

---

# **labibi Documentation**

***Release 5.0***

**C. Titus Brown**

August 03, 2014



<b>1</b>	<b>Cheat sheet for starting up an EC2 instance</b>	<b>3</b>
<b>2</b>	<b>Dramatis personae</b>	<b>5</b>
<b>3</b>	<b>Papers and References</b>	<b>7</b>
3.1	Books . . . . .	7
3.2	RNAseq . . . . .	7
3.3	Computing and Data . . . . .	7
<b>4</b>	<b>Links</b>	<b>9</b>
4.1	Humor . . . . .	9
4.2	Resources . . . . .	9
4.3	Blogs . . . . .	9



**Warning:** These documents are not maintained and their instructions may be out of date. However the GED Lab does maintain the [khmer protocols](#) which may cover similar topics. See also the [installation instructions for the current version of the khmer project](#).

This is the schedule for the [2013 MSU NGS course](#), which ran from June 10th to June 20th, 2013. If you're interested in this course in 2014, please see [the 2014 announcement](#).

Day	Schedule
Monday 6/10	<ul style="list-style-type: none"> <li>1:30pm tutorial: <code>day1</code> (Adina)</li> <li>7pm: research presentations</li> </ul>
Tuesday 6/11	<ul style="list-style-type: none"> <li><code>day2</code></li> <li>9:30am: lecture, Welcome! (Titus)</li> <li>10:45am: <code>running-command-line-blast</code> tutorial,</li> <li>1:15pm: <code>short-read-quality-evaluation</code> tutorial,</li> <li>Evening: <i>firepit social</i></li> </ul>
Wednesday 6/12	<ul style="list-style-type: none"> <li><code>day3</code></li> <li>9:30am: lecture, Mapping. (Titus)</li> <li>10:45am: tutorial, <code>bwa-tutorial</code> (Likit)</li> <li>1:15pm: <code>plot-mapping-mismatches</code> tutorial cont'd; also,</li> <li>8pm: General bioinformatics overview (Istvan)</li> </ul>
Thursday 6/13	<ul style="list-style-type: none"> <li>9:15am: lecture, Assembly. (Titus)</li> <li>10:45am: <code>assembling-ecoli-with-velvet</code> tutorial,</li> <li>1:15pm: tutorial, cont'd, evaluating assemblies.</li> <li>Evening: <i>brew pub in Kalamazoo</i>.</li> </ul>
Friday 6/14	<ul style="list-style-type: none"> <li>9:15am: lecture, Intervals (Istvan)</li> <li>10:45am: <code>interval-analysis-tutorial</code> tutorial, (Istvan)</li> <li>1:15pm: lecture/tutorial, Statistics (Ian)</li> <li>8 Tutorial: <code>teach-me-intervals</code> (Istvan)</li> </ul>
Saturday 6/15	<ul style="list-style-type: none"> <li>9:15am: lecture, Pipelines and Automation (Titus)</li> <li>10:45am: tutorial: Shell scripts and pipelines.</li> <li>1:15pm: tutorial, R (<code>text   code</code>) (Josh)</li> <li>Evening: BBQ/dinner.</li> </ul>
Sunday 6/16	Day of rest. Brunch in the morning; takeout dinner in evening.
Monday 6/17	<ul style="list-style-type: none"> <li>9:15am: lecture, mRNAseq I (Ian Dworkin)</li> <li>Tutorials 10:45am, 1:15pm as usual; topics: <ol style="list-style-type: none"> <li><code>rnaseq_bwa</code></li> <li><code>rnaseq_tophat</code></li> <li><code>DGE_analysis_with_MISO_cuffdiff</code></li> <li><code>transcriptome_de_novo_assembly</code></li> </ol> </li> <li>7:30pm: <code>git-koans</code> (Titus)</li> <li>8-9pm: look busy</li> <li>9pm: <code>firepit</code></li> </ul>
Tuesday 6/18	<ul style="list-style-type: none"> <li>9:15am: lecture, mRNAseq II (Ian Dworkin)</li> </ul>
2	<ul style="list-style-type: none"> <li>Impromptu tutorial: <code>seqtk_tools</code></li> <li>Tutorials 10:45am, 1:15pm as usual; mRNAseq, continued.</li> <li><code>rnaseq_bwa_counting</code> (edgeR code)</li> <li>8:30pm: Single-cell mRNAseq (Erich Schwarz)</li> </ul>

---

## Cheat sheet for starting up an EC2 instance

---

- use Amazon Machine Instance “ami-c17ec8a8”;
- m1.large or larger;
- make sure you are in the US East zone (Virginia) – see upper right;
- make sure the security group you use has SSH and HTTPS enabled for inbound;



---

## Dramatis personae

---

Instructors:

- Istvan Albert
- C Titus Brown
- Ian Dworkin

TAs:

- Amanda Charbonneau
- Michael Crusoe
- Tristan De Buysscher
- Joshua Herr
- Elijah Lowe
- Likit Preeyanon

Lecturers:

- Nick Beckloff
- Rayan Chaikhi
- Chris Chandler
- Adina Chuang Howe
- Erich Schwarz



---

## Papers and References

---

### 3.1 Books

- [Practical Computing for Biologists](#)

This is a highly recommended book for people looking for a systematic presentation on shell scripting, programming, UNIX, etc.

### 3.2 RNAseq

- [Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks](#), Trapnell et al., Nat. Protocols.  
One paper that outlines a pipeline with the tophat, cufflinks, cuffdiffs and some associated R scripts.
- [Statistical design and analysis of RNA sequencing data.](#), Auer and Doerge, Genetics, 2010.
- [A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in \*Saccharomyces cerevisiae\*](#). Nookaew et al., Nucleic Acids Res. 2012.
- [Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments](#) Vijay et al., 2012.
- [Computational methods for transcriptome annotation and quantification using RNA-seq](#), Garber et al., Nat. Methods, 2011.
- [Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.](#), Bullard et al., 2010.
- [A comparison of methods for differential expression analysis of RNA-seq data](#), Sonesson and Delorenzi, BMC Bioinformatics, 2013.
- [Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples.](#), Wagner et al., Theory Biosci, 2012. Also see [this blog post](#) explaining the paper in detail.

### 3.3 Computing and Data

- [A Quick Guide to Organizing Computational Biology Projects](#), Noble, PLoS Comp Biology, 2009.

- Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results, Wicherts et al., PLoS One, 2011.
- Got replicability?, McCullough, Economics in Practice, 2007.

Also see this great pair of blog posts on [organizing projects](#) and [research workflow](#).

## **4.1 Humor**

- [Data Sharing and Management Snafu in 3 Short Acts](#)

## **4.2 Resources**

- [Biostar](#)  
A high quality question & answer Web site.
- [SEQanswers](#)  
A discussion and information site for next-generation sequencing.
- [Software Carpentry lessons](#)  
A large number of open and reusable tutorials on the shell, programming, version control, etc.

## **4.3 Blogs**

- <http://www.genomesunzipped.org/>  
Genomes Unzipped.
- <http://ivory.idyll.org/blog/>  
Titus's blog.
- <http://bcbio.wordpress.com/>  
Blue Collar Bioinformatics
- <http://massgenomics.org/>  
Mass Genomics
- <http://blog.nextgenetics.net/>  
Next Genetics
- <http://gettinggeneticsdone.blogspot.com/>  
Getting Genetics Done

- <http://omicsomics.blogspot.com/>  
Omics! Omics!