

CNAViz: User-guided local and global copy-number segmentation for tumor sequencing data

05-20-2022

Zubair Lalani^{*, 1}, Gillian Chu^{*, 1}, Silas Hsu¹, Simone Zaccaria^{†, 2, 3}, and Mohammed El-Kebir^{†, 1, 4}

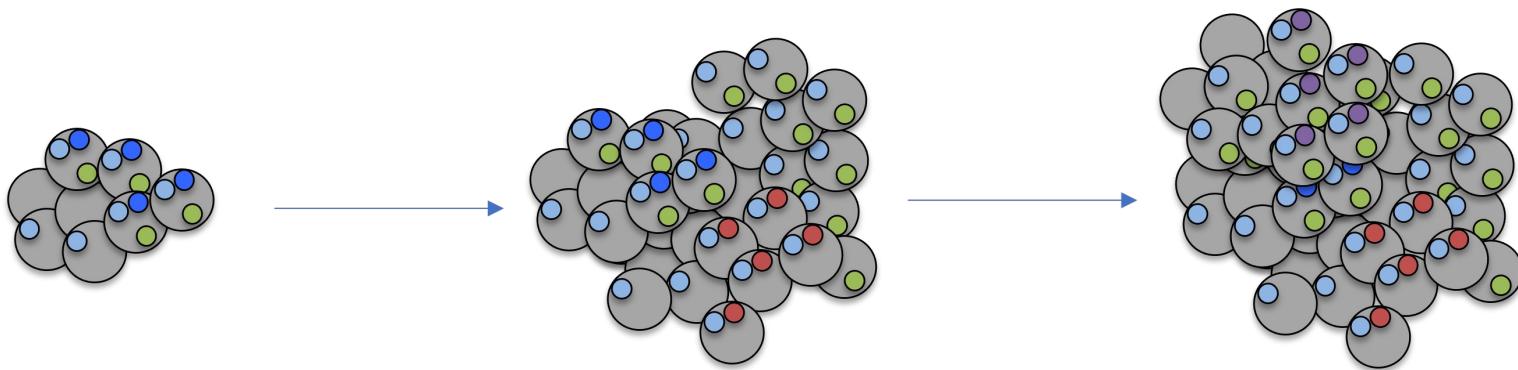
¹ Department of Computer Science, University of Illinois at Urbana-Champaign, ² Computational Cancer Genomics Group, University College London Cancer Institute ³, Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, ⁴ Cancer Center at Illinois, University of Illinois Urbana-Champaign

* joint first-author, † corresponding author

- Background & Motivation
- CNAViz Toolkit
 - Overview
 - Coordinate Plots
 - Linear Plots
- Exploration of Simulated Data
- Results on Biological Data
- Concluding Remarks

Cancer Evolution

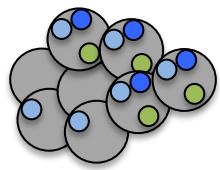
Copy number aberrations (CNAs) are ubiquitous in cancer. To understand cancer evolution and progression, we need to study these CNAs.



Clonal Evolution Theory of Cancer
[Nowell, 1976]

Cancer Evolution through Copy Number Aberrations (CNAs)

