

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

 @Ty\_Garvin

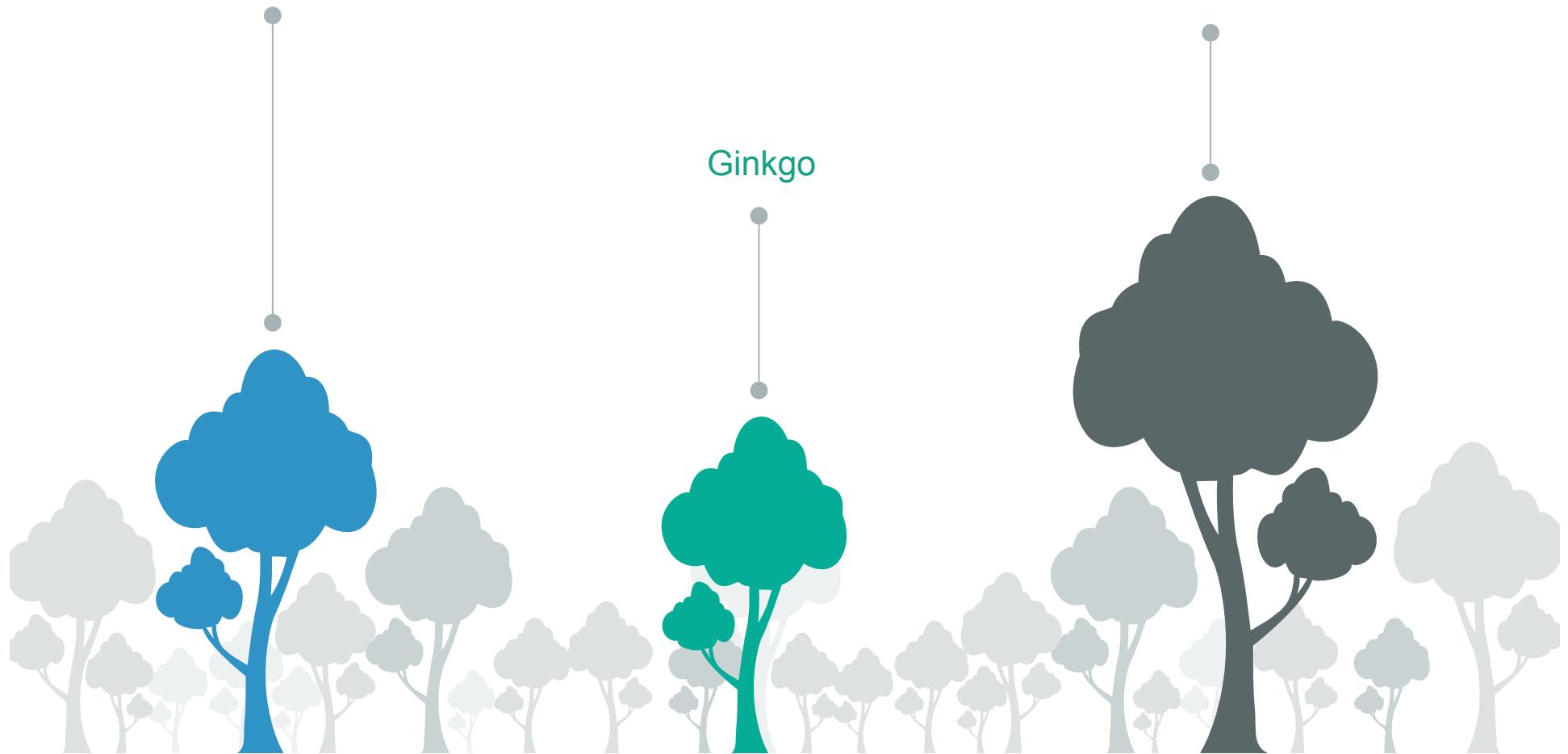
Tyler Garvin, Robert Aboukhalil, Jude Kendall, Timour Baslan,  
Gurinder S. Atwal, Jim Hicks, Michael Wigler, Michael C. Schatz



Cold Spring Harbor Laboratory

# Outline

Introduction

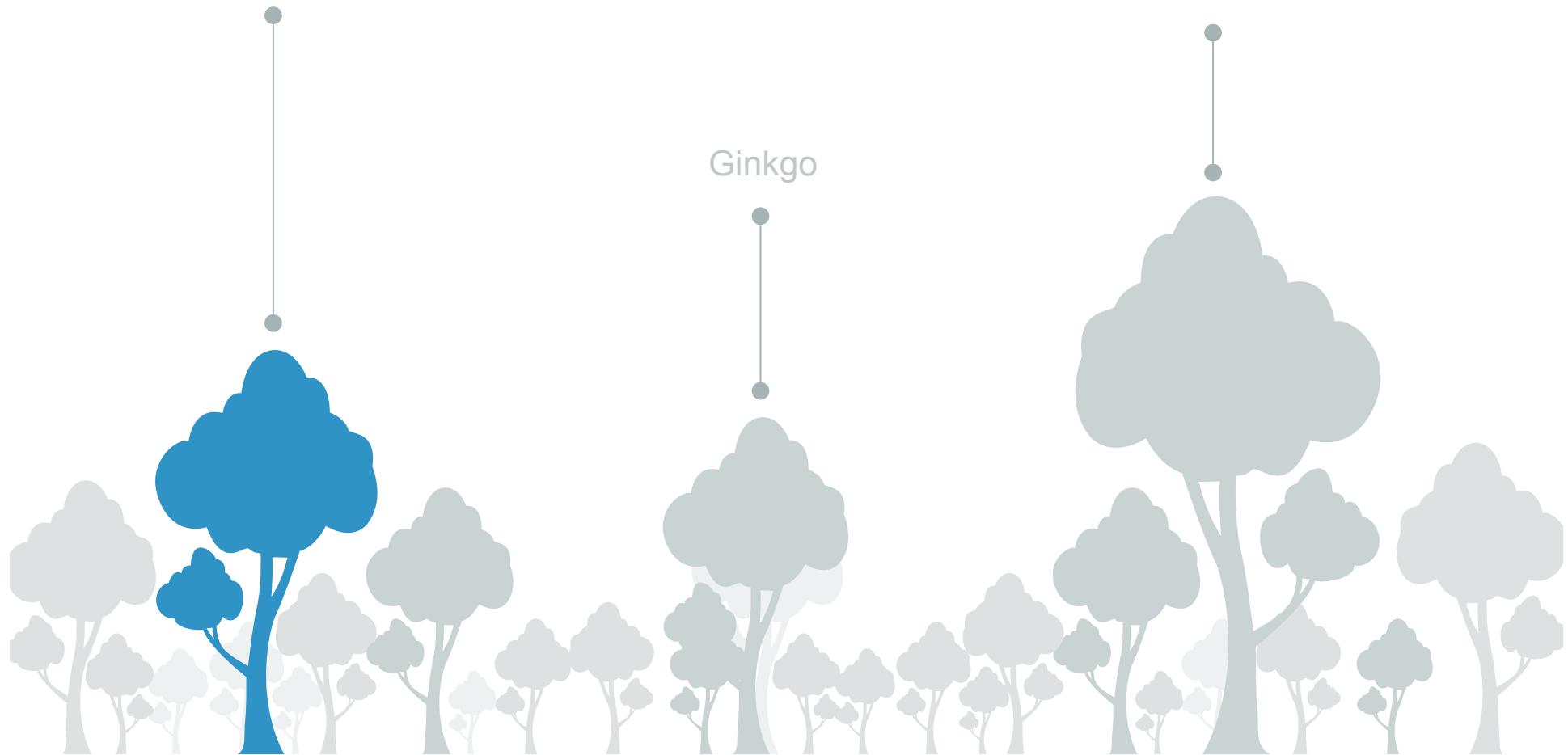


Comparison of  
WGA methods

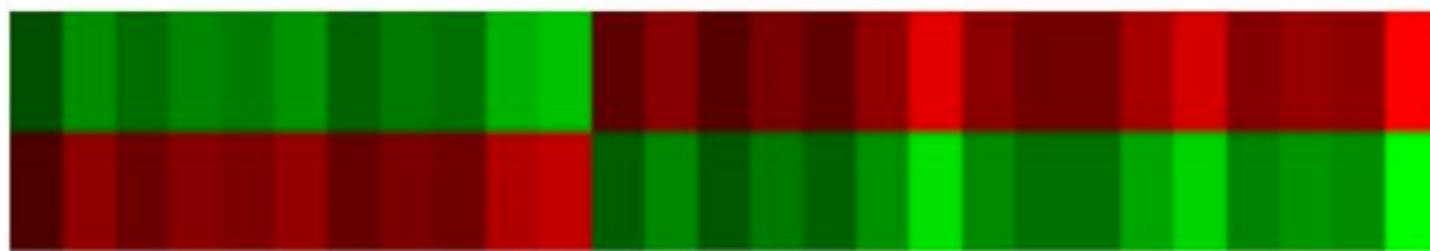
Ginkgo

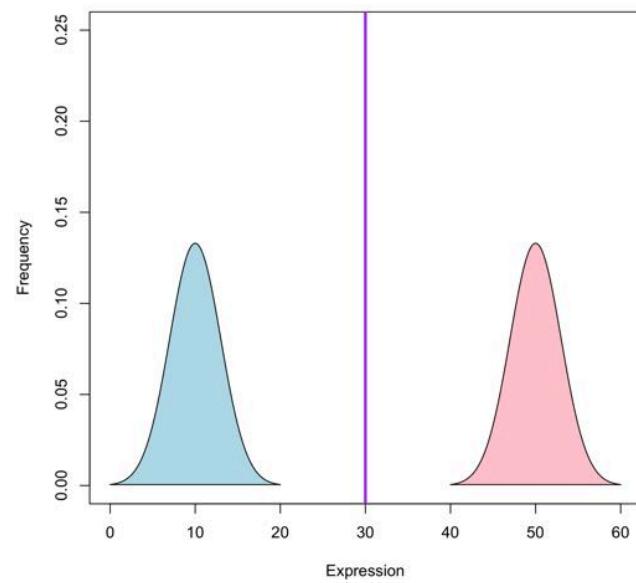
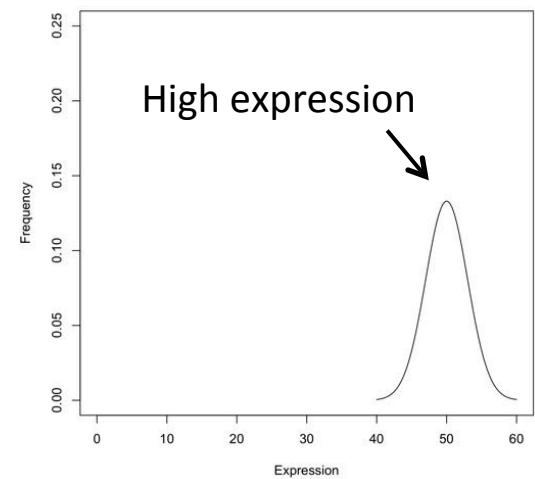
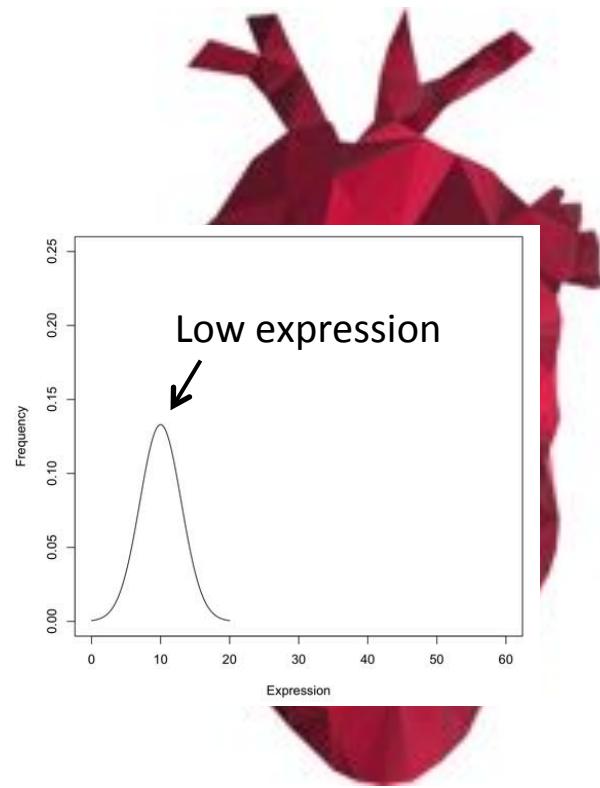
# Outline

Introduction

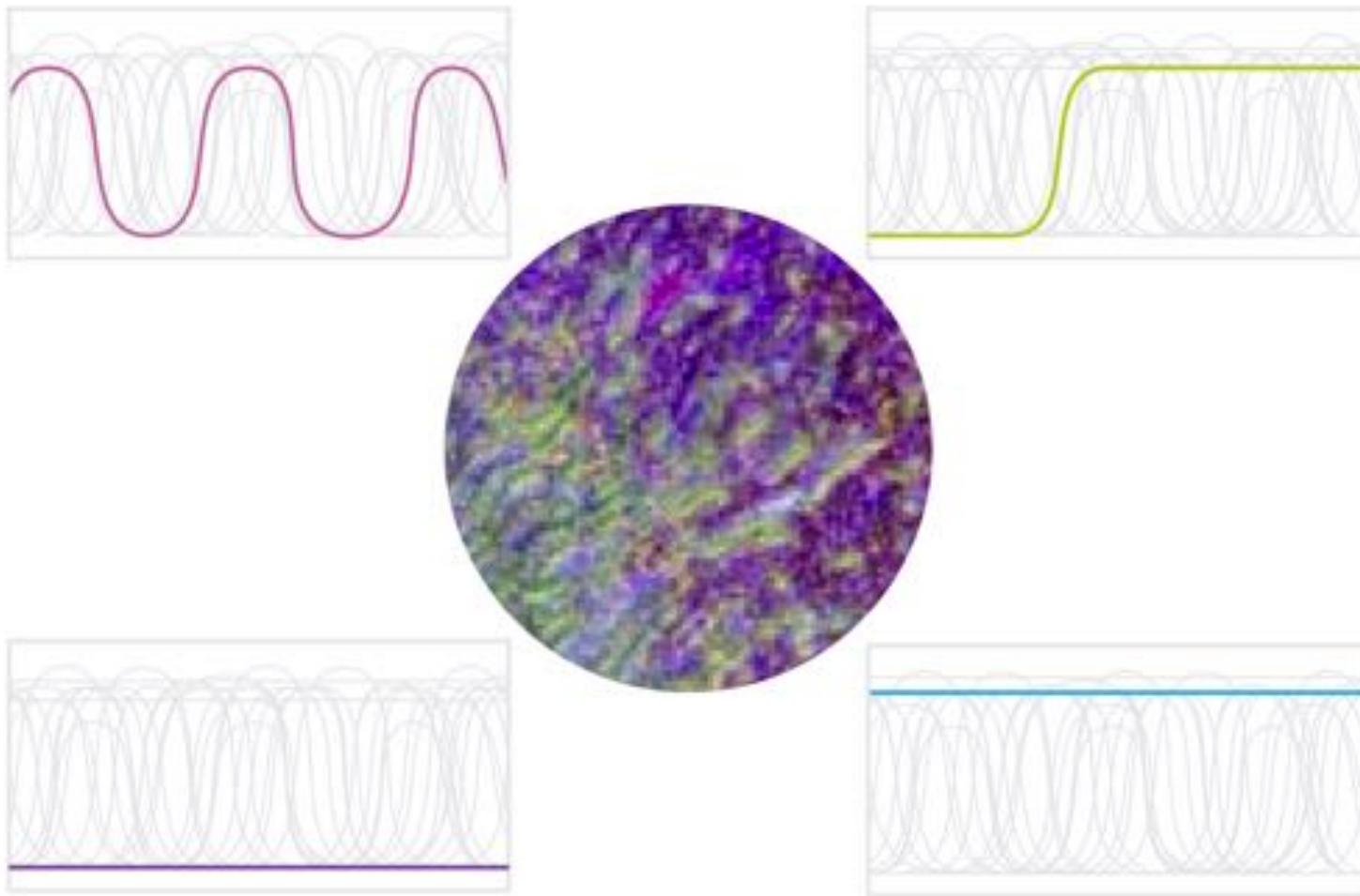


Why should we use single-cell sequencing over bulk sequencing?





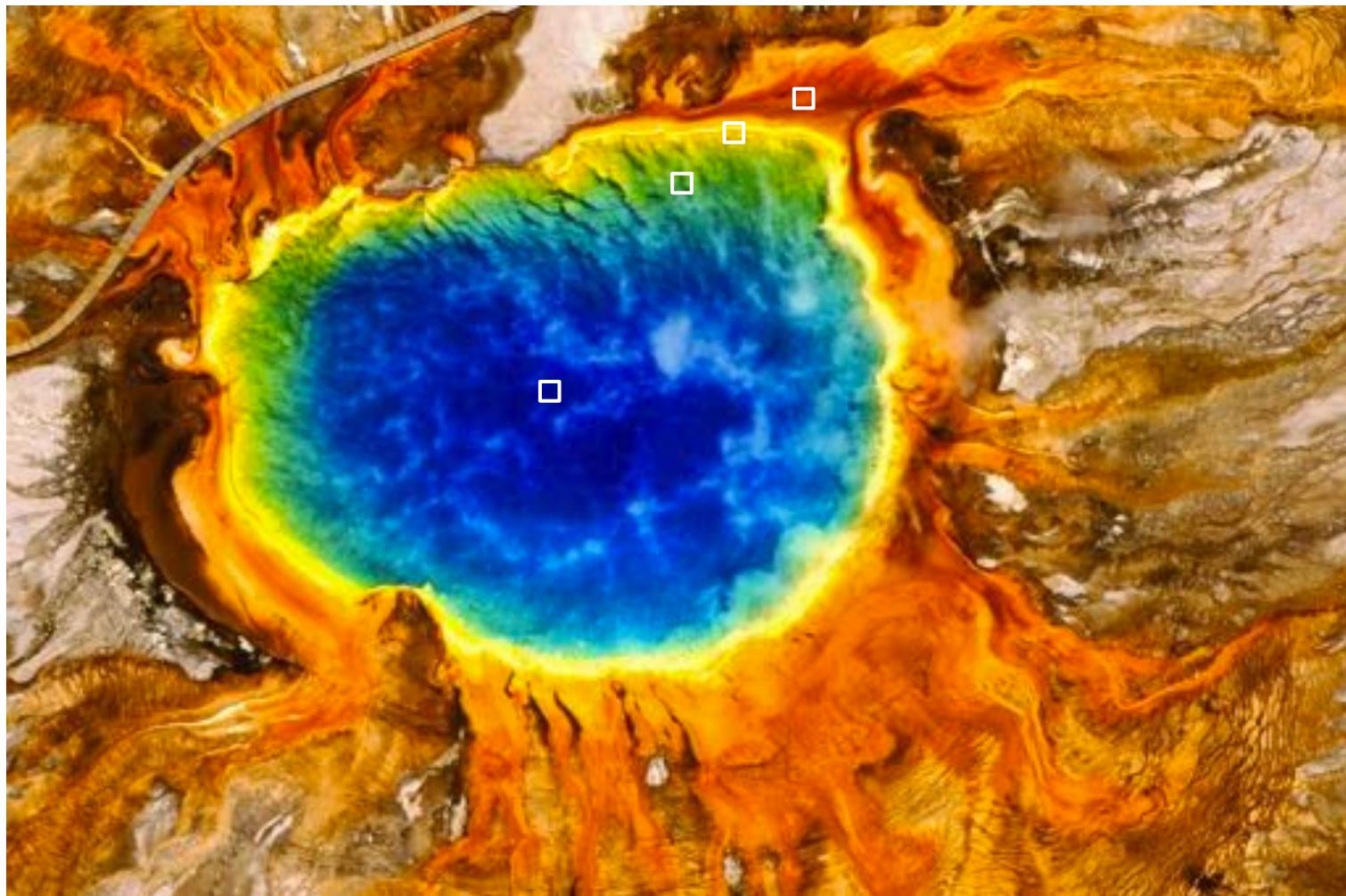
# Single cell expression profiles



*Single cell research. Illumina.*

# Heterogeneity

# Metagenomics



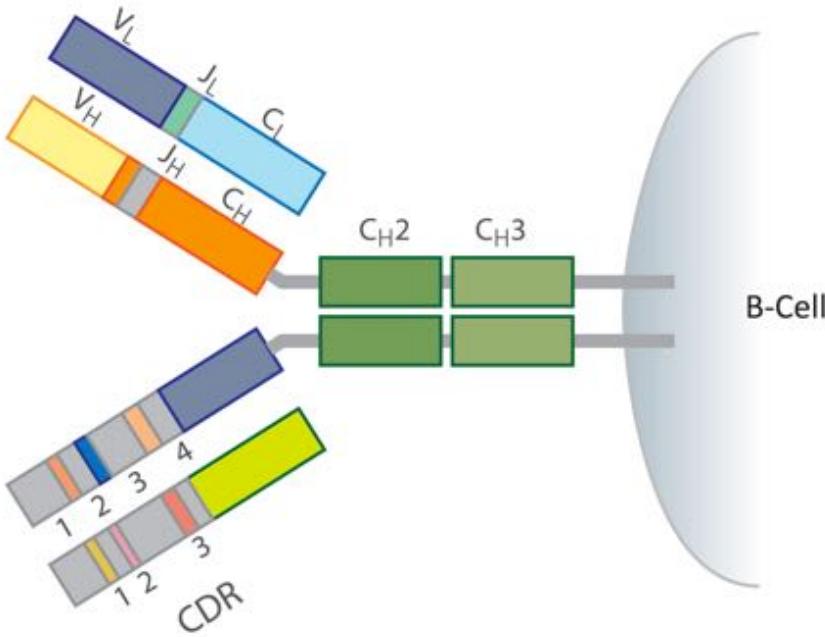
*Single cell research. Illumina.*



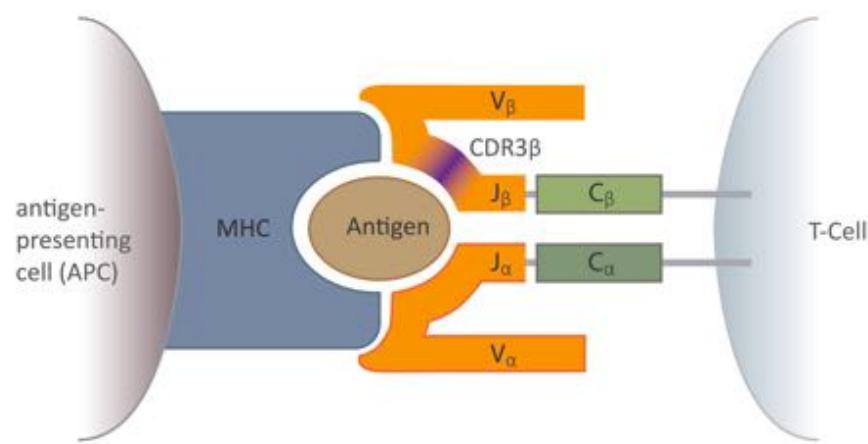
## Microbiome

- Human cells
- Bacteria
- Fungi
- Virus
- Temporal dynamics
- Diet-dependent dynamics
- Stress-related dynamics

# Immunology



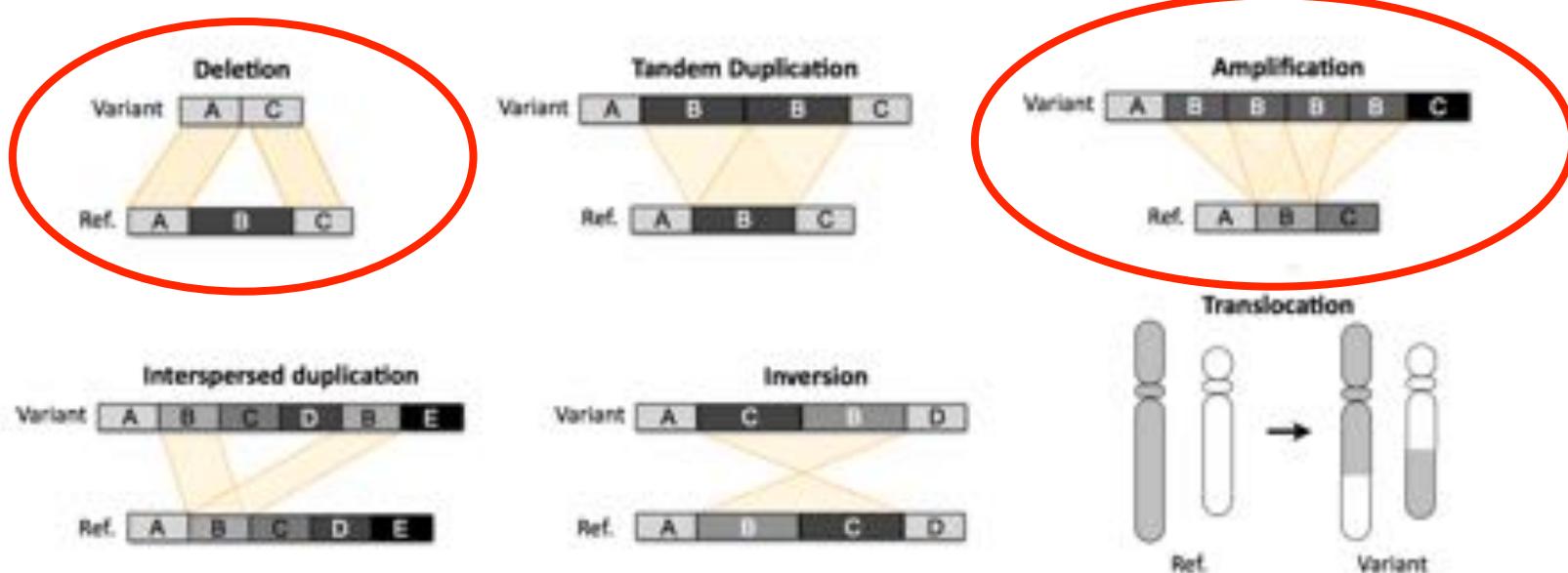
- **Massive diversity rivaled only by germ cells**
- **Somatic recombination**



