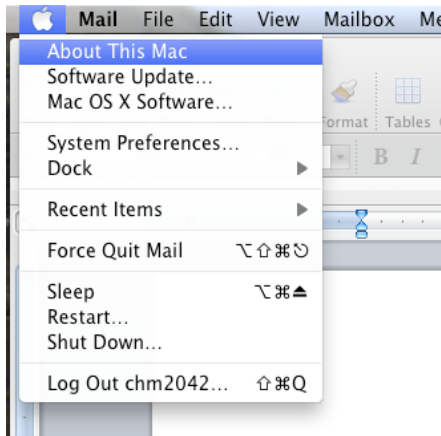


For tomorrow's class, you need to have five things done and on your Desktop

1. **XCode Installed**
2. **Human Genome (hg19) downloaded and cleaned**
3. **BWA downloaded and compiled**
4. **Build Human Genome BWT Index**
5. **Download 1000Genomes Data**

## 1. Xcode

Determine the type of OS X operating system that you have:



Get the appropriate version of Xcode.

<http://physiology.med.cornell.edu/faculty/mason/lab/data/xcode>

Xcode 2.5 for Tiger (10.4.X)

Xcode 3.1.4 for Leopard (10.5.X)

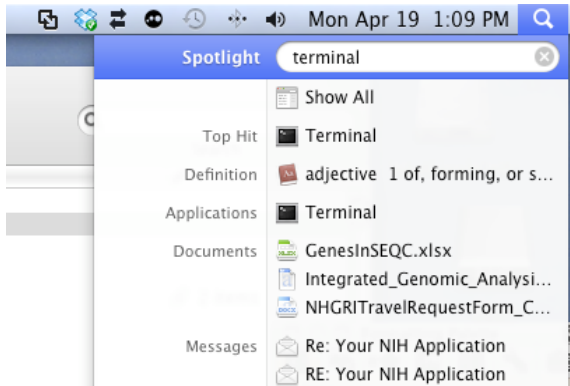
Xcode 3.2.2 for Snow Leopard (10.6.X)

To install, just double-click the .dmg file and follow the instructions.

Note: These are normally found at <http://connect.apple.com> where you can sign up for a free account to develop tools on your Mac.

## 2. Human Genome (hg19)

Open the terminal application:



Move to your Desktop to begin working there:

```
]cd Desktop
```

Make a new folder on your desktop called “genomes”

```
]mkdir genomes
```

Download the human genome to your Desktop:

Go to: <http://genome.ucsc.edu/>

Go to Downloads → Human → Full Data Set → [chromFa.tar.gz](http://hgdownload.soe.ucsc.edu/genome/hg19/chromFa.tar.gz)

Move the chromFa.tar.gz file into the genomes directory

```
]mv chromFa.tar.gz genomes/
```

Go into the genomes directory:

```
]cd genomes
```

Gunzip the tarball:

```
]gunzip chromFa.tar.gz
```

```
]tar -xvf chromFa.tar
```

Remove the haplotype, unmapped, and random chromosomes

```
]rm *random*
```

```
]rm *Un*
```

```
]rm *hap*
```

Concatenate the different chromosomes into one chromosome. The “>” sign indicates the name of the new file that will be created from the concatenation.

```
]cat chr*.fa >hg19.fa
```

### 3. BWA

Download the aligner program to your Desktop. The file is here:

<http://sourceforge.net/projects/bio-bwa/files/bwa-0.5.7.tar.bz2/download>

Unzip the file:

```
]bunzip2 bwa-0.5.7.tar.bz2
```

Untar the file:

```
]tar -xvf bwa-0.5.7.tar
```

Go into the directory with the program, then compile the Program:

```
] cd bwa-0.5.7
```

```
] make
```

#### **4. Make your Burrows-Wheeler Transformed (BWT) Index for the Genome**

You should still be inside the bwa directory. From here, we will run the bwa program to build the index on the hg19.fa file. To run a program, you must first type “./” before the name of the program, as seen below. This should take 1-3 hours to run.

```
] ./bwa index -a bwtsv ../genomes/hg19.fa
```

