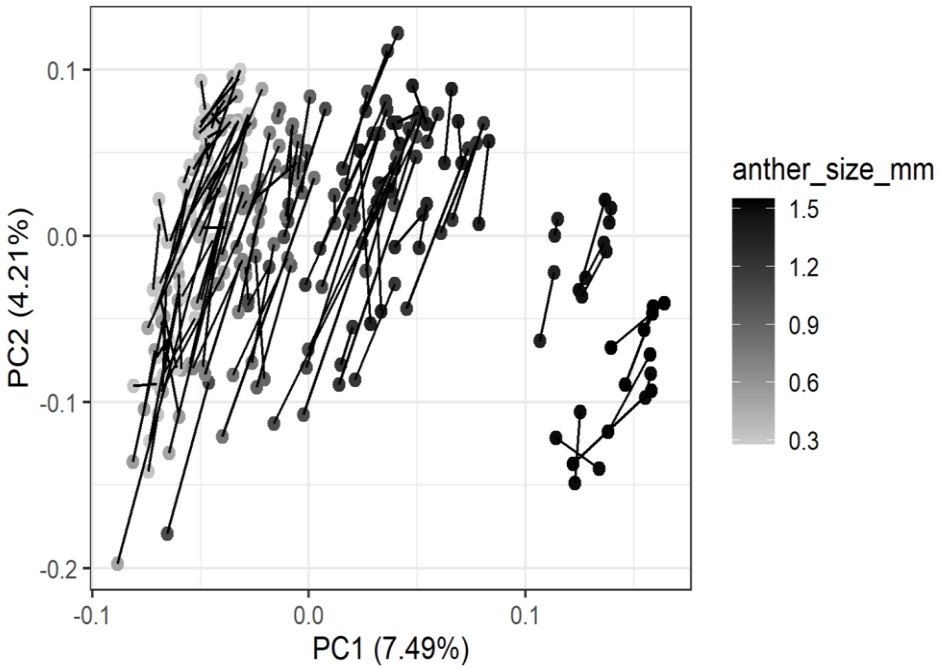
~2,000 genes

princurve

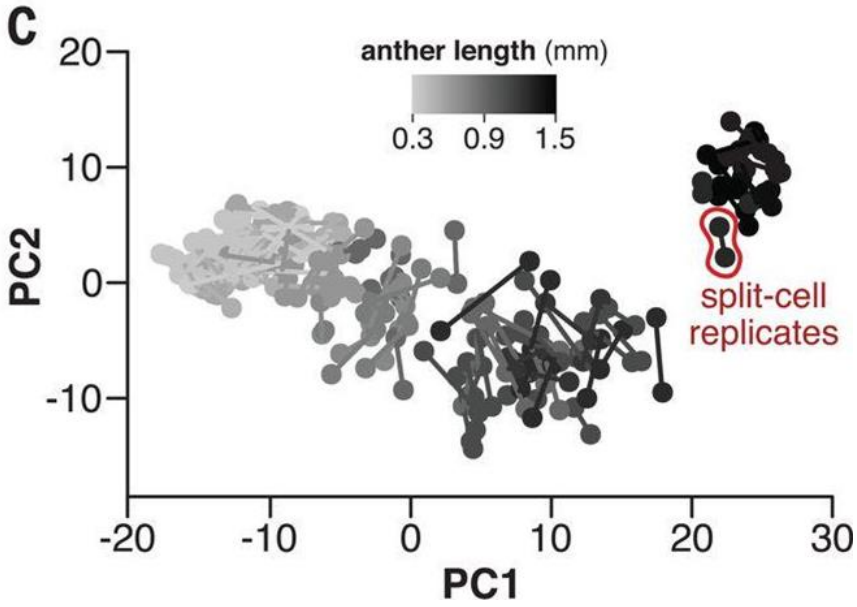
pheatmap

PC Analysis

Ours



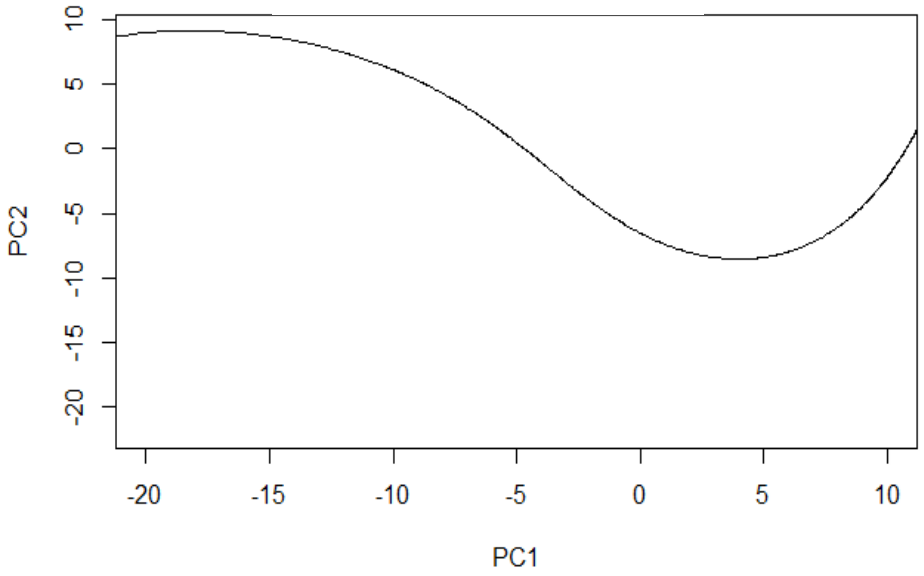
Nelms et al., 2019



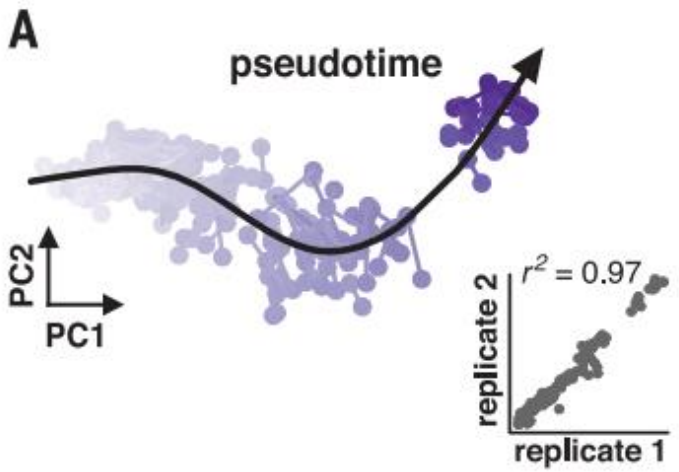
Pseudotime

- Take normalized data with 128 cells and then select the genes with the 2000 highest variances