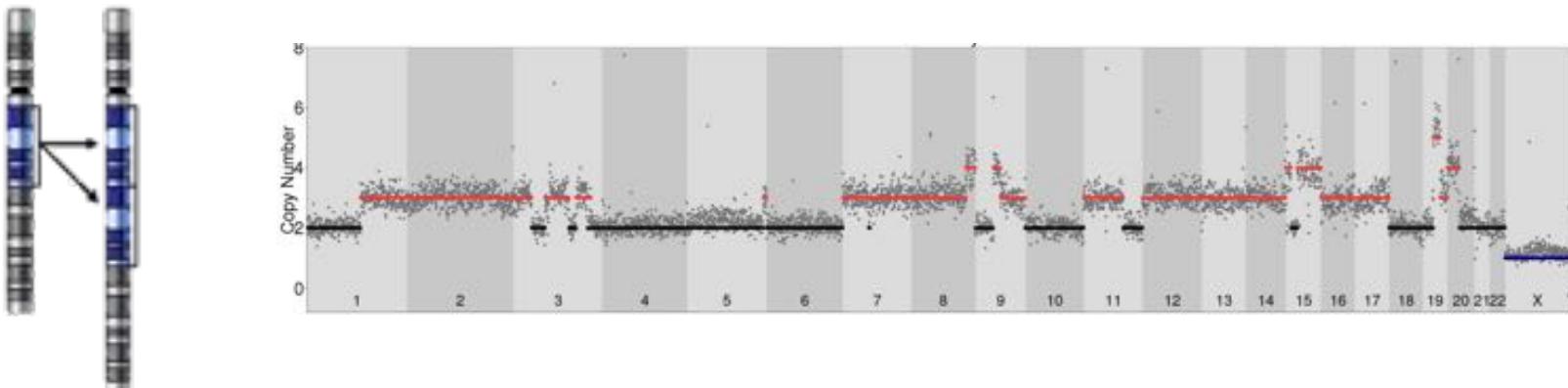
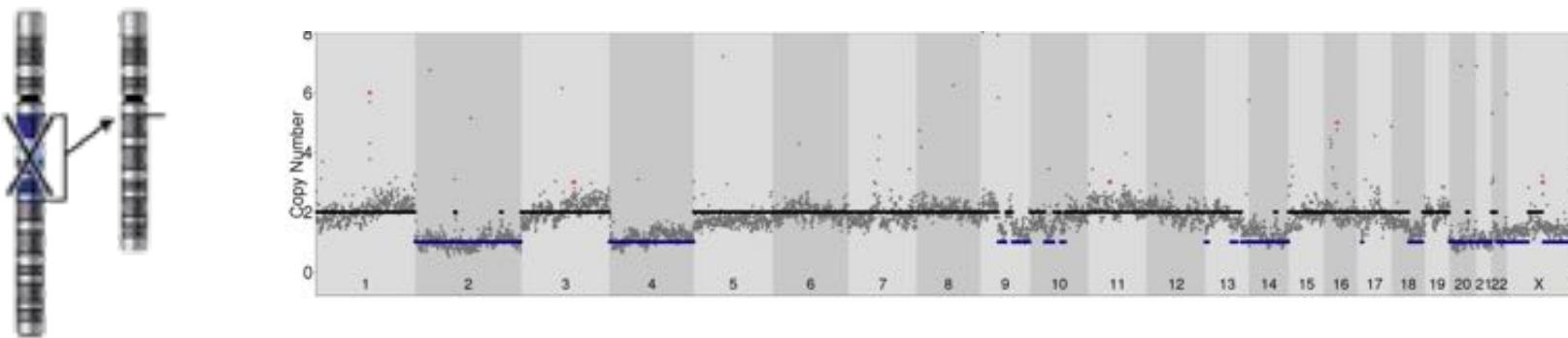
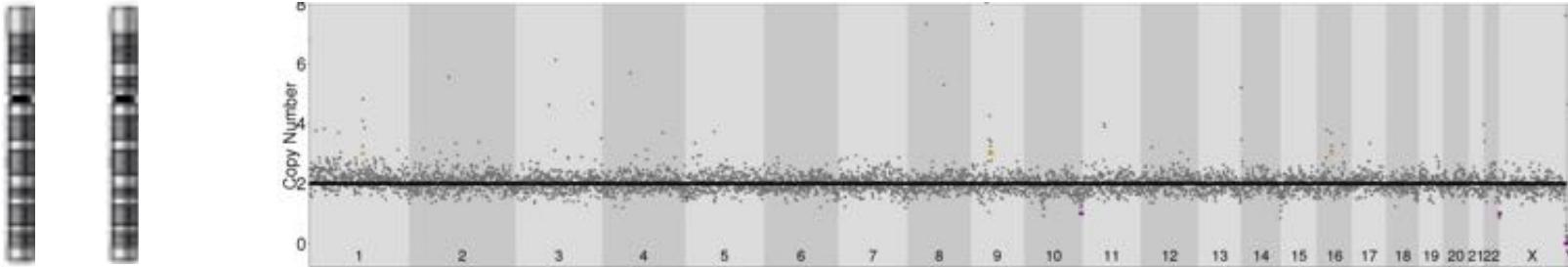
- **B cells – antibody generation**
- **T cells – antigen response**

# What is structural variation?

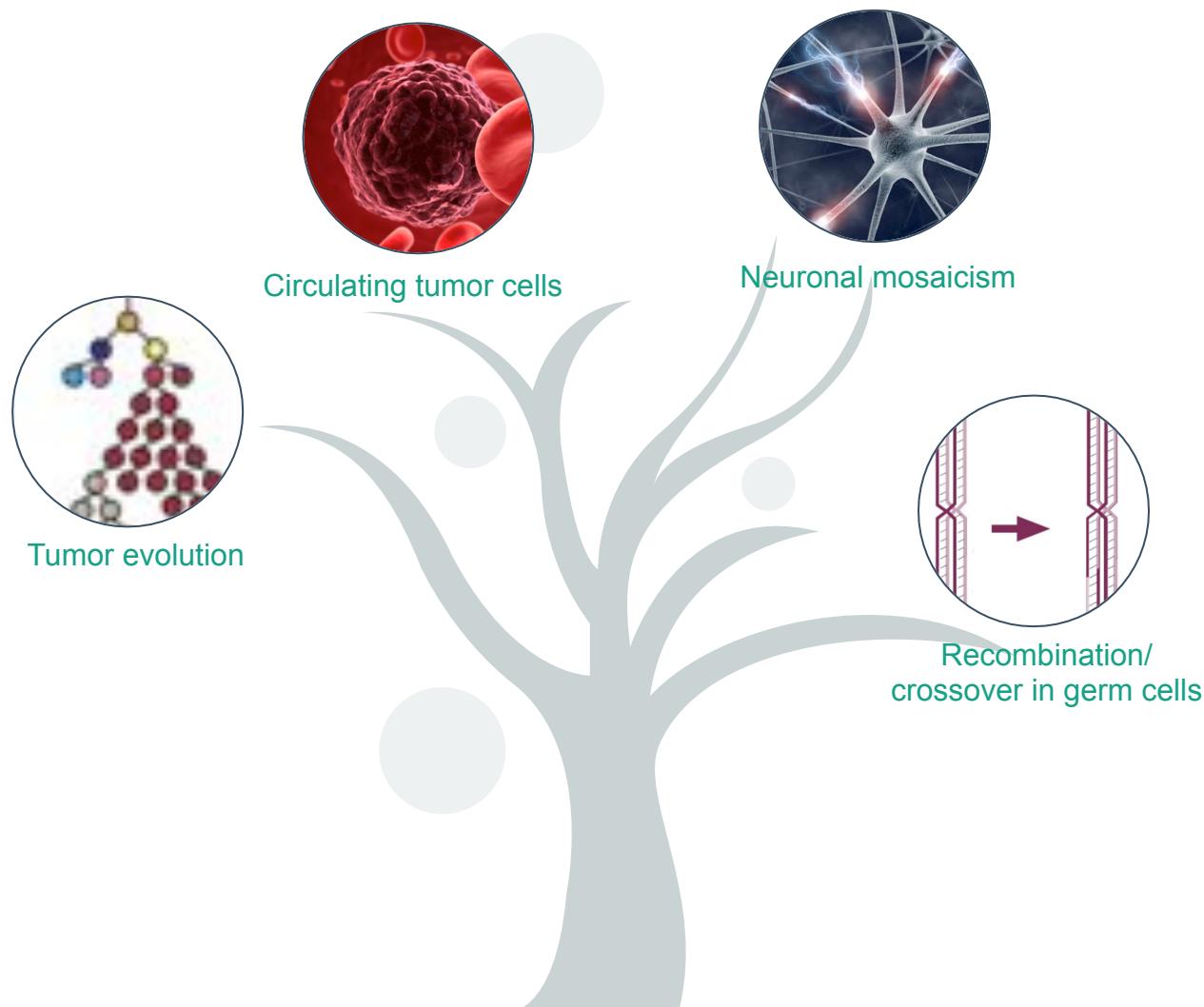
Difference in copy number, orientation, or location of any genomic sequence over 50 bp in size



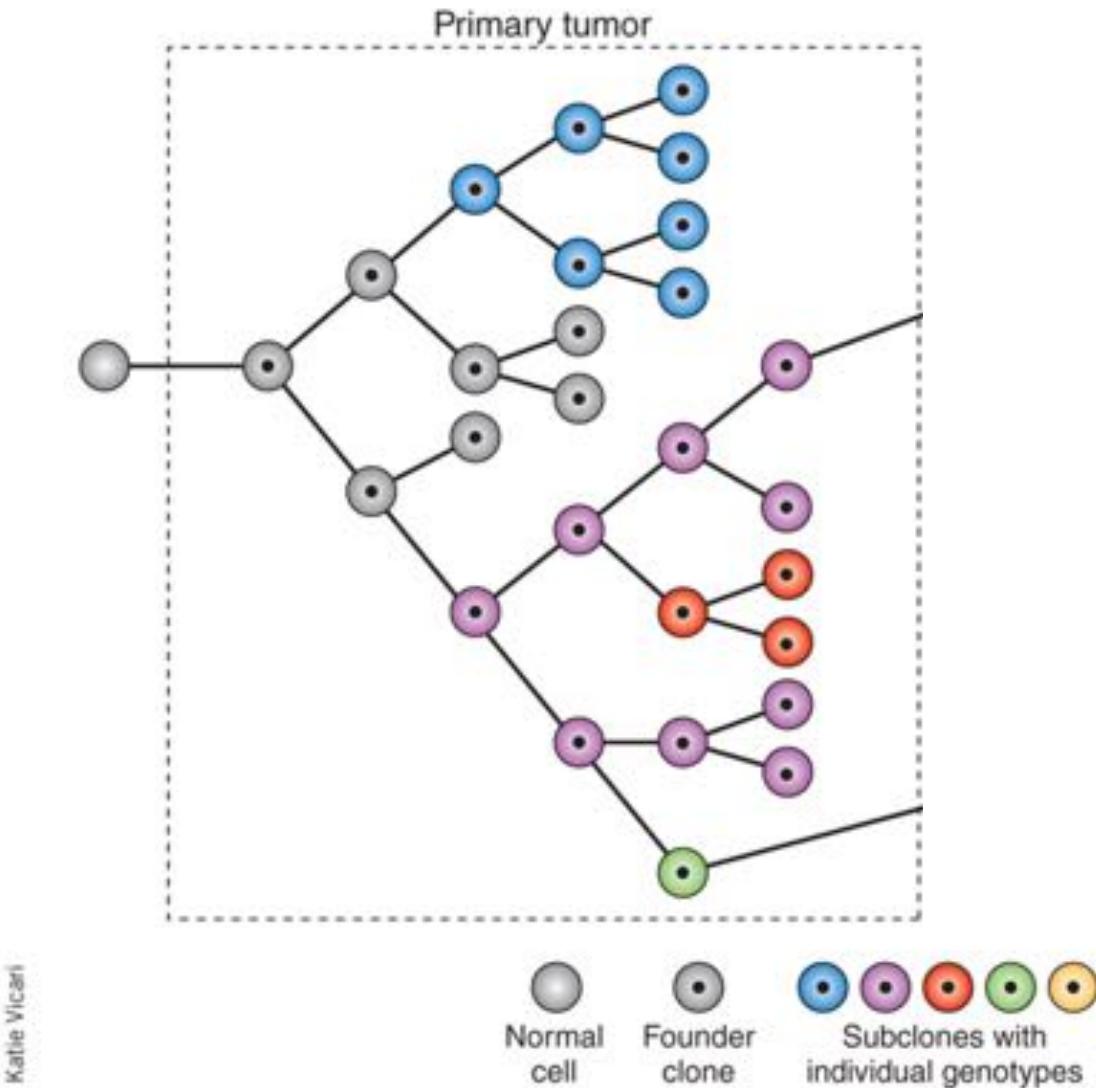
# What are CNVs?



