

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



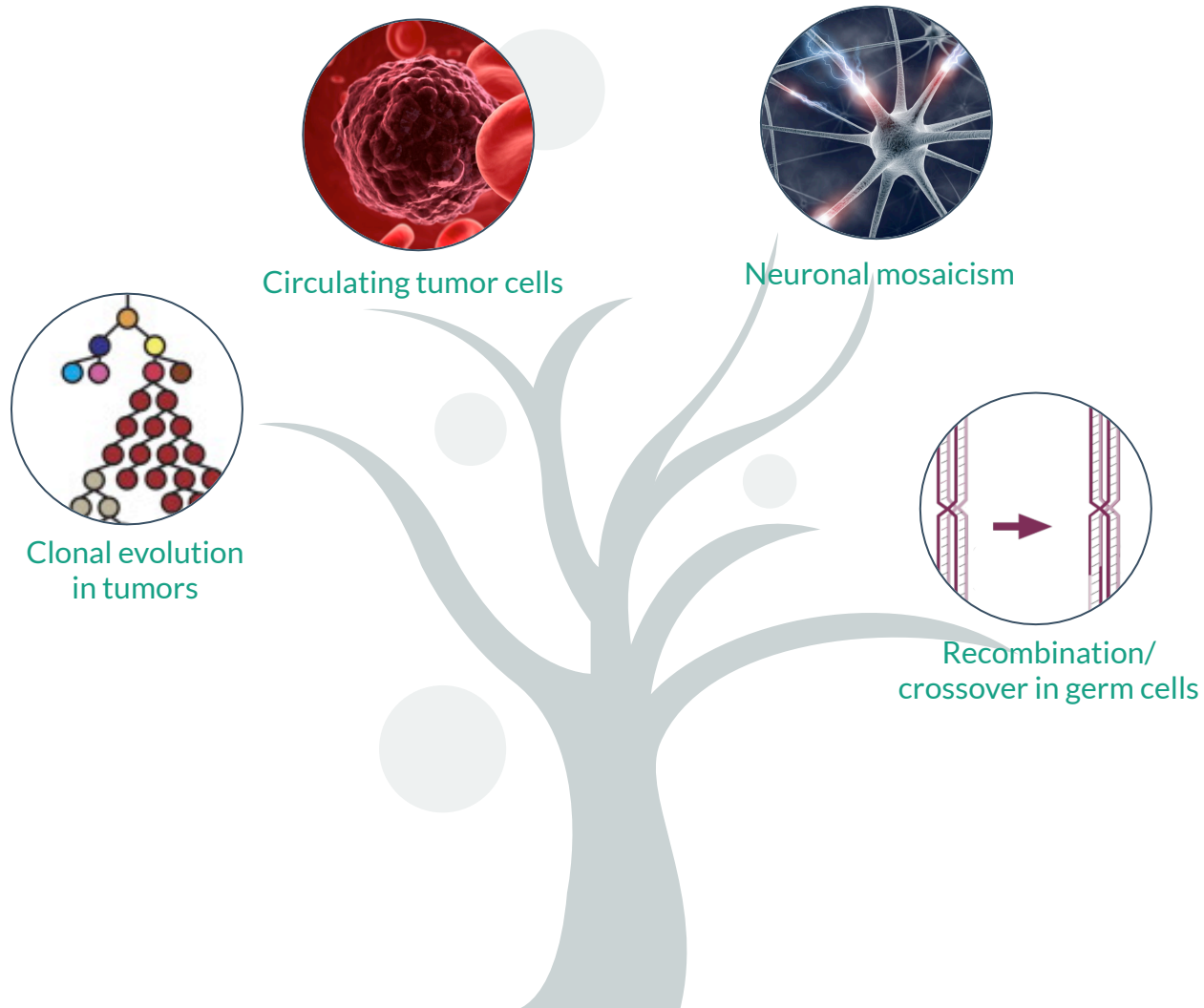
Ginkgo



Comparison of
WGA methods



Single-cell sequencing

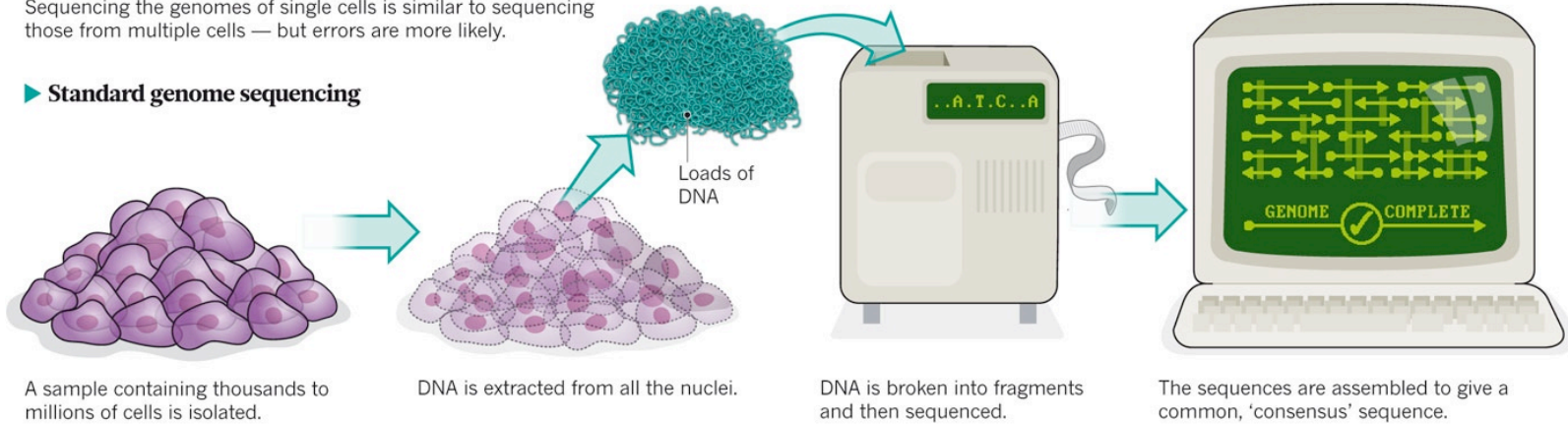


Single-cell vs. bulk sequencing

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



A sample containing thousands to millions of cells is isolated.

DNA is extracted from all the nuclei.

DNA is broken into fragments and then sequenced.

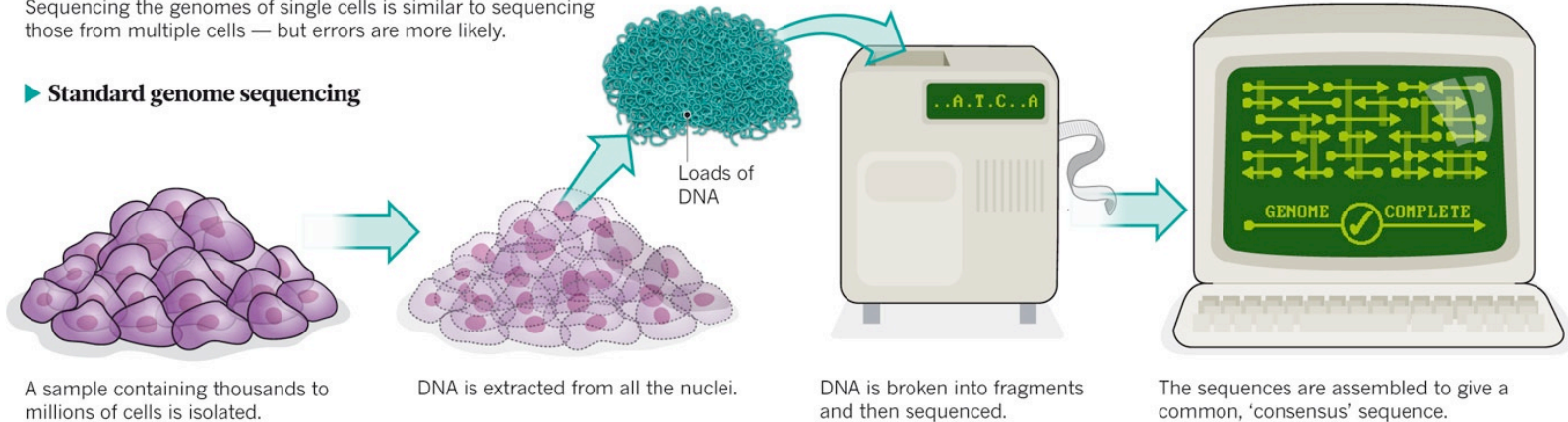
The sequences are assembled to give a common, 'consensus' sequence.

