

Ginkgo—Interactive analysis and quality assessment of single-cell CNV data

 @RobAboukhalil

Robert Aboukhalil, Tyler Garvin, Jude Kendall, Timour Baslan,
Gurinder S. Atwal, Jim Hicks, Michael Wigler, Michael C. Schatz



Outline

Introduction



Ginkgo



**Comparison of
WGA methods**



Outline

Introduction



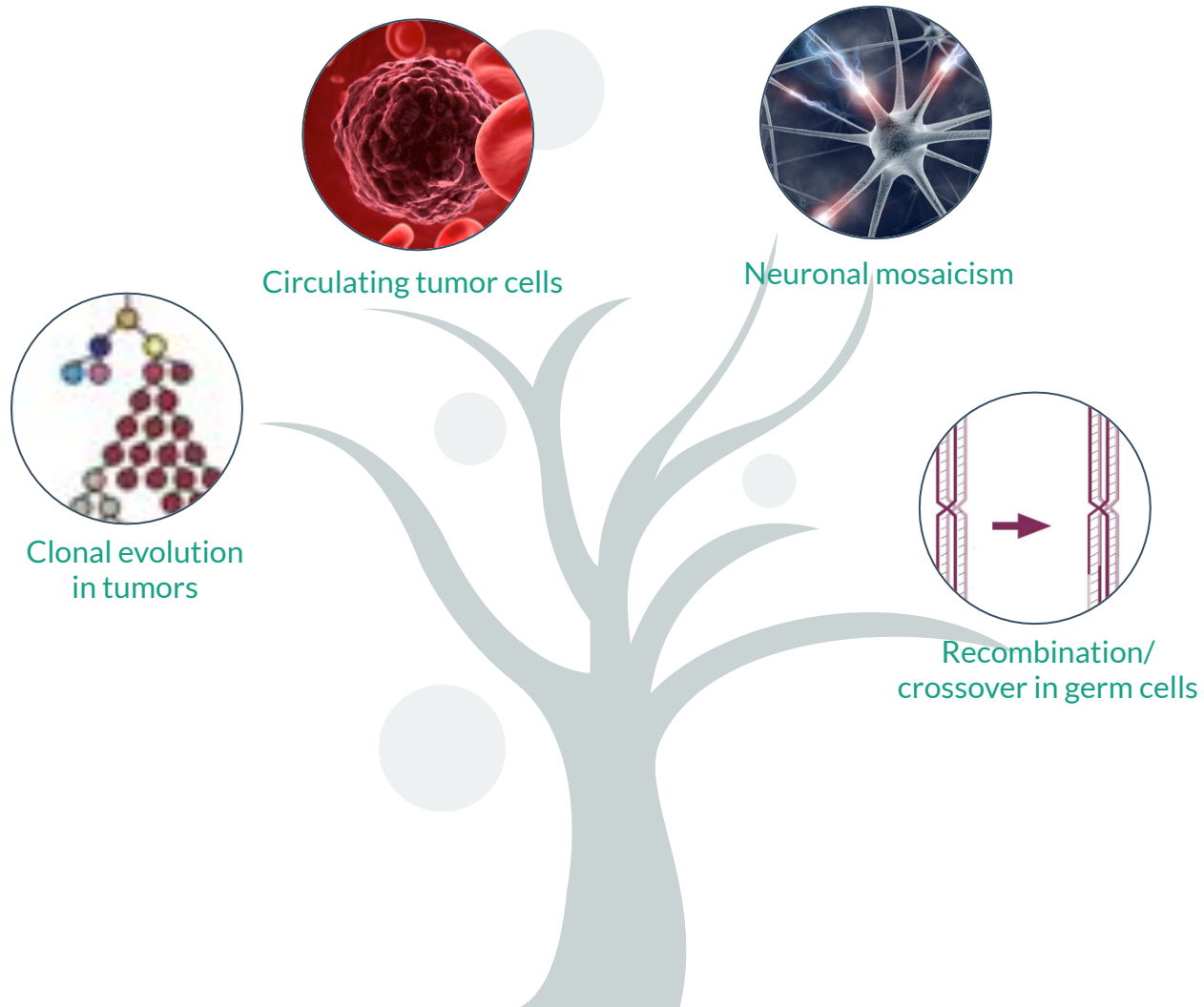
Ginkgo



Comparison of
WGA methods



Single-cell sequencing

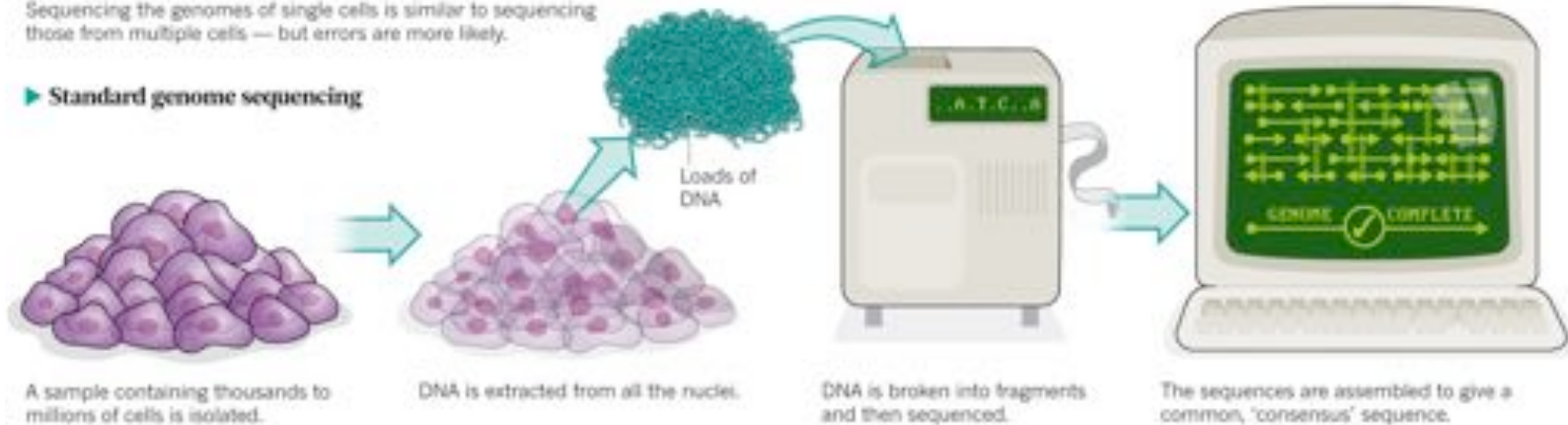


Single-cell vs. bulk sequencing

ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

► Standard genome sequencing

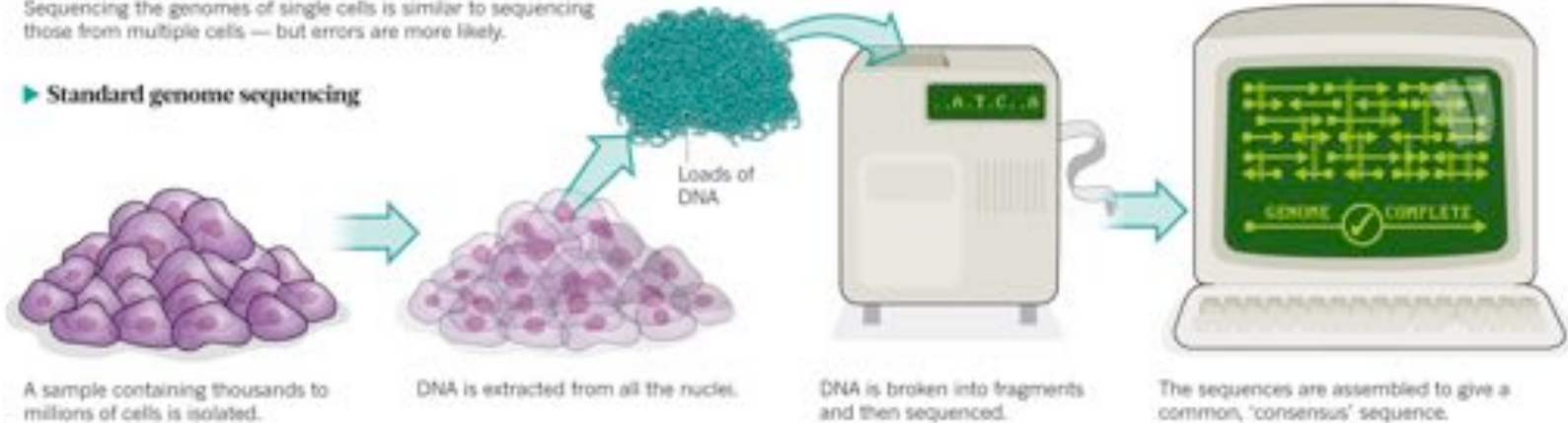


Single-cell vs. bulk sequencing

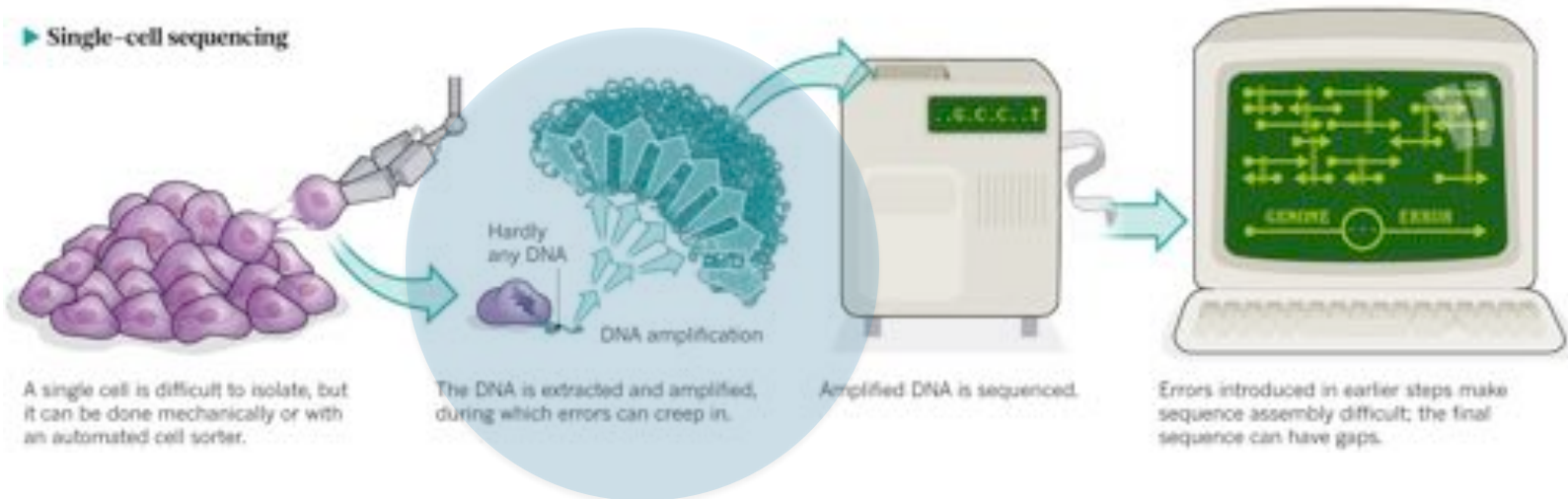
ONE GENOME FROM MANY

Sequencing the genomes of single cells is similar to sequencing those from multiple cells — but errors are more likely.

