# **Papers to Read**

Ideal scenario: study each paper over the span of two weeks. I’d probably read the paper the weekend before, then spend the following two weeks studying and replicating what the authors did.

**“Bayesian estimation of Differential Transcript Usage from RNA-seq data”**

The authors created new methodology for identifying genes consisting of differentially expressed transcripts that relies on Bayesian statistics rather than frequentist statistics (which the existing technologies, cjBitSeq and DRIMSeq use).

* Working with RNA-seq data is something I learned a bit about this summer, so I already have a knowledge basis
* Bayesian techniques would be a cool way to integrate statistics into this project; I know basic statistics but have never learned about Bayesian stats before
* Paper uses synthetic data, which is an interesting concept

**“Algorithms to Model Single Gene, Single Chromosome, and Whole Genome Copy Number Changes Jointly in Tumor Phylogenetics”**

The authors construct phylogenetic models of cellular-level indications of tumor progression, like gene copy number changes.

* Phylogenetics would be cool to learn about
* Discusses cancer biology, which from what I can tell is a hot topic in computational biology

**“COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease.”**

The authors model progression of Alzheimer’s using multiple machine learning methods such as support vector machine.

* Not based in cell and molecular biology, unlike previous papers considered
* It’d be cool to learn about more clinical applications of bioinfo

**“Metabolic Network Prediction of Drug Side Effects”**

The authors use a combination of medical informatics data and a human genome-scale metabolic model to predict drug side effects.

* Learning how to work with networks would be useful
* Has some machine learning involved

**“Data-analysis strategies for image-based cell profiling”**

The authors create a workflow for analyzing images of cells in order to quantify differences in phenotype.

* I’d be interested in learning about image processing
* This paper is more removed from public health/biomedicine than the other papers

I’ll probably select a few additional papers, just in case I cover these papers more quickly than I anticipate to, or in case one of the above papers is too challenging/relies on data I don’t have access to.

# **Background Info (read over winter break)**

**“Practical computational reproducibility in the life sciences”**

* The idea of reproducibility should be considered throughout this project

**“A primer on molecular biology”**

* It *might* be useful to know a thing or two about molecular biology…...

**At least a few of the papers from** [**MIT’s Foundations of Computational Biology course**](https://ocw.mit.edu/courses/biology/7-91j-foundations-of-computational-and-systems-biology-spring-2014/readings/)

* Some of these papers are introductions to certain fields (e.g. sequence alignment) while others are research papers in related areas that might provide more insight into the topic, or that could give an idea of a new direction for existing research to go in?