Single-cell vs. bulk sequencing

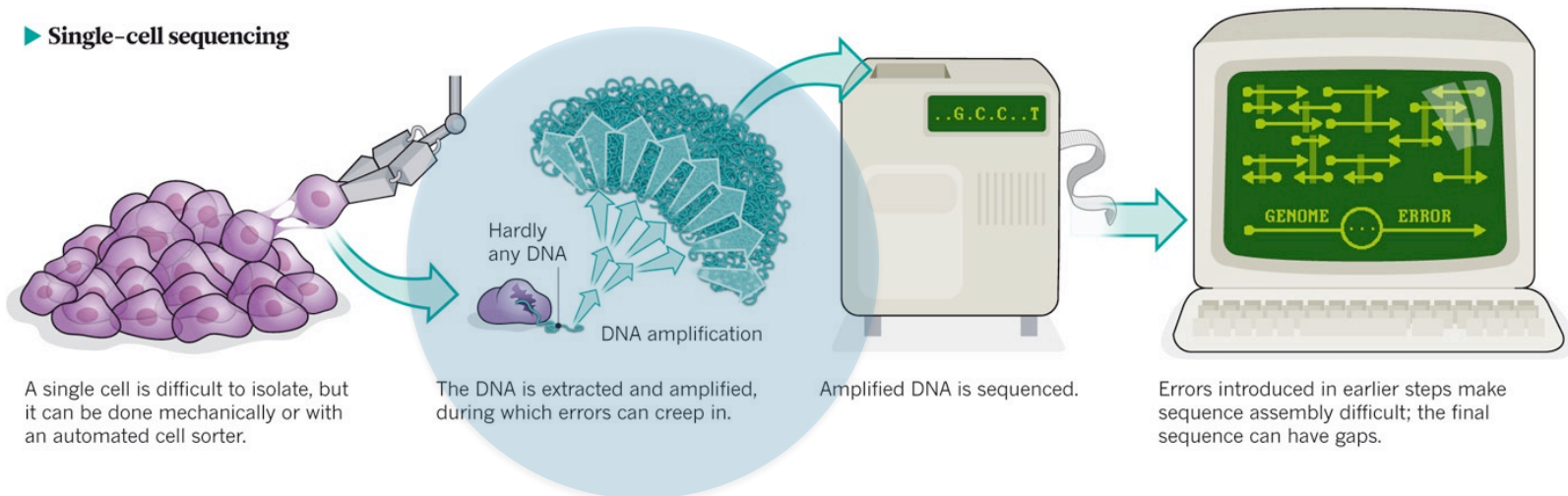
ONE GENOME FROM MANY

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

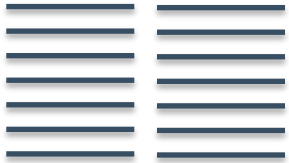
► Standard genome sequencing



► 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

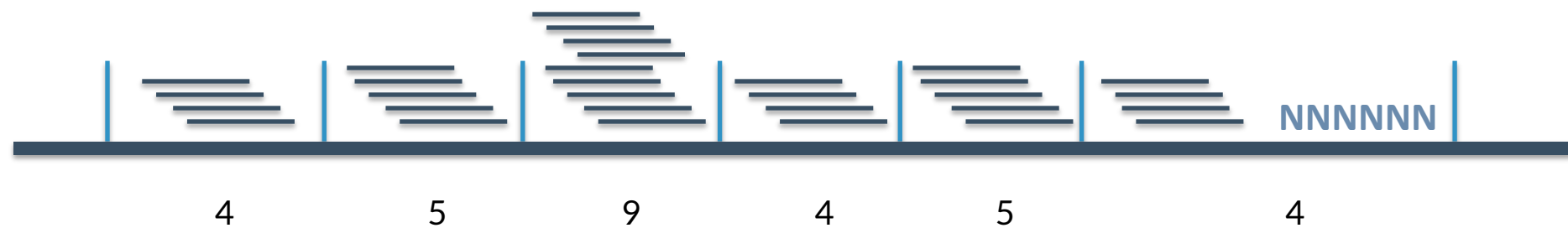
<1X coverage, often <0.1X

Copy-number variant analysis

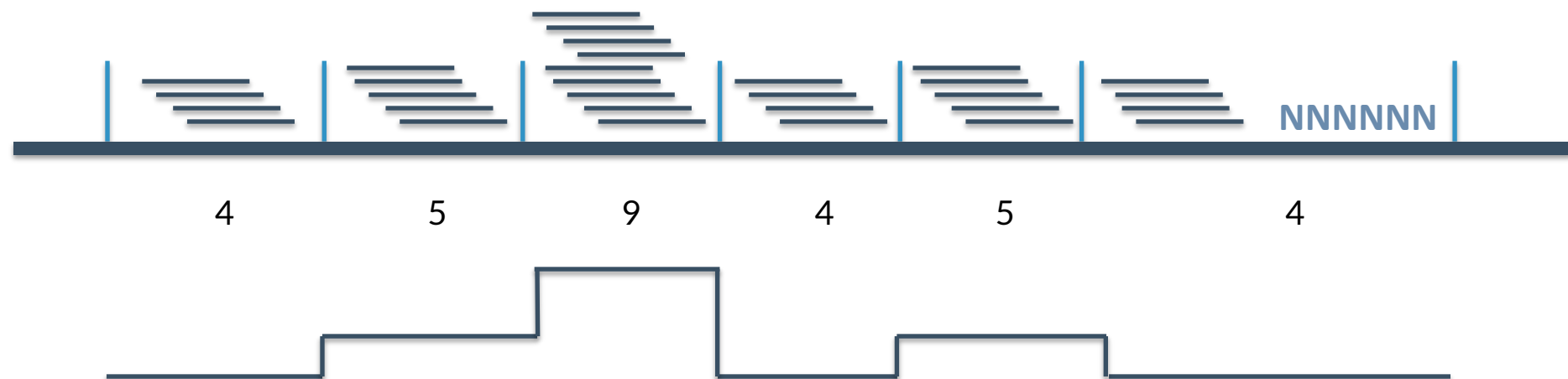


Divide genome into “bins” with ~50 – 100 reads / bin

Copy-number variant analysis



Copy-number variant analysis



Copy-number variant analysis



4

5

9

4

5

4



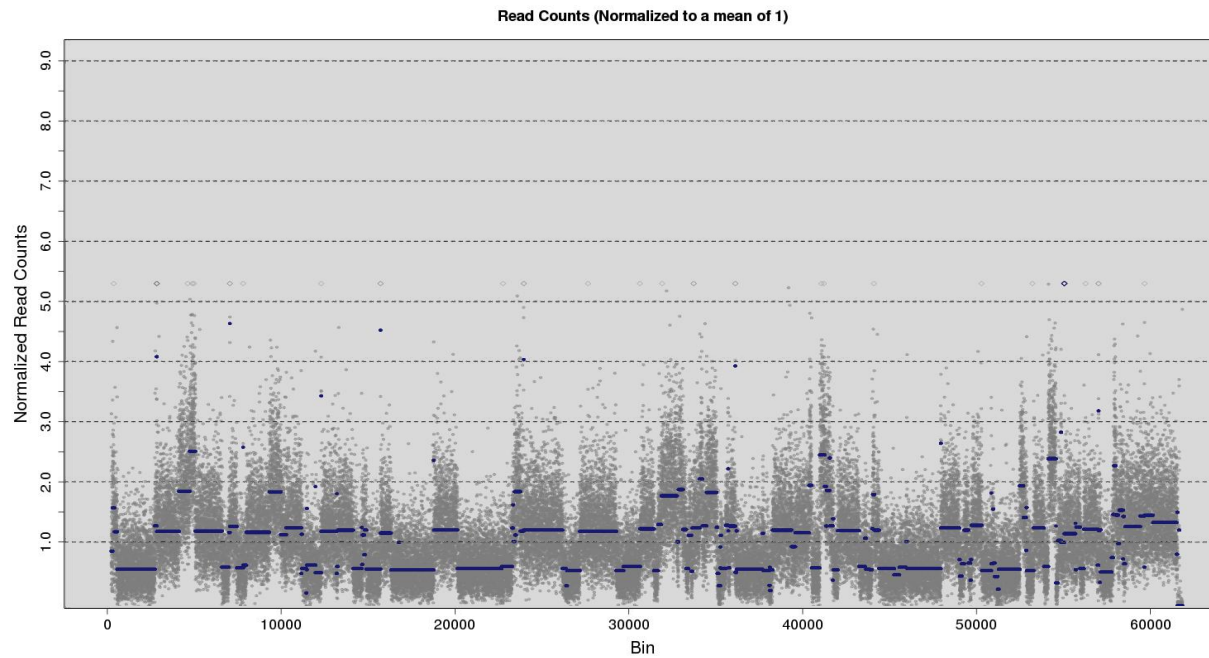
4.5

9

4

Circular Binary Segmentation (CBS)
to reduce noise in data

Copy-number variant analysis



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

Outline

Introduction



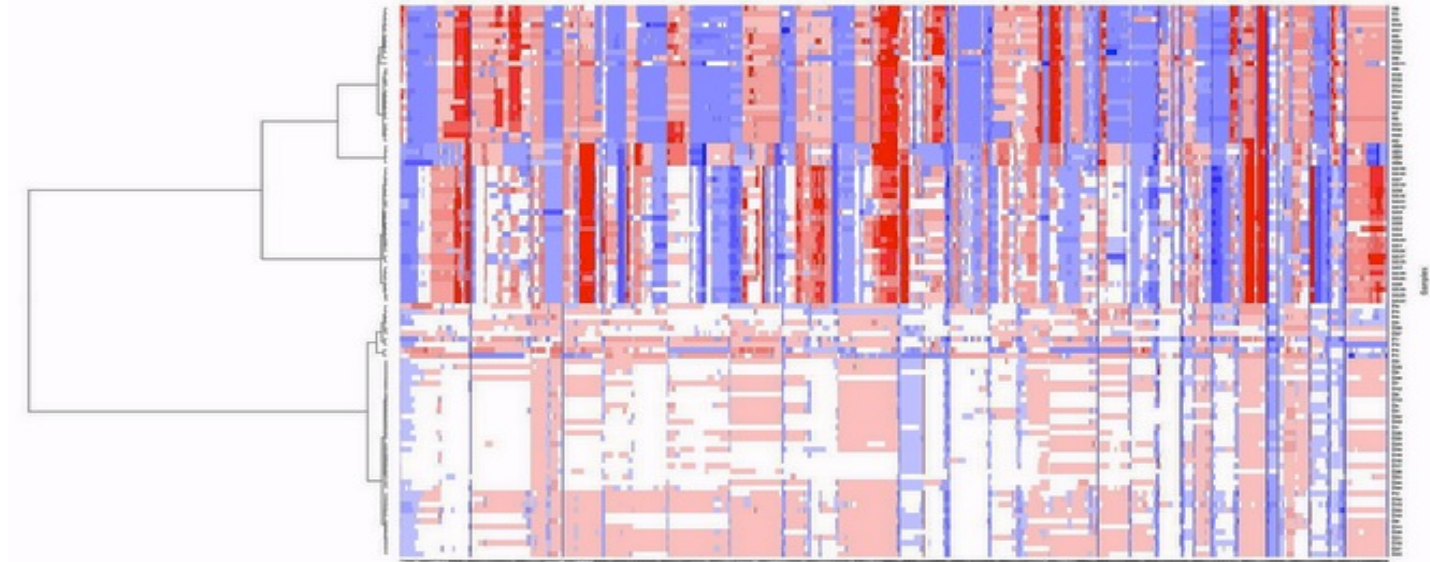
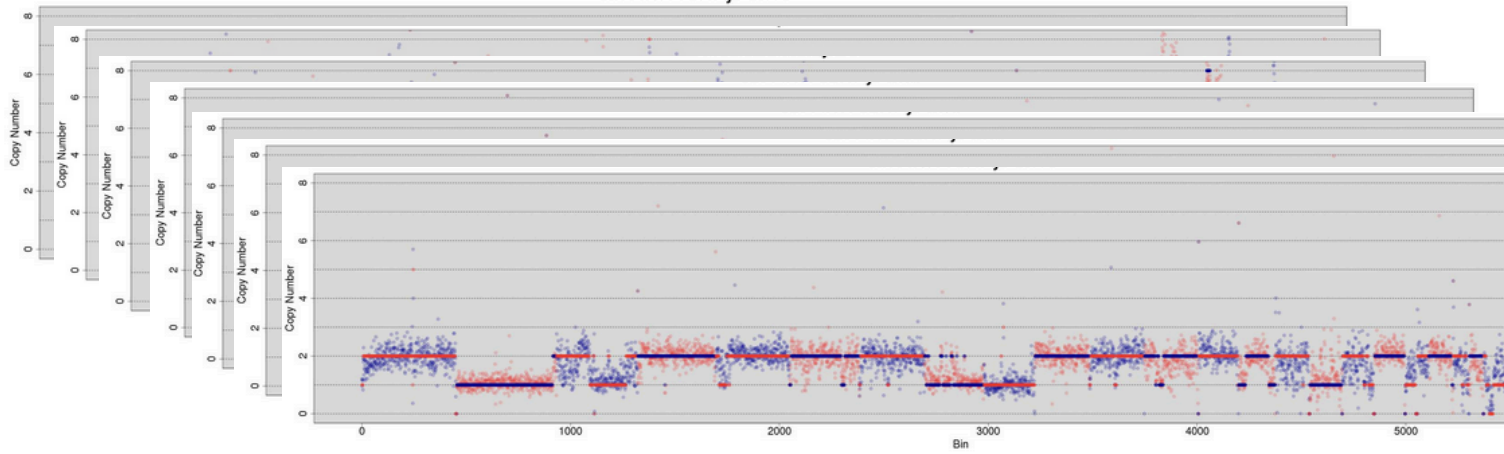
Ginkgo



