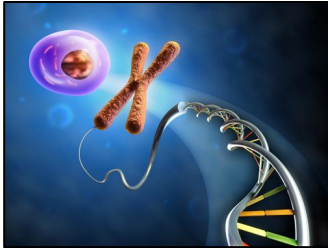


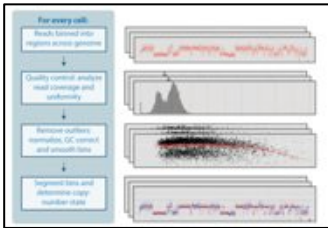
Single Cell Sequencing: Techniques, Analysis, and Visualization.

Tyler Garvin

Roadmap



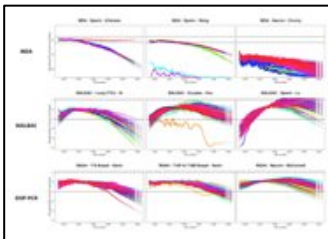
Introduction



Implementation



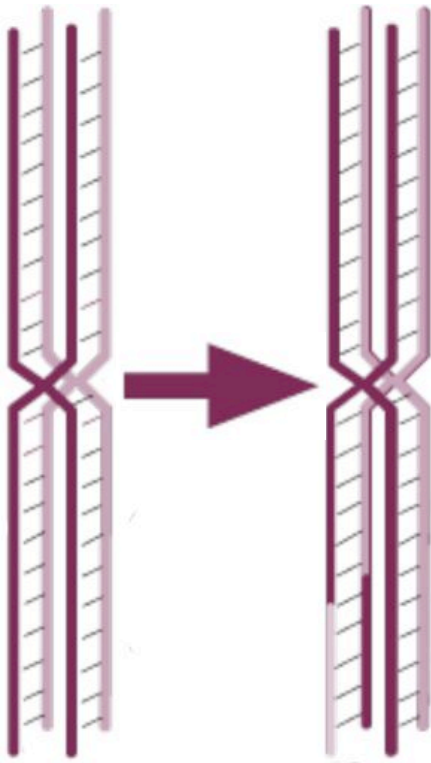
Live Demonstration



Validation and Recommendations

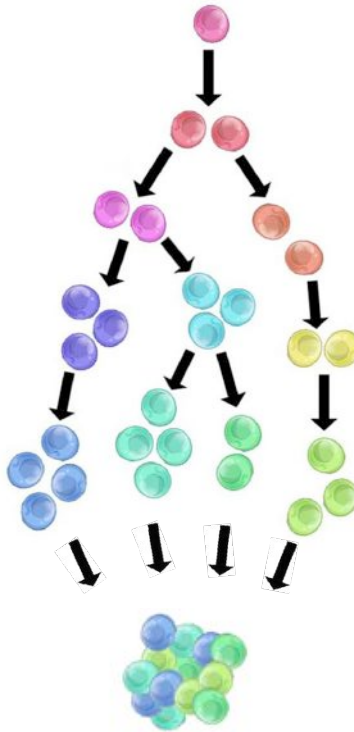
Tissue Targets

Germ Cells



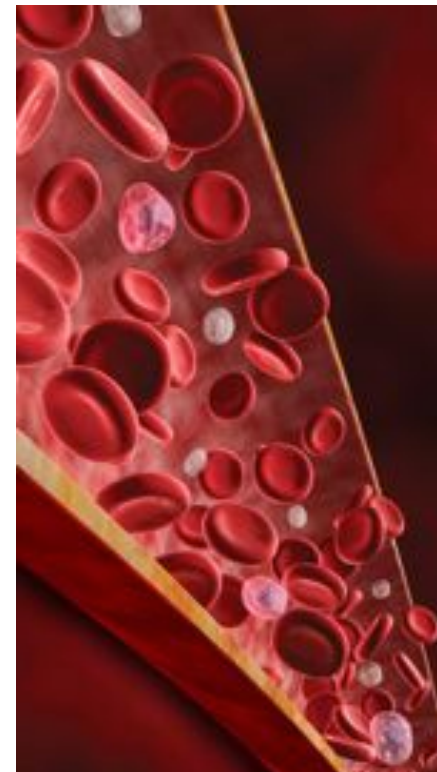
Recombination & crossover events

Heterogeneous Tumors



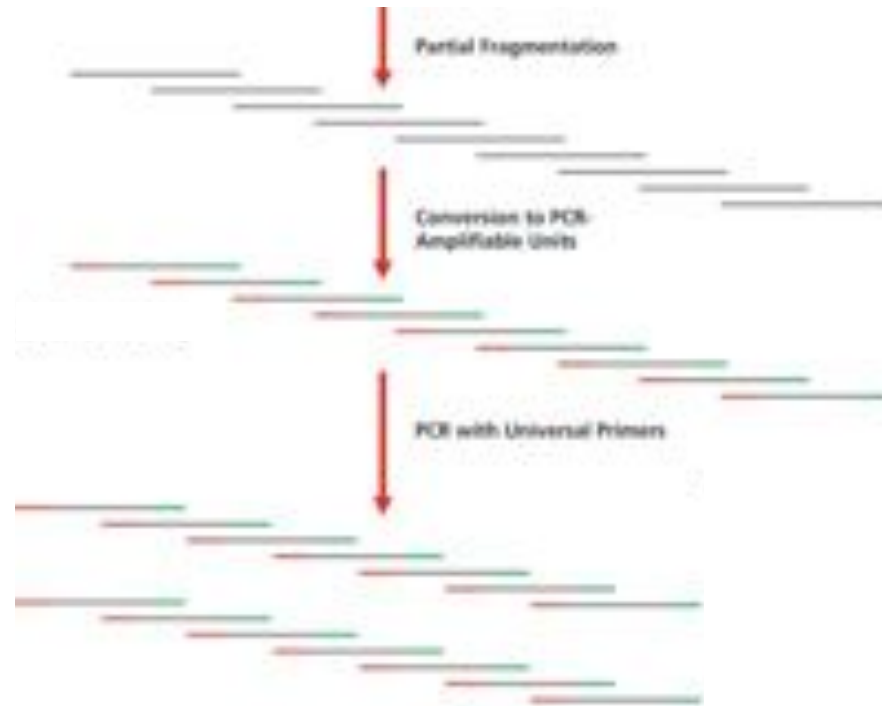
Clonal expansion

Heterogeneous Tissues:
blood and lymph



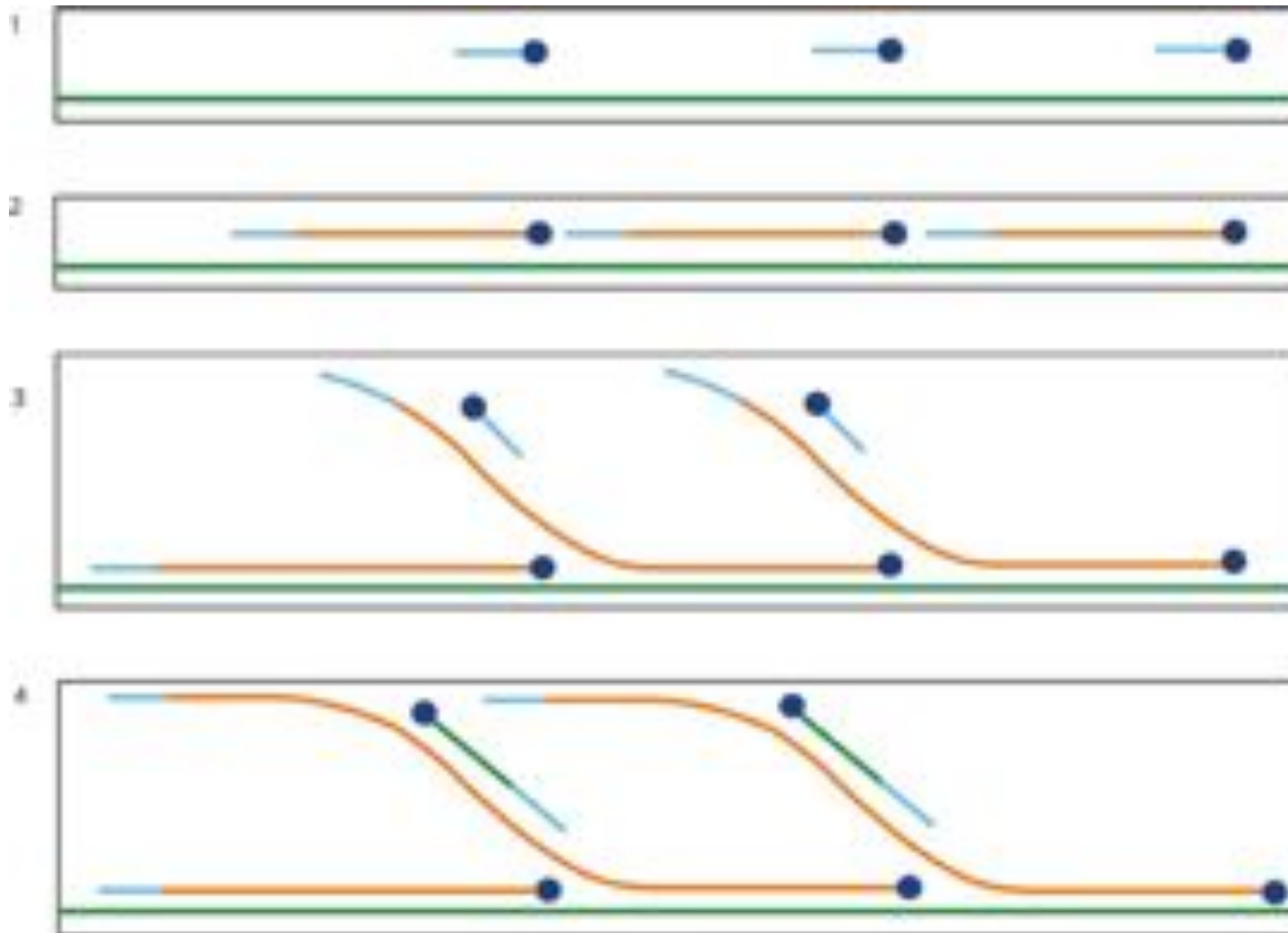
Isolating unique
cell types

Whole Genome Amplification

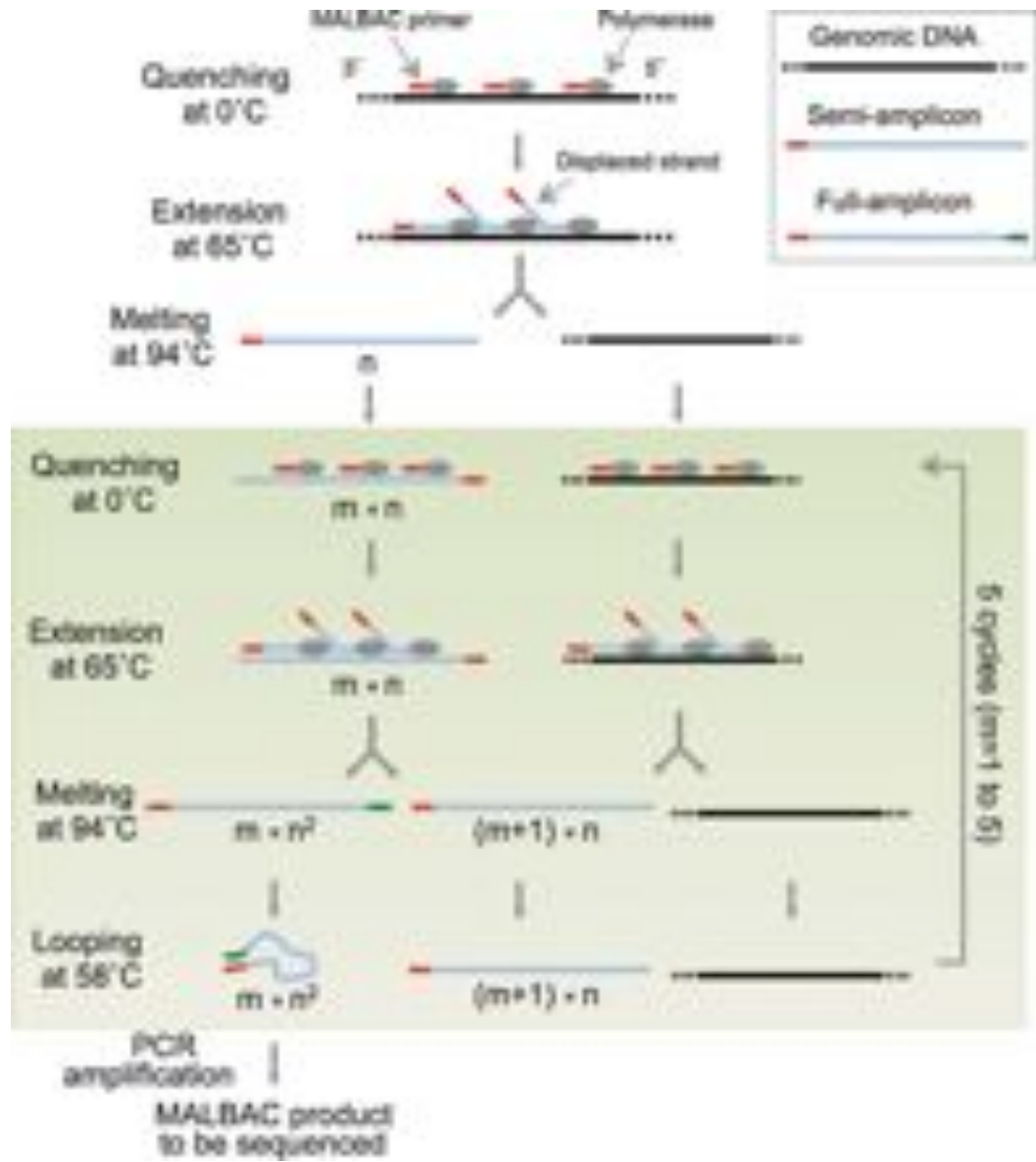


- 1) MDA: Multiple displacement amplification
- 2) DOP-PCR: Degenerate oligonucleotide-primed PCR
- 3) MALBAC: Multiple annealing and looping-based amplification

