

# SSC Intro to NGS Bioinformatics Course

May 2013



We will meet in Room 101B of the [Flawn Academic Center \(FAC\) building](#). We STRONGLY encourage you to use the computers provided in the classroom, but you may also bring your personal laptops.

- [May 2013](#)
  - [Your Instructors](#)
  - [Day 1: Linux/TACC Introduction and Read Mapping](#)
    - [Part 1: Linux/TACC Introduction](#)
    - [Part 2: Read Mapping](#)
      - [Enrichment modules \(4:30-5:30\)](#)
      - [Extras](#)
  - [Day 2: Handling Raw and Aligned sequences, and Calling Genome Variants](#)
    - [Part 1. Handling Raw and Aligned sequences](#)
    - [Part 2. Calling Genome Variants](#)
      - [Enrichment module \(12:30-1:30\)](#)
      - [Enrichment modules \(4:30-5:30\)](#)
      - [Extras](#)
  - [Day 3: RNA-seq](#)
    - [Part 1. Introduction to RNA-seq Counting](#)
    - [Part 2. The Tuxedo RNA-seq Pipeline \(Tophat & Cufflinks\)](#)
      - [Enrichment module \(12:30-1:30\)](#)
      - [Enrichment modules \(4:30-5:30\)](#)
      - [Extras](#)
  - [Day 4: Assembly and Annotation](#)
    - [Part 1. Genome Assembly](#)
    - [Part 2. Assembly Annotation](#)
      - [Enrichment module \(12:30-1:30\)](#)
      - [Enrichment module \(4:30-5:30\)](#)
  - [Resources](#)

## [Resources tool list, file formats & more](#)

Link to Etherpad: <https://etherpad.mozilla.org/g2NxIEAFWL>

Use this to post any questions you have about the lessons and tutorials.

## Your Instructors

Name	Initials	Affiliation	Expertise
<a href="#">Scott Hunicke-Smith</a>	SPHS	Director GSAF	Everything, if loosely defined (but especially <code>awk</code> )
<a href="#">Jeff Barrick</a>	JB	Asst. Prof. Biochemistry	Microbes, Perl, C++, Mac, miscellanea
<a href="#">Dhivya Arasappan (<i>in absentia</i>)</a>	DA	GSAF	RNA-seq, transcriptome assembly
<a href="#">Anna Battenhouse</a>	AB	Iyer Lab	Eukaryotes, Bash scripting, UCSC Genome Browser
<a href="#">Daechan Park</a>	DP	Iyer Lab	Eukaryotes, ChIP-seq, Python, Samtools
<a href="#">Nichole Bennett</a>	NB	Parmesan/Singer Labs	Python, R, Unix
<a href="#">Dan Deatherage</a>	DD	Barrick Lab	Unix, Python, NGS Library Prep
<a href="#">Nathan Abell</a>	NA	Iyer Lab	Eukaryotes, RNA-Seq

## [instructor action item list](#)

Day 1a: Scott 1b: Jeff  
Day 2a: Jeff, Daechan, Anna, 2b: Scott  
Day 3a: Jeff 3b: Iyer lab  
Day 4a: Jeff, 4b: Scott




Instructors: meet 9am Monday for final check

Each Part 1/Part 2 section needs to be standardized with:




- \*Learning Objectives
- \*Theory
- \*Workflow diagram (data, toolbox/recipe, exercises)
- \*Tutorial (bulk of time here)
- \*Recap learning objectives
- \*Next steps...

