

MSBI 32400 – LAB 3

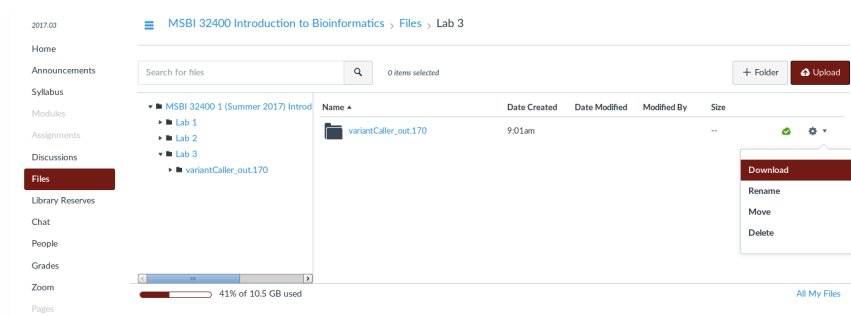
LARRY HELSETH PHD &
JASON EDELSTEIN

July 5, 2017

Replace some missing files first

2

- Sign into Canvas from your VM and download the variantCaller_out.170 folder from Lab 3 files:



MSBI 32400 Lab 3 7/5/17

Extract the folder into /data

3

- cd /data then unzip ~/Downloads/Lab_3_export.zip

```

student@MSBI32400Lab:/data
File Edit View Search Terminal Help
[student@MSBI32400Lab data]$ unzip -h
UnZip 6.00 of 20 April 2009, by Info-ZIP. Maintained by C. Spieler. Send
bug reports using http://www.info-zip.org/zip-bug.html; see README for details.

Usage: unzip [-Z] [-opts[modifiers]] file[.zip] [list] [-x xlist] [-d exdir]
Default action is to extract files in list, except those in xlist, to exdir;
file[.zip] may be a wildcard. -Z => ZipInfo mode ("unzip -Z" for usage).

-p extract files to pipe, no messages      -l list files (short format)
-f freshen existing files, create none     -t test compressed archive data
-u update files, create if necessary        -z display archive comment only
-v list verbosely/show version info        -T timestamp archive to latest
-x exclude files that follow (in xlist)    -d extract files into exdir

modifiers:
-n never overwrite existing files          -q quiet mode (-qq => quieter)
-o overwrite files WITHOUT prompting       -a auto-convert any text files
-j junk paths (do not make directories)    -aa treat ALL files as text
-U use escapes for all non-ASCII Unicode   -UI ignore any Unicode fields
-C match filenames case-insensitively     -L make (some) names lowercase
-K restore UID/GID info                   -V retain VMS version numbers
-k keep setuid/setgid/sticky permissions  -M pipe through "more" pager
-O CHARSET specify a character encoding for DOS, Windows and OS/2 archives
-I CHARSET specify a character encoding for UNIX and other archives

See "unzip -hh" or unzip.txt for more help. Examples:
unzip data1 -x joe => extract all files except joe from zipfile data1.zip
unzip -p foo | more => send contents of foo.zip via pipe into program more
unzip -fo foo README => quietly replace existing README if archive file newer
[student@MSBI32400Lab data]$ unzip -l ~/Downloads/Lab_3_export.zip
Archive:  /home/student/Downloads/Lab_3_export.zip
Length Date   Time    Name
-----
0 07-04-2017 14:12 variantCaller.out.17b/
0 07-04-2017 14:12 variantCaller.out.17b/IonXpress.003/
174185 07-04-2017 14:12 variantCaller.out.17b/IonXpress.003/TSVC_variants.bed
1020855 07-04-2017 14:12 variantCaller.out.17b/IonXpress.003/TSVC_variants.vcf
1698896 07-04-2017 14:12 variantCaller.out.17b/IonXpress.003/IonXpress.003_rawlib_processed.bam.bai
5844398 07-04-2017 14:12 variantCaller.out.17b/IonXpress.003/IonXpress.003_rawlib_processed.bam
-----
61343434 0 files
[student@MSBI32400Lab data]$