MDA



MALBAC



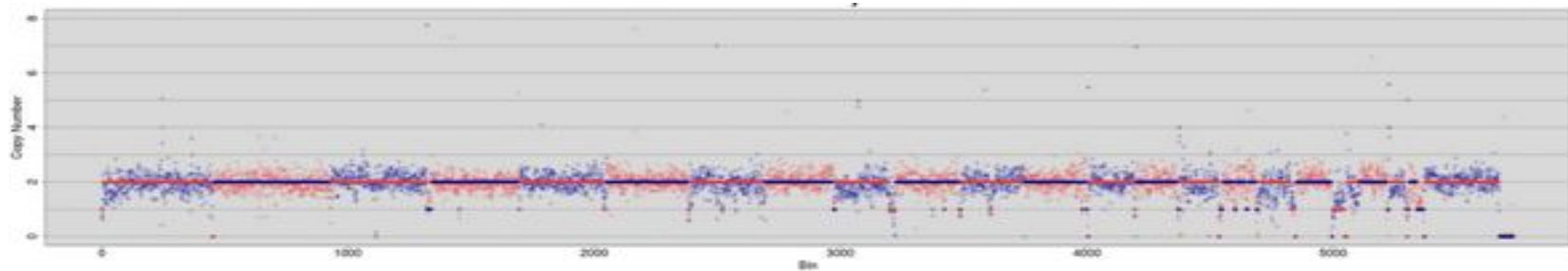
DOP-PCR

- Degenerate base pairing



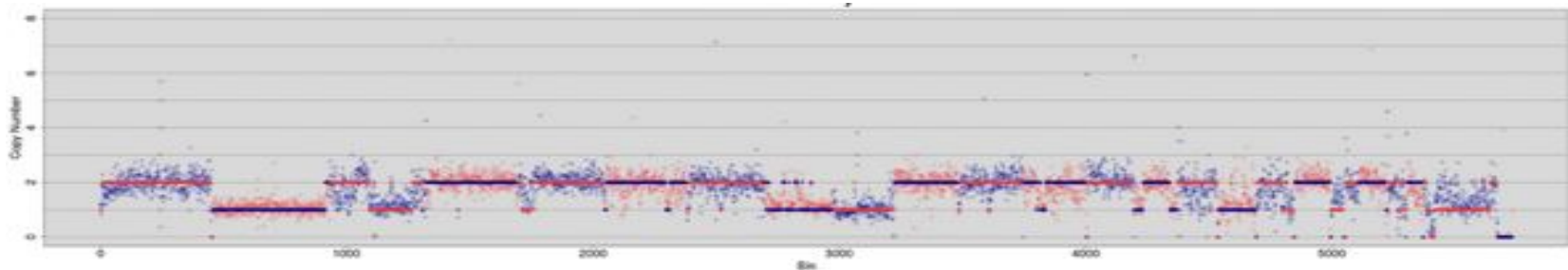
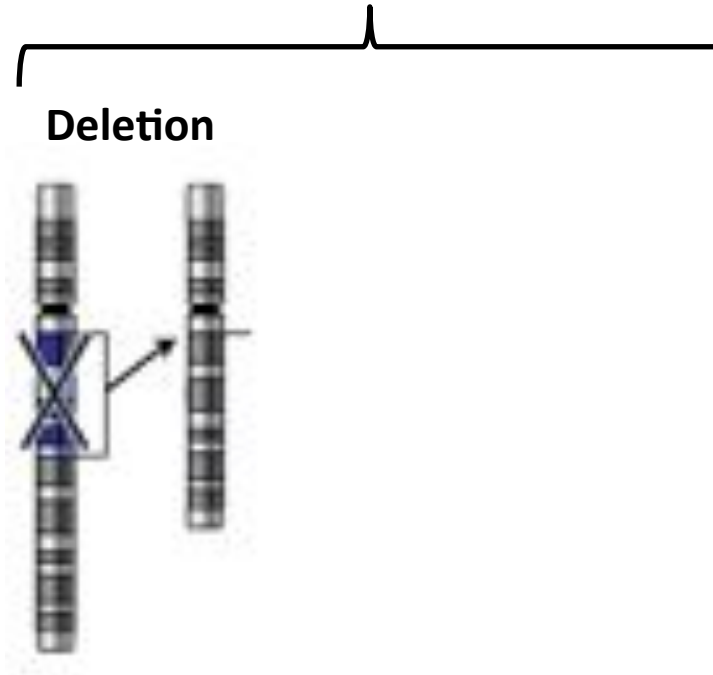
- Low initial annealing temperatures

Copy Number Variants



Copy Number Variants

Structural Variation

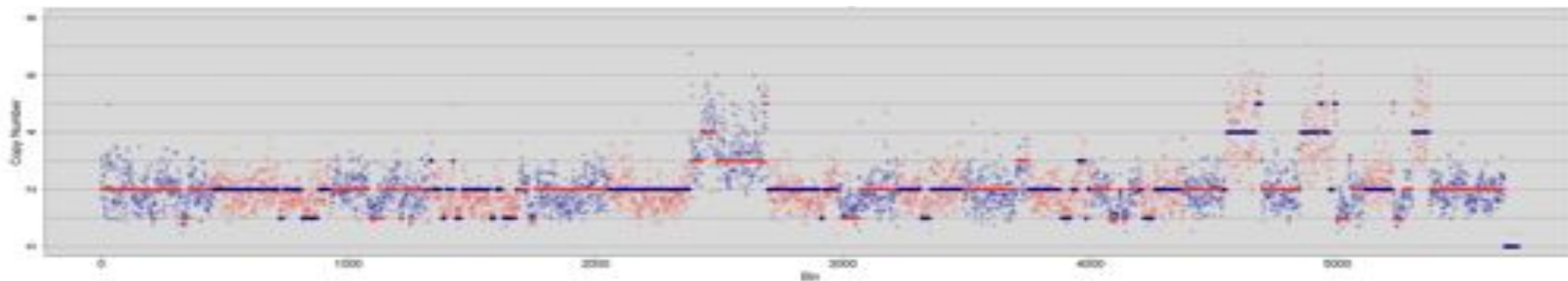


Copy Number Variants

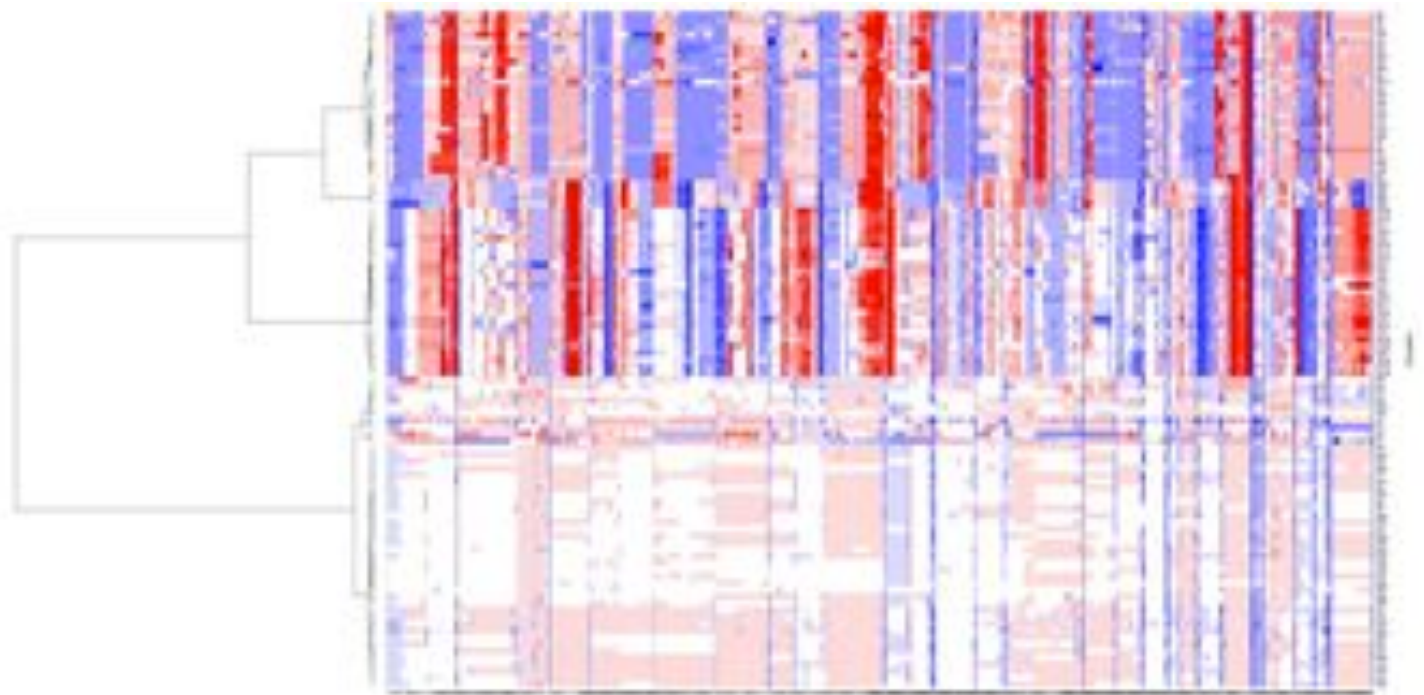
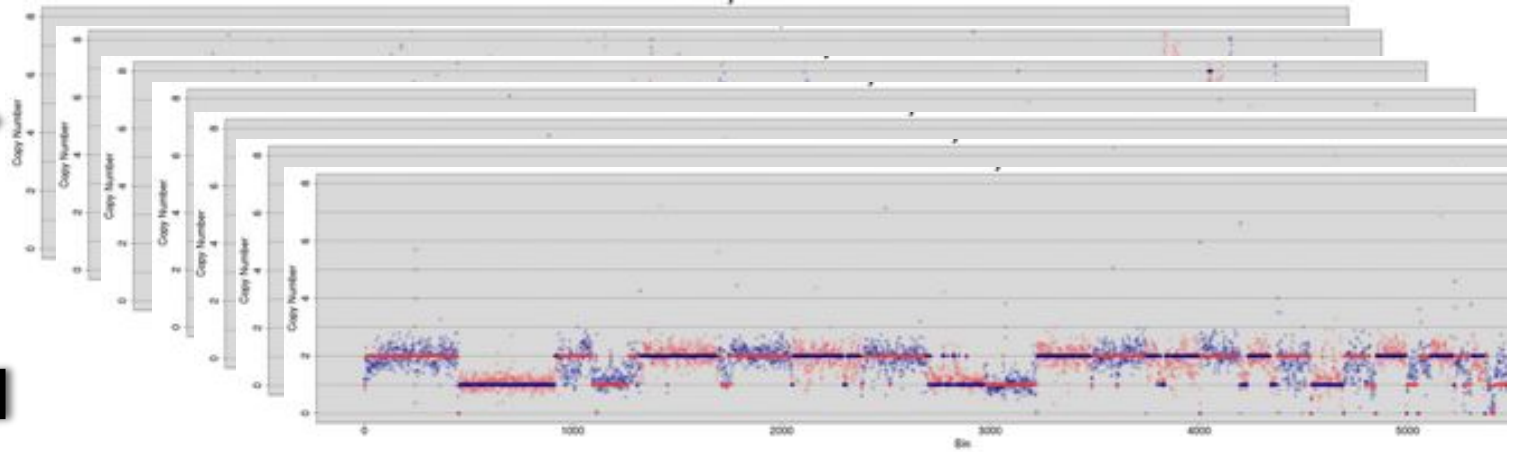
Structural Variation



