Goal: What are the mutations? At generation 1 vs generation 100?

Materials:

SRA for gen0: SRA0 = **control (E1\_gen007=****SRR515969)**

SRA for gen100: SRA100 = **case (E1\_gen133=SRR519089)**

(Echo $PATH

Export PATH=$PATH:/home/fzhang/Research/2013\_research\_plan/Yeast\_experiment/bwa-0.7.5a)

1. **Reference sequence for GSY1135**

* Download GSY1135 from [27]

<http://sra.dnanexus.com/studies/SRP002895/runs> (SRA020606.1 -> SRR063399)

* Map GSY1135 to S288C (**Chr10**) using BWA (GSY1135: reads; S288C: references)
  + - Align indexd fastq: bwa aln –B 6 –t 8 ./Chr10.fa ./Reference\_GSY1135/SRR063399.fastq > ./SAI\_files/SRR063399.sai
    - bwa samse ./Chr10.fa ./SAI\_files/SRR063399.sai ./Reference\_GSY1135/SRR063399.fastq > ./Reference\_GSY1135/ SRR063399.sam
    - Convert SAM to **BAM**: samtools view –bt ./Chr10.fa –o ./Reference\_GSY1135/SRR063399bam ./ Reference\_GSY1135/SRR063399.sam
    - Sort the BAM file: samtools sort ./SRR063399.bam ./SRR063399.sort
    - Index the sorted BAM file:  samtools index ./SRR063399.sort.bam **rename the reference -> GSY1135\_Chr10**
* Call SNPs using GATK UnifiedGenotyper <http://www.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_sting_gatk_walkers_genotyper_UnifiedGenotyper.html>
  + - Download picard-tools-1.104.zip from: <http://sourceforge.net/projects/picard/files/>
    - Create Dictionary: Java –jar CreateSequeneDictionary.jar R=Chr10.fa O=Chr10.dict
    - Add head: Java –jar AddOrReplaceGroups.jar I=SRR063399.sort.bam O=./headSRR063399.sort.bam LB=whatever PL=illumine PU=whatever SM=whatever
    - Index: Samtools index ./headSRR063399.sort.bam
    - Call SNPs: Java –jar GenomeAnalysisTK.jar –R Chr100.fa –T UnifiedGenotyper –I ./headSRR063399.sort.bam –o SRR063399.vcf –stand\_call\_conf 50.0 –stand\_emit\_conf 10.0
* Create a FASTA GSY1135 reference using GATK FastaAlternative (GSY1135\_Chr10)

<http://www.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_sting_gatk_walkers_fasta_FastaAlternateReferenceMaker.html>

* java –Xmx2g –jar GenomeAnalysisTK.jar –R Chr10.fa –T FastaAlternateReferenceMaker –O GSY1135\_Chr10.fasta –variant ./Reference\_GSY1135/SRR063399.vcf

1. **Download SRAs from Dan’s Paper** <http://sra.dnanexus.com/studies/SRP013879/runs>
2. **Convert SRA0 and SRA100 to FASTQ format**
3. **Remove WT population using** **FASTX Barcode Splitter**

<http://hannonlab.cshl.edu/fastx_toolkit/commandline.html#fastx_barcode_splitter_usage>

* + - cat s\_2\_100.txt | /usr/local/bin/fastx\_barcode\_splitter.pl --bcfile mybarcodes.txt --bol --mismatches 2 \ --prefix /tmp/bla\_ --suffix ".txt

1. **Trim Nextera tag using Cutadapt**

<https://pypi.python.org/pypi/cutadapt>

* + - cutadapt –a CAAGCAGAAGACGGCATACGAGATNNNNNNCGGTCTGCCTTGCCAGCCCGCTCAG –m 15 Gen007\_unmatched.fastq > testGen007.fastq

1. **Map FASTQ0 and FASTQ100 to GSY1135 reference using BWA.**

(Quality score: 50)

(Choose GSY1135\_Chr10. Look for where is **CYR1** in E1)

* + - Align indexd fastq: bwa aln –B 6 –t 8 **–I –q 10** ./Reference\_GSY1135/GSY1135\_Chr10.fa ./BARCODE/Trimed\_Gen007.fastq > ./SAI\_files/Gen007.sai
    - bwa samse ./Reference\_GSY1135/GSY1135\_Chr10.fa ./SAI\_files/Gen007.sai ./BARCODE/Trimed\_Gen007.fastq > ./Control\_Case\_files/ Gen007.sam
    - Convert SAM to BAM: samtools view –bt ./Reference\_GSY1135/ GSY1135\_Chr10.fa –o ./Control\_Case\_files/ Gen007.bam ./Control\_Case\_files/ Gen007.sam
    - Sorted BAM file was created with Picard v1.45 FixMateInformation <http://picard.sourceforge.net/>

1. **Make pileup files for FASTA0 and FASTQ100 using Samtools** <http://www.biostars.org/p/63429/#73512>
   * + ~~Sort the BAM file: samtools sort ./Gen007.bam ./Gen007.sort~~
     + Index: Samtools index ./Gen007.sort.bam
     + samtools mpileup –C 50 -uf ./Reference\_GSY1135/GSY1135.fasta ./Gen007.bam > file.mpileup
2. **Make depthchart files using pileup2dc**
   * + cd src/pileup2dc
     + gcc -o pileup2dc main.c
     + mv pileup2dc ../../bin/
3. **Estimate models for case + control**
4. **Test for variants between case + control -> vcf files**