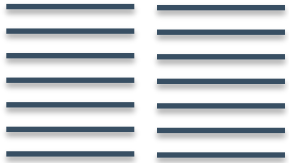
► Standard genome sequencing



► Single-cell sequencing



Copy-number variant analysis



Low coverage allows us to study copy-number variants

<1X coverage, often <0.1X

Copy-number variant analysis



Low coverage allows us to study copy-number variants

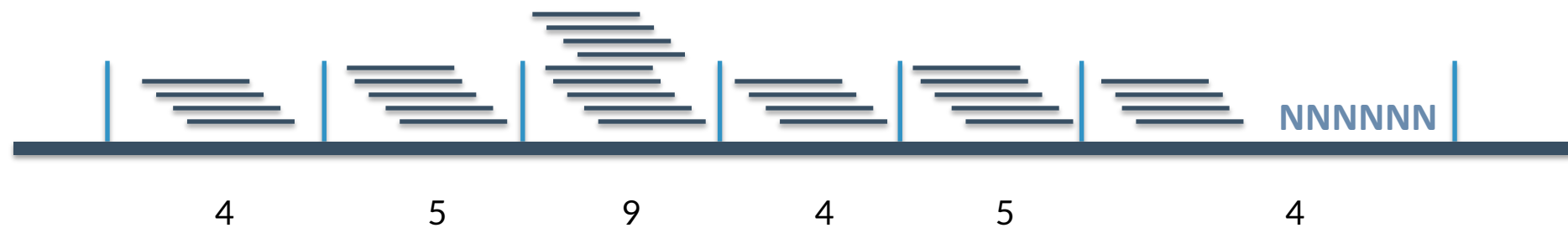
<1X coverage, often <0.1X

Copy-number variant analysis

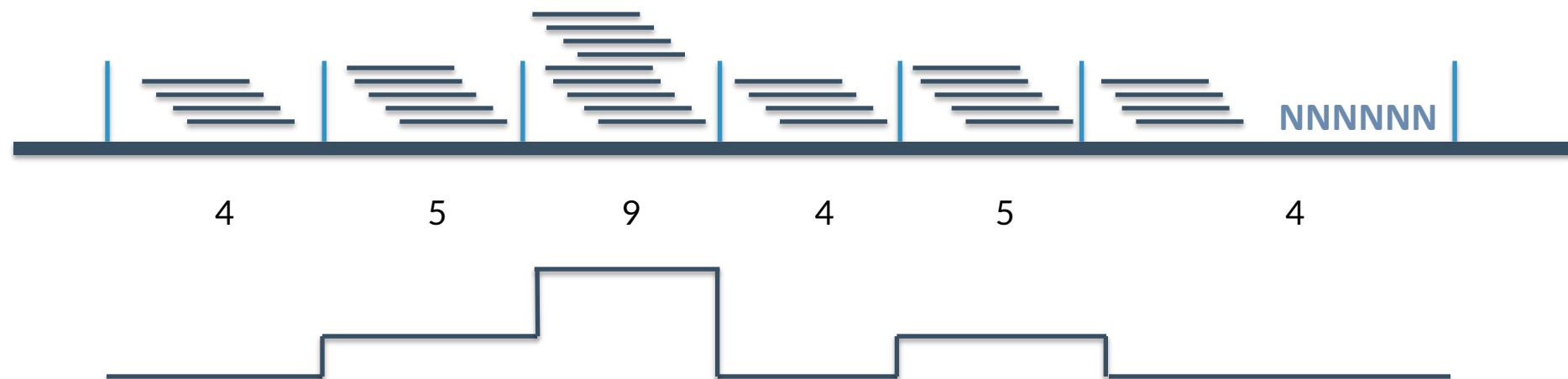


Divide genome into “bins” with ~50 – 100 reads / bin

Copy-number variant analysis



Copy-number variant analysis



Copy-number variant analysis



4

5

9

4

5

4



Circular Binary Segmentation (CBS)
to reduce noise in data

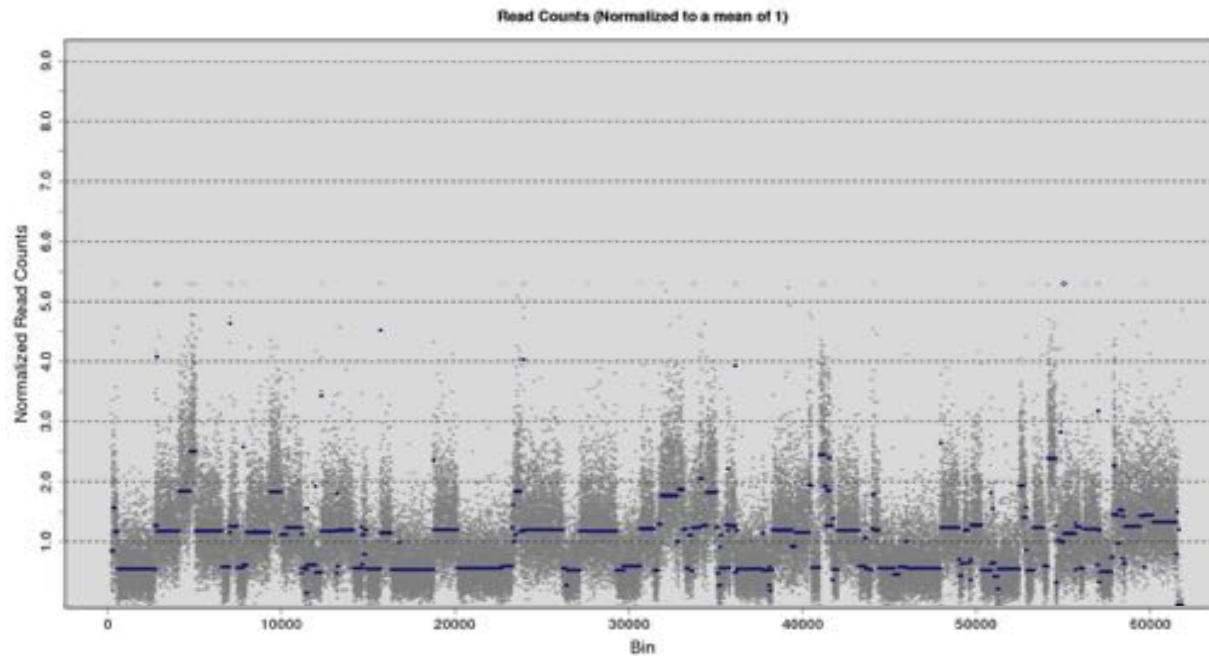
4.5

9

4



Copy-number variant analysis



We can estimate integer copy-number states by scaling the profile and minimizing the sum of squares error

Outline

Introduction



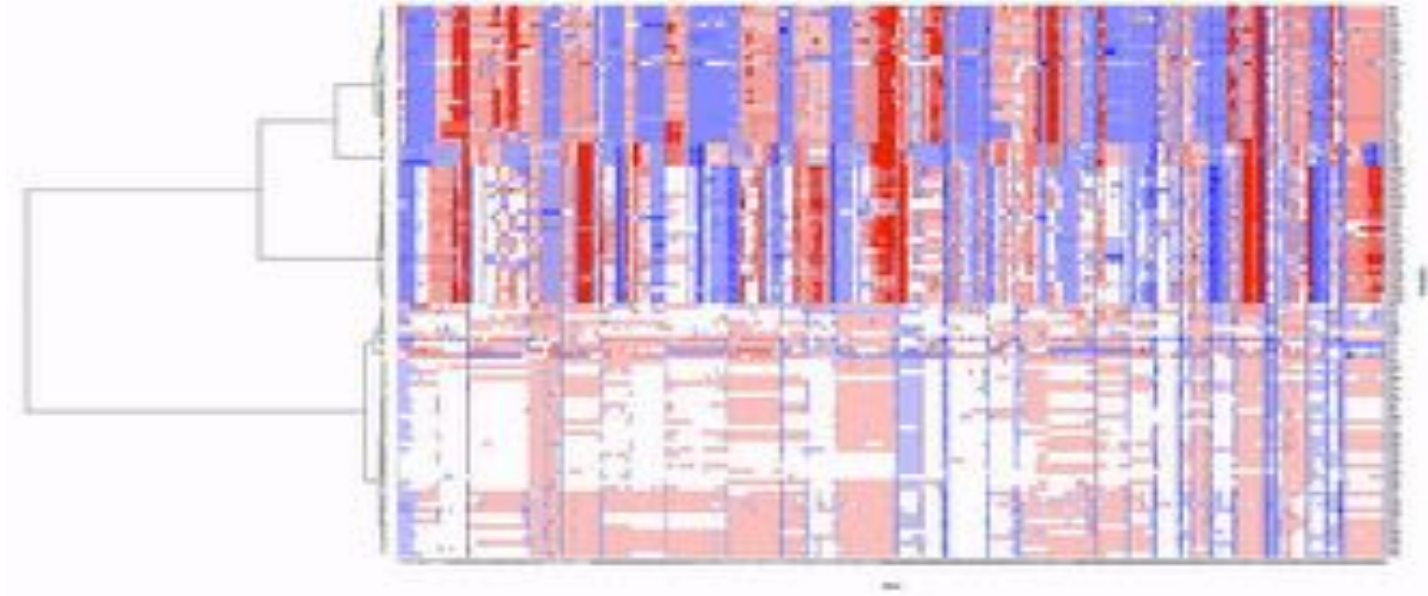
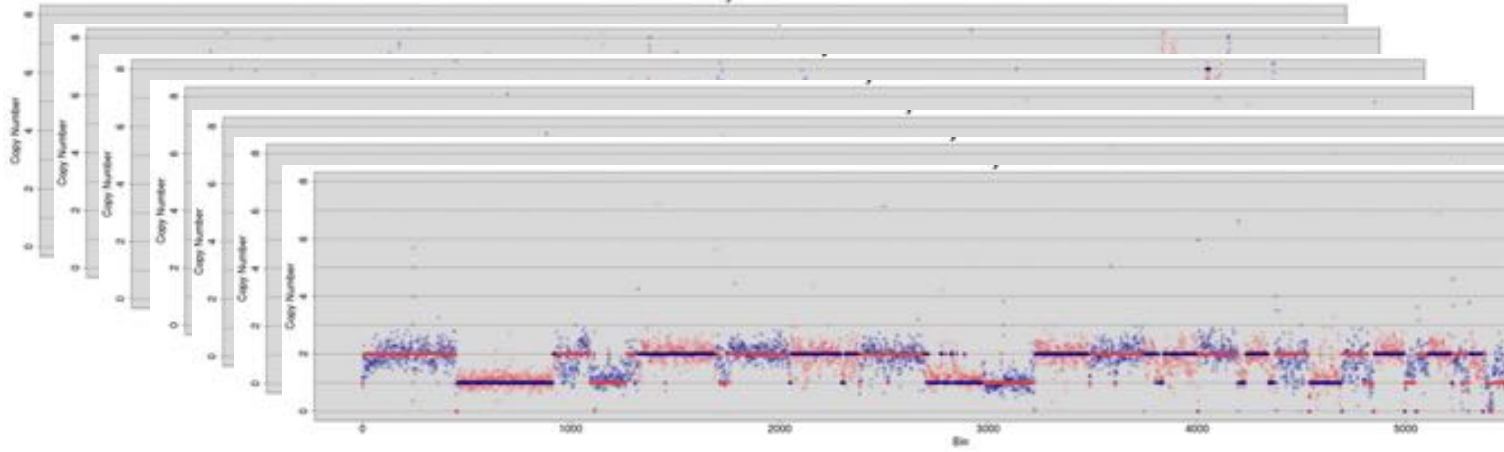
Ginkgo



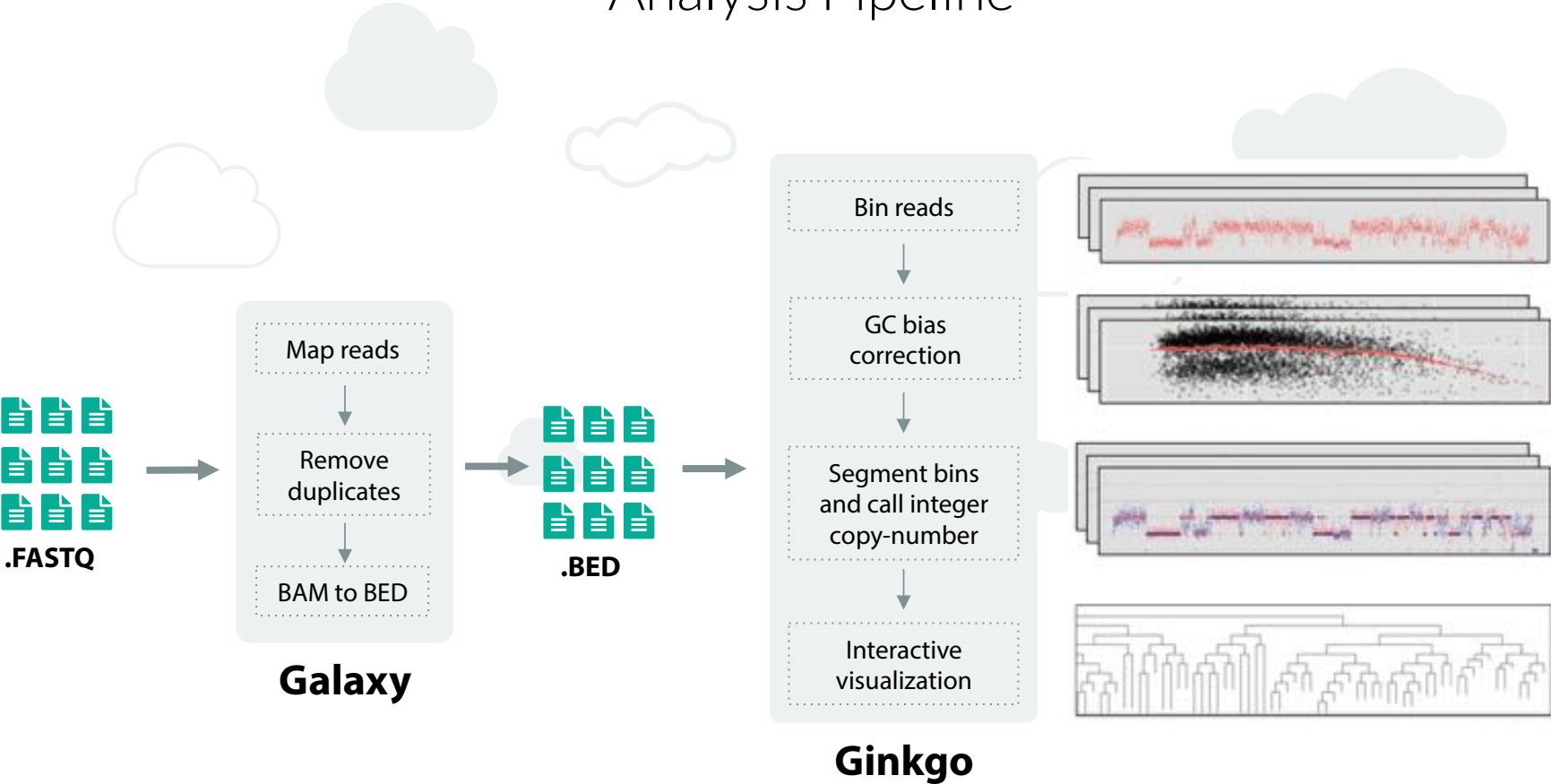
Comparison of
WGA methods



Copy-number variant analysis



Analysis Pipeline



Ginkgo Demo

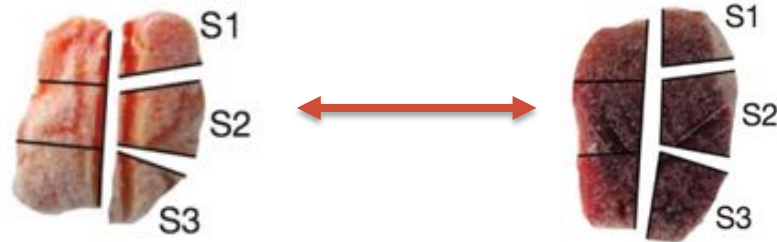
Sample dataset

LETTER

doi:10.1038/nature09807

Tumour evolution inferred by single-cell sequencing

Nicholas Navin^{1,2}, Jude Kendall¹, Jennifer Troge¹, Peter Andrews¹, Linda Rodgers¹, Jeanne McIndoo¹, Kerry Cook¹, Asya Stepansky¹, Dan Levy¹, Diane Esposito¹, Lakshmi Muthuswamy³, Alex Krasnitz¹, W. Richard McComble¹, James Hicks¹ & Michael Wigler¹



Triple-negative (ER⁻, PR⁻, HER2⁻) ductal carcinoma

Ginkgo

qb.cshl.edu/ginkgo/?q=/NHeshEpiCXcE7hqmLx2c

Ginkgo

A web tool for analyzing single-cell sequencing data.

Sample analyses - Load previous analysis -

STEP 1 Upload your .bed files (We accept *.bed and *.bed.gz, max 1GB/file, min 3 cells)

+ Add file... Cancel upload

382.21 Mbb/s | 00:00:04 | 56.03 % | 270.04 MB / 481.98 MB

File Name	Size	Progress	Action
M16_SRR089711.bed.gz	56.85 MB	100%	Delete
M16_SRR090212.bed.gz	41.11 MB	100%	Delete
P16_SRR089688.bed.gz	55.07 MB	100%	Remove
P16_SRR089693.bed.gz	51.54 MB	~10%	Cancel
P16_SRR089604.bed.gz	43.70 MB	100%	Delete
P16_SRR089646.bed.gz	43.32 MB	~10%	Cancel
P16_SRR089659.bed.gz	40.86 MB	~10%	Cancel

Next step >

View analysis later

Access your results later at the following address:

<https://qb.cshl.edu/ginkgo/?q=/NHeshEpiCXcE7hqmLx2c>

Help

Sample .bed file

chrom	chromStart	chromEnd
chr1	555485	555533
chr1	676584	676632
chr1	745138	745184

How to make .bed files

If your mapped reads are saved in the file `reads.bam`:

```
bamToBed -i reads.bam > reads.bed
```

Detailed instructions

Ginkgo -

Ginkgo

Single-cell analysis

3% complete.

Step 1: 10% Mapping reads to bins... (M16 SRR090212 bed.gz)

View results

[← Analysis Options](#)

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NHeoHlEpiCXcE7hqmLx2c>

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHestHpiCXcE7hqmLx2c

Ginkgo

Single-cell analysis

95% complete.