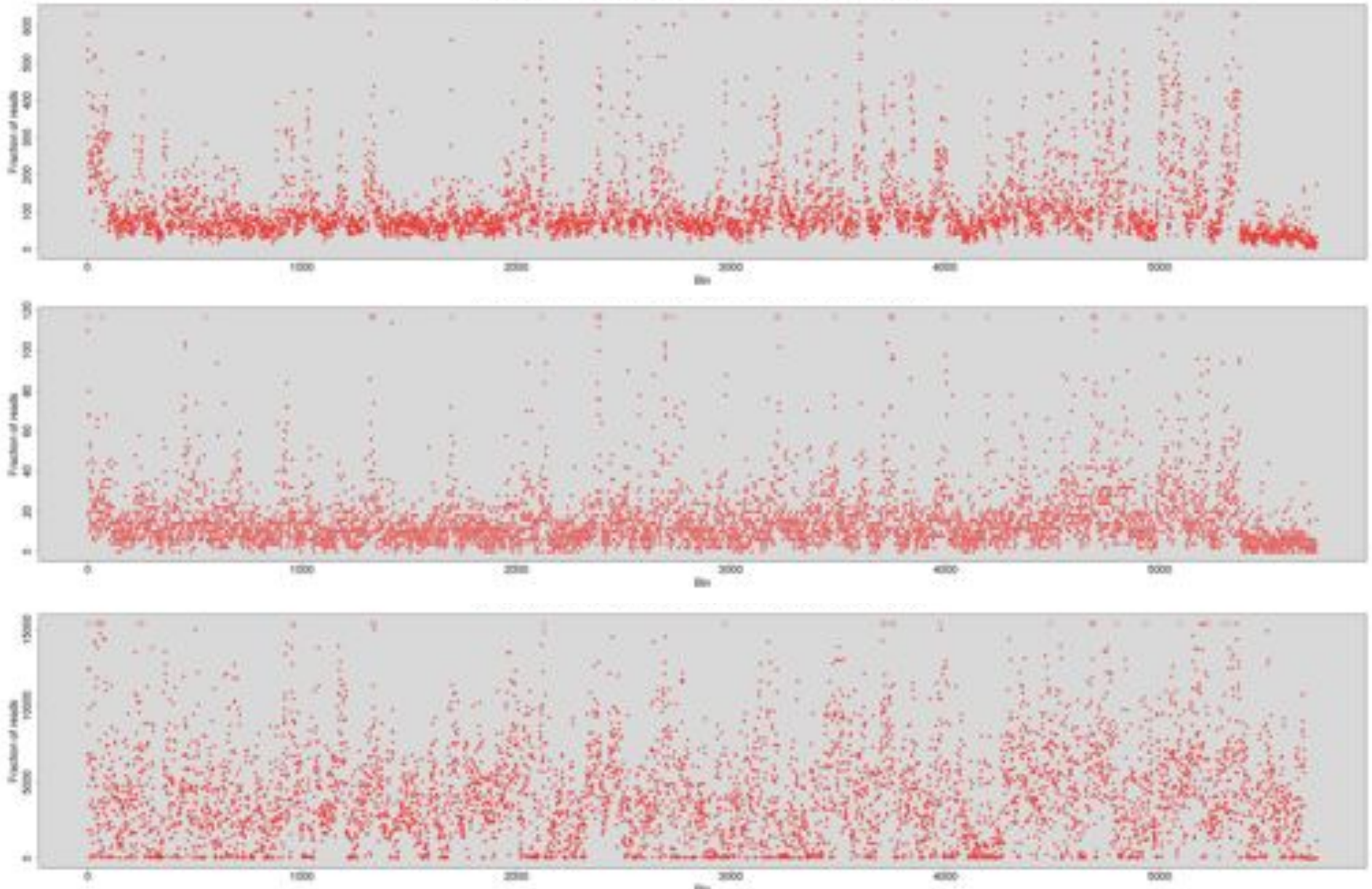
Amplification



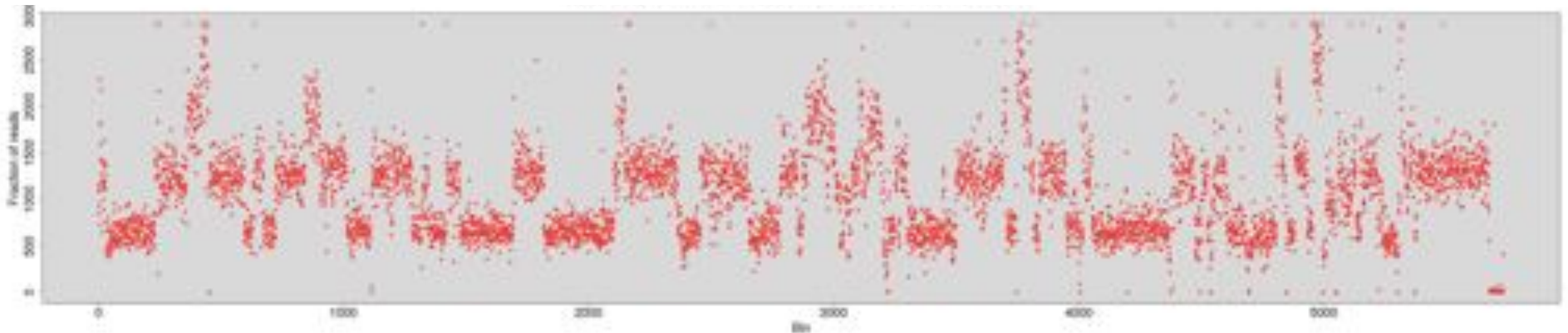
Underlying Concept



Data is noisy

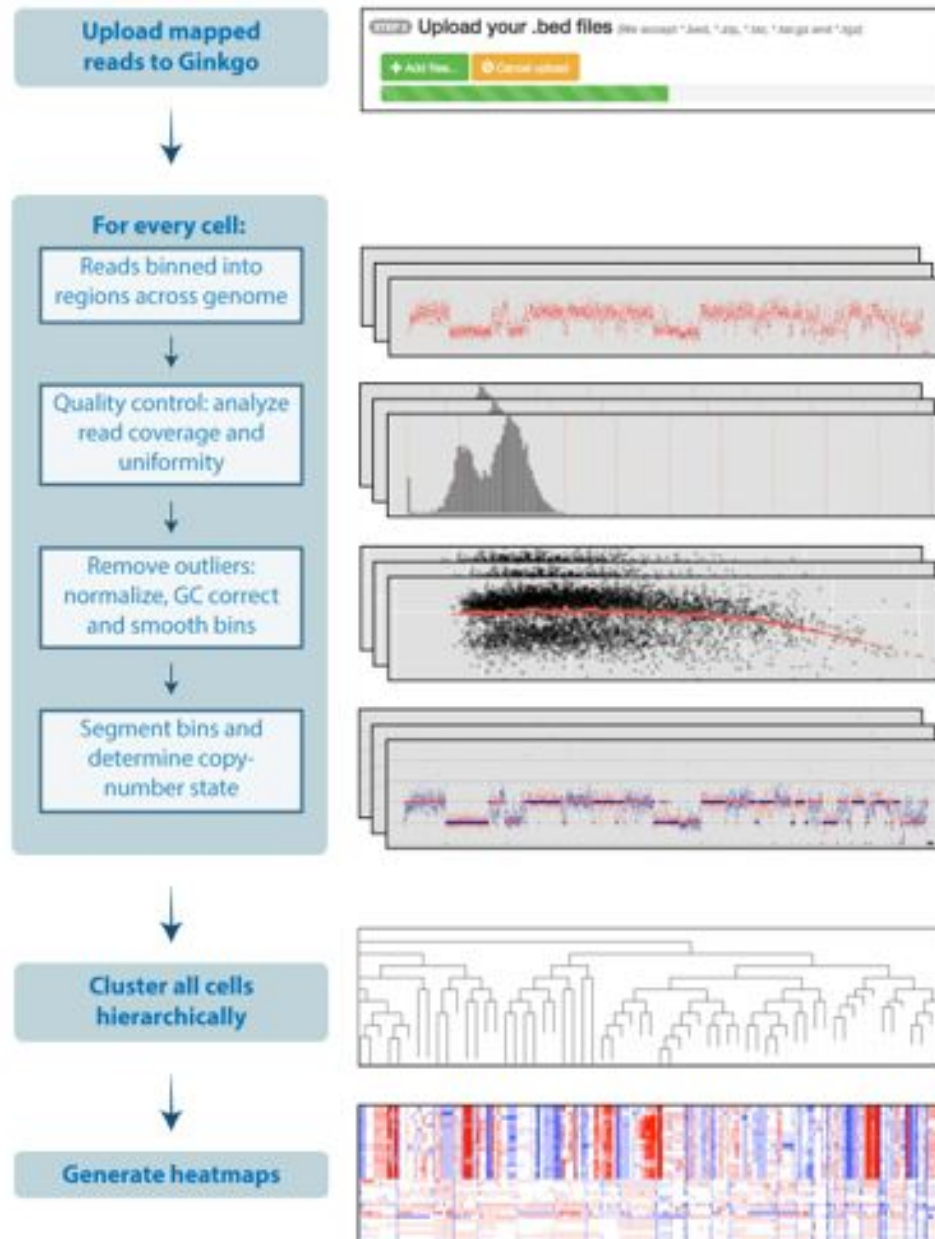


Data is noisy

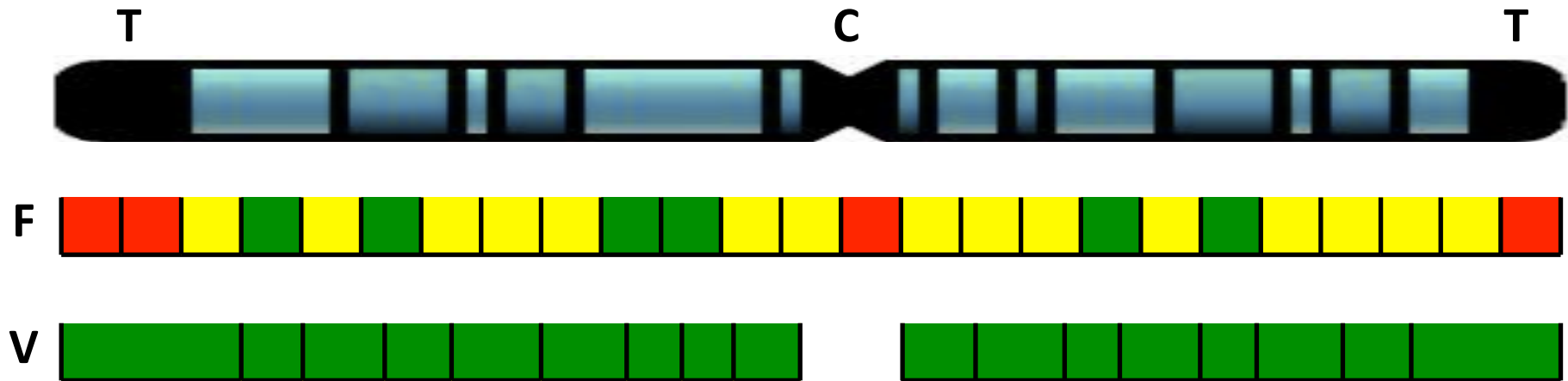


- Riddled with biases
 - **WGA**
 - Wet bench
 - Sequencing
 - Introduced through downstream analysis

Implementation

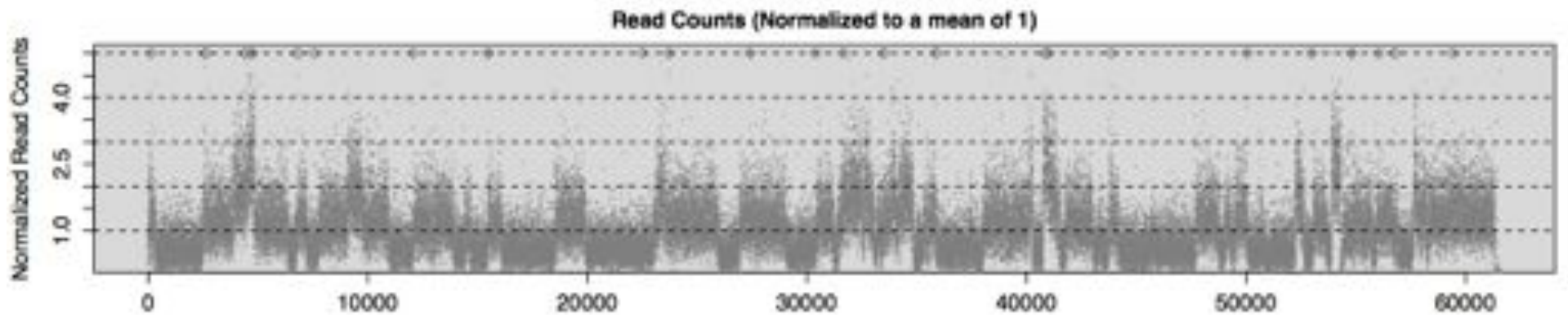
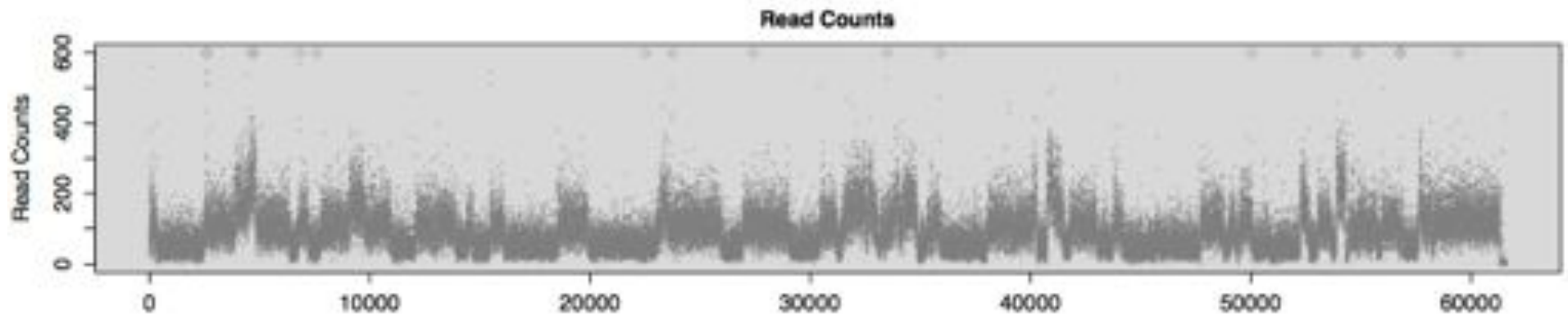


1) Binning: fixed vs. variable length bins

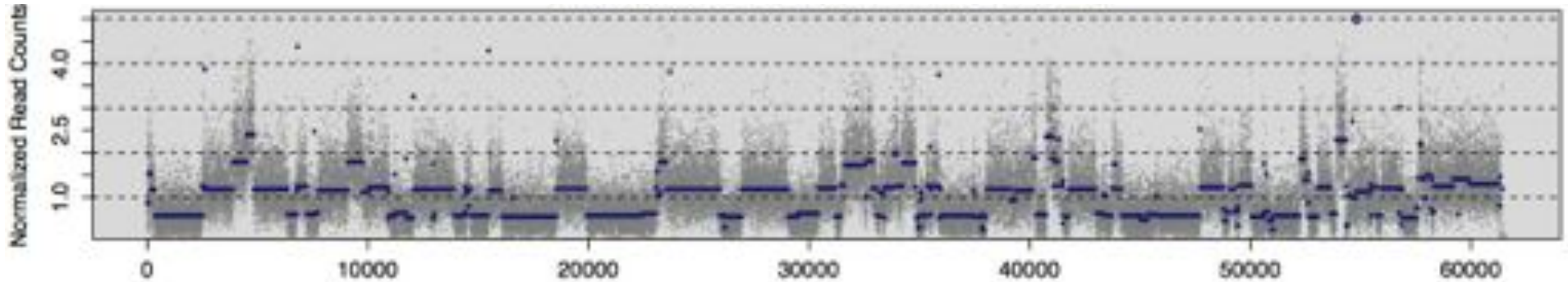


- Variable length bins
 - Extract read from each position across genome (3.2 billion in total)
 - Uniquely map reads to genome
 - Generate bins such that each bin shares the same number of uniquely mappable positions.

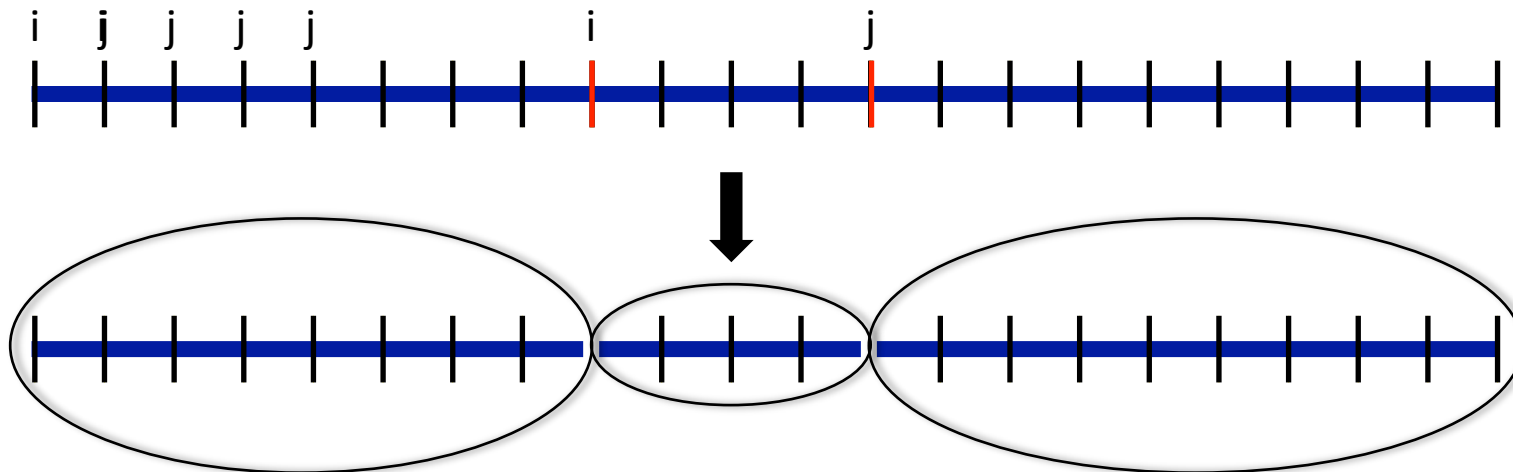
2) Normalization



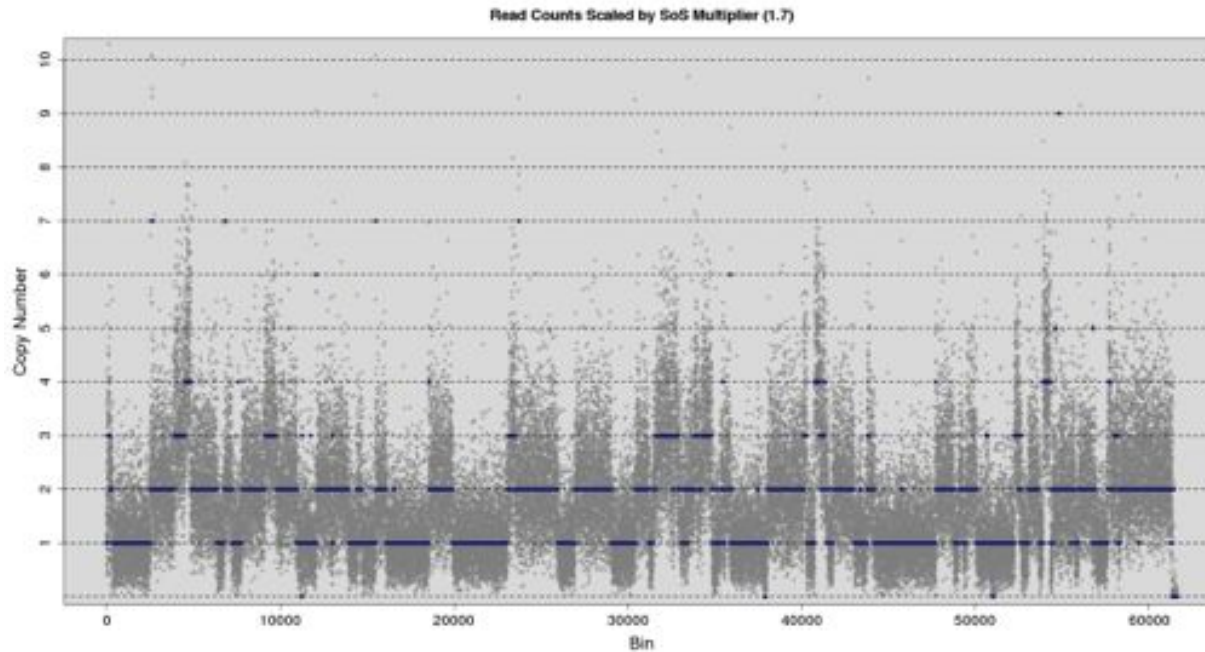
3) Segmentation



Circular Binary Segmentation (CBS)