Deletion



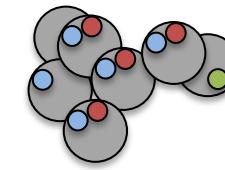
Copy Number: 1

Normal Somatic Cell



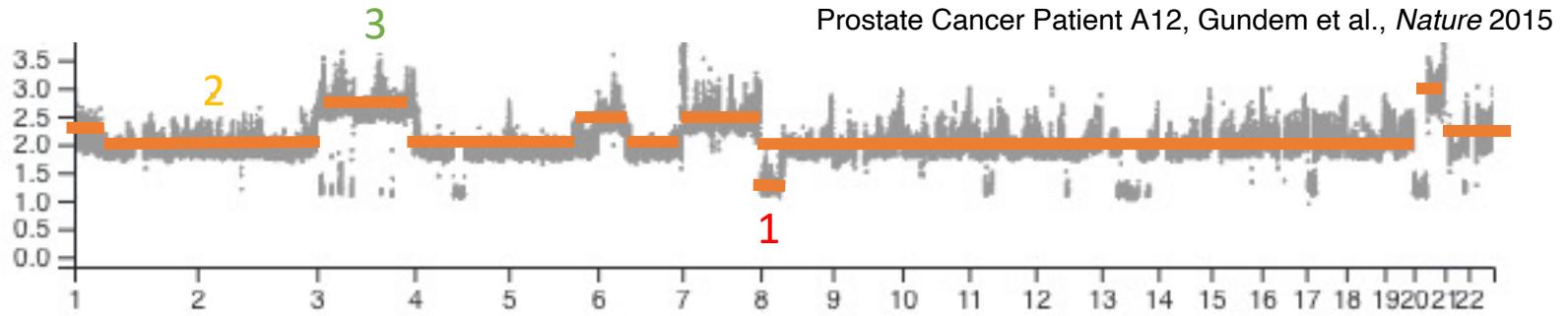
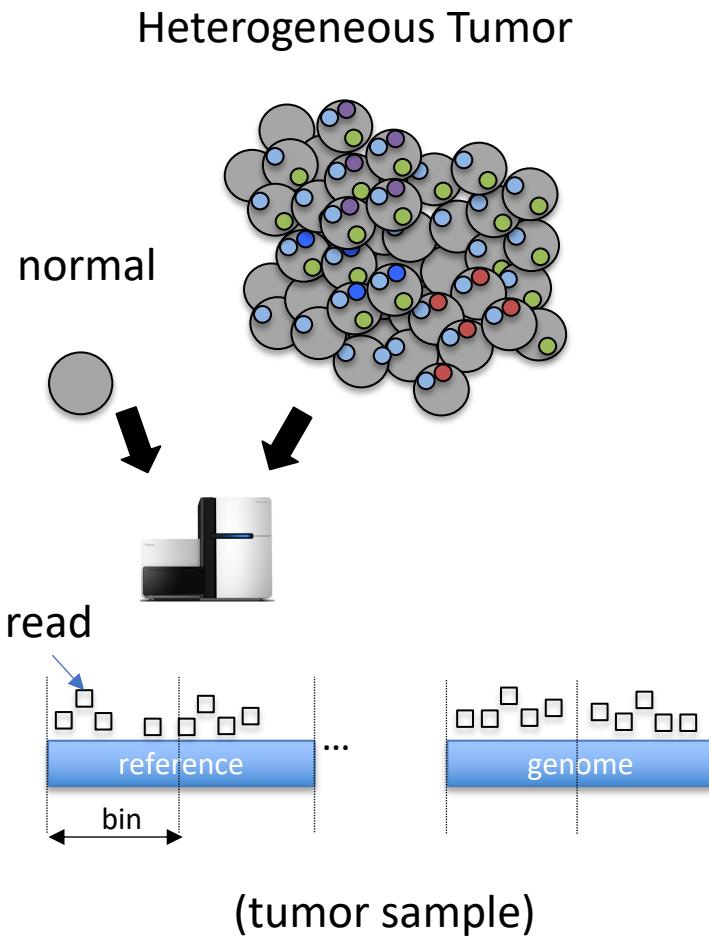
Copy Number: 2