Step 2: 85% Calling copy number events... (Computing Cluster (Copy Number))

STEP 2 View results

← Analysis Options

View analysis later

Access your results later at the following address:

`https://qb.cshl.edu/ginkgo/?q=results/NHestHpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (variable/100kb reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Ginkgo Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
P16_SRR089586
 P16_SRR089646
 P16_SRR089645
 P16_SRR089644
 P16_SRR089643
 P16_SRR089642
 P16_SRR089641
 P16_SRR089640
 P16_SRR089639
 P16_SRR089638
 P16_SRR089637
 P16_SRR089636
 P16_SRR089635
 P16_SRR089634
 P16_SRR089633
 P16_SRR089632
 P16_SRR089631
 P16_SRR089630
 P16_SRR089629
 P16_SRR089628
 P16_SRR089627
 P16_SRR089626
 P16_SRR089625
 P16_SRR089624
 P16_SRR089623
 P16_SRR089622
 P16_SRR089621
 P16_SRR089620
 P16_SRR089619
 P16_SRR089618
 P16_SRR089617
 P16_SRR089616
 P16_SRR089615
 P16_SRR089614
 P16_SRR089613
 P16_SRR089612
 P16_SRR089611
 P16_SRR089610
 P16_SRR089609
 P16_SRR089608
 P16_SRR089607
 P16_SRR089606
 P16_SRR089605
 P16_SRR089604

View analysis later

Access your results later at the following address:

<http://q5.cshl.edu/ginkgo/>
www.11e/70wuk8p1C3dFT9q642e

Parameters

Clustering: variable bin of 500kb size (see below)
 Segmentation: using normalized read counts
 Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts (rawick | sm | pdf | png)
 Copy-number (rawick | sm | pdf | png)
 Correlations (rawick | sm | pdf | png)

Download processed data

Statistics: Bin count statistics for each cell (1 KB)
Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (283 KB)
Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (283 KB)
Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.48 MB)

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
<input type="checkbox"/> M16_SRR089711		15,311,446	2,855.02	1,023.36	0.36
<input type="checkbox"/> M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
<input type="checkbox"/> M16_SRR090144		11,757,180	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,582,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	656.59	0.55
<input type="checkbox"/> P16_SRR089586		1,308,333	1,852.61	1,023.36	0.55

Diploids from both primary and metastasis cluster

Ginkgo Single-cell analysis

Analysis complete!

View results

Tree

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NH5eHlEpiCXxE7hqmLx2c>

Bin size: 500kb

Segmentation: using normalized read counts

Clustering: ward linkage, euclidean distance

Tree display

- Normalized read counts [view] [xml] [pdf] [png]
- Copy number [view] [xml] [pdf] [png]
- Correlations [view] [xml] [pdf] [png]

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (283 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (283 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.48 MB)

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
<input type="checkbox"/> M16_SRR089711		15,311,446	2,855.02	1,023.36	0.36
<input type="checkbox"/> M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
<input type="checkbox"/> M16_SRR090144		11,757,180	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,582,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	656.59	0.55
<input type="checkbox"/> P16_SRR089664		1,308,333	1,852.61	1,023.36	0.55

Non-diploids from primary and metastasis segregate

Ginkgo -

q5.cshl.edu/ginkgo/?q=results/NHeshEpiCXxE7hqmLx2c

Ginkgo

Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
 P16_SRR089585
 P16_SRR089646
 P16_SRR089664
 P16_SRR089663
 P16_SRR089659
P16_SRR089662
 M16_SRR090210
 M16_SRR090211
 M16_SRR090212
 M16_SRR090213
 M16_SRR090214
 M16_SRR090215

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NHeshEpiCXxE7hqmLx2c>

Analysis Parameters

Bin size of 500kb size (default) bins using normalized read counts linkage, euclidean distance

Tree display

Normalized read counts (raw) | vml | pdf | png
 Copy number (raw) | vml | pdf | png
 Correlations (raw) | vml | pdf | png

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (282 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.48 MB)

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711		15,311,446	2,855.02	1,023.26	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
M16_SRR090144		11,757,180	2,192.28	1,252.84	0.57
M16_SRR090156		7,582,032	1,415.63	766.63	0.54
M16_SRR090158		6,414,000	1,195.97	656.59	0.55
M16_SRR089662		1,708,733	1,855.61	1,855.61	0.55

Non-diploids from primary and metastasis segregate



Ginkgo Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
 P16_SRR089585
 P16_SRR089646
 P16_SRR089664
 P16_SRR089663
 P16_SRR089659
 P16_SRR089662
 M16_SRR090210
M16_SRR090158
 M16_SRR090212
 M16_SRR090158
 M16_SRR090158

View analysis later

Access your results later at the following address:

<https://gdc.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c>

Analysis Parameters

Genome: hg19

Bin size: 500Kb

Linkage: euclidean

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711		15,311,446	2,855.02	1,023.36	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
M16_SRR090144		11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156		7,582,032	1,415.63	766.63	0.54
M16_SRR090158		6,414,000	1,195.97	656.59	0.55

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (282 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.48 MB)

Non-diploids from primary and metastasis segregate

Ginkgo Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
 P16_SRR089585
 P16_SRR089646
 P16_SRR089664
 P16_SRR089663
 P16_SRR089659
 P16_SRR089662
 M16_SRR090210
 M16_SRR090158
 M16_SRR090132
M16_SRR090144
 M16_SRR090156

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NH5eHlEpiCXcE7hqmLx2c>

Analysis Parameters

Genome: hg19
 Binning: variable bins of 500kb size (see [binning](#) for details)
 Segmentation: using normalized read counts (see [segmentation](#) for details), euclidean distance

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
<input type="checkbox"/> M16_SRR089711		15,311,446	2,855.02	1,023.36	0.36
<input type="checkbox"/> M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
<input type="checkbox"/> M16_SRR090144		11,757,180	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,582,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	656.59	0.55

Normalized read counts (raw) | [vml](#) | [pdf](#) | [png](#)
 Copy number (raw) | [vml](#) | [pdf](#) | [png](#)
 Correlations (raw) | [vml](#) | [pdf](#) | [png](#)

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (282 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (282 KB)

Non-diploids from primary and metastasis segregate



Ginkgo -

q5.cshl.edu/ginkgo/?q=results/NHeshEpiCXxE7hqmLx2c

Ginkgo

Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
 P16_SRR089585
 P16_SRR089646
 P16_SRR089664
 P16_SRR089663
 P16_SRR089659
 P16_SRR089662
 M16_SRR090210
 M16_SRR090158
 M16_SRR090112
 M16_SRR090144
 M16_SRR090156

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NHeshEpiCXxE7hqmLx2c>

Analysis Parameters

Genome: hg19
 Binning: variable bins of 500kb size (see below for details)
 Segmentation: using normalized read counts
 Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts (rawick | sm | pdf | png)
 Copy number (rawick | sm | pdf | png)
 Correlations (rawick | sm | pdf | png)

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (282 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.46 MB)

Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
<input type="checkbox"/> M16_SRR089711		15,311,446	2,855.02	1,023.36	0.36
<input type="checkbox"/> M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
<input type="checkbox"/> M16_SRR090144		11,757,180	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,582,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	656.59	0.55
<input type="checkbox"/> M16_SRR090210		18,171,821	3,768.73	1,850.61	0.55

Analysis completed

View results

Tree

View analysis later

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=www.1x4/0000001C00T70g0000>

Analysis Parameters

Genome: hg19
 Binning: variable bins of 500kb size
 Segmentation: using normalized read counts
 Clustering: ward linkage, euclidean distance

Summary

Cell	Reads	Bin Count	Copy Number	Correlation	Score
<input type="checkbox"/> M16_SRR089733	17,395,808	3,243.67	1,147.06	0.35	
<input checked="" type="checkbox"/> M16_SRR090144	11,757,193	2,192.26	1,252.64	0.57	
<input checked="" type="checkbox"/> M16_SRR090156	7,592,532	1,415.63	766.63	0.54	
<input type="checkbox"/> M16_SRR090158	6,414,000	1,195.97	656.59	0.55	
<input type="checkbox"/> M16_SRR090210	18,171,621	3,388.33	1,866.61	0.55	
<input type="checkbox"/> M16_SRR090212	17,658,321	3,292.81	1,741.57	0.53	
<input type="checkbox"/> P16_SRR089604	15,674,790	3,810.66	610.36	0.32	

With selected cells, plot:

Tree display

Normalized read counts [treeview | smi | pdf | gviz]
 Copy-number [treeview | smi | pdf | gviz]
 Correlations [treeview | smi | pdf | gviz]

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (262 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (263 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

Normalized and Segmented Counts: Normalized and segmented bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

Analysis completed

View results

Tree

- P16_SRR089604
- P16_SRR089593
- M16_SRR089733
- M16_SRR089711
- P16_SRR089586
- P16_SRR089646
- P16_SRR089664
- P16_SRR089663
- P16_SRR089659
- P16_SRR089662
- M16_SRR090210
- M16_SRR090158
- M16_SRR090212
- M16_SRR090144
- M16_SRR090156

View analysis later

Access your results later at the following address:

<http://q5.cshl.edu/ginkgo/1q=results/NH5eHlEpiCXxE7hqmLx2c>

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Summary

Cell	Reads	1000x	100x	10x	1x
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> M16_SRR090210		16,171,621	3,388.33	1,866.61	0.55
<input type="checkbox"/> M16_SRR090212		17,659,321	3,292.81	1,741.57	0.53
<input type="checkbox"/> P16_SRR089666		15,074,760	2,810.68	910.36	0.32
<input type="checkbox"/> P16_SRR089693		16,176,073	3,016.24	1,035.63	0.34
<input type="checkbox"/> P16_SRR089604		5,760,274	1,074.06	357.84	0.33

Tree display

Normalized read counts [treeview | smi | pdf | gseq]

Copy-number [treeview | smi | pdf | gseq]

Correlations [treeview | smi | pdf | gseq]

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (262 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (263 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

Normalized and Segmented Counts: Normalized and segmented bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

With selected cells, plot:

Analysis completed

View results

Tree

View analysis later

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=www1/cv/000001/CA011746046>

Analysis Parameters

Genome: hg19
 Binning: variable bins of 500kb size
 Segmentation: using normalized read counts
 Clustering: ward linkage, euclidean distance

Summary

Cell	Reads	Bin Count	Copy Number	MAD	GC Bias
M16_SRR090210		18,171,621	3,355.33	1,866.61	0.55
M16_SRR090212		17,859,321	3,292.61	1,741.57	0.53
P16_SRR089586		15,074,760	2,810.68	910.36	0.32
P16_SRR089660		16,176,073	3,016.24	1,035.63	0.34
P16_SRR089604		5,760,274	1,074.06	357.84	0.33
P16_SRR089646		8,276,396	1,543.61	834.04	0.54
P16_SRR089658		14,513,017	2,706.25	1,459.16	0.54

Tree display

Normalized read counts [treeview | smi | pdf | gviz]

Copy-number [treeview | smi | pdf | gviz]

Correlations [treeview | smi | pdf | gviz]

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (262 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (263 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

Normalized and Segmented Counts: Normalized and segmented bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

With selected cells, plot:

Analysis completed

View results

Tree

- P16_SRR089604
- P16_SRR089593
- M16_SRR089733
- M16_SRR089711
- P16_SRR089586
- P16_SRR089646
- P16_SRR089664
- P16_SRR089663
- P16_SRR089659
- P16_SRR089662
- M16_SRR090210
- M16_SRR090158
- M16_SRR090212
- M16_SRR090144
- M16_SRR090156

View analysis later

Access your results later at the following address:

<http://q5.cshl.edu/ginkgo/?q=results/NHhoshEpiCXcE7hqmLx2c>

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Summary

Cell	Reads	Bin Count	GC Bias	MAD	Correlation
M16_SRR090210		18,171,621	3,355.33	1,866.61	0.55
M16_SRR090212		17,659,321	3,292.61	1,741.57	0.53
P16_SRR089586		15,074,760	2,810.68	910.36	0.32
P16_SRR089660		16,176,073	3,016.24	1,035.63	0.34
P16_SRR089604		5,760,274	1,074.06	357.84	0.33
P16_SRR089646		8,276,396	1,543.61	834.04	0.54
P16_SRR089658		14,513,017	2,706.25	1,459.16	0.54

Tree display

- Normalized read counts [treeview | smi | pdf | gseq]
- Copy-number [treeview | smi | pdf | gseq]
- Correlations [treeview | smi | pdf | gseq]

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

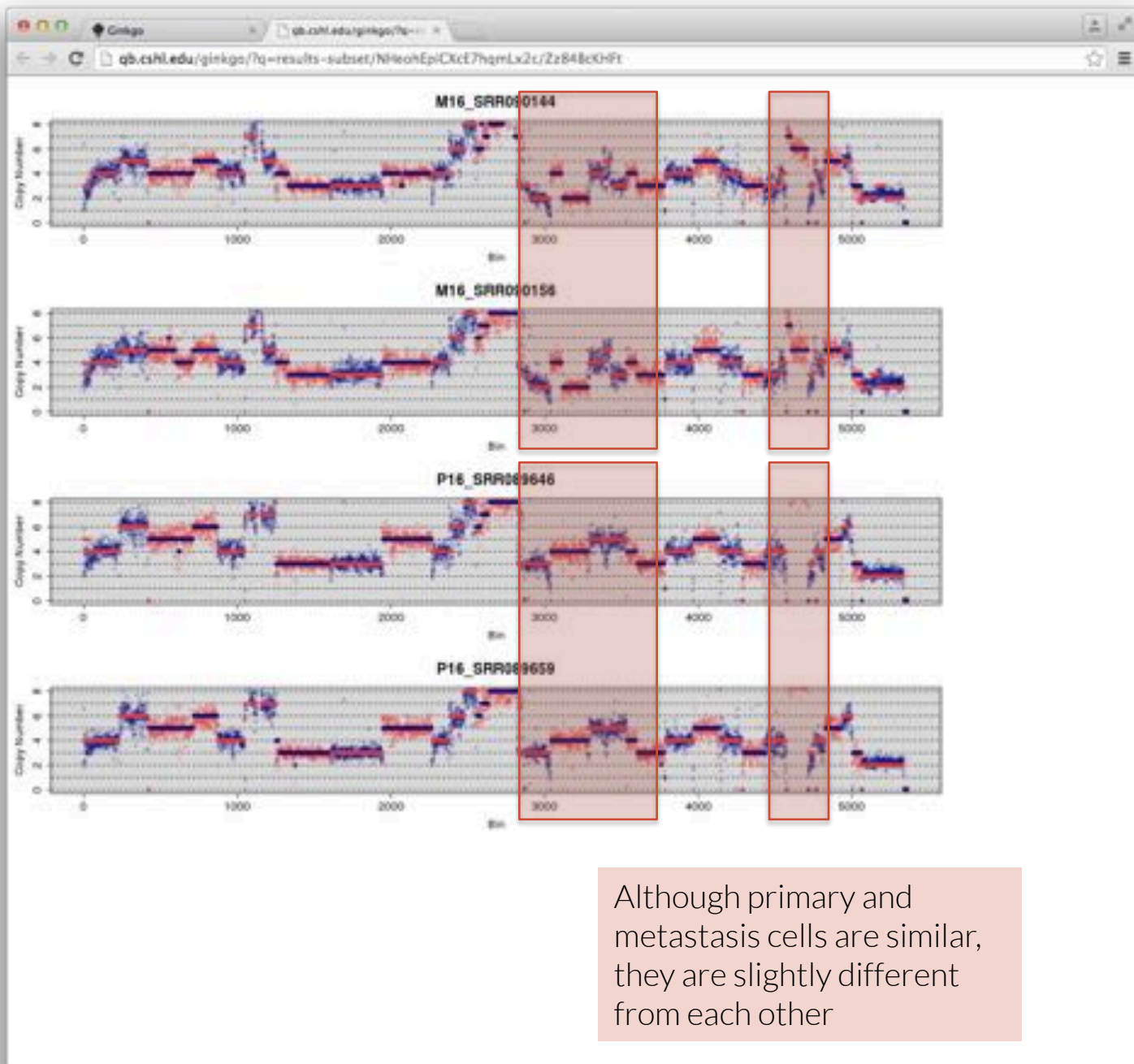
Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (262 KB)