- Run a PCA on the data and then take the first 10 PC's for the principal curve

princurve



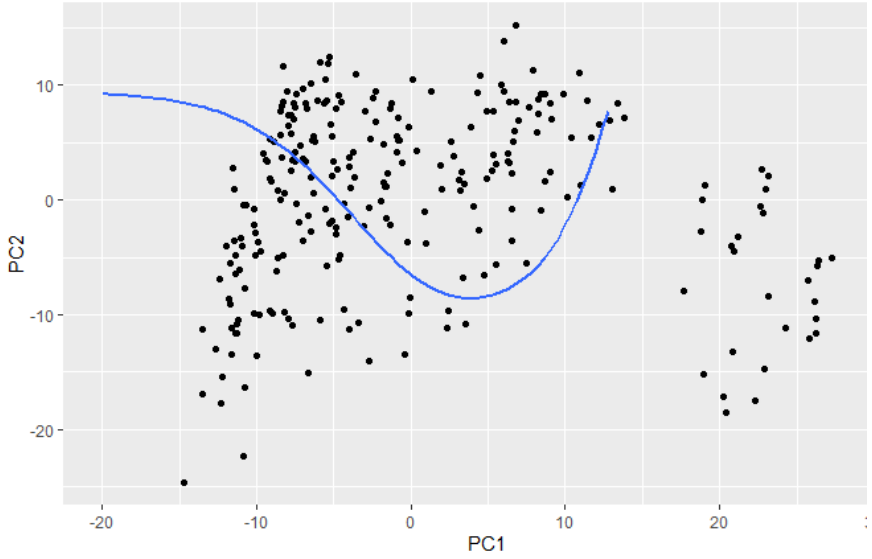
Nelms et al., 2019



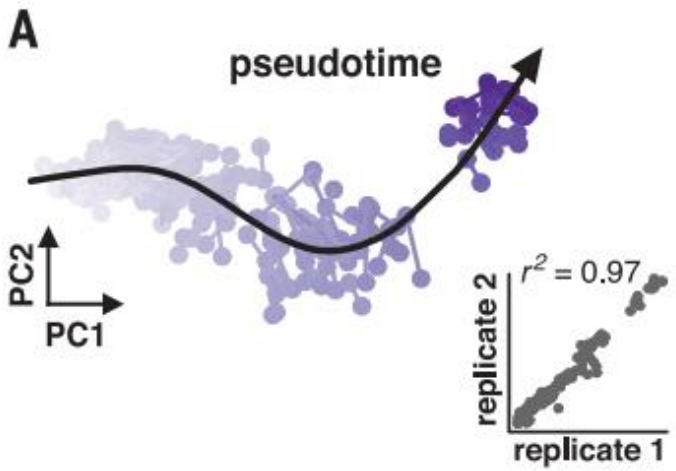
Pseudotime

- Take normalized data with 128 cells and then select the genes with the 2000 highest variances