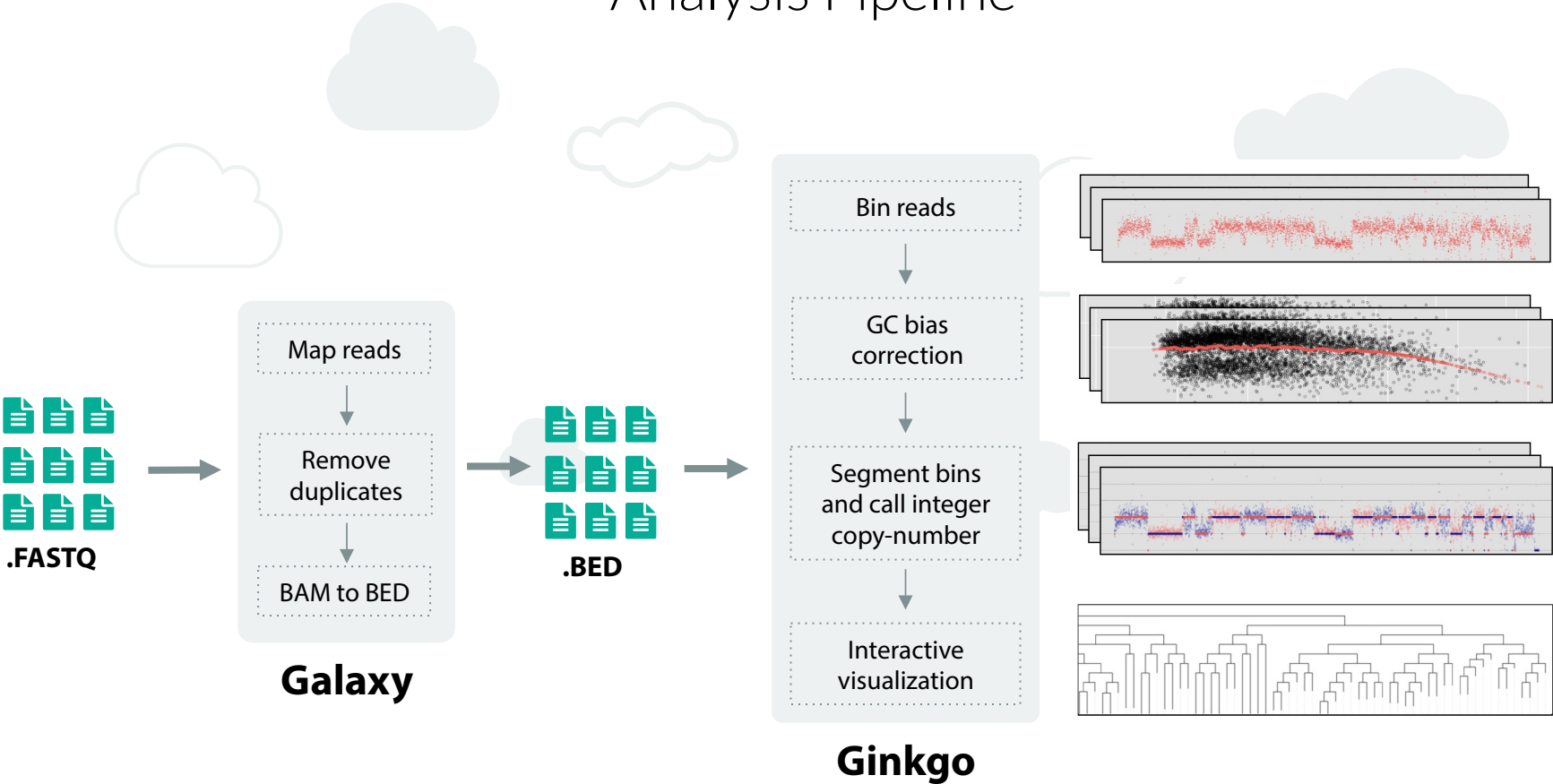
Comparison of
WGA methods



Copy-number variant analysis



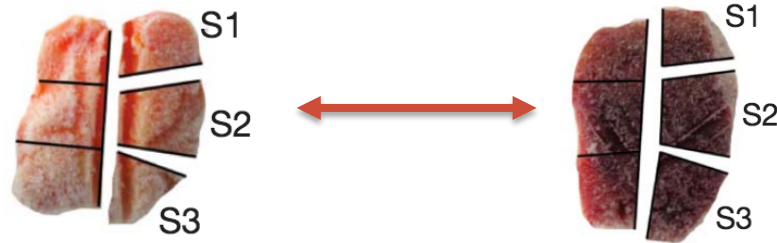
Analysis Pipeline



Ginkgo Demo

Tumour evolution inferred by single-cell sequencing

Nicholas Navin^{1,2}, Jude Kendall¹, Jennifer Troge¹, Peter Andrews¹, Linda Rodgers¹, Jeanne McIndoo¹, Kerry Cook¹, Asya Stepanky¹, Dan Levy¹, Diane Esposito¹, Lakshmi Muthuswamy³, Alex Krasnitz¹, W. Richard McCombie¹, James Hicks¹ & Michael Wigler¹



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

Ginkgo

qb.cshl.edu/ginkgo/?q=/NHeohEpiCXcE7hqmLx2c

Ginkgo

A web tool for analyzing single-cell sequencing data.

Sample analyses ▾ Load previous analysis ▾

STEP 0 Upload your .bed files (We accept *.bed and *.bed.gz, max 1GB/file, min 3 cells)

+ Add files... Cancel upload

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

M16_SRR089711.bed.gz	56.85 MB	Delete
M16_SRR090212.bed.gz	41.11 MB	Delete
P16_SRR089586.bed.gz	53.07 MB	Cancel
P16_SRR089593.bed.gz	51.54 MB	Cancel
P16_SRR089604.bed.gz	43.70 MB	Delete
P16_SRR089646.bed.gz	43.32 MB	Cancel
P16_SRR089659.bed.gz	40.86 MB	Cancel

Next step >

View analysis later

Access your results later at the following address:

```
http://qb.cshl.edu/ginkgo?
q=results/NHeohEpiCXcE7hqmLx2c
```

Help

Sample .bed file

chrom	chromStart	chromEnd
chr1	555485	555533
chr1	676584	676632
chr1	745136	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

qb.cshl.edu/ginkgo/?q=dashboard/NHeohEpiCXcE7hqmLx2c

Ginkgo

Your files are uploaded. Now let's do some analysis:

STEP 1 Choose cells for analysis

Select all cells

- M16_SRR089711.bed.gz
- M16_SRR090212.bed.gz

View analysis later

Access your results later at the following address:

```
http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c
```

STEP 2 Set analysis options

Job name:

Genome:

STEP 3 E-mail notification

If you want to be notified once the analysis is done, enter your e-mail here:

OPTIONAL Advanced parameters

[Manage Files](#) [Start Analysis](#)

Ginkgo

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

Ginkgo

Ginkgo

Single-cell analysis

3% complete.

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

STEP 3 View results

← Analysis Options

View analysis later

Access your results later at the following address:

```
http://qb.cshl.edu/ginkgo?  
q=results/NHeohEpiCXcE7hqmLx2c
```

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

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

Ginkgo

Ginkgo

Single-cell analysis

95% complete.

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

STEP 3 View results

← Analysis Options

View analysis later

Access your results later at the following address:

```
http://qb.cshl.edu/ginkgo?  
q=results/NHeohEpiCXcE7hqmLx2c
```

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Ginkgo Single-cell analysis

Analysis complete!

Diploids from both primary and metastasis cluster

STEP 3 View results

Tree

P16_SRR089604
P16_SRR089593
M16_SRR089733
M16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
P16_SRR089711
M16_SRR089711
M16_SRR089711
M16_SRR090158
M16_SRR090212
M16_SRR090144
M16_SRR090156

View analysis later

Access your results later at the following address:

[http://qb.cshl.edu/ginkgo?](http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c)

Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55

Tree display

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/P16_SRR089593

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

Ginkgo

Single-cell analysis

Analysis complete!

