UF HPC Training
Working with NGS data
December 1, 2011

1. Get some data:
    a. Shared Data: Data Libraries: Training datasets: filtered.100K.fastq
2. NGS: QC and manipulation: **FATQ Groomer**
    a. Input FASTQ quality scores type: Illumina 1.3+
3. NGS: QC and manipulation:Fastqc: **Fastqc QC**
    a. Use the defaults or add a title for easier reference later
4. NGS: Assembly: **velveth**
    a. Hash Length: 29
    b. Click Add new Input Files
        i. File format: fastq
        ii. Dataset: Select your reads file
5. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velvet h output
6. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velvet h output
    b. Set minimum contig length: yes
        i. Minimum contig length: 500
7. NGS: Mapping: **Map with Bowtie for Illumina**
    a. Will you select a reference genome from your history or use a built-in index?: Use one from the history
    b. Select the reference genome: Use the velvetg results with 500bp min contig length
8. Save the SAM file and the contigs used as reference, and open in Tablet (available at http://bioinf.scri.ac.uk/tablet)

From the command line:

1. Login to submit
2. Go to you scratch space, make a directory called velvet_test and cd into it
3. Copy /scratch/hpc/bio/training/01Dec11/filtered.100K.fastq to velvet_test
4. Copy /scratch/hpc/bio/training/01Dec11/velet.qsub to velvet_test
5. Edit the velvet.qsub file to have your e-mail.
6. qsub velvet.qsub