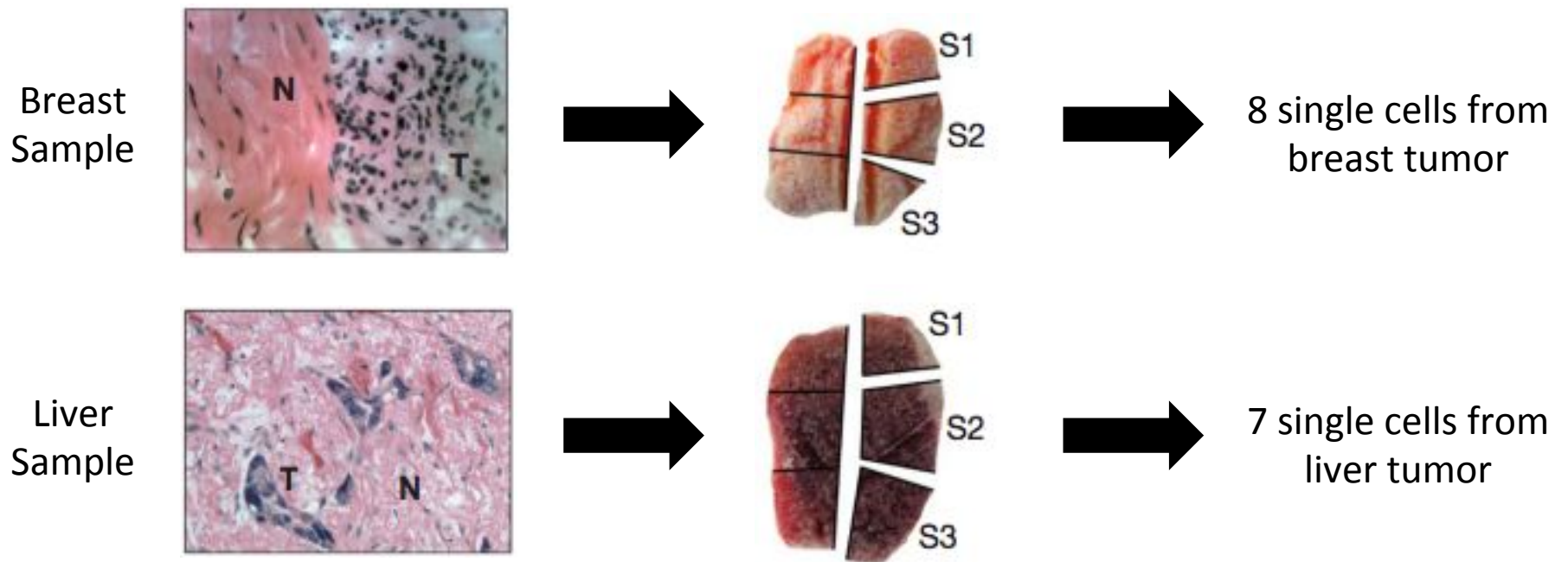


4) Estimating Copy Number



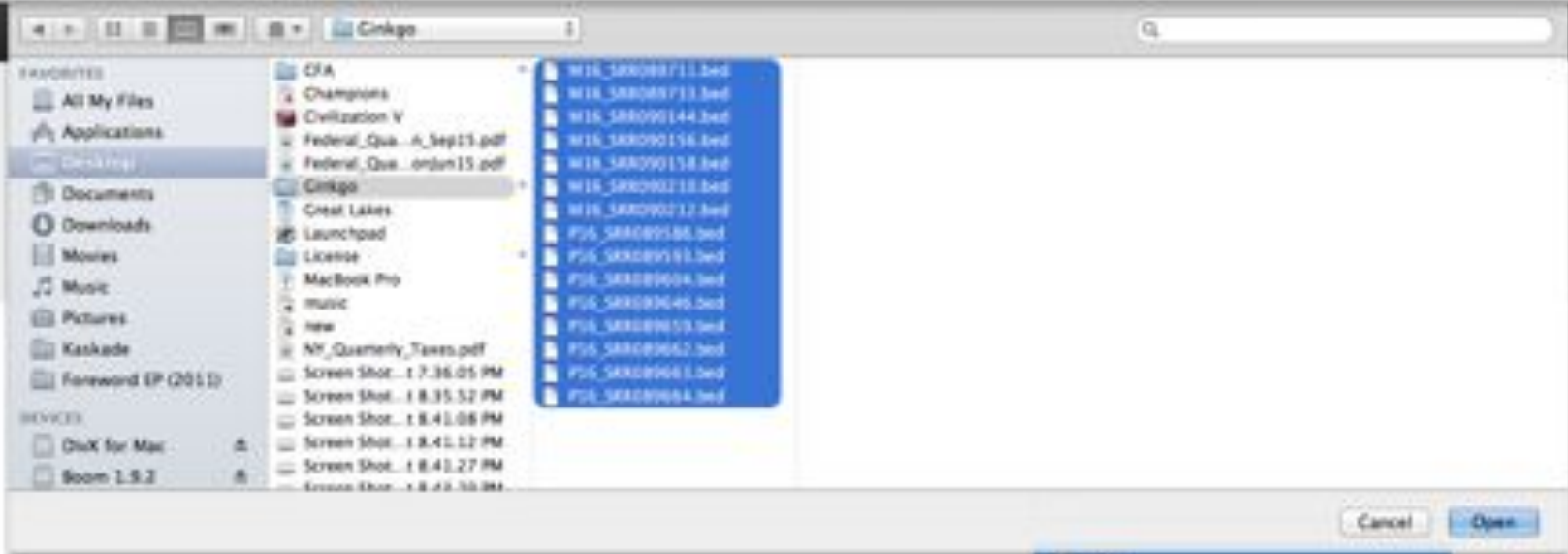
$$CN = \operatorname{argmin} \left\{ \sum_{i,j} (\hat{Y}_{i,j} - Y_{i,j}) \right\}$$

Demo: tumor metastasis



Tumor evolution inferred from single cell sequencing.
Navin, N. *et al.* (2011) *Nature*.

<http://qb.cshl.edu/ginkgo/>



Next step →

Help

How to make .bed files

Open a terminal and navigate to your data folder:

```
$ bowtie2 file - file.bed  
$ bowfish -i file.bed - file.bed
```

What a .bed file should look like

chrom	chromStart	chromEnd
chr1	555485	555523

Ginkgo

A web tool for analyzing single-cell sequencing data.

[Sample analysis](#) [Load previous analysis](#)

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

[Add files](#) [Cancel upload](#)

25.14 MB/s | 00:58:43 | 3.27 % | 300.43 MB / 9.17 GB

M16_SFF089711.bed	741.27 MB	<div style="width: 3.27%;"></div>	
M16_SFF089733.bed	843.27 MB	<div style="width: 3.27%;"></div>	
M16_SFF090144.bed	570.56 MB	<div style="width: 3.27%;"></div>	

[Next step](#)

View analysis later

Access your results later at the following address:

<http://gh.uwbl.edu/analysis/300031e14311f0a046460d0d70>

Help

How to make .bed files

Open a terminal and navigate to your data folder:

```
$ bowtie2 file - file.bed
$ bowfish -i file.bed - file.bed
```

What a .bed file should look like

Ginkgo

A web tool for analyzing single-cell sequencing data.

[Sample analysis](#) [Load previous analysis](#)

Upload your .bed files (We accept *.bed and *.bed.gz, max 100 files)

[Add files](#) [Cancel upload](#)

MHG_SRP089711.bed	741.27 MB	
MHG_SRP089733.bed	843.27 MB	
MHG_SRP089744.bed	570.56 MB	
MHG_SRP089755.bed	366.78 MB	
MHG_SRP089758.bed	309.42 MB	

[Next step](#)

View analysis later

Access your results later at the following address:

<http://gh.ucsf.edu/stage/ginkgo/161116080460bedfiles>

Help

How to make .bed files

Open a terminal and navigate to your data folder:

```
$ bowtie2 file - file.bed  
$ bowtie2 -i file.bed - file.bed
```

What a .bed file should look like

Chrome File Edit View History Bookmarks Window Help

MS 57% Sun Oct 5 8:41 PM Tyler Carvin

ginkgo - gavin.kyle@y... Google Calendar Ginkgo

gib.cshLeds/ginkgo/?q=dashboard/rvW311kxduAbZTndOv98a

Ginkgo - [UNOFT MAINTENANCE: Oct 4-6]

Ginkgo

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

1 Choose cells for analysis

Select all cells

M16_SRR090212.bed

P16_SRR090205.bed

P16_SRR090203.bed

2 Set analysis options

Job name: Single-cells from HeLa 3

Genome: Human hg19

3 Email notification

View analysis later

Access your results later at the following address:

<https://gib.cshLeds/ginkgo/ginkgo216/rvW311kxduAbZTndOv98a>

Chrome File Edit View History Bookmarks Window Help

MS 57% Sun Oct 5 8:41 PM Tyler Carvin

ginkgo - gavin.kyle@yale.edu Google Calendar Ginkgo

gib.cshLeda/ginkgo/?q=dashboard/rvW311kxduAbZTndOv9f8a

Ginkgo - [UNOFT MAINTENANCE: Oct 4-6]

Ginkgo

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

1 Choose cells for analysis

Select all cells

M16_SRR090212.bed

P16_SRR090205.bed

P16_SRR090203.bed

2 Set analysis options

Job name: Single-cells from HeLa 3

Genome: Human hg19

3 E-mail notification

View analysis later

Access your results later at the following address:

<https://gib.cshLeda.ginkgo.org/analysis/abZTndOv9f8a>

Chrome File Edit View History Bookmarks Window Help MS 57% Sun Oct 5 8:41 PM Tyler Carvin

gibbs - garvin.tyler@pm... Google Calendar Ginkgo

gb.cshLedsa/ginkgo/?q=dashboard/vw311kxduAbZTndOx9f8a

Ginkgo - [UNDER MAINTENANCE: Oct 4-8]

1/3 Choose cells for analysis

Select all cells

M16_SRR0990212.bed

P16_SRR0990986.bed

P16_SRR0990985.bed

2/3 Set analysis options

Job name: Ginkgo Demo

Genome: Human hg38

- Chimpanzee panTro4
- Mac mMacaca panTro
- S. manganesi chG
- D. rerio gmel10
- Other genome
- Human hg38
- Chimpanzee panTro4

3/3 E-mail notification

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

my@email.com

4/3 Advanced parameters

View analysis later

Access your results later at the following address:

<http://gb.cshLedsa/ginkgo/greenCell/vw311kxduAbZTndOx9f8a>

Chrome File Edit View History Bookmarks Window Help

MS [System Icons] Sun Oct 5 8:43 PM Tyler Carvin

gibbs - garvin.tyler@purdue.edu Google Calendar OnPage

gibbs.cs.hawaii.edu/ginkgo/?p=dashboard/vW31Y1kxdu4bZTndOo78o#parameters

OnPage [UNOIR MAINTENANCE: Oct 4-6]

Advanced parameters

Sample Parameters

CHW Profile Color Scheme Use color scheme.

General Binning Options Use a bin size of size.

Binning Simulation Options Bins based on simulations of bp reads, mapped with .

Segmentation Use method to segment.

Mask bad bins (optional)

Clustering Parameters


Clustering Use clustering.

Distance metric Use distance.

Include sex chromosomes?
(not recommended for mixed gender samples)

FACS File

FACS file



Chrome File Edit View History Bookmarks Window Help

MS [System Icons] Sun Oct 5 8:49 PM Tyler Carvin

ginkgo - gavin.tyler@y... Google Calendar Ginkgo

gib.cshL.edu/ginkgo/?q=results/MS3T7kxdsAb2TndOofSo

Ginkgo - [UNOFT MAINTENANCE: Oct 4-6]

Ginkgo

Ginkgo Demo

15% complete.

Step 1: 40% Mapping reads to bins... (P16 SFR080586.bed)

[View results](#)

[← Analysis Options](#)


[View analysis later](#)

Access your results later at the following address:

<https://gib.cshL.edu/ginkgo/gincvsc3sa/MS3T7kxdsAb2TndOofSo>

[Analysis Parameters](#)

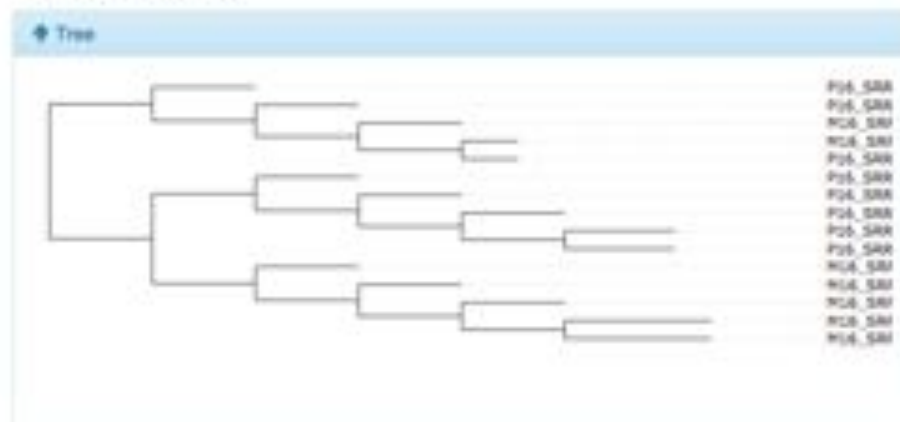
Binning: variable bins of 500bp size (using time-reso)
Segmentation: using normalized read counts
Clustering: single linkage, euclidean distance



Ginkgo Ginkgo Demo

Analysis complete!

View results



View analysis later

Access your results later at the following address:

<http://qb.cshLeda.edu/ginkgo/analysis/M3T7fxxdsAb25ndOof8o>

Analysis Parameters

Binning: variable bins of 500kb size [view more info](#)
Segmentation: using normalized read counts
Clustering: single linkage, euclidean distance

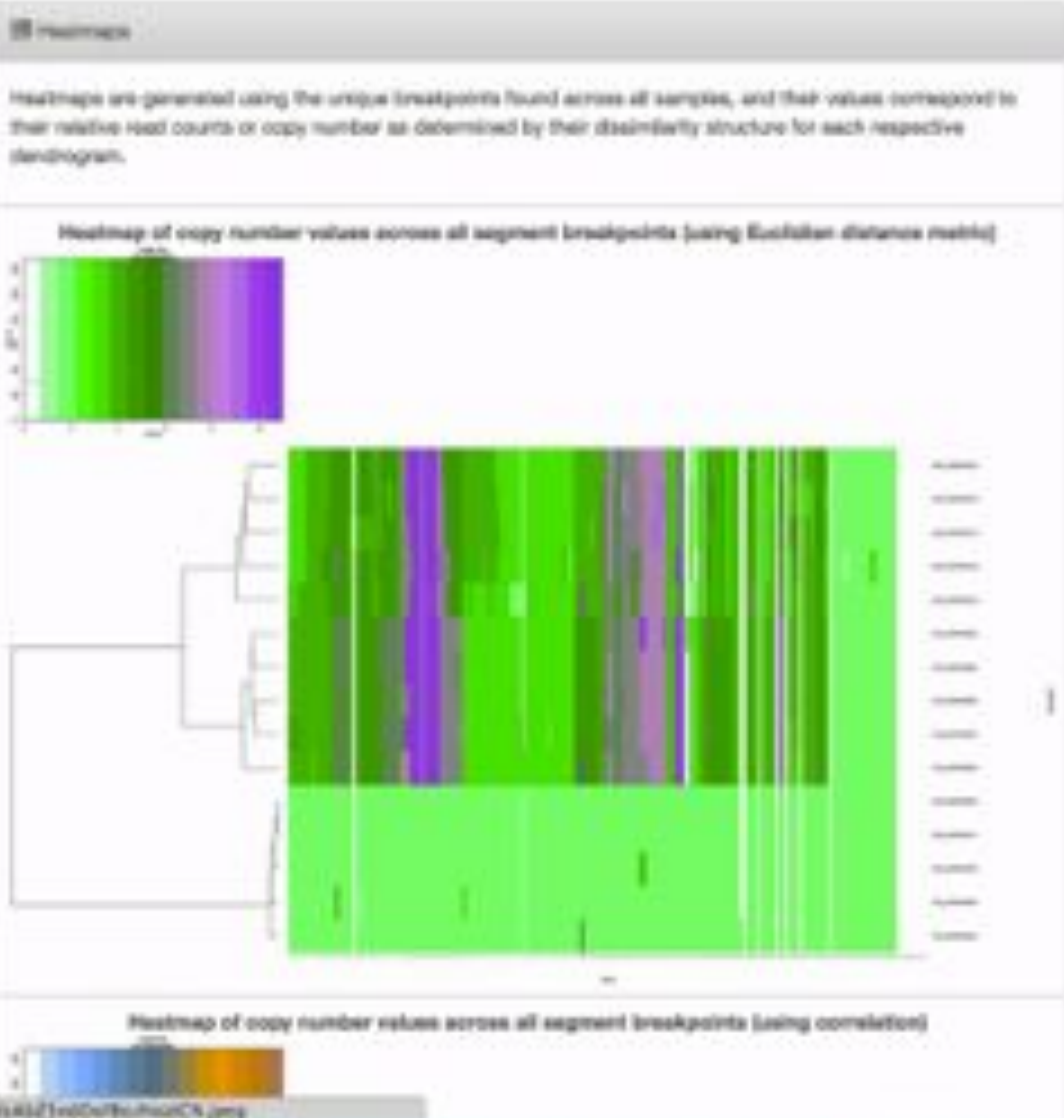
- ### Tree display
- Normalized read counts [view](#) | [pdf](#) | [png](#)
 - Copy number [view](#) | [pdf](#) | [png](#)
 - Correlations [view](#) | [pdf](#) | [png](#)

