UF HPC Training

Working with NGS data in Galaxy

June 20, 2012

1. Log into UF HPC's Galaxy instance: http://galaxy.hpc.ufl.edu/
2. Get some data:
    a. Shared Data: Data Libraries: Training datasets: wine_yeast.100K.fq
3. NGS: QC and manipulation: **FASTQ Groomer**
    a. Input FASTQ quality scores type: Sanger
4. NGS: QC and manipulation:Fastqc: **Fastqc QC**
    a. Use the defaults or add a title for easier reference later
    b. Notice poor quality at ends of reads
5. NGS:QC and manipulation:**FASTQ Quality Timmer**
    a. Window size: 4
    b. Quality score: 30
    c. Rerun Fastqc QC on trimmed dataset
6. NGS: Assembly: **velveth**
    a. Hash Length: 29
    b. Click Add new Input Files
        i. File format: fastq
        ii. Dataset: Select your reads file
7. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velveth output
8. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velveth output
    b. Set minimum contig length: yes
        i. Minimum contig length: 500
9. NGS: Mapping: **Map with Bowtie for Illumina**
    a. Will you select a reference genome from your history or use a built-in index?:
    b. Select S. cerevisiae (CGD) 2011
10. NGS:SAM tools:SAM-BAM
    a. Convert your SAM file to BAM using the defaults
11. Click on the View in Trackster icon in the BAM results window
12. Select a chromosome and see where reads mapped

13. *Changing datasets*, let's look at a visualization I've made and published
    a. Shared Data: Published Visualizations : Cufflinks demo
    b. Explore options