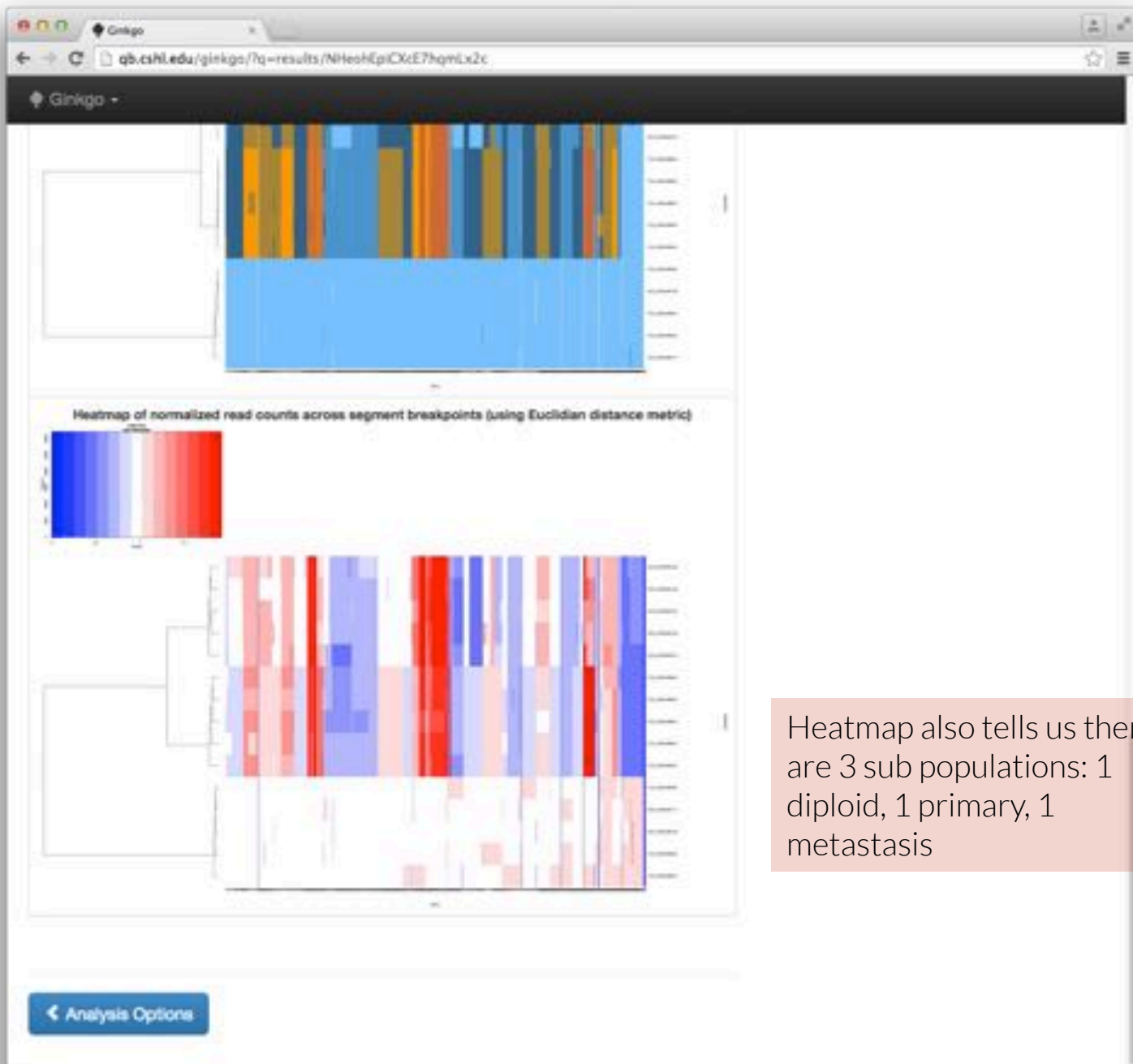
Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (263 KB)

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

Normalized and Segmented Counts: Normalized and segmented bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 MB)

With selected cells, plot: IV plots Lorenz curve GC bias MAD





Heatmap also tells us there are 3 sub populations: 1 diploid, 1 primary, 1 metastasis

Ginkgo -

q5.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c

Ginkgo

Single-cell analysis

Analysis complete!

View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
 P16_SRR089585
 P16_SRR089646
 P16_SRR089664
 P16_SRR089663
 P16_SRR089659
 P16_SRR089682
 M16_SRR090210
 M16_SRR090210
 M16_SRR090210
 M16_SRR090210
 M16_SRR090210

View analysis later

Access your results later at the following address:

<https://q5.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c>

Analysis Parameters

Binomial hold

Window of 500kb size (with 100bp resolution)

Using normalized read counts

Linkage, euclidean distance

Tree display

Normalized read counts (newick | vml | pdf | png)
 Copy number (newick | vml | pdf | png)
 Correlations (newick | vml | pdf | png)

Download processed data

Statistics: Bin count statistics for each cell (1 KB)

Breakpoints: Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (283 KB)

Copy Number: Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (283 KB)

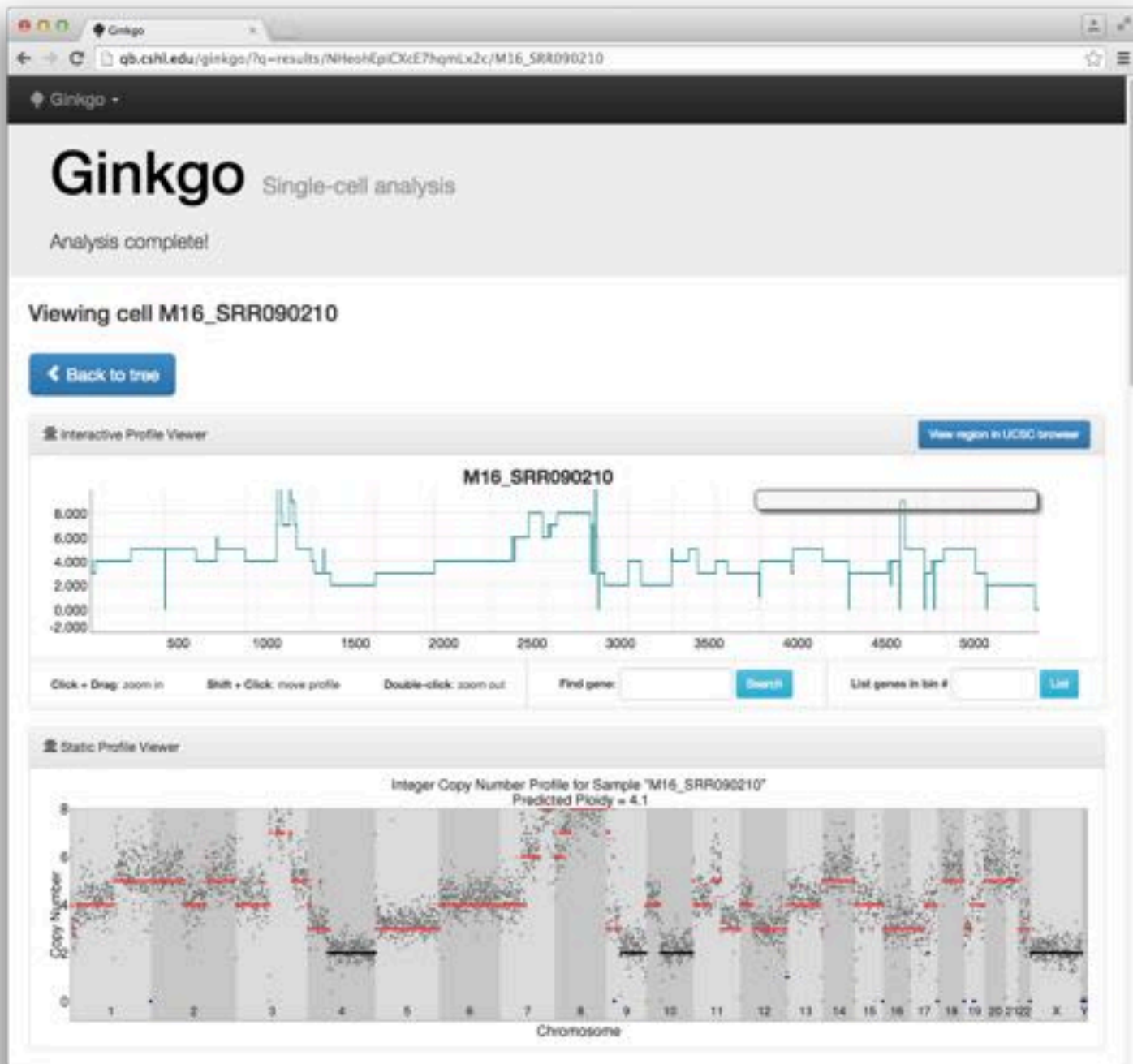
Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (11.48 MB)

Summary

Cell ID	Genomic Plot	Bin Count	Copy Number	Correlation	Other Metric
M16_SRR090210		16,171,621	3,388.33	1,806.61	0.55
M16_SRR090212		17,858,321	3,292.61	1,741.57	0.53
P16_SRR089646		16,674,760	2,810.66	910.36	0.32
P16_SRR089653		16,176,073	3,016.24	1,036.63	0.34
P16_SRR089604		5,760,274	1,074.06	357.84	0.33
P16_SRR089646		8,276,396	1,543.61	834.04	0.54
P16_SRR089659		14,513,617	2,706.25	1,459.16	0.54

q5.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210

Can click on cells to get in-depth view



Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

Interactive Profile Viewer [View region in UCSC browser](#)

M16_SRR090210

Bin 4590
Position: chr16:5878154-7306581
Copy-Number: 0

Click + Drag: zoom in Shift + Click: move profile Double-click: zoom out Find gene: [Search](#) List genes in bin #: [List](#)

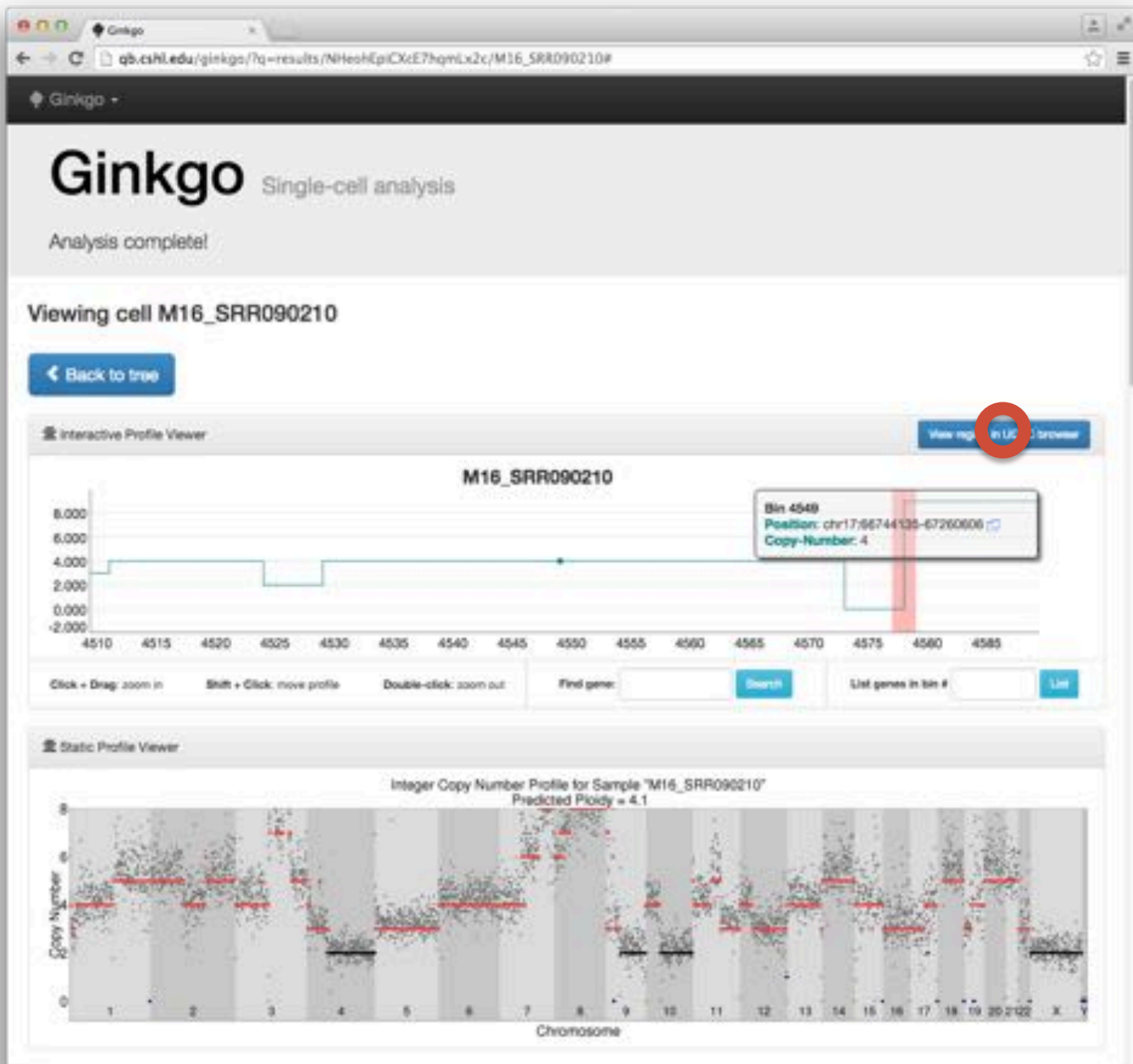
Static Profile Viewer

Integer Copy Number Profile for Sample "M16_SRR090210"
Predicted Ploidy = 4.1

Copy Number

Chromosome

Interactive profile viewer allows you to zoom into a region of interest



UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

chr17:46,497,635-81,195,210 34,697,576 bp. [hg19 replaces hg12 as default human assembly](#)

UCSC Genes RefSeq Genes AceView Genes Augustus CCDS Ensembl Genes

EvoFold Exoniphy GENCODE... Genelid Genes Genscan Genes H-inv 7.0

KMC Genes Mapped lncRNAs... LRG Transcripts MGC Genes N-SCAN Old UCSC Genes

ORFome Clones Other RefSeq Pfam in UCSC Gene Retroposed Genes SGP Genes SIB Genes

We automatically upload amplification and deletion tracks to the UCSC browser

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

← Back to tree

Interactive Profile Viewer [View region in UCSC browser](#)

M16_SRR090210

Bin 4590
Position: chr16:6876154-7306561
Copy-Number: 0

Click + Drag: zoom in Shift + Click: move profile Double-click: zoom out Find gene: List genes in bin #:

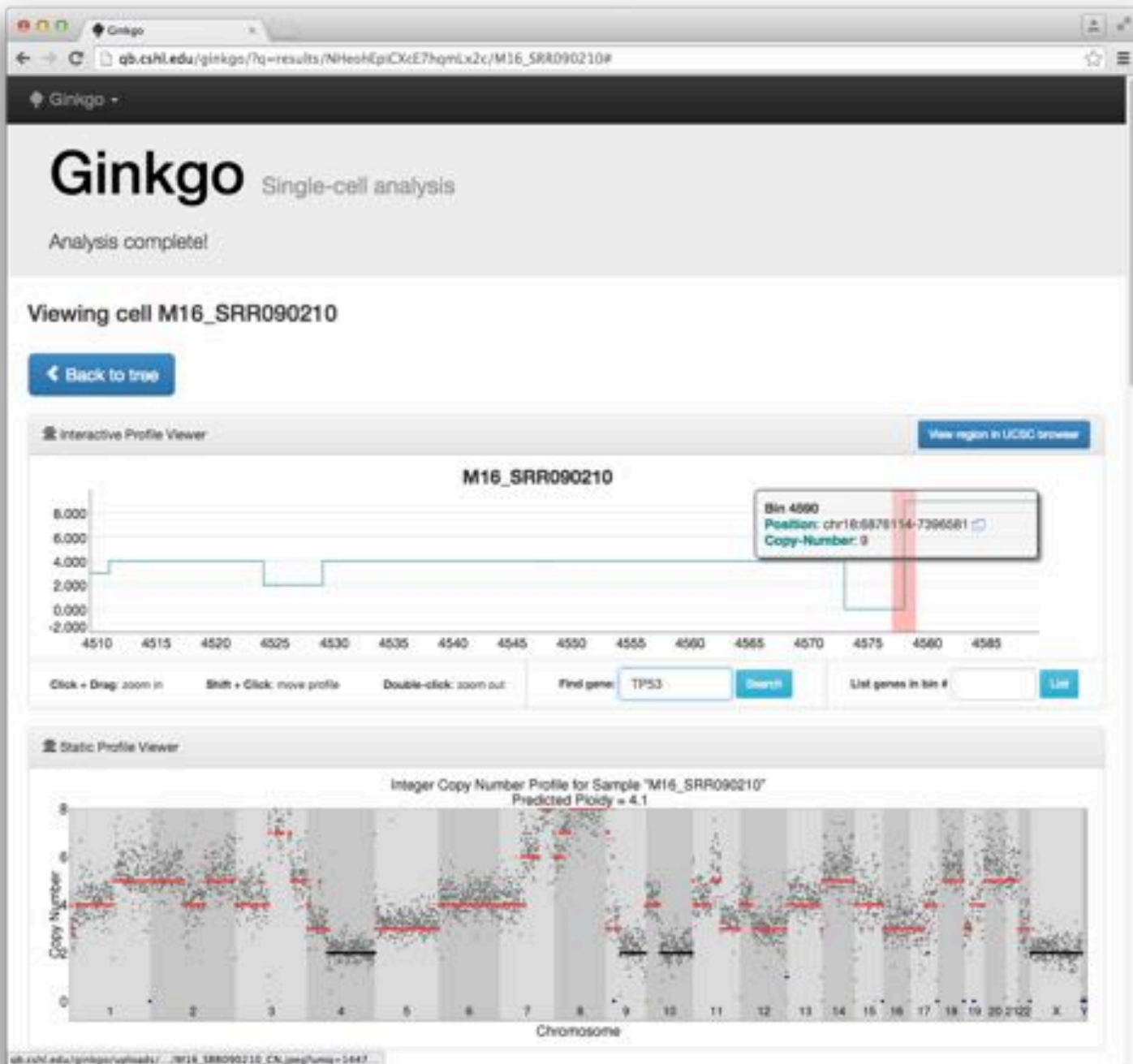
Static Profile Viewer

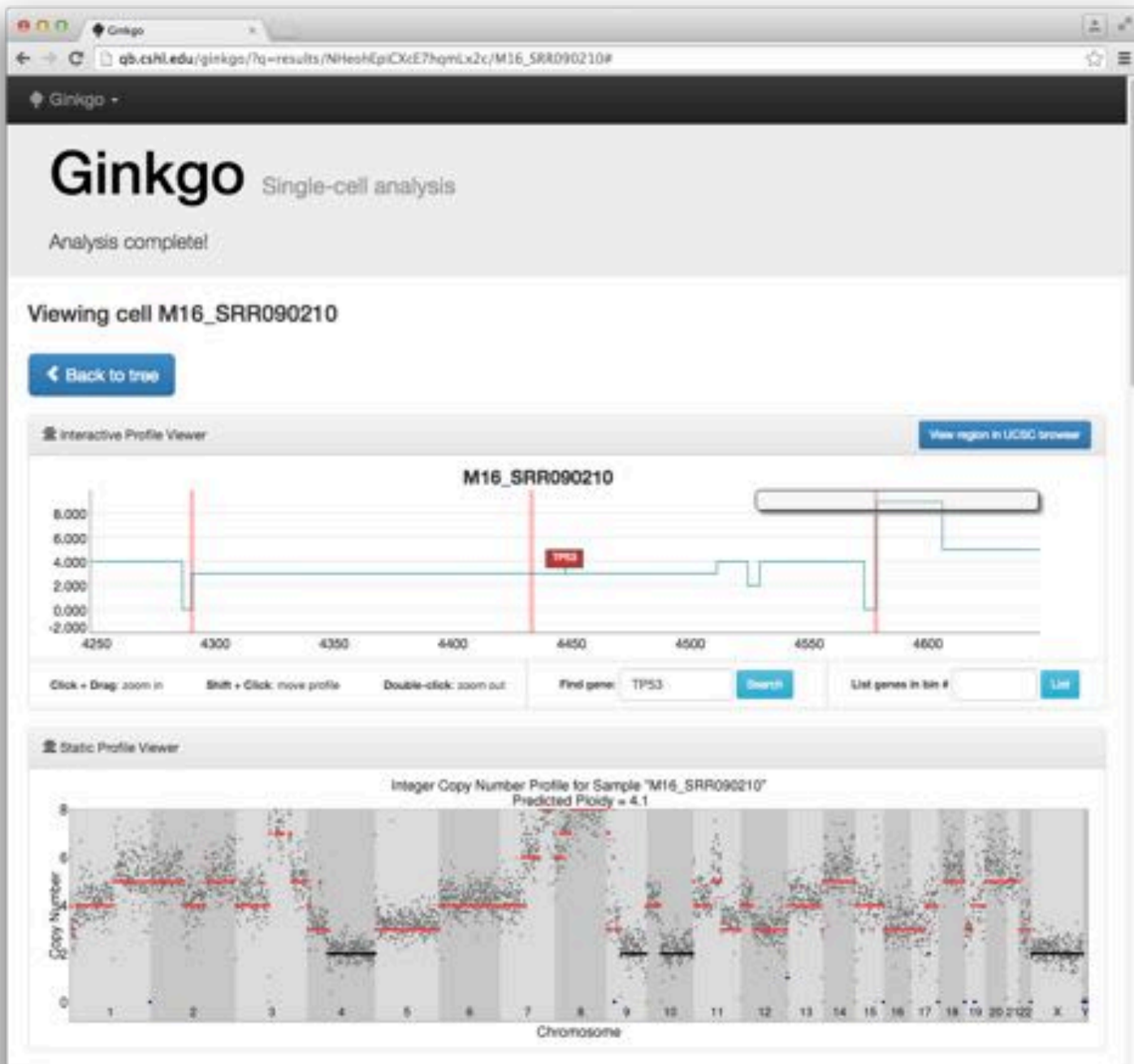
Integer Copy Number Profile for Sample "M16_SRR090210"
Predicted Ploidy = 4.1