You should see this:

```

[BWTIncConstructFromPacked] 260 iterations done. 2599999999 characters processed.
[BWTIncConstructFromPacked] 270 iterations done. 2699999999 characters processed.
[BWTIncConstructFromPacked] 280 iterations done. 2799999999 characters processed.
[BWTIncConstructFromPacked] 290 iterations done. 2899999999 characters processed.
[BWTIncConstructFromPacked] 300 iterations done. 2999999999 characters processed.
[BWTIncConstructFromPacked] 310 iterations done. 3092429455 characters processed.
[bwt_gen] Finished constructing BWT in 311 iterations.
[bwa_index] 2631.92 seconds elapse.
[bwa_index] Construct BWT for the reverse packed sequence...
[BWTIncConstructFromPacked] 10 iterations done. 999999999 characters processed.
[BWTIncConstructFromPacked] 20 iterations done. 1999999999 characters processed.
[BWTIncConstructFromPacked] 30 iterations done. 2999999999 characters processed.
[BWTIncConstructFromPacked] 40 iterations done. 3999999999 characters processed.
[BWTIncConstructFromPacked] 50 iterations done. 4999999999 characters processed.
[BWTIncConstructFromPacked] 60 iterations done. 5999999999 characters processed.
[BWTIncConstructFromPacked] 70 iterations done. 6999999999 characters processed.
[BWTIncConstructFromPacked] 80 iterations done. 7999999999 characters processed.
[BWTIncConstructFromPacked] 90 iterations done. 8999999999 characters processed.
[BWTIncConstructFromPacked] 100 iterations done. 9999999999 characters processed.
[BWTIncConstructFromPacked] 110 iterations done. 10999999999 characters processed.
[BWTIncConstructFromPacked] 120 iterations done. 11999999999 characters processed.
[BWTIncConstructFromPacked] 130 iterations done. 12999999999 characters processed.
[BWTIncConstructFromPacked] 140 iterations done. 13999999999 characters processed.
[BWTIncConstructFromPacked] 150 iterations done. 14999999999 characters processed.
[BWTIncConstructFromPacked] 160 iterations done. 15999999999 characters processed.
[BWTIncConstructFromPacked] 170 iterations done. 16999999999 characters processed.
[BWTIncConstructFromPacked] 180 iterations done. 17999999999 characters processed.
[BWTIncConstructFromPacked] 190 iterations done. 18999999999 characters processed.
[BWTIncConstructFromPacked] 200 iterations done. 19999999999 characters processed.
[BWTIncConstructFromPacked] 210 iterations done. 20999999999 characters processed.
[BWTIncConstructFromPacked] 220 iterations done. 21999999999 characters processed.
[BWTIncConstructFromPacked] 230 iterations done. 22999999999 characters processed.
[BWTIncConstructFromPacked] 240 iterations done. 23999999999 characters processed.
[BWTIncConstructFromPacked] 250 iterations done. 24999999999 characters processed.
[BWTIncConstructFromPacked] 260 iterations done. 25999999999 characters processed.
[BWTIncConstructFromPacked] 270 iterations done. 26999999999 characters processed.
[BWTIncConstructFromPacked] 280 iterations done. 27999999999 characters processed.
[BWTIncConstructFromPacked] 290 iterations done. 28999999999 characters processed.
[BWTIncConstructFromPacked] 300 iterations done. 29999999999 characters processed.
[BWTIncConstructFromPacked] 310 iterations done. 3092429455 characters processed.
[bwt_gen] Finished constructing BWT in 311 iterations.
[bwa_index] 2652.96 seconds elapse.
[bwa_index] Update BWT... 16.42 sec
[bwa_index] Update reverse BWT... 16.63 sec
[bwa_index] Construct SA from BWT and Occ... 787.74 sec
[bwa_index] Construct SA from reverse BWT and Occ... 786.83 sec
[ngs19@capek genomes]$ ls

```

## 5. Download 1000genomes data

Move back up to the desktop

```
] cd ..
```

Make a new directory on the Desktop for these data called 1KG:

```
] mkdir 1KG
```

The FTP site for these files are listed here:

[ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data/NA06985/sequence\\_read/ERR001014.1.filt.fastq.gz](ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data/NA06985/sequence_read/ERR001014.1.filt.fastq.gz)

[ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data/NA06985/sequence\\_read/ERR001014\\_2.filt.fastq.gz](ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/data/NA06985/sequence_read/ERR001014_2.filt.fastq.gz)

Once they have downloaded, move the files to the 1KG folder.

```
] mv *.gz 1KG/
```

Then, uncompress the files:

```
] gunzip *.gz
```