- Ran a PCA on the 2000 genes and then took the first 10 PC's for the principal curve



Nelms et al., 2019



Attempting to Develop Pseudotime Through Other Means

An R toolkit that specializes in analysis of RNA expression data
Had references for other expression analysis done through software
Data set up in Nelms et al. does not conform to needed format for monocle
Data between supplementary materials is different

The CellDataSet class

Monocle holds single cell expression data in objects of the `CellDataSet` class. The class is derived from the Bioconductor `ExpressionSet` class, which provides a common interface familiar to those who have analyzed microarray experiments with Bioconductor. The class requires three input files:

- `exprs`, a numeric matrix of expression values, where rows are genes, and columns are cells
- `phenoData`, an `AnnotatedDataFrame` object, where rows are cells, and columns are cell attributes (such as cell type, culture condition, day captured, etc.)
- `featureData`, an `AnnotatedDataFrame` object, where rows are features (e.g. genes), and columns are gene attributes, such as biotype, gc content, etc.

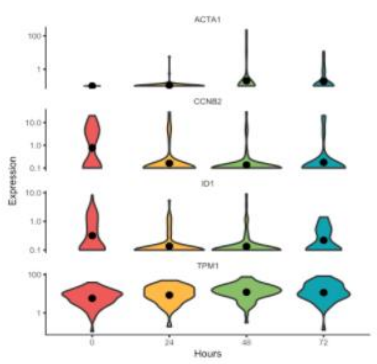
Required dimensions for input files

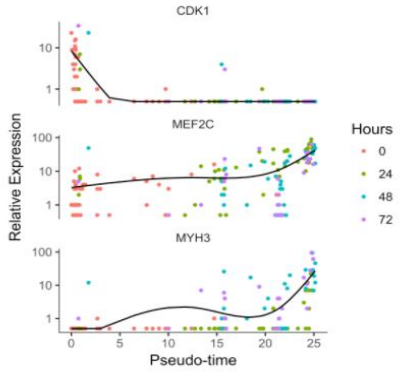
The expression value matrix **must**:

- have the same number of columns as the `phenoData` has rows.
- have the same number of rows as the `featureData` data frame has rows.

Additionally:

- row names of the `phenoData` object should match the column names of the expression matrix.
- row names of the `featureData` object should match row names of the expression matrix.
- one of the columns of the `featureData` should be named "gene_short_name".