# Single-cell sequencing for CNV analysis



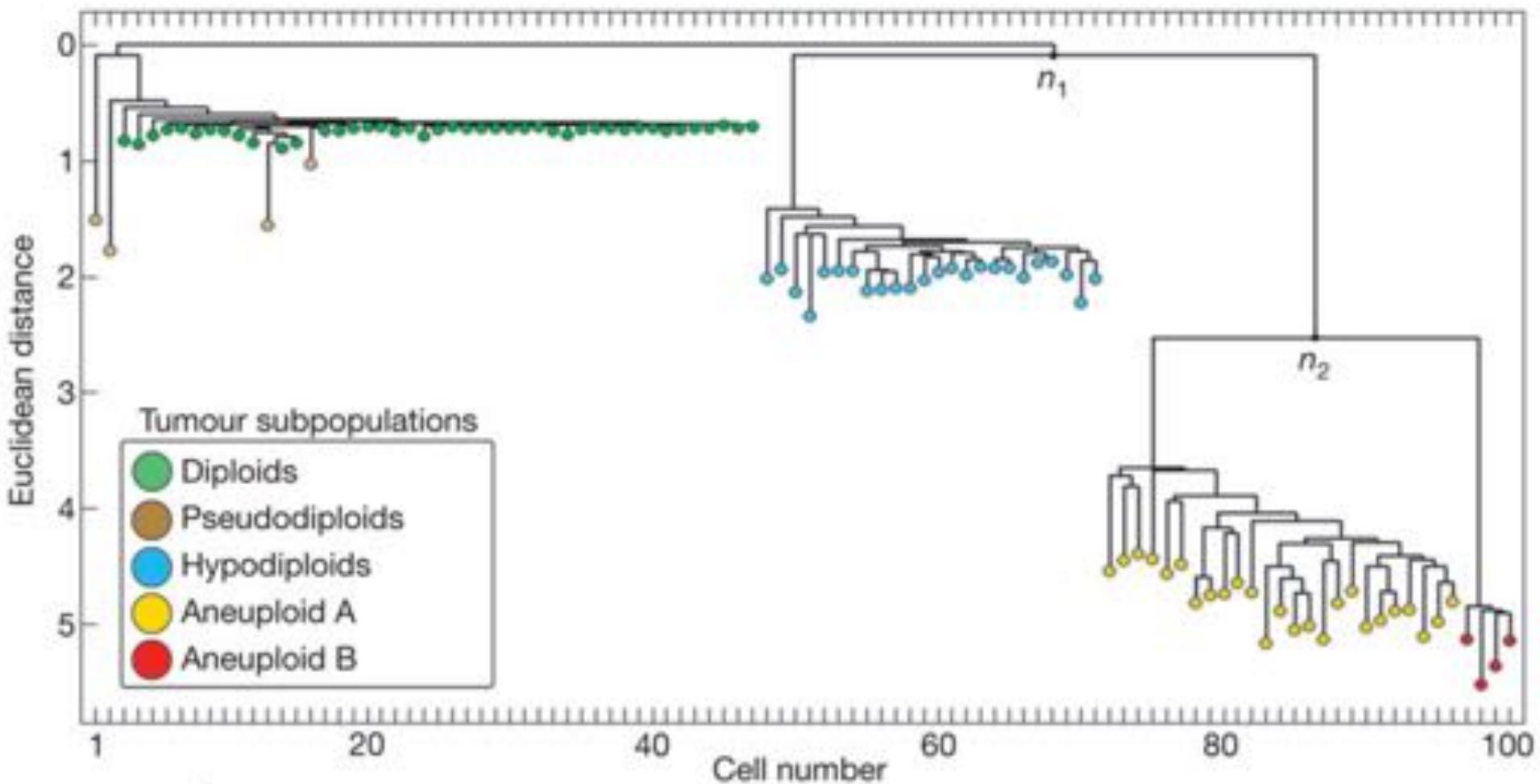
# Clonal evolution in tumors



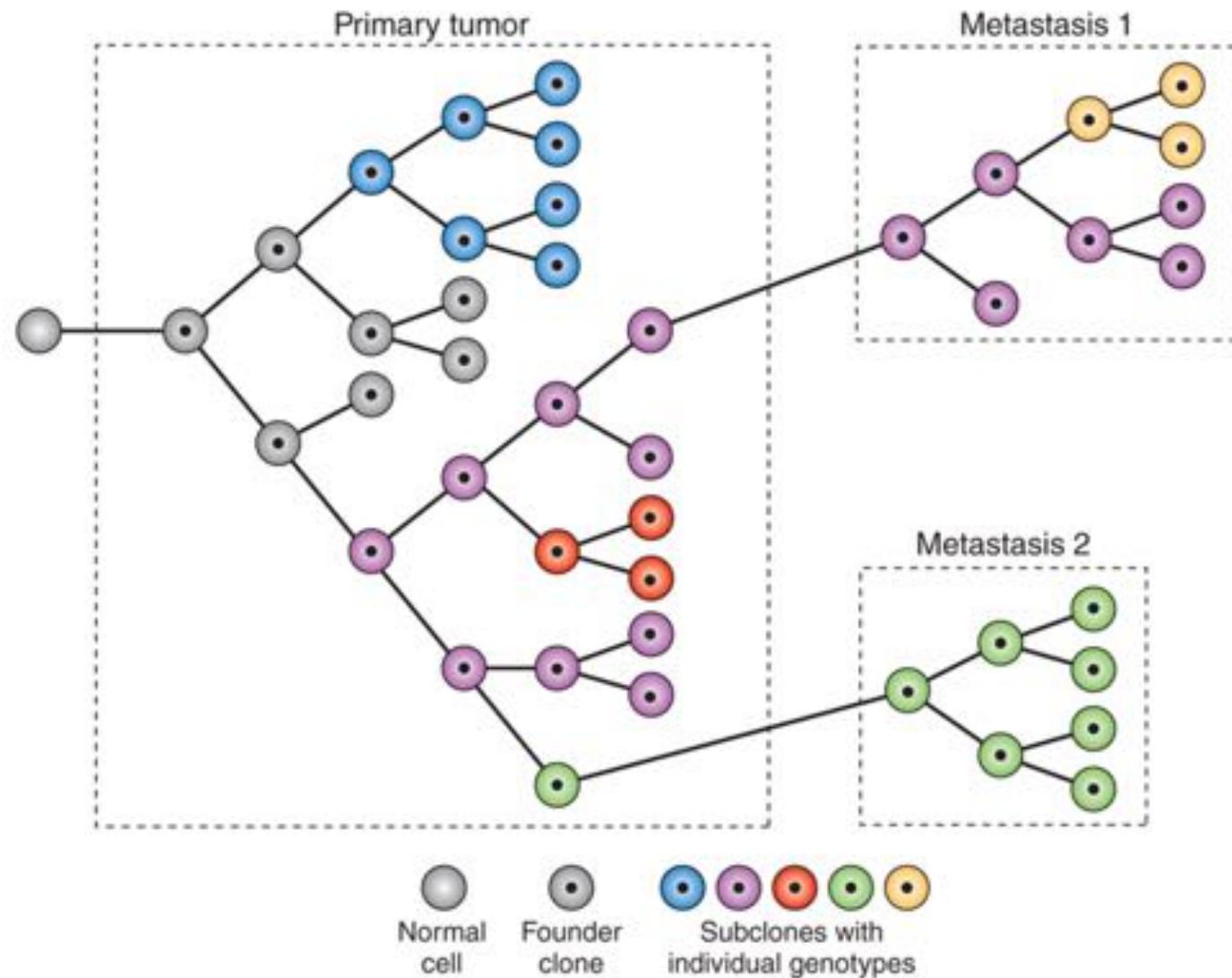
Katie Vicari

Carlos Caldas, Nature Biotech 2012

# Clonal evolution in tumors



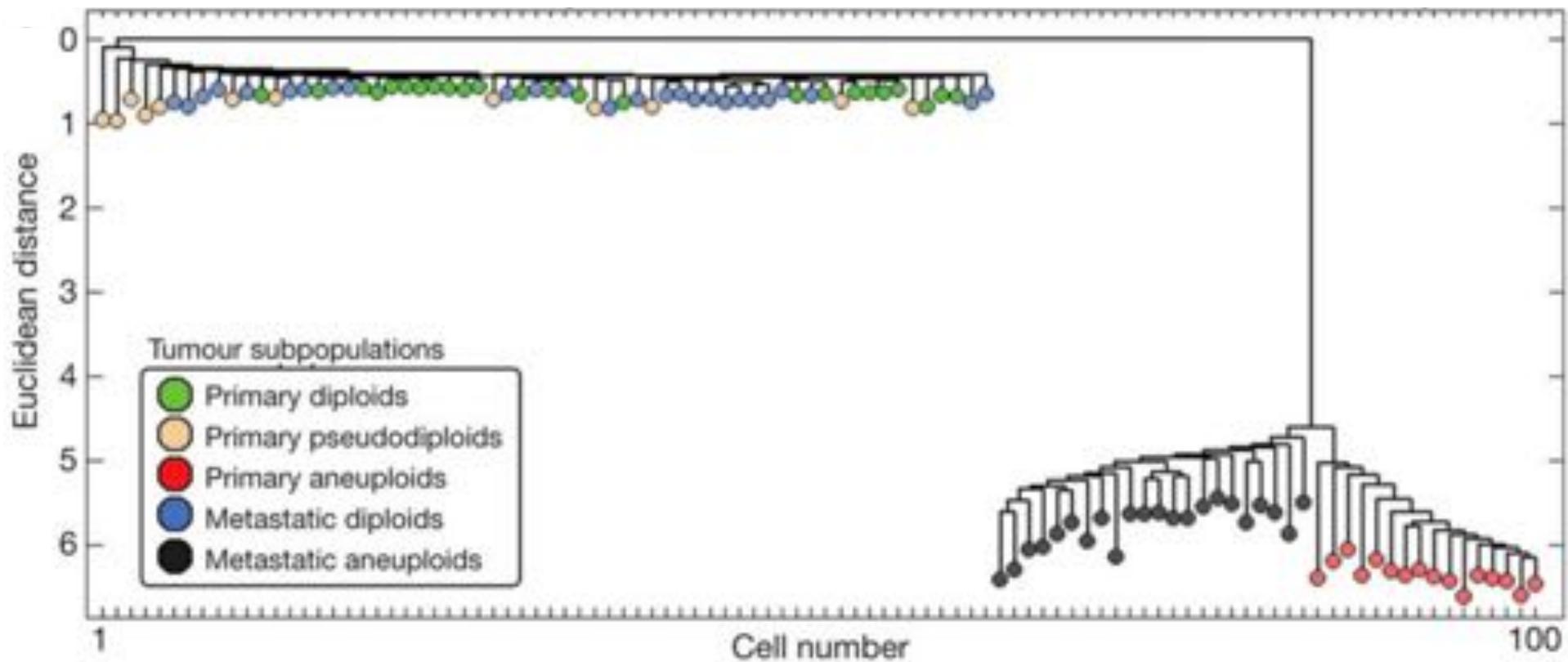
# Cancer metastasis



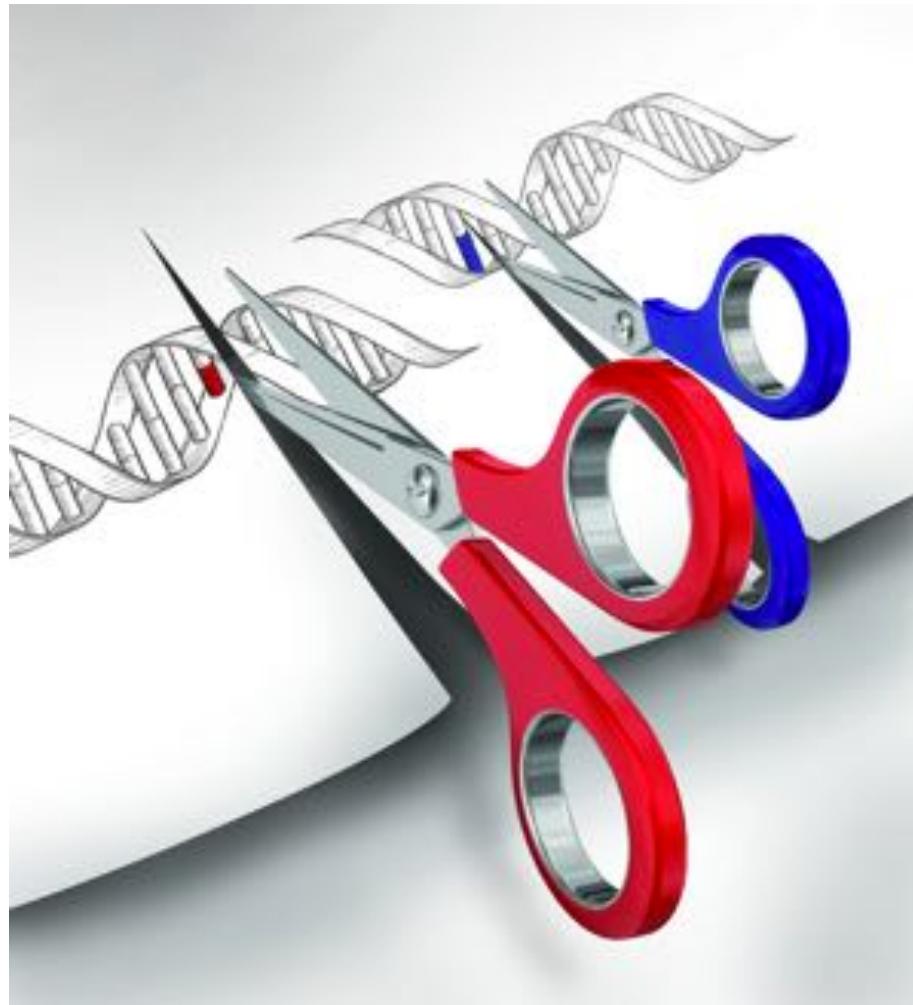
Katie Vicari

Carlos Caldas, Nature Biotech 2012

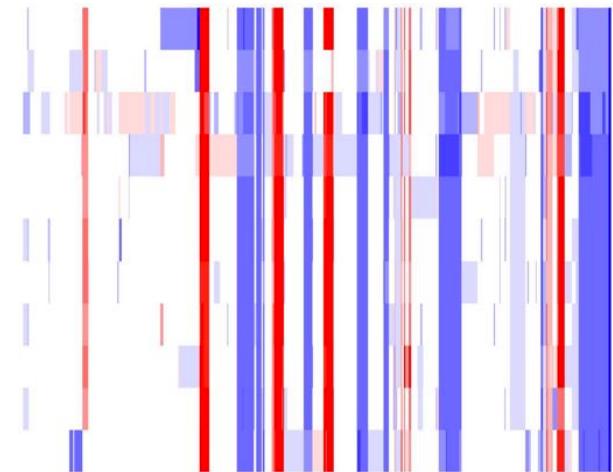
# Cancer metastasis



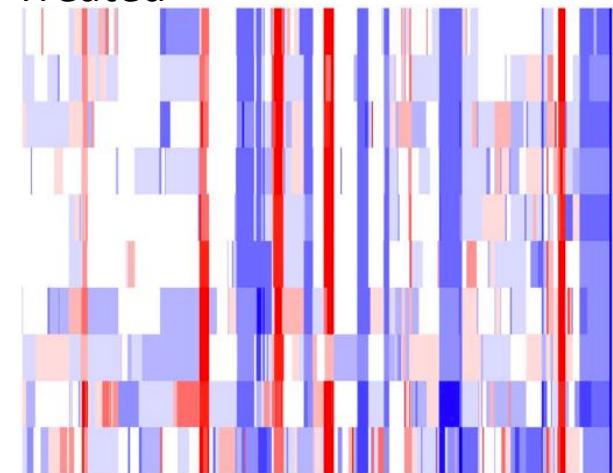
# DNA repair in cancer



Untreated

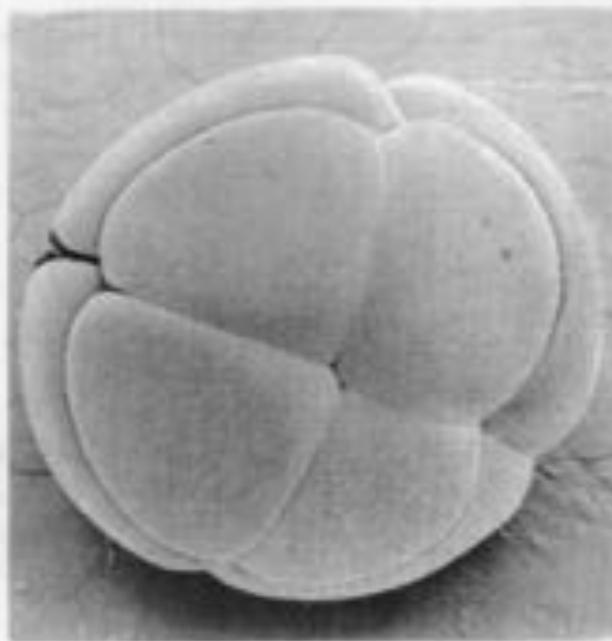


Treated

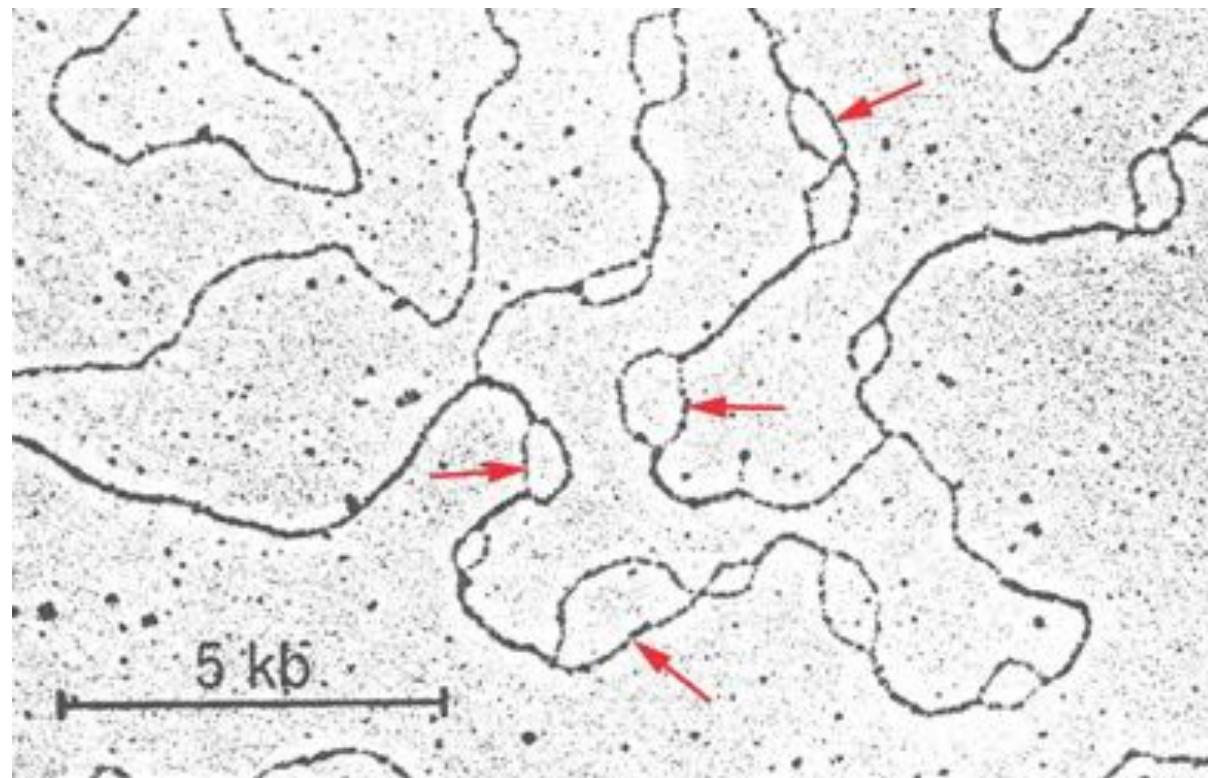


← →  
Genome

## In-vitro fertilization



## DNA replication & cell cycle

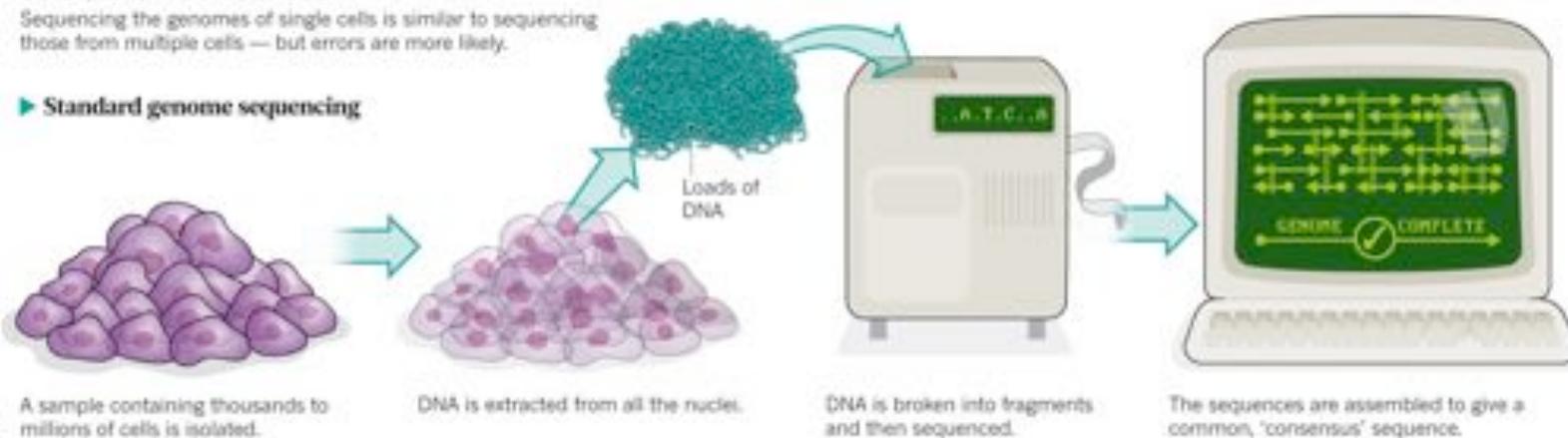


late vs. early replicating regions

# 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.

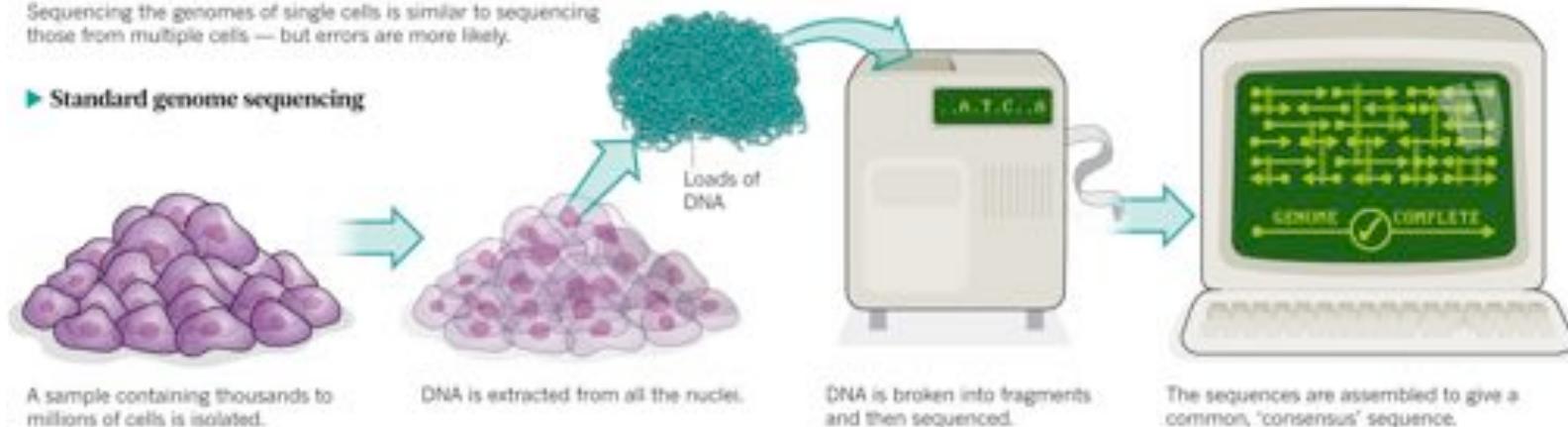


# Single-cell vs. bulk sequencing

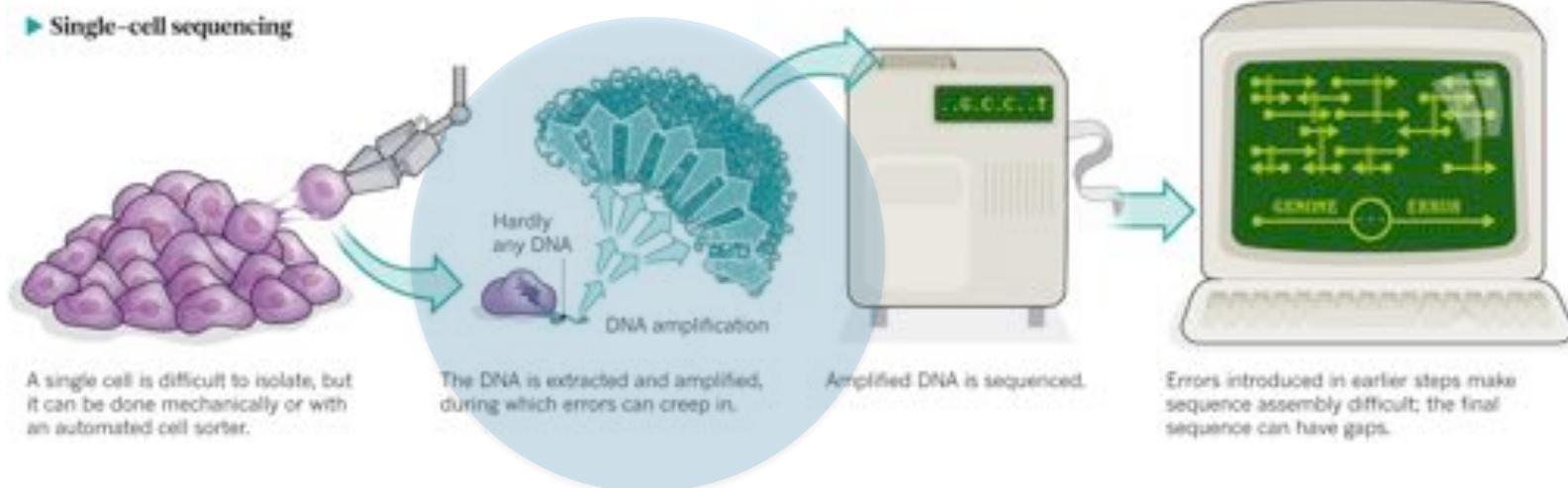
## ONE GENOME FROM MANY

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

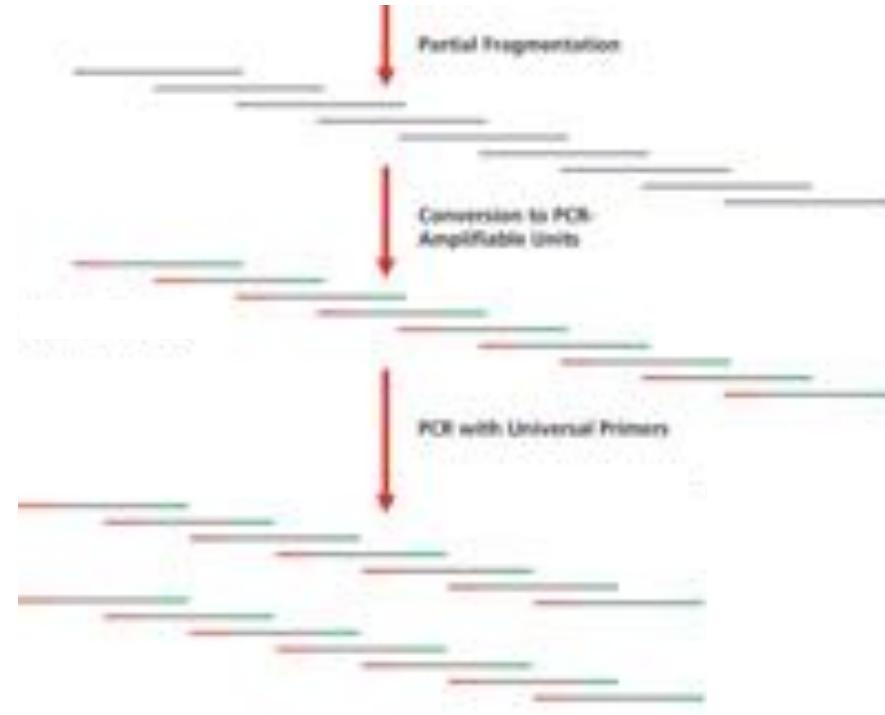
### ► Standard genome sequencing



### ► Single-cell sequencing

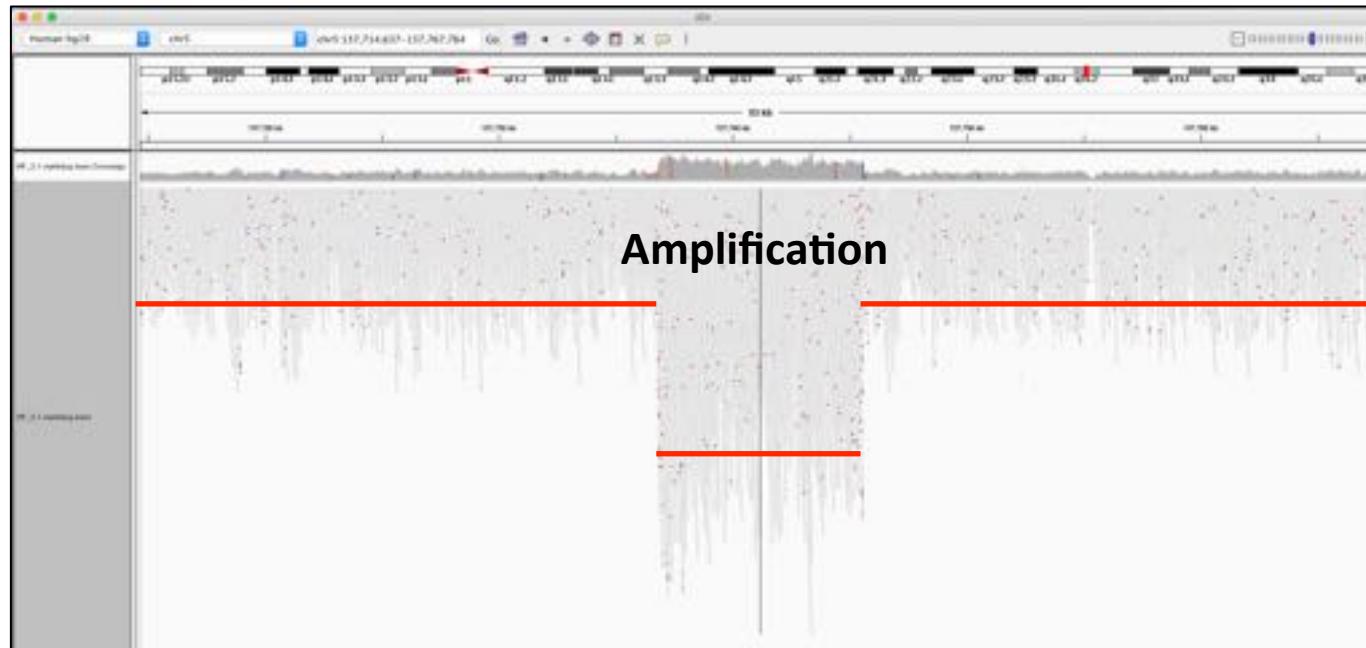
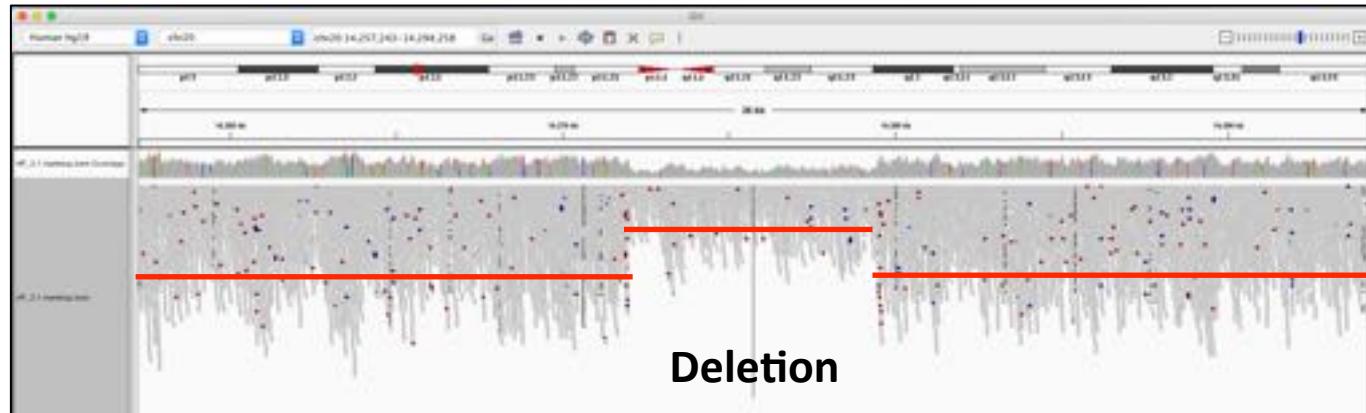


# Whole Genome Amplification

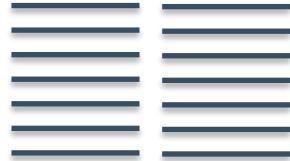


- 1) MDA: Multiple displacement amplification
- 2) DOP-PCR: Degenerate oligonucleotide-primed PCR
- 3) MALBAC: Multiple annealing and looping-based amplification