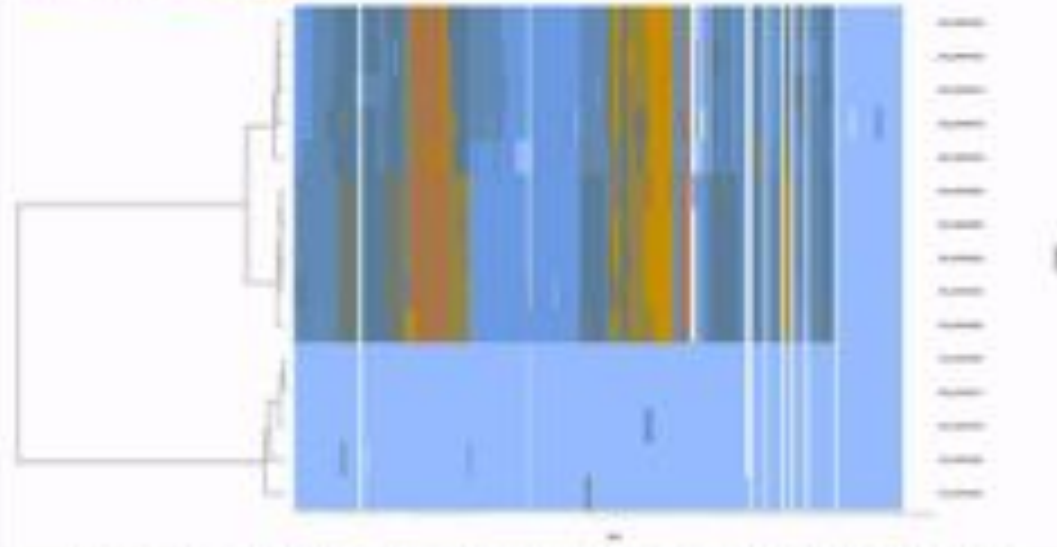
Download processed data

[logfiles](#)

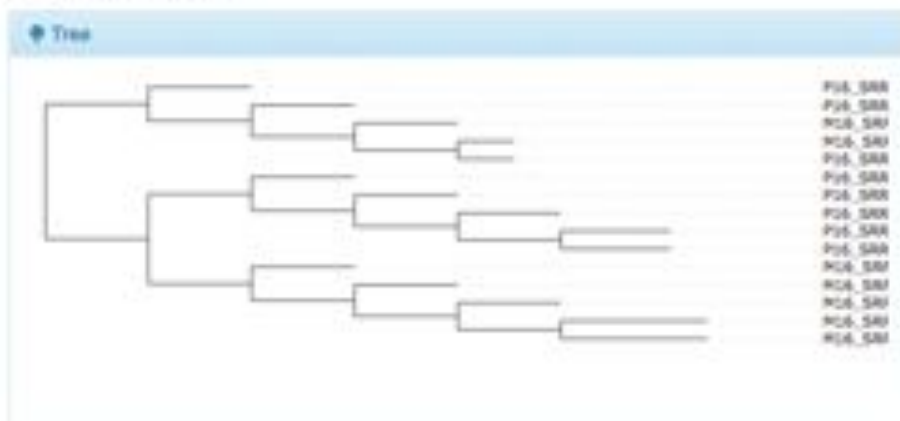
Summary

Cell	CNV Profile	# Reads	Mean read count	Read count variance	Index of dispersion
M16_SRR090711		18,211,447	2,670.78	984.12	0.37
M16_SRR090712		17,895,898	2,074.33	1,117.35	0.37





VIEW RESULTS



View analysis later

Access your results later at the following address:

<https://qb.csh.ledu/ginkgo/analysis/16337f7xdsAb25ndOofSo>

Analysis Parameters

Binning: variable bins of 500kb size (uniformly near)

Segmentation: using normalized read counts

Clustering: single linkage, euclidean distance

Tree display

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

Download processed data

- [Seg.html](#)
- [Seg.html](#)
- [SegCopy](#)
- [Seg.html](#)
- [Seg.html](#)

Summary

P16_S1628960		16,176,074	2,621.57	1,009.4	0.36
P16_S1628964		5,780,276	1,054.76	349.39	0.35
P16_S1628968		8,276,397	1,443.99	799.7	0.55
P16_S1628969		14,513,618	2,531.59	1,399.26	0.55
P16_S1628962		9,797,632	1,893.29	952.81	0.53

Download detailed quality assessment: [.txt](#)

LINKAGE -
 CLUST VIEW RESULTS



View analysis later

Access your results later at the following address:

gb.csh.edu/ginkgo/?q=results/MS377xdsAb25ndOofSo/P16_S000904

Binning: variable bins of 500kb size based on read
Segmentation: using normalized read counts
Clustering: single linkage, euclidean distance

Summary

Sample ID	Reads	Bin Size	Copy Number	Segmentation Score	Clustering Score
P16_S000900	18,178,074	2,621.57	1,009.4	0.38	
P16_S000904	5,780,278	1,054.78	349.38	0.38	
P16_S000905	8,278,387	1,443.99	799.7	0.59	
P16_S000909	14,513,618	2,531.59	1,398.26	0.59	
P16_S000902	9,797,632	1,893.29	952.81	0.53	

Tree display

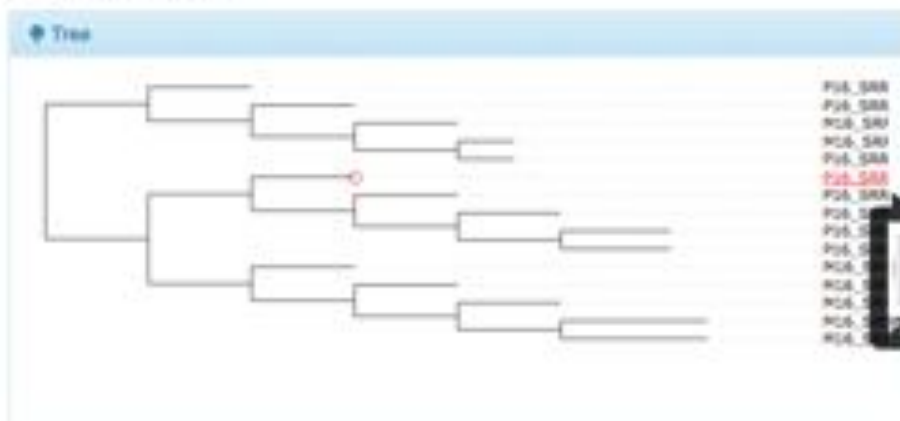
- Normalized read counts (html) (xml) (pdf) (png)
- Copy number (html) (xml) (pdf) (png)
- Correlations (html) (xml) (pdf) (png)

Download processed data

- Seg.html
- Seg.html
- SegCopy
- Seg.html
- Seg.html

Download detailed results: [results.html](#)

VIEW RESULTS



P16_S37
P16_S38
P16_S39
P16_S40
P16_S41
P16_S37029602
P16_S37029603
P16_S37029604
P16_S37029605
P16_S37029606
P16_S37029607
P16_S37029608
P16_S37029609
P16_S37029610
P16_S37029611
P16_S37029612

View analysis later

Access your results later at the following address:

<https://qb.csh.edu/ginkgo/analysis/MS377xdsAbZ5ndOofSo>



Summary

Sample ID	Reads	Length	GC	GC	GC
P16_S37029600	16,176,074	2,621.57	1,009.4	0.36	
P16_S37029604	5,780,276	1,054.76	349.39	0.36	
P16_S37029605	8,276,397	1,443.99	799.7	0.59	
P16_S37029609	14,513,618	2,531.59	1,399.26	0.59	
P16_S37029602	9,797,632	1,693.29	952.81	0.53	

- Tree display**
- Normalized read counts (html) (xml) (pdf) (png)
 - Copy number (html) (xml) (pdf) (png)
 - Correlations (html) (xml) (pdf) (png)

- Download processed data**
- SeqStats
 - SeqBreaks
 - SeqCopy
 - SeqNorm
 - SeqFlank

Download detailed results: [analysis/MS377xdsAbZ5ndOofSo/P16_S37029602](#)

Viewing cell P16_SRR089662

[← Back to tree](#)



Cytoscape

← Back to tree

Interactive Profile Viewer

View region in UCSC browser

P16_SRR089662

Bin 2009
Position: chr5:134503732-134891821
Copy Number: 3

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 "P16_SRR089662"

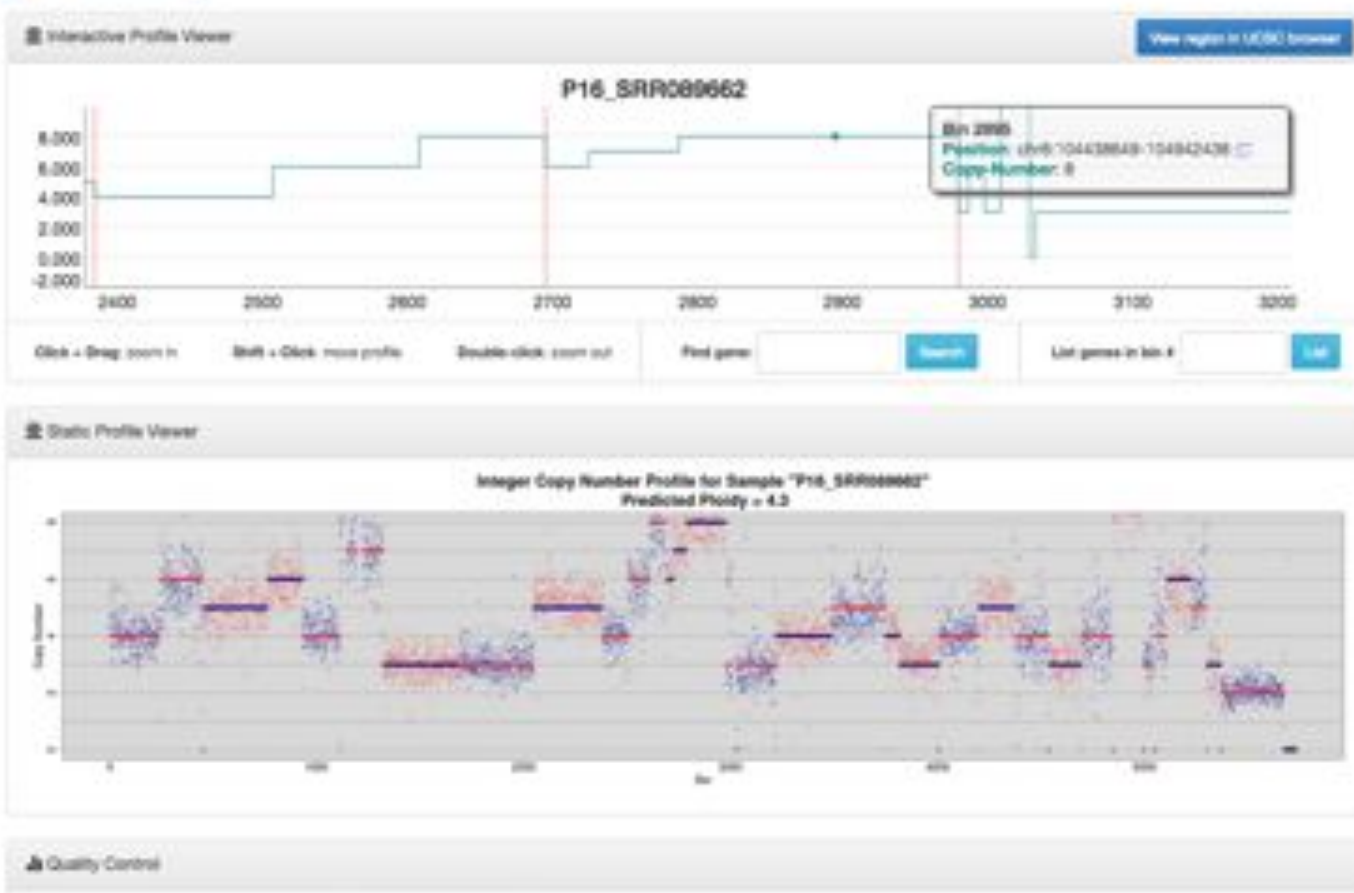
Predicted Ploidy = 4.2

Quality Control

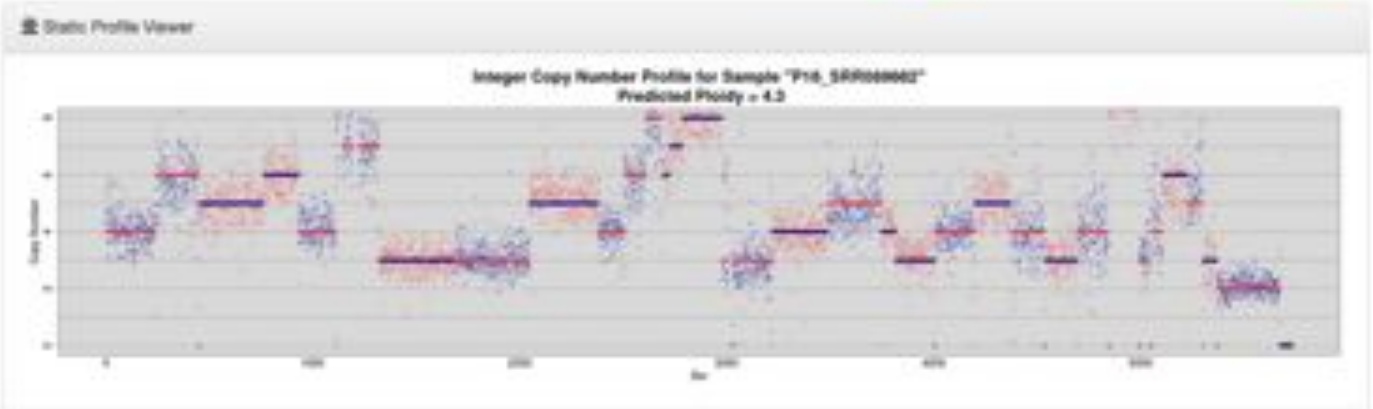
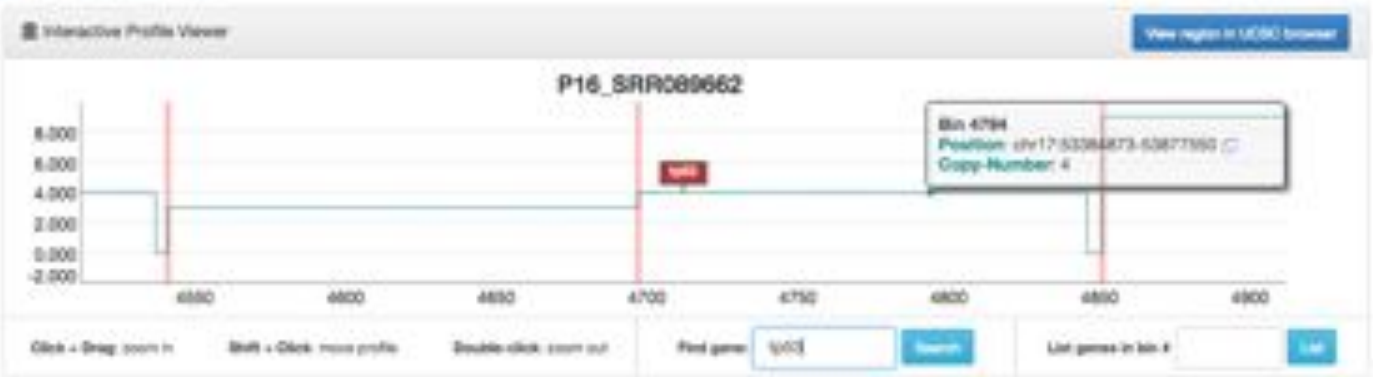