```

MSBI 32400 Lab 3 7/5/17

Outline

4

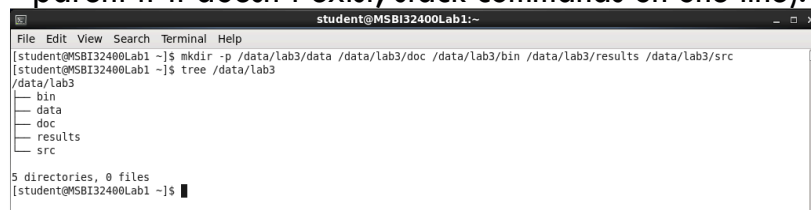
- First we'll download some files from Pevsner's website
- Use some text commands to trim files
- Generate BED files from UCSC
- Then install bedtools on the VM
- Use bedtools to sort and merge bed file
- Use samtools to work with BAM files

MSBI 32400 Lab 3 7/5/17

Pevsner's book companion site

5

- Create your project directories on the VM ('-p' creates parent if it doesn't exist; stack commands on one line):



```
student@MSBI32400Lab1:~$ mkdir -p /data/lab3/data /data/lab3/doc /data/lab3/bin /data/lab3/results /data/lab3/src
student@MSBI32400Lab1:~$ tree /data/lab3
/data/lab3
├── bin
├── data
├── doc
├── results
└── src

5 directories, 0 files
student@MSBI32400Lab1:~$
```

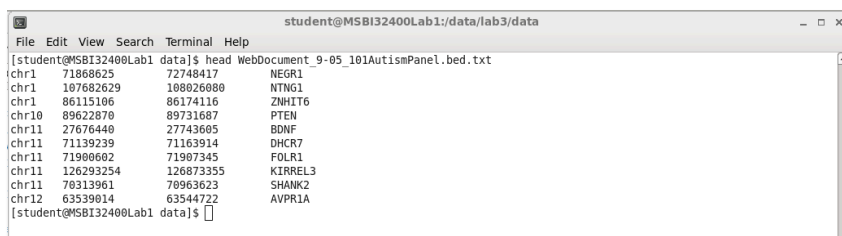
- From the VM, go to <http://www.bioinfbook.org/php/C9E3k> and download 9.5 and both 9.7 files into /data/lab3/data
 - May need to "mv" files from ~/Download to /data/lab3/data (hint- "mv <path to files> ." moves to current location)

MSBI 32400 Lab 3 7/5/17

Let's get the genes from file 9.5

6

- Head the document to see structure



```
student@MSBI32400Lab1:/data/lab3/data$ head WebDocument_9-05_101AutismPanel.bed.txt
chr1 71868625 72748417 NEGR1
chr1 107682629 108026080 NTNG1
chr1 86115106 86174116 ZNHIT6
chr10 89622870 89731687 PTEN
chr11 27676440 27743605 BDNF
chr11 71139239 71163914 DHCR7
chr11 71900602 71907345 FOLR1
chr11 126293254 126873355 KIRREL3
chr11 70313961 70963623 SHANK2
chr12 63539014 63544722 AVPR1A
student@MSBI32400Lab1:/data/lab3/data$
```

- Use the "cut" command to extract the 4th column (gene symbols) and send it to a text file:
 - cut -f4 WebDocument_9-05_101AutismPanel.bed.txt > genelist.txt

MSBI 32400 Lab 3 7/5/17

```

student@MSB132400Lab1:/data/lab3/data
File Edit View Search Terminal Help
[student@MSB132400Lab1 data]$ cut -f4 WebDocument_9-05_101AutismPanel.bed.txt | more
WDR1
NTNG1
ZNF116
PTEN
BDNF
DCHCR7
POLR1
KIRREL3
SHANK2
AVPR1A
CACNA1C
GRIN2B
PTPNI1
SOD5
PCDH9
CHD8
FOXG1
GABRB3
SNRPN
UBE3A
ANKRD11
CREBBP
KCTD13
RFXO1
TSC2
PAFAH1B1
RAI1
SLC6A4
SMG6
C18orf1
KATNAL2
TCF4
PINK1
ZNF507
CNTNAP5
DPP10
MBD5
NRXN1
SATB2
SCN1A
SCN2A
SPAST
ZEB2
--More--

