SeqMonk tools for methylation analysis

Simon Andrews simon.andrews@babraham.ac.uk @simon_Andrews 2016-11



SeqMonk





SeqMonk Data Model

- Conventional data (ChIP-Seq, RNA-Seq etc)
 - Data is reads (BAM files etc)
 - Strand indicates genomic strand
- BS-Seq and related data
 - Data is methylation calls
 - All 'reads' are 1bp in length
 - Strand indicates meth state (+=meth -=unmeth)
 - Original strand comes from the imported file



Raw Data

Mus musculus NCBIM37 chr7:88379990-88381219 (1.2 kbp)



Red = Meth Blue = Unmeth



Basic Movement Controls

- Move left right
 - Drag bottom scrollbar
 - Mouse scroll wheel
 - Left/Right arrows
- Zoom In
 - Drag a box and release
 - Up arrow
- Zoom Out
 - Right mouse button
 - Down arrow
- Find a feature
 - Edit > Find Feature
 - Control+F
- Change chromosome
 - Edit > Goto Position
 - Drag a box in the genome view





Raw Data Display



Bioinformatics

View > Data Track Display

Quantitation Model

- Probe = Location to make a measurement
- ProbeSet = Collection of probes
- Quantitation associates a value with each probe for each data set.
- Define Probes > Quantitate Probes > Visualise



Probe Generation



Targeting Measurement Features

😤 Define Probes			
Probe Generator Options			
Running Window Generator	Features to design around	CGI 👻	
Feature Probe Generator Feature Percentile Probe Generator Contig Probe Generator Even Coverage Probe Generator	Split into subfeatures (exons)		
Read Position Probe Generator Random Position Generator	Remove exact duplicates		
Existing Probe List Generator Interstitial Probe Generator Deduplication Probe Generator	Ignore feature strand information		
Merge Consecutive Probe Generator Current Region Generator	Make probes	Centered on feature From - 2000 to + 2000 bp	
Close Create Probes			

Data > Define Probes > Feature Probe Generator



Targeting Measurement Fixed Windows

😤 Define Probes		
Probe Generator Options		
Running Window Generator		
Feature Probe Generator	Window Size (bp)	5000
Feature Percentile Probe Generator		
Contig Probe Generator		
Even Coverage Probe Generator		
MACS peak caller		
Read Position Probe Generator	Step Size (bp)	2500
Random Position Generator		
Existing Probe List Generator		
Deduplication Probe Cenerator		
Merge Consecutive Probe Generator		
Current Region Generator	Limit by region	C Active Probe List
	Close	Create Probes

Data > Define Probes > Running Window Generator



Targeting Measurement Fixed number of calls

Probe Generator Options	Stores to use		
Running Window Generator	Oocyte_reimport		
Feature Probe Generator	Serum_reimport		
Feature Percentile Probe Generator			
Contig Probe Generator			
Even Coverage Probe Generator			
MACS peak caller			
Read Position Probe Generator			
Random Position Generator			
Existing Probe List Generator Interstitial Probe Generator	Use reads on strand	All Reads	
Deduplication Probe Generator	Ignore duplicate reads		
Merge Consecutive Probe Generator Current Region Generator	Target read count	1000	
	Max Probe Size (bp)	0	

Data > Define Probes > Even Coverage Probe Generator



Targeting Measurement Fixed number of call positions

😤 Define Probes		—
Probe Generator Options Running Window Generator Feature Probe Generator Feature Percentile Probe Generator Contig Probe Generator Even Coverage Probe Generator MACS peak caller Read Position Probe Generator	Stores to use Oocyte_reimport Serum_reimport	
Random Position Generator Existing Probe List Generator Interstitial Probe Generator Deduplication Probe Generator Merge Consecutive Probe Generator Current Region Generator	Use reads on strand Minimum read count per position Valid positions per window Limit to currently visible region Ignore Strand	All Reads I 50
	Close Create Probes	

Data > Define Probes > Read Position Probe Generator



Quantitation



Methylation Measurement Simple percentage of all calls

😤 Define Quantitation					X
Quantitation Options Enrichment Quantitation Read Count Quantitation Base Pair Quantitation Exact Overlap Count Quantitation Difference Quantitation % Coverage Quantitation Coverage Depth Quantitation Distance to Feature Quantitation Probe Length Quantitation Eived Value Quantitation		Calculate difference as Ignore duplicates	Forward Only	✓ As Percentage of	▼ All Reads
Only quantitate visible stores					
Ciose Quantitate					

Data > Quantitate Existing Probes > Difference Quantitation



Methylation Measurement More complex corrected measure

😤 Define Quantitation		X
Quantitation Options RNA-Seq quantitation pipeline Active transcription quantitation pipeline	Features to quantitate	[Existing Probes]
Gene trap quantitation pipeline Wiggle Plot for Initial Data Inspection	Minimum count to include position	1
Bisulphite methylation over features Splicing efficiency quantitation	Apply min count over all stores	
Antisense transcription pipeline Codon Bias Pipeline	Minimum observations to include feature	10
Only quantitate visible stores	Combined value to report	Mean 👻
	Close Run Pipeline	

Data > Quantitation Pipelines > Bisulphite methylation over features



Visualisation of quantitated methylation



7,500,000

7,550,000

7,600,000

7,650,000

7,700,000

450,000

Bioinformatics

Distributions



Plots > Probe value histogram



Plots > Cumulative Distribution Plot





Plots > Beanplot

Comparisons



Plots > Scatter plot



Plot is interactive – mouse over a point for information Double click on a point to move there in the chromosome view 18

Trend Plots

Plots > Quantitation Trend Plot

📽 Quantitation Trend Preferences		×
Features to design around	gene	•
Split into subfeatures (exons)		
Remove exact duplicates		
Ignore feature strand information		
Make probes	Over feature From - 5000 to + 5000	bp
[Cancel Create Plot	



Clustering

Correlation based (per probe normalised)

Euclidean





Plots > Hierarchical Clusters