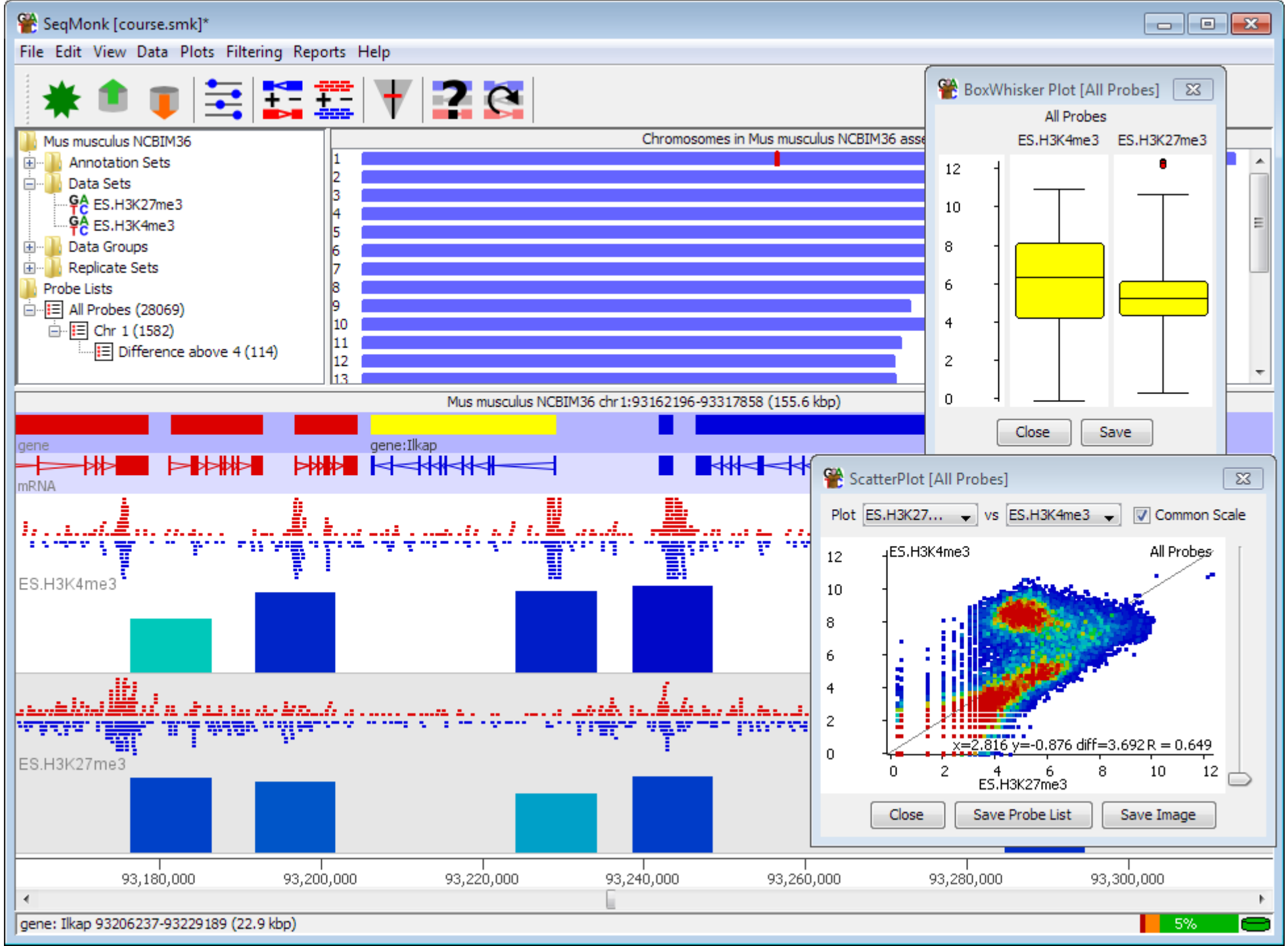
An Introduction to SeqMonk

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

Babraham Bioinformatics


SeqMonk [course.smk]*
File Edit View Data Plots Filtering Reports Help
Mus musculus NCBIM36
Annotation Sets
Data Sets
ES.H3K27me3
ES.H3K4me3
Data Groups
Replicate Sets
Probe Lists
All Probes (28069)
Chr 1 (1582)
Difference above 4 (114)
Chromosomes in Mus musculus NCBIM36 asse
Mus musculus NCBIM36 chr1:93162196-93317858 (155.6 kbp)
gene
gene:Ilkap
mRNA
ES.H3K4me3
ES.H3K27me3
93,180,000
93,200,000
93,220,000
93,240,000
93,260,000
93,280,000
93,300,000
gene: Ilkap 93206237-93229189 (22.9 kbp)
5%
BoxWhisker Plot [All Probes]
All Probes
ES.H3K4me3
ES.H3K27me3
Close
Save
ScatterPlot [All Probes]
Plot ES.H3K27...
vs ES.H3K4me3
Common Scale
ES.H3K4me3
All Probes
x=2.816 y=-0.876 diff=3.692 R = 0.649
ES.H3K27me3
Close
Save Probe List
Save Image

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

Babraham Bioinformatics

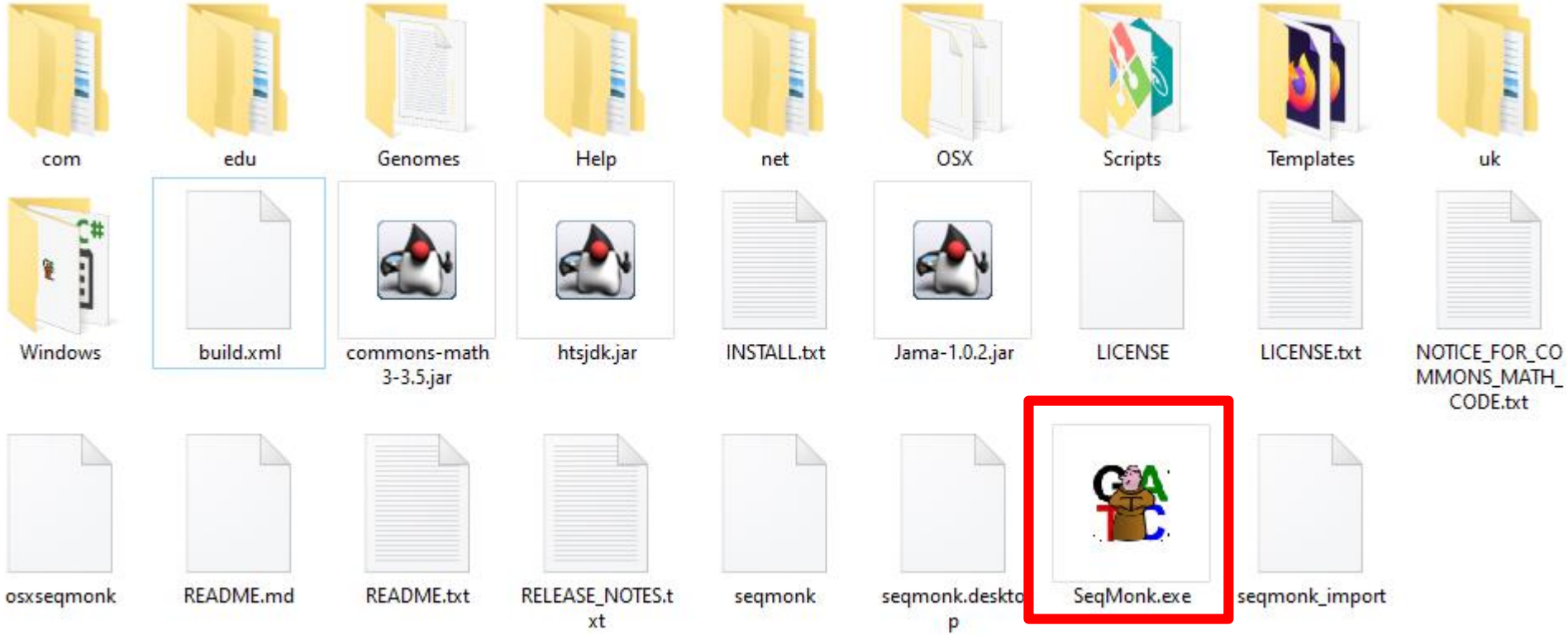
https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/

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 OSX

Windows

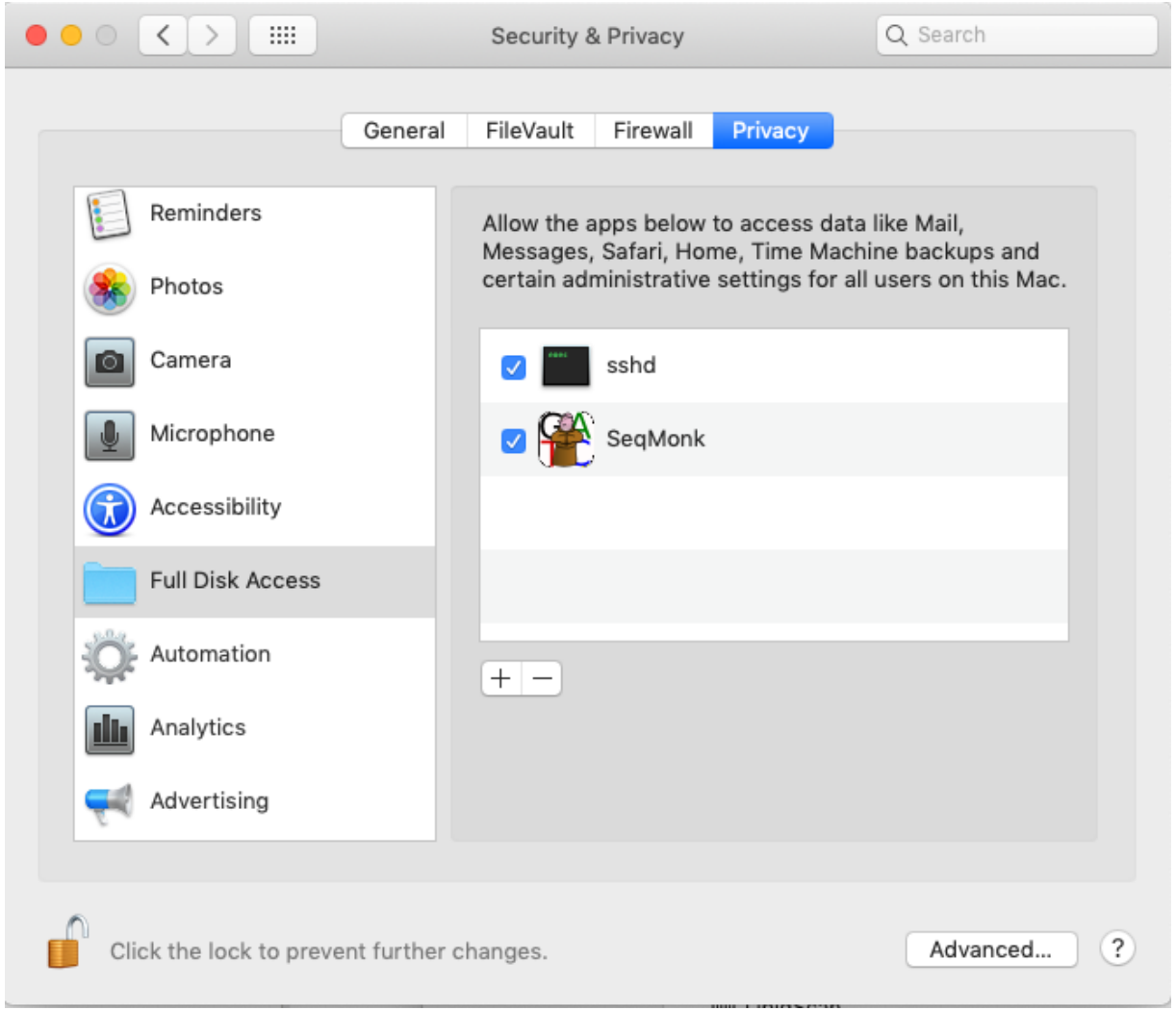
- Unzip zip file



Mac OSX

- Download and run the DMG file
- Copy App to the Applications folder

Mac OSX


Security & Privacy
Search
General
FileVault
Firewall
Privacy
Reminders
Photos
Camera
Microphone
Accessibility
Full Disk Access
Automation
Analytics
Advertising
Allow the apps below to access data like Mail, Messages, Safari, Home, Time Machine backups and certain administrative settings for all users on this Mac.
sshd
SeqMonk
Click the lock to prevent further changes.
Advanced...