Diploids from both primary and metastasis cluster

STEP 3 View results

Tree

P16_SRR089604
 P16_SRR089593
 M16_SRR089733
 M16_SRR089711
P16_SRR089586
 P16_SRR089646
 P16_SRR089564
 P16_SRR089564
 P16_SRR089564
 P16_SRR089564
 P16_SRR089564
 P16_SRR089564
 M16_SRR090144
 M16_SRR090156

View analysis later

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c>

Parameters

Genome: hg19

Binning: variable bins of 500kb size (bowtie/101bp reads)

Segmentation: using normalized read counts

Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts (newick | xml | pdf | jpeg)
 Copy-number (newick | xml | pdf | jpeg)
 Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> M16_SRR089586		6,211,621	3,388.33	1,866.61	0.55

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

Ginkgo

Single-cell analysis

Analysis complete!

Non-diploids from primary and metastasis segregate

STEP 3 View results

Tree

P16_SRR089604
P16_SRR089593
M16_SRR089733
M16_SRR089711
P16_SRR089586
P16_SRR089646
P16_SRR089664
P16_SRR089663
P16_SRR089662
P16_SRR089661
P16_SRR089660
M16_SRR089659
M16_SRR089658
M16_SRR089657
M16_SRR089656
M16_SRR089655
M16_SRR089654
M16_SRR089653
M16_SRR089652
M16_SRR089651
M16_SRR089650

View analysis later

Access your results later at the following address:

`http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Parameters

Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> P16_SRR089646		1,621	3,388.33	1,866.61	0.55

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

Ginkgo

Single-cell analysis

Analysis complete!

Non-diploids from primary and metastasis segregate

STEP 3 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_SRR090156
M16_SRR090154
M16_SRR090152
M16_SRR090150

View analysis later

Access your results later at the following address:

`http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Bin size: 500kb
Reads per bin: 101bp
Normalization: normalized read counts
Distance metric: linkage, 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> P16_SRR089662		3,388,333	1,866.61	1,866.61	0.55

Tree display

- Normalized read counts (newick | xml | pdf | jpeg)
- Copy-number (newick | xml | pdf | jpeg)
- Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

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

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

Ginkgo

Single-cell analysis

Analysis complete!

Non-diploids from primary and metastasis segregate

STEP 3 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_SRR090158

View analysis later

Access your results later at the following address:

`http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads) using normalized read counts
Linkage: 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090158

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

Ginkgo

Single-cell analysis

Analysis complete!

Non-diploids from primary and metastasis segregate

STEP 3 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://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts (e.g., 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090144

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

Ginkgo

Single-cell analysis

Analysis complete!

STEP 3 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://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (1.46 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		11,757,193	2,192.28	1,252.84	0.57
<input type="checkbox"/> M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> M16_SRR090210		18,171,621	3,388.33	1,866.61	0.55

Analysis complete!

STEP 3 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://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Copy-number ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Correlations ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

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 (**283 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**).

Summary

<input type="checkbox"/>	M16_SRR089733		17,395,808	3,243.67	1,147.08	0.35
<input checked="" type="checkbox"/>	M16_SRR090144		11,757,193	2,192.28	1,252.84	0.57
<input checked="" type="checkbox"/>	M16_SRR090156		7,592,032	1,415.63	766.63	0.54
<input type="checkbox"/>	M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/>	M16_SRR090210		18,171,621	3,388.33	1,866.61	0.55
<input type="checkbox"/>	M16_SRR090212		17,659,321	3,292.81	1,741.57	0.53
<input type="checkbox"/>	P16_SRR089586		15,074,760	2,810.88	910.36	0.32

With selected cells, plot:

Analysis complete!

STEP 3 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://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Tree display

Normalized read counts ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Copy-number ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Correlations ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Summary

Cell	Reads	Bin Count	Mean	Stdev	CV
<input type="checkbox"/> M16_SRR090158		6,414,000	1,195.97	658.59	0.55
<input type="checkbox"/> M16_SRR090210		18,171,621	3,388.33	1,866.61	0.55
<input type="checkbox"/> M16_SRR090212		17,659,321	3,292.81	1,741.57	0.53
<input type="checkbox"/> P16_SRR089586		15,074,760	2,810.88	910.36	0.32
<input type="checkbox"/> P16_SRR089593		16,176,073	3,016.24	1,035.63	0.34
<input type="checkbox"/> P16_SRR089604		5,760,274	1,074.08	357.84	0.33

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 (**283 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 complete!

STEP 3 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://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Summary

<input type="checkbox"/>	M16_SRR090210		18,171,621	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_SRR089586		15,074,760	2,810.88	910.36	0.32
<input type="checkbox"/>	P16_SRR089593		16,176,073	3,016.24	1,035.63	0.34
<input type="checkbox"/>	P16_SRR089604		5,760,274	1,074.08	357.84	0.33
<input checked="" type="checkbox"/>	P16_SRR089646		8,278,396	1,543.61	834.04	0.54
<input checked="" type="checkbox"/>	P16_SRR089659		14,513,617	2,706.25	1,459.16	0.54

Tree display

Normalized read counts ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Copy-number ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

Correlations ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

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 (**283 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 complete!

STEP 3 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

Summary

Cell ID	Reads	Bin Count	Mean	Stdev	CV
<input type="checkbox"/> M16_SRR090210		18,171,621	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_SRR089586		15,074,760	2,810.88	910.36	0.32
<input type="checkbox"/> P16_SRR089593		16,176,073	3,016.24	1,035.63	0.34
<input type="checkbox"/> P16_SRR089604		5,760,274	1,074.08	357.84	0.33
<input checked="" type="checkbox"/> P16_SRR089646		8,278,396	1,543.61	834.04	0.54
<input checked="" type="checkbox"/> P16_SRR089659		14,513,617	2,706.25	1,459.16	0.54

With selected cells, plot: **Read count** Lorenz curve GC bias MAD

View analysis later

Access your results later at the following address:

`http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

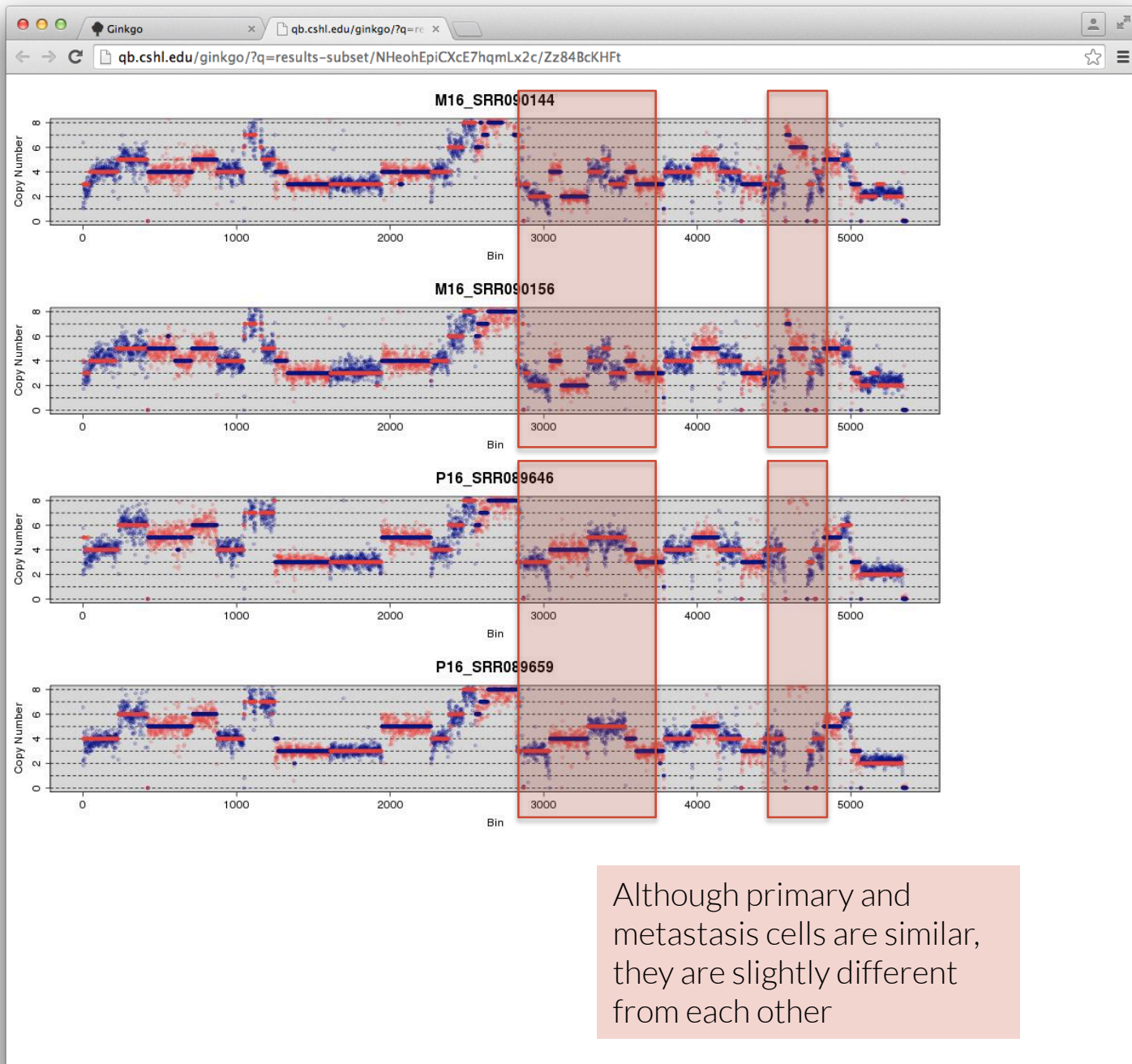
Genome: hg19
Binning: variable bins of 500kb size (bowtie/101bp reads)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