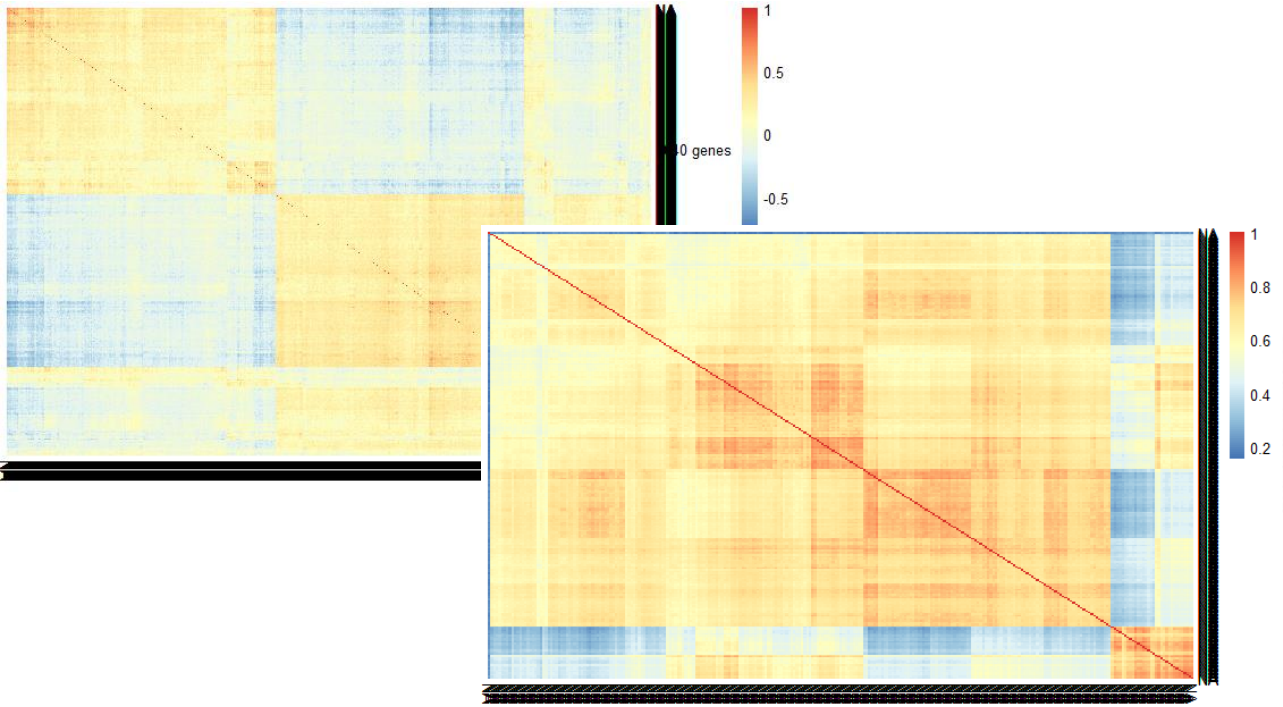




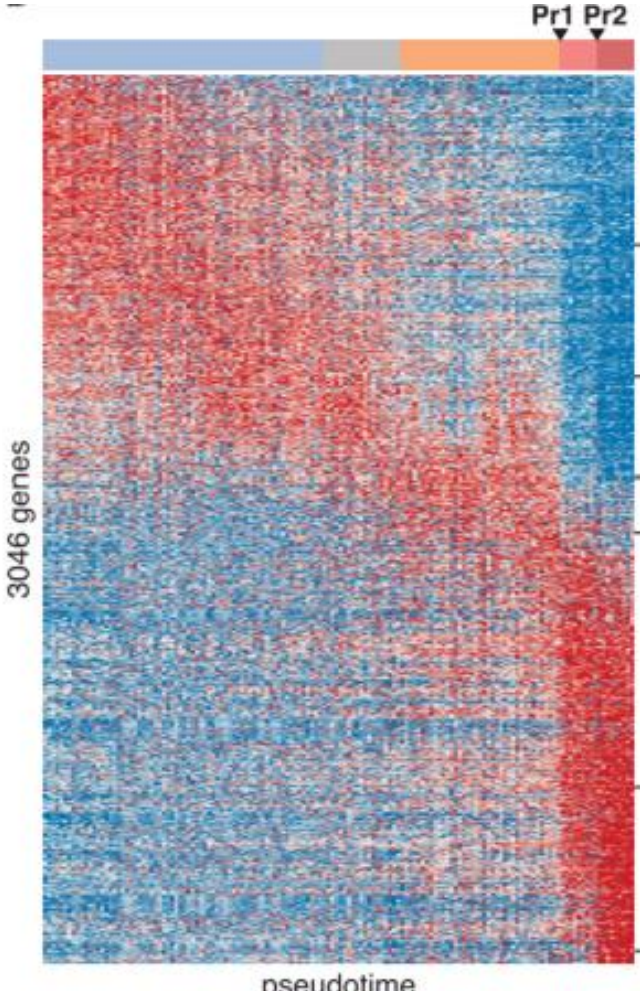
Pseudotime heatmap

- 3040 genes and 144 cells.
- Use pseudotime in x-axis and and gene expression in Y.

