Practical I

Isheng Jason Tsai

Introduction to NGS Data and Analysis
Lecture 12

Practical outline I

Linux basics:
http://wiki.bits.vib.be/index.php/Introduction_to_Linux_for_bioinformatics

The command line exercises
Gentle introduction to the command line
http://wiki.bits.vib.be/index.php/Gentle_introduction_to_the_command_line

Downloading and storing bioinformatics data
http://wiki.bits.vib.be/index.php/Downloading_and_storing_bioinformatics_data

Practical outline I

Managing data

Compression and archiving
http://wiki.bits.vib.be/index.php/Compression_and_archiving

Symbolic links
http://wiki.bits.vib.be/index.php/Symbolic_links

Tips

 # Unzip bz file and untar
tar -xvjf file.tar.bz2

 # Unzip gz file and untar
tar -zxvf file.tar.gz

 # Pipes are useful
For example, print the second column of file
less file | awk ‘{print $2}’ | less

Exercise

Can you download and install the following into your linux environment?

Install FastQC http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Install bwa https://sourceforge.net/projects/bio-bwa/files/
install mummer http://mummer.sourceforge.net/

Tip: It is easier to just download the binaries (already executables). If not possible, download the source code and try compile yourself

Installation

Always look for compiled binaries
(already executable when downloaded)

 # Two simple commands
 # For Mac, you need to install xcode first
make
make install

Sometimes need cmake
Available here: https://cmake.org/download/

Practical outline II

Exercise

Mapping with bwa
Assembly using minia
Mummerplot
Load bam into artemis

BWA - exercises

http://bio-bwa.sourceforge.net/

```
# You need reference file (ref.fa),
# Paired end fastqs (A42_1.fq F42_2.fq)

# Index the genome (make a database)
bwa index ref.fa

# Mapping with bwa
bwa mem -R '@RG\tID:foo\tSM:bar\tLB:library1' ref.fa A42_1.fq A42_2.fq > A42.sam

# Fix flags and sort
samtools fixmate -O bam A42.sam A42_fixmate.bam
samtools sort -O bam -o A42_sorted.bam -T ./ A42_fixmate.bam

# Index the reads
samtools index A42_sorted.bam

# Indexed bam files can be further processed using samtools
samtools view A42_sorted.bam
```

BWA and linux - exercises

 # How many reads are in A42_1.fq (Tip: use less and wc -l)
 # How many reads are in A42_2.fq?
 # Are they the same? Why?

 # Do a less of A42.sam.
 # Check the second column. What does the number mean?
 # Tip: http://broadinstitute.github.io/picard/explain-flags.html

 # Can you count the number of second column from A42.sam file?
 # Tip: use awk, sort and uniq –c

 # Sorted bam versus original sam. How is it different?
samtools view A42_sorted.bam | less

 # More usage of samtools.
 # What do they do?
samtools depth A42_sorted.bam
samtools mpileup A42_sorted.bam

More information

http://www.htslib.org/workflow/#mapping_to_variant

Assembly with minia and assess with QUAST

Can you install it youself?
#Note: For Mac people it's easier to copy minia executables from Dropbox folder
http://minia.genouest.org/

http://bioinf.spbau.ru/quast (Tip: check the manual)

 # Merge all fastq files into one
cat A42_1.fq A42_2.fq > merged.fq

 # One command
minia -in merged.fq -kmer-size 31 -abundance-min 3 -out minia

 # What does the assembly file look like?
Tip: less

 # Run QUAST to assess the assembly
INSTALLATION_PATH/quast.py -R ref.fa minia.contigs.fa

 # Explore around the data

Exercise: install nucmer and mummerplot

http://bioinf.spbau.ru/en/content/spades-download-0

 # Download MUMmer3.23.tar.gz
https://sourceforge.net/projects/mummer/files/

 # Unzip and untar
tar -zxvf MUMmer3.23.tar.gz

 # Install
cd MUMmer3.23/ ; make

 # Go back to data directory and run nucmer
cd PATHOFYOURDATA
PATHOFMUMMER3/nucmer ref.fa minia.contigs.fa -p nucmeroutput