# Identifying CNVs: bulk vs. single-cell



## Identifying CNVs: bulk vs. single-cell



Low coverage allows us to study copy-number variants

<1X coverage, often <0.1X

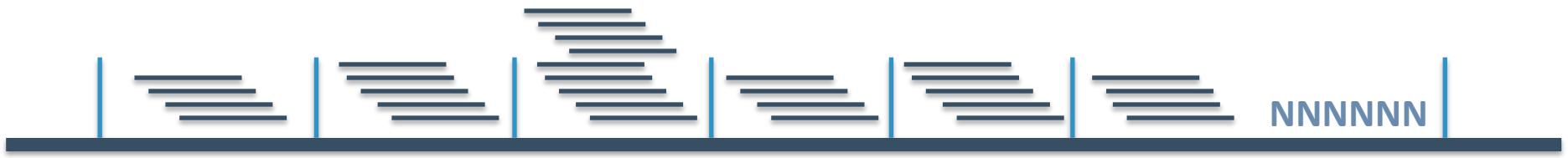
# Copy-number variant analysis



Low coverage allows us to study copy-number variants

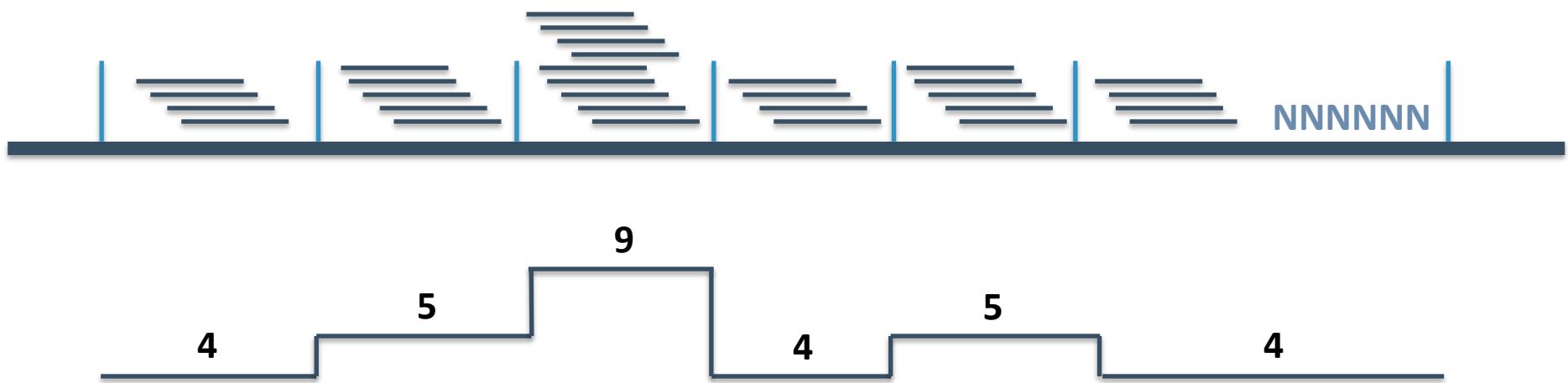
<1X coverage, often <0.1X

# Copy-number variant analysis

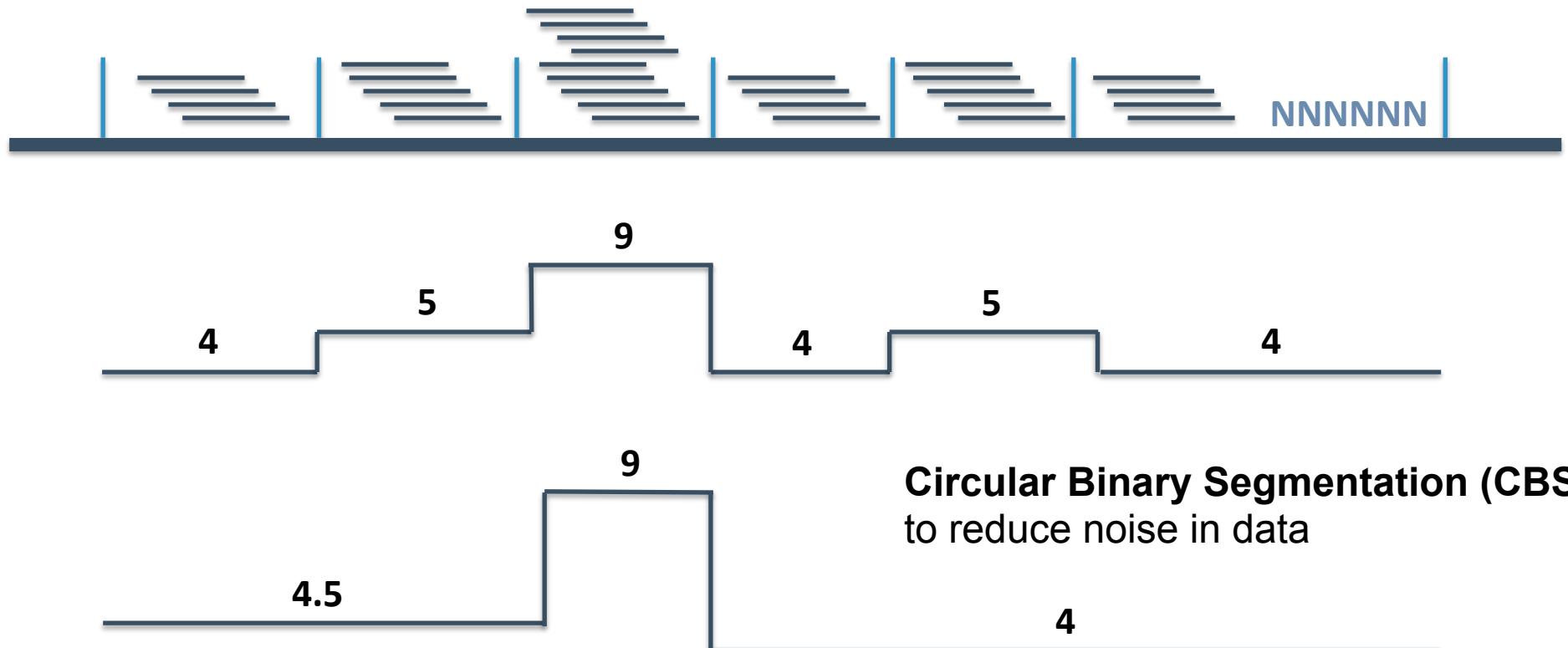


Divide genome into “bins” with ~50 – 100 reads / bin

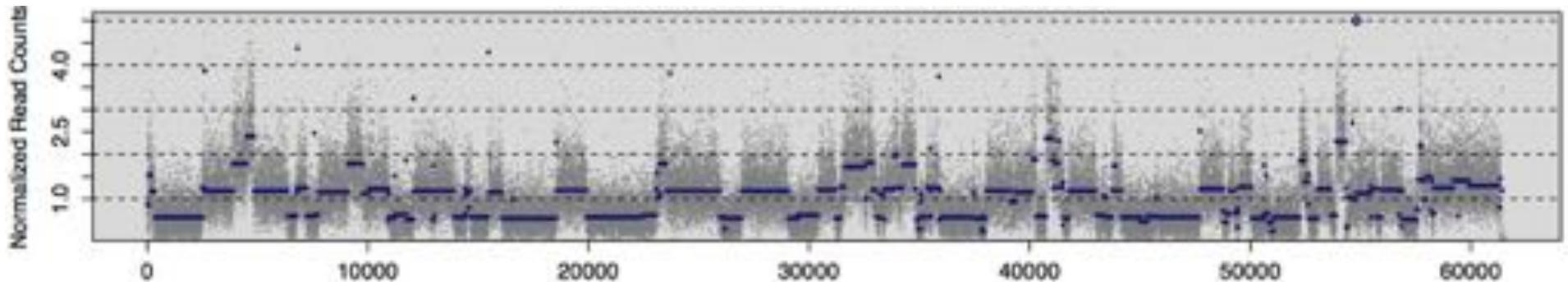
# Copy-number variant analysis



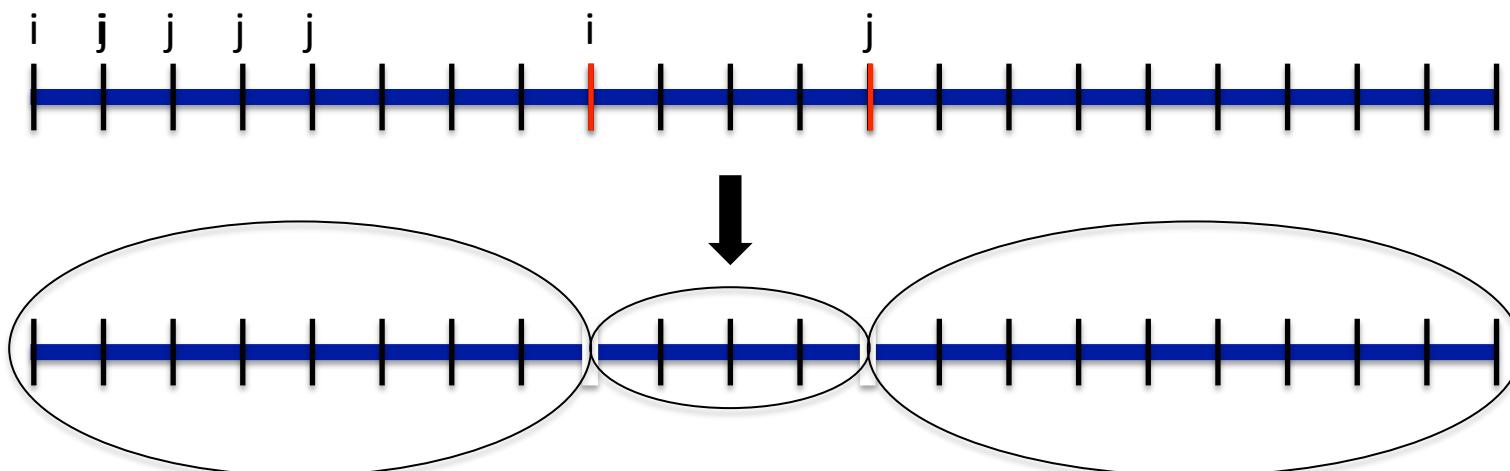
# Copy-number variant analysis



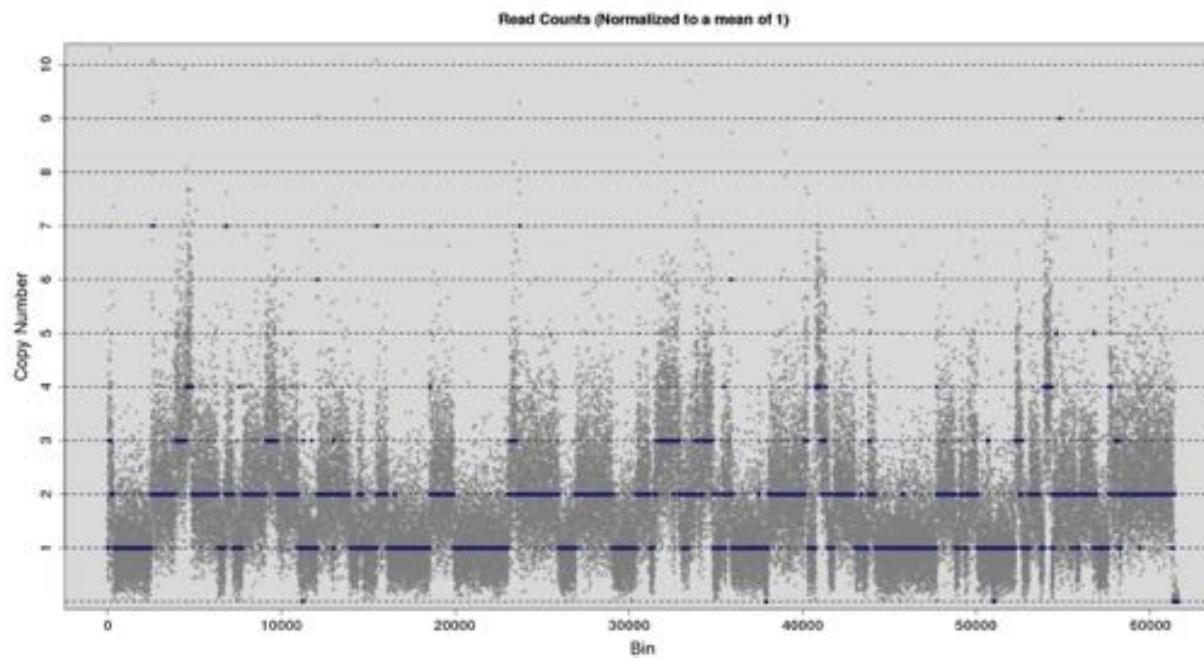
# Segmentation



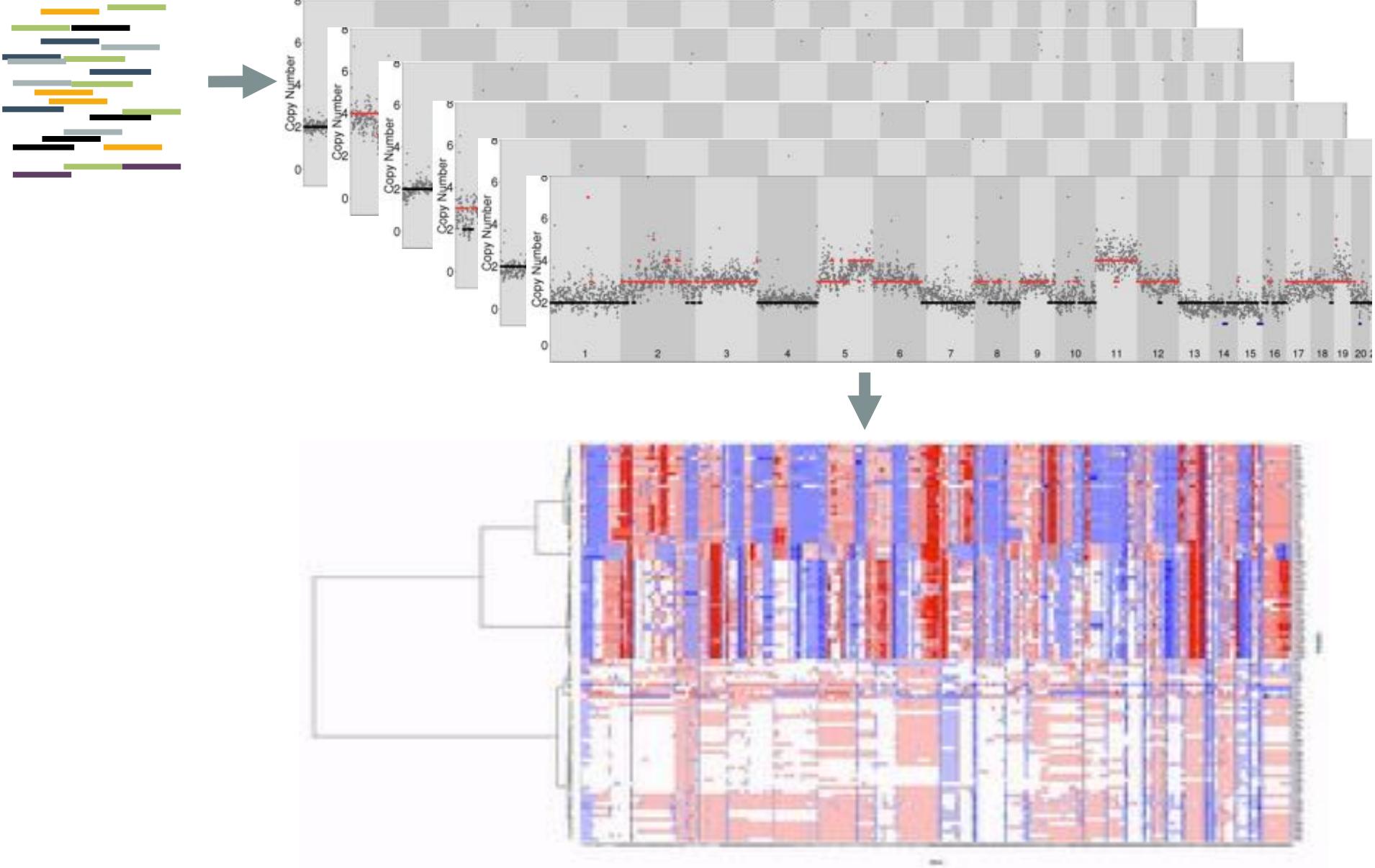
## Circular Binary Segmentation (CBS)



# Copy-number variant analysis



## Copy-number variant analysis

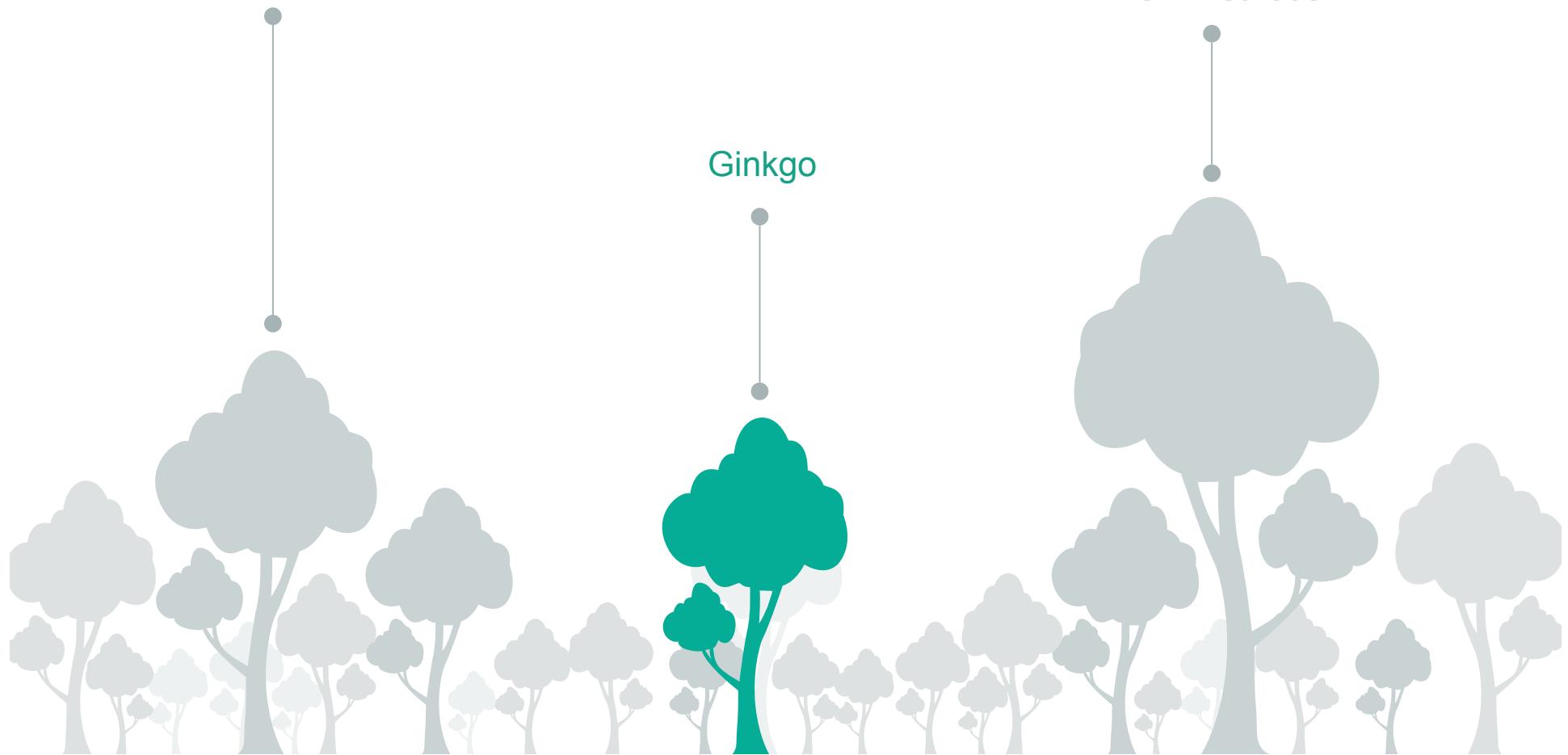


# Outline

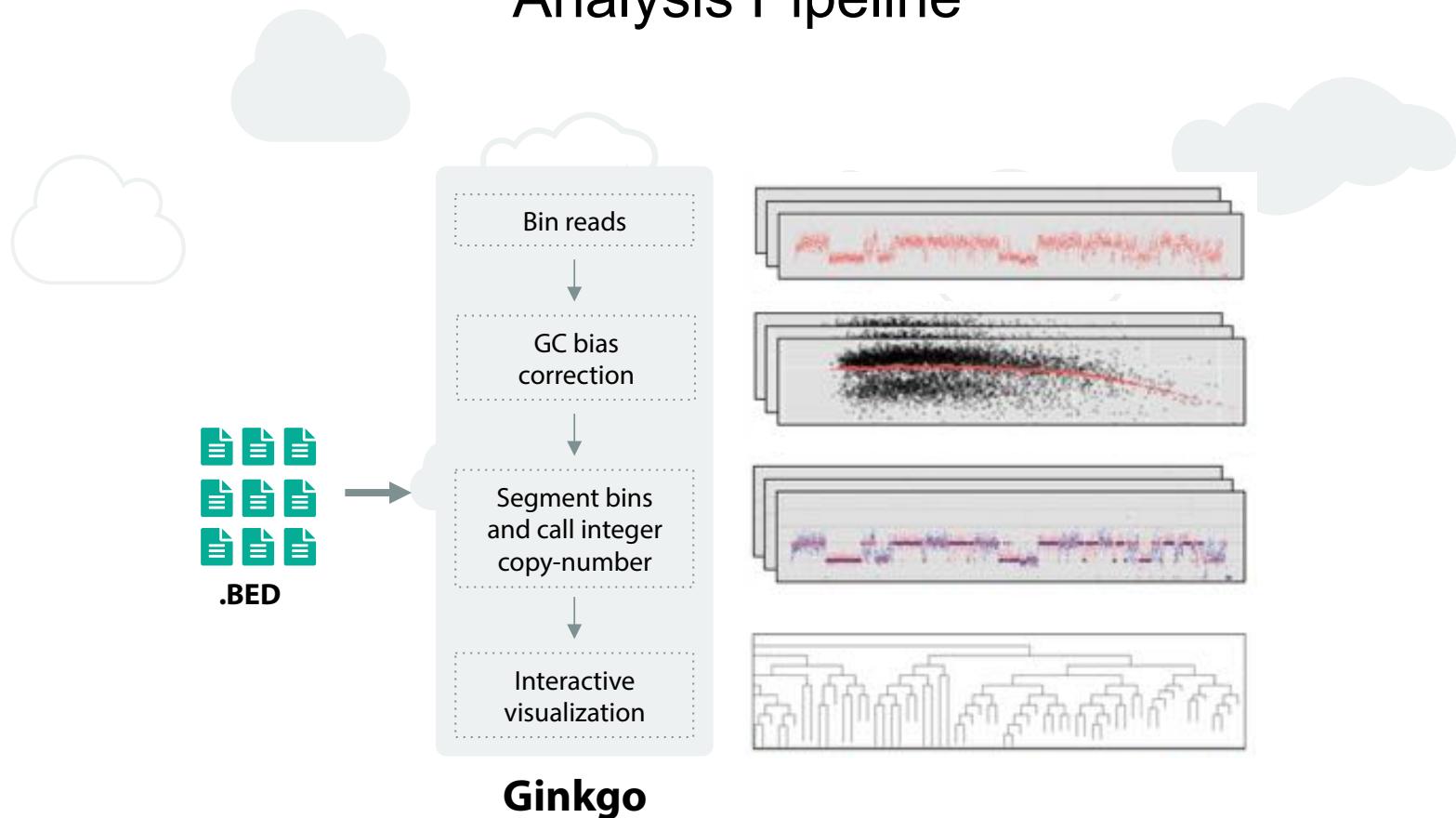
Introduction

Ginkgo

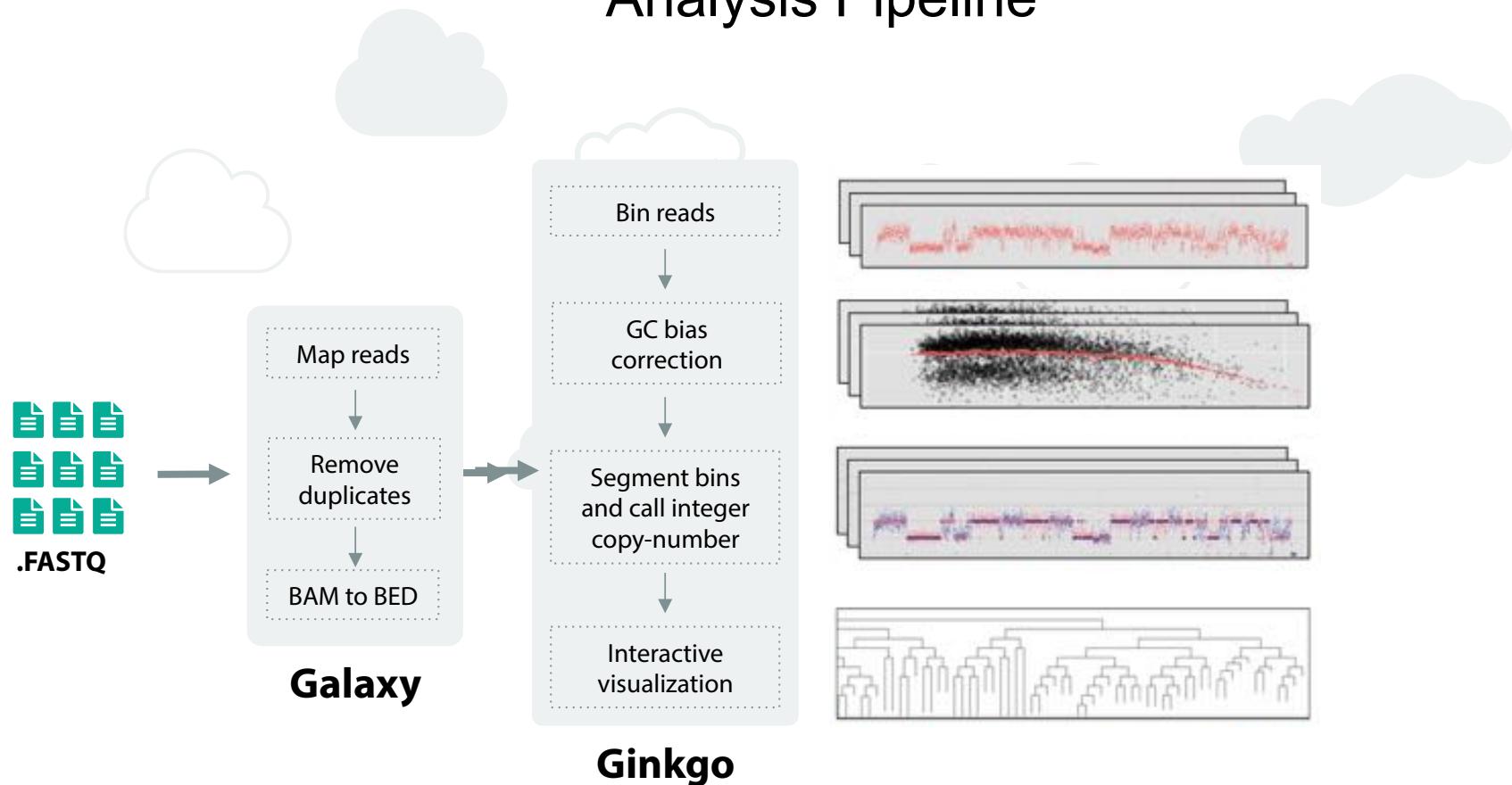
Comparison of  
WGA methods



# Analysis Pipeline



# Analysis Pipeline



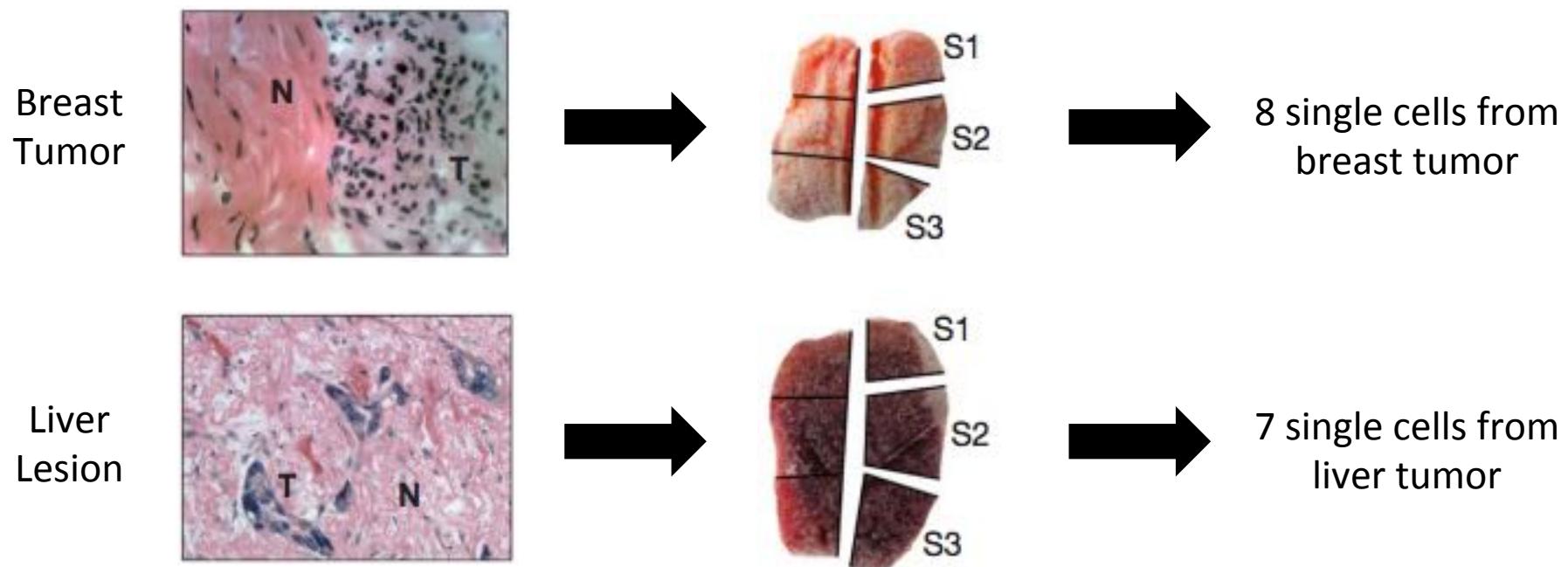
# **Ginkgo Demo**

# LETTER

doi:10.1038/nature09807

## Tumour evolution inferred by single-cell sequencing

Nicholas Navin<sup>1,2</sup>, Jude Kendall<sup>1</sup>, Jennifer Troge<sup>1</sup>, Peter Andrews<sup>1</sup>, Linda Rodgers<sup>1</sup>, Jeanne McIndoo<sup>1</sup>, Kerry Cook<sup>1</sup>, Asya Stepansky<sup>1</sup>, Dan Levy<sup>1</sup>, Diane Esposito<sup>1</sup>, Lakshmi Muthuswamy<sup>2</sup>, Alex Krasnitz<sup>1</sup>, W. Richard McCombie<sup>1</sup>, James Hicks<sup>2</sup> & Michael Wigler<sup>1</sup>



Ginkgo

A web tool for analyzing single-cell sequencing data.

Sample analyses • Load previous analysis •

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

Add file... Cancel upload

706.33 Mbit/s | 00:00:04 | 19.08 % | 91.85 MB / 481.98 MB

M16_SRR089711.bed.gz	56.85 MB	Cancel
M16_SRR0890212.bed.gz	41.11 MB	Cancel
P16_SRR089586.bed.gz	53.07 MB	Cancel
P16_SRR089583.bed.gz	51.54 MB	Cancel
P16_SRR089604.bed.gz	43.70 MB	Cancel
P16_SRR089646.bed.gz	43.32 MB	Cancel

View analysis later

Access your results later at the following address:

<http://gb.cshl.edu/ginkgo/?q=/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

Next step ➤

Ginkgo

A web tool for analyzing single-cell sequencing data.

Sample analyses • Load previous analysis •

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

Add file... Cancel upload

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

M16_SPR088711.bed.gz	56.85 MB	Delete
M16_SPR090212.bed.gz	41.11 MB	Delete
P16_SPR089583.bed.gz	51.54 MB	Cancel
P16_SPR089604.bed.gz	43.70 MB	Delete
P16_SPR089646.bed.gz	43.32 MB	Cancel
P16_SPR089659.bed.gz	40.86 MB	Cancel

View analysis later

Access your results later at the following address:

<http://gb.csHL.edu/ginkgoF/qceoeltp/1kcl0fhepa3r>

Help

Sample .bed file

chrom	chromStart	chromEnd
chr1	555485	555633
chr1	676584	676632
chr1	745136	745184

How to make .bed files

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

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

Detailed Instructions

Next step ➤

Ginkgo

A web tool for analyzing single-cell sequencing data.

Sample analyses • Load previous analysis •

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

Add file... Cancel upload

865.17 Mb/s | 00:00:00 | 91.49 % | 440.96 MB / 481.98 MB

M16_SRR088711.bed.gz	56.85 MB	Delete
M16_SRR0880212.bed.gz	41.11 MB	Delete
P16_SRR088586.bed.gz	53.07 MB	Delete
P16_SRR088593.bed.gz	51.54 MB	Cancel
P16_SRR088604.bed.gz	43.70 MB	Delete
P16_SRR088644.bed.gz	43.33 MB	Delete
P16_SRR088659.bed.gz	40.88 MB	Delete
P16_SRR088662.bed.gz	49.09 MB	Cancel
Next step >		

View analysis later

Access your results later at the following address:

<http://gb.csHL.edu/ginkgoF/qceoeltp/1kcllhepa3r>

Help

Sample .bed file

chrom	chromStart	chromEnd
chr1	555485	555633
chr1	676584	676632
chr1	745136	745184

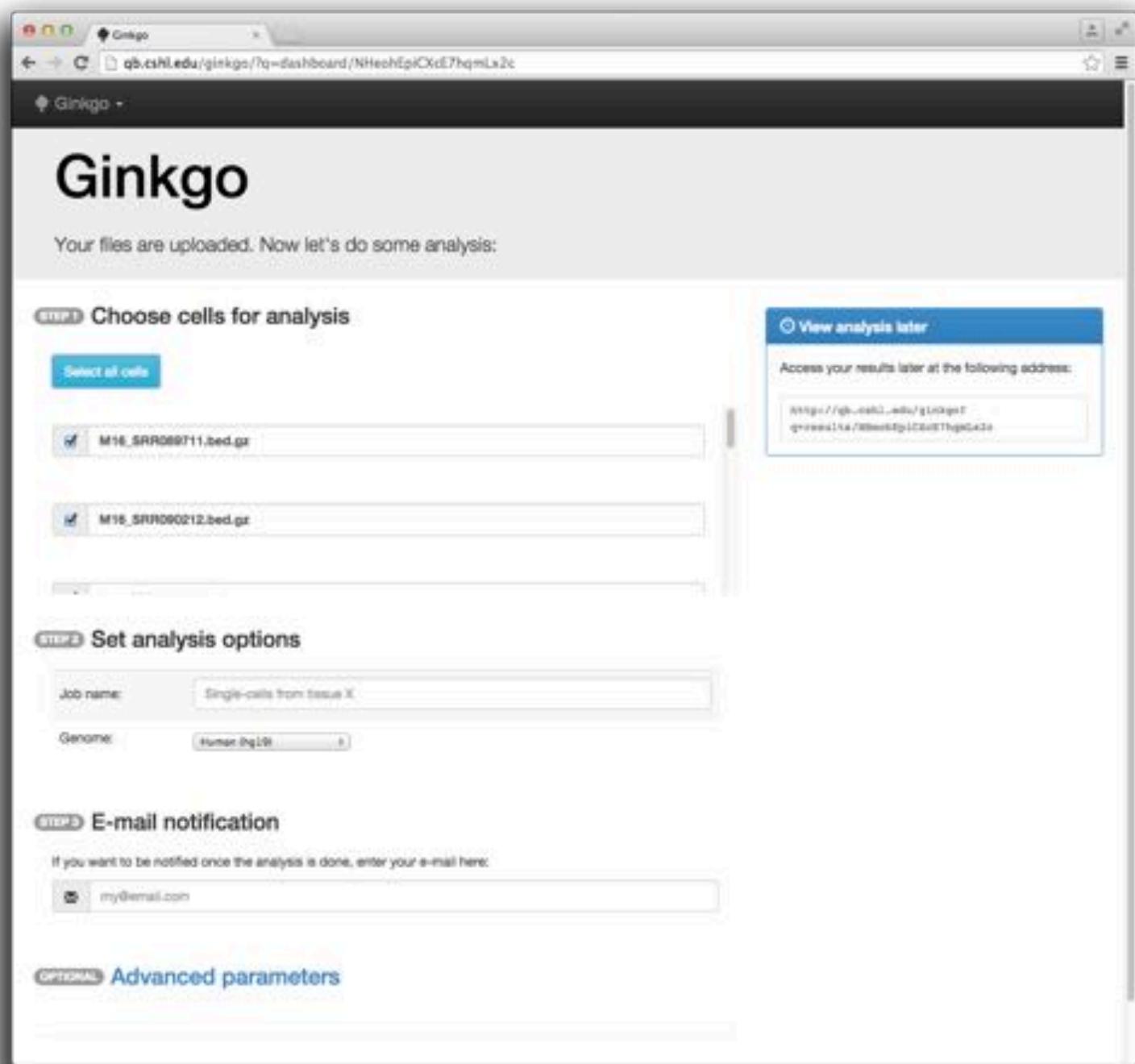
How to make .bed files

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

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

Detailed Instructions

The screenshot shows the Ginkgo web application interface. On the left, a list of .bed files is displayed with their sizes and delete/cancel buttons. A progress bar indicates the upload status for each file. On the right, there's a sidebar with links to view analysis later and help, which includes sample .bed file data and instructions on how to make .bed files from a bam file.