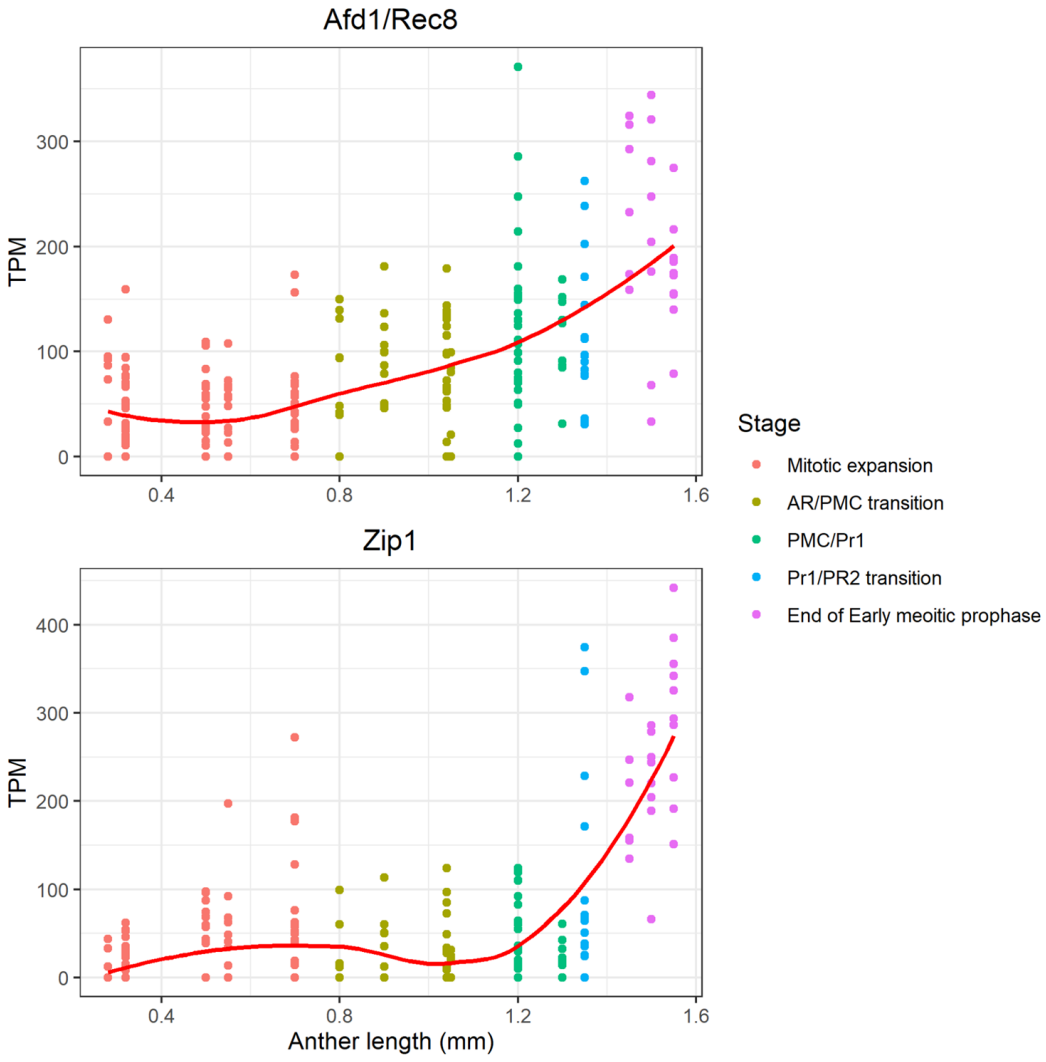
Pheatmap in R Studio

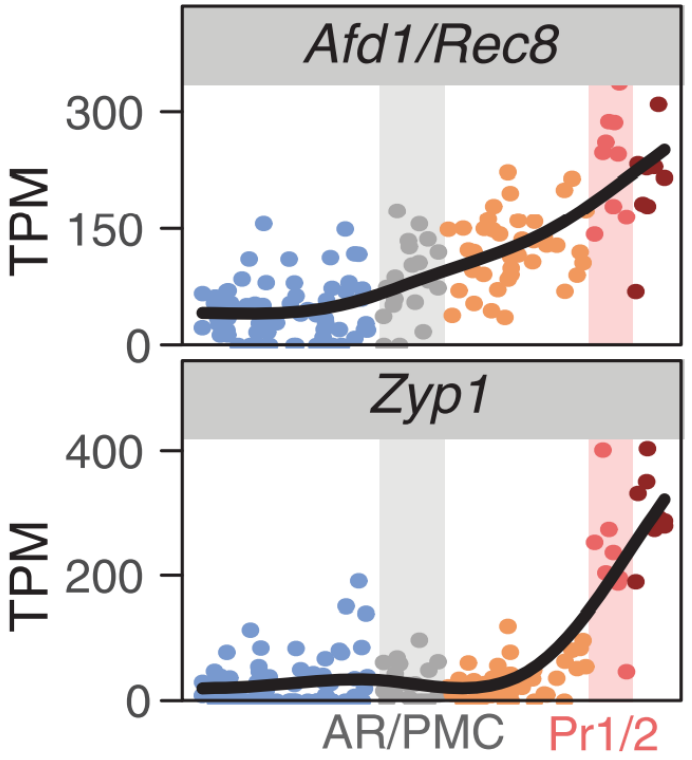


Nelms et al., 2019



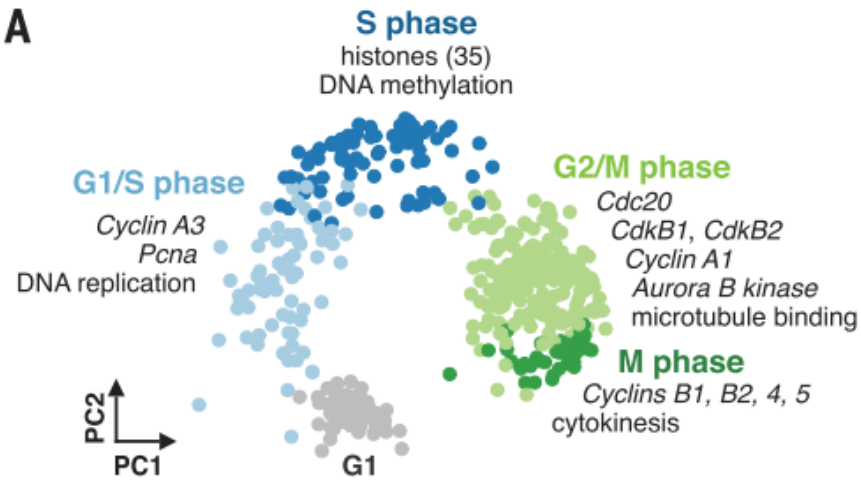
Cell Cycle–Regulated Genes