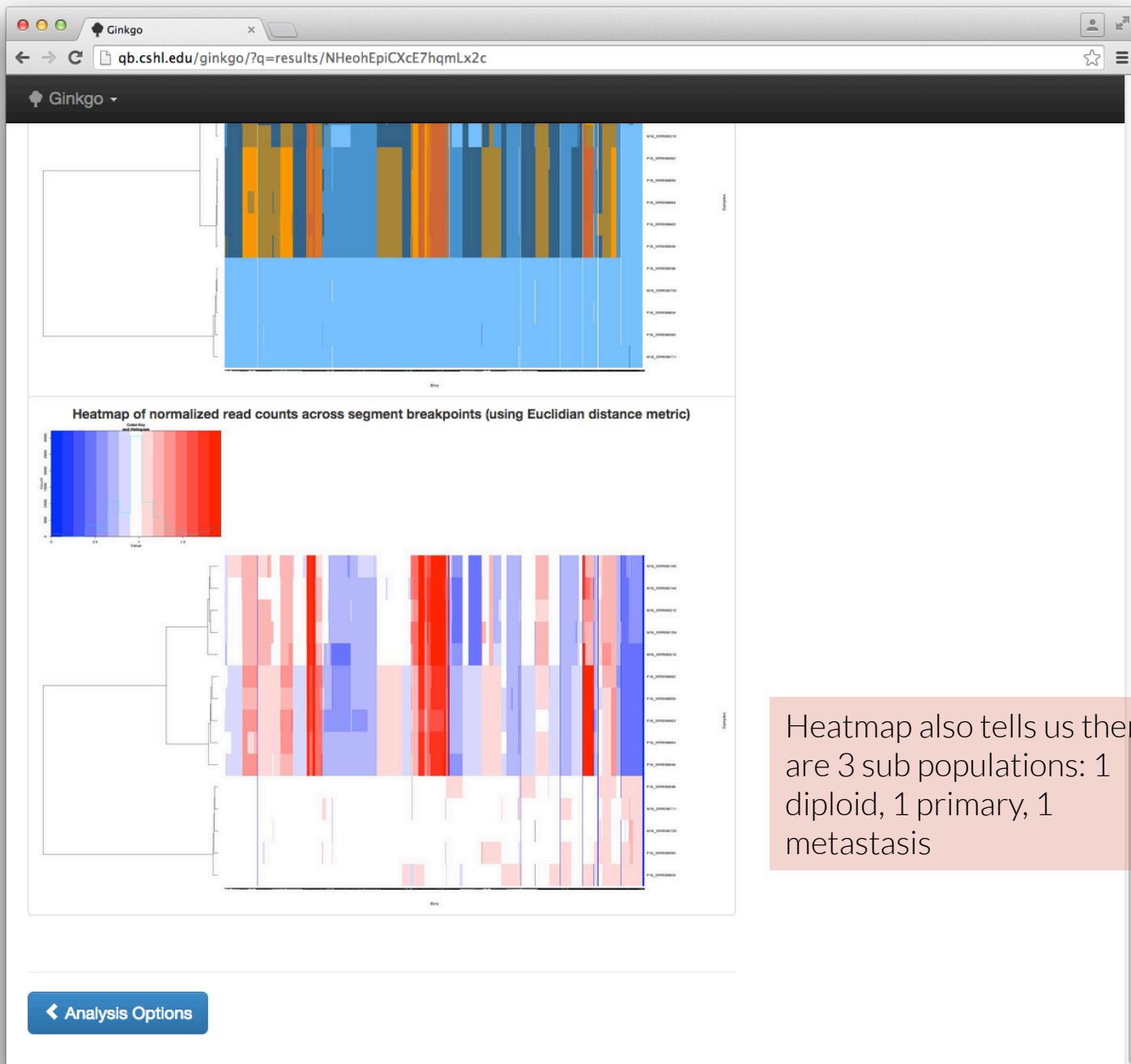
Tree display

Normalized read counts (newick | xml | pdf | jpeg)
 Copy-number (newick | xml | pdf | jpeg)
 Correlations (newick | xml | pdf | jpeg)

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 (283 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).





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

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

Ginkgo

Single-cell analysis

Analysis complete!

Can click on cells to get in-depth view

STEP 3 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_SRR090210
M16_SRR090210

View analysis later

Access your results later at the following address:

`http://qb.cshl.edu/ginkgo?q=results/NHeohEpiCXcE7hqmLx2c`

Analysis Parameters

Genome: hg19

bins of 500kb size (bowtie/101bp reads)

using normalized read counts

linkage, euclidean distance

Summary

Cell	Genomic Plot	Chromosome	Start (bp)	End (bp)	Reads	Distance
<input type="checkbox"/> M16_SRR090210		18,171,621	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_SRR089586		15,074,760	2,810.88	910.36	0.32	
<input type="checkbox"/> P16_SRR089593		16,176,073	3,016.24	1,035.63	0.34	
<input type="checkbox"/> P16_SRR089604		5,760,274	1,074.08	357.84	0.33	
<input checked="" type="checkbox"/> P16_SRR089646		8,278,396	1,543.61	834.04	0.54	
<input checked="" type="checkbox"/> P16_SRR089659		14,513,617	2,706.25	1,459.16	0.54	

Tree display

Normalized read counts (newick | xml | pdf | jpeg)

Copy-number (newick | xml | pdf | jpeg)

Correlations (newick | xml | pdf | jpeg)

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 (283 KB).

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

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

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

Ginkgo

Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210#

Ginkgo

Single-cell analysis

Analysis complete!

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

Viewing cell M16_SRR090210

[← Back to tree](#)

Interactive Profile Viewer

[View region in UCSC browser](#)

M16_SRR090210

Bin 4590
Position: chr18:6876114-7396581
Copy-Number: 9

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

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

Bin 4549
Position: chr17:66744135-67260606
Copy-Number: 4

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

Ginkgo x Human chr17:46497635- x

https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&position=chr17%3A46497635-81195210&hgsid=451141399_24ZAgO6j5KRDCeKcmwG...

Genomes Genome Browser Tools Mirrors Downloads My Data View Help About Us

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

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr17:46,497,635-81,195,210 34,697,576 bp. enter position, gene symbol or search terms go **hg38 replaces hg19 as default human assembly**

chr17 (q21.32-q25.3) 13.1 17p12 p11.2 q11.2 17q12 17q22 24.3 25.1 q25.3

Scale chr17: 50,000,000 55,000,000 60,000,000 65,000,000 70,000,000 75,000,000 80,000,000 hg19

Amplifications M16_SRR090210

Deletions

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

RefSeq Genes Publications: Sequences in Scientific Articles

Sequences Human mRNAs From GenBank

Human mRNAs Human ESTs That Have Been Spliced

Spliced ESTs 100

Layered H3K27Ac HSK27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

DNase Clusters DNase I Hypersensitivity Clusters in 125 cell types from ENCODE (V3)

Txn Factor ChIP 4.66 Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs

100 Vert. Cons 100 vertebrates Basewise Conservation by PhyloP

Rhesus Mouse Dog Elephant Chicken X_tropicalis Zebrafish Lambrey Multiz Alignments of 100 Vertebrates

Common SNPs (142) Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples

RepeatMasker Repeating Elements by RepeatMasker

move start Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end

< 2.0 >

track search default tracks default order hide all manage custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

Custom Tracks refresh

Amplifications Deletions

dense dense

Mapping and Sequencing refresh

Genes and Gene Predictions refresh

UCSC Genes RefSeq Genes AceView Genes Augustus CCDS Ensembl Genes

dense dense hide hide hide hide

17 EvoFold Exoniphy GENCODE... Geneid Genes Genscan Genes H-Inv 7.0

hide hide hide hide hide hide

IKMC Genes Mapped lincRNAs... LRG Transcripts MGC Genes N-SCAN Old UCSC Genes

hide hide hide hide hide hide

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

hide hide hide hide hide hide

We automatically upload amplification and deletion tracks to the UCSC browser

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210#

Ginkgo

Single-cell analysis

Analysis complete!

We can also annotate the CNV profile with genes of interest

Viewing cell M16_SRR090210

[← Back to tree](#)

Interactive Profile Viewer

[View region in UCSC browser](#)

M16_SRR090210

Bin 4590
Position: chr18:6876114-7396581
Copy-Number: 9

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210#

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: chr18:6876114-7396581
Copy-Number: 9

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

qb.cshl.edu/ginkgo/uploads/.../M16_SRR090210_CN.jpeg?uniq=1447...

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

Click + Drag: zoom in Shift + Click: move profile Double-click: zoom out Find gene: TP53 [Search](#) List genes in bin # [List](#)

Static Profile Viewer

Integer Copy Number Profile for Sample "M16_SRR090210"

Predicted Ploidy = 4.1

Copy Number

Chromosome

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

TP53

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

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

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

Ginkgo Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

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

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210#

Ginkgo

Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

BRCA2

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

qb.cshl.edu/ginkgo/?q=results/NHeohEpiCXcE7hqmLx2c/M16_SRR090210#

Ginkgo

Single-cell analysis

Analysis complete!

Viewing cell M16_SRR090210

[← Back to tree](#)

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

M16_SRR090210

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"