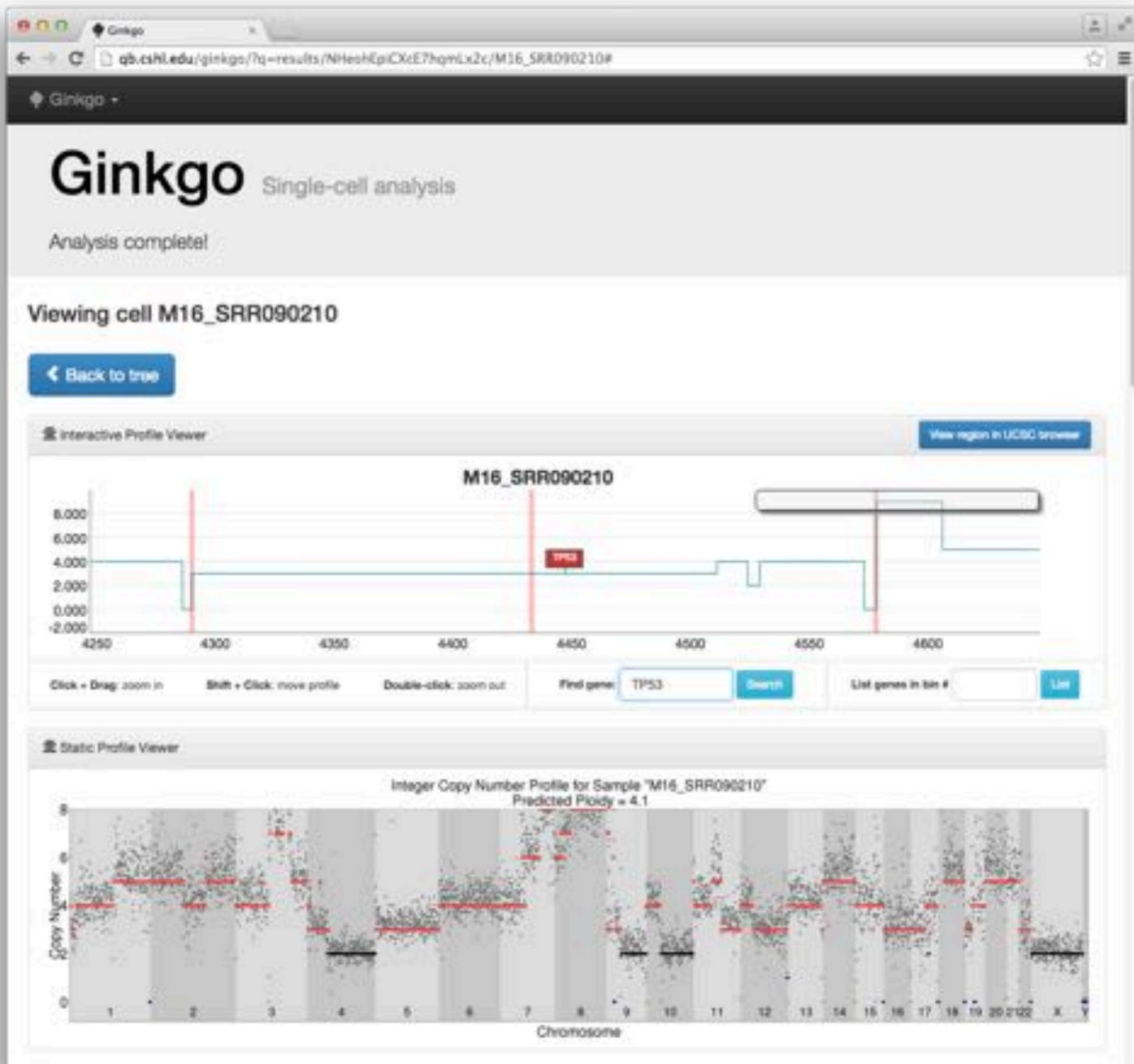
Copy Number

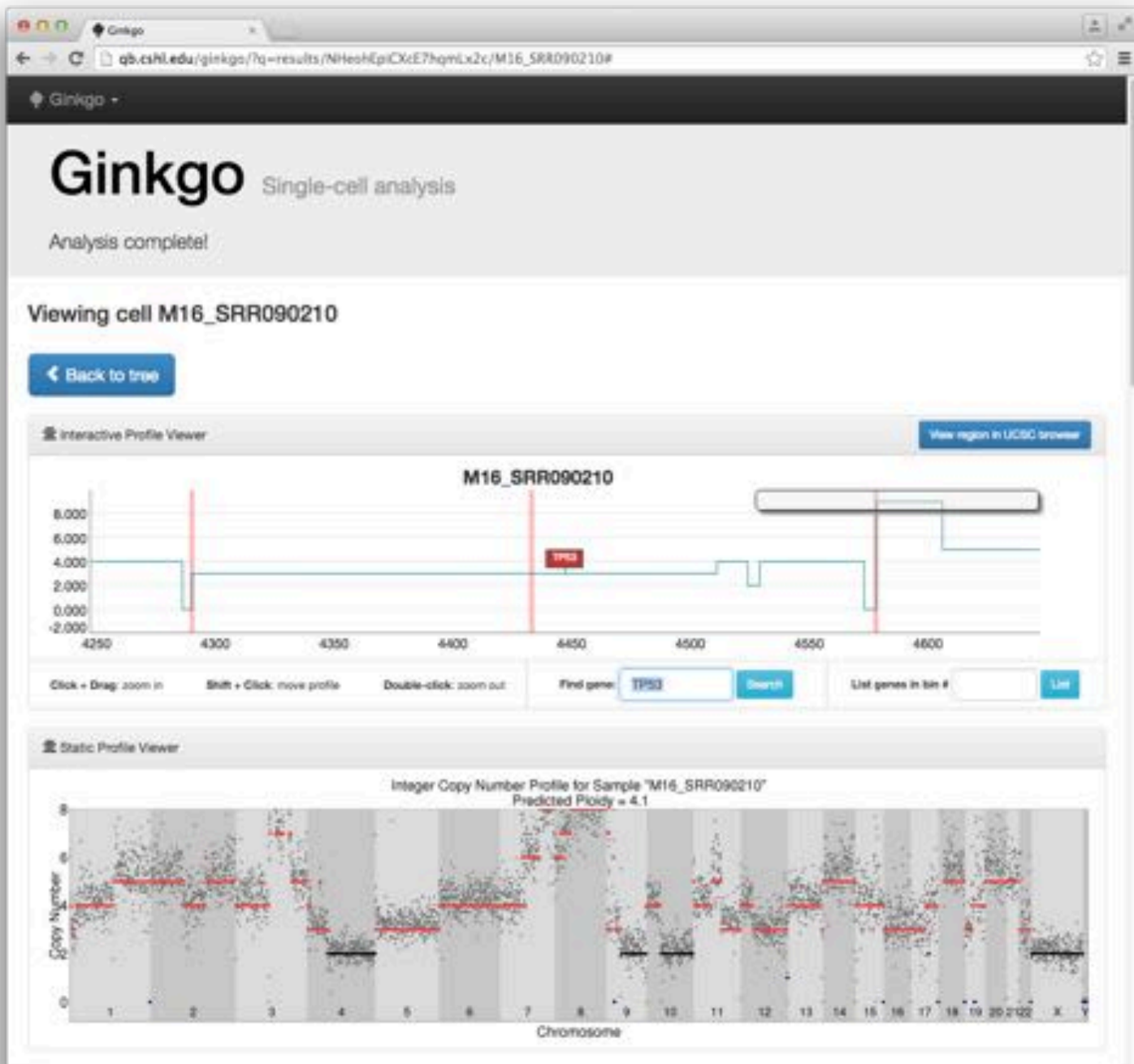
Chromosome

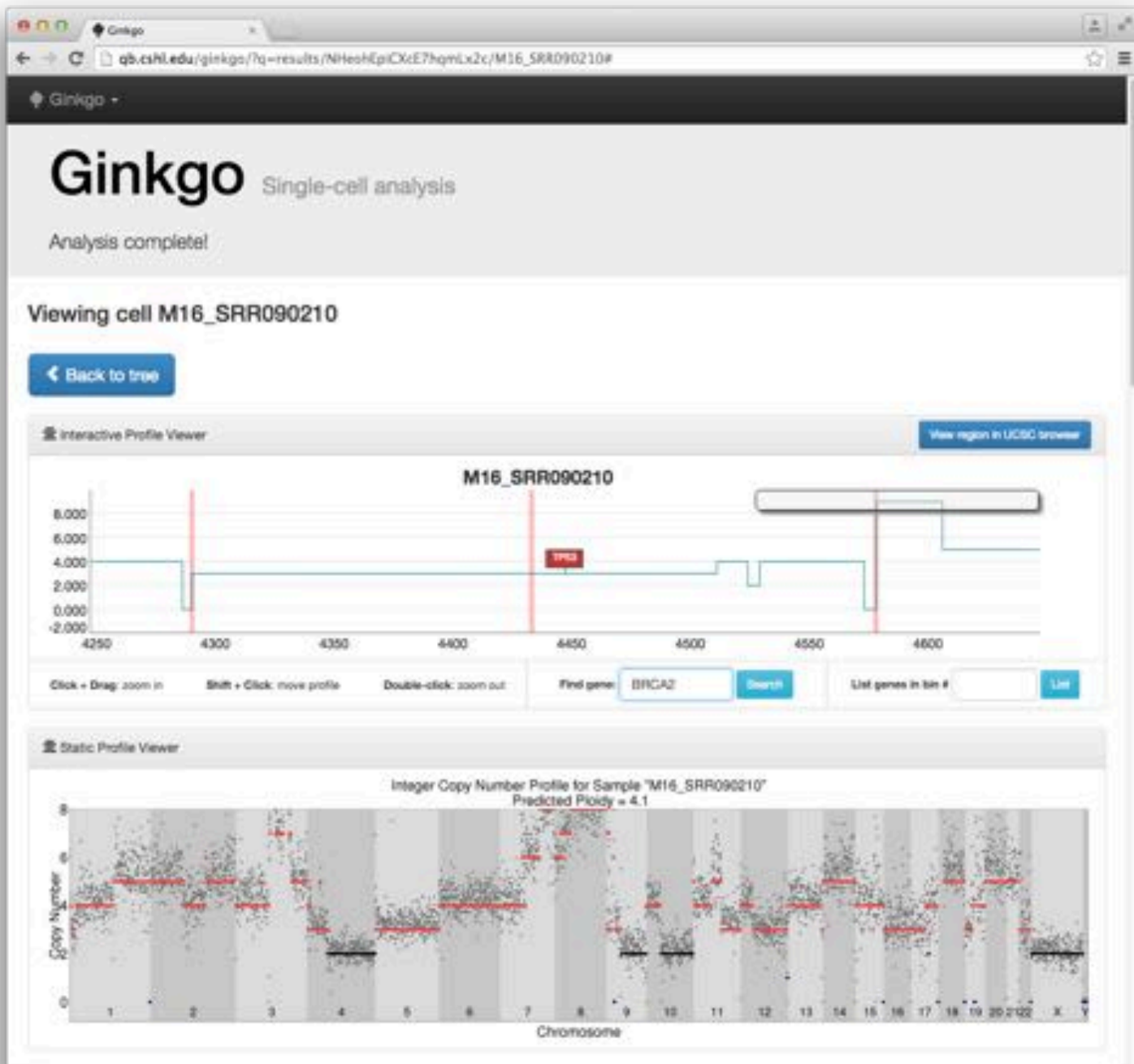
We can also annotate the CNV profile with genes of interest