```

MSB1 32400 Lab 3 7/5/17

Now, go to UCSC and upload genes

- <http://genome.ucsc.edu/cgi-bin/hgTables>
- Change Assembly to hg19
- “Upload list” in Identifiers and browse to your new **genelist.txt** & Submit
- Change output type to BED
- Give a filename and click **get output**

assembly: Dec. 2013 (GRCh38/hg38)
 Dec. 2013 (GRCh38/hg38)
 Feb. 2009 (GRCh37/hg19)
 Mar. 2006 (NCBI36/hg18)
 May 2004 (NCBI35/hg17)
 July 2003 (NCBI34/hg16)
 ma
 lookup

output format: all fields from selected table
 output file: all fields from selected table
 selected fields from primary and related tables
 file type returned: sequence
 GTF - gene transfer format
 CDS FASTA alignment from multiple alignment
 BED - browser extensible data
 custom track
 To reset all user hyperlinks to Genome Browser

output format: BED - browser extensible data
 output file: 101AutismGenelistExons.bed (leave blank to keep output in
 file type returned: ☒ plain text ☐ gzip compressed
 get output summary/statistics
 To reset all user cart settings (including custom tracks), [click here](#).

MSB1 32400 Lab 3 7/5/17

Change output type to exon + 10bp

9

Output knownGene as BED

☐ Include custom track header:
 name=
 description=
 visibility=
 url=

Create one BED record per:

- ☐ Whole Gene
- ☐ Upstream by bases
- ☒ Exons plus bases at each end
- ☐ Introns plus bases at each end
- ☐ 5' UTR Exons
- ☐ Coding Exons
- ☐ 3' UTR Exons
- ☐ Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream past the edge of the chromosome.

MSBI 32400 Lab 3 7/5/17

Move the new BED file to lab3/data

10

```

student@MSBI32400Lab1:/data/lab3/data
File Edit View Search Terminal Help
[student@MSBI32400Lab1 data]$ ls -ltr
total 336832
-rw-rw-r-- 1 student student 3126 Jan 15 14:07 WebDocument_9-05_101AutismPanel.bed.txt
-rw-rw-r-- 1 student student 3131856 Jan 15 14:08 WebDocument_9-7_mysample1.bai
-rw-rw-r-- 1 student student 341767805 Jan 15 14:26 WebDocument_9-7_mysample1.bam
-rw-rw-r-- 1 student student 715 Jan 15 14:48 genelists.txt
[student@MSBI32400Lab1 data]$ ls -ltr ~/Downloads/
total 1452
-rw-rw-r-- 1 student student 711132 Dec 22 21:03 setuptools-32.2.0.zip
-rw-rw-r-- 1 student student 3102 Dec 22 21:29 README_install_notes.md
-rw-rw-r-- 1 student student 4116 Dec 22 21:29 README_install_notes.html
-rw-rw-r-- 1 student student 186570 Dec 27 15:01 crossroads_bioinformatics_by_tuxedopengu.jpg
-rw-rw-r-- 1 student student 572374 Jan 15 14:52 101AutismGenelistExons.bed
[student@MSBI32400Lab1 data]$ mv ~/Downloads/101AutismGenelistExons.bed .
[student@MSBI32400Lab1 data]$

```

MSBI 32400 Lab 3 7/5/17

Open IGV then add BAM + BEDs

11

- File/Open in IGV then browse to /data/lab3/data
- Go to a gene like MEFC2 and compare BED coverages MEF2C***
 - ▣ Pevsner's file is full gene, ours is just exons + 10 bp
 - ▣ Zoom in to compare the coverage
 - ▣ If you hover over one of the exons in BED file you'll see multiple transcripts
 - Need bedtools to clean that up!