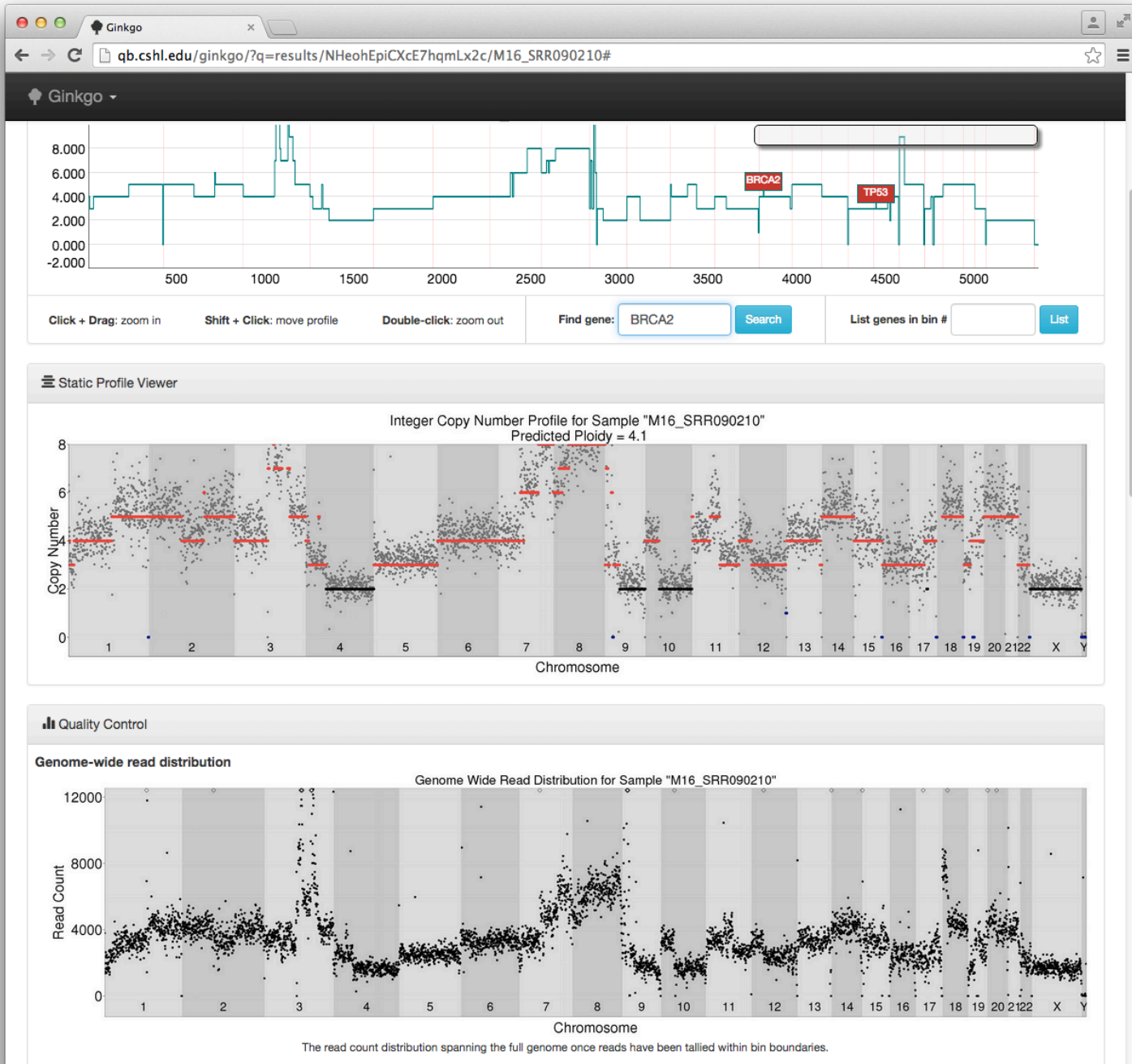
Predicted Ploidy = 4.1

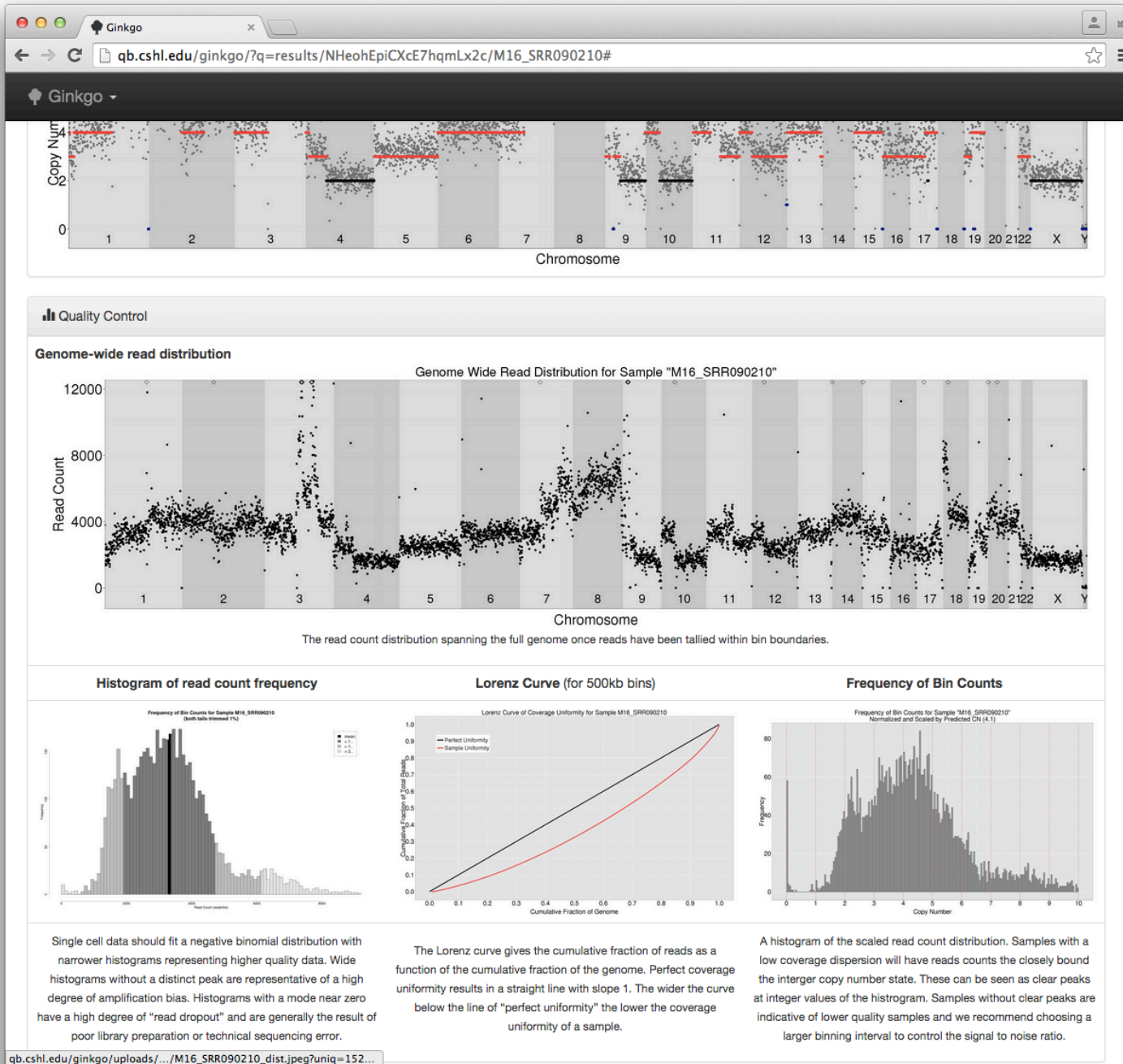
Copy Number

Chromosome



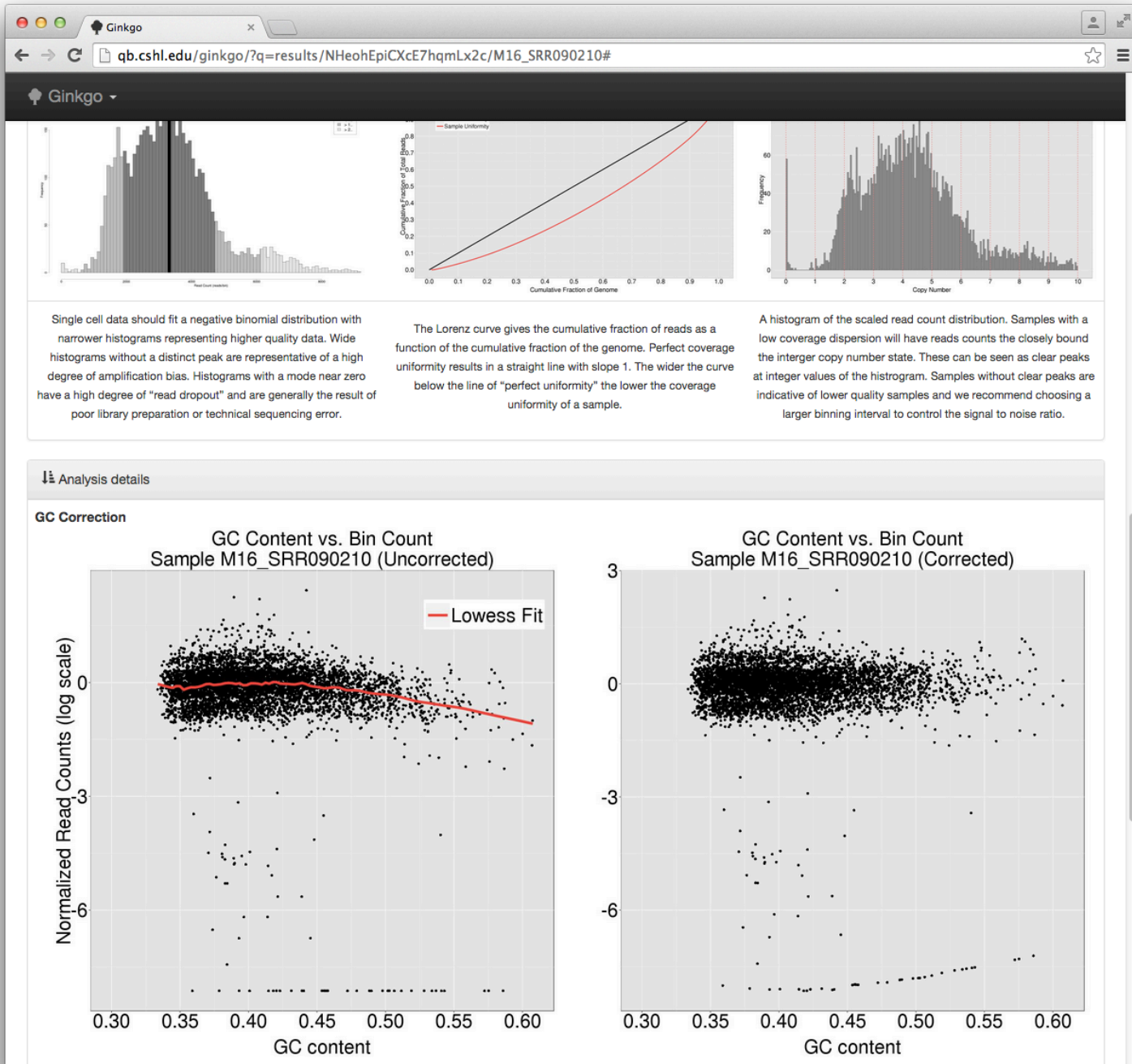
Static CNV profile



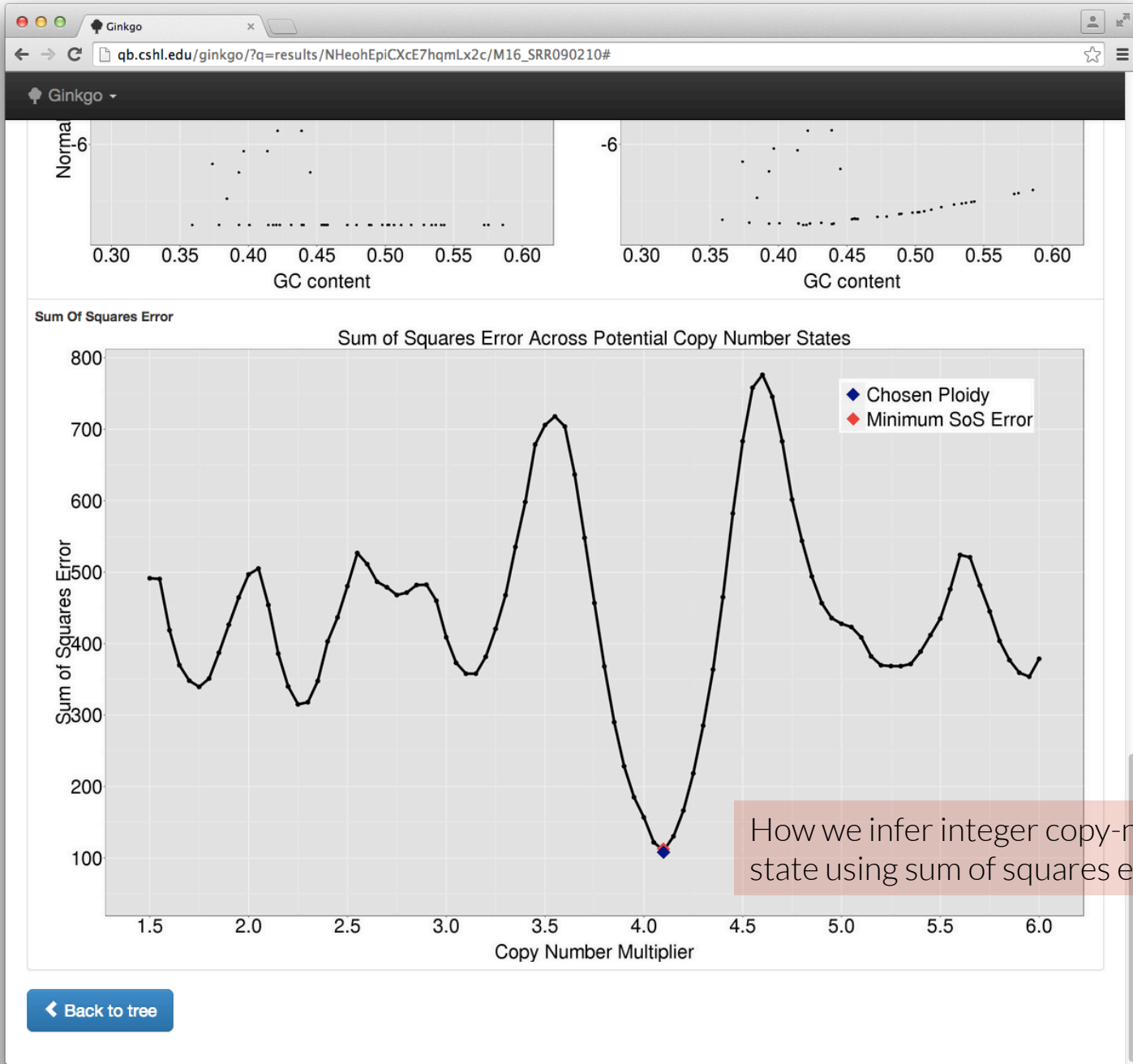


Original read count data before analysis

QC metrics



GC bias,
before and
after
correction



Outline

Introduction