eb.cshLeda

← Back to tree

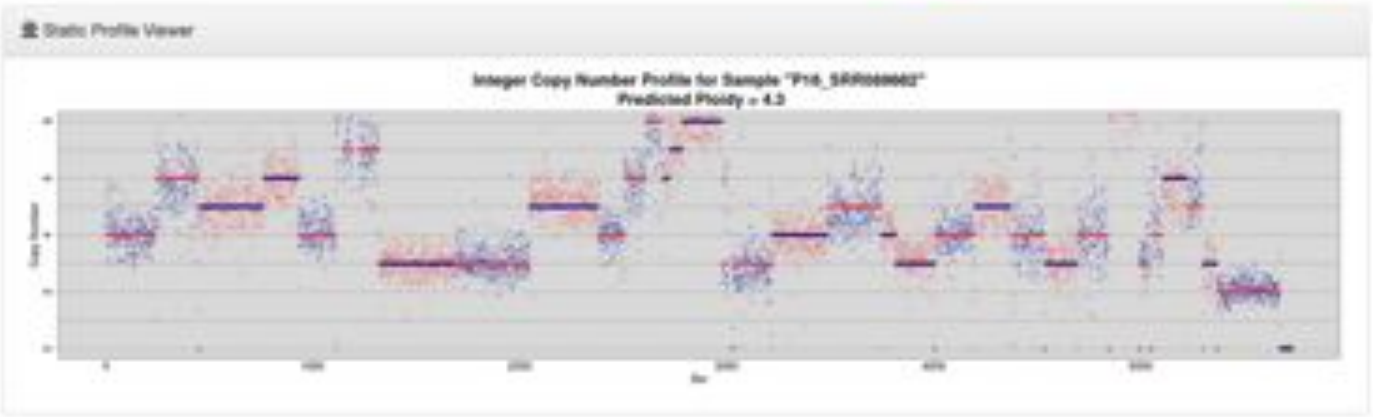


← Back to tree



Quality Control

◀ Back to tree

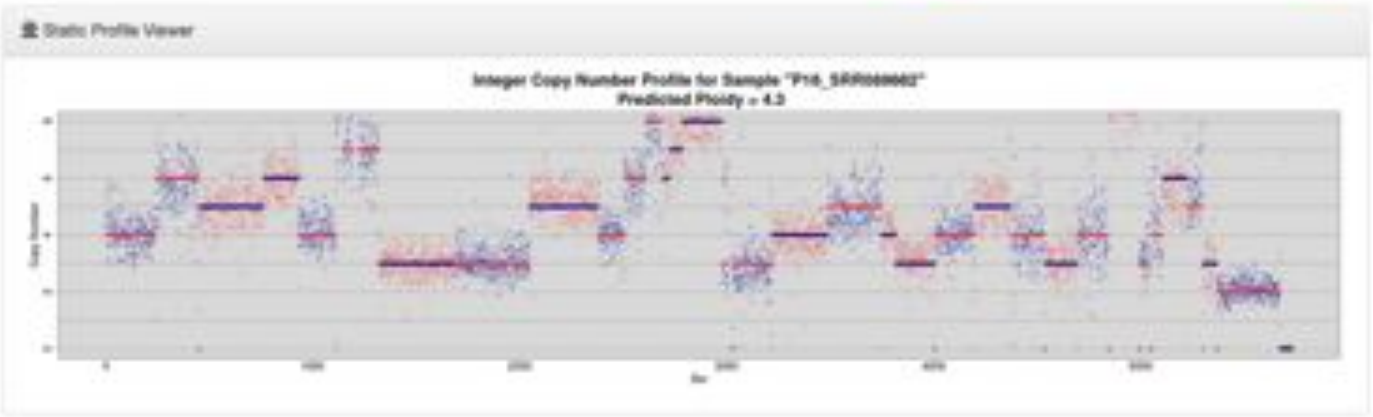


Quality Control



Cytoscape

← Back to tree



Quality Control



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

chr7:121,372,751-138,989,376 17,616,626 bp.

Probing the limits of single cell analysis

- Validation
- Exploring the effects of WGA on data quality
 - 1) GC biases
 - 2) Coverage dispersion
 - 3) Detecting integer copy number states
- Recommendations for single cell projects

Validating Ginkgo (5 independent studies)

nature Tumour evolution inferred by single-cell sequencing

Nicholas Navin^{1,2}, Jude Kendall¹, Jennifer Troge¹, Peter Andrews¹, Linda Rodgers¹, Jeanne McIndoo¹, Kerry Cook¹, Asya Stepansky¹, Dan Levy¹, Diane Esposito¹, Lakshmi Muthuswamy¹, Alex Krasnitz¹, W. Richard McCombie¹, James Hicks¹ & Michael Wigler¹

PNAS Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients

Xiaohui Ni^{a,b,1}, Minglei Zhuo^{c,1}, Zhe Su^{a,1}, Jianchun Duan^{c,1}, Yan Gao^{a,1}, Zhijie Wang^{c,1}, Chenghang Zong^{b,1,2}, Hua Bai^c, Alec R. Chapman^{b,d}, Jun Zhao^c, Liya Xu^a, Tongtong An^c, Qi Ma^a, Yuyan Wang^c, Meina Wu^c, Yu Sun^e, Shuhang Wang^c, Zhenxiang Li^c, Xiaodan Yang^c, Jun Yong^b, Xiao-Dong Su^a, Youyong Lu^f, Fan Bai^{a,3}, X. Sunney Xie^{a,b,3}, and Jie Wang^{c,3}



Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing

Sijia Lu *et al.*
Science **338**, 1627 (2012);
DOI: 10.1126/science.1229112



Mosaic Copy Number Variation in Human Neurons

Michael J. McConnell *et al.*
Science **342**, 632 (2013);
DOI: 10.1126/science.1243472

Cell Genome Analyses of Single Human Oocytes

Yu Hou,^{1,6} Wei Fan,^{1,4,6} Liying Yan,^{1,6} Rong Li,¹ Ying Lian,¹ Jin Huang,¹ Jinsen Li,¹ Liya Xu,¹ Fuchou Tang,^{1,5,*} X. Sunney Xie,^{1,2,*} and Jie Qiao^{1,3,*}

Home
Github

- Polygenomic breast tumor -- Naim et al, 2011
- Breast cancer + liver metastasis -- Naim et al, 2011
- Neurons -- McConnell et al, 2013
- Circulating lung tumor cells -- Ai et al, 2013
- Oocytes -- Hou et al, 2013
- Sperm -- Lu et al, 2012
- Sperm -- Kirkness et al, 2013
- Sperm -- Wang et al, 2012
- Neurons -- Florjanczyk et al, 2012

Quail
Thesis - Prostate gtl2
4T1_VQ
MDA - Sperm - Wang et al.
Sperm
Bone marrow - Hou et al.
Kidney - Lu et al.
MFI4-31 -61 48bp bins 100kb bins
PIET_and_P15M
Oocyte - Hou et al.
Ginkgo Demo

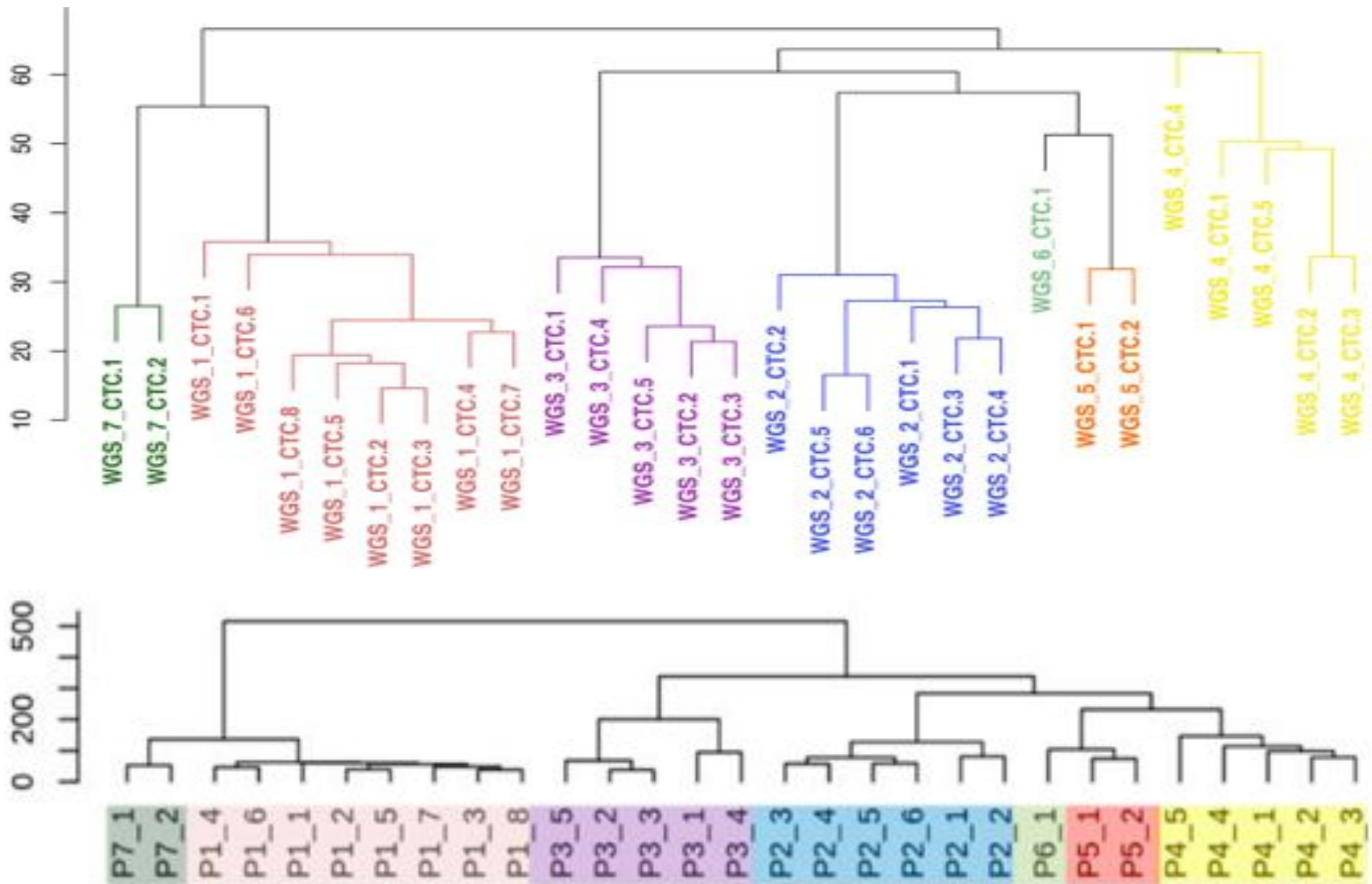
R6
R3
R1
D1
R6
D04
D17
D7
D22
D34
D33
D2
D08
R2
D19
D20
D27
D3
D37
D39
D9
D30
D31
D1
D16
R7
D38
D8
D12
D4
D23
D14
D32
D29
D24

View analysis later
Access your results later at the following address:
https://gb.cshl.edu/ginkgo/?q=results/_x10breast_navin

Analysis Parameters
Binning: variable bins of 500kb size [view info](#) [view](#)
Segmentation: using normalized read counts
Clustering: ward linkage, euclidean distance

Tree display
Normalized read counts [\[view\]](#) [\[url\]](#) [\[png\]](#)
Copy number [\[view\]](#) [\[url\]](#) [\[png\]](#)
Correlations [\[view\]](#) [\[url\]](#) [\[png\]](#)

Download processed data
[Sign Out](#)



Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients.
 Ni, Xiaohui *et al.* (2013) *PNAS*.

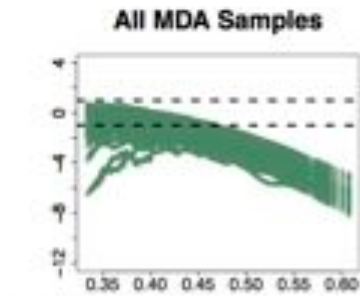
Comparing Single Cell Sequencing Techniques

Study	WGA Method	Disease State	Tissue Type	Accession
Kirkness et al.	MDA	None	Sperm	SRP017516
Wang et al.	MDA	None	Sperm	SRA053375
Evrony et al.	MDA	None	Neuron	SRA056303
Lu et al.	MALBAC	None	Sperm	SRA060945
Ni et al.	MALBAC	Cancer	Lung	SRP029757
Hou et al.	MALBAC	None	Oocyte	SRA091188
Navin et al.	DOP-PCR	Cancer	Breast (T10)	SRX021401
Navin et al.	DOP-PCR	Cancer	Breast (T16P/M)	SRX037035/132
McConnell et al.	DOP-PCR	None	Neuron	SRP030642