 # Try show-coords to visualise nucmeroutput.delta file
 # Check different options
show-coords nucmeroutput.delta

Advanced: dotplot using nucmer and mummerplot

 # gnuplot needs to be installed
 # Type gnuplot in command line if it's installed

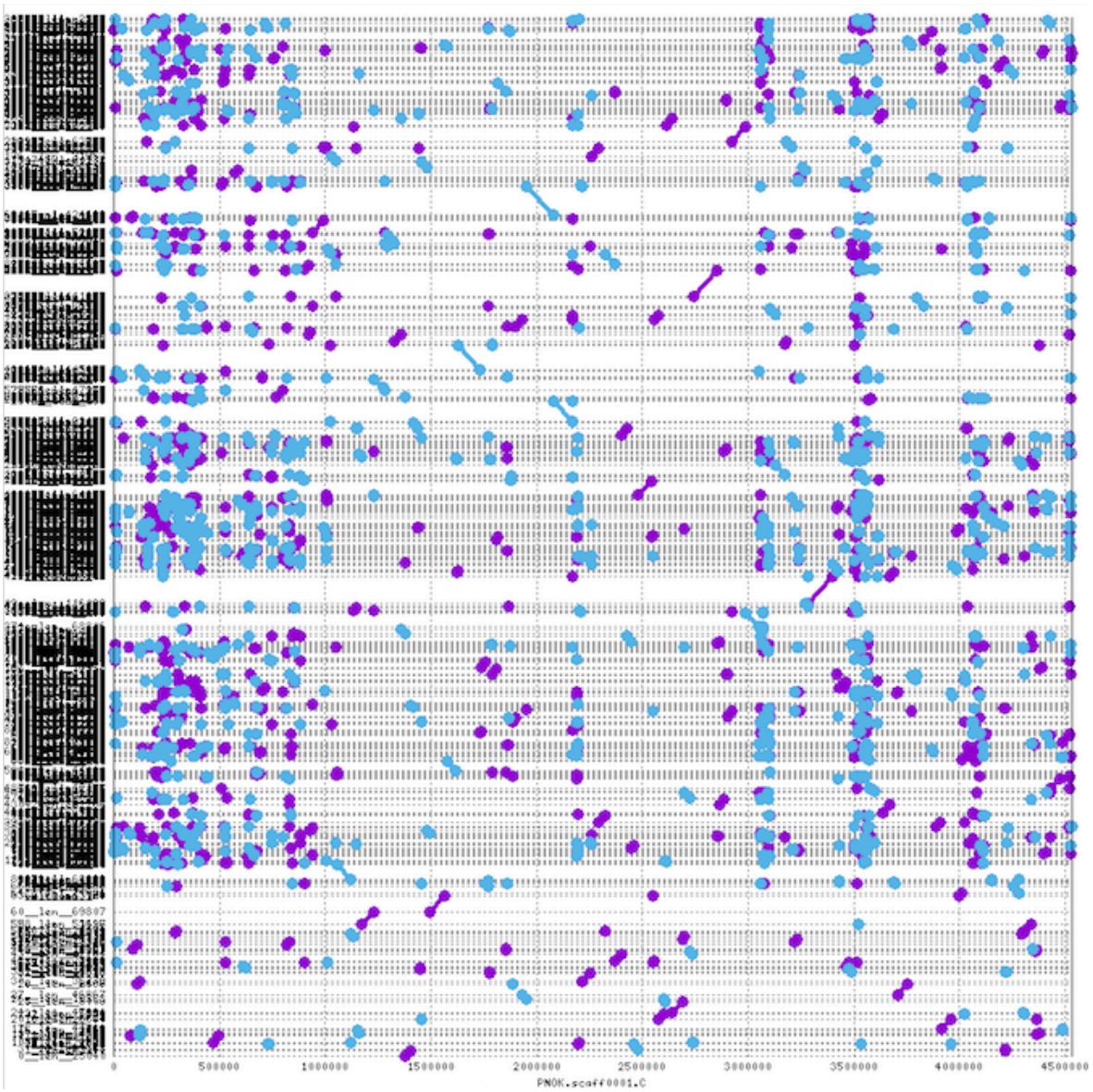
 # For Mac, you need to install gnuplot
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
brew install gnuplot

Note: You need to comment out three lines to make mummerplot work in Mac
#$P_FORMAT .= "\nset mouse format \"$TFORMAT\"";
#$P_FORMAT .= "\nset mouse mouseformat \"$MFORMAT\"";
#$P_FORMAT .= "\nset mouse clipboardformat \"$MFORMAT\"";

 # Mummerplot
PATHOFMUMMER3/mummerplot -png nucmeroutput.delta

Q: Do you understand the dotplot?

Dotplot


0
500000
1000000
1500000
2000000
2500000
3000000
3500000
4000000
4500000
PNOK.scaff0001.C

Artemis

 # Download website
http://www.sanger.ac.uk/science/tools/artemis

 # Load reference file and gff
1. Open File Manager -> find ref.fa and double click
2. Drag ref.gff into the window

 # Load the BAM
1. Read BAM/VCF

Q: Can you see where the SNPs are?