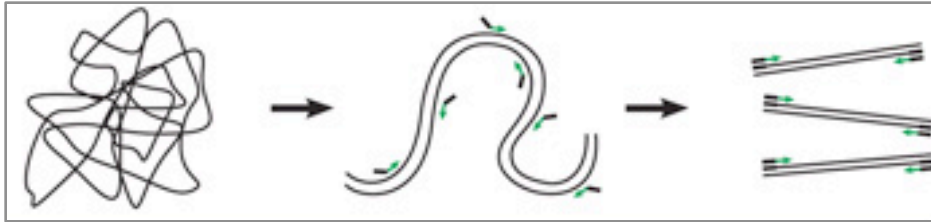
Ginkgo



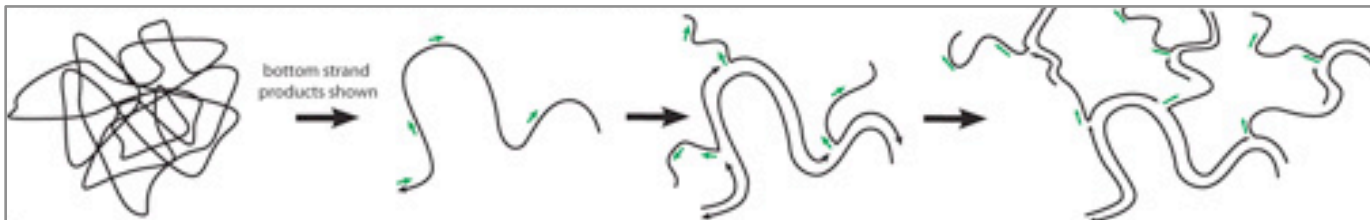
**Comparison of
WGA methods**



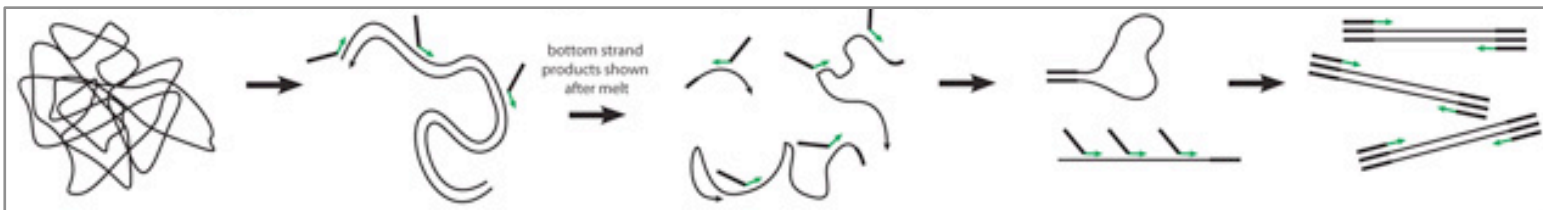
Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)



MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

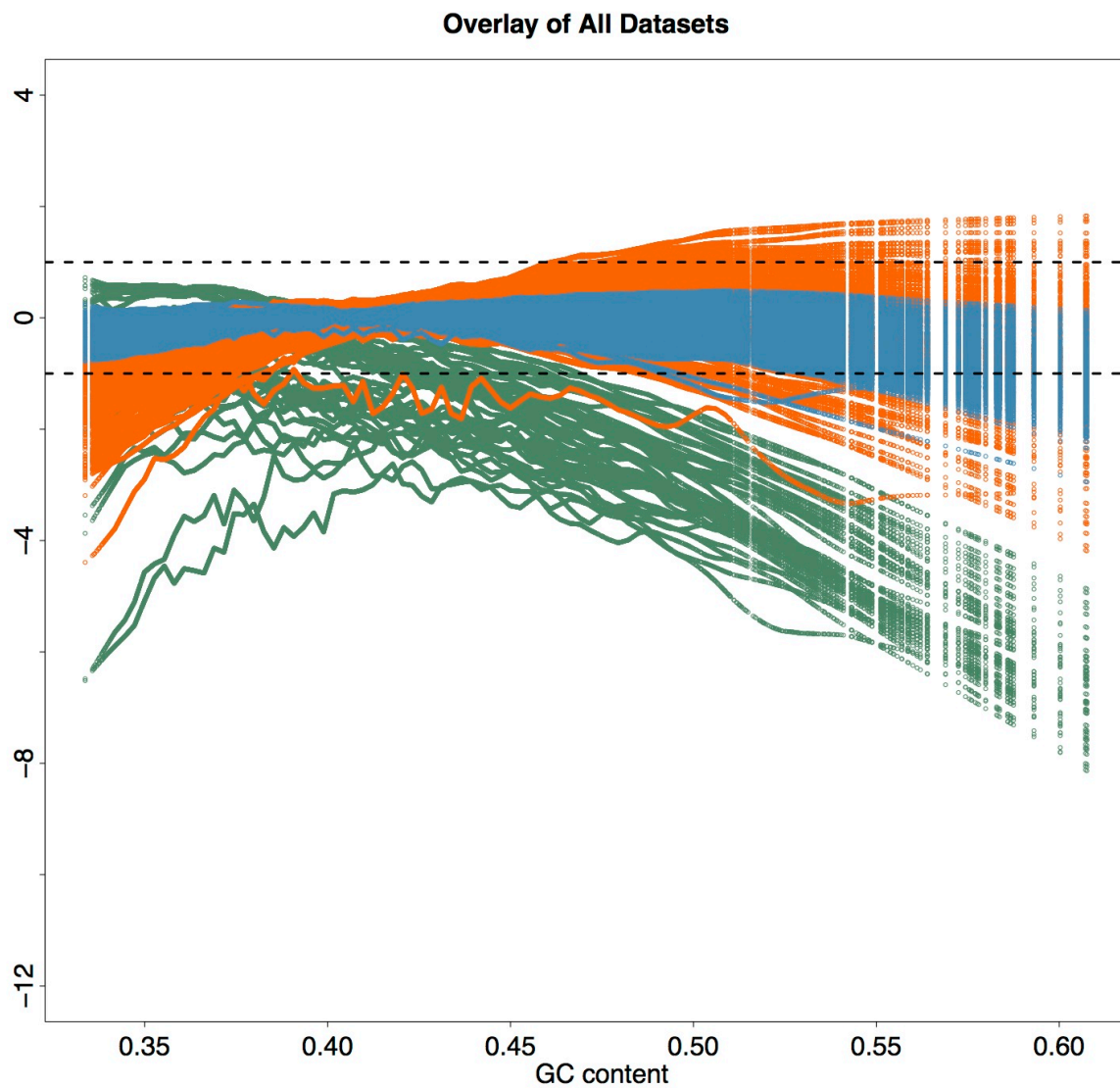
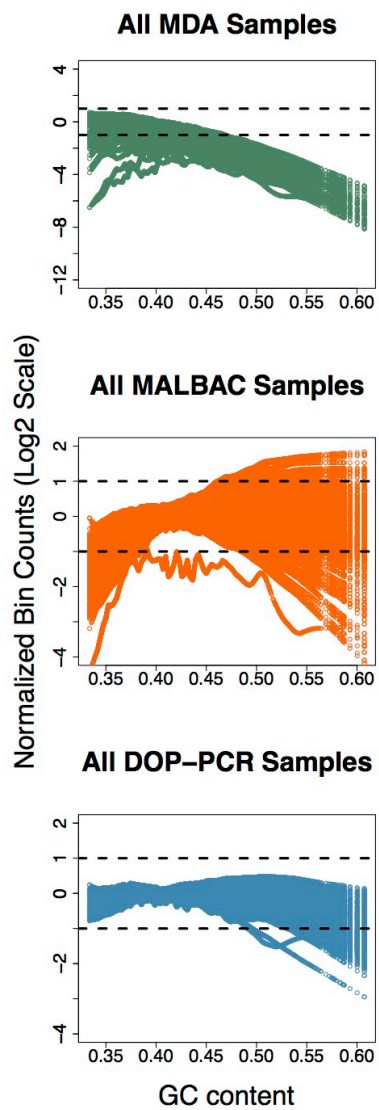
Comparison of WGA methods