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

**STEP 1 Choose cells for analysis**

M16\_SRR09011.bed.gz  
 M16\_SRR09012.bed.gz

**STEP 2 Set analysis options**

Job name: Single-cells from tissue X  
Genome: human (hg19)

**STEP 3 E-mail notification**

If you want to be notified once the analysis is done, enter your e-mail here:  
 my@email.com

**OPTIONAL Advanced parameters**

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

**STEP 1 Choose cells for analysis**

M16\_SRR09011.bed.gz  
 M16\_SRR09012.bed.gz

**STEP 2 Set analysis options**

Job name: Single-cells from tissue X  
Genome: human hg19

**STEP 3 E-mail notification**

If you want to be notified once the analysis is done, enter your e-mail here:  
 my@email.com

**OPTIONAL Advanced parameters**

**View analysis later**

Access your results later at the following address:  
<http://qb.csail.mit.edu/ginkgo?qrwww1ta/8Bmklip1Qk8fThqgla2r>

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

**STEP 1 Choose cells for analysis**

M16\_SRR09011.bed.gz  
 M16\_SRR09012.bed.gz

**STEP 2 Set analysis options**

Job name: Single-cell analysis  
Genome: human (hg19)

**STEP 3 E-mail notification**

If you want to be notified once the analysis is done, enter your e-mail here:  
 my@email.com

**OPTIONAL Advanced parameters**

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

**STEP 1 Choose cells for analysis**

M16\_SRR09011.bed.gz  
 M16\_SRR09012.bed.gz

**STEP 2 Set analysis options**

Job name: Single-cell analysis  
Genome: Human (hg19)

**STEP 3 E-mail notification**

If you want to be notified once the analysis is done, enter your e-mail here:  
 my@email.com

**OPTIONAL Advanced parameters**

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

**STEP 1 Choose cells for analysis**

M16\_SRR09011.bed.gz  
 M16\_SRR09012.bed.gz

**STEP 2 Set analysis options**

Job name: Single-cell analysis  
Genome: human (hg19)

**E-mail notification**  
If you want to be notified once the analysis is done, enter your e-mail here:  
 my@email.com

**OPTIONAL Advanced parameters**

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHcpICXcE7hqmlx2c

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

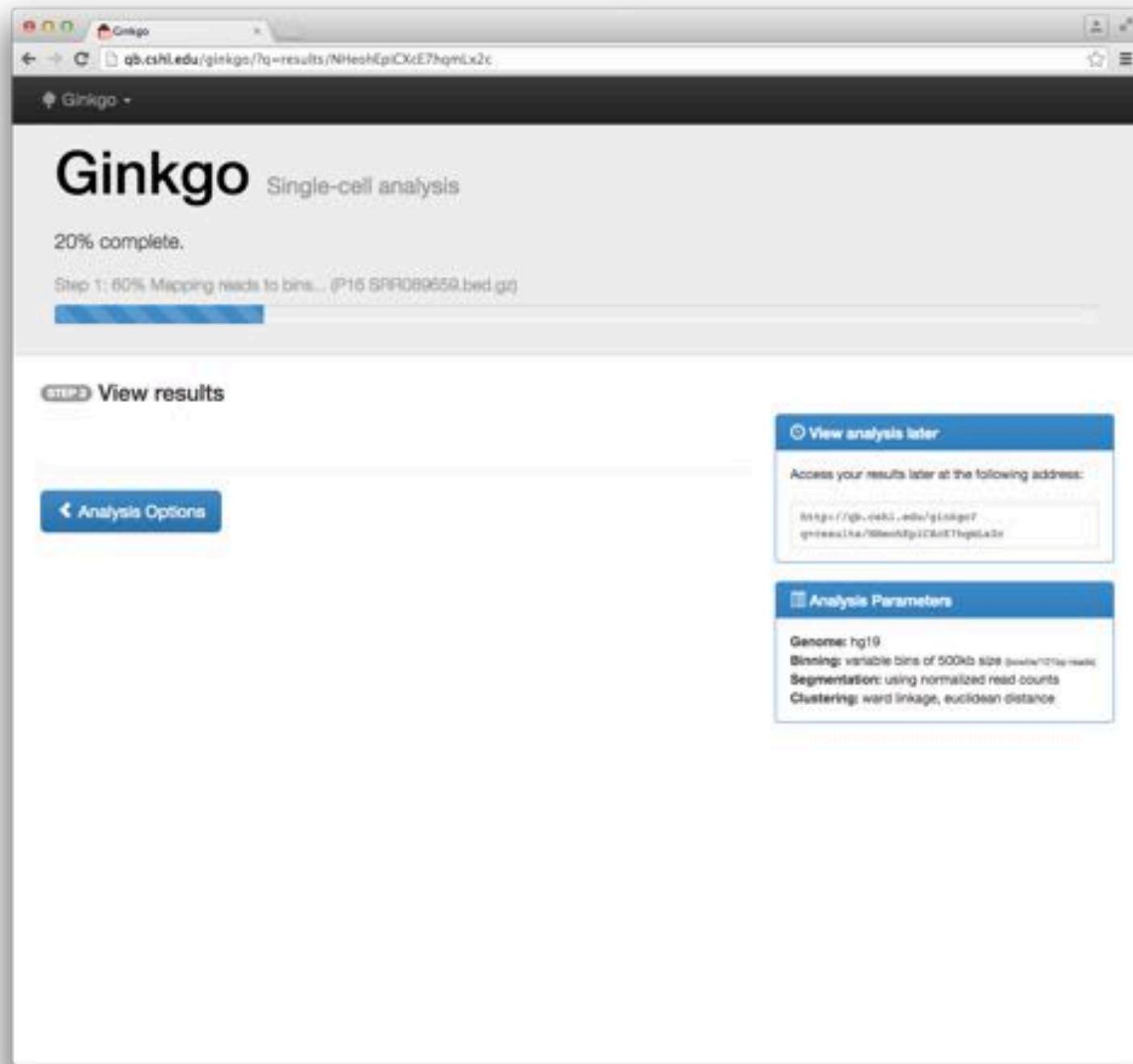
Access your results later at the following address:

<http://ginkgo.csdl.uchicago.edu/ginkgo/?q=results/NHeoHcpICXcE7hqmlx2c>

Analysis Parameters

Genome: hg19  
Binning: variable bins of 500kb size (parallel 10% reads)  
Segmentation: using normalized read counts  
Clustering: ward linkage, euclidean distance

The screenshot shows a web browser window for the Ginkgo tool. The title bar says 'Ginkgo'. The address bar shows the URL 'qb.cshl.edu/ginkgo/?q=results/NHeoHcpICXcE7hqmlx2c'. The main content area has a large 'Ginkgo' logo and the text 'Single-cell analysis'. Below that, it says '3% complete.' and 'Step 1: 10% Mapping reads to bins... (M16\_SRR090212.bed.gz)'. There is a progress bar with a blue segment. On the left, there's a 'View results' button and an 'Analysis Options' button. On the right, there's a 'View analysis later' button (which is highlighted in blue), and a box containing the URL 'http://ginkgo.csdl.uchicago.edu/ginkgo/?q=results/NHeoHcpICXcE7hqmlx2c'. Below that is another box titled 'Analysis Parameters' with details about the genome (hg19), binning (variable bins of 500kb size), segmentation (using normalized read counts), and clustering (ward linkage, euclidean distance).



Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

95% complete.

Step 2: 65% Calling copy number events... (Computing Cluster (Copy Number))



 View results

 Analysis Options

 View analysis later

Access your results later at the following address:

<http://ginkgo.csdl.kit.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

 Analysis Parameters

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

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090156  
M16\_SRR090158  
M16\_SRR090112  
M16\_SRR090144  
M16\_SRR090156

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

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

**Tree display**

Normalized read counts [newick | xml | pdf | png]

Copy-number [newick | xml | pdf | png]

Correlations [newick | xml | pdf | png]

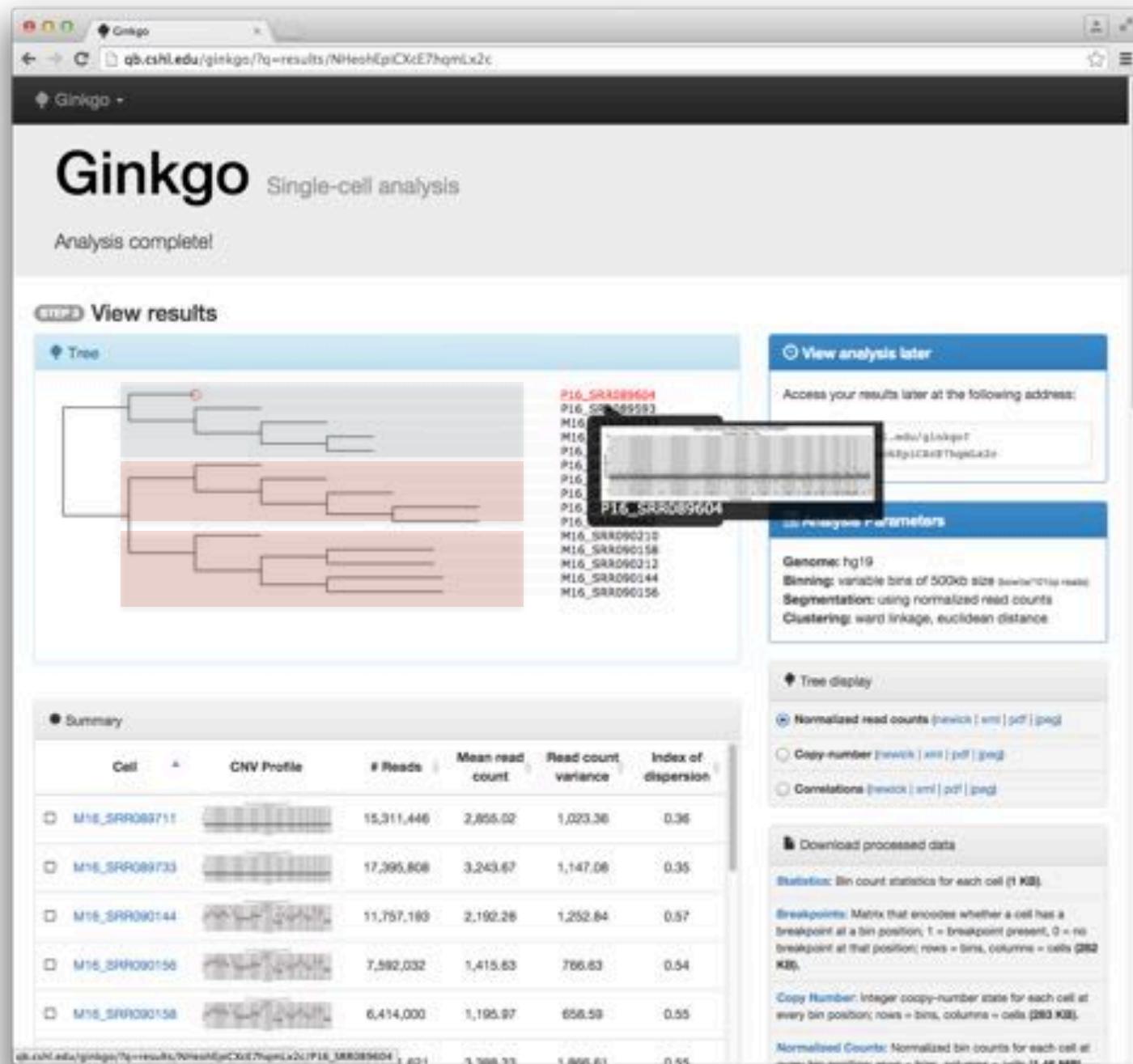
**Download processed data**

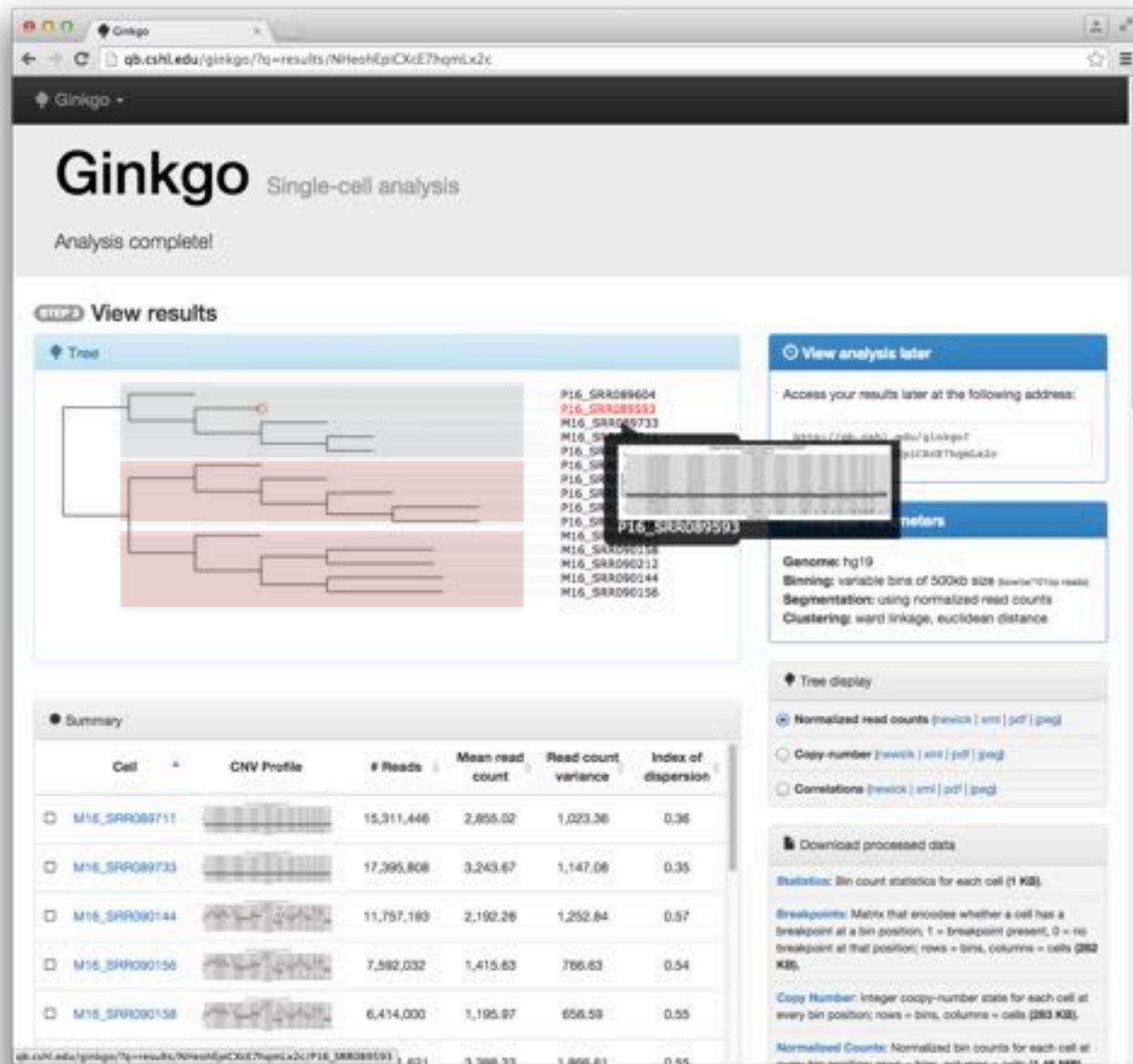
**Breakpoints:** Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB)

**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 (28 KB)

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile image]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile image]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[CNV profile image]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[CNV profile image]	7,582,032	1,415.63	766.63	0.54
M16_SRR090158	[CNV profile image]	6,414,000	1,195.97	656.59	0.55
M16_SRR090010	[CNV profile image]	18,171,621	3,308.33	1,860.61	0.55





Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

Tree

View analysis later

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Parameters**

Genome: hg19  
 Binning: variable bins of 500kb size (over 100k 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**

**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 (283 KB)

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHpiCXcE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

M16\_SRR089711

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

**View analysis later**

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=results/NHeoHpiCXcE7hqmlx2c>

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711		15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.06	0.35
M16_SRR089144		11,757,180	2,192.26	1,252.84	0.57
M16_SRR089156		7,582,032	1,415.63	766.63	0.54
M16_SRR089158		6,414,000	1,195.97	656.59	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**

**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 (28 KB).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Parameters**

Segmenting: variable bins of 500kb size (over 100k 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
M16_SRR089711		15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.06	0.35
M16_SRR089144		11,757,180	2,192.26	1,252.84	0.57
M16_SRR089156		7,582,032	1,415.63	766.63	0.54
M16_SRR089158		6,414,000	1,195.97	656.59	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**

**Breakpoints:** 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 (28 KB).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090218  
M16\_SRR090158  
M16\_SRR090112  
M16\_SRR090144  
M16\_SRR090156

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c>

**Analysis Parameters**

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

**Tree display**

Normalized read counts [newick | xml | pdf | png]

Copy-number [newick | xml | pdf | png]

Correlations [newick | xml | pdf | png]

**Download processed data**

**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 (28 KB)

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile image]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile image]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[CNV profile image]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[CNV profile image]	7,582,032	1,415.63	766.63	0.54
M16_SRR090158	[CNV profile image]	6,414,000	1,195.97	656.59	0.55
M16_SRR090210	[CNV profile image]	18,171,621	3,308.33	1,860.61	0.55

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
**P16\_SRR089546**  
P16\_SRR089664  
P16\_SRR089546  
P16\_SRR089546  
M16\_SRR089546  
M16\_SRR089546  
M16\_SRR089546  
M16\_SRR089546  
M16\_SRR089546

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**parameters**

bin of 500kb size (overlapped reads)  
Segmentation: using normalized read counts  
Clustering: ward linkage, euclidean distance

**Tree display**

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

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

Correlations (newick | xml | pdf | png)

**Download processed data**

Breakpoints: 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, column = cells (283 KB).

Normalized Counts: Normalized bin counts for each cell at every bin position; rows = bins, column = cells (283 KB).

http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c/P16\_SRR089546\_1.pdf 1 / 21

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089664  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089664  
P16\_SRR089663  
P16\_SRR089658  
P16\_SRR089657  
M16\_SRR089656  
M16\_SRR089655  
M16\_SRR089654  
M16\_SRR089653  
M16\_SRR089652

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Refers**

Normalized read counts (500kb size binning) (100k reads)  
Normalized read counts (Clustering: ward linkage, euclidean distance)

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile bar]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile bar]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR089744	[CNV profile bar]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR089756	[CNV profile bar]	7,582,032	1,415.63	766.83	0.54
M16_SRR089758	[CNV profile bar]	6,414,000	1,195.97	656.59	0.55

**Tree display**

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

**Download processed data**

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 (28 KB).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089663  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089664  
**P16\_SRR089663**  
P16\_SRR089659  
P16\_SRR089653  
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/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

Cells of 500kb size (bowtie2) (10 reads)  
using normalized read counts  
distance measure, euclidean distance

**P16\_SRR089663**

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711		15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.06	0.35
M16_SRR089744		11,757,180	2,192.26	1,252.84	0.57
M16_SRR089756		7,582,032	1,415.63	766.83	0.54
M16_SRR089758		6,414,000	1,195.97	656.59	0.55

**Tree display**

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

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

Correlations (newick | xml | pdf | png)

**Download processed data**

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 (28 KB).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089662

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

bin size: 500kb (size binning) (100 bins)  
using normalized read counts  
linkage, euclidean distance

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile bar]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile bar]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR089144	[CNV profile bar]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR089156	[CNV profile bar]	7,582,032	1,415.63	766.63	0.54
M16_SRR089158	[CNV profile bar]	6,414,000	1,195.97	656.59	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**

Breakpoints: 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 (28 KB).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090218  
M16\_SRR090158  
M16\_SRR090112  
M16\_SRR090144  
M16\_SRR090156

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c>

**Analysis Parameters**

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

**Tree display**

Normalized read counts [newick | xml | pdf | png]

Copy-number [newick | xml | pdf | png]

Correlations [newick | xml | pdf | png]

**Download processed data**

**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 (28 KB)

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[redacted]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[redacted]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[redacted]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[redacted]	7,582,032	1,415.63	766.83	0.54
M16_SRR090158	[redacted]	6,414,000	1,195.97	656.59	0.55
M16_SRR090210	[redacted]	18,171,621	3,308.33	1,860.61	0.55

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

M16\_SRR090210

M16\_SRR090210  
M16\_SRR090158  
M16\_SRR090156  
M16\_SRR090144  
M16\_SRR090133  
M16\_SRR090111  
P16\_SRR090593  
P16\_SRR090586  
P16\_SRR090585  
P16\_SRR090584  
P16\_SRR090583  
P16\_SRR090582  
P16\_SRR090581  
P16\_SRR090580  
P16\_SRR090579  
P16\_SRR090578  
P16\_SRR090577  
P16\_SRR090576  
P16\_SRR090575  
P16\_SRR090574  
P16\_SRR090573  
P16\_SRR090572  
P16\_SRR090571  
P16\_SRR090570  
P16\_SRR090569  
P16\_SRR090568  
P16\_SRR090567  
P16\_SRR090566  
P16\_SRR090565  
P16\_SRR090564  
P16\_SRR090563  
P16\_SRR090562  
P16\_SRR090561  
P16\_SRR090560  
P16\_SRR090559  
P16\_SRR090558  
P16\_SRR090557  
P16\_SRR090556  
P16\_SRR090555  
P16\_SRR090554  
P16\_SRR090553  
P16\_SRR090552  
P16\_SRR090551  
P16\_SRR090550  
P16\_SRR090549  
P16\_SRR090548  
P16\_SRR090547  
P16\_SRR090546  
P16\_SRR090545  
P16\_SRR090544  
P16\_SRR090543  
P16\_SRR090542  
P16\_SRR090541  
P16\_SRR090540  
P16\_SRR090539  
P16\_SRR090538  
P16\_SRR090537  
P16\_SRR090536  
P16\_SRR090535  
P16\_SRR090534  
P16\_SRR090533  
P16\_SRR090532  
P16\_SRR090531  
P16\_SRR090530  
P16\_SRR090529  
P16\_SRR090528  
P16\_SRR090527  
P16\_SRR090526  
P16\_SRR090525  
P16\_SRR090524  
P16\_SRR090523  
P16\_SRR090522  
P16\_SRR090521  
P16\_SRR090520  
P16\_SRR090519  
P16\_SRR090518  
P16\_SRR090517  
P16\_SRR090516  
P16\_SRR090515  
P16\_SRR090514  
P16\_SRR090513  
P16\_SRR090512  
P16\_SRR090511  
P16\_SRR090510  
P16\_SRR090509  
P16\_SRR090508  
P16\_SRR090507  
P16\_SRR090506  
P16\_SRR090505  
P16\_SRR090504  
P16\_SRR090503  
P16\_SRR090502  
P16\_SRR090501  
P16\_SRR090500

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

Genome: hg19  
Cell size: 500kb size bins (100k reads)  
Using normalized read counts  
Metric: linkage, euclidean distance

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR090111	[CNV profile bar]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR090133	[CNV profile bar]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[CNV profile bar]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[CNV profile bar]	7,582,032	1,415.63	766.63	0.54
M16_SRR090158	[CNV profile bar]	6,414,000	1,195.97	656.59	0.55
		1,300,33	1,860,61	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**

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 (283 KB).

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

M16\_SRR090158

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090210  
**M16\_SRR090158**  
M16\_SRR090212  
M16\_SRR090214

**View analysis later**

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

Genome: hg19  
Binsize: 500kb bins of 500kb size (overlaps 100kb)  
using normalized read counts  
linkage, euclidean distance

**M16\_SRR090158**

**Tree display**

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile bar]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile bar]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[CNV profile bar]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[CNV profile bar]	7,582,032	1,415.63	766.63	0.54
<b>M16_SRR090158</b>	[CNV profile bar]	6,414,000	1,195.97	656.59	0.55

**Normalized read counts** [newick | xml | pdf | jpeg]

**Copy-number** [newick | xml | pdf | jpeg]

**Correlations** [newick | xml | pdf | jpeg]

**Download processed data**

**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 (28 KB)

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHpiCXcE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090218  
M16\_SRR090158  
**M16\_SRR090212**  
M16\_SRR090144  
M16\_SRR090145

**View analysis later**

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=results/NHeoHpiCXcE7hqmlx2c>

**Analysis Parameters**

Genome: hg19  
Binning: variable bins of 500kb size (over 100k reads)  
using normalized read counts  
linkage, euclidean distance

**M16\_SRR090212**

**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	1,147.06	0.35
M16_SRR090144		11,757,180	2,192.26	1,252.84	0.57
M16_SRR090158		7,582,032	1,415.63	766.63	0.54
M16_SRR090158		6,414,000	1,195.97	656.59	0.55

**Normalized read counts** [newick | xml | pdf | jpeg]

**Copy-number** [newick | xml | pdf | jpeg]

**Correlations** [newick | xml | pdf | jpeg]

**Download processed data**

**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 (28 KB)

Ginkgo

qb.cshl.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090218  
M16\_SRR090158  
M16\_SRR090212  
**M16\_SRR090244**  
M16\_SRR090156

**View analysis later**

Access your results later at the following address:

<http://qb.cshl.edu/ginkgo/?q=results/NHeoHepICxEd7hqmlx2c>

**Analysis Parameters**

Genome: hg19  
Binning: variable bins of 500kb size (overlaps 100 reads)  
Segmentation: using normalized read counts  
Distance metric: euclidean distance

**Summary**

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711		15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733		17,395,808	3,243.67	1,147.06	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

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

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

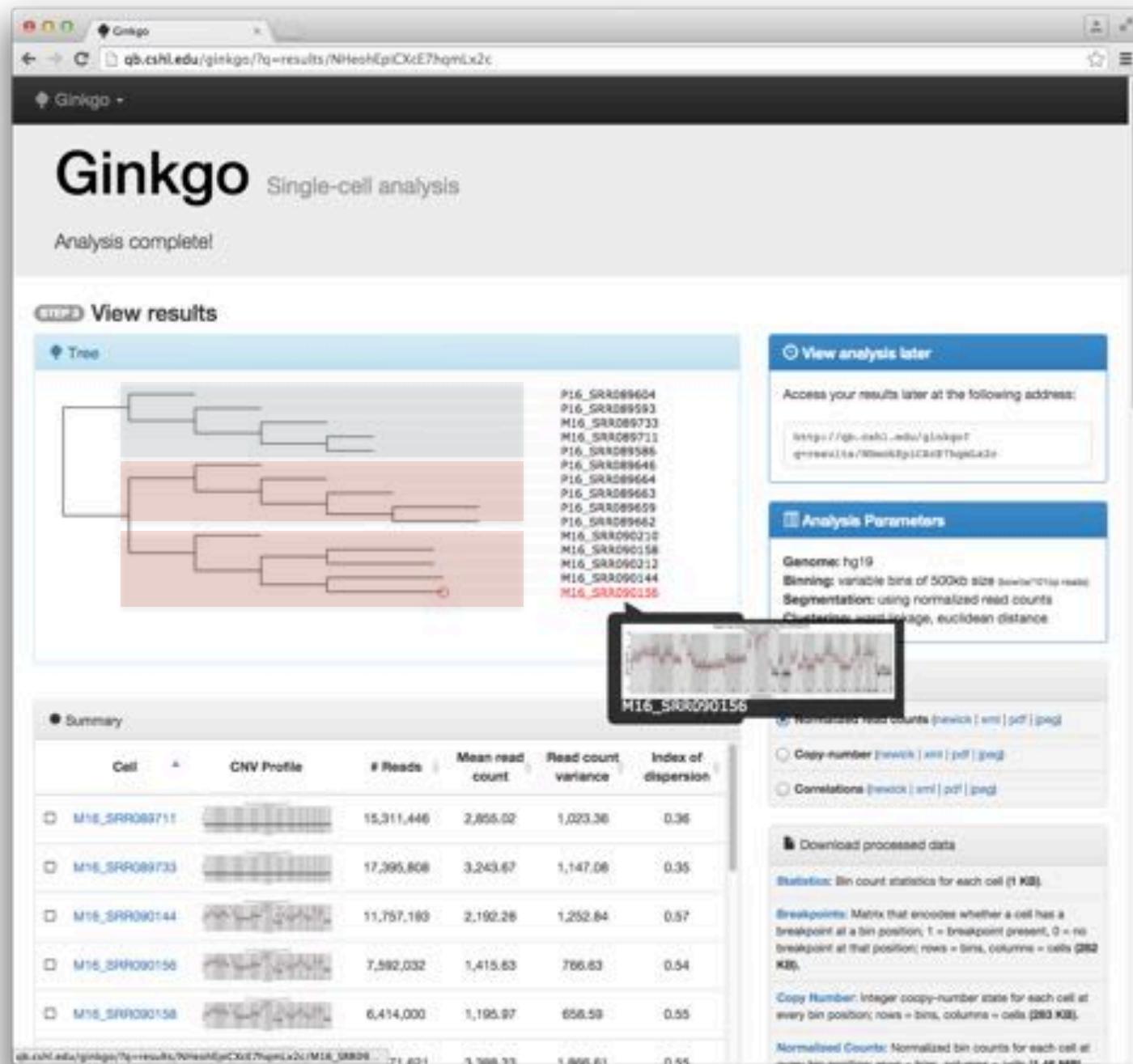
Correlations ([newick](#) | [xml](#) | [pdf](#) | [jpeg](#))

**Download processed data**

**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 (28 KB)



Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090218  
M16\_SRR090158  
M16\_SRR090112  
M16\_SRR090144  
M16\_SRR090156

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c>

**Analysis Parameters**

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

**Tree display**

Normalized read counts [newick | xml | pdf | png]

Copy-number [newick | xml | pdf | png]

Correlations [newick | xml | pdf | png]

**Download processed data**

**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 (28 KB)

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR089711	[CNV profile image]	15,311,446	2,865.02	1,023.36	0.36
M16_SRR089733	[CNV profile image]	17,395,808	3,243.67	1,147.06	0.35
M16_SRR090144	[CNV profile image]	11,757,180	2,192.26	1,252.84	0.57
M16_SRR090156	[CNV profile image]	7,582,032	1,415.63	766.63	0.54
M16_SRR090158	[CNV profile image]	6,414,000	1,195.97	656.59	0.55
M16_SRR090210	[CNV profile image]	18,171,621	3,308.33	1,860.61	0.55

Ginkgo

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://gb.csHL.edu/ginkgo/>  
<http://gb.csHL.edu/ginkgo/1C8071bgkzad8>

**Analysis Parameters**

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

**Tree display**

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

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

Correlations (newick | 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 (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: CNV profiles | Lorenz curve | GC bias | MAD | Select all | Deselect all

Ginkgo

qbs.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c

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:

www://gb.csHL.edu/ginkgo/  
q=results/NHeoHepICxE7hqmlx2c

**Analysis Parameters**

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

**Tree display**

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

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

Correlations (newick | 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 (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: CNV profiles | Lorenz curve | GC bias | MAD | Select all | Deselect all

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHcpICx7hqmlx2c

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://qb.csHL.edu/ginkgo/?q=results/NHeoHcpICx7hqmlx2c>

**Analysis Parameters**

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

**Summary**

Sample	Normalized read counts	Copy number	Correlations	Breakpoints
M16_SRR090158	6,414,000	1,195.97	658.59	0.55
M16_SRR090210	18,171,621	3,388.33	1,866.61	0.55
M16_SRR090212	17,659,321	3,292.81	1,741.57	0.53
P16_SRR089666	15,074,760	2,810.88	910.38	0.32
P16_SRR089693	16,176,073	3,016.24	1,035.63	0.34
P16_SRR089604	9,760,274	1,074.06	357.84	0.33

With selected cells, plot: [CNV profiles](#) [Lorenz curve](#) [GC bias](#) [MAD](#) [Select all](#) [Deselect all](#)

**Tree display**

Normalized read counts ([newick](#) | [xml](#) | [pdf](#) | [png](#))  
 Copy-number ([newick](#) | [xml](#) | [pdf](#) | [png](#))  
 Correlations ([newick](#) | [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 (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).

Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHcpICxEd7hqmlx2c

Analysis completed.

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089654  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR089210  
M16\_SRR089158  
M16\_SRR089012  
M16\_SRR089044  
M16\_SRR089056

**Summary**

Sample ID	Number of Cells	Total Bins	Mean Bins	Median Bins	CV
M16_SRR0890210	18,171,621	3,388.33	1,866.61	0.55	
M16_SRR0890212	17,659,321	3,292.81	1,741.57	0.53	
P16_SRR089686	15,674,760	2,810.88	910.36	0.32	
P16_SRR089640	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	8,278,396	1,543.61	834.04	0.54	
P16_SRR089656	14,513,617	2,705.25	1,459.16	0.54	

With selected cells, plot: CNV profiles | Lorenz curve | GC bias | MAD | Select all | Deselect all

**View analysis later**

Access your results later at the following address:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHcpICxEd7hqmlx2c>

**Analysis Parameters**

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

**Tree display**

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

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

Correlations (newick | 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 (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).

Ginkgo

Analysis completed.

### View results

Tree

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089654  
P16\_SRR089663  
P16\_SRR089659  
P16\_SRR089662  
M16\_SRR090210  
M16\_SRR090158  
M16\_SRR090212  
M16\_SRR090144  
M16\_SRR090156

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_SRR089606		15,074,760	2,810.88	910.36	0.32
<input type="checkbox"/>	P16_SRR089640		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_SRR089658		14,513,617	2,705.25	1,459.16	0.54

With selected cells, plot: CNV profiles | Lorenz curve | GC bias | MAD | Select all | Deselect all

View analysis later

Access your results later at the following address:

<http://gb.cshl.edu/ginkgo/>  
<http://gb.cshl.edu/ginkgo/?q=results/NHeoHcpICxCE7hqmlx2c>

Analysis Parameters

Genome: hg19  
Binning: variable bins of 500kb size (chromosome level)  
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

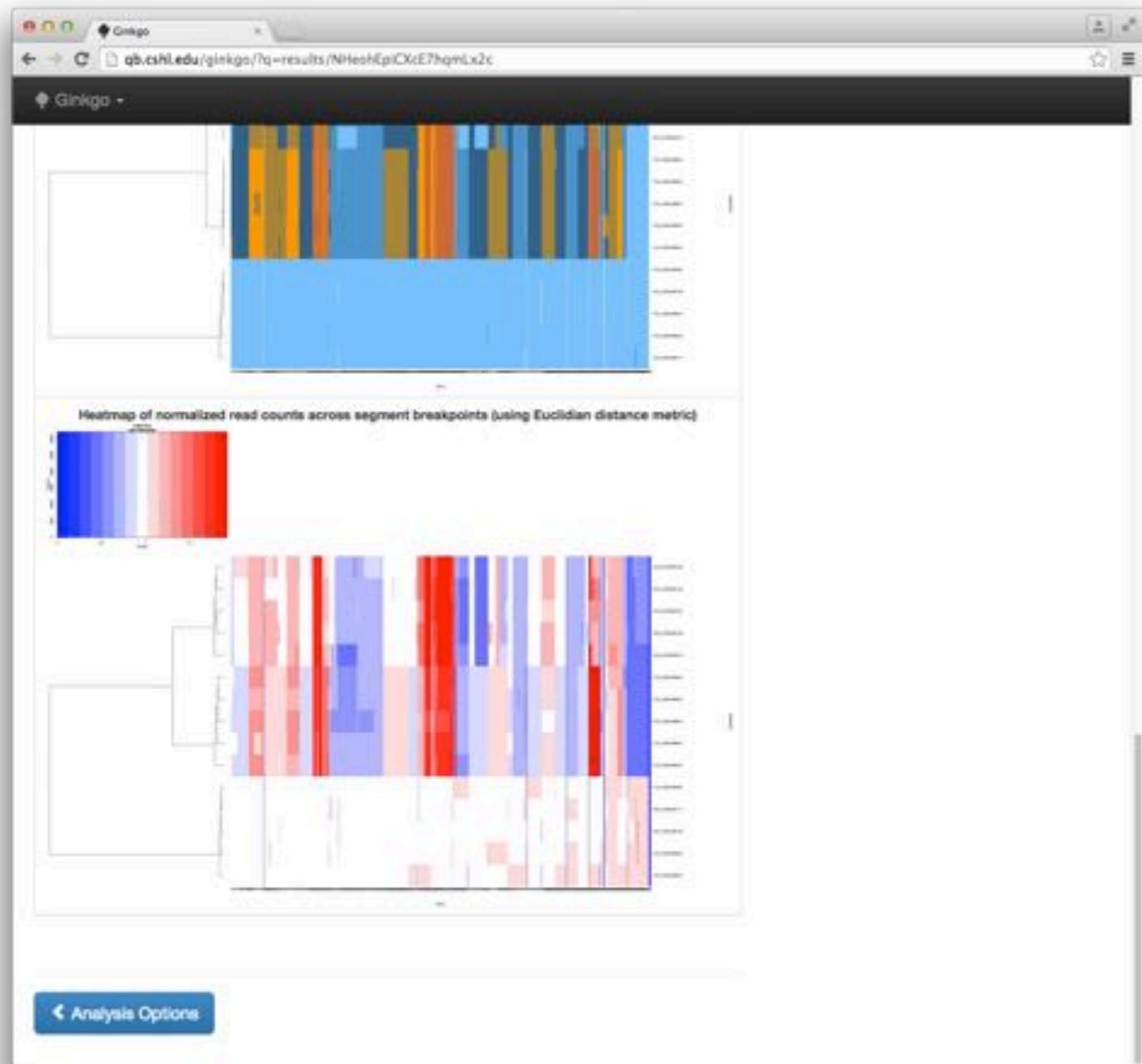
Statistics: Bin count statistics for each cell (1 KB).

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).



Ginkgo

qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c

Ginkgo

# Ginkgo

Single-cell analysis

Analysis complete!

**View results**

**Tree**

P16\_SRR089604  
P16\_SRR089593  
M16\_SRR089733  
M16\_SRR089711  
P16\_SRR089586  
P16\_SRR089646  
P16\_SRR089644  
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:

<http://qb.csHL.edu/ginkgo/?q=results/NHeoHepICxE7hqmlx2c>

**Analysis Parameters**

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

**Summary**

<input type="checkbox"/>	M16_SRR090210		16,171,621	3,368.33	1,866.61	0.55
<input type="checkbox"/>	M16_SRR090112		17,659,321	3,292.81	1,741.57	0.53
<input type="checkbox"/>	P16_SRR089646		15,674,760	2,810.66	910.36	0.32
<input type="checkbox"/>	P16_SRR089603		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
<input checked="" type="checkbox"/>	P16_SRR089645		8,276,396	1,543.61	834.04	0.54
<input checked="" type="checkbox"/>	P16_SRR089609		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](#))

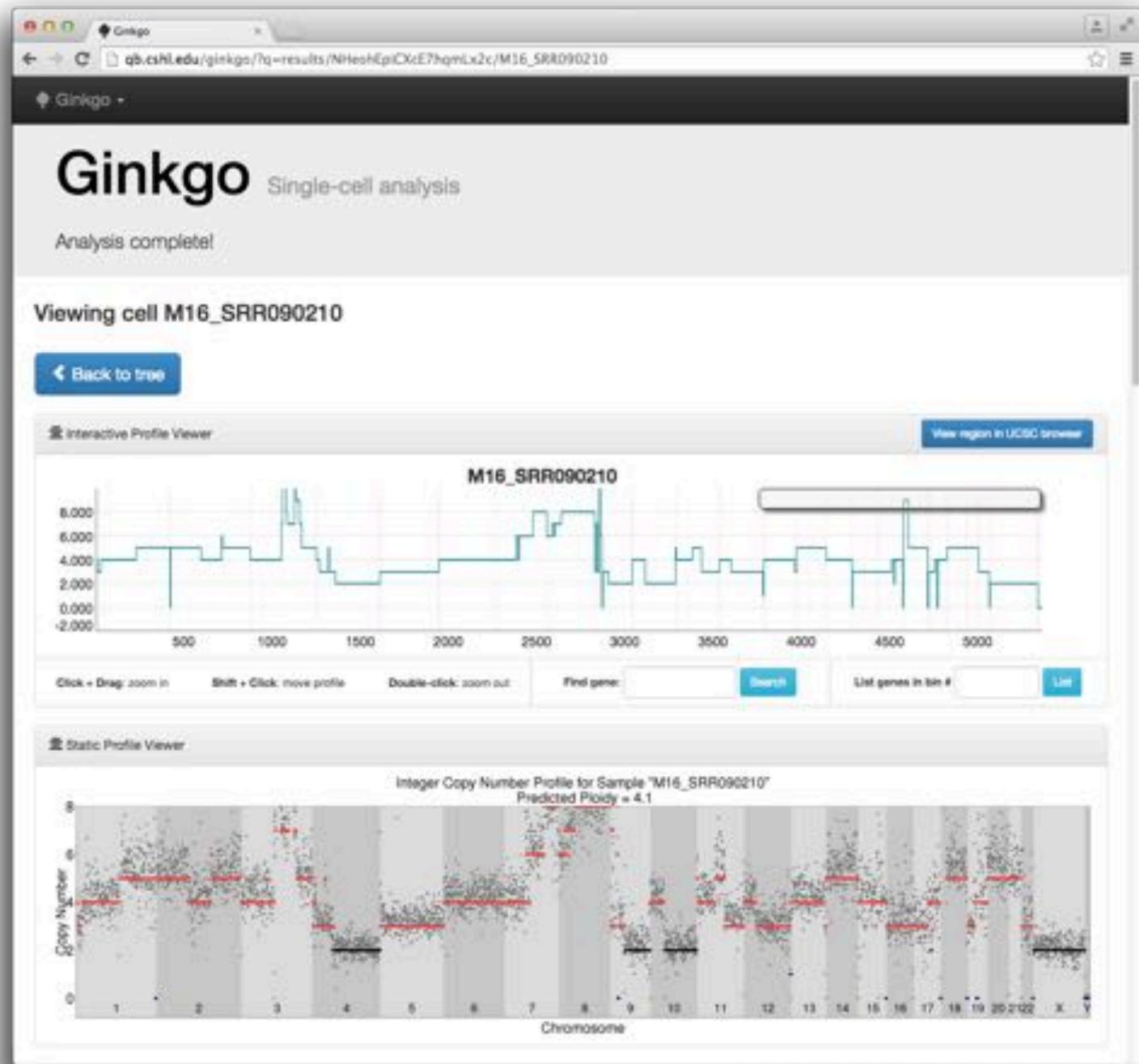
**Download processed data**

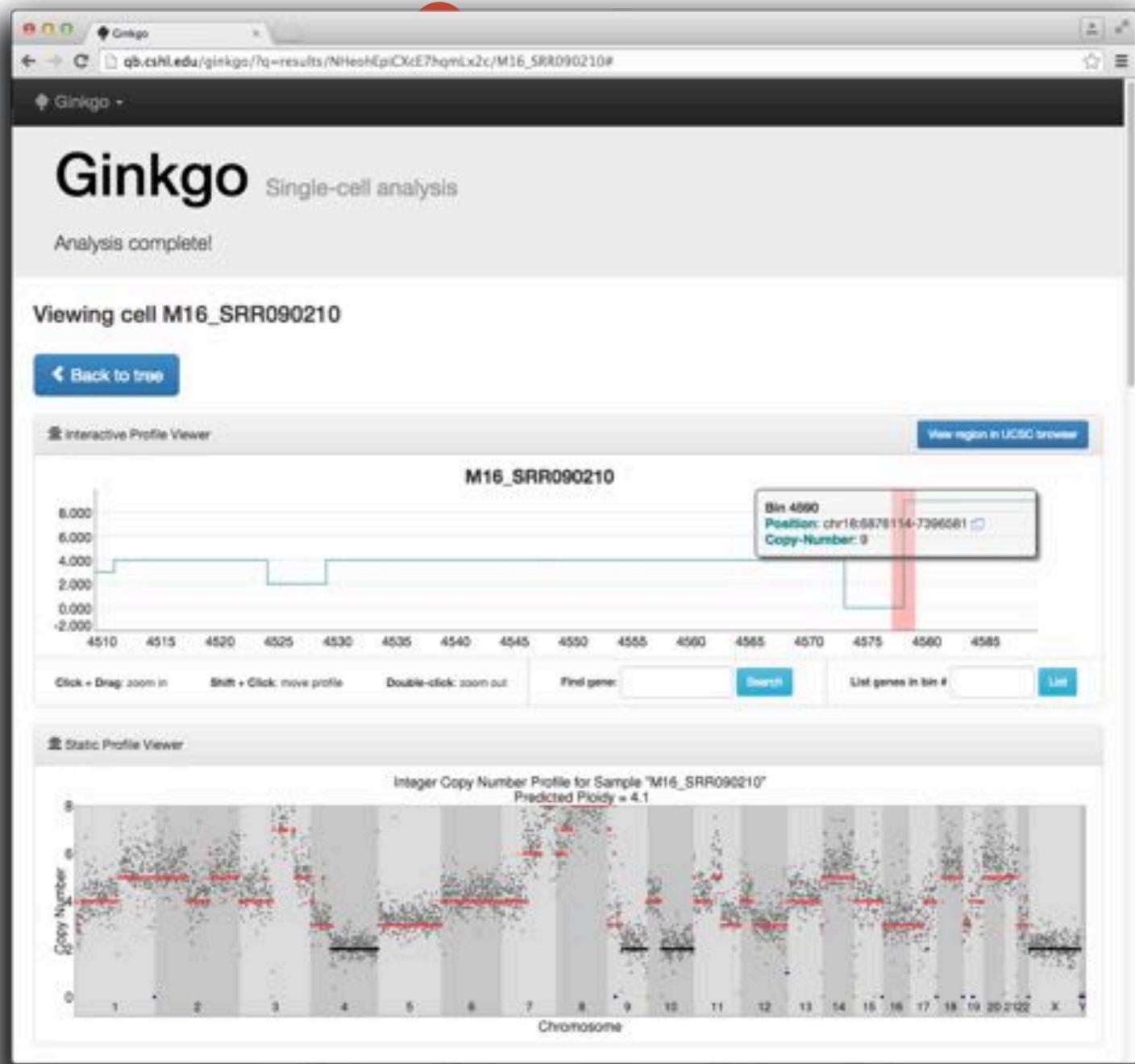
**Statistics:** Bin count statistics for each cell (1 KB).

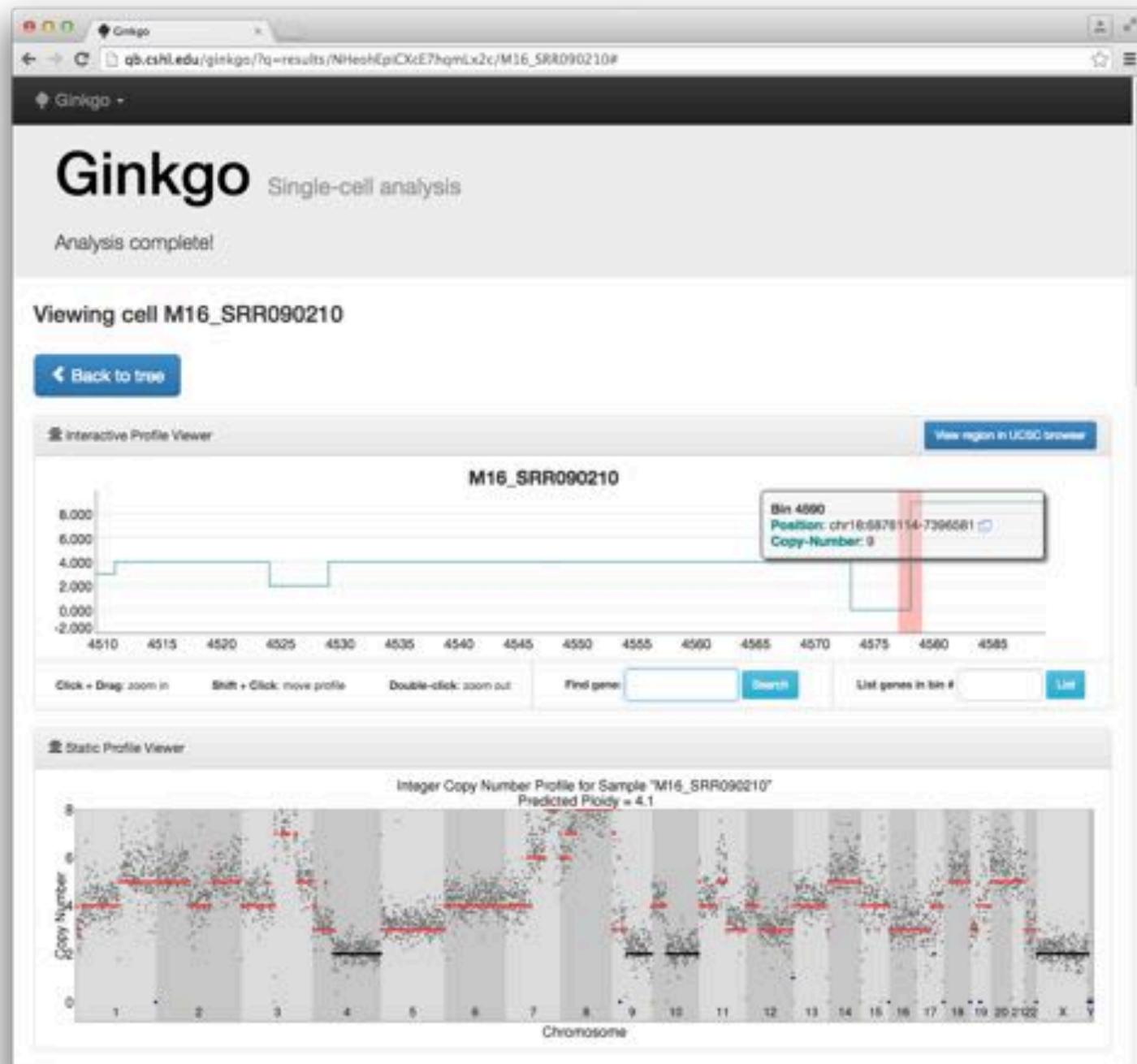
**Breakpoints:** Matrix that encodes whether a cell has a breakpoint at a bin position; 1 = breakpoint present, 0 = no breakpoint at that position; rows = bins, columns = cells (282 KB).

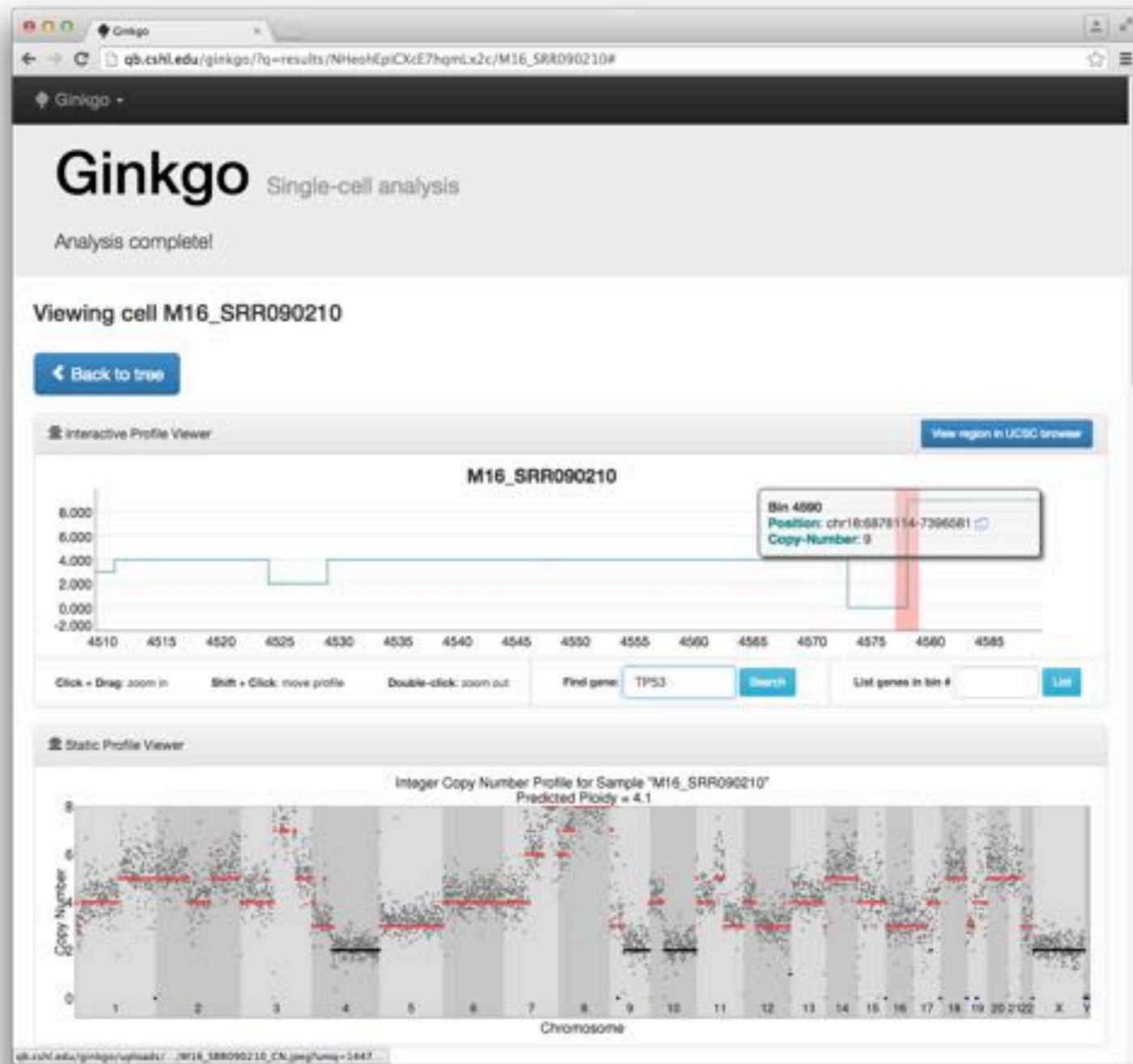
**Copy Number:** Integer copy-number state for each cell at every bin position; rows = bins, columns = cells (283 KB).

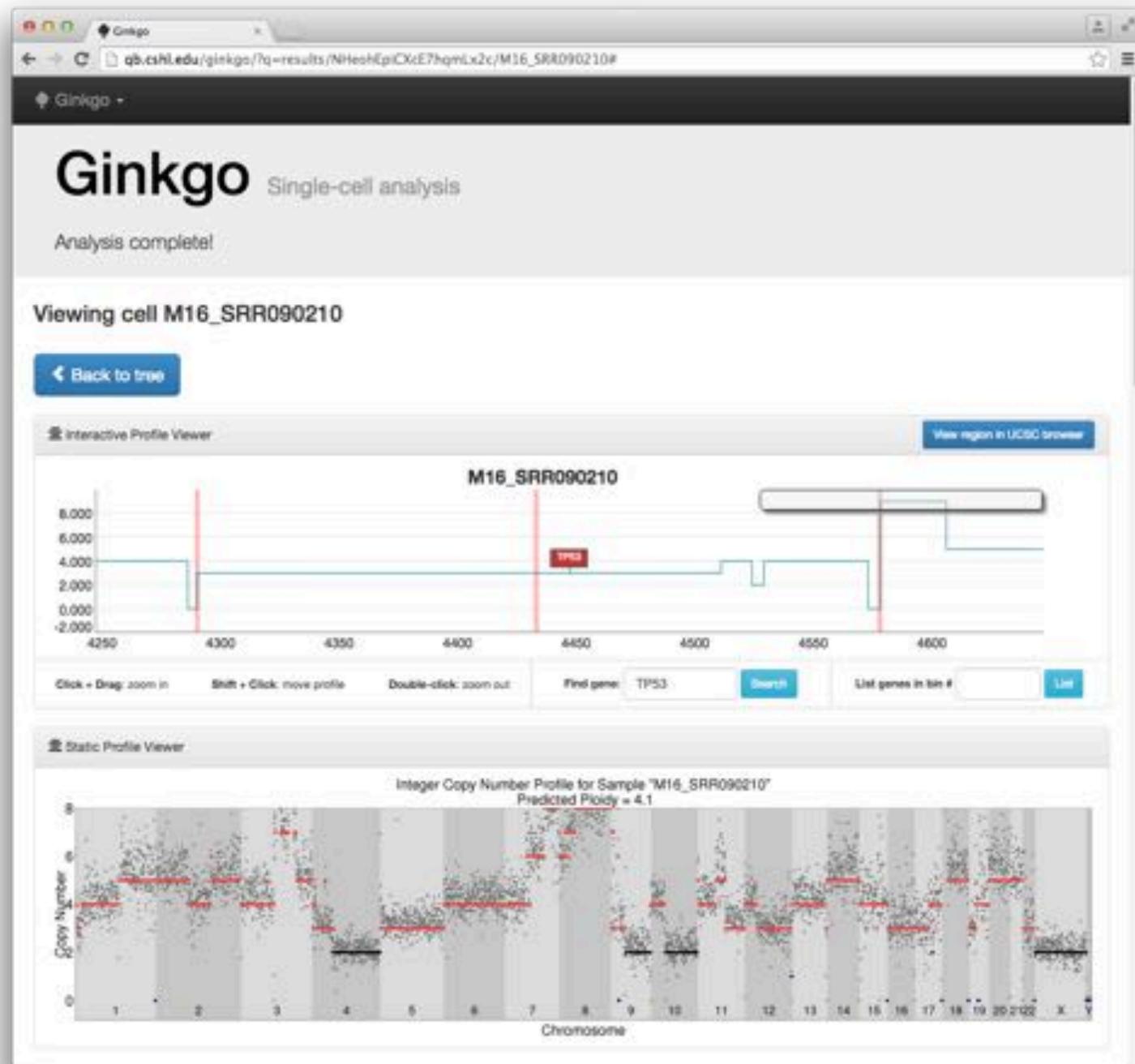
**Normalized Counts:** Normalized bin counts for each cell at every bin position; rows = bins, columns = cells (14.4K KB).