MSBI 32400 Lab 3 7/5/17

BED file has overlapping transcripts

12

```

student@MSBI32400Lab3:~$ cat MEF2C.bed
chr1 215740711 215741121 NM_001319294 exon 8 10 chr1 215740722 f 0 +
chr1 215747118 215747192 NM_001319294 exon 1 10 chr1 215747129 f 0 +
chr1 215747405 215747471 NM_001319294 exon 2 10 chr1 215747416 f 0 +
chr1 215748233 215749327 NM_001319294 exon 3 10 chr1 215749244 f 0 +
chr1 215751802 215751881 NM_001319294 exon 4 10 chr1 215751813 f 0 +
chr1 215753332 215753434 NM_001319294 exon 5 10 chr1 215753244 f 0 +
chr1 215753332 215752400 NM_001319294 exon 6 10 chr1 215752343 f 0 +
chr1 215753241 215753352 NM_001319294 exon 7 10 chr1 215753252 f 0 +
chr1 215759827 215760838 NM_001319294 exon 8 10 chr1 215759838 f 0 +
chr1 215768687 215768823 NM_001319294 exon 9 10 chr1 215768698 f 0 +
chr1 215775198 215775306 NM_001319294 exon 10 10 chr1 215775209 f 0 +
chr1 215775416 215775553 NM_001319294 exon 11 10 chr1 215775427 f 0 +
chr1 215777463 215777654 NM_001319294 exon 12 10 chr1 215777474 f 0 +
chr1 215781348 215781524 NM_001319294 exon 13 10 chr1 215781359 f 0 +
chr1 215785156 215785273 NM_001319294 exon 14 10 chr1 215785167 f 0 +
chr1 215792217 215792422 NM_001319294 exon 15 10 chr1 215792228 f 0 +
chr1 215792498 215792643 NM_001319294 exon 16 10 chr1 215792501 f 0 +
chr1 215793088 215795173 NM_001319294 exon 17 10 chr1 215793099 f 0 +
chr1 215740711 215741121 NM_001319295 exon 8 10 chr1 215740722 f 0 +
chr1 215747118 215747192 NM_001319295 exon 1 10 chr1 215747129 f 0 +
chr1 215747405 215747471 NM_001319295 exon 2 10 chr1 215747416 f 0 +
chr1 215748233 215749327 NM_001319295 exon 3 10 chr1 215749244 f 0 +
chr1 215751802 215751881 NM_001319295 exon 4 10 chr1 215751813 f 0 +
chr1 215753332 215752400 NM_001319295 exon 5 10 chr1 215753244 f 0 +
chr1 215753241 215753352 NM_001319295 exon 6 10 chr1 215753252 f 0 +
chr1 215759827 215760838 NM_001319295 exon 7 10 chr1 215759838 f 0 +
chr1 215768687 215768823 NM_001319295 exon 8 10 chr1 215768698 f 0 +
chr1 215775198 215775306 NM_001319295 exon 9 10 chr1 215775209 f 0 +
chr1 215775416 215775553 NM_001319295 exon 10 10 chr1 215775427 f 0 +
chr1 215777463 215777654 NM_001319295 exon 11 10 chr1 215777474 f 0 +
chr1 215781348 215781524 NM_001319295 exon 12 10 chr1 215781359 f 0 +
chr1 215785156 215785273 NM_001319295 exon 13 10 chr1 215785167 f 0 +
chr1 215792217 215792422 NM_001319295 exon 14 10 chr1 215792228 f 0 +
chr1 215792498 215792643 NM_001319295 exon 15 10 chr1 215792501 f 0 +
chr1 215793088 215795173 NM_001319295 exon 16 10 chr1 215793099 f 0 +
chr1 215740711 215741121 NM_016121 exon 8 10 chr1 215740722 f 0 +
chr1 215747118 215747192 NM_016121 exon 1 10 chr1 215747129 f 0 +
chr1 215747405 215747471 NM_016121 exon 2 10 chr1 215747416 f 0 +
chr1 215748233 215749327 NM_016121 exon 3 10 chr1 215749244 f 0 +
chr1 215751802 215751881 NM_016121 exon 4 10 chr1 215751813 f 0 +
chr1 215753332 215752400 NM_016121 exon 5 10 chr1 215753244 f 0 +
chr1 215753241 215753352 NM_016121 exon 6 10 chr1 215753252 f 0 +
chr1 215759827 215760838 NM_016121 exon 7 10 chr1 215759838 f 0 +
chr1 215768687 215768823 NM_016121 exon 8 10 chr1 215768698 f 0 +
chr1 215775198 215775306 NM_016121 exon 9 10 chr1 215775209 f 0 +
chr1 215775416 215775553 NM_016121 exon 10 10 chr1 215775427 f 0 +
chr1 215777463 215777654 NM_016121 exon 11 10 chr1 215777474 f 0 +
chr1 215781348 215781524 NM_016121 exon 12 10 chr1 215781359 f 0 +
chr1 215785156 215785273 NM_016121 exon 13 10 chr1 215785167 f 0 +
chr1 215792217 215792422 NM_016121 exon 14 10 chr1 215792228 f 0 +
chr1 215792498 215792643 NM_016121 exon 15 10 chr1 215792501 f 0 +
chr1 215793088 215795173 NM_016121 exon 16 10 chr1 215793099 f 0 +