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:

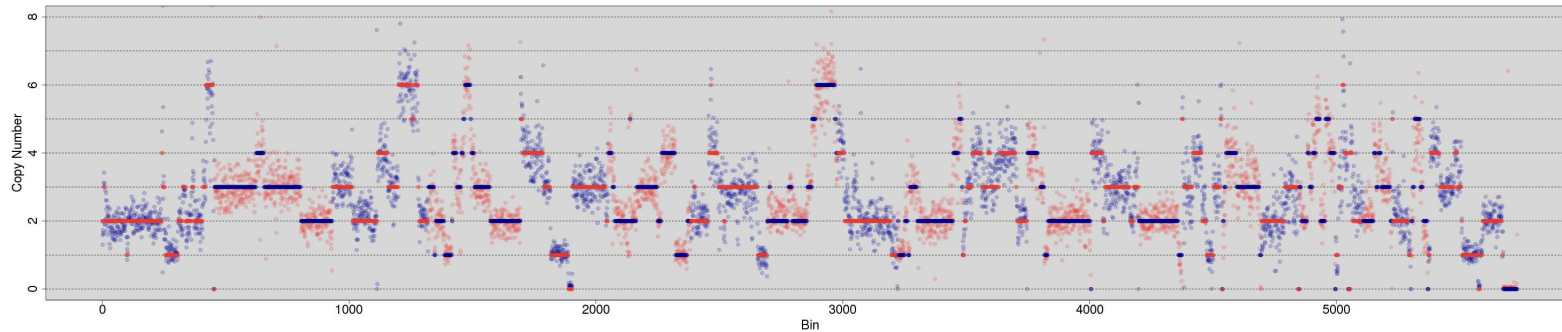
- 1) GC bias
- 2) Coverage dispersion

GC Bias

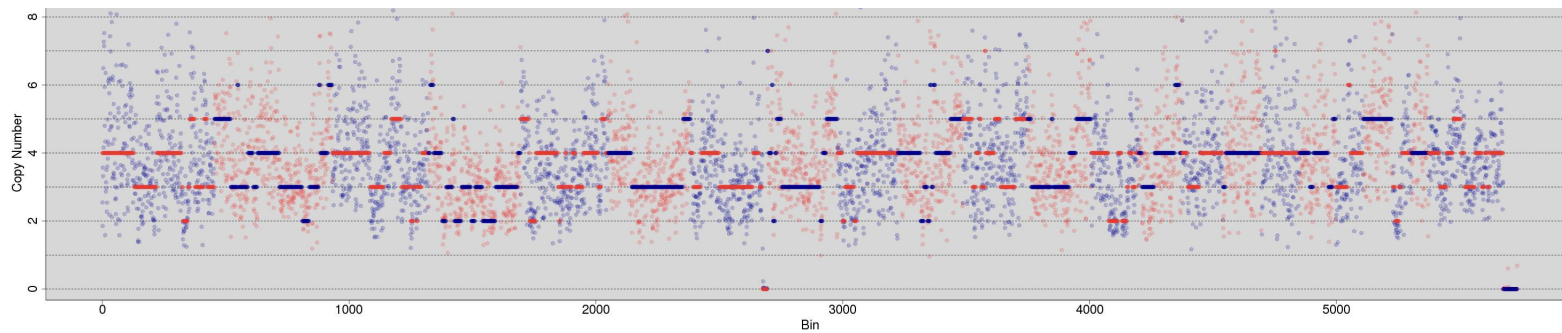


Coverage Dispersion

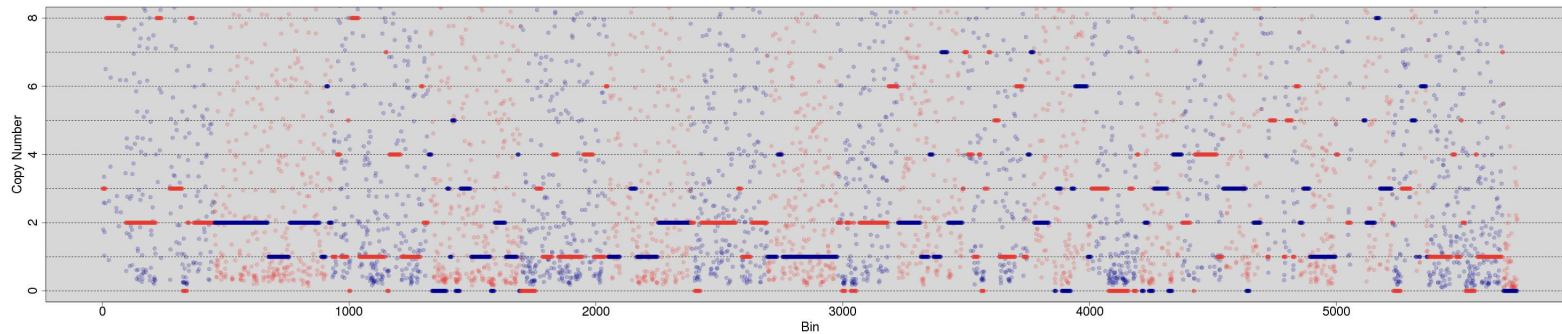
DOP-PCR



MALBAC

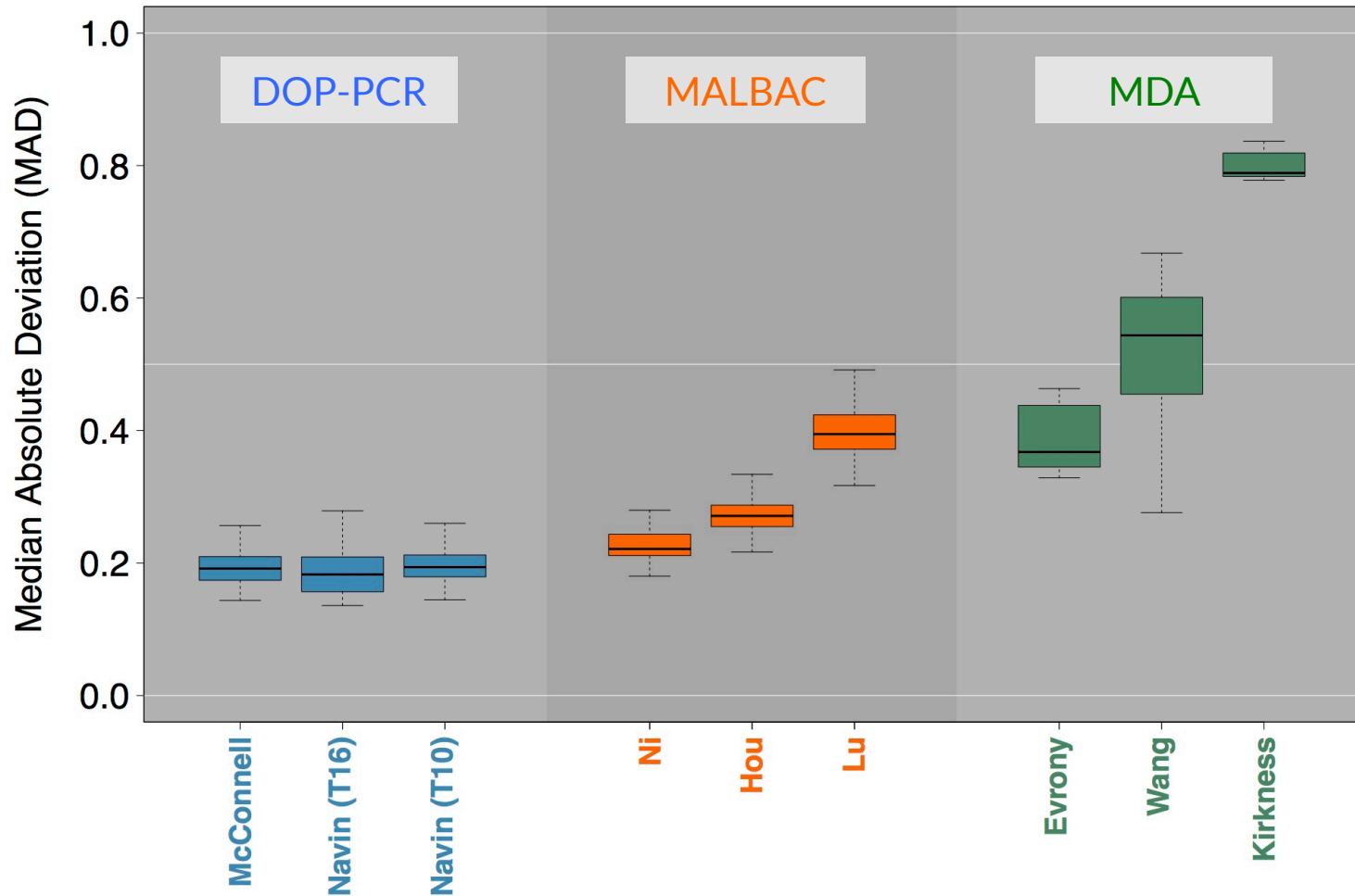


MDA



Coverage Dispersion

Median Absolute Deviation of Neighboring Bins



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

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