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/comm
ons-math3-3.5.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/seqmonk_prefs.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.S
eqMonk.SeqMonkApplication
```

Installing R

- https://cran.r-project.org/

Download and Install R

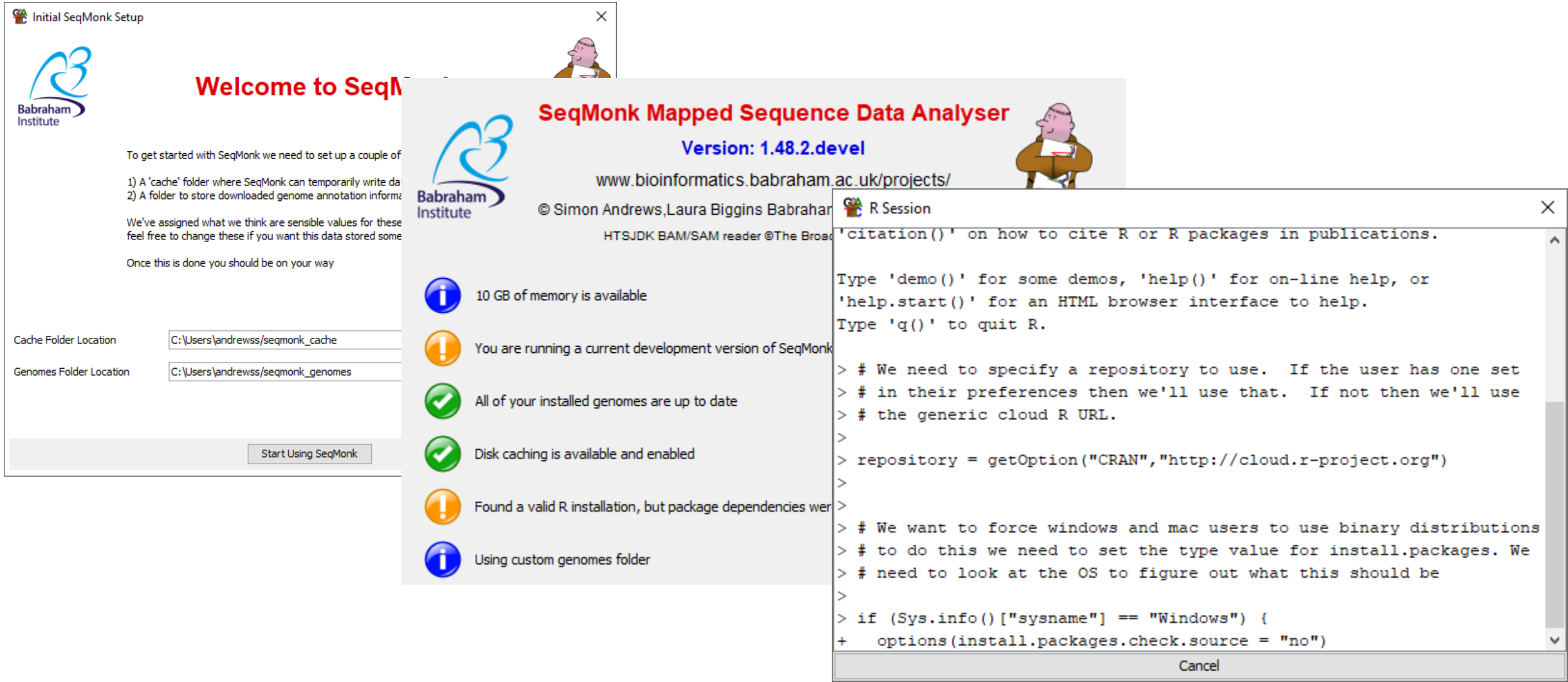
Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

- Linux users need to install development versions of libssl, libcurl4, libxml2

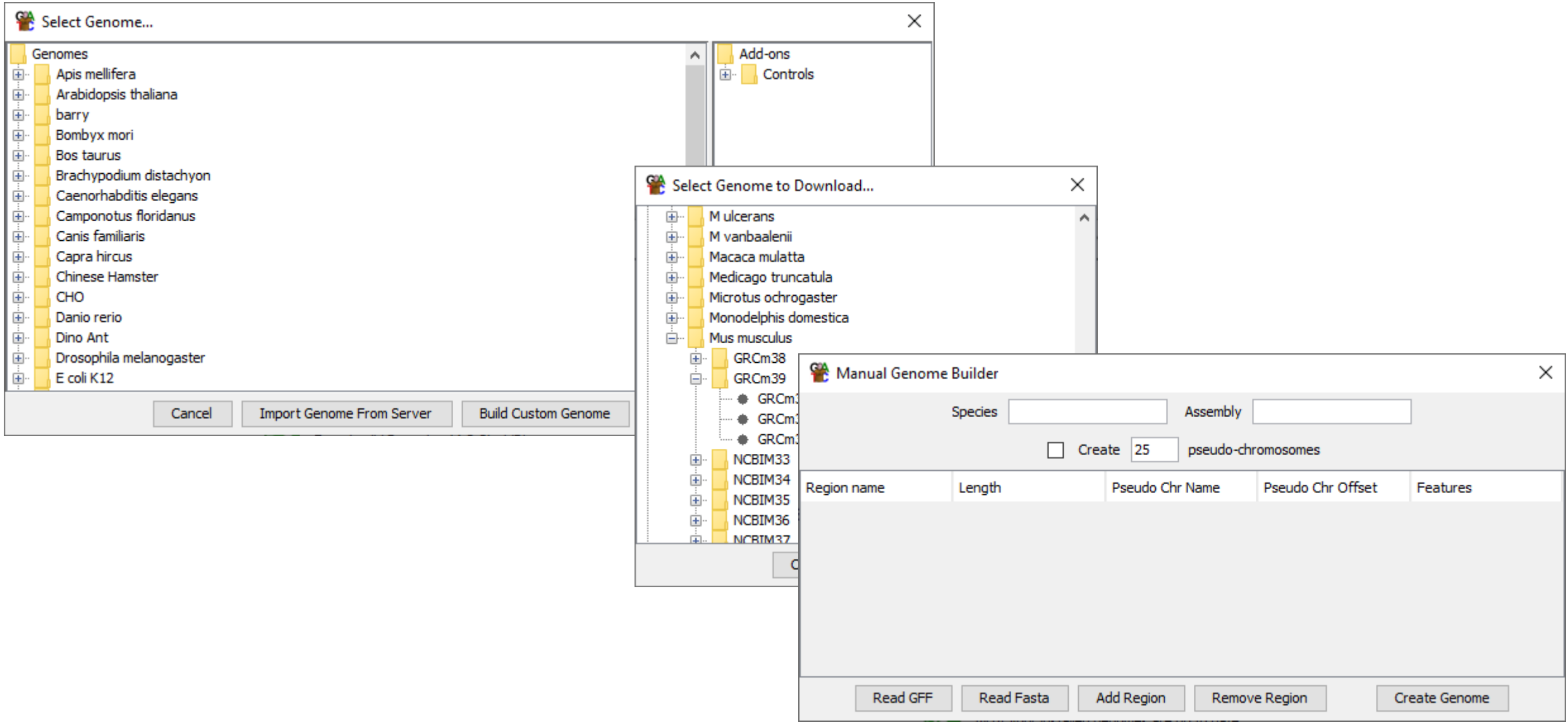
First Launch



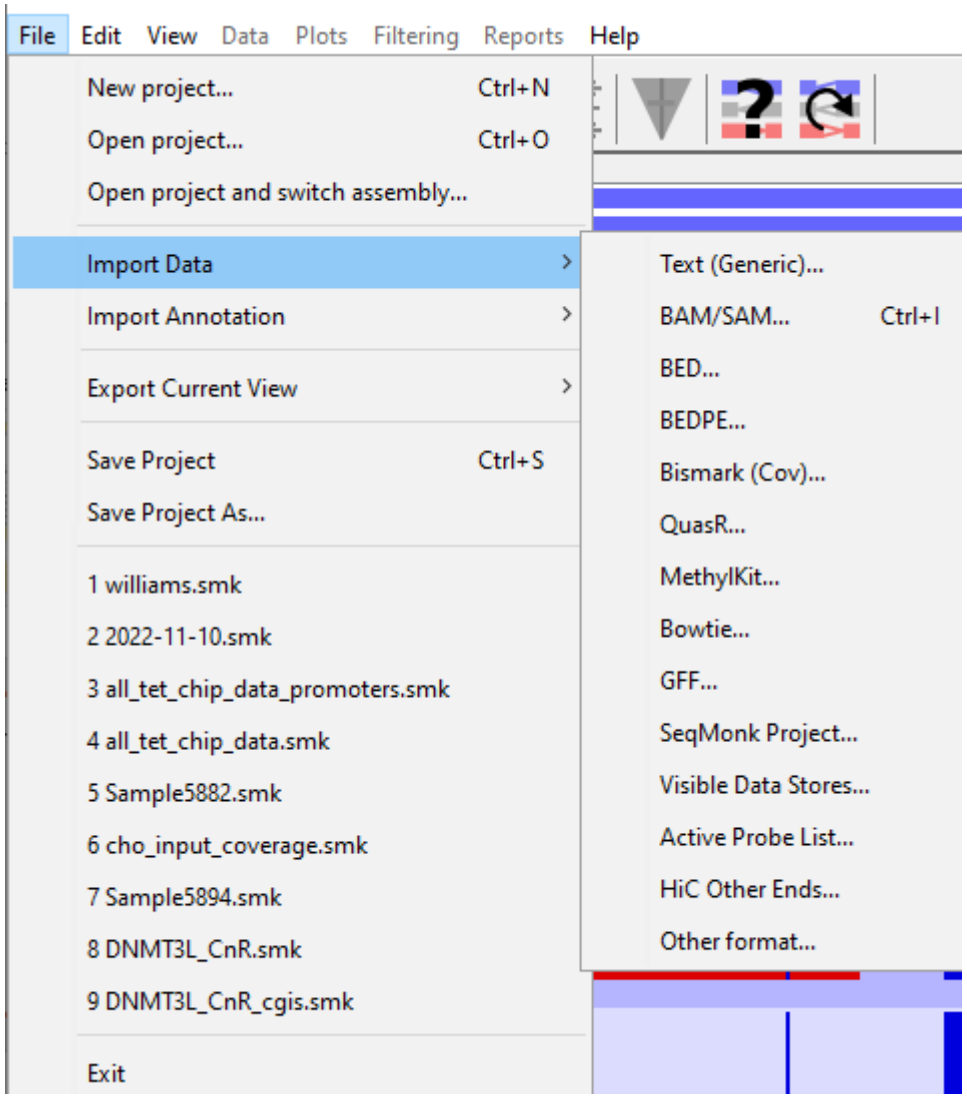
Creating a Project and Importing Data

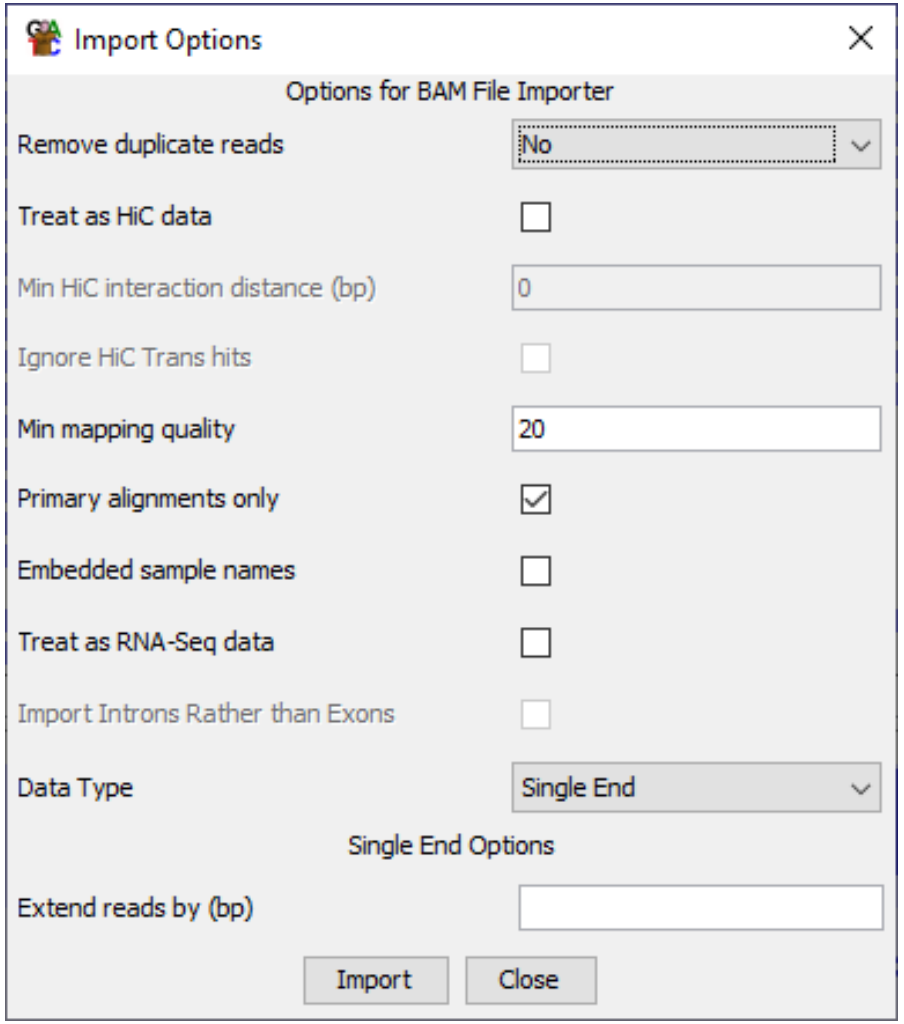
Babraham Bioinformatics

File > New Project



File > Import Data

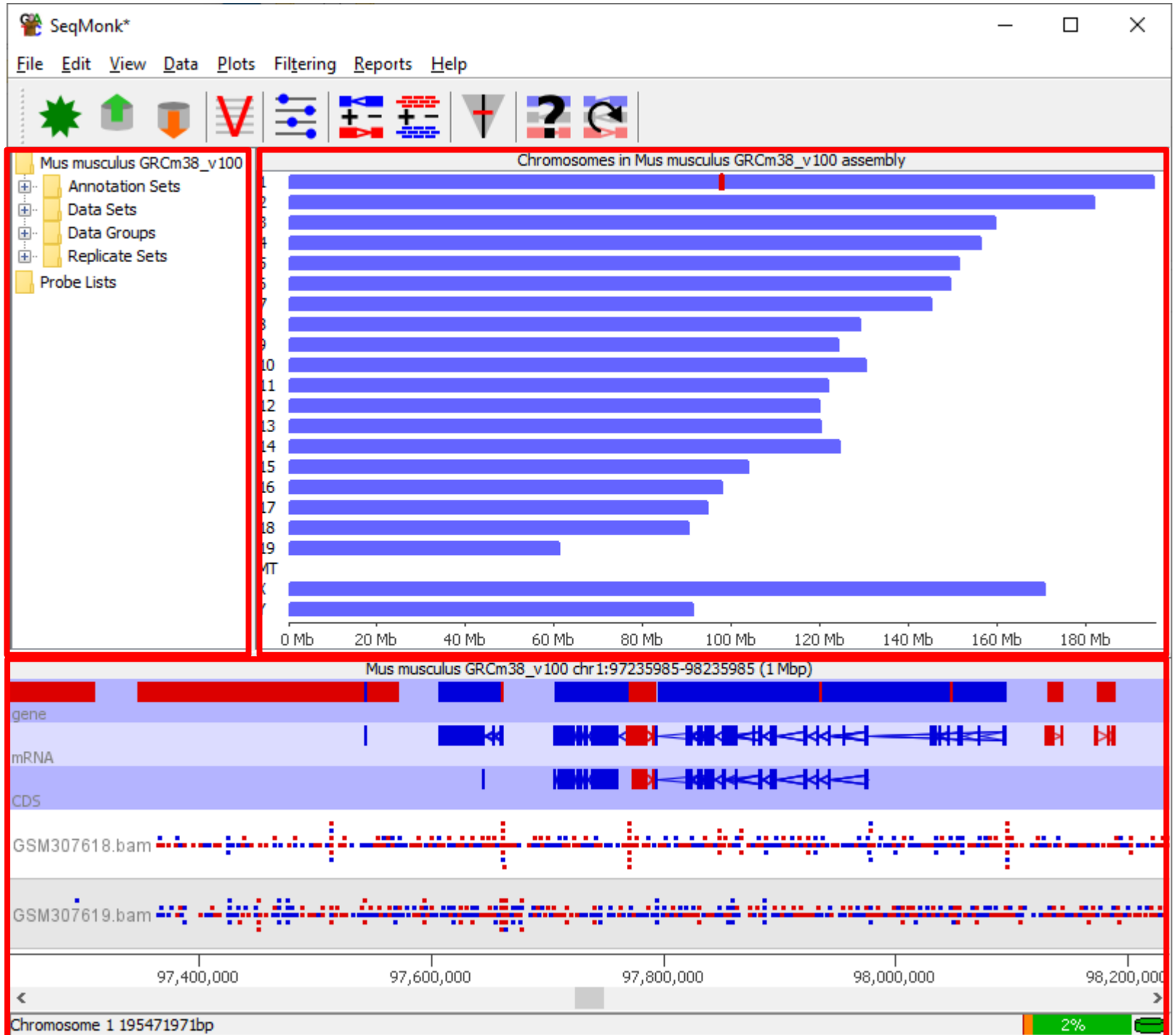




UI Layout and Basic Controls

Babraham Bioinformatics

Data View



Genome View

Chromosome View

Data Types

Annotation

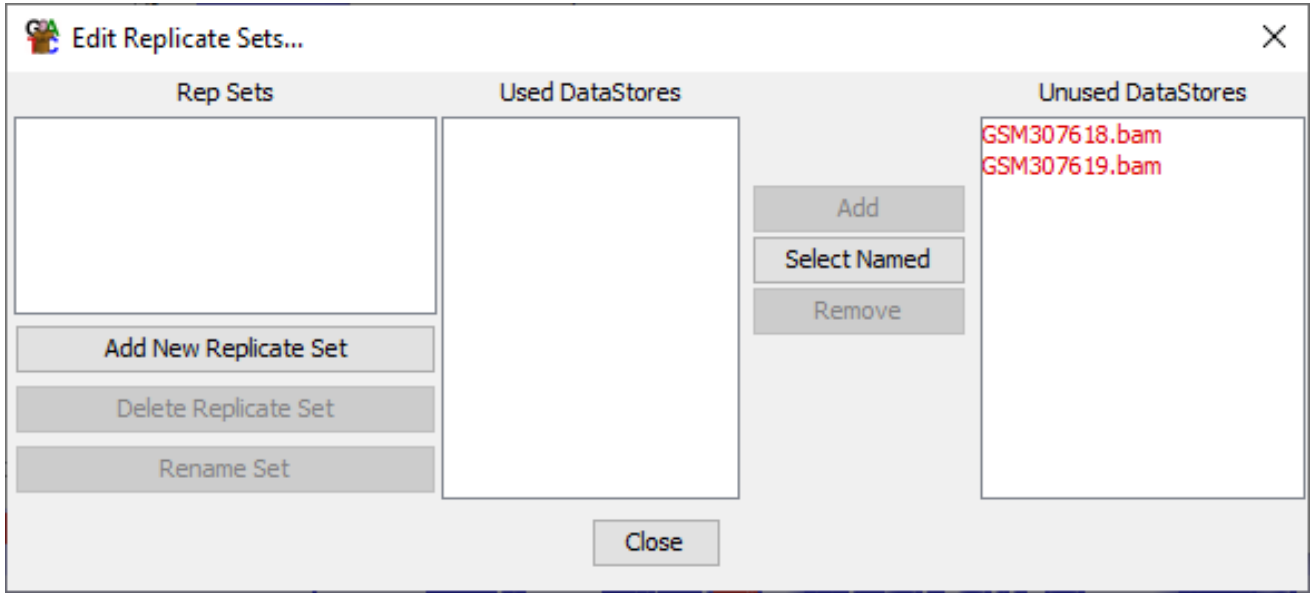
- Annotation collection
  - All of the annotation in the project
- Annotation set
  - A collection of features of varying types which came from the same source
- Annotation track
  - A set of features of the same type which might be drawn from several annotation sets

Reads

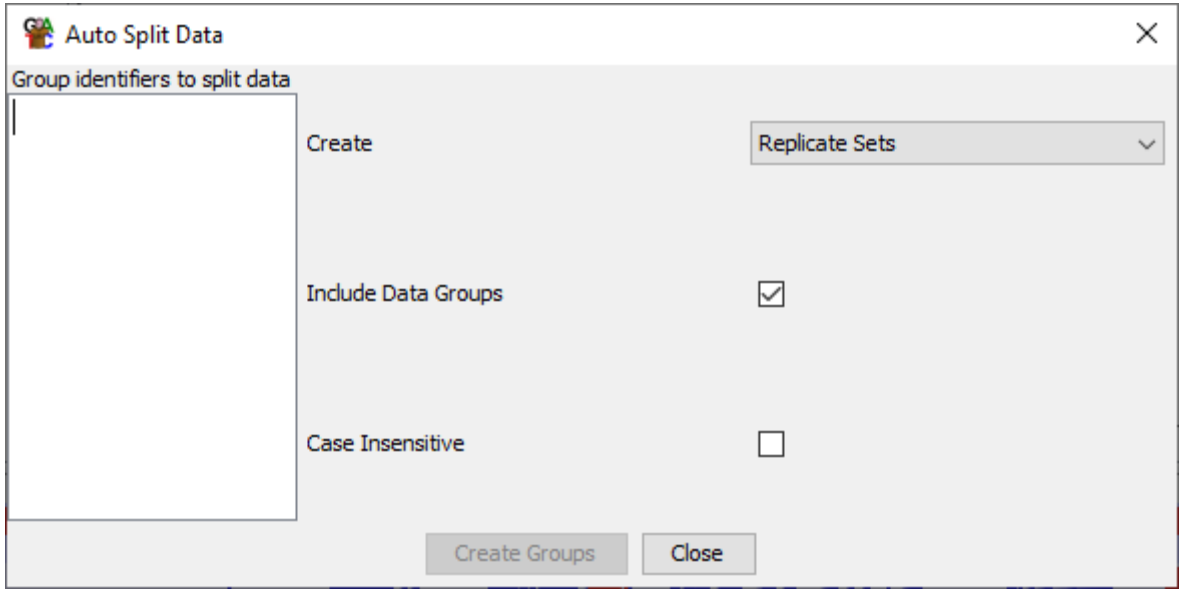
- Data Set
  - A set of reads which came from one source (usually file)
- Data Group
  - A set of reads merged together from multiple datasets.