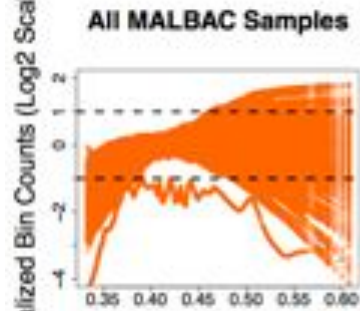
**% Bins w/
Bounded
GC Bias**

GC Biases

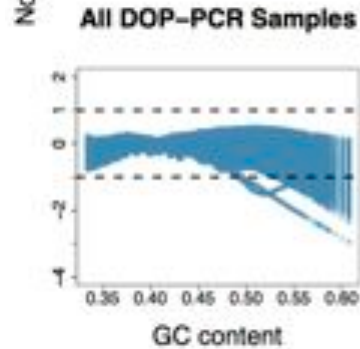
45.9%



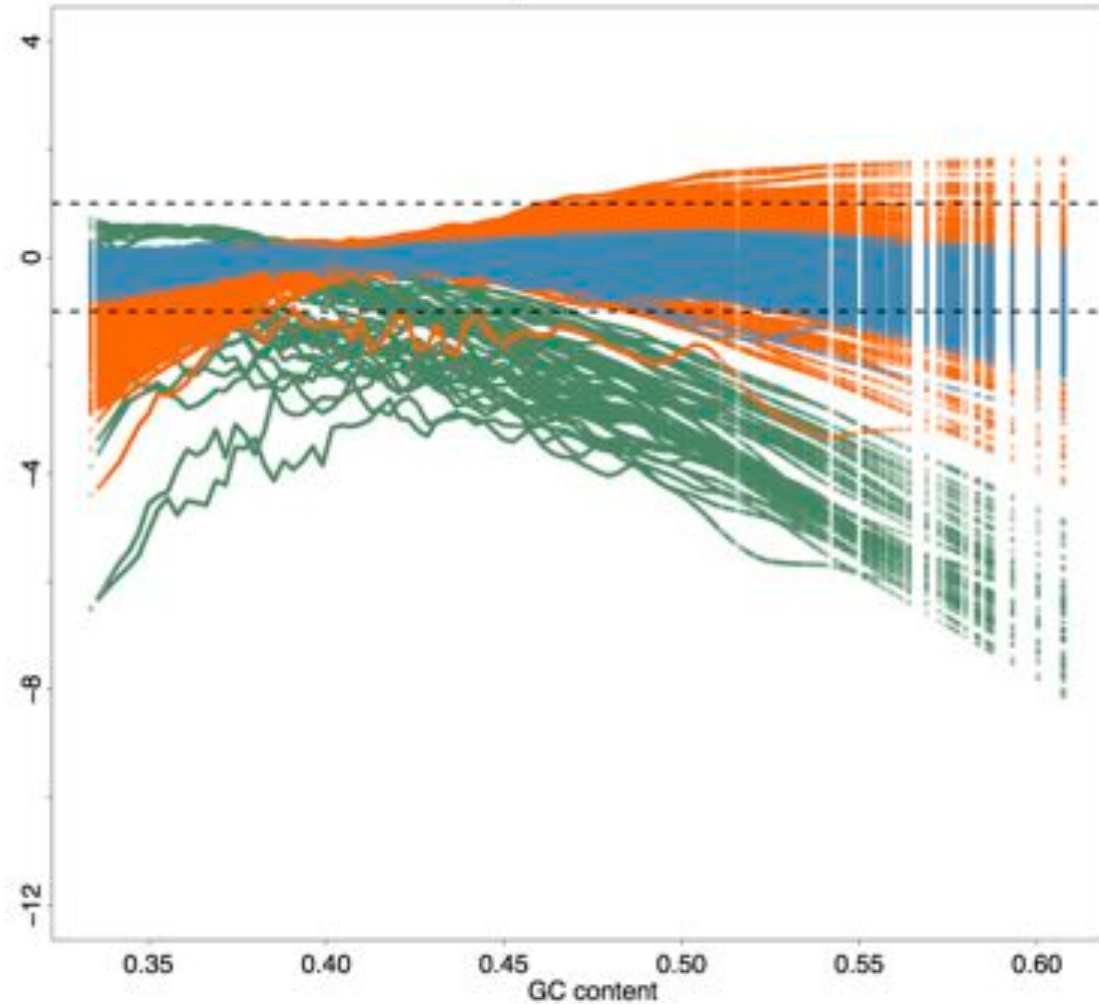
94.0%



99.6%

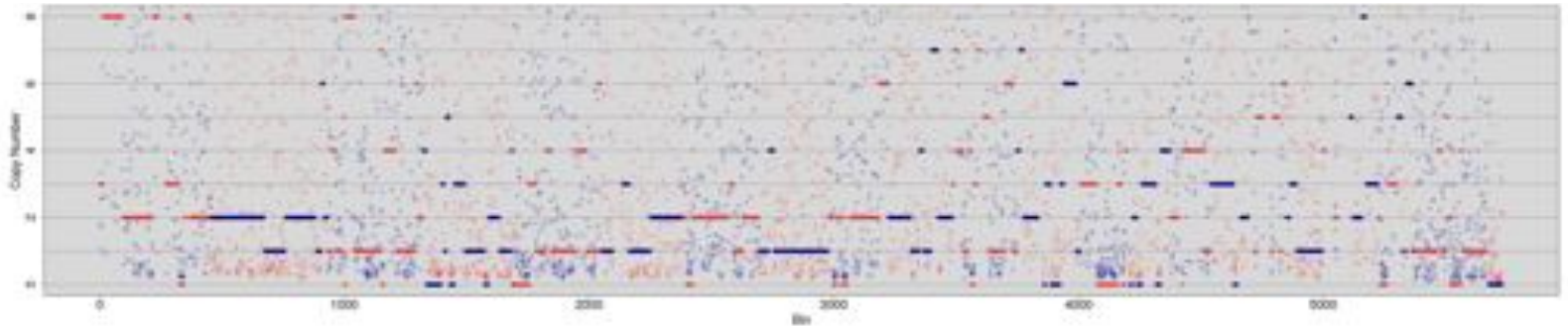


Overlay of All Datasets

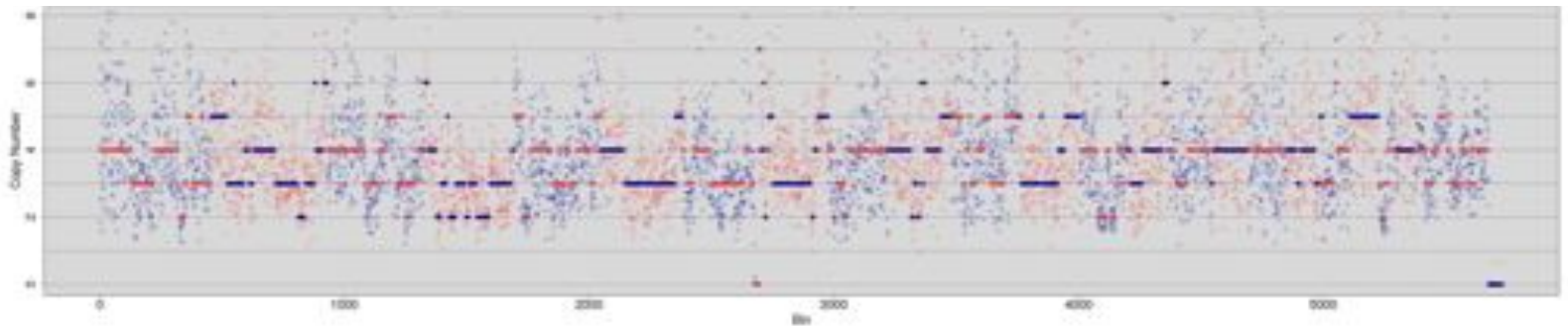


Coverage Dispersion

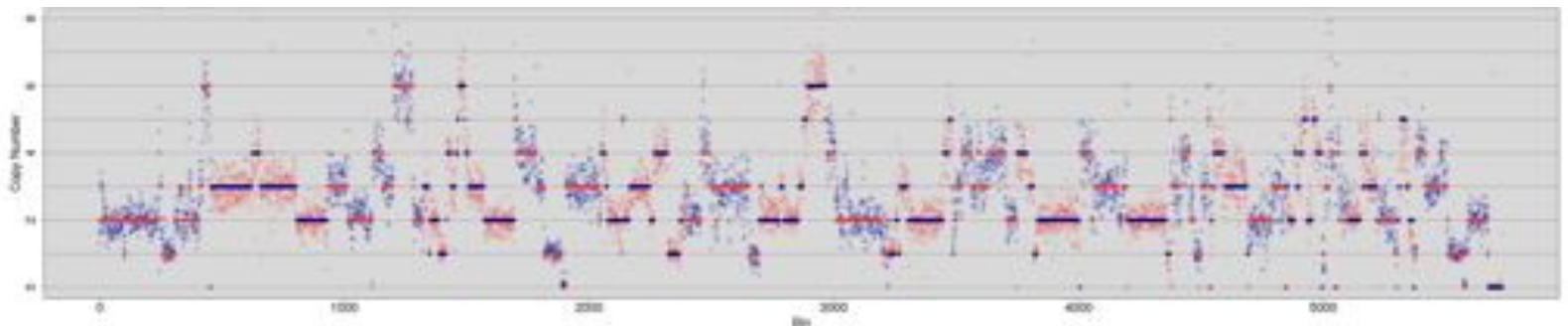
MDA



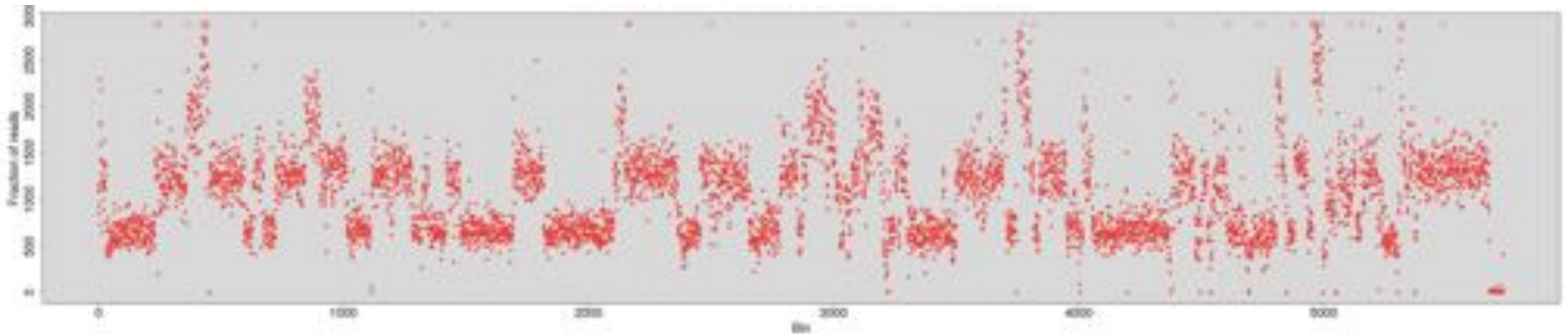
MALBAC



DOP-PCR



Measuring Coverage Dispersion

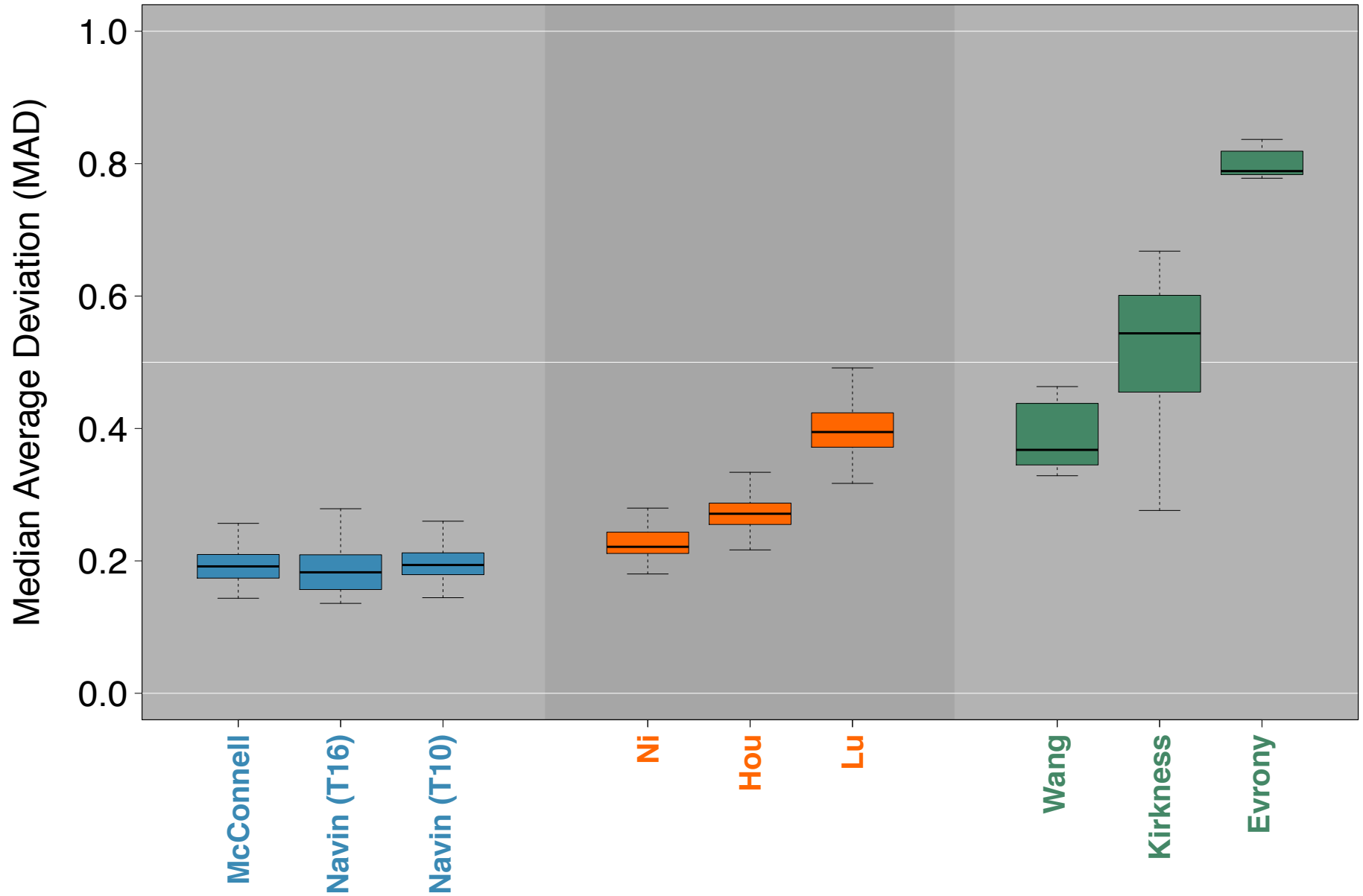


Median absolute deviation (MAD) – median of the absolute deviation from the data's median. Measure of the statistical dispersion.

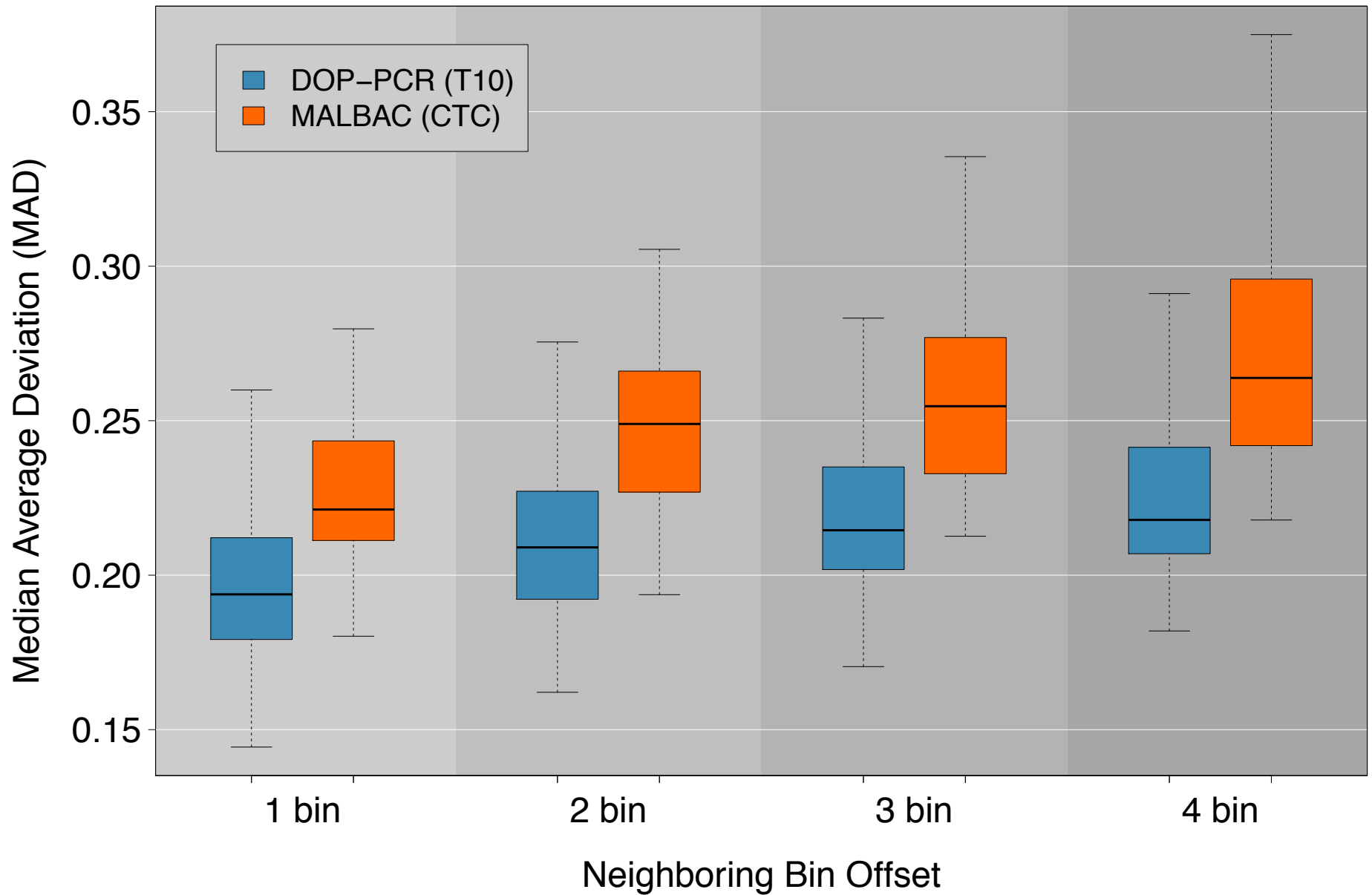
Calculate all pairwise differences between neighboring bins (d)

