

# Storey Lab Notebook

Peter Edge

June 16, 2014

## June 9, 2014

Today I did not yet have Emily's code available to study or access to the cetus data, so instead I focused on preparing my workstation and reading literature related to my project. I focused on Skelly et al (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3202289/>) because it is perhaps the most relevant to what I am doing.

## June 10, 2014

Today I briefly introduced myself to Emily's snp\_pipeline code before spending most of my day at lab safety training.

## June 11, 2014

Today I focused on reading Emily's code in the snp\_pipeline.py file, and familiarizing myself with the libraries, tools, and paradigms it uses. I read through the documentation for ruffus (<http://www.ruffus.org.uk>) as well as reading up a bit on python iterators.

## June 12, 2014

Today I finished studying the details of Emily's snp\_pipeline and began trying to use pysam myself for the manipulation of sam files. I created directories RM\_bowtie\_test and S288C\_bowtie\_test in SeqSorter/sample\_data and used them to align the test reads to the RM and S288C genomes using bowtie:

```
cd git/SeqSorter/sample_data
mkdir RM_bowtie_test S288C_bowtie_test
bowtie2-build S288C_reference_genome_R64-1-1_20110203/S288C_reference_sequence_R64-1-1_20110203.fsa S288C_reference_genome_R64-1-1_20110203
bowtie2-build RM11_1A/assembly/genome.fa RM_bowtie_test/RM_index
bowtie2-build S288C_reference_genome_R64-1-1_20110203/S288C_reference_sequence_R64-1-1_20110203.fsa S288C_bowtie_test/BY_index
bowtie2 -x RM_bowtie_test/RM_index -U E2a0_sample.fastq
-S RM_bowtie_test/RM_bowtie_out.sam
bowtie2 -x S288C_bowtie_test/BY_index -U E2a0_sample.fastq
-S S288C_bowtie_test/BY_bowtie_out.sam
```

I then started experimenting with pysam to understand how the library is used to access data from the alignment files.

## **June 13, 2014**

Today I narrowed my focus to the method `compare_mappings(infile, outfile)`, since this is where the initial focus of my project will be. I familiarized myself with using pysam for navigating SAM files, using the alignment SAMs from the test dataset (created yesterday).