NGS Course Resources

A healthy taste of resources available, specifically for this course - not a comprehensive catalog.

Technology videos

Roche/454
Illumina (Solexa) Genome Analyzer and HiSeq
Life Technologies SOLiD
Pacific Biosciences

Community Resources

- SEQAnwers forum - many NGS sequencing questions answered here
- UCSC Genome Browser - visualize and download NGS data (see more below)
- Galaxy website for online sequencing data analysis
- Broad Institute Integrated Genomcs Viewer (**IGV**) - especially good for ***bam*** files

Getting started with Linux and Perl

- Unix and Perl for Biologists website
    - tutorial primer (pdf)
    - running this tutorial at TACC, on this wiki
- Cheat sheet of useful Unix commands
- A funny SEQAnwers post about biologists starting to analyze NGS data
    - http://seqanswers.com/forums/showthread.php?t=4589

Fastq analysis/manipulation

- Wikipedia FASTQ format page
- FastQC from Babraham Bioinformatics; produces nice quality report for fastq files.
- Cutadapt - An excellent command line tool for adapter sequence removal.
- FASTX Toolkit - Command line tools for fastq analysis and manipulation
- Illumina library construction on GSAF user wiki - useful for contaminent detection or adapter removal.

Alignment

- Comparison of different aligners
    - by Heng Li, developer of BWA and MAQ
    - by Nils Homer, developer of BFAST
- Aligners
    - **bowtie (http:bowtie-bio.sourceforge.net/)** - very fast, not very sensitive
    - **BFAST** wiki & manual - slow and relatively complicated, but tunable sensitivity
    - **bwa** - fast, sensitive and easy to use
    - **bowtie2** - fast, sensitive, configurable, easy to use
- File formats
    - fastq format
    - The **SAM** (Sequence Alignment Map) format specification (pdf)

Alignment analysis

- SAM (Sequence Alignment Map) format specification (pdf)
- ***sam/bam*** tools
    - samtools - ***sam/bam*** conversion, flag filtering, bam sort/index
    - Picard - sam/bam utilities that are read-group aware
    - Translate SAM file flags - type in a decimal number to see which flags are set

- SAMstat - produces detailed graphical statistics for ***sam/bam*** files.
- BEDTools - region overlap, merge, coverage & much more, w/***bed, bam, vcf, gff*** support
    - BEDTools user manual (pdf)

UCSC Genome Browser

- intro on this wiki
- Main UCSC Genome Browser web site
- Beta Test browser site - most up-to-date datasets and features; can be buggy
- File formats - BED format especially is widely used
- Table browser - Browse and download data in different formats

Variant calling

- The 1000 Genomes project - catalog of human genetic variants
- Tools
    - Broad institute **GATK** - complex but powerful; used by 1000 Genomes
- File formats
    - **VCF** (Variant Call Format) v4.0 - developed by 1000 Genomes project
        - list of tools and format specs

Transcriptome analysis

- The **Tuxedo** pipeline: RNAseq with tophat/cufflinks
    - tophat - exon-aware sequence alignment (uses bowtie)
    - cufflinks - transcript assembly, differential expression & regulation
    - RNAseq analysis protocol article in Nature Protocols
        - cufflinks resource bundles for selected organisms (gff annotations, pre-built bowtie references, etc.)

Format converters and miscellaneous tools

- SRA (Sequence Read Archive) from NCBI
    - overview on this wiki
    - SRA search home page
    - SRA Toolkit
        - NCBI documentation
        - SRA toolkit downloads
- Mason program for simulating second-generation sequencing reads.

De novo assembly

- <put something here>

Other courses with online tutorials

- 2012 Next-Gen Sequence Analysis Workshop (Michigan State University) has similar tutorials to our course, but also includes introductions to using the Amazon EC2 where you can "rent" Linux machines (useful if you don't have access to TACC), Python, R, ChIP-Seq, etc.