

# Graduation Timeline

Nicholas Knoblauch

May 20, 2020

My plan is that by the end of July I have convinced the committee that no substantial additional work needs to be done before I defend my thesis. I've put down times for when I would like to have certain items finished so as to make that deadline. I believe the timeline is aggressive but (hopefully) achievable. **TODO** items are highlighted in red, and where there are multiple items in a category, they are prioritized by letter, (with "A" being the highest).

## Functional Genetics by Expectation Maximization Project

My plan is for this chapter to consist of a submitted paper (targeting Bioinformatics). **DEADLINE:** *<2020-06-06 Sat>*

### **TODO Finish analyses**

**DEADLINE:** *<2020-05-22 Fri>*

- **DONE** #A Run FGEM on putative cancer genes from driverMAPS paper **DEADLINE:** *<2020-05-07 Thu>*  
I ran FGEM on 20 lists of gene-level Bayes Factors obtained from [1]
- **TODO** #A Validate FGEM features against gold standard (IntO-Gen) **DEADLINE:** *<2020-05-21 Thu>*  
I will validate the set of gene-level features I identify by estimating enrichment for cancer genes in the intOGen database [2]
- **DONE** #B Make a decision about feature collinearity issue **DEADLINE:** *<2020-05-19 Tue>*

If the features from genes are highly correlated, the parameter estimates can be unstable. The current plan is to detect and merge or remove highly correlated features from the analysis

**TODO Finish draft**

**DEADLINE:** *<2020-05-29 Fri>*

## Polygenic RSS

As it stands there will probably more than one paper's worth of work to be written up for this project by the time it is finished. The current plan is for this project is to get one submitted paper out of the project and for the rest of the work to go in the dissertation.

**Write Paper**

**DEADLINE:** *<2020-06-30 Tue>*

**TODO Finish analyses**

**DEADLINE:** *<2020-06-04 Thu>*

- **DONE** #A Simulate UK Biobank GWAS with polygenic genetic architecture
- **TODO** #A Simulate UK Biobank GWAS with sparse genetic architecture
- **TODO** #B Finish UK-biobank/Genbank GWAS real-data analysis

## Preterm Birth Fine Mapping

**DEADLINE:** *<2020-07-30 Thu>*

My plan is to write up the analyses that I've already done (that went in to the first March of Dimes paper) as my third chapter. Depending on how things go, additional analyses that need to be done can go in to another paper or simply into the chapter.

**TODO** #A Write up analyses for chapter

**DEADLINE:** <2020-07-15 Wed>

**TODO** #B Finish additional analyses

**TODO** #C Incorporate TCM term vs preterm ATAC annotations into a new model

**TODO** #D Incorporate TF ChIP-seq annotations into a new model

**TODO** #D Compare which loci are implicated by which models

## References

- [1] Siming Zhao, Jun Liu, Pranav Nanga, Yuwen Liu, A. Ercument Cicek, Nicholas Knoblauch, Chuan He, Matthew Stephens, and Xin He. Detailed modeling of positive selection improves detection of cancer driver genes. *Nature Communications*, 10(1):3399, 2019.
- [2] Abel Gonzalez-Perez, Christian Perez-Llamas, Jordi Deu-Pons, David Tamborero, Michael P Schroeder, Alba Jene-Sanz, Alberto Santos, and Nuria Lopez-Bigas. Intogen-mutations identifies cancer drivers across tumor types. *Nature Methods*, 10(11):1081–1082, 2013.