- Replicate Set
  - A collection of data sets / groups which come from the same biological condition

Creating Data Groups / Replicate Sets

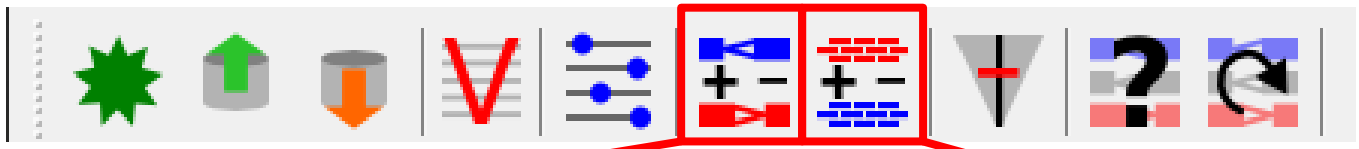
Data > Edit Replicate Sets

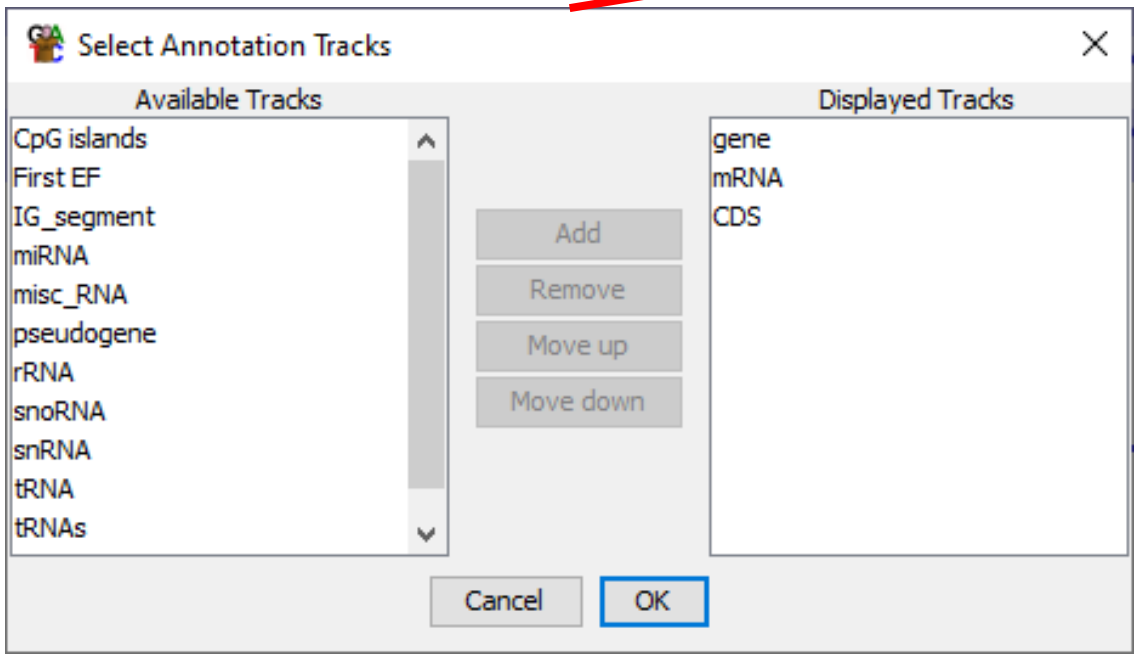


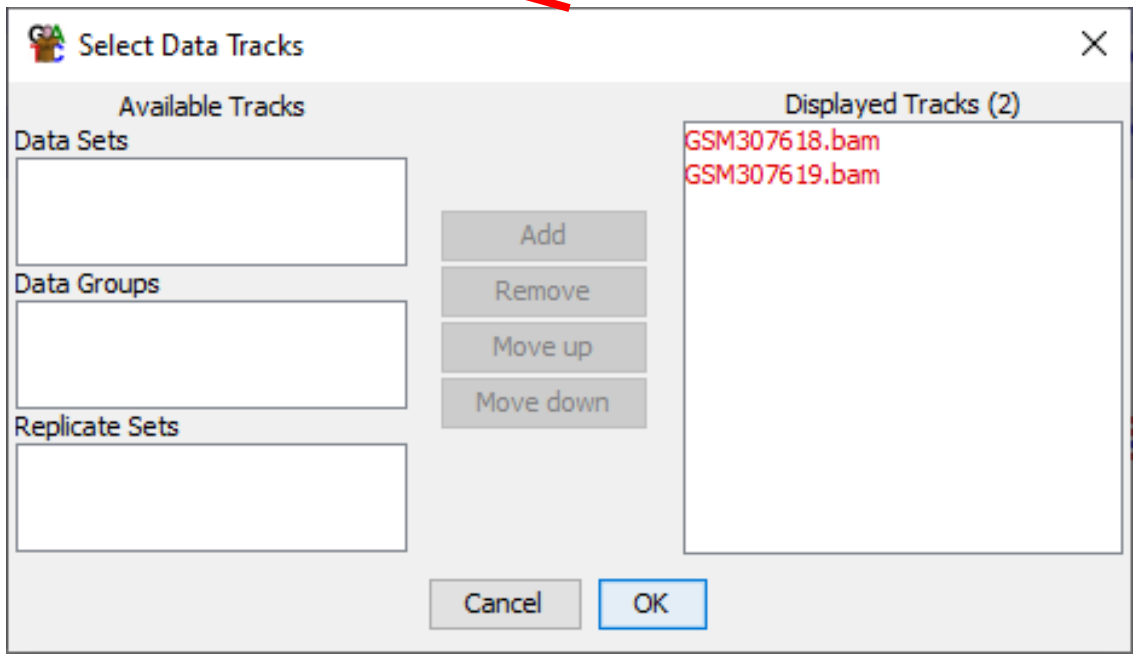
Data > Auto Create Groups/Sets



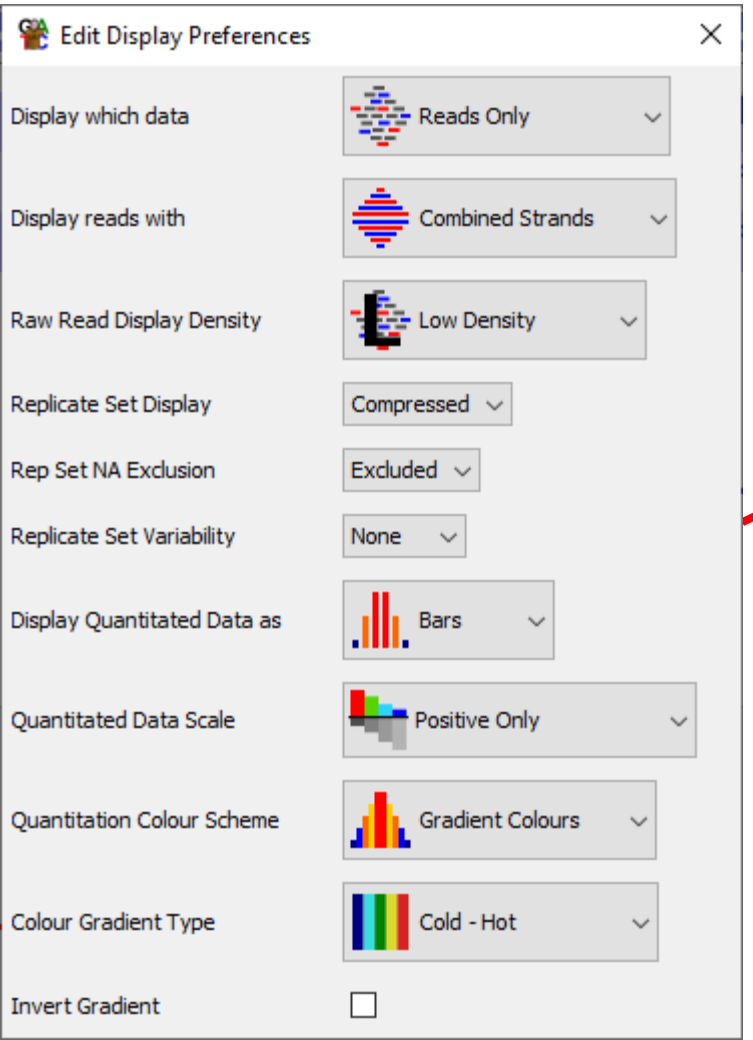
Changing Chromosome Tracks

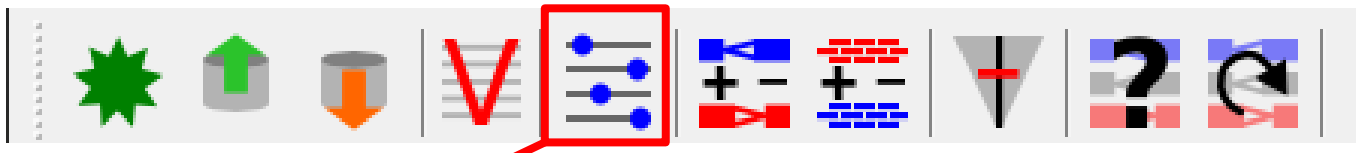






Changing Display Preferences





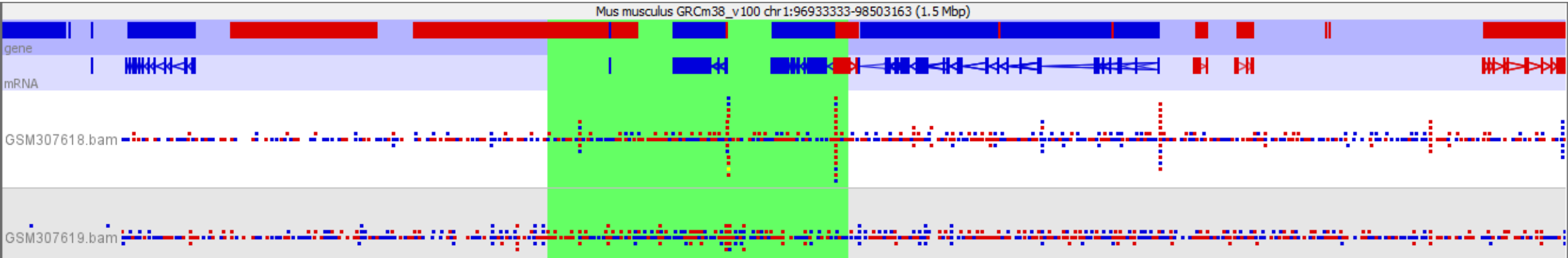
Movement Controls

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

Babraham Bioinformatics

Terminology for Quantitation

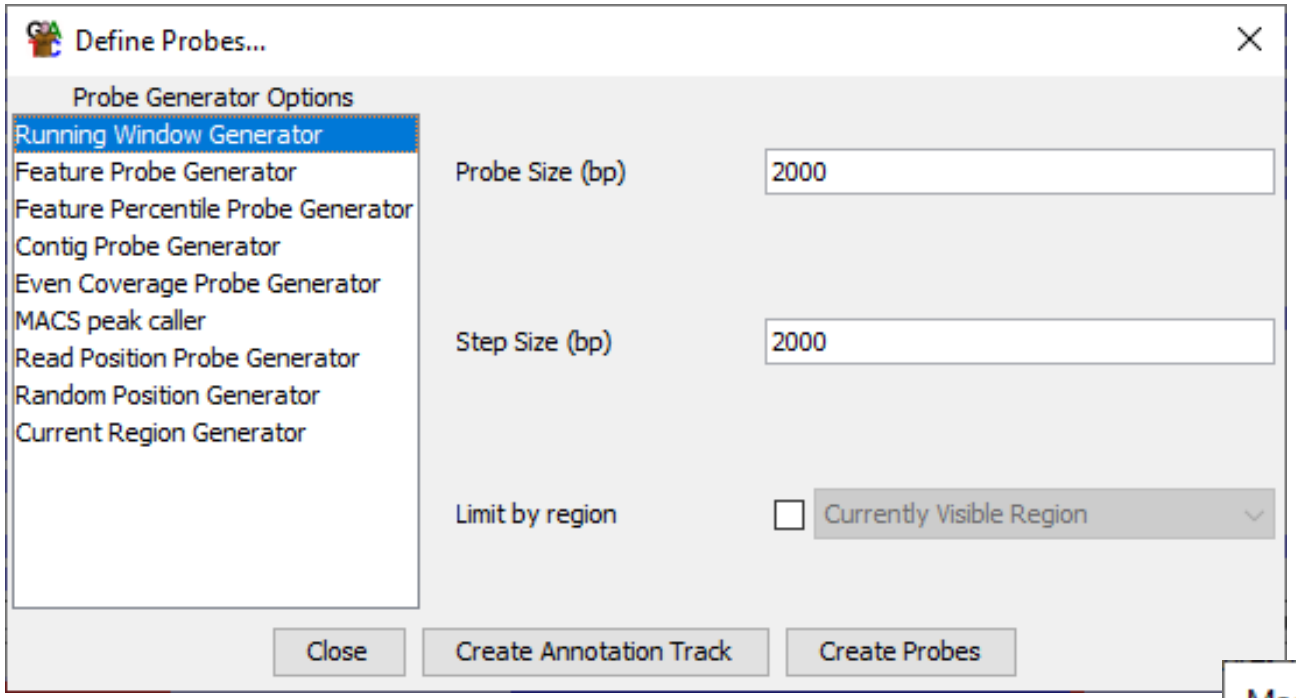
- **Probe**
  - A region of the genome where a measurement will be made. Has a start and end, and optionally a strand
- **Probe Set**
  - The full set of probes currently being used for quantitation (eg all the promoters in the genome)
- **Probe List**
  - A subset of probes drawn from within the current Probe set (eg all of the promoters on chromosome 1)

Quantitation Rules