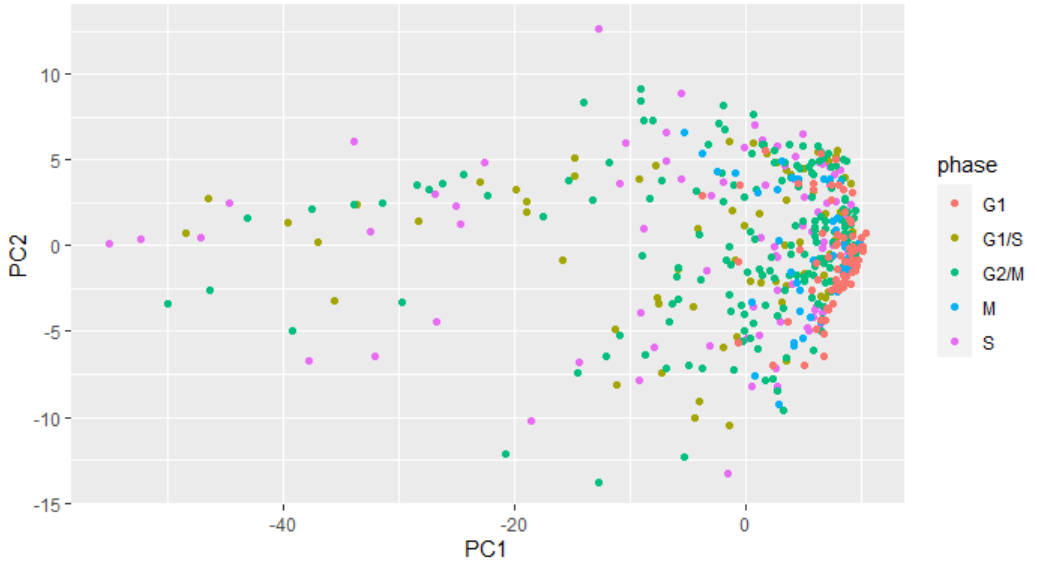


Cell Cycle–Regulated Genes

Regulated genes filter threshold?
Normalized method?



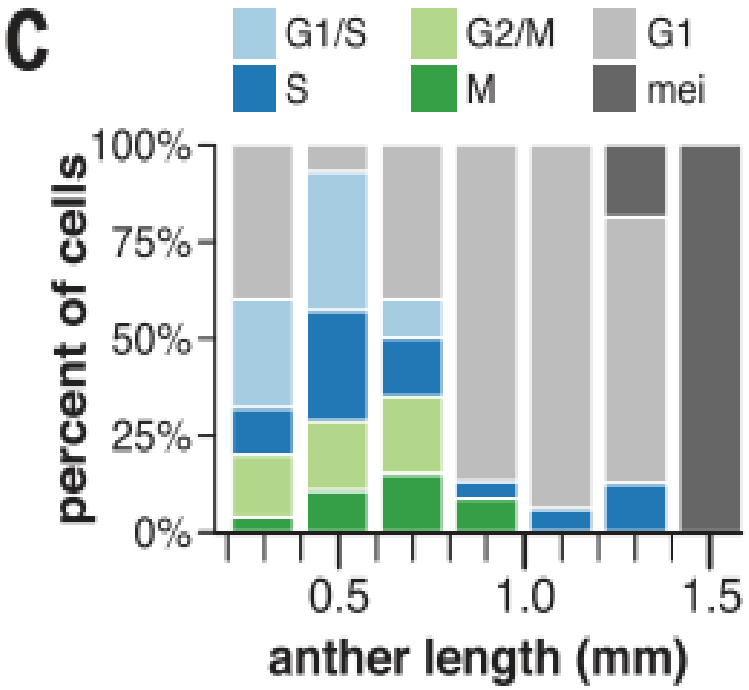
Principal component plot of cell cycle–regulated genes
Figure 5A