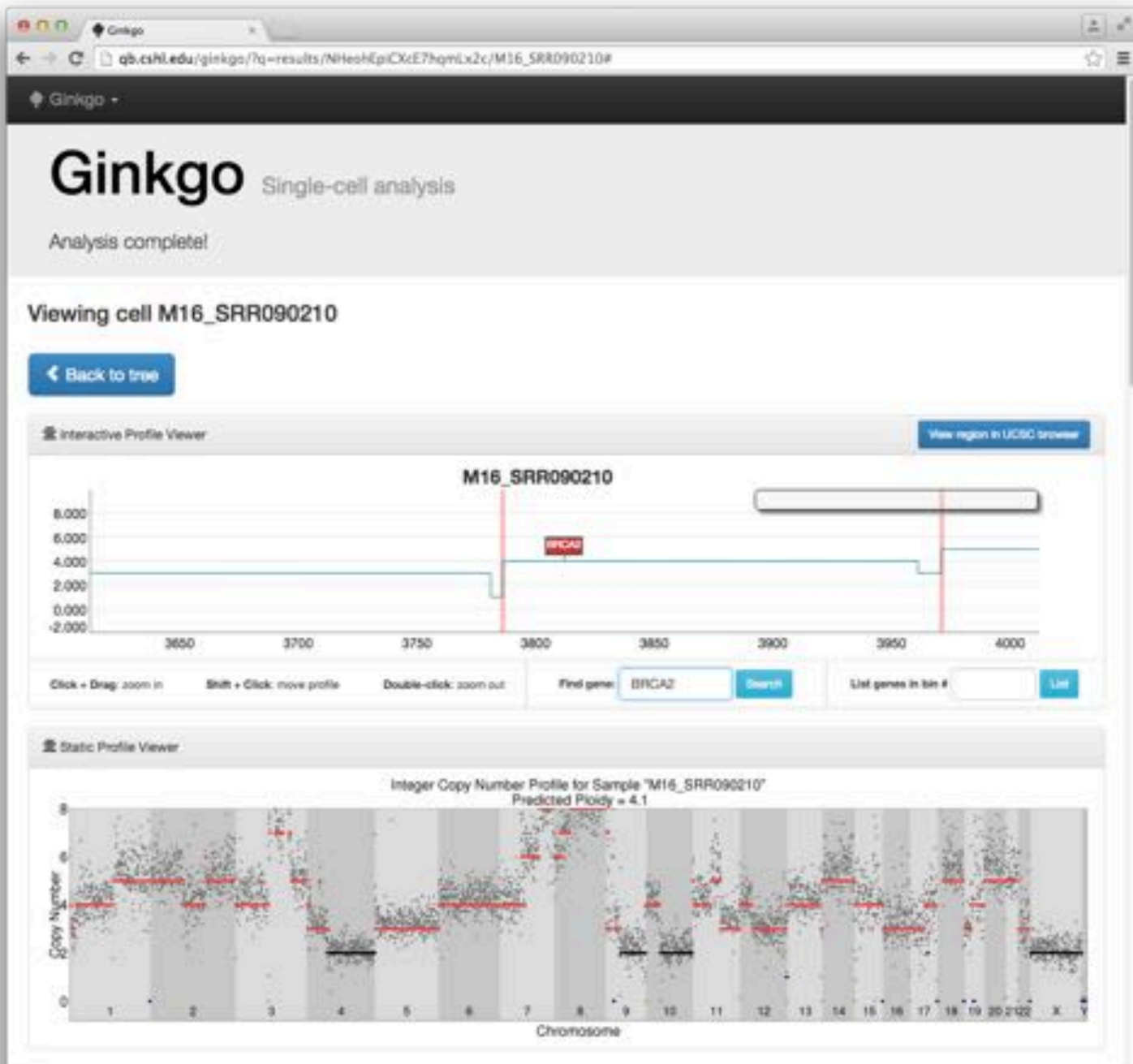


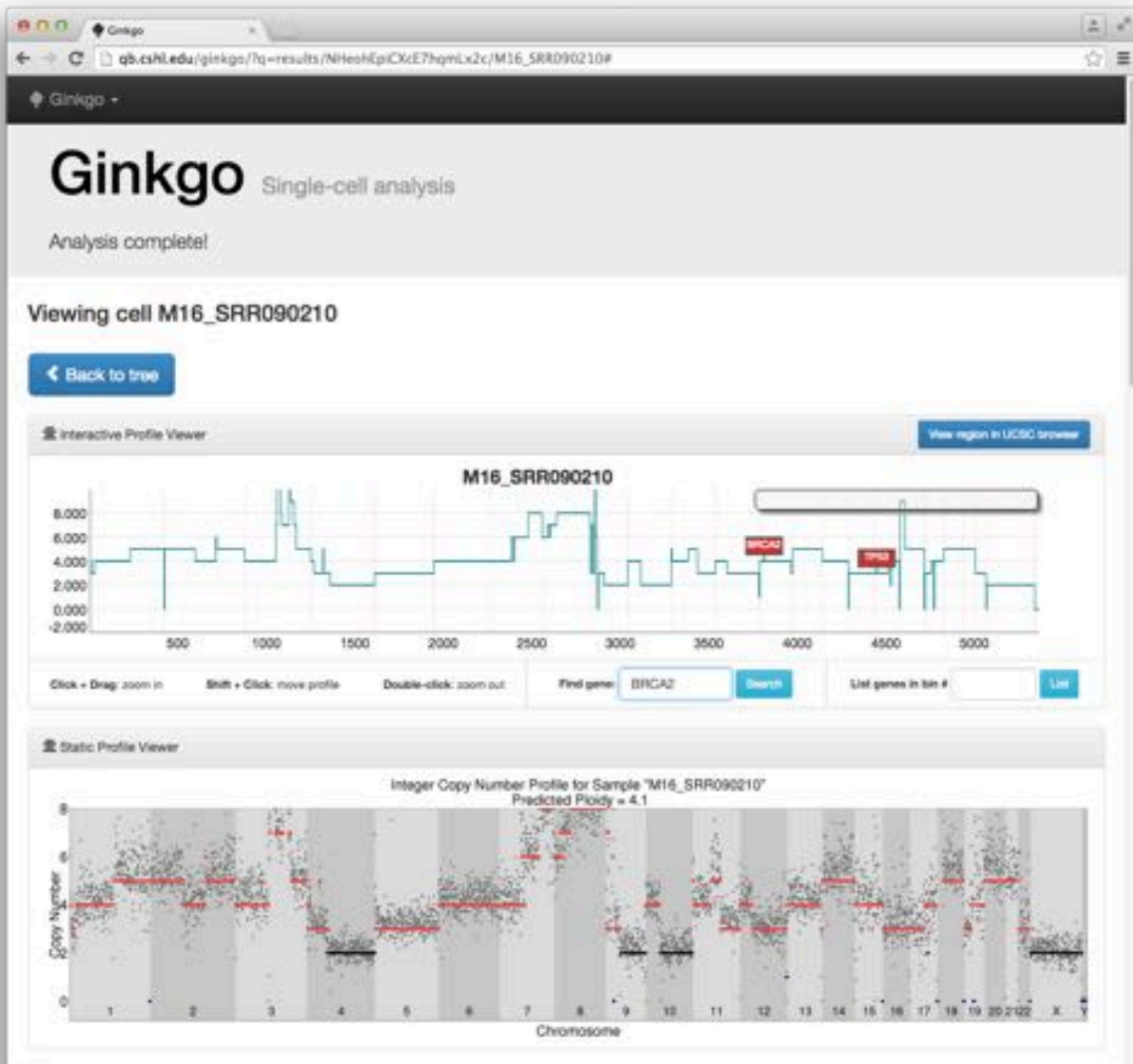


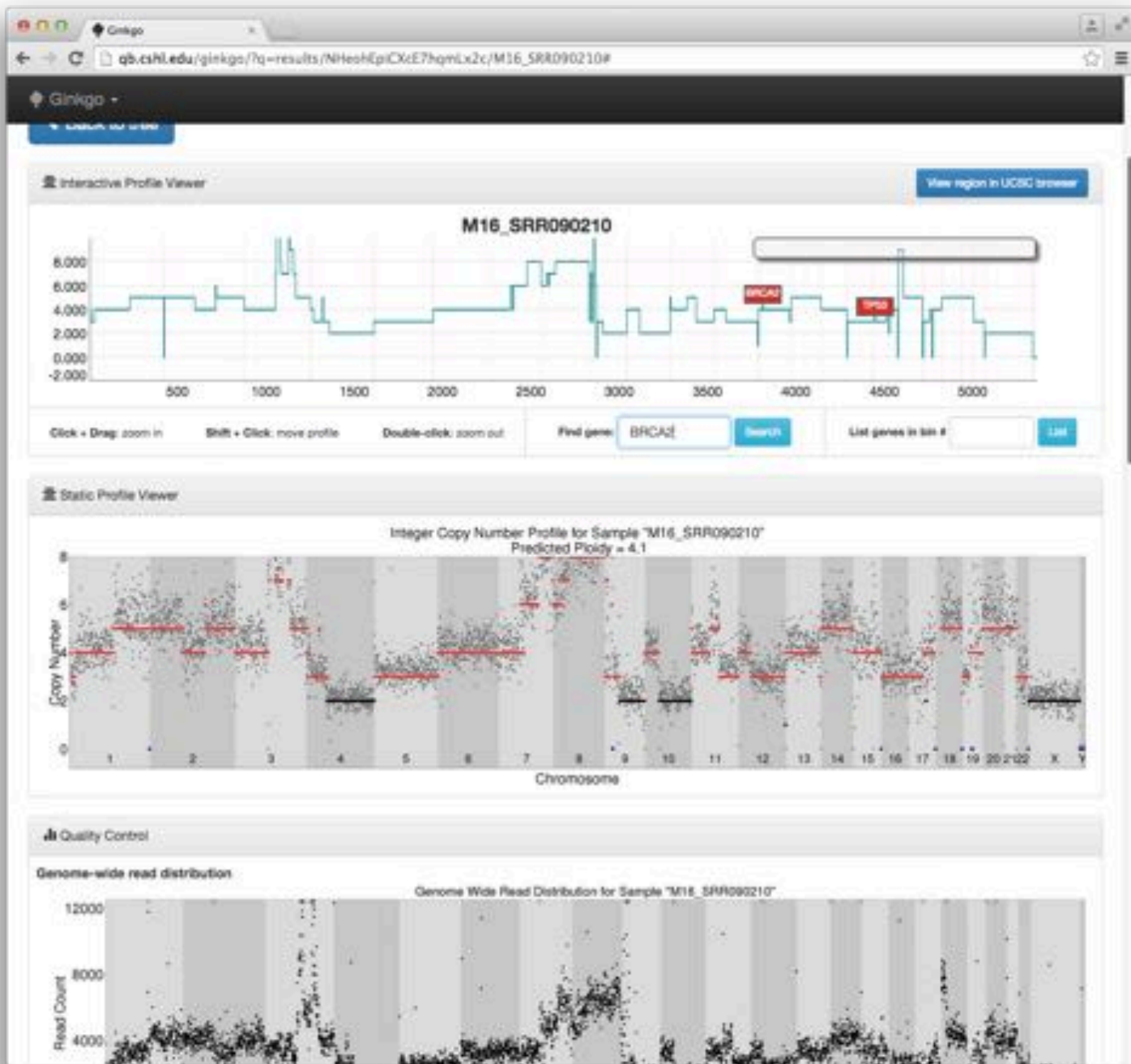




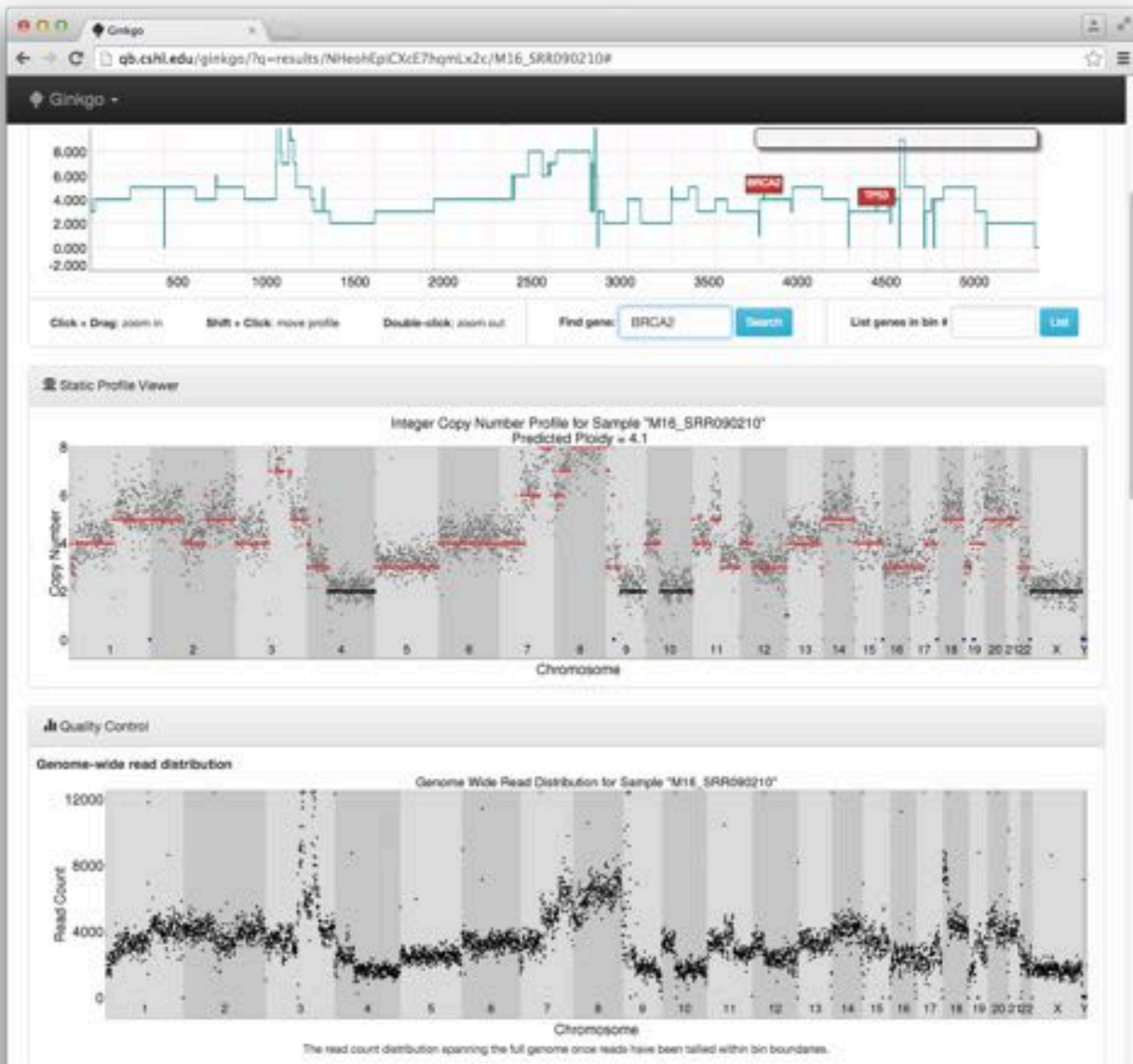


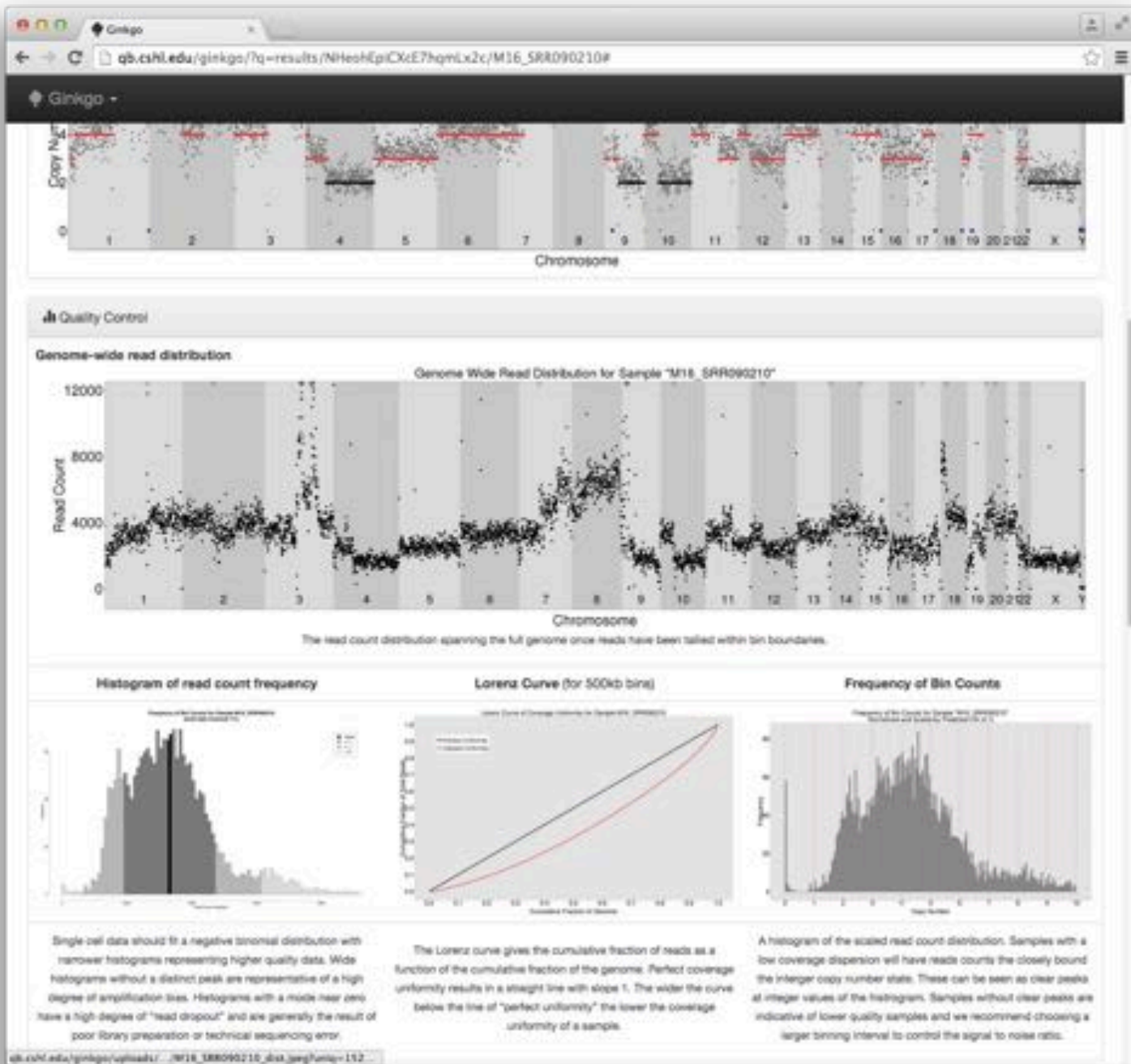






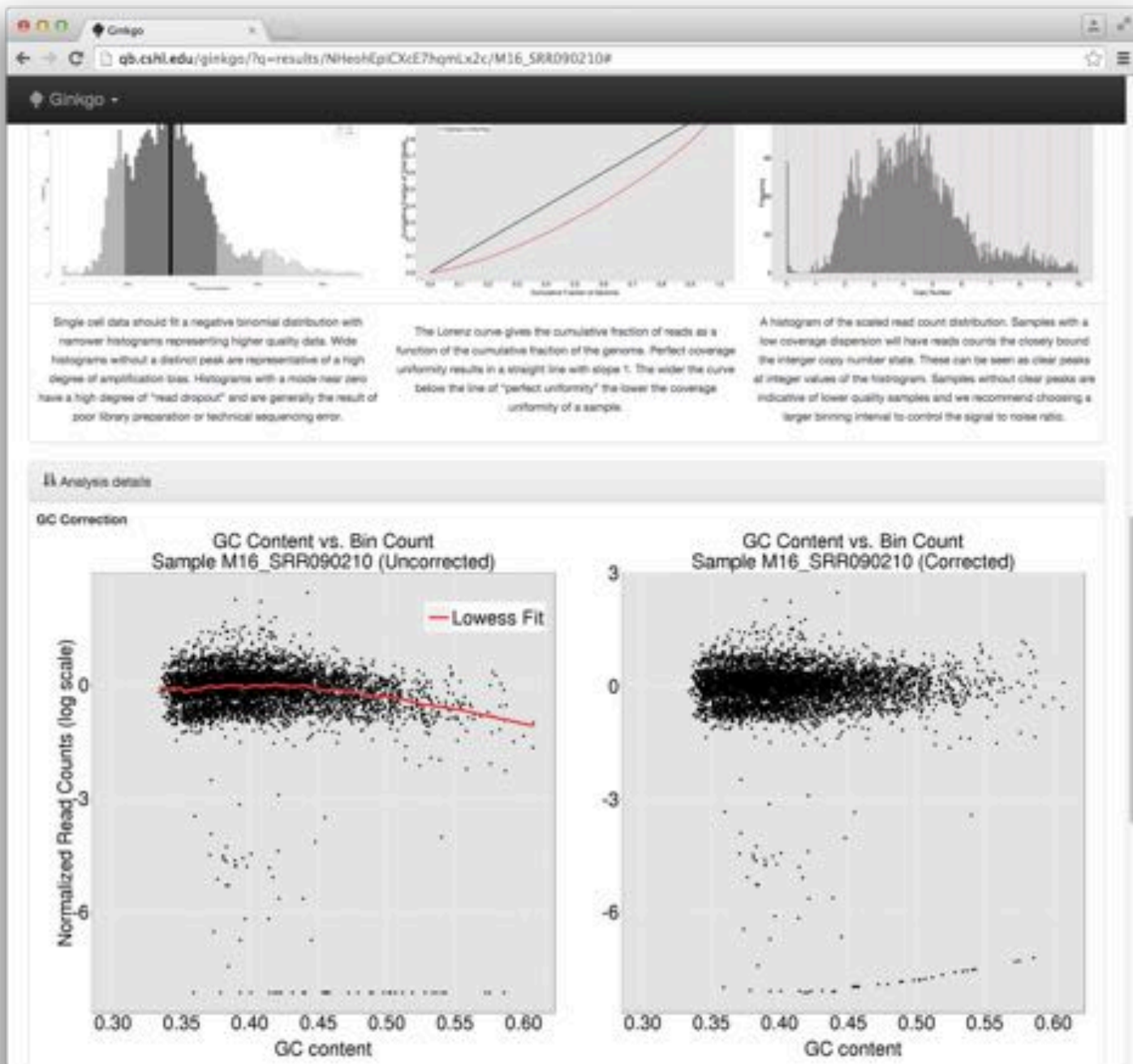
Static CNV profile



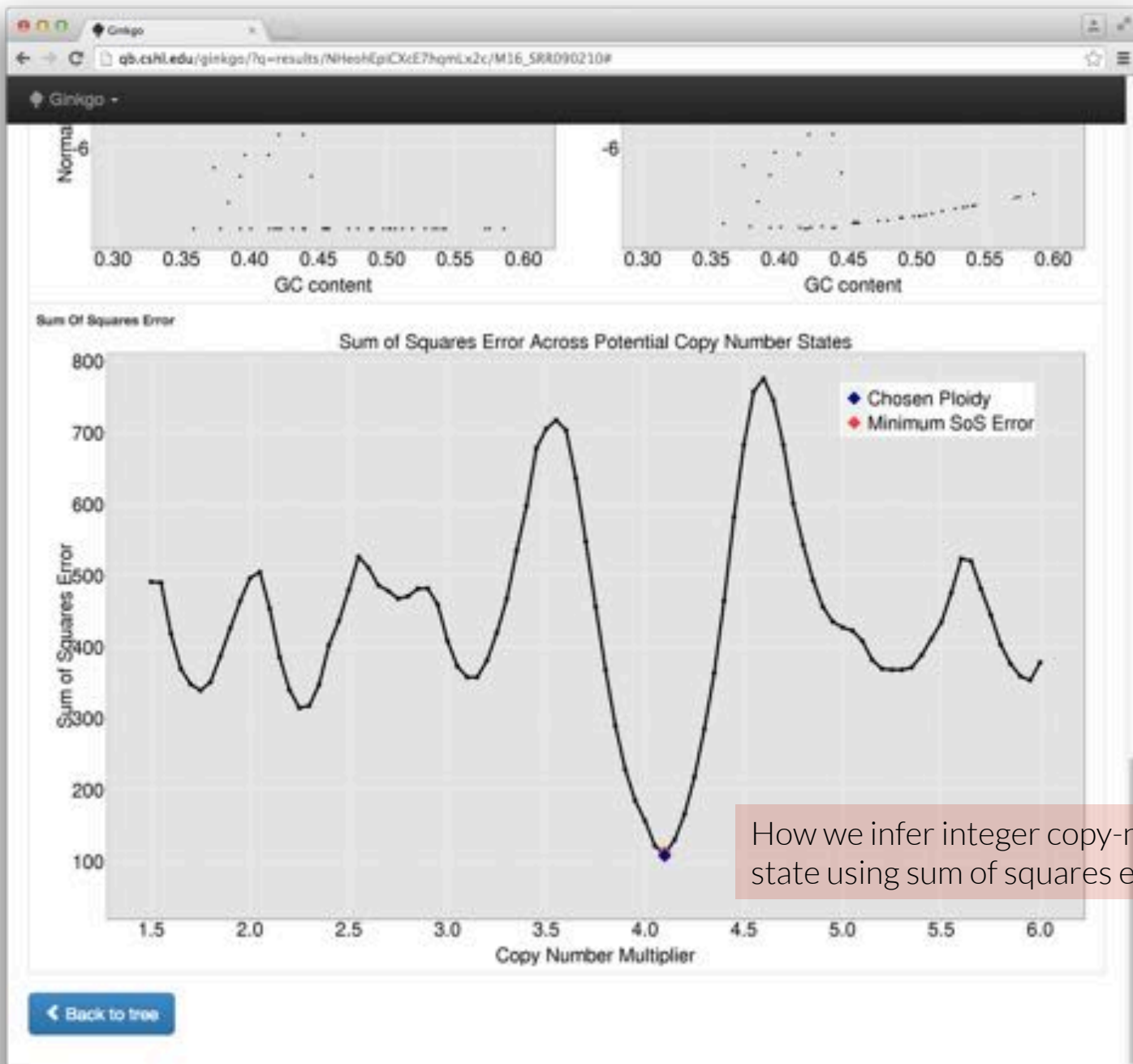


Original read count data before analysis

QC metrics



GC bias,
before and
after
correction



How we infer integer copy-number state using sum of squares error

Outline

Introduction



Ginkgo



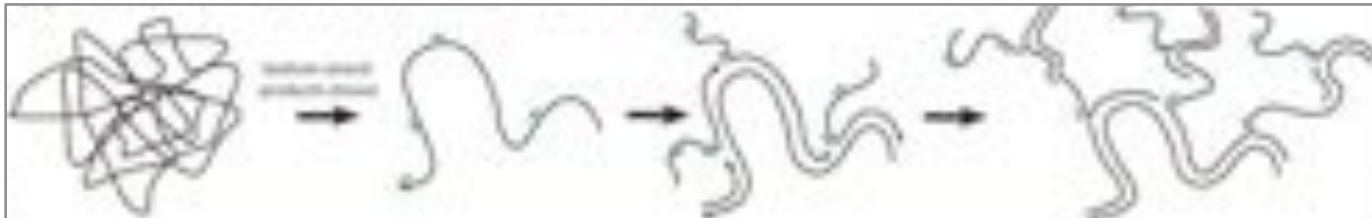
**Comparison of
WGA methods**



Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)



MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

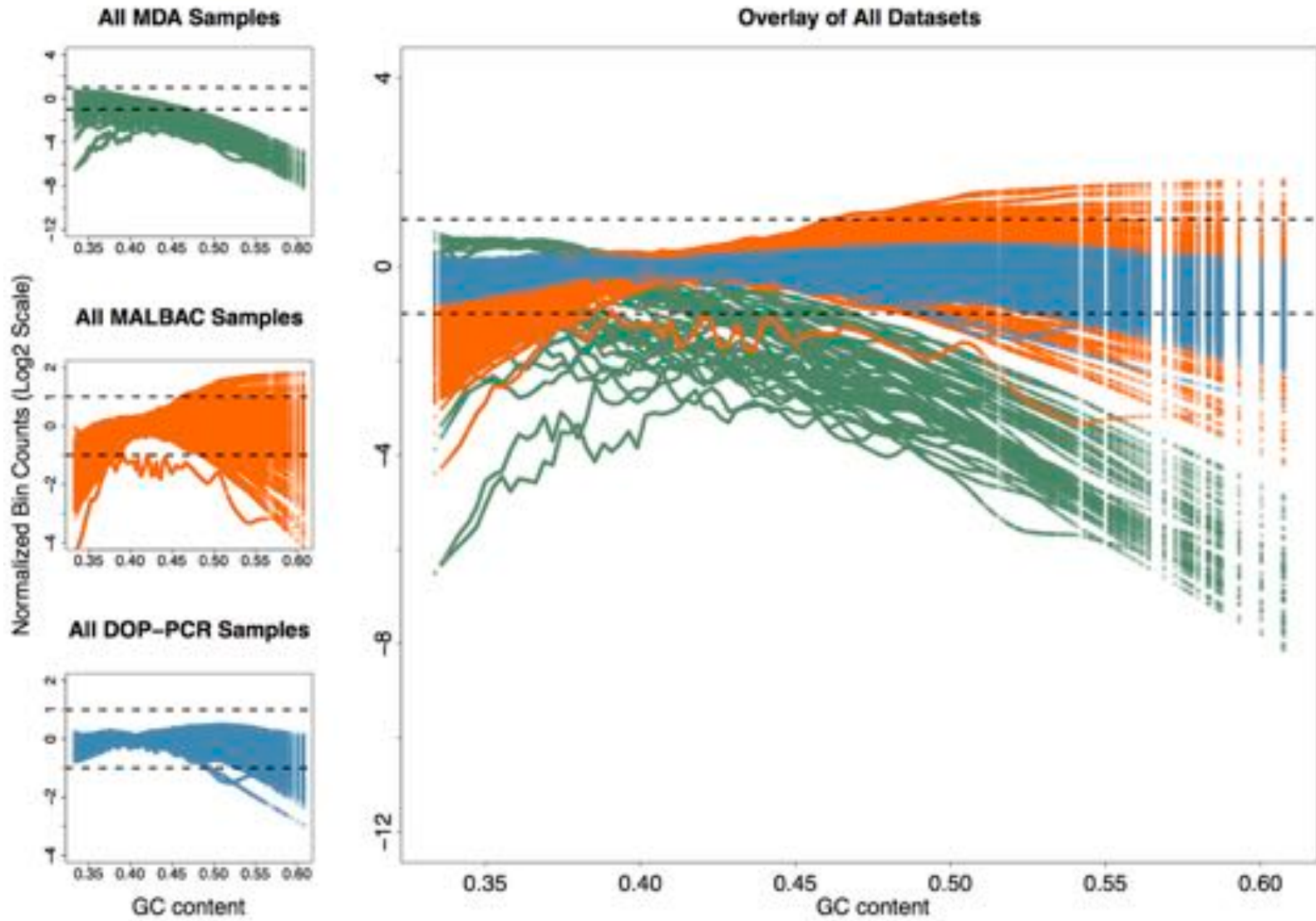
Comparison of WGA methods

Paper	WGA Method	Tissue
Navin et al., 2011	DOP-PCR	Breast (T10)
Navin et al., 2011	DOP-PCR	Breast (T16P/M)
McConnell et al., 2013	DOP-PCR	Neuron
Lu et al., 2012	MALBAC	Sperm
Ni et al., 2013	MALBAC	Lung
Hou et al., 2013	MALBAC	Oocyte
Kirkness et al., 2013	MDA	Sperm
Wang et al., 2012	MDA	Sperm
Evrony et al., 2012	MDA	Neuron

Explore the effects of WGA method on data quality:

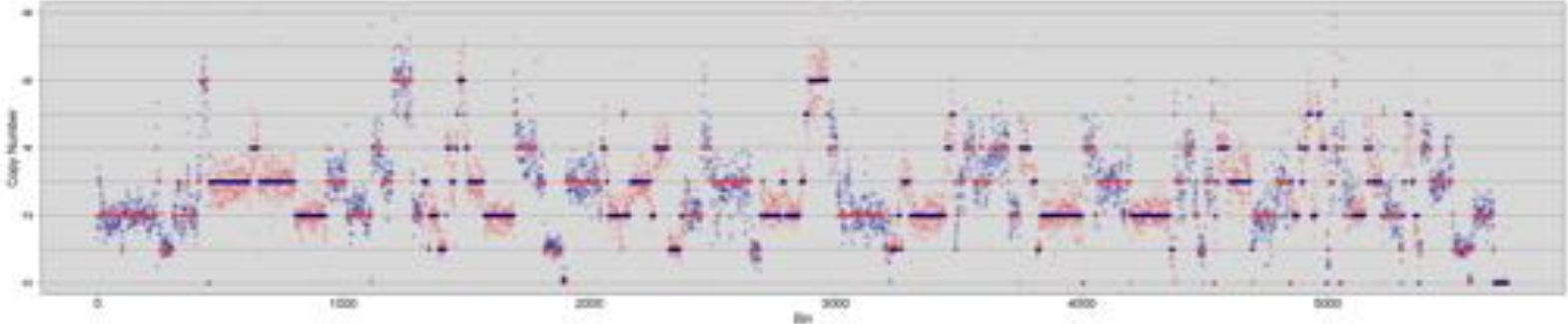
- 1) GC bias
- 2) Coverage dispersion

GC Bias

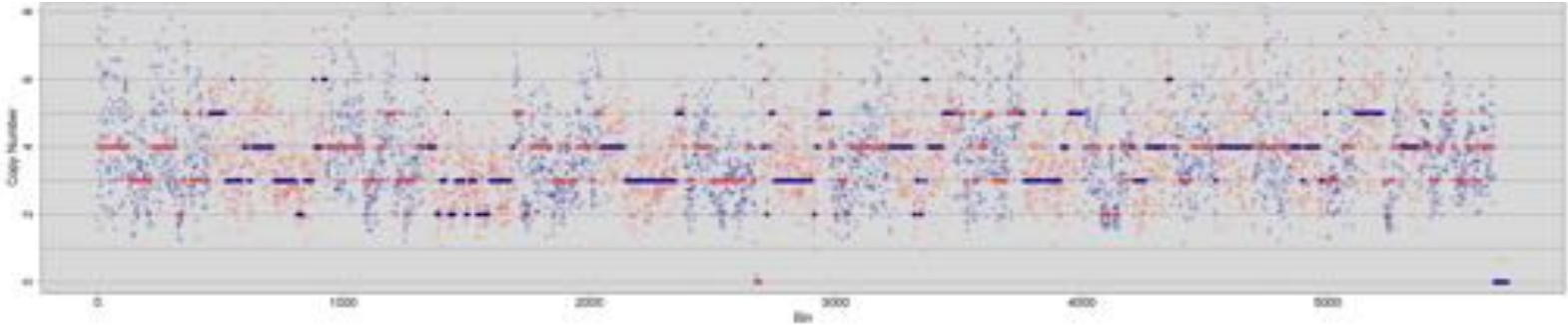


Coverage Dispersion

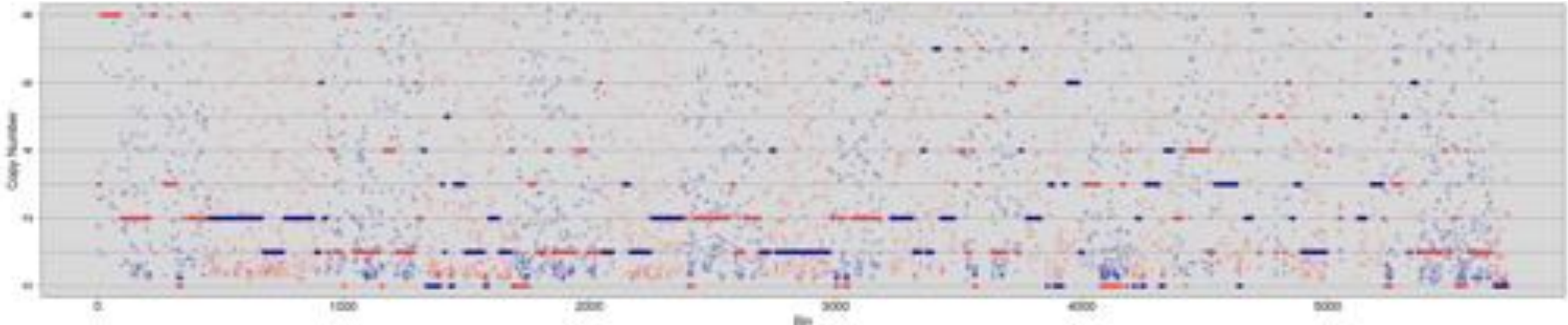
DOP-PCR



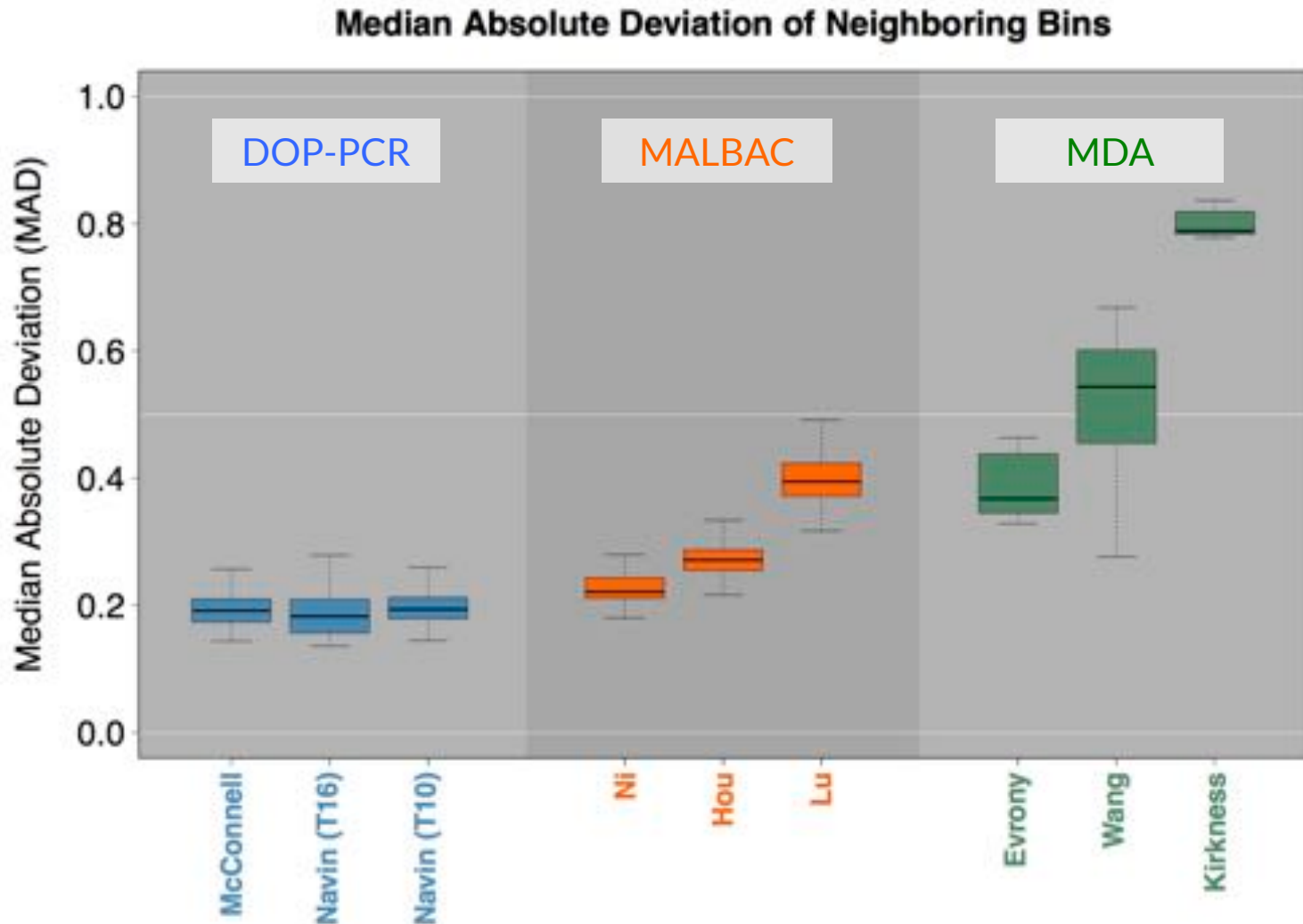
MALBAC



MDA



Coverage Dispersion



$$\text{MAD} = \text{median}_i (|X_i - \text{median}_j(X_j)|),$$

Summary

- Ginkgo is a platform for single-cell CNV analysis and visualization
- For copy-number analysis, we recommend DOP-PCR
- Check out Ginkgo and give us feedback
 - qb.cshl.edu/ginkgo
 - *Garvin and Aboukhalil et al., Nature Methods, 2015*

> 1K
USERS

21,500
PAGEVIEWS

10 MIN
TIME SPENT



Nov 3, San Francisco

Thanks

Ginkgo Team

Tyler Garvin

Jude Kendall

Timour Baslan

Jim Hicks

Gurinder S. Atwal

Michael Wigler

Michael C. Schatz

qb.cshl.edu/ginkgo



Cold
Spring
Harbor
Laboratory