$$\text{MAD} = \text{median} (| d_i - \text{median}(d) |)$$

Coverage Dispersion

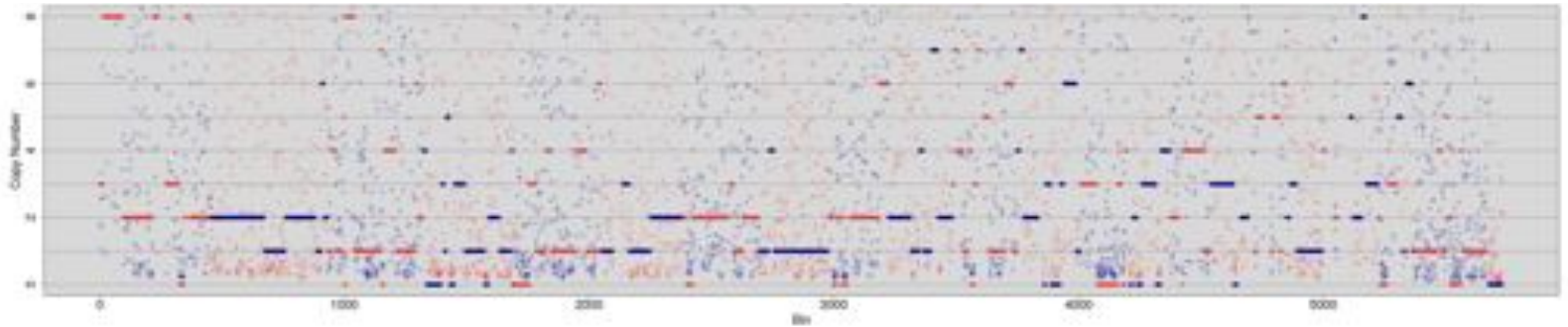


Coverage Dispersion

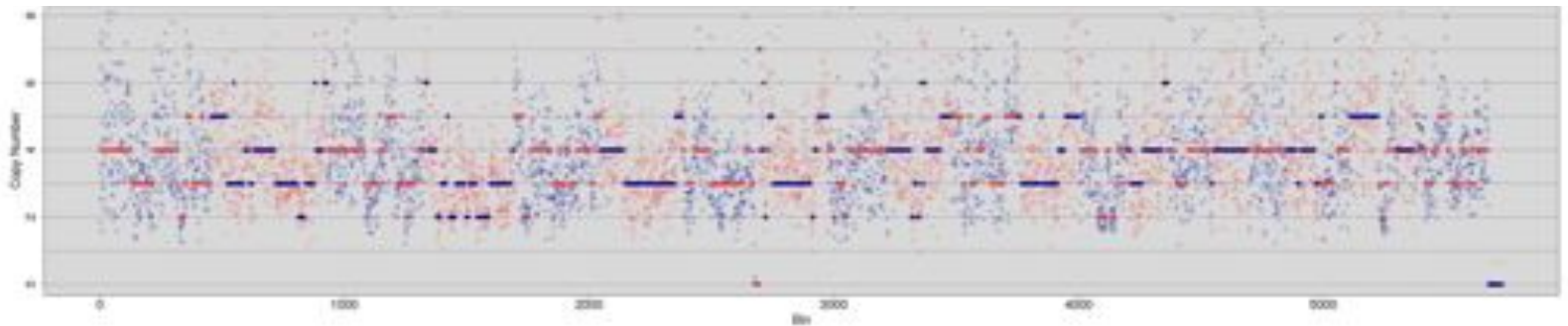


Coverage Dispersion

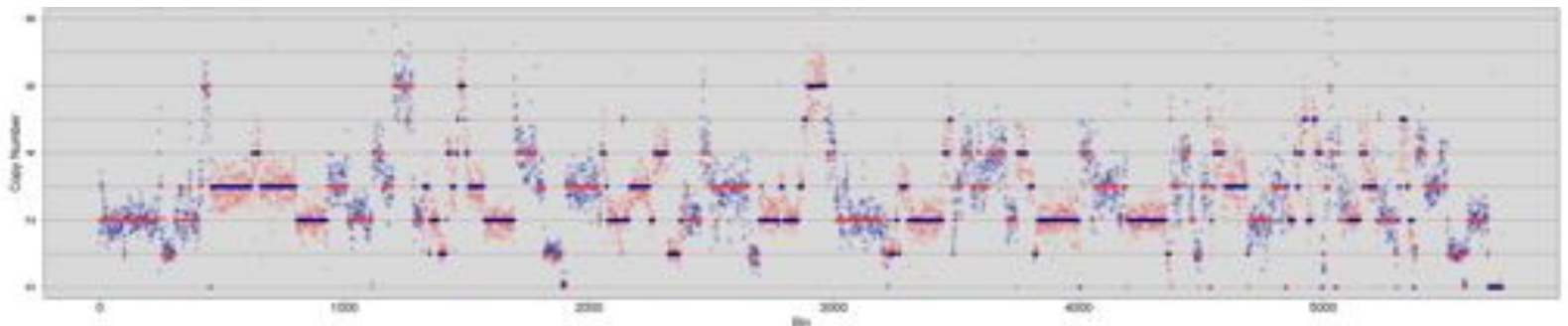
MDA



MALBAC

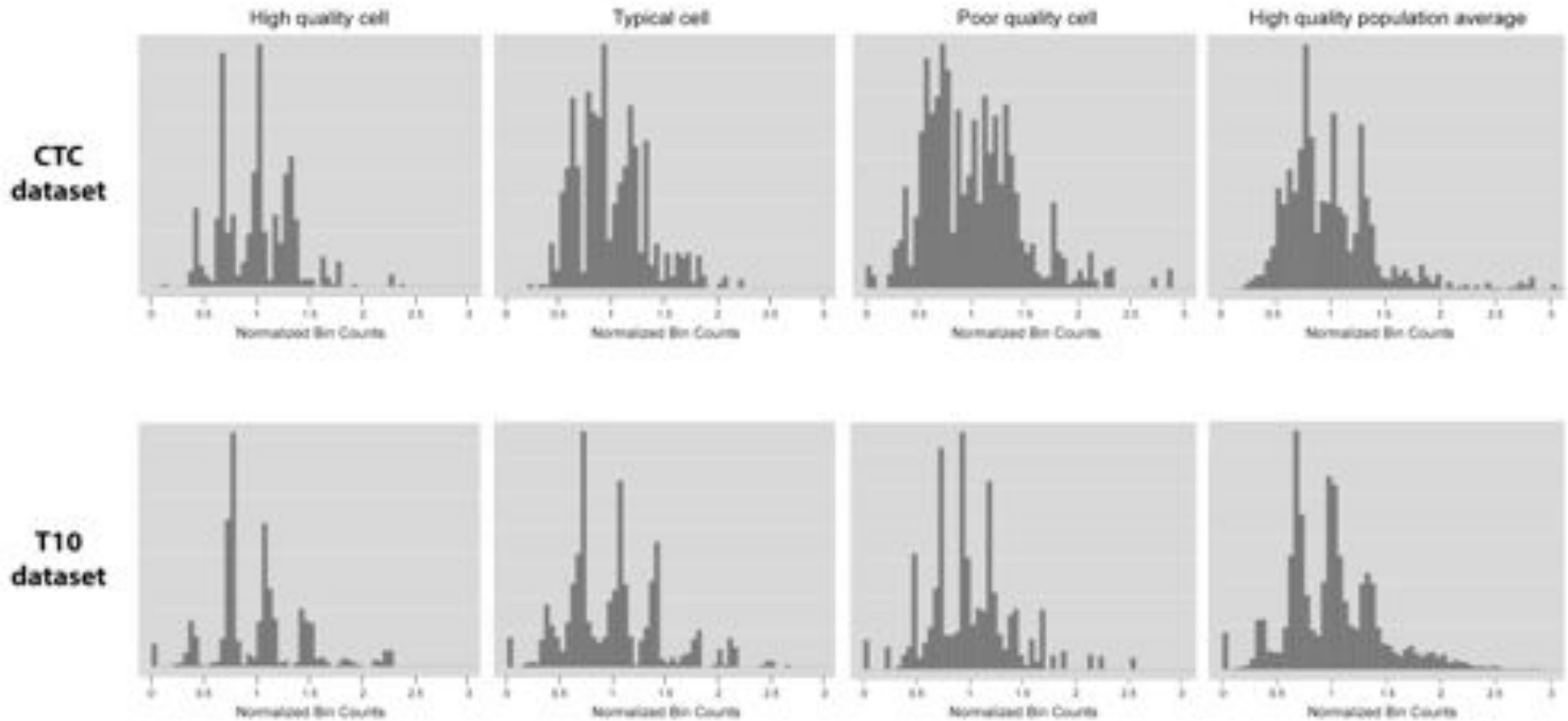


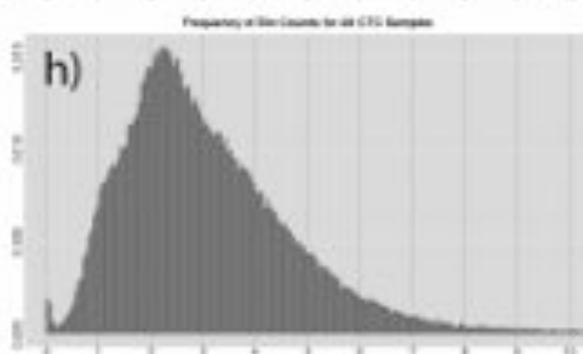
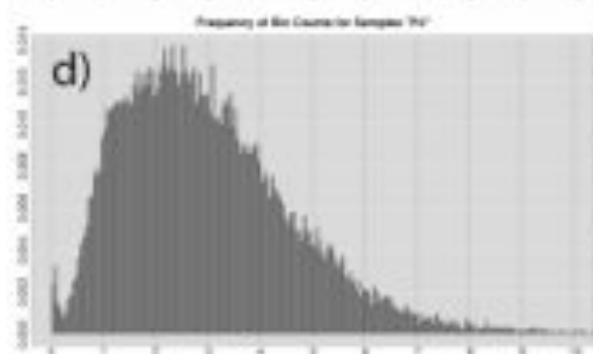
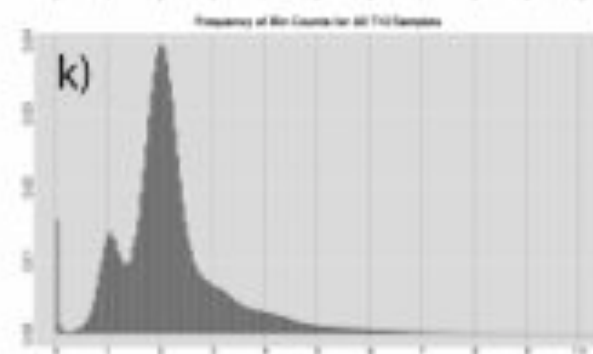
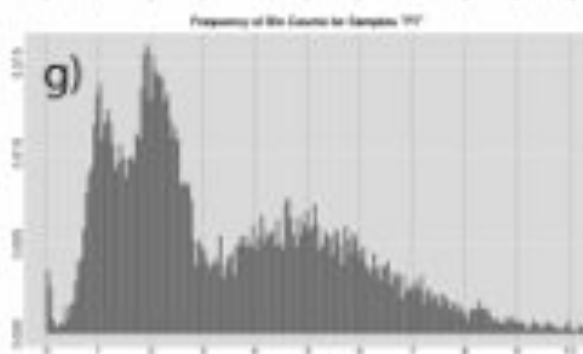
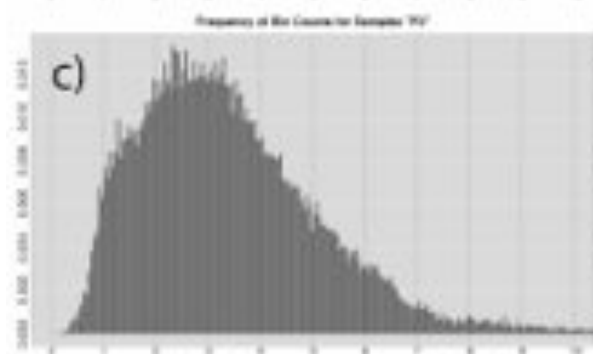
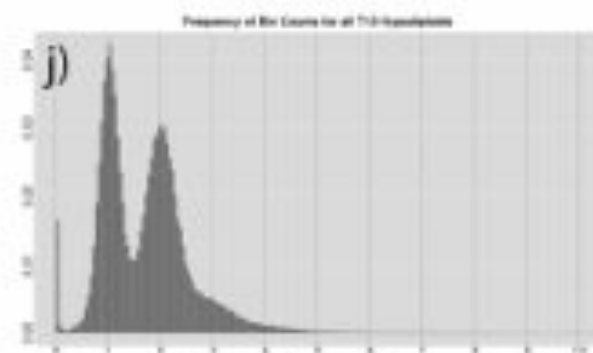
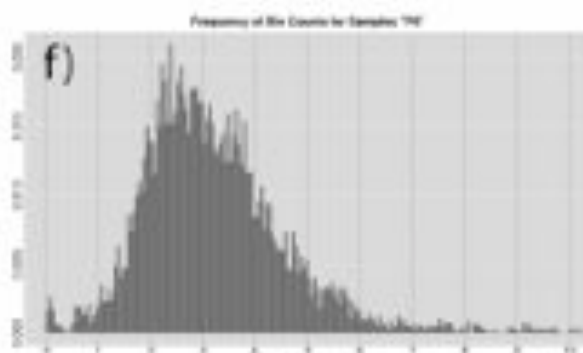
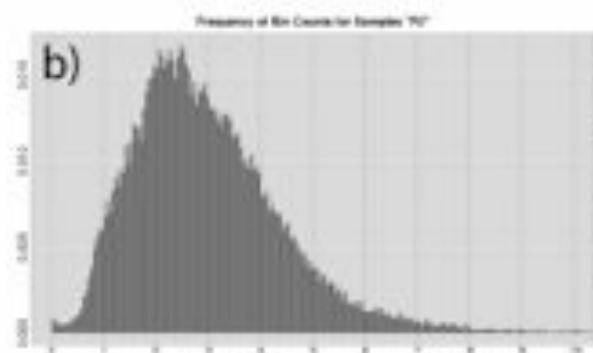
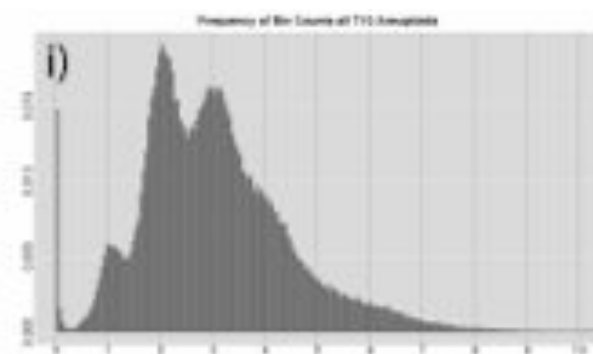
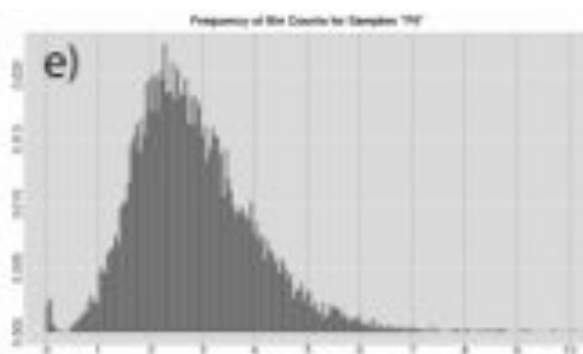
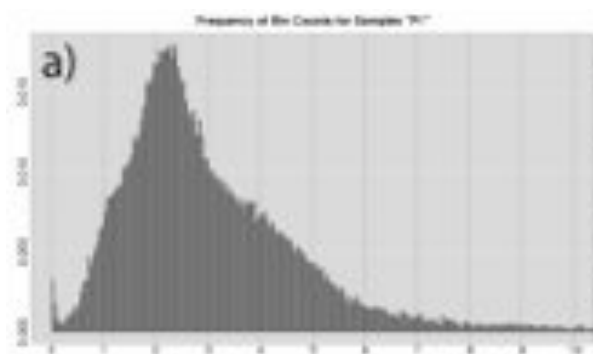
DOP-PCR



Detecting Integer Copy Number States

Histograms of Normalized Bin Counts





Conclusions and Recommendations

- Don't use MDA! For CNV analysis we recommend using DOP-PCR for WGA.
 - Sequence healthy diploid cells for reference
 - Plan for 25% of cells to be filtered out
 - Save your FACS data
 - Start with 500kb bins: 100 reads/bin (~1-1.5 million reads per cell)
 - Control for gender

Acknowledgments

Mike Schatz

Rob Aboukhalil

James Gurtowski

Srividya Ramakrishnan

Hayan Lee

Giuseppe Narzisi

Shoshana Marcus

Maria Nattestad

Han Fang

John and Amy Phelan Fellowship

Ginkgo Team

Jude Kendall

Timour Baslan

Mickey Atwal

Jim Hicks

Mike Wigler

Committee and Collaborators

Dick McCombie

Dave Tuveson

Zach Lippman

Lloyd Trotman

Josh Huang



Cold Spring Harbor Laboratory

WATSON SCHOOL
of BIOLOGICAL SCIENCES

