An Introduction to SeqMonk

Simon Andrews
simon.andrews@babraham.ac.uk
v2023-06
Course Programme

• Installing SeqMonk and Dependencies
• Creating a Project and Importing Data
• UI layout and basic controls
• Probes and Quantitation
• Plotting Figures
• Filtering Probes
• Saving, Reporting and Vistories
Installing SeqMonk and Dependencies
SeqMonk Mapped Sequence Analysis Tool

- README
- INSTALL Installation instructions for the program.
- Release Notes Please read these before using the program.
- SeqMonk v1.48.1 for 64-bit Windows
- SeqMonk v1.48.1 for 64-bit Linux
- SeqMonk v1.48.1 for 64-bit Mac OS X
Windows

• Unzip zip file
Mac OSX

• Download and run the DMG file
• Copy App to the Applications folder
Mac OSX
Linux

- Download tar file
- Run the launcher

```
student@ip-172-31-23-170:~$ wget --quiet https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/seqmonk_v1.48.1_linux64.tar.gz
student@ip-172-31-23-170:~$ tar -xzf seqmonk_v1.48.1_linux64.tar.gz
student@ip-172-31-23-170:~$ cd SeqMonk/
student@ip-172-31-23-170:~/SeqMonk$ ./seqmonk
CLASSPATH is: /home/student/SeqMonk:/home/student/SeqMonk/htsjdk.jar:/home/student/SeqMonk/Jama-1.0.2.jar:/home/student/SeqMonk/common-lib.jar:/home/student/SeqMonk/CommLib-jdk1.7-3.5.jar
Java interpreter is `/home/student/SeqMonk/jre/bin/java'
openjdk version "13.0.2" 2020-01-14
OpenJDK Runtime Environment AdoptOpenJDK (build 13.0.2+8)
OpenJDK 64-Bit Server VM AdoptOpenJDK (build 13.0.2+8, mixed mode, sharing)
Prefs file is at: /home/student/seqmonkPrefs.txt
Set memory to 0 from prefs file
Memory ceiling is 10240
Raw physical memory is 7847
Using 5231 MB of RAM to launch seqmonk
Correcting for VM actual requested allocation for 5231 is 5230
Command is: /home/student/SeqMonk/jre/bin/java -Xss4m -Xmx5230m -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true uk.ac.babraham.SeqMonk.SeqMonkApplication
```
Installing R

- [https://cran.r-project.org/](https://cran.r-project.org/)

- Linux users need to install development versions of libssl, libcurl4, libxml2
First Launch
Creating a Project and Importing Data
File > New Project
File > Import Data
UI Layout and Basic Controls
Data Types

**Annotation**
- Annotation collection
  - All of the annotation in the project
- Annotation set
  - A collection of features of varying types which came from the same source
- Annotation track
  - A set of features of the same type which might be drawn from several annotation sets

**Reads**
- Data Set
  - A set of reads which came from one source (usually file)
- Data Group
  - A set of reads merged together from multiple datasets.
- Replicate Set
  - A collection of data sets / groups which come from the same biological condition
Creating Data Groups / Replicate Sets

Data > Edit Replicate Sets

Data > Auto Create Groups/Sets
Changing Chromosome Tracks
Changing Display Preferences
Movement Controls

Mouse
- Scroll Wheel to move left / right
- Click and drag to zoom in
- Right click to zoom out
- Double click for feature details

Keyboard
- Up arrow to zoom in
- Down arrow to zoom out
- Left / Right arrows to move along
- Control +F to search (find)
- Control +G to jump to position (goto)
Probes and Quantitation
Terminology for Quantitation

• **Probe**
  – A region of the genome where a measurement will be made. Has a start and end, and optionally a strand

• **Probe Set**
  – The full set of probes currently being used for quantitation (eg all the promoters in the genome)

• **Probe List**
  – A subset of probes drawn from within the current Probe set (eg all of the promoters on chromosome 1)
Quantitation Rules

• A project can only have a single probe set, and the same probe set is used to quantitate all data

• Each probe has a quantitative value associated with it in every Data Set and Data Group

• Replicate Sets show the mean quantitation of the Data Sets within them

• The chromosome view will show only the currently selected probe list, and most plots only use data from the current probe list
Data > Define Probes

Probe Generators
(Different ways of defining a probe set)

Generator Options
(Options specific to the currently selected generator)
Data > Quantitate Existing Probes

• Opens automatically after defining new probes
• Can be rerun on existing probes without changing them

Quantitation Methods
(Different ways of assigning a value to a probe)

Quantitation Options
(Options specific to the currently selected quantitation method)
Quantitation Example
Quantitation Example

Bars = Probes
Check the positions match with what you expect

Height = Value = Colour
Capped at 95th Percentile by default
Quantitation Pipelines

- Data > Quantitation Pipelines
- Combine Probe Generation and Quantitation

![Pipeline Options]

- Define Quantitation...