```

MSBI 32400 Lab 3 7/5/17

Installing bedtools (as non-root)

13

- Modified from <http://bedtools.readthedocs.io/en/latest/content/installation.html> (see also Pevsner pg. 413)
- Let's use the **wget** command from ~/Downloads folder:
 - `wget https://github.com/arq5x/bedtools2/releases/download/v2.26.0/bedtools-2.26.0.tar.gz`
- Next extract using **tar** command
 - `tar -zxvf bedtools-2.26.0.tar.gz`
- `cd` to new bedtools2 directory
- Type **"make"** and wait while code is compiled, then copy the entire bin subdirectory to your home directory
 - `cp -rp bin/ ~`
- Add your new bin directory to your PATH
 - `export PATH=~/.bin:$PATH` ("echo \$PATH" to verify)

MSBI 32400 Lab 3 7/5/17

Let's clean up the BED file

14

- `bedtools sort -i 101AutismGenelistExons.bed > 101AutismGenelistExons_sort.bed`
- `bedtools merge -c 4 -o collapse -i 101AutismGenelistExons_sort.bed > 101AutismGenelistExons_sort_merged.bed`

```

student@MSBI32400Lab1:~/data/lab3/data
File Edit View Search Terminal Help
[student@MSBI32400Lab1 data]$ head 101AutismGenelistExons_sort.bed
chr1 71868614 71873263 uc010oqs.2 exon 8 chr1 71868625 r 0 -
chr1 71868614 71873263 uc010dfv.3 exon 8 chr1 71868625 r 0 -
chr1 71868614 71873263 uc010dfv.3 exon 8 chr1 71868625 r 0 -
chr1 72058489 72058661 uc010oqs.2 exon 1 chr1 72058500 r 0 -
chr1 72058489 72058661 uc010dfv.3 exon 1 chr1 72058500 r 0 -
chr1 72058489 72058661 uc010dfv.3 exon 1 chr1 72058500 r 0 -
chr1 72076698 72076839 uc010oqs.2 exon 2 chr1 72076709 r 0 -
chr1 72076698 72076839 uc010dfv.3 exon 2 chr1 72076709 r 0 -
chr1 72076698 72076839 uc010dfv.3 exon 2 chr1 72076709 r 0 -
chr1 72163689 72163832 uc010dfv.3 exon 3 chr1 72163691 r 0 -
[student@MSBI32400Lab1 data]$ bedtools merge -c 4 -o collapse -i 101AutismGenelistExons_sort.bed > 101AutismGenelistExons_sort_merged.bed
[student@MSBI32400Lab1 data]$ head 101AutismGenelistExons_sort_merged.bed
chr1 71868614 71873263 uc010oqs.2 exon 8 chr1 71868625 r,uc010dfv.3 exon 8 chr1 71868625 r 0 -
chr1 72058489 72058661 uc010oqs.2 exon 1 chr1 72058500 r,uc010dfv.3 exon 1 chr1 72058500 r 0 -
chr1 72076698 72076839 uc010oqs.2 exon 2 chr1 72076709 r,uc010dfv.3 exon 2 chr1 72076709 r 0 -
chr1 72163689 72163832 uc010dfv.3 exon 3 chr1 72163691 r,uc010dfv.3 exon 3 chr1 72163691 r 0 -
chr1 72241844 72241990 uc010dfv.3 exon 4 chr1 72241855 r,uc010dfv.3 exon 4 chr1 72241855 r 0 -
chr1 72480751 72480862 uc010dfv.3 exon 5 chr1 72480762 r,uc010oqs.2 exon 5 chr1 72480762 r 0 -
chr1 72566417 72566624 uc010dfv.3 exon 6 chr1 72566428 r 0 -
chr1 72747791 72748415 uc010dfv.3 exon 6 chr1 72748082 r,uc010oqs.2 exon 5 chr1 72748082 r 0 -
chr1 86112595 86112791 uc010dfv.3 exon 8 chr1 86113106 r,uc010osc.2 exon 8 chr1 86113106 r 0 -
chr1 86123519 86123664 uc010dfv.3 exon 10 chr1 86123538 r,uc010osc.2 exon 10 chr1 86123538 r 0 -
[student@MSBI32400Lab1 data]$