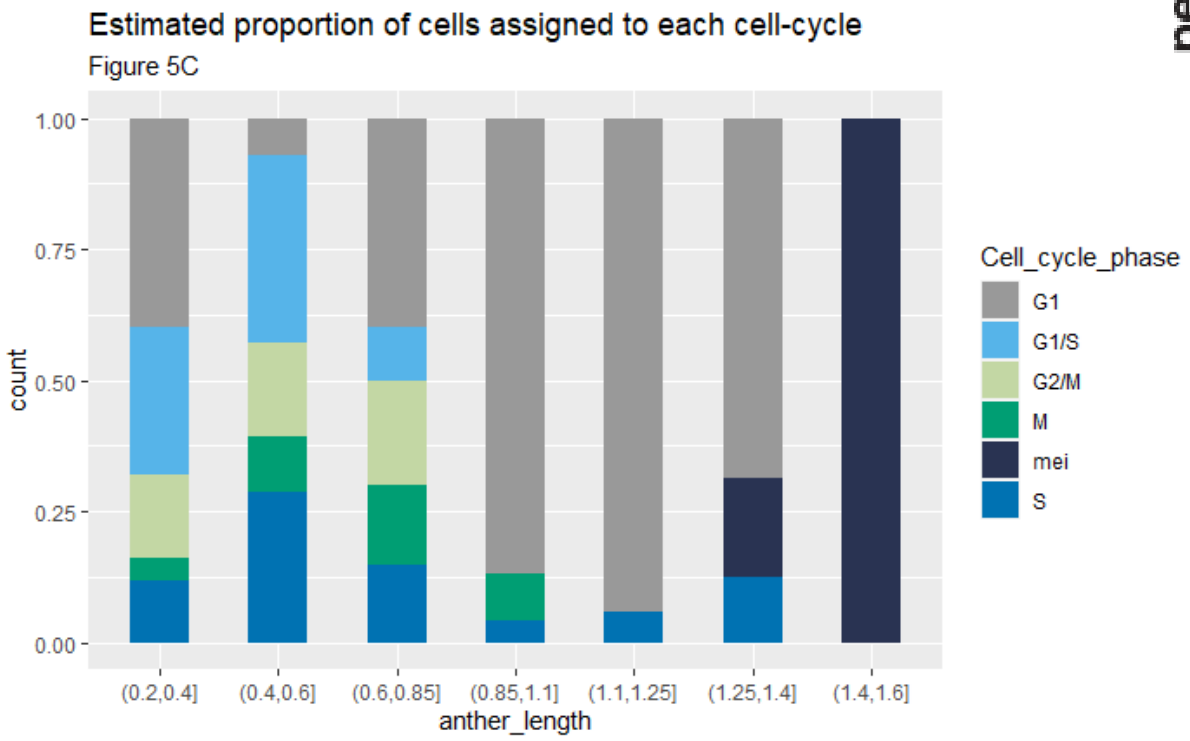


448 expresses genes in meiotic cycle

Cell Cycle–Regulated Genes

Gene expression profile of 144 cells in different anther length





Expression of Marker Genes by Pseudotime

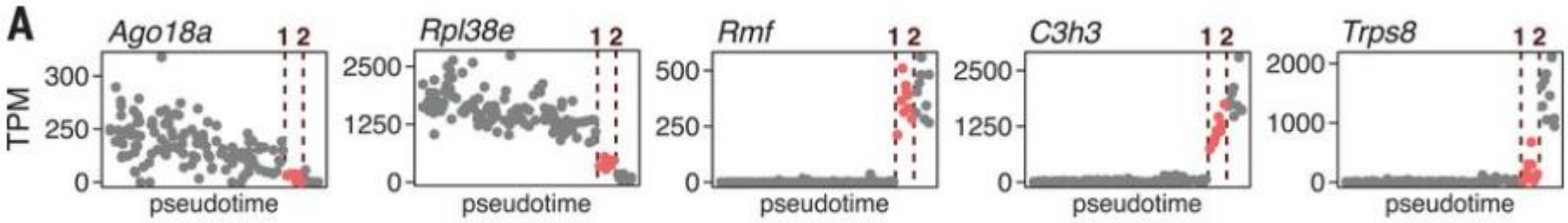


Figure 3A of Marker genes Ago18a etc. includes Expression data put with pseudotime.

Was not able to replicate graphs

No description of specific graph formation and did not include marker gene information within supplementary materials