# Day 1: Linux/TACC Introduction and Read Mapping

## Part 1: Linux/TACC Introduction

- [General introduction](#) (SPHS)
  - [Introduction to Bioinformatics Prezi](#)
- [Linux refresher](#) (SPHS)
  - [Step 1: Start somewhere](#)
  - [Step 2: Establishing a profile on Lonestar](#)
  - [Step 3: Editing files](#)
  - [Step 4: Final explanations](#)
  - 
- [Using TACC's Lonestar Cluster](#) (SPHS)
  - [Diagram of Lonestar's directories](#)
  - 
  - [Diagram of running a job on Lonestar](#)
  - 
  - [Using SFTP for file browsing on Linux.](#)
- [Recap and "for further study"](#)

## Part 2: Read Mapping

- [Introduction to next-gen sequencing technologies](#) (JB)
  - [Paper comparing NGS technologies](#) (Liu et al., 2012)
  - [Official Illumina video](#) | [Another Illumina video](#)
  - [Broad Center GA Boot Camp](#)
- [Variant calling workflow diagram:](#)  
-  (JB)
- [Mapping tutorial \(bowtie, BWA, bowtie2\)](#) (JB)

## Enrichment modules (4:30-5:30)

- [Sharing Linux tricks - linux one-liners](#) (SPHS)
- [Working on TACC from your Mac or PC](#) (AB)
  - [Editing files, more detail](#)


## Extras

- [Tutorial - Start diploid mapping for Day 2](#)
- [Running \*Unix & Perl for Biologists\* tutorial at TACC](#)
- [Installing Virtual machine & Linux on Windows](#) (DP)




# Day 2: Handling Raw and Aligned sequences, and Calling Genome Variants

## Part 1. Handling Raw and Aligned sequences

- [Overview and Resources](#)
- [Evaluating your raw sequencing data](#) (AB)
  - [GSAF adaptor and barcode sequence resource](#)
- [Mapped read data evaluation \(SAMtools\)](#) (DP)

 Unknown macro: 'hidden-data'

## Part 2. Calling Genome Variants

-  (JB)
  - [Variant calling workflow diagram:](#)  
- [Variant calling tutorial \(SAMtools\)](#) (JB)
- [Integrative Genomics Viewer \(IGV\) tutorial](#) (JB)


## Enrichment module (12:30-1:30)

- [Shell Scripting](#) (SPHS/AB)

## Enrichment modules (4:30-5:30)


- [Installing Linux tools](#) (JB)

## Extras

- [Calling variants in diploid genomes](#) (SPHS)
- [Introduction to genome variation](#)
- [Variant calling with GATK](#) (SPHS)
- [Annotating variants](#) (SPHS)
-  (JB)
- [Visualize mapped data at UCSC genome browser](#) (AB)
- [Genome variation in mixed samples \(FreeBayes, deepSNV\)](#) (JB)
- [Identifying structural variants \(SVDetect\)](#) (JB)
- [SRA toolkit](#) (AB)

## Day 3: RNA-seq

### Part 1. Introduction to RNA-seq Counting

- 
- [Differential gene expression analysis](#) (JB)

### Part 2. The Tuxedo RNA-seq Pipeline (Tophat & Cufflinks)

- [Differential expression with splice variant analysis](#)

#### Enrichment module (12:30-1:30)

- [Identifying mutations in microbial genomes \(breseq\)](#) (JB)

#### Enrichment modules (4:30-5:30)


- [Start tophat by submitting to lonestar](#) (DA)

## Extras

- [Visualize mapped data at UCSC genome browser](#) (AB)
- [non-coding RNA analysis](#)

## Day 4: Assembly and Annotation

### Part 1. Genome Assembly

- 
- [Genome Assembly Examples](#) (SPHS)
- [Tutorial: Genome Assembly \(velvet\)](#) (SPHS)

### Part 2. Assembly Annotation

- [Genome Annotation \(Glimmer3\)](#) (SPHS)
- [Evaluating & Visualizing assemblies](#) (bacterial, SPHS)

#### Enrichment module (12:30-1:30)

- Office hours: "I want to learn how to install and use this tool called \_\_\_\_\_ that we didn't talk about in class." (JB).

#### Enrichment module (4:30-5:30)

- [Transcriptome assembly & annotation](#)
- [Protein functional classification...](#)
- [Custom Genome Databases](#)

## Resources

- [Scott's list of linux one-liners](#)
- [Example BWA alignment script](#)

- Exercises
- Key take home points
- Resources tool list, file formats & more

As you're getting settled