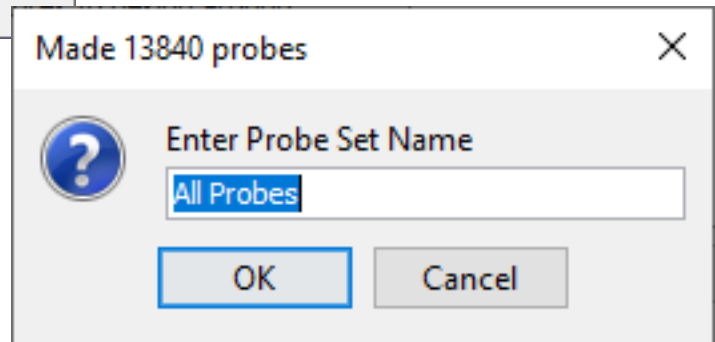
- A project can only have a single probe set, and the same probe set is used to quantitate all data
- Each probe has a quantitative value associated with it in every Data Set and Data Group
- Replicate Sets show the mean quantitation of the Data Sets within them
- The chromosome view will show only the 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)



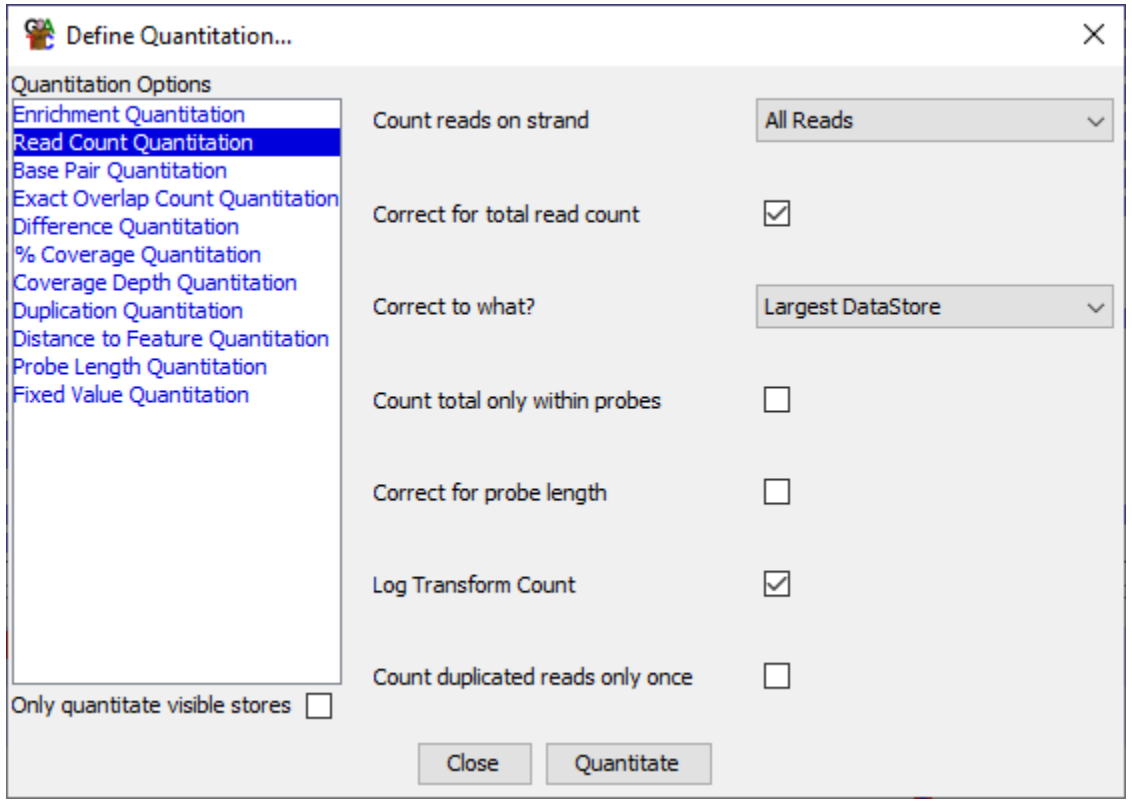
Generator Options
(Options specific to the currently selected generator)



Data > Quantitate Existing Probes

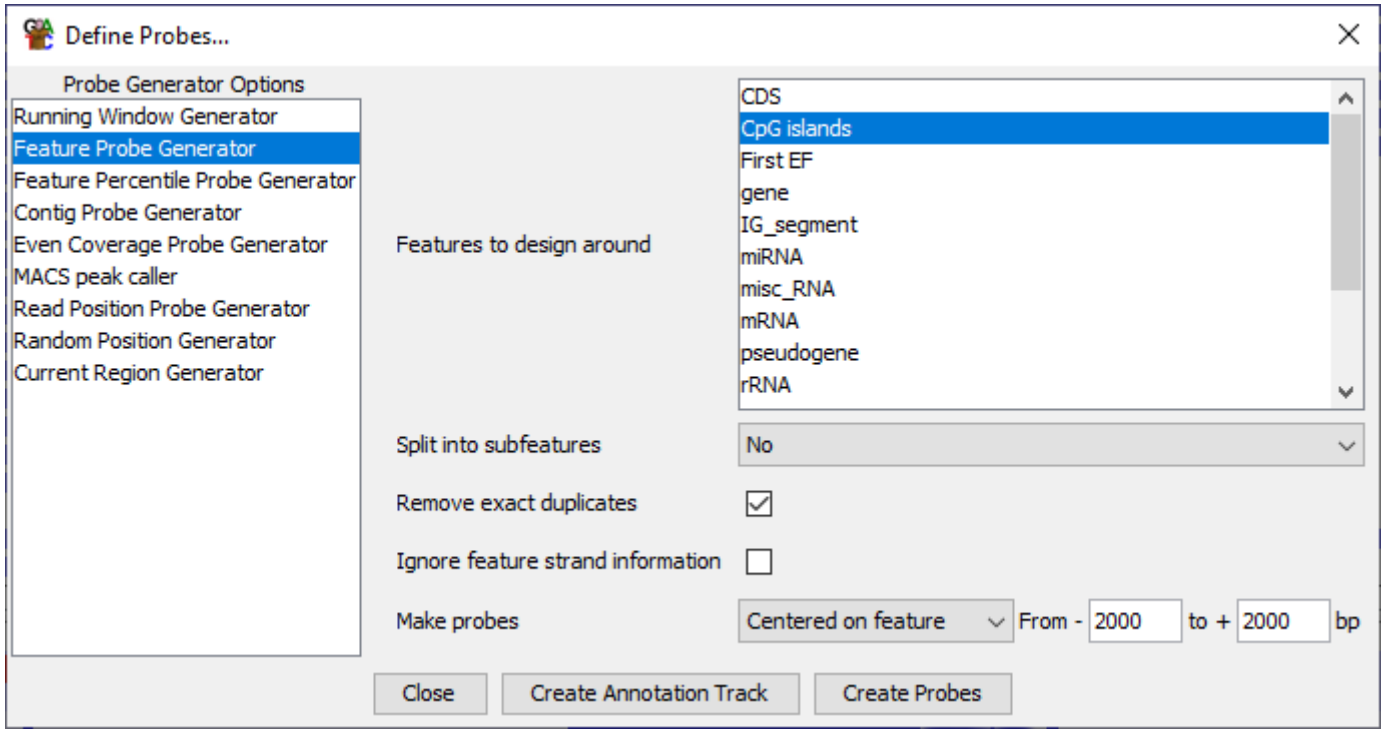
- Opens automatically after defining new probes
- Can be rerun on existing probes without changing them

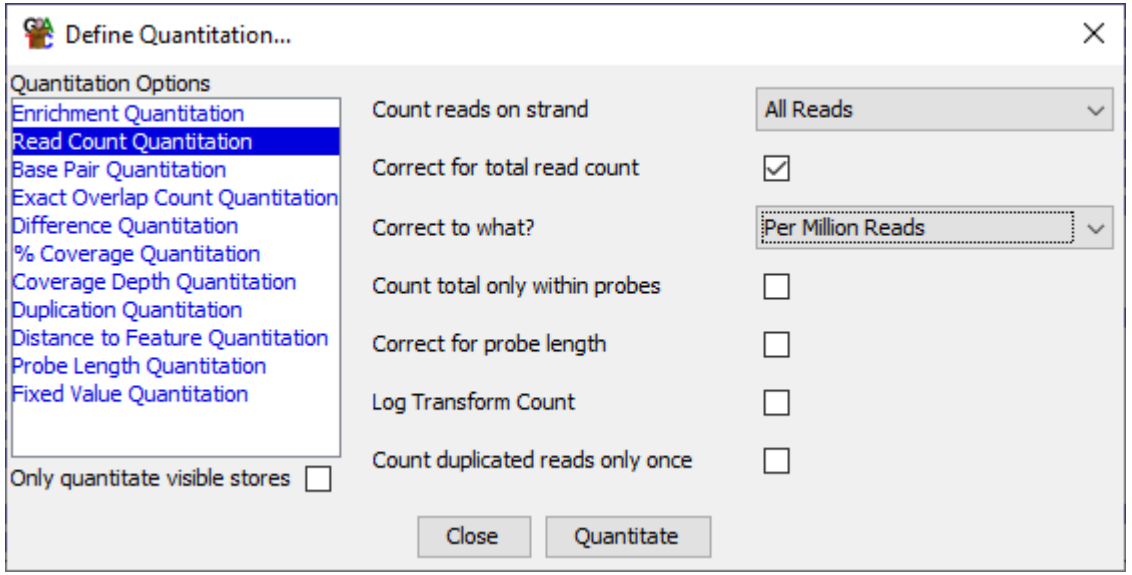
Quantitation Methods
(Different ways of assigning a value to a probe)

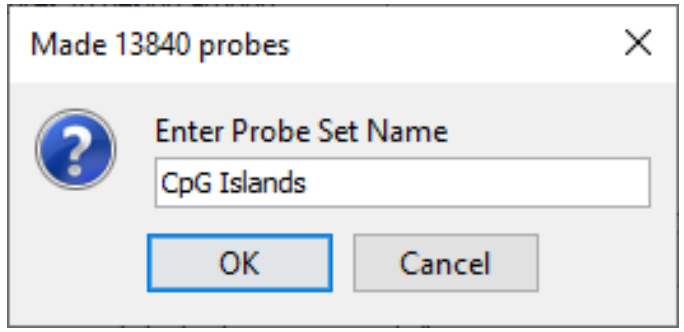


Quantitation Options
(Options specific to the currently selected quantitation method)

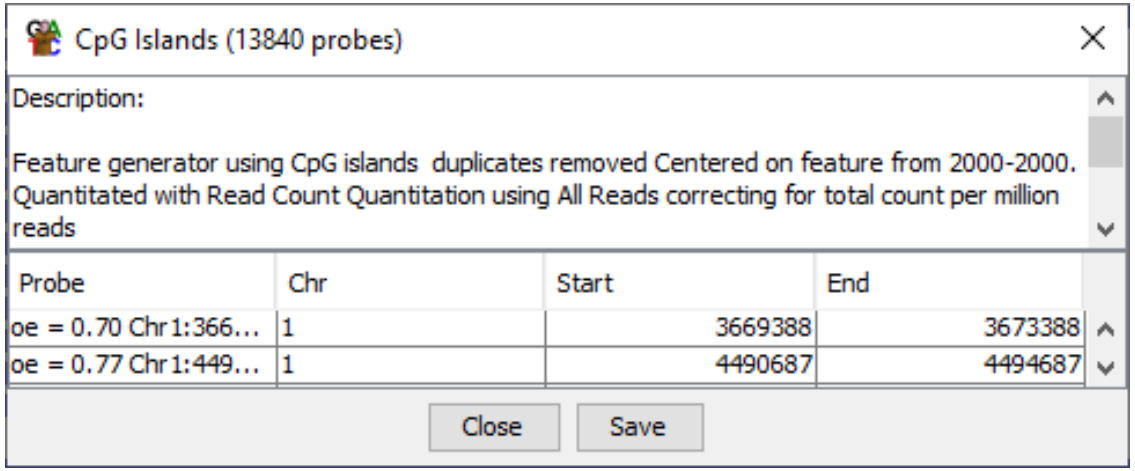
Quantitation Example

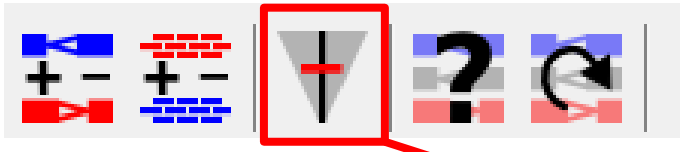


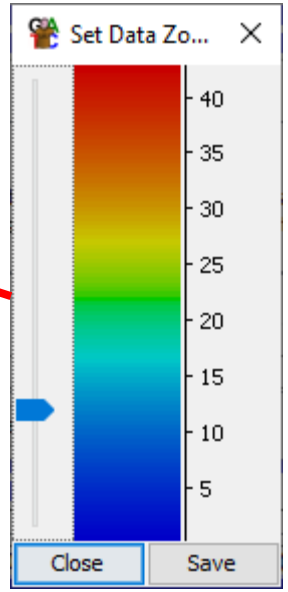