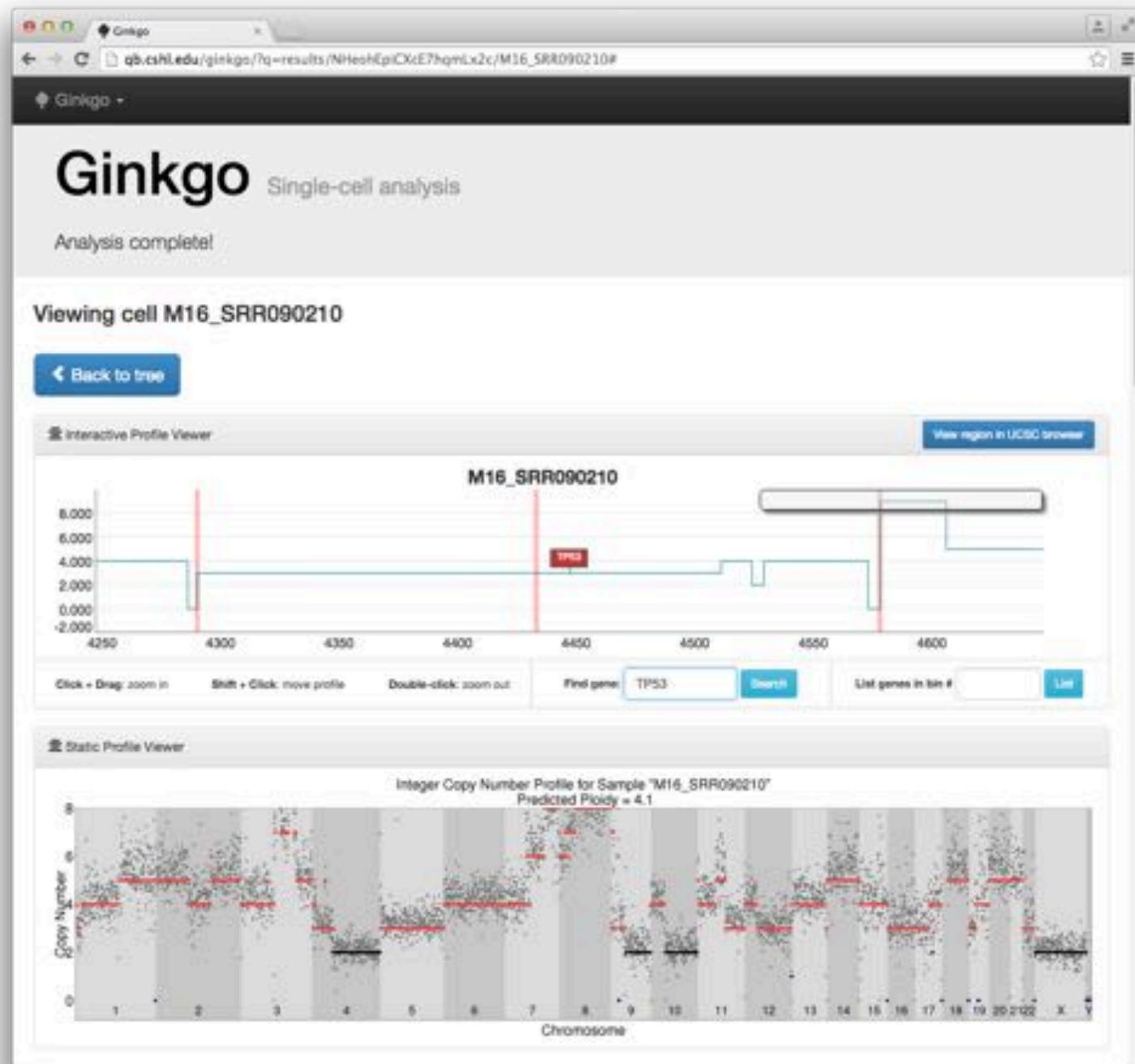


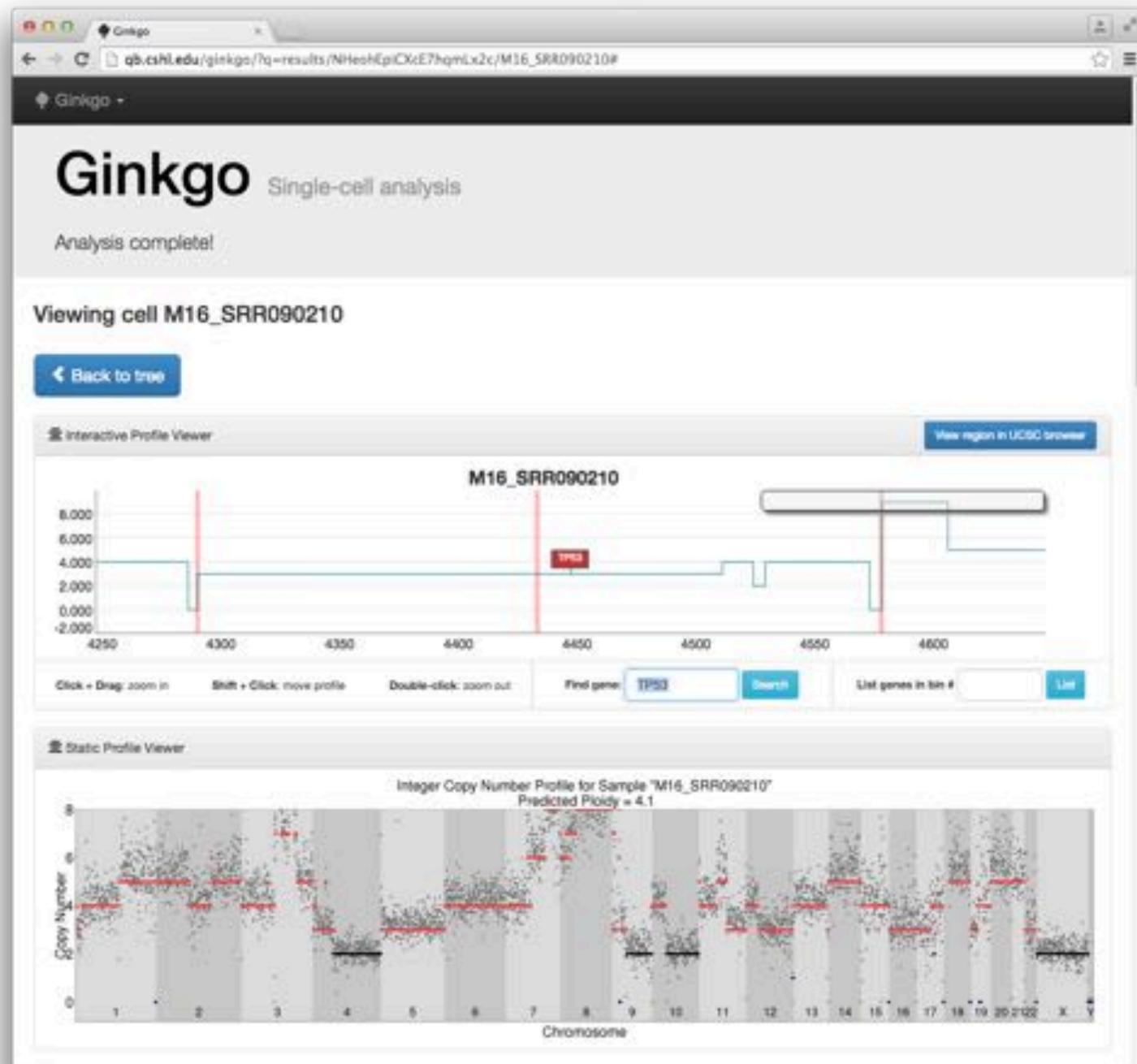


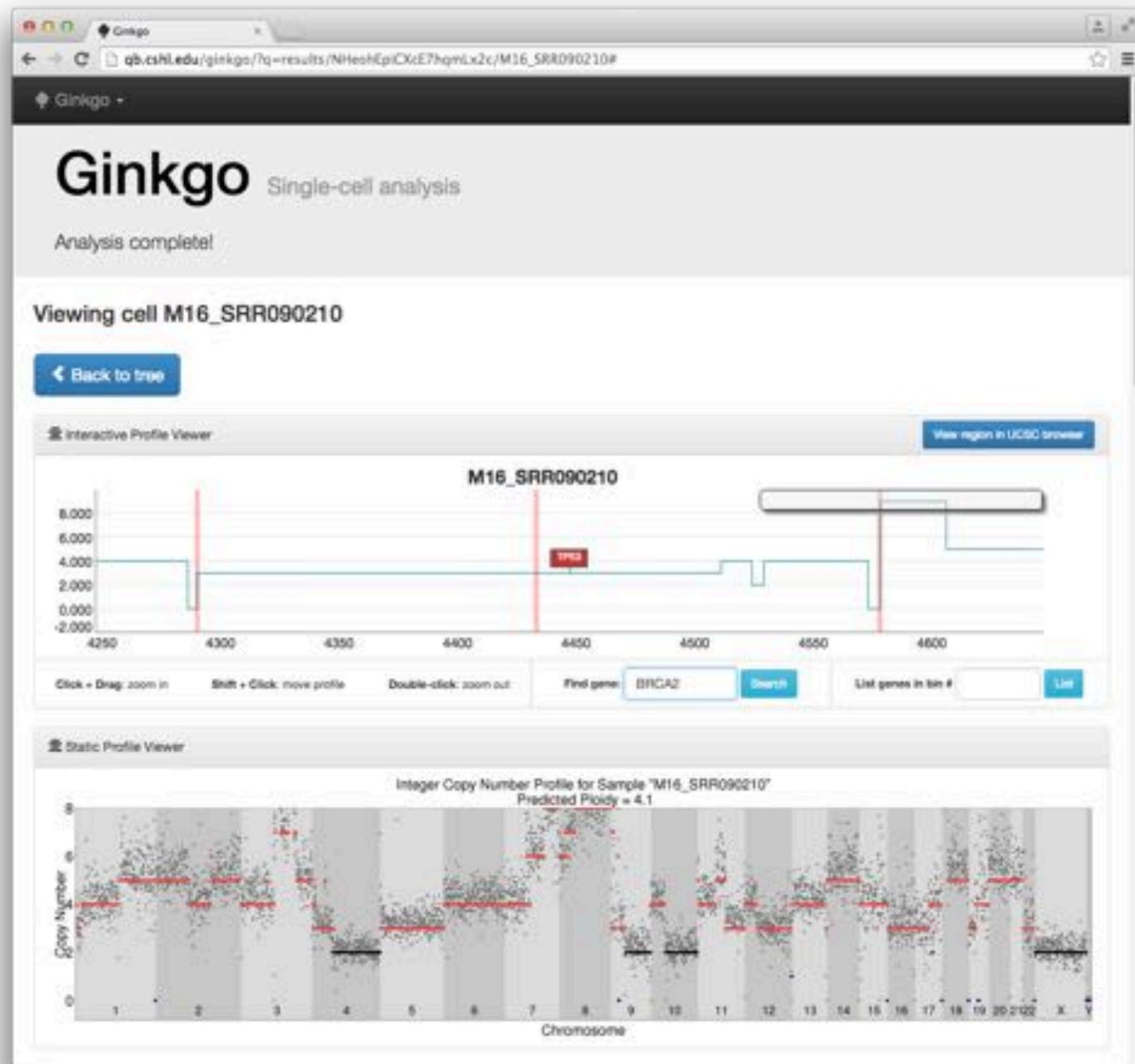


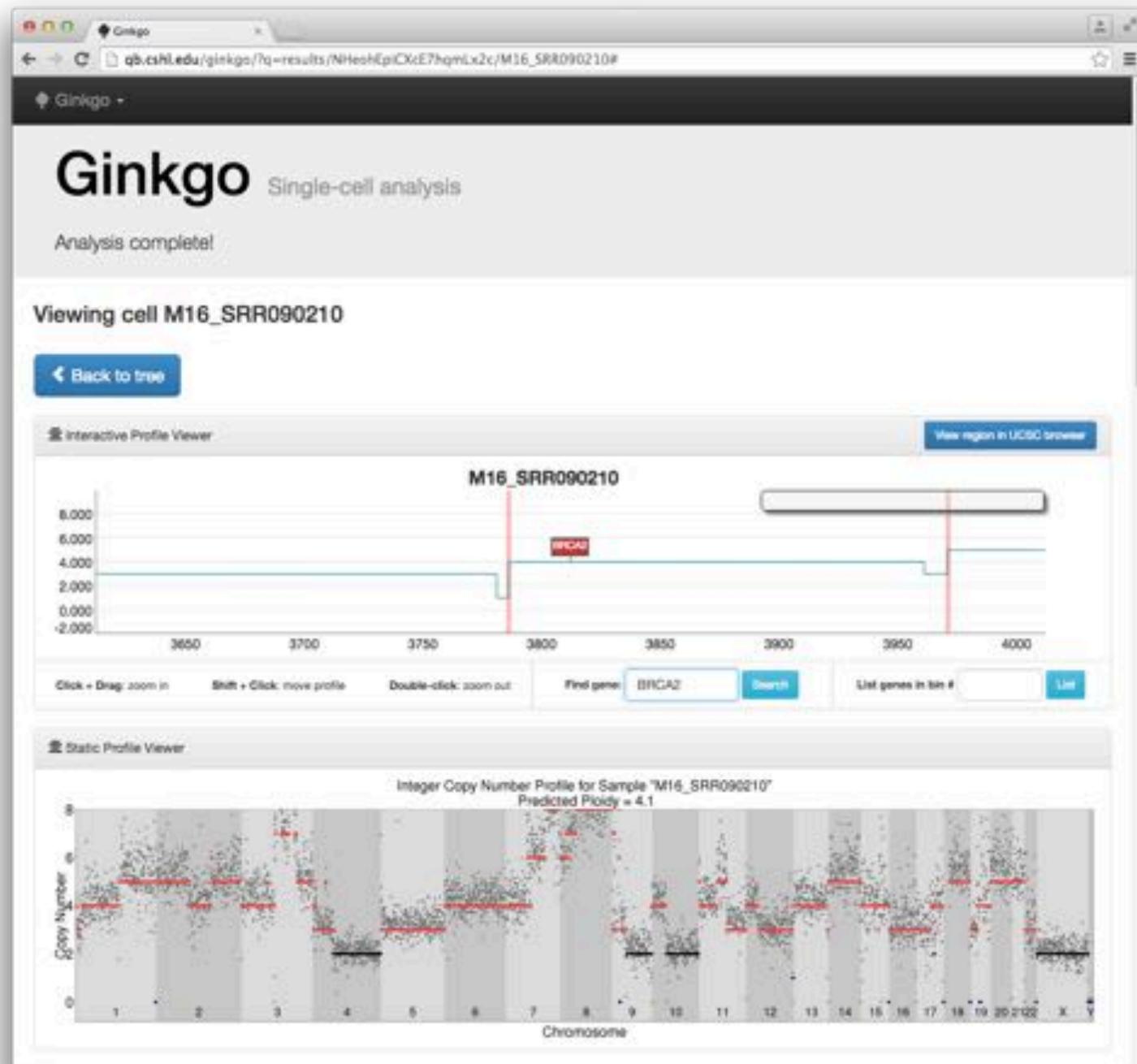


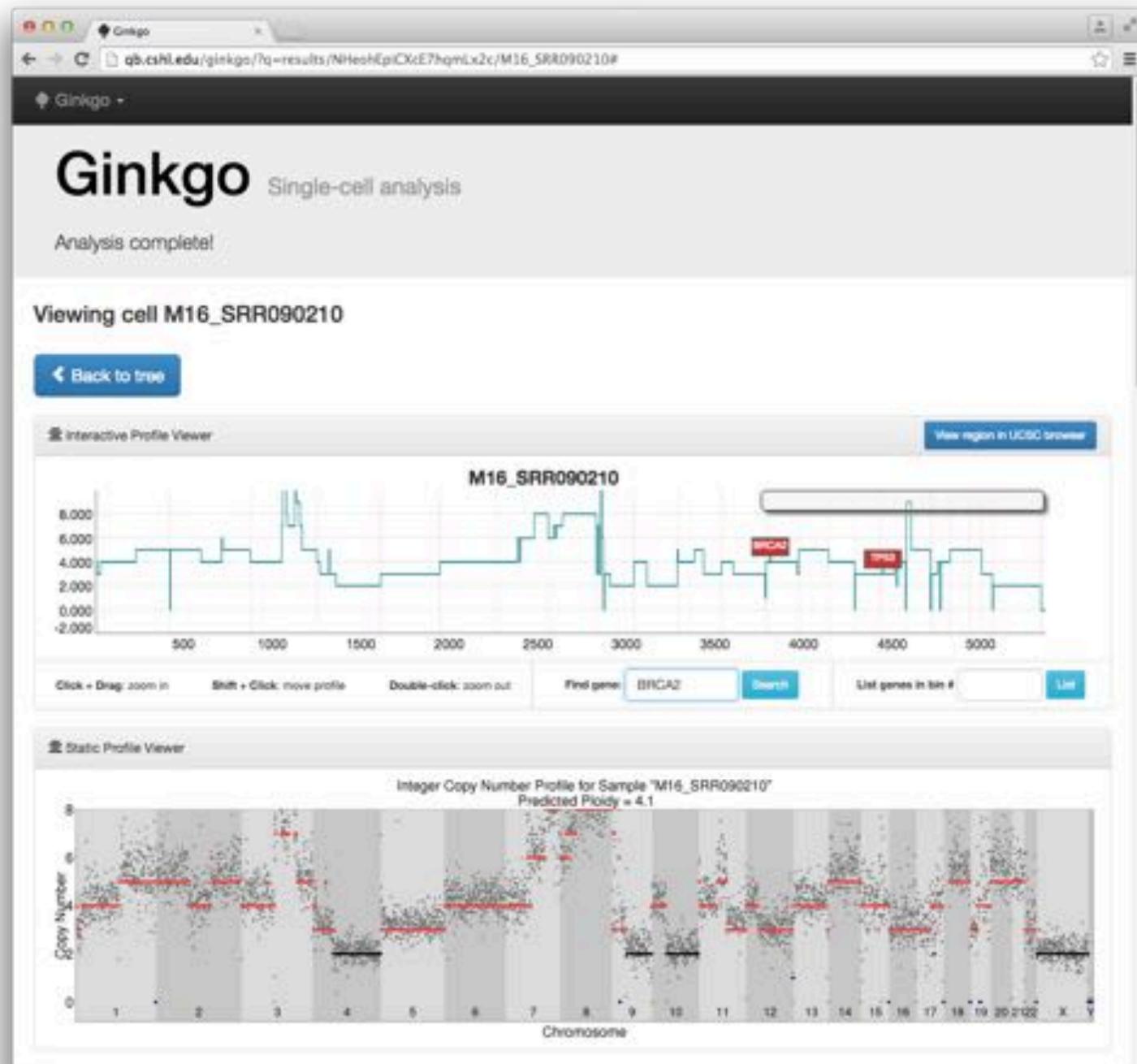


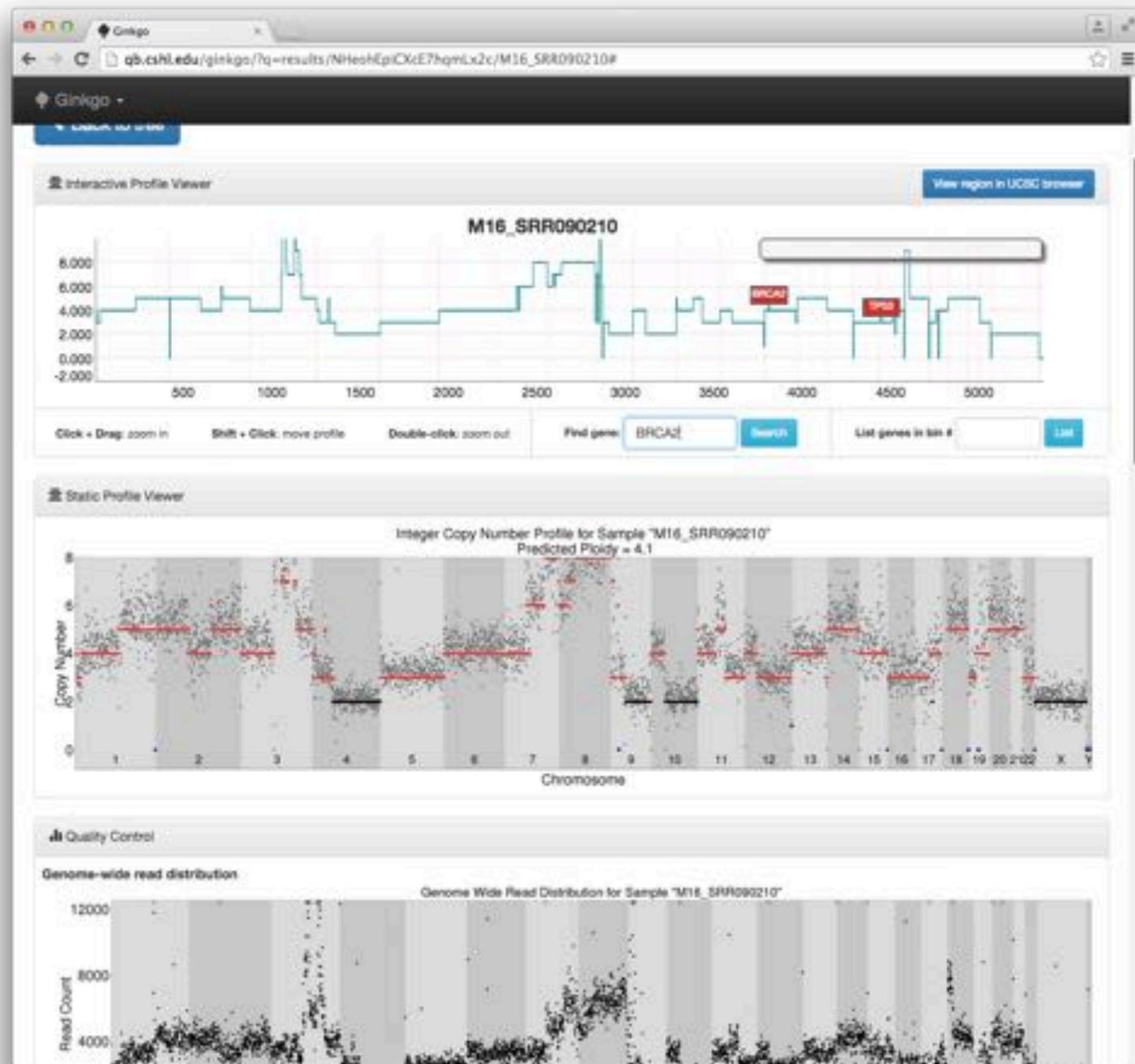


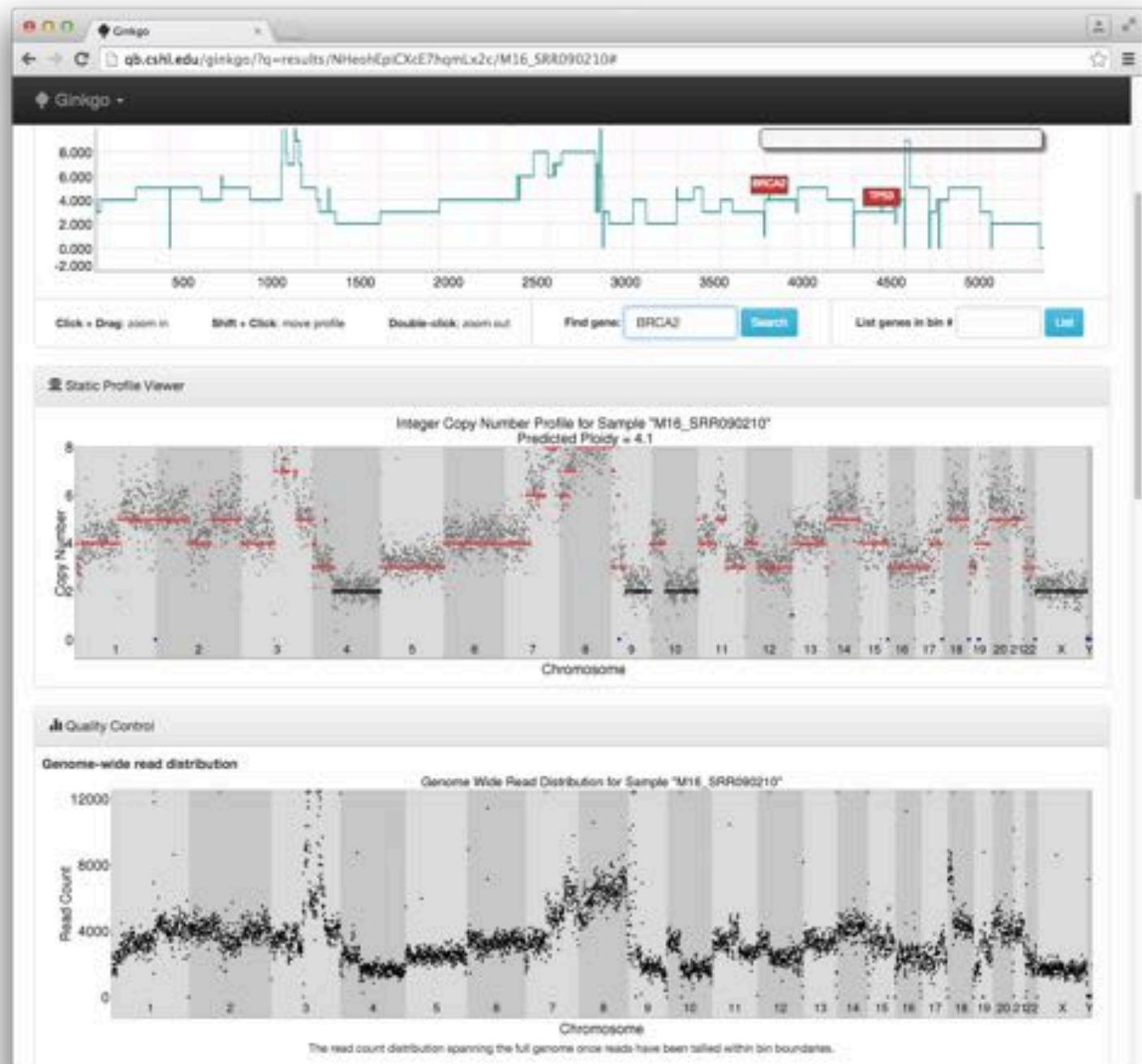


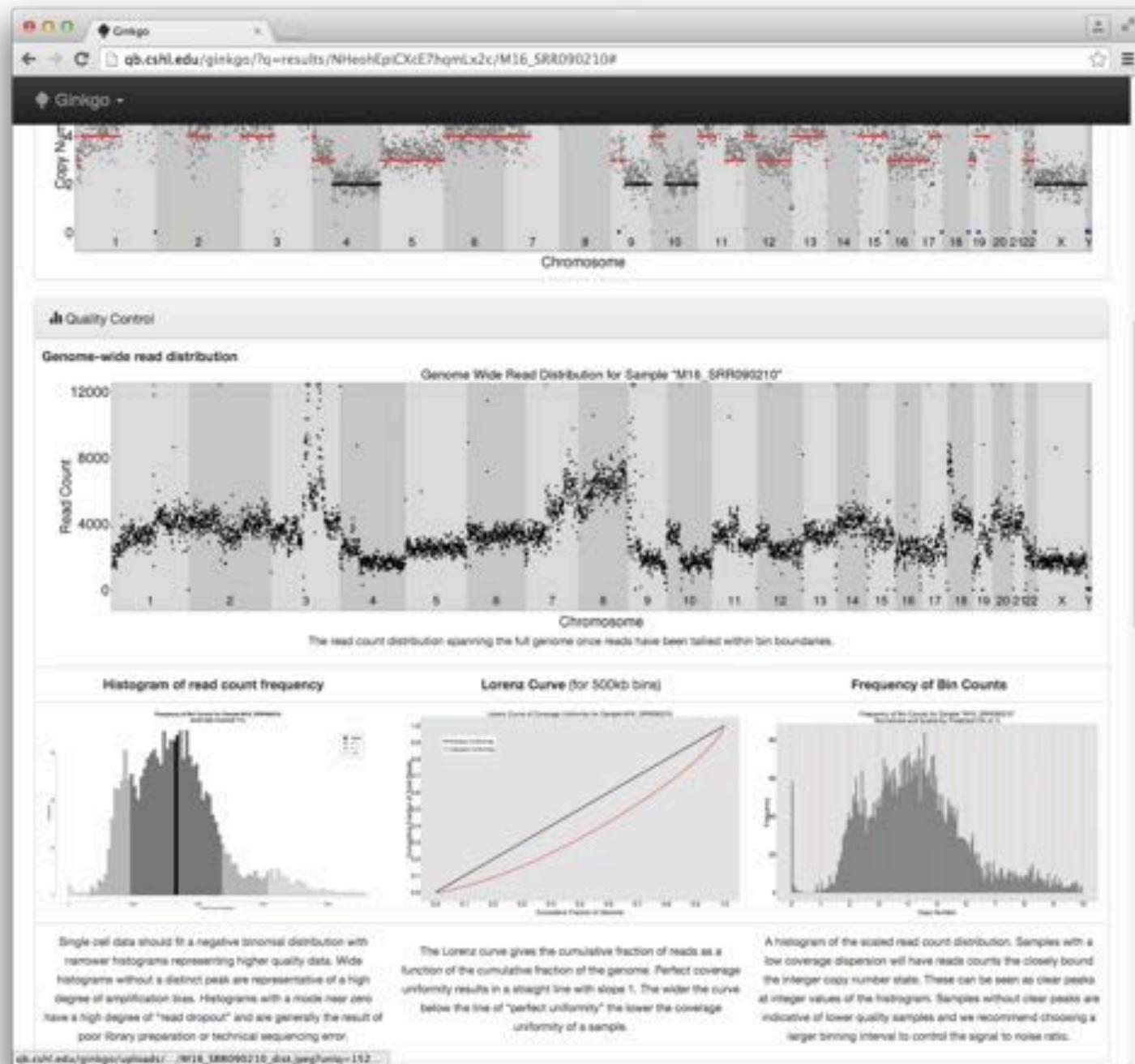


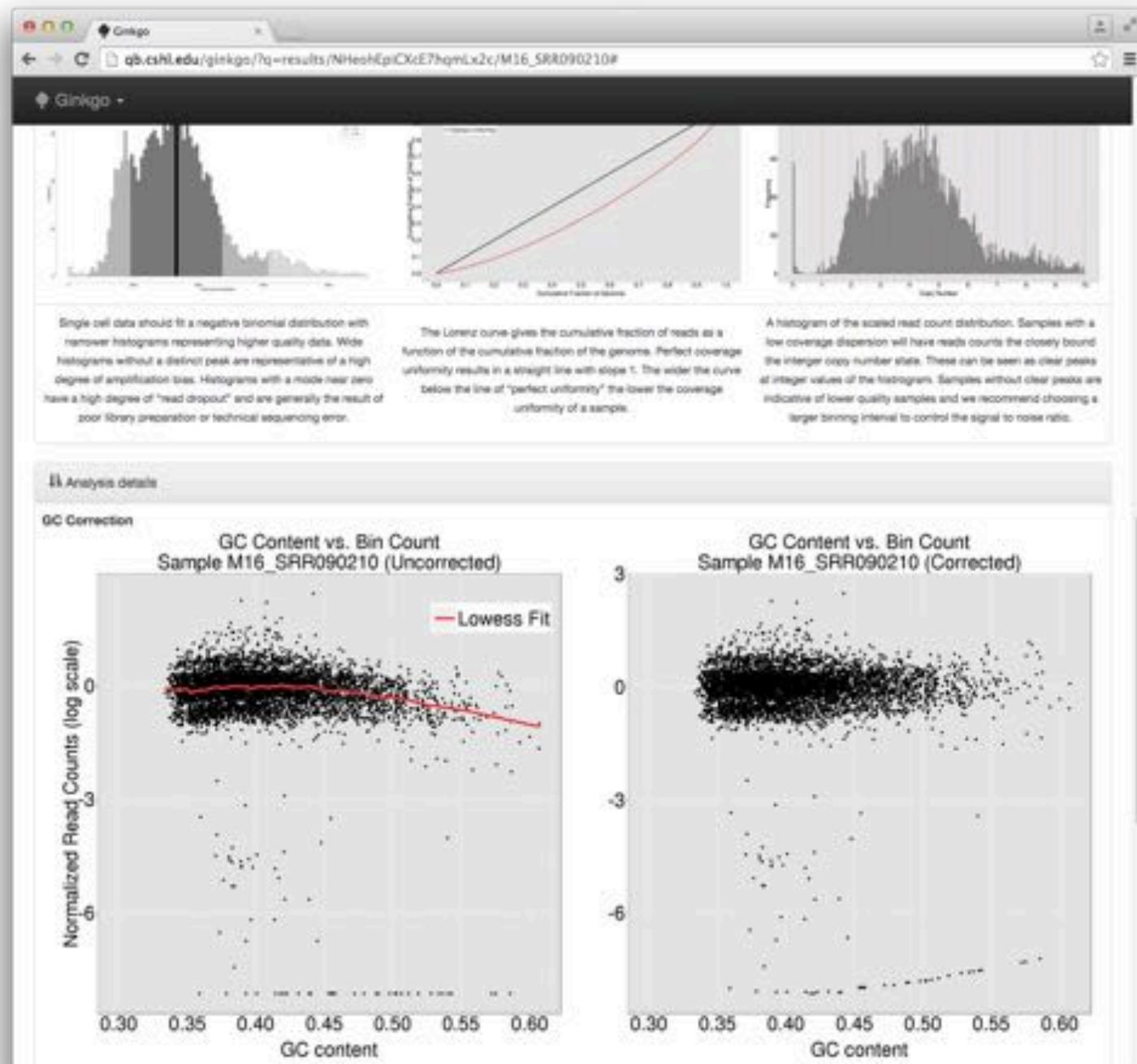


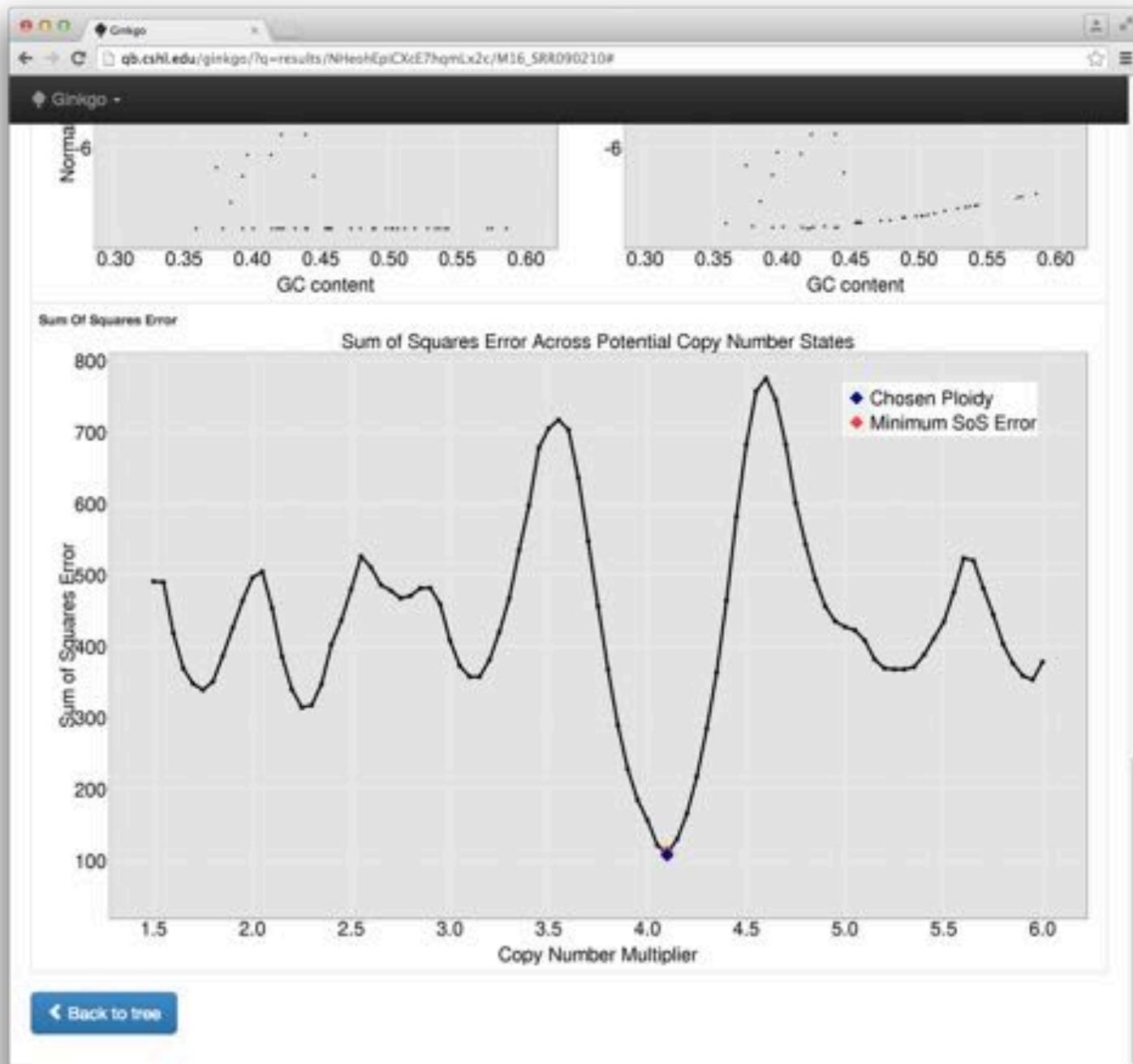










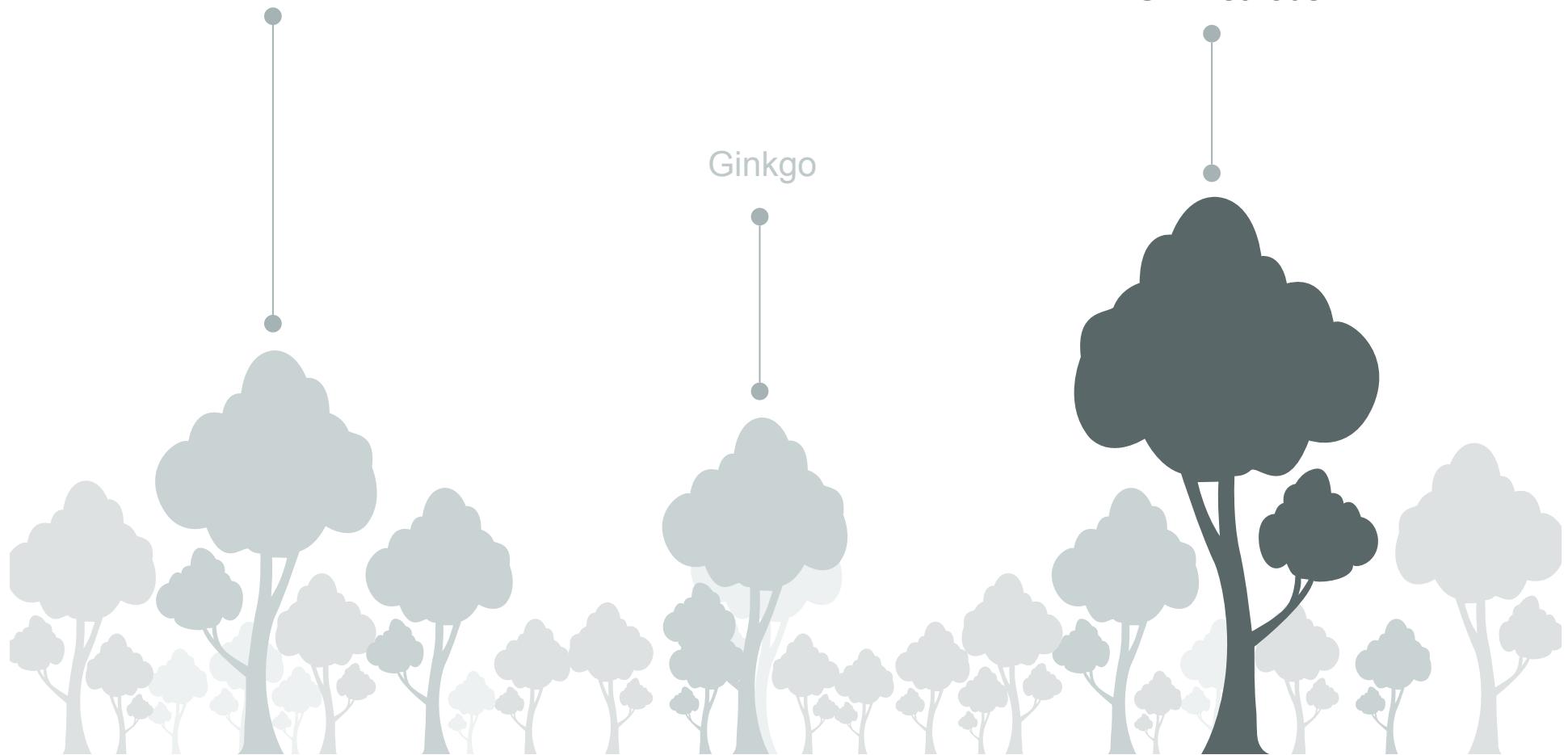


# Outline

Introduction

Ginkgo

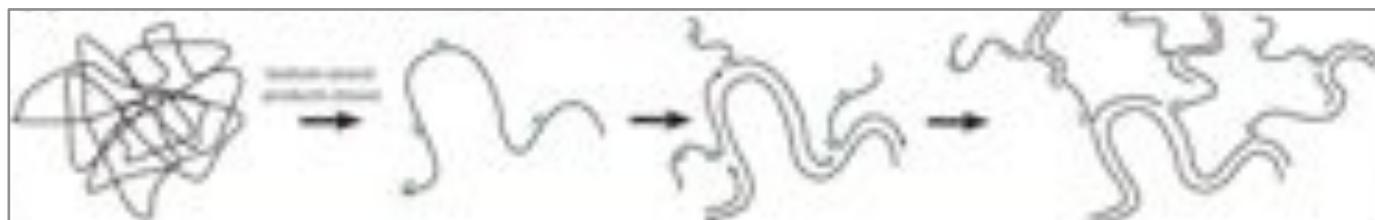
Comparison of  
WGA methods



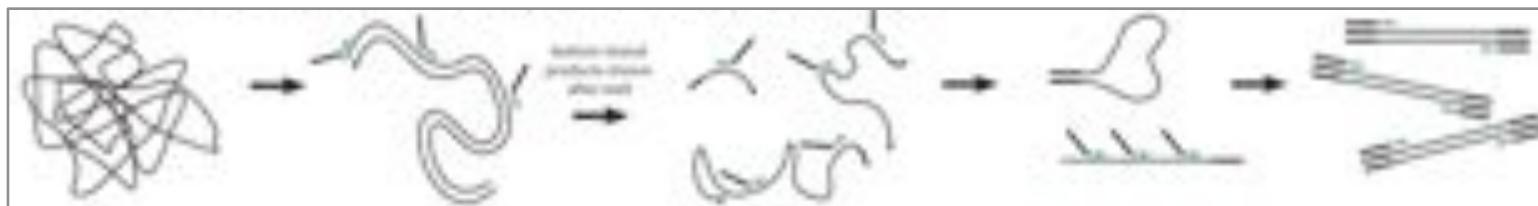
# Whole Genome Amplification (WGA) methods



DOP-PCR (Degenerate Oligonucleotide Primed PCR)



MDA (Multiple Displacement Amplification)



MALBAC (Multiple Annealing and Looping Based Amplification Cycles)