```

MSBI 32400 Lab 3 7/5/17

Load merged BED in IGV

15

- Point to an exon and compare information with that from original UCSC downloaded BED file
- Advantage of merging is single track name
- Disadvantage is can't expand track to see all transcripts
- Your approach should vary based upon needs

MSBI 32400 Lab 3 7/5/17

SAMtools and other toys

16

- Copy /data/bds-files/chapter-11-alignment/NA12891_CEU_sample.vcf.gz to lab3/data folder
- Look at the file structure using **zcat** (pipe to more)
- Extract the text VCF using **gunzip**
NA12891_CEU_sample.vcf.gz
- Count lines without comments using **grep**
 - ▣ `grep -v "^#" | wc -l`
 - ▣ Can also use: `zcat NA12891_CEU_sample.vcf.gz | grep -v '^#' | wc -l` if you didn't gunzip file
- Look for the samtoolsCommand & reference file in VCF header (it tells how VCF was generated)

MSBI 32400 Lab 3 7/5/17

SAMtools

17

- Use samtools to check the sample name:
 - ▣ `samtools view -H WebDocument_9-7_mysample1.bam | grep '@RG'`
- Use samtools to extract the FASTQ and SAM from BAM file (3620774 reads)
 - ▣ `samtools fastq WebDocument_9-7_mysample1.bam > WebDocument_9-7_mysample1.fastq`
- Use samtools to view SAM file
 - ▣ `samtools view -h WebDocument_9-7_mysample1.bam > WebDocument_9-7_mysample1.sam`
- Compare file sizes and send to Jason then delete .fastq to conserve space

MSBI 32400 Lab 3 7/5/17

Monitor what's going on

18

- Open a second terminal and use **top**
- Use **time** before your command to time the process
- My BAM → SAM took ~40 seconds.
- My SAM → sorted BAM took ~9 1/2 minutes
- Can use second terminal to monitor intermediate file sizes
 - ▣ Click File/Open Terminal to open second terminal in same directory (can also Open Tab but harder to switch between using Alt-Tab)
 - ▣ Hint-Can open Terminal from File Explorer

MSBI 32400 Lab 3 7/5/17

19

The screenshot shows a terminal window titled "student@MSBI32400Lab1:~" with the following output:

```
top - 08:05:48 up 1 day, 17:01, 3 users, load average: 1.67, 0.78, 0.29
Tasks: 155 total, 2 running, 137 sleeping, 0 stopped, 0 zombie
Cpu(s): 96.8%us, 3.2%sy, 0.0%ni, 0.0%id, 0.0%wa, 0.0%hi, 0.0%st
Mem: 392256k total, 357648k used, 35193k free, 41506k buffers
Swap: 2097148k total, 0k used, 2097148k free, 218728k cached
```

Below the statistics is a table of processes:

PID	USER	PR	NI	VIRT	RES	SHR	S	TCPU	MEM	TIME+	COMMAND
5308	student	20	0	10788	1412	804	R	93.4	0.0	2:26.37	santools
5301	student	20	0	9206	6308	892	S	7.3	21.9	0:35.78	santools
5326	student	20	0	15832	1268	936	R	2.0	0.0	0:00.26	top
36	root	20	19	0	0	0	S	0.7	0.0	0:05.00	khugapaged
206	root	20	0	0	0	0	S	0.3	0.0	0:49.09	scsi_eh_1
2296	student	20	0	2046	860	440	S	0.3	0.0	4:07.97	VBoxClient
1	root	20	0	3546	928	492	S	0.8	0.0	0:03.81	init
2	root	20	0	0	0	0	S	0.0	0.0	0:00.01	kthreadd
3	root	RT	0	0	0	0	S	0.0	0.0	0:00.00	migration/0
4	root	20	0	0	0	0	S	0.0	0.0	0:01.94	ksoftirqd/0
5	root	RT	0	0	0	0	S	0.0	0.0	0:00.00	stopper/0
6	root	RT	0	0	0	0	S	0.0	0.0	0:01.77	watchdog/0
7	root	20	0	0	0	0	S	0.0	0.0	0:56.85	events/0
8	root	20	0	0	0	0	S	0.0	0.0	0:00.00	events/0
9	root	20	0	0	0	0	S	0.0	0.0	0:00.00	events/long/0
10	root	20	0	0	0	0	S	0.0	0.0	0:00.00	events/power_of
11	root	20	0	0	0	0	S	0.0	0.0	0:00.00	cgroup

MSBI 32400 Lab 3 7/5/17

20

The screenshot shows a terminal window titled "student@MSBI32400Lab1:~" with the following output:

```
(student@MSBI32400Lab1 data)$ time santools view -b5 WebDocument_9-7_mysample1.sam | santools sort -o WebDocument_9-7_mysample1_file_sorted
[am_sort_core] merging from 3 files...
```

Below the command output is a list of files and their permissions:

```
total 2324132
-rw-rw-r-- 1 student student 3126 Jan 15 14:07 WebDocument_9-05_101AutismPanel.bed.txt
-rw-rw-r-- 1 student student 3131056 Jan 15 14:08 WebDocument_9-7_mysample1.bal
-rw-rw-r-- 1 student student 341707895 Jan 15 14:26 WebDocument_9-7_mysample1.bam
-rw-rw-r-- 1 student student 715 Jan 15 14:48 genelists.txt
-rw-rw-r-- 1 student student 572374 Jan 15 14:52 101AutismGenelists.txt
-rw-rw-r-- 1 student student 572374 Jan 15 15:54 101AutismGenelists.txt.sorted.bed
-rw-rw-r-- 1 student student 301490 Jan 15 15:58 101AutismGenelists.txt.sorted.merged.bed
-rw-rw-r-- 1 student student 3404 Jan 15 16:04 NA12893_CEU_sample.vcf.gz
-rw-rw-r-- 1 student student 162856781 Jan 16 07:55 WebDocument_9-7_mysample1.sam
-rw-rw-r-- 1 student student 155079434 Jan 18 08:04 WebDocument_9-7_mysample1_file_sorted.tmp.0000.b
am
-rw-rw-r-- 1 student student 83786629 Jan 18 08:07 WebDocument_9-7_mysample1_file_sorted.tmp.0001.b
am
-rw-rw-r-- 1 student student 12136393 Jan 18 08:08 WebDocument_9-7_mysample1_file_sorted
(student@MSBI32400Lab1 data)$
```

MSBI 32400 Lab 3 7/5/17

Create a new BAM file from SAM

21

- `samtools view -bS WebDocument_9-7_mysample1.sam | samtools sort -o WebDocument_9-7_mysample1_file_sorted.bam`
- `samtools index WebDocument_9-7_mysample1_file_sorted.bam`
- Check that header information is intact:
 - `samtools view -H WebDocument_9-7_mysample1_file_sorted.bam`
- Compare the file sizes of the regenerated files with the original BAMs (send to Jason). Delete the SAM and new BAM files to conserve space on the VM.

MSBI 32400 Lab 3 7/5/17

Homework

22

- E-mail Jason (jasone@uchicago.edu) the screenshots and file information requested above before next class with “**Lab #3**” in the subject line

MSBI 32400 Lab 3 7/5/17