Quantitation Example

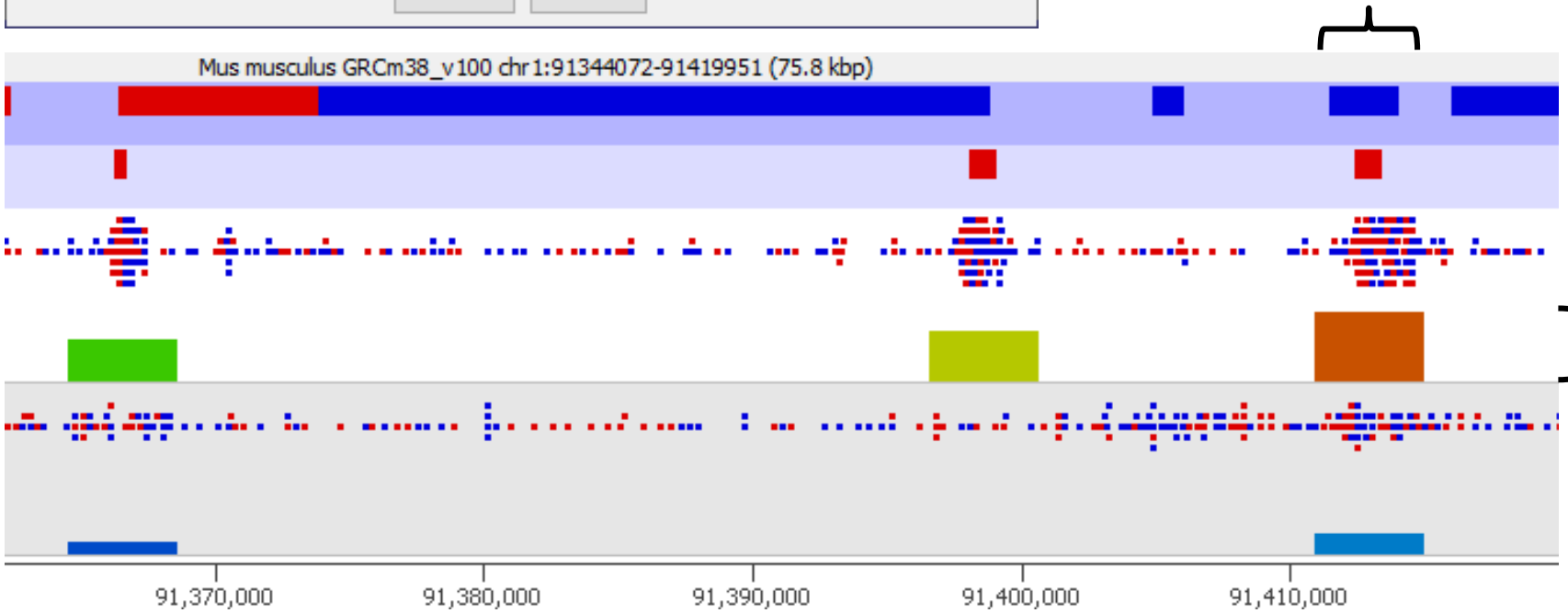






Bars = Probes

Check the positions match with what you expect



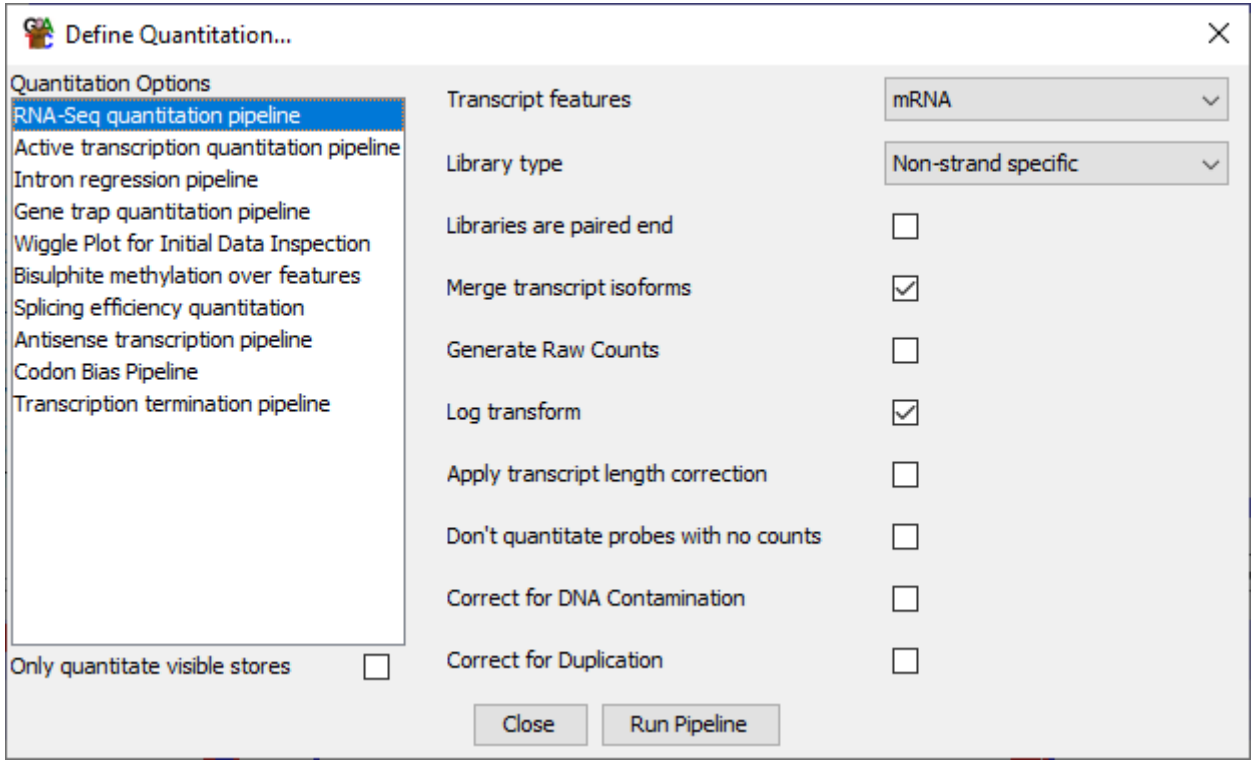
Height = Value = Colour

Capped at 95th Percentile by default

Quantitation Pipelines

- Data > Quantitation Pipelines
- Combine Probe Generation and Quantitation

Pipeline



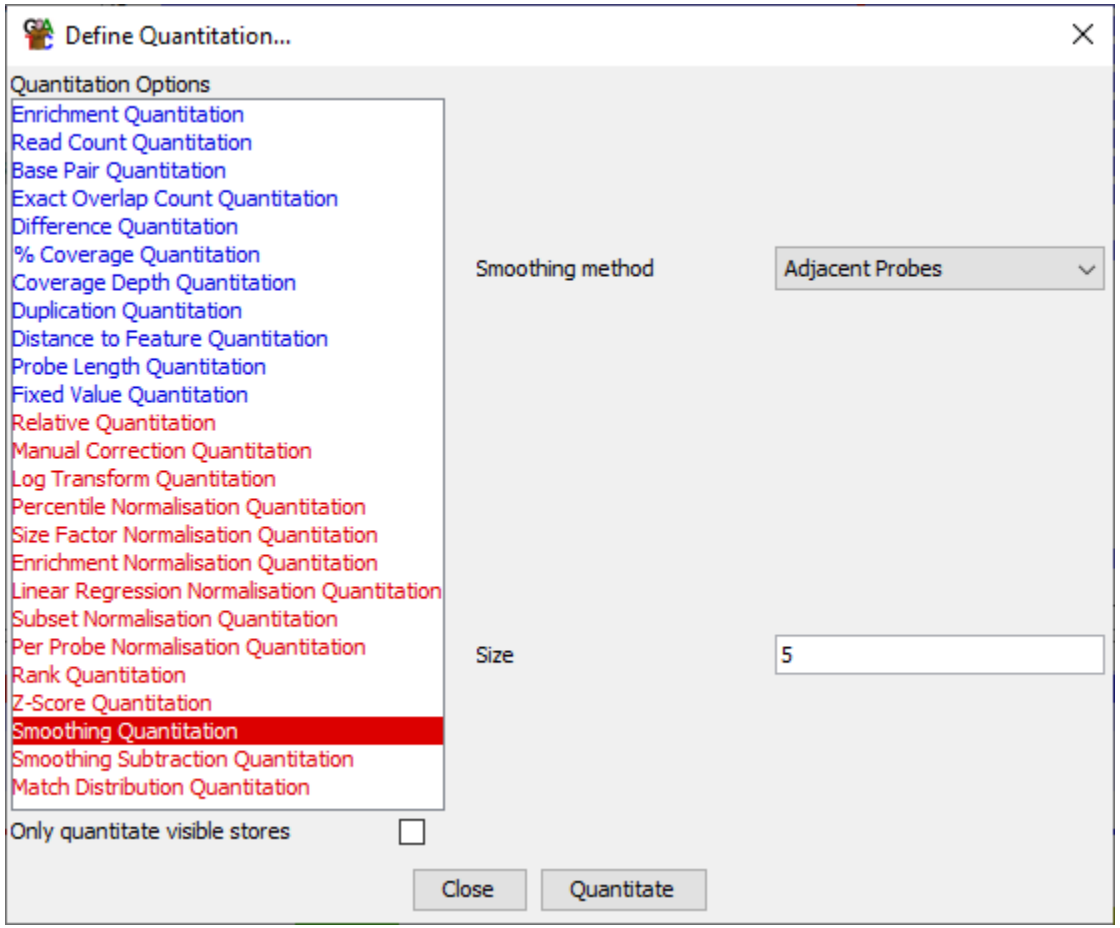
Pipeline Options

Quantitation Adjustment

- Additional Options once you have a quantitation

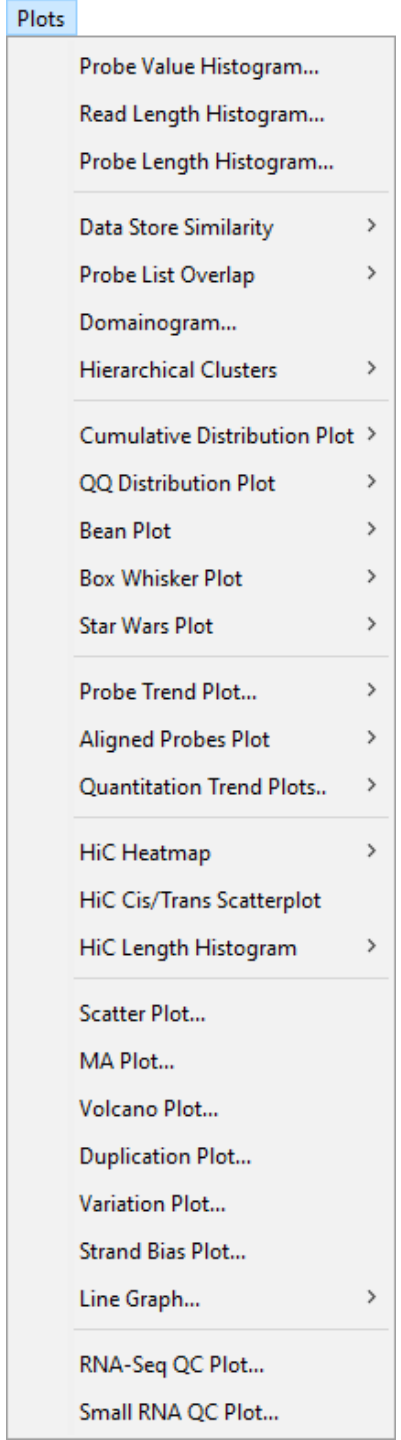
Blue
Fresh quantitation from the raw read data

Red
Methods to normalise / scale / adjust the existing quantitation



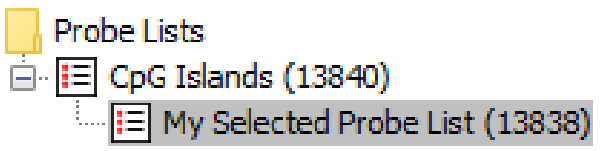
Plotting Figures

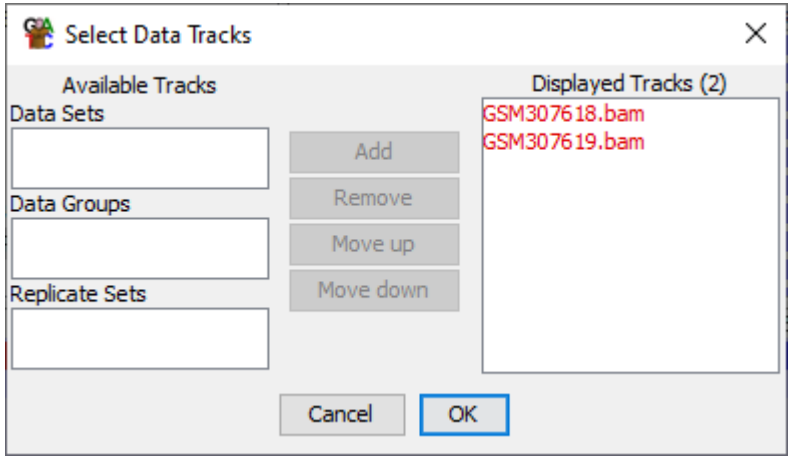
Babraham Bioinformatics

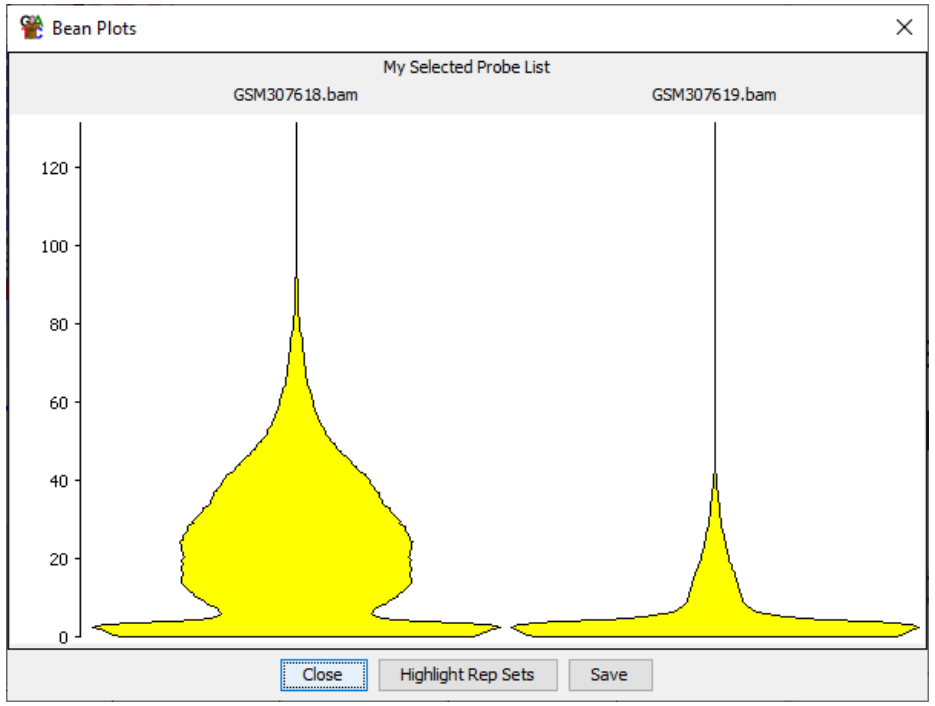


Plotting

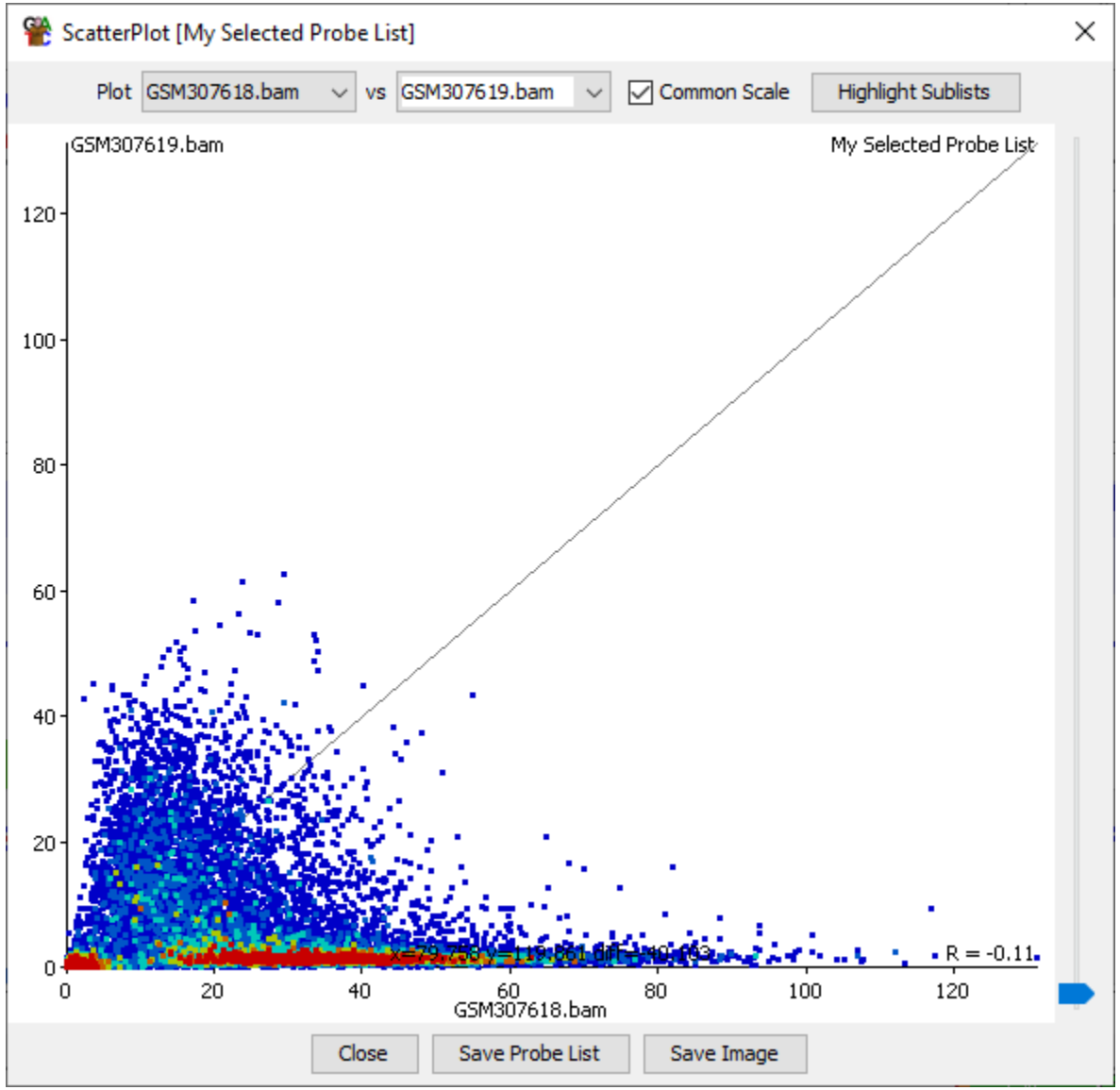
- By Default
  - Uses the data stores shown in the chromosome view
  - Uses the probes in the selected probe list





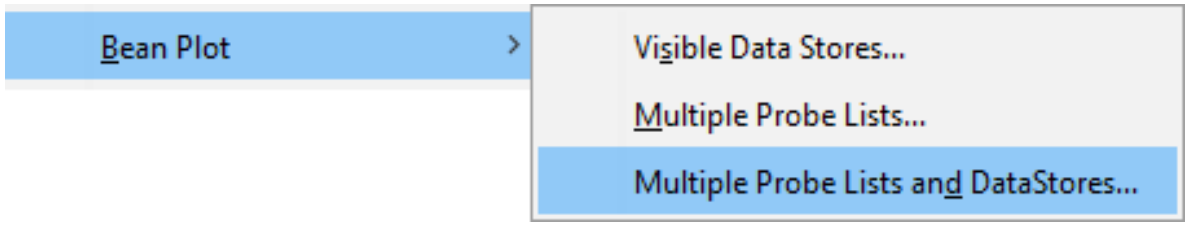


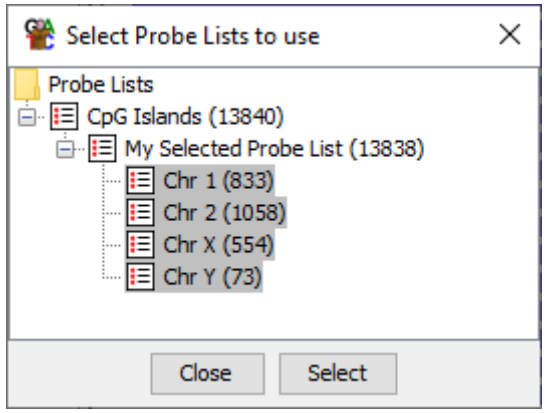
Some Plots are Interactive