Amplification

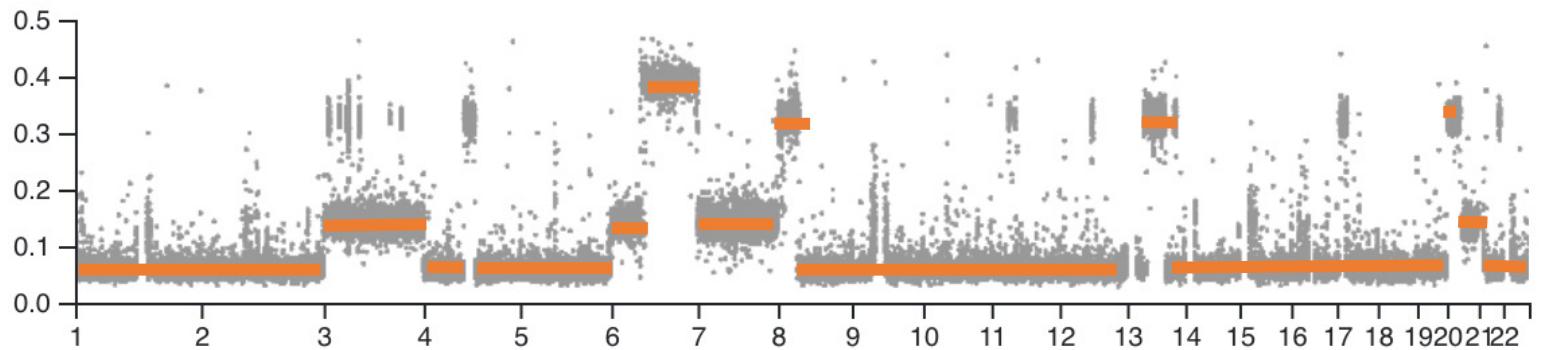


Copy Number: 3

Copy Number Calling: Signals for Segmentation



1. Read Depth Ratio (RDR) – number of sequencing reads we observe across the genome



2. B-Allele Frequency (BAF) – frequency of the nucleotides



1/2



2/3



0

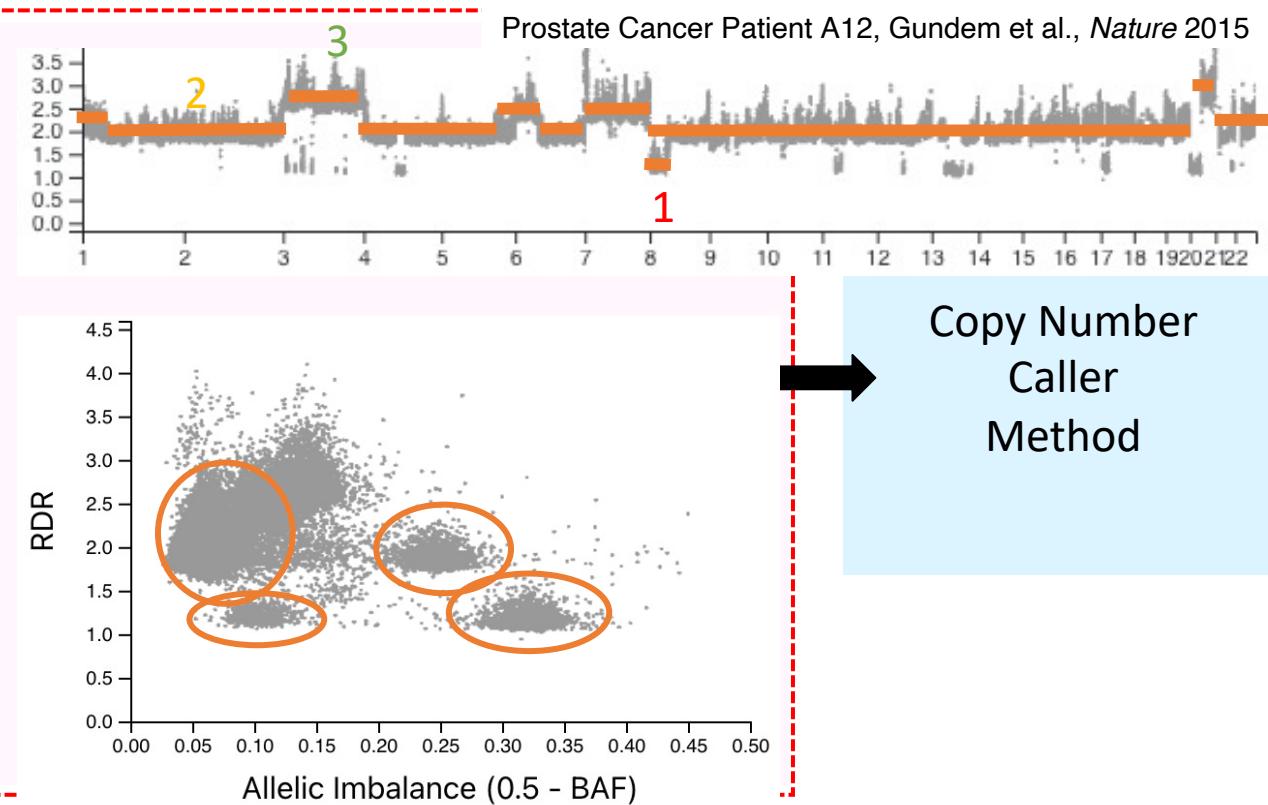
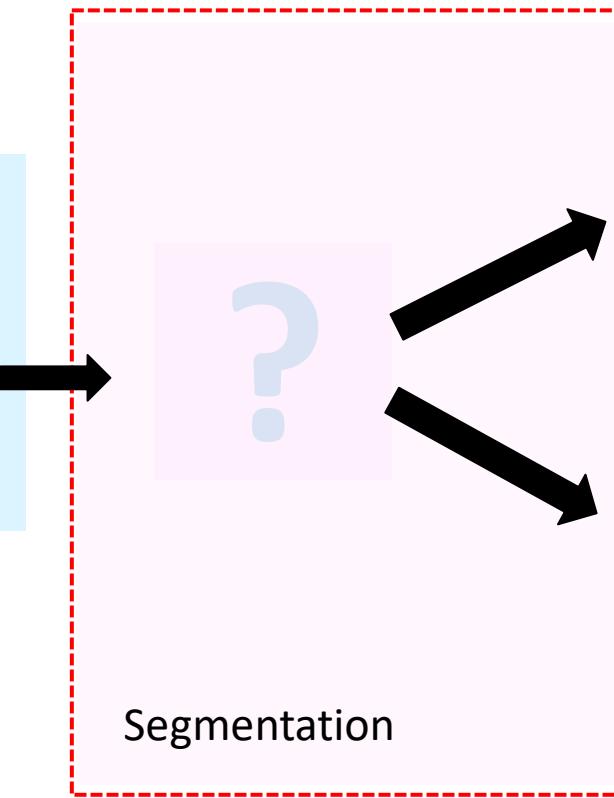
CNAViz: User-guided Unification of Local and Global Insights

A cluster/segment is a set of bins with shared copy numbers. This is a key preprocessing since input data is noisy.