  - Quantitation Options
    - RNA-Seq quantitation pipeline
    - Active transcription quantitation pipeline
    - Intron regression pipeline
    - Gene trap quantitation pipeline
    - Wiggie Plot for Initial Data Inspection
    - Bisulfite methylation over features
    - Splicing efficiency quantitation
    - Antisense transcription pipeline
    - Codon Bias Pipeline
    - Transcription termination pipeline

  - Transcript features
    - mRNA

  - Library type
    - Non-strand specific

  - Libraries are paired end
    - [ ]

  - Merge transcript isoforms
    - [ ]

  - Generate Raw Counts
    - [ ]

  - Log transform
    - [ ]

  - Apply transcript length correction
    - [ ]

  - Don't quantitate probes with no counts
    - [ ]

  - Correct for DNA Contamination
    - [ ]

  - Correct for Duplication
    - [ ]

[Close] [Run Pipeline]
Quantitation Adjustment

- Additional Options once you have a quantitation

**Blue**
- Fresh quantitation from the raw read data

**Red**
- Methods to normalise / scale / adjust the existing quantitation
Plotting Figures
Plotting

• By Default
  – Uses the data stores shown in the chromosome view
  – Uses the probes in the selected probe list
Some Plots are Interactive

- Hover to see label
- Click to fix label
- Double click to show probe in Chromosome View
- Triple Click to clear labels
Some Plots can be Duplicated
Filtering Probes
Filtering Concepts

- Start from an existing Probe List
- Run a Filter to select a subset of those probes
- Create a new Probe List as a child of the original list
- Build up a tree of filtered Probe Lists
Filters

Starting List
Check this is what you expect

Number of passed probes

New List Name
Filter Details

- Rename list (doesn't change contents)
- View list (shows options used)
- Delete list (and any children)
Saving, Reporting and Vistories
Saving SeqMonk Projects

• File > Save Project
• Saves Everything
  – Data
  – Quantitations
  – Probes / Filters
  – Current View
• Single file with .smk file extension
• Can be moved to another machine and opened*

*Unless using a custom genome – you need to copy that separately
Saving Images

- PNG – Bitmap – Screenshot
- SVG – Vector – Editable
Editing SVG Images
Creating Reports

• **Annotated Probe Report**
  – Generates a report for every probe in a probe list. Can annotate it with a feature from a chosen annotation track

• **Probe Group Report**
  – Generates a report from a probe list but can group together probes which are close to each other

• **Feature Report**
  – Generates a report for all features in an annotation track. Relates them to probes in a probe list

• **Data Store Summary**
  – Gives statistics about the data and quantitation in the currently selected probe list.
Creating Reports

- Sort by clicking headers
- Double click to change view
- Save to file (tab delimited text)
Vistories

- A way to formally record your activities in SeqMonk
- Generates an HTML report
- Created automatically
- You can add commentary, images, reports and summaries
- Easy way to record and share your analysis
Vistories

Example Vistory

This is an example which shows what you can do with vistories.

Load the data

New Probe Set: Cpg Islands
Feature generator using Cpg islands duplicates removed. Centered on feature from 2000-2000

Probes Quantitated

Read Count Quantitation using All Reads correcting for total count per million reads

Probes Filtered

New Probe Set: My Selected Probe List (1353 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome 1 on strand All Probes

Probes Quantitated

Read Count Quantitation using All Reads correcting for total count per million reads

Filter

New Probe Set: My Selected Probe List (1353 probes)
[Probe Value Filter] Filter on probes in Cpg Islands where exactly 1 of GSM307618.bam, GSM307619.bam had a value below 200.0. Quantitation was Read Count Quantitation using All Reads correcting for total count per million reads

New Probe Set: Chr X (934 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome X on strand All Probes

New Probe Set: Ovlapping gene (671 probes)
[Feature Filter] Filter probes in Chr 1 on region based on gene Over feature relationship is Ovlapping

New Probe Set: Random Probes (100 probes)
Random Probe Filter A random subset of 100 probes from Cpg Islands
• Save as .smv file (editable)
• Save as HTML (report)
https://www.bioinformatics.babraham.ac.uk/vistorydb/