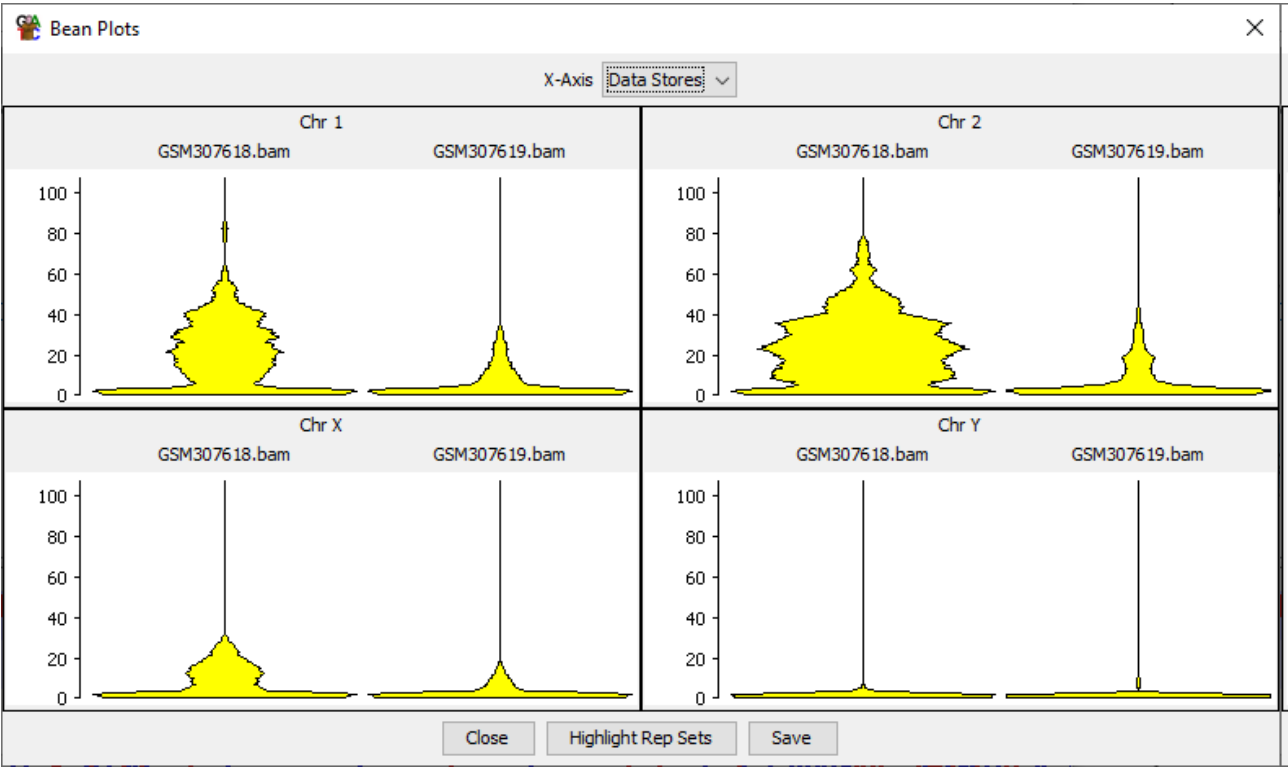


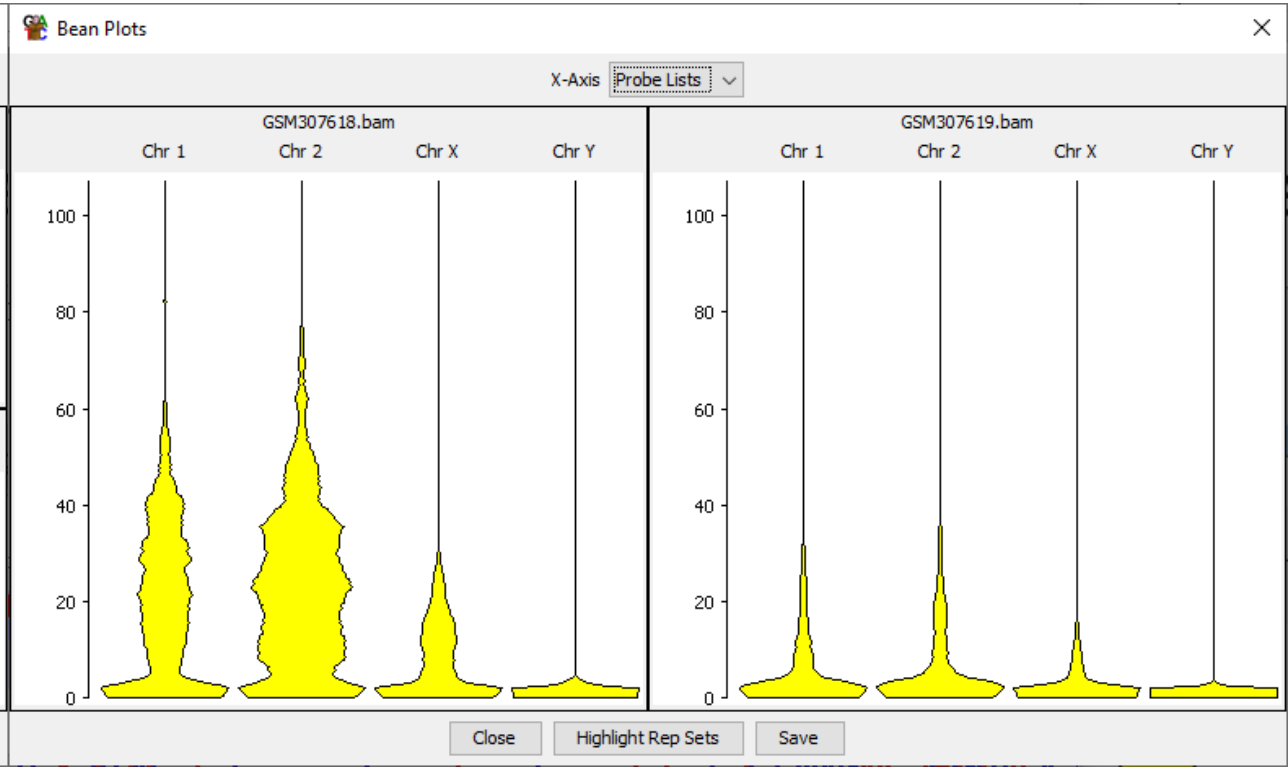
- Hover to see label
- Click to fix label
- Double click to show probe in Chromosome View
- Triple Click to clear labels

Some Plots can be Duplicated







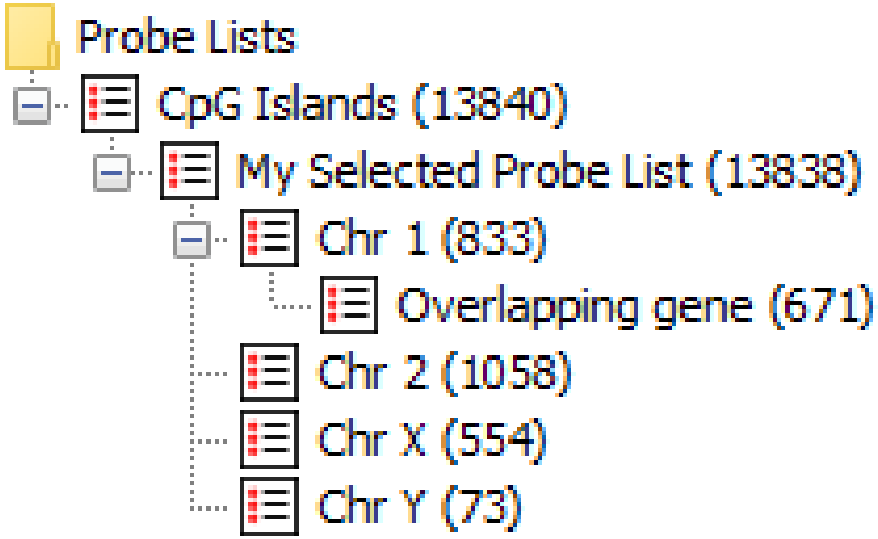


Filtering Probes

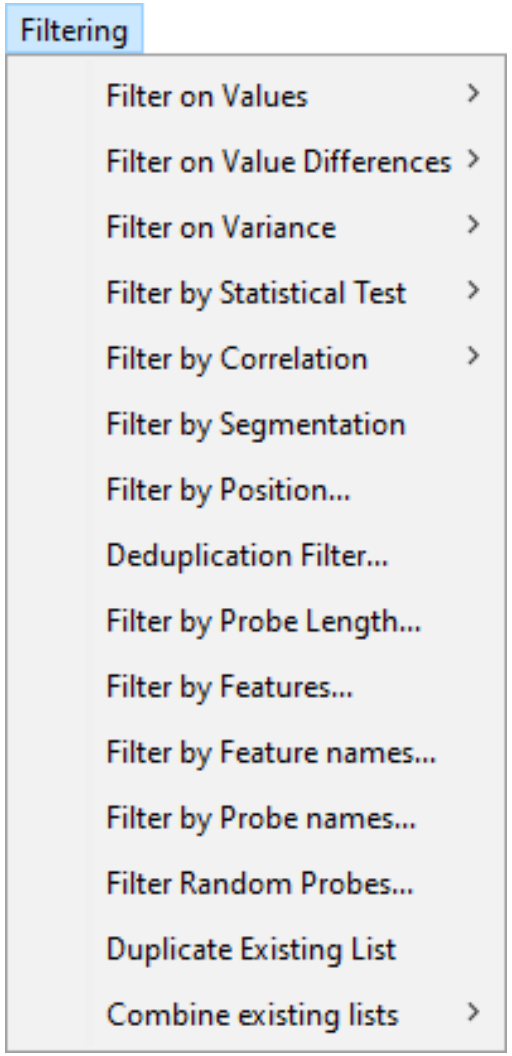
Babraham Bioinformatics

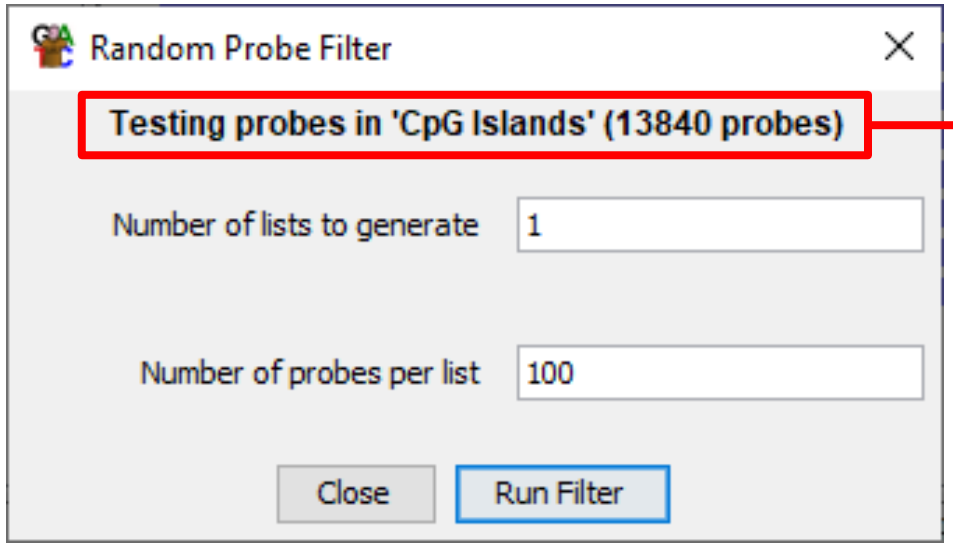
Filtering Concepts

- 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



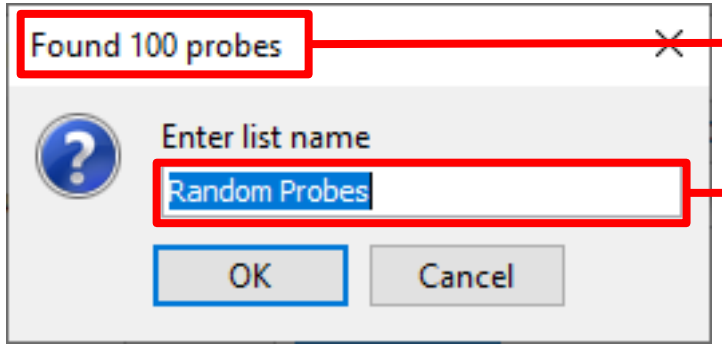
Filters





Starting List

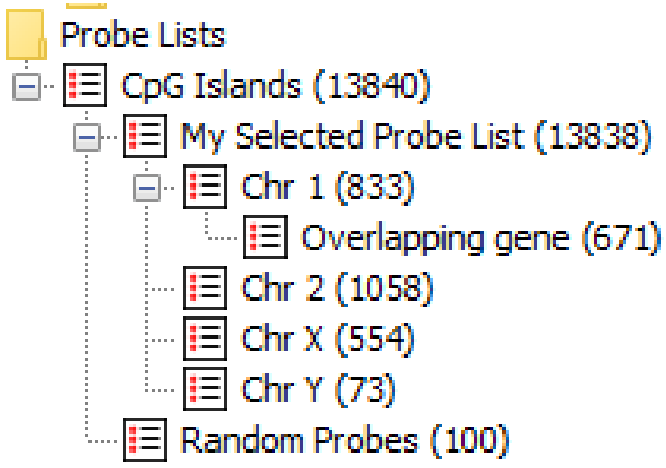
Check this is what you expect



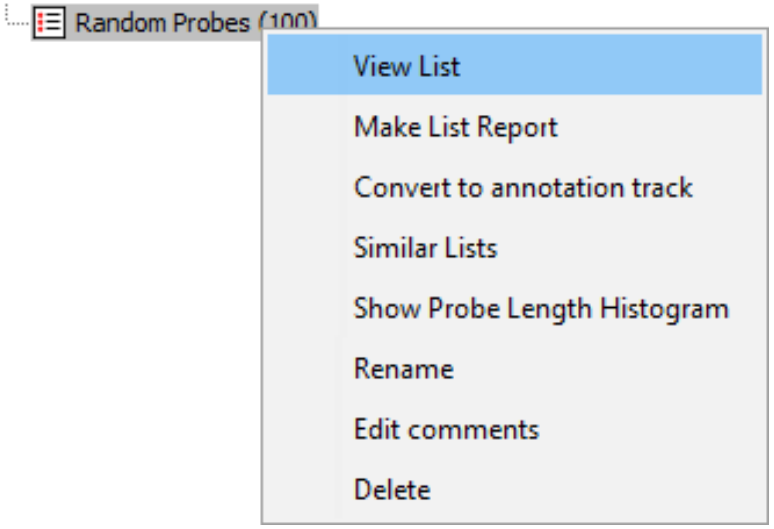
Number of passed probes

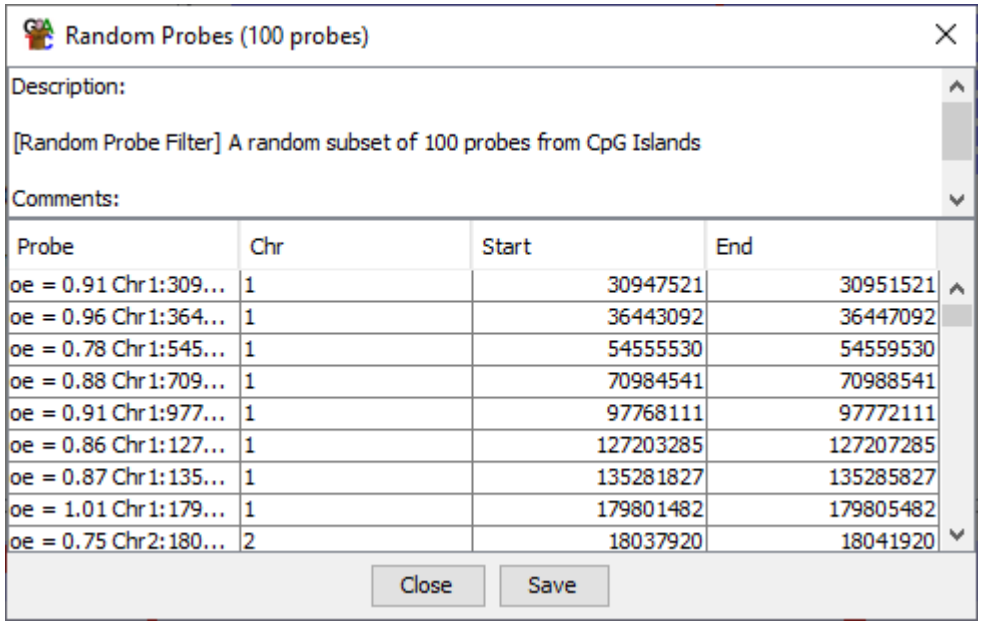
New List Name

Filter Details



- Rename list (doesn't change contents)
- View list (shows options used)
- Delete list (and any children)





Probe	Chr	Start	End
oe = 0.91 Chr1:309...	1	30947521	30951521
oe = 0.96 Chr1:364...	1	36443092	36447092
oe = 0.78 Chr1:545...	1	54555530	54559530
oe = 0.88 Chr1:709...	1	70984541	70988541