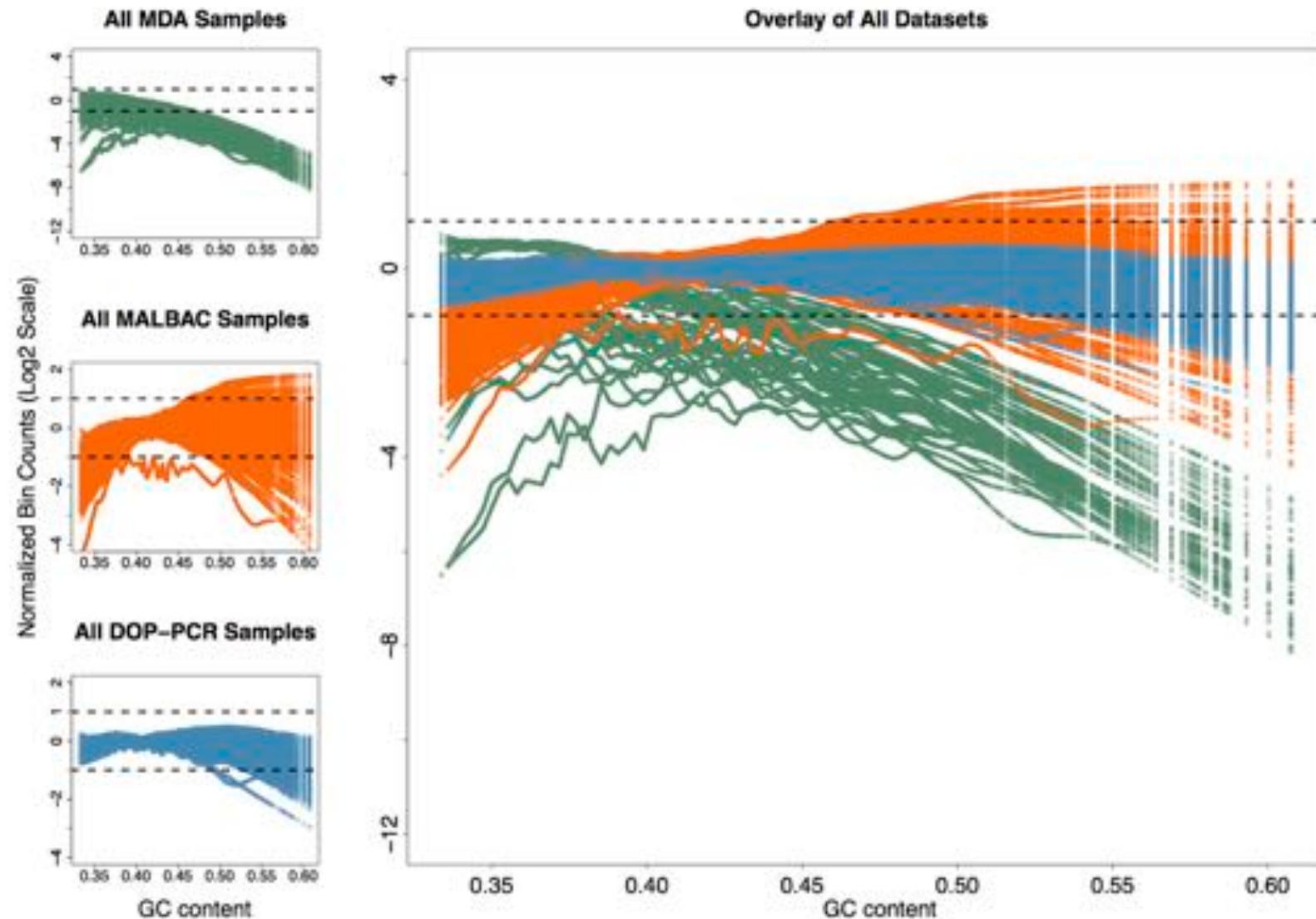
# Comparison of WGA methods

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:**

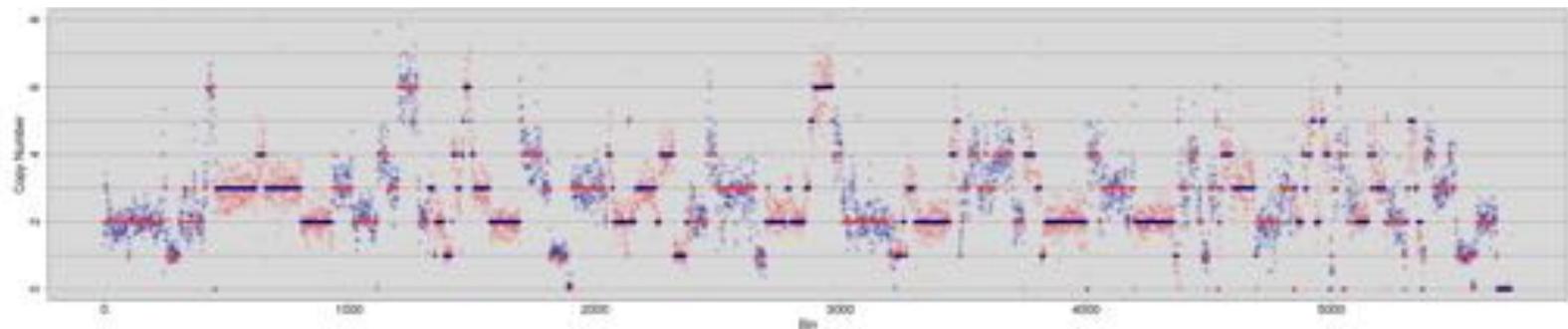
- 1) GC bias**
- 2) Coverage dispersion**

# GC Bias

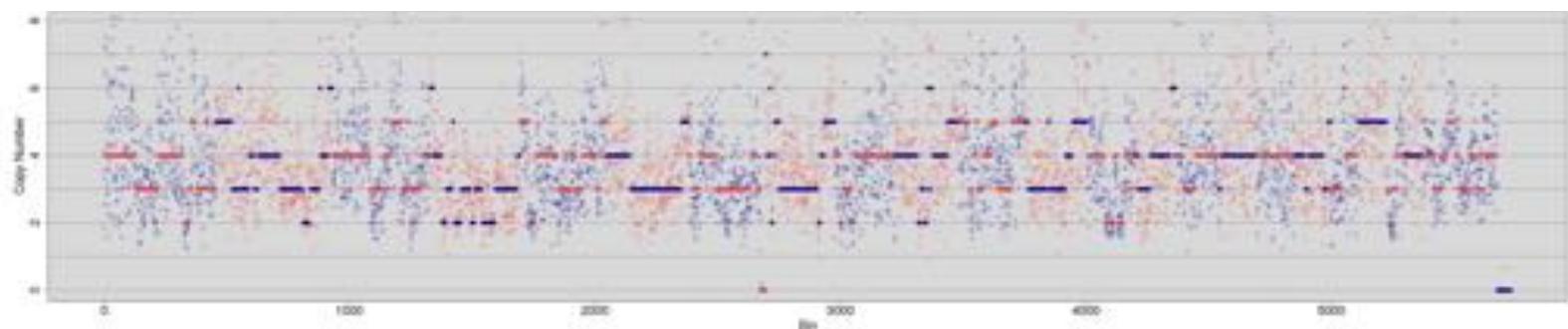


# Coverage Dispersion

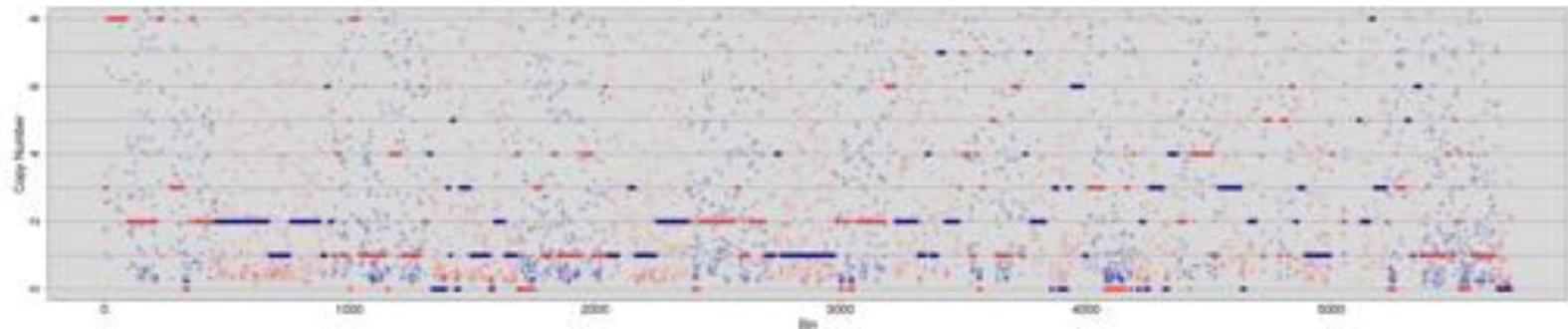
DOP-PCR



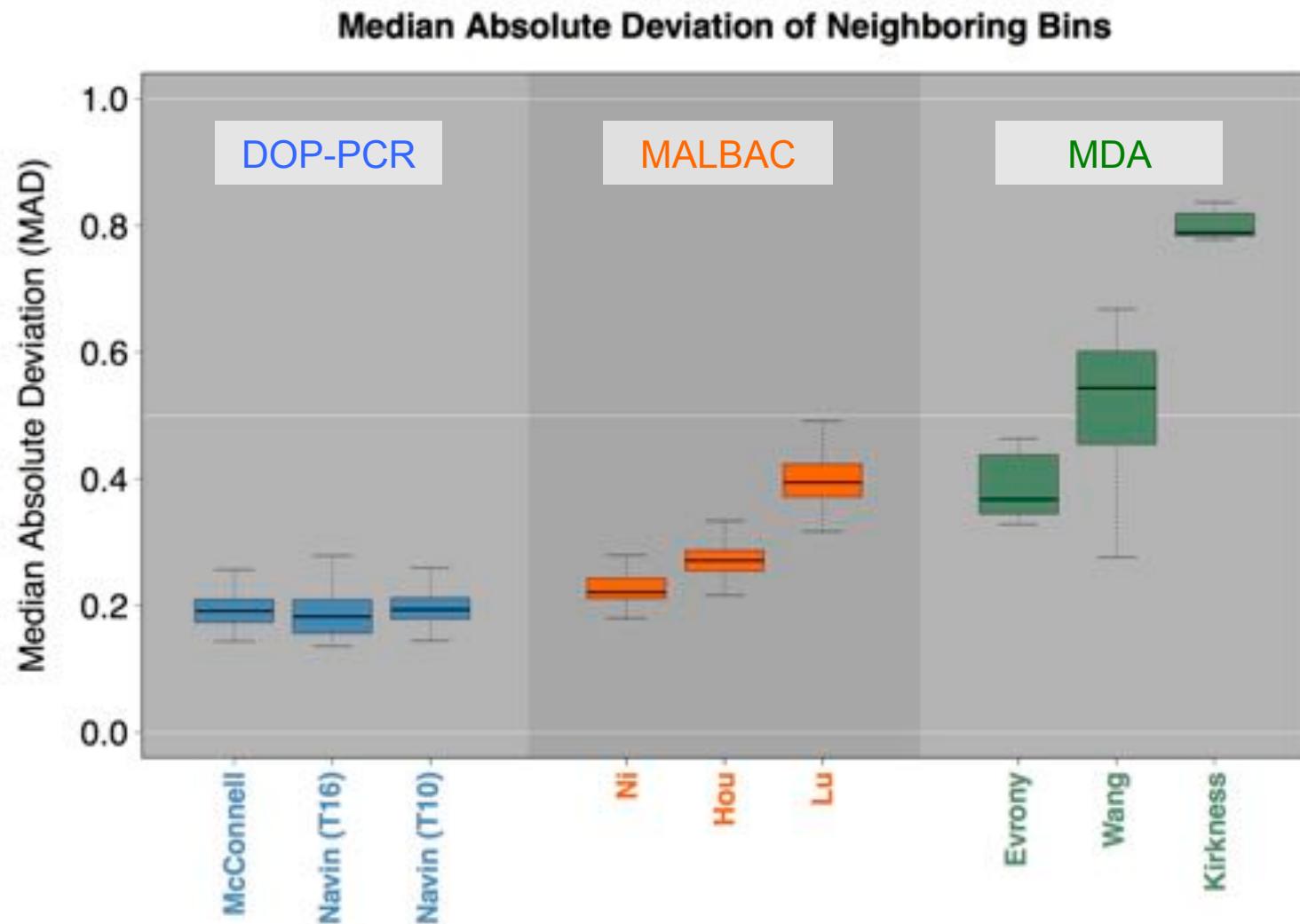
MALBAC



MDA



# Coverage Dispersion



# 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](http://qb.cshl.edu/ginkgo)
  - <http://qb.cshl.edu/ginkgo/workshop/fog.pdf>
  - *Garvin and Aboukhalil et al., Nature Methods, 2015*

> 1K  
USERS

21,500  
PAGEVIEWS

10 MIN  
TIME SPENT

# Thanks

## Ginkgo Team

**Rob Aboukhalil**

Jude Kendall

Timour Baslan

Jim Hicks

Gurinder S. Atwal

Michael Wigler

Michael C. Schatz

[qb.cshl.edu/ginkgo](http://qb.cshl.edu/ginkgo)