Information on Rpl38e found through search does not show in data given

Isolate Marker gene expression data

↓

Normalize data --- TPM

↓

Merge with pseudotime

↓

Develop graph of expression over pseudotime

Heatmap by Anther Size

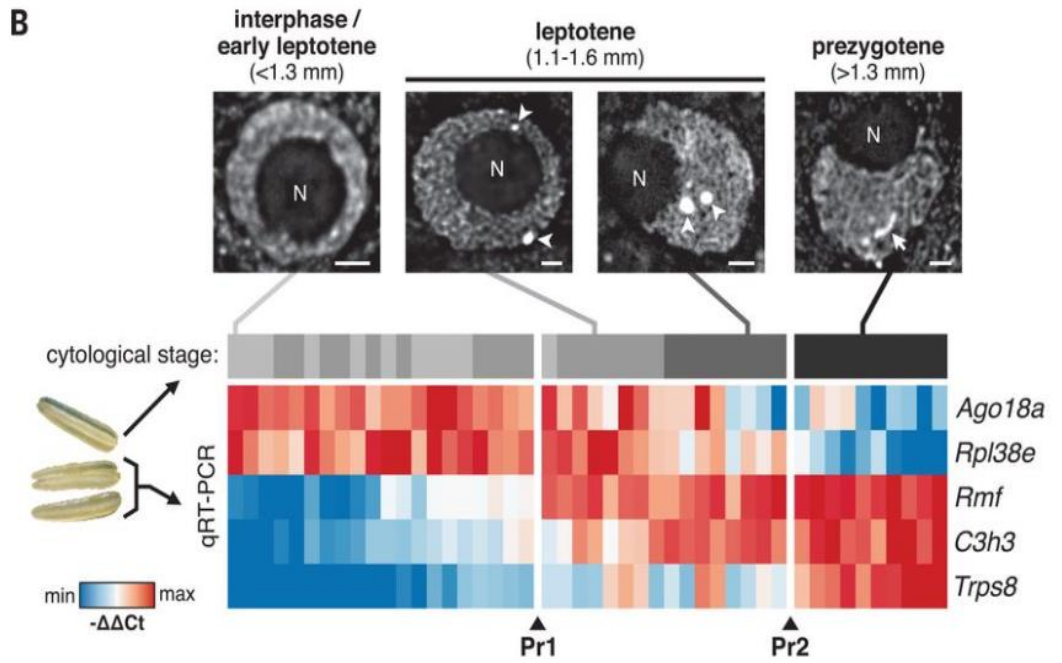
- Was able to produce a heatmap of product
- Stark differences between Original and recreated products
- Differences in amount measures (e.g 46 measurements vs 216)
- No specifics on graph structure and components
- Missing genes from Zm numbers and Quality control

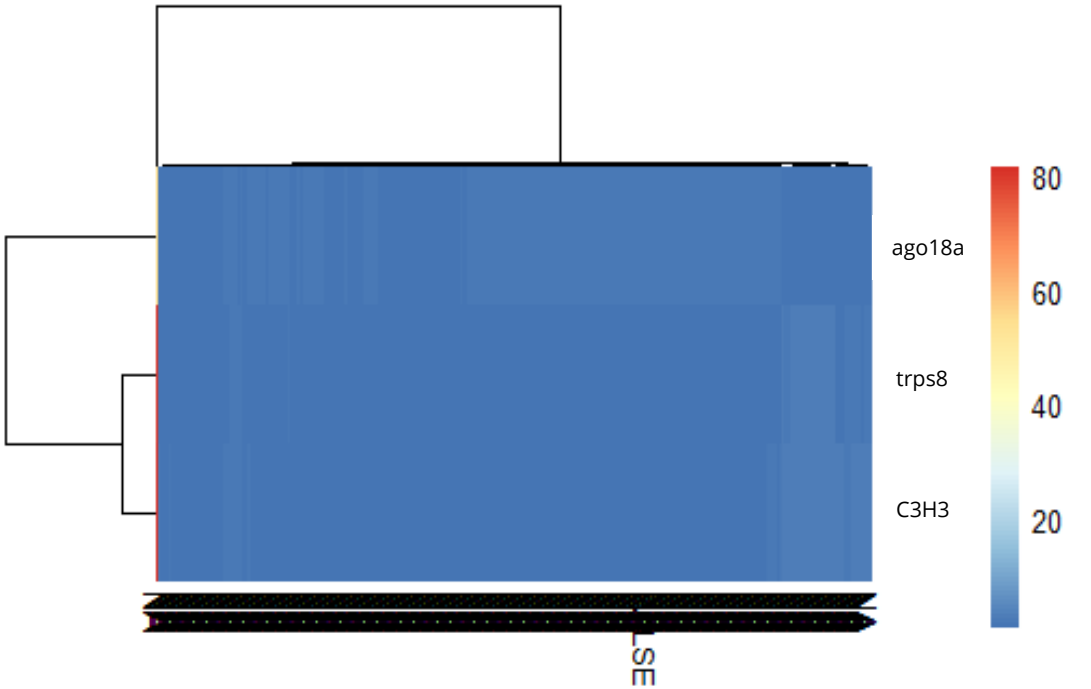
Arrange Anther Data and Marker Genes

Normalize Marker Gene Expression

Merge Datasets

Produce Heatmap





Conclusion

Challenges to reproduce the results in the project

- Cloud-based repositories are imperative for reproduction of results
- Deprecated packages provide enough difficulties without making parameter and linguistic “guesses”
- Had to source alternative methods to develop the data due to trouble reproducing data development described in methods
- Clear and concise documentation of methods is really helpful for the reproducibility