oe = 0.91 Chr1:977...	1	97768111	97772111
oe = 0.86 Chr1:127...	1	127203285	127207285
oe = 0.87 Chr1:135...	1	135281827	135285827
oe = 1.01 Chr1:179...	1	179801482	179805482
oe = 0.75 Chr2:180...	2	18037920	18041920

Saving, Reporting and Vistories

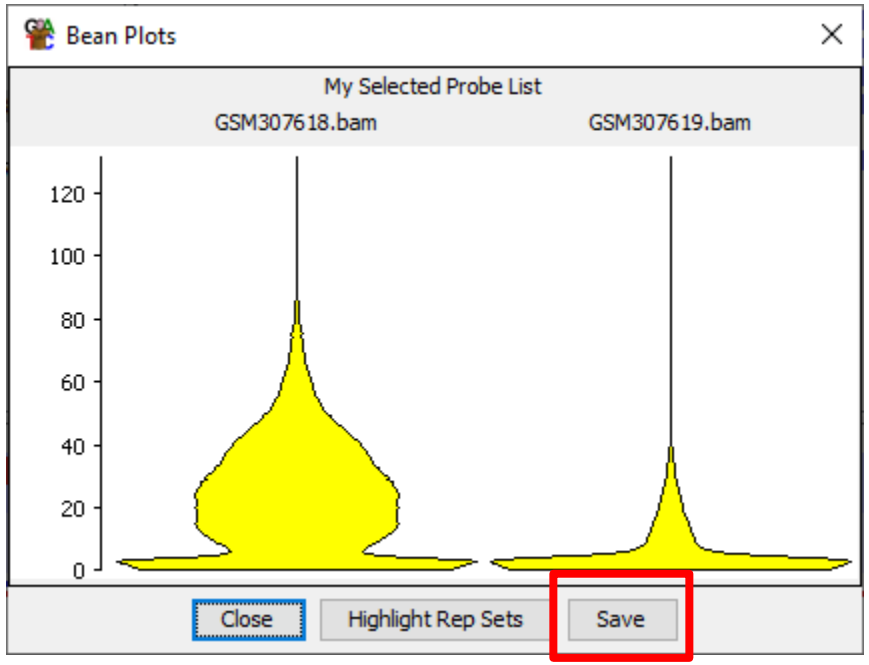
Babraham Bioinformatics

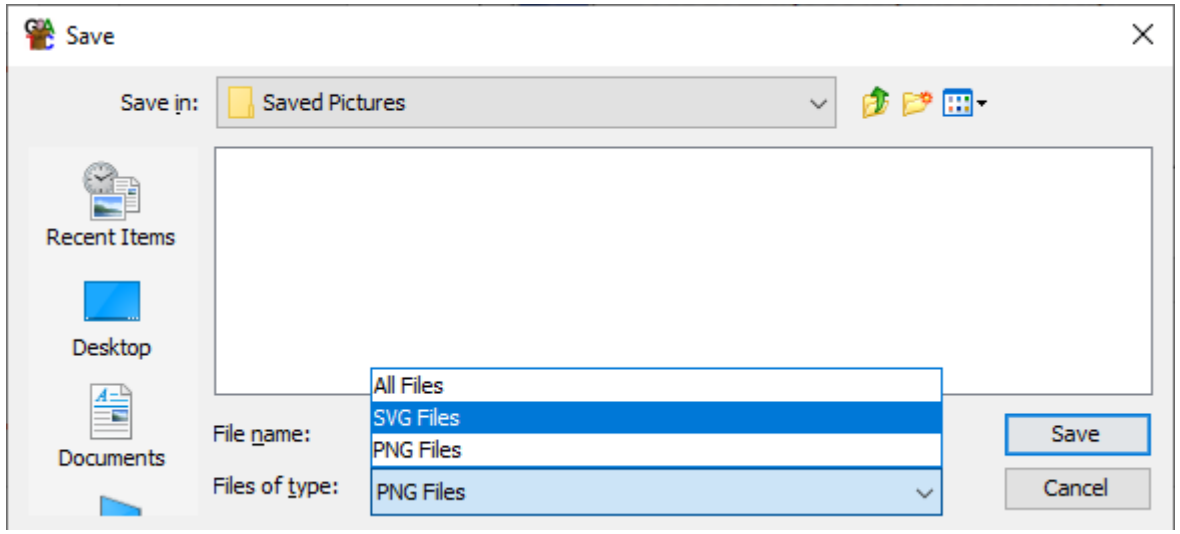
Saving SeqMonk Projects

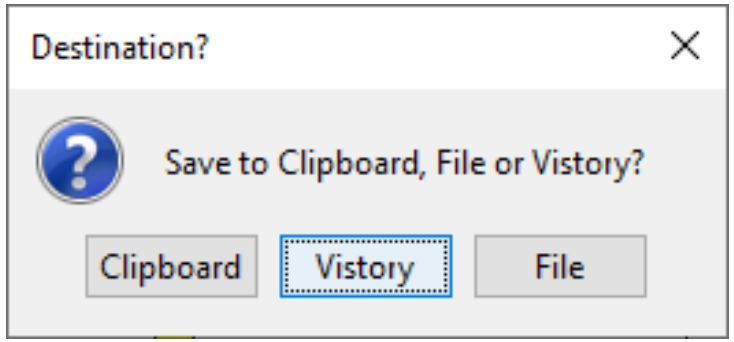
- File > Save Project
- Saves Everything
  - Data
  - Quantitations
  - Probes / Filters
  - Current View
- Single file with .smk file extension
- Can be moved to another machine and opened*

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

Saving Images

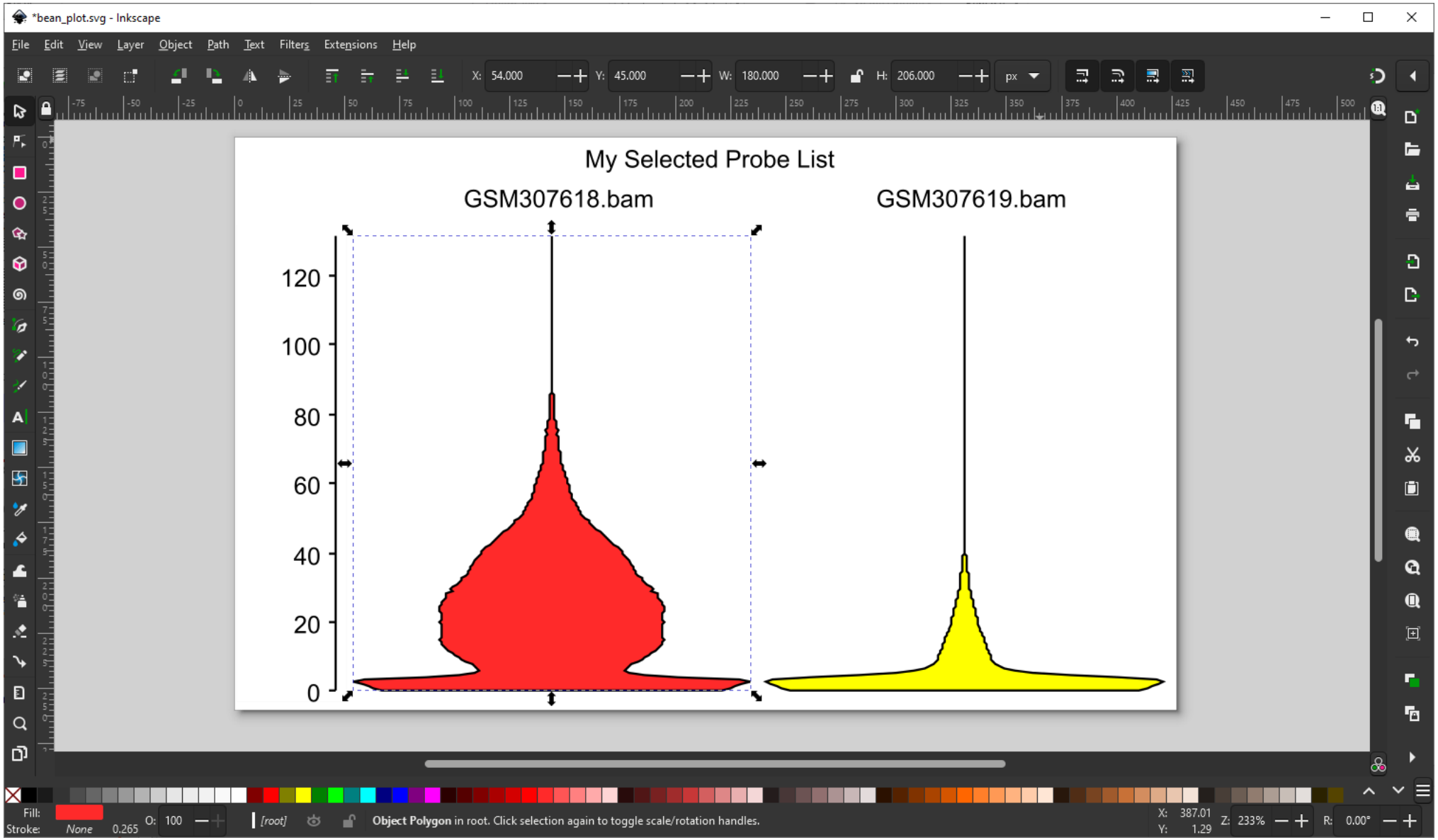






- PNG – Bitmap – Screenshot
- SVG – Vector – Editable

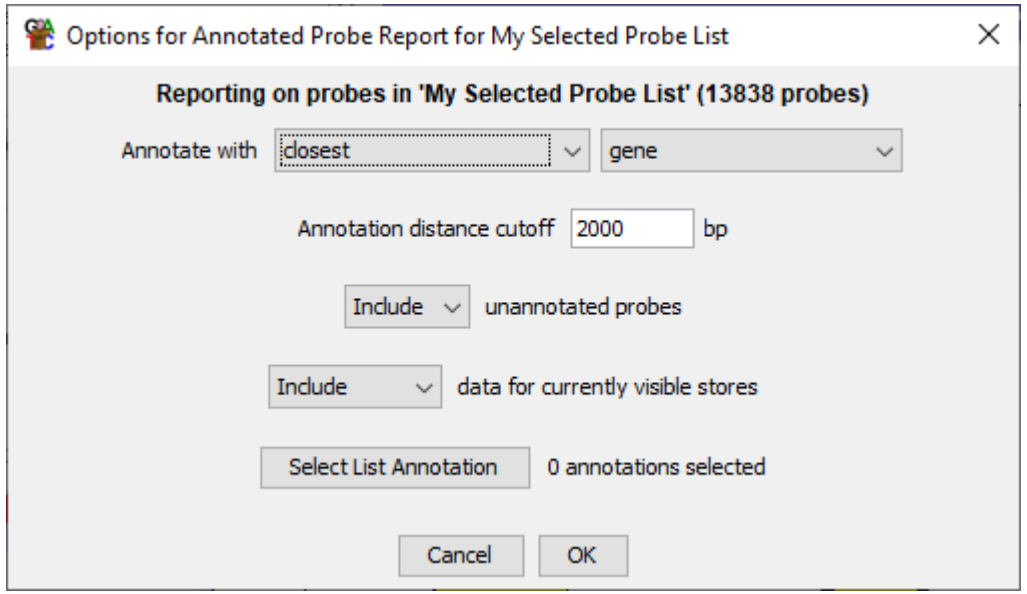
Editing SVG Images



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)

Annotated Probe Report for My Selected Probe List

Probe	Chromosome	Start	End	Probe Strand	Feature	ID	Description	Feature Str...	Type	Feature Ori...	Distance	GSM307618...	GSM30761...
oe = 0.70	1	3669388	3673388	+	Xkr4	ENSMUSG00...	X-linked Kx bl...	-	gene	overlapping	0	22.471	11.923
oe = 0.77	1	4490687	4494687	+	Sox17	ENSMUSG00...	SRY (sex det...	-	gene	overlapping	0	33.279	22.162
oe = 1.00	1	4495278	4499278	+	Sox17	ENSMUSG00...	SRY (sex det...	-	gene	overlapping	0	9.845	27.346
oe = 0.61	1	4558325	4562325	+						Not found	0	0.214	0
oe = 0.78	1	4565280	4569280	+						Not found	0	0.428	0.518
oe = 0.77	1	4569858	4573858	+						Not found	0	12.948	6.61
oe = 0.80	1	4783595	4787595	+	Mrpl15	ENSMUSG00...	mitochondrial...	-	gene	overlapping	0	49.223	0.778
oe = 0.92	1	4805831	4809831	+	Lypla1	ENSMUSG00...	lysophospholi...	+	gene	overlapping	0	27.073	0.648
oe = 0.81	1	4855919	4859919	+	Gm37988	ENSMUSG00...	predicted ge...	+	gene	overlapping	0	33.921	1.296
oe = 0.76	1	5017134	5021134	+	Rgs20	ENSMUSG00...	regulator of ...	-	gene	overlapping	0	27.073	25.142
oe = 0.82	1	5081288	5085288	+	Atp6v1h	ENSMUSG00...	ATPase, H+ ...	+	gene	overlapping	0	15.944	0.518
oe = 0.67	1	5348726	5352726	+						Not found	0	0.856	0.13

Close | Save to File | Save to Vistory

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

Vistory

Load Save Save As Title Subtitle Text Paste Image Clear Summary Save HTML

DataSet Added
GSM307618.bam E:\Illumina Analysis\SeqMonk_Course_Data\GSM307618.bam Library type Single End Dedup=No MAPQ>=20 Primary alignments only.

DataSet Added
GSM307619.bam E:\Illumina Analysis\SeqMonk_Course_Data\GSM307619.bam Library type Single End Dedup=No MAPQ>=20 Primary alignments only.

New Probe Set:CpG Islands (13840 probes)
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

New Probe List: My Selected Probe List (13838 probes)
[Probe Values Filter] Filter on probes in CpG Islands where exactly 2 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 List: Chr 1 (833 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome 1 on strand All Probes

New Probe List: Chr 2 (1058 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome 2 on strand All Probes

New Probe List: Chr X (554 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome X on strand All Probes

New Probe List: Chr Y (73 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome Y on strand All Probes

New Probe List: Overlapping gene (671 probes)
[Feature Filter] Filter probes in Chr 1 on regions based on gene Over feature relationship is Overlapping

New Probe List: Random Probes (100 probes)
[Random Probe Filter] A random subset of 100 probes from CpG Islands

Vistory

Load Save Save As Title Subtitle Text Paste Image Clear Summary Save HTML

Example Vistory

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

Load the data

DataSet Added
GSM307618.bam E:\Illumina Analysis\SeqMonk_Course_Data\GSM307618.bam Library type Single End Dedup=No MAPQ>=20 Primary alignments only.

DataSet Added
GSM307619.bam E:\Illumina Analysis\SeqMonk_Course_Data\GSM307619.bam Library type Single End Dedup=No MAPQ>=20 Primary alignments only.

Make Probes

New Probe Set:CpG Islands (13840 probes)
Feature generator using CpG islands duplicates removed Centered on feature from 2000-2000

Quantitate

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

Filter

New Probe List: My Selected Probe List (13838 probes)
[Probe Values Filter] Filter on probes in CpG Islands where exactly 2 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

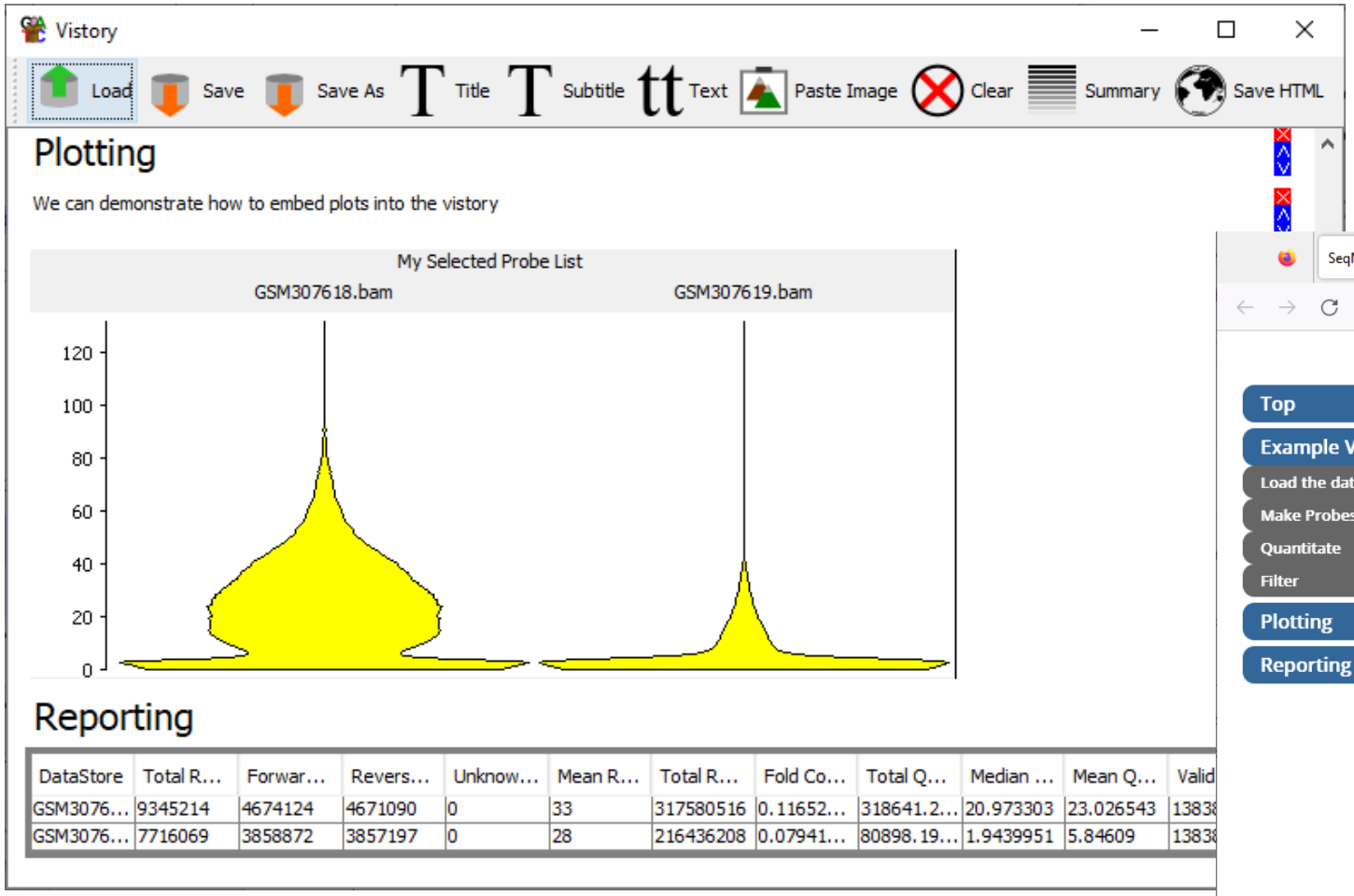
New Probe List: Chr 1 (833 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome 1 on strand All Probes

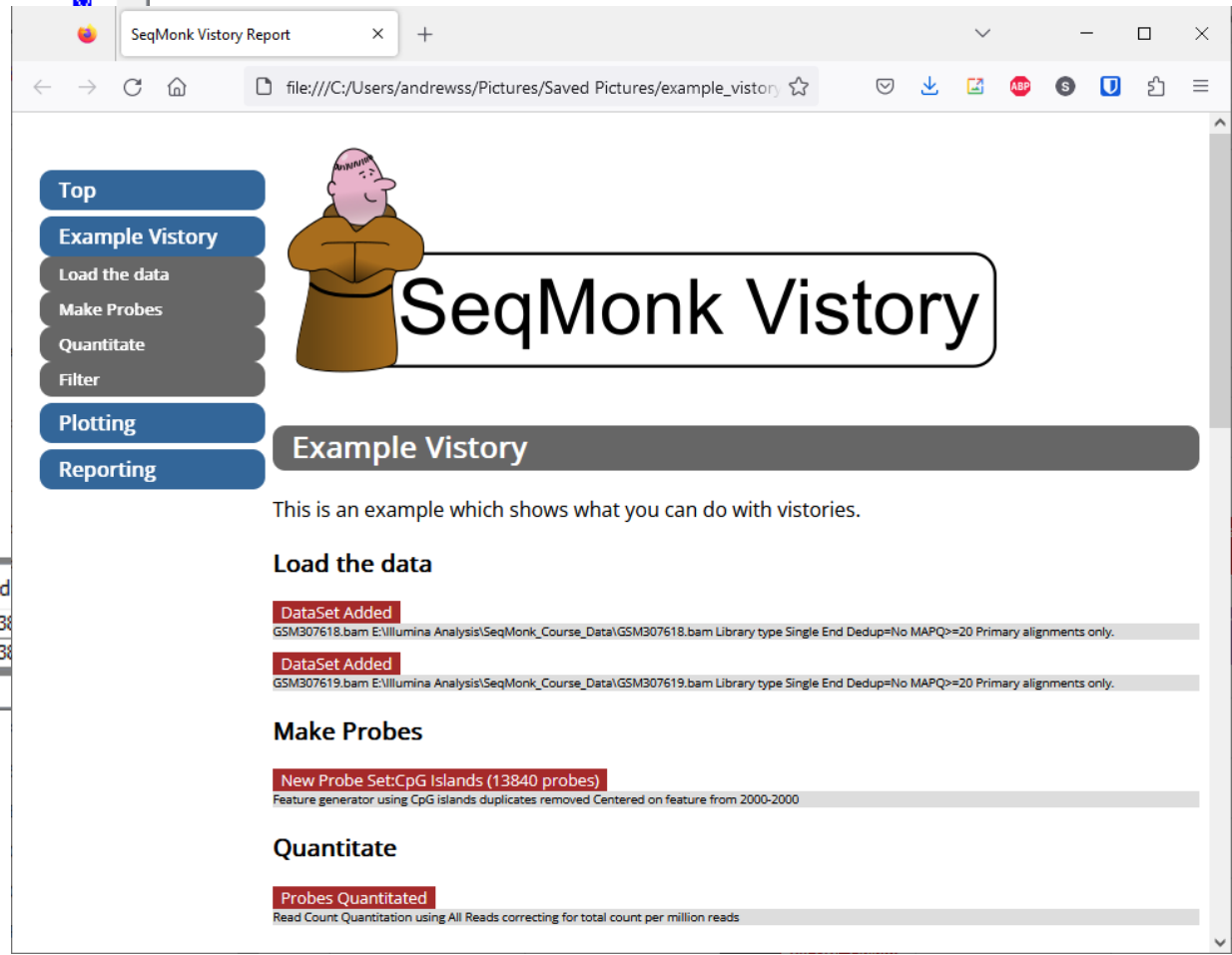
New Probe List: Chr 2 (1058 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome 2 on strand All Probes

New Probe List: Chr X (554 probes)
[Position Filter] Probes from My Selected Probe List which are on chromosome X on strand All Probes

New Probe List: Chr Y (73 probes)

Vistories





- Save as .smv file (editable)
- Save as HTML (report)

https://www.bioinformatics.babraham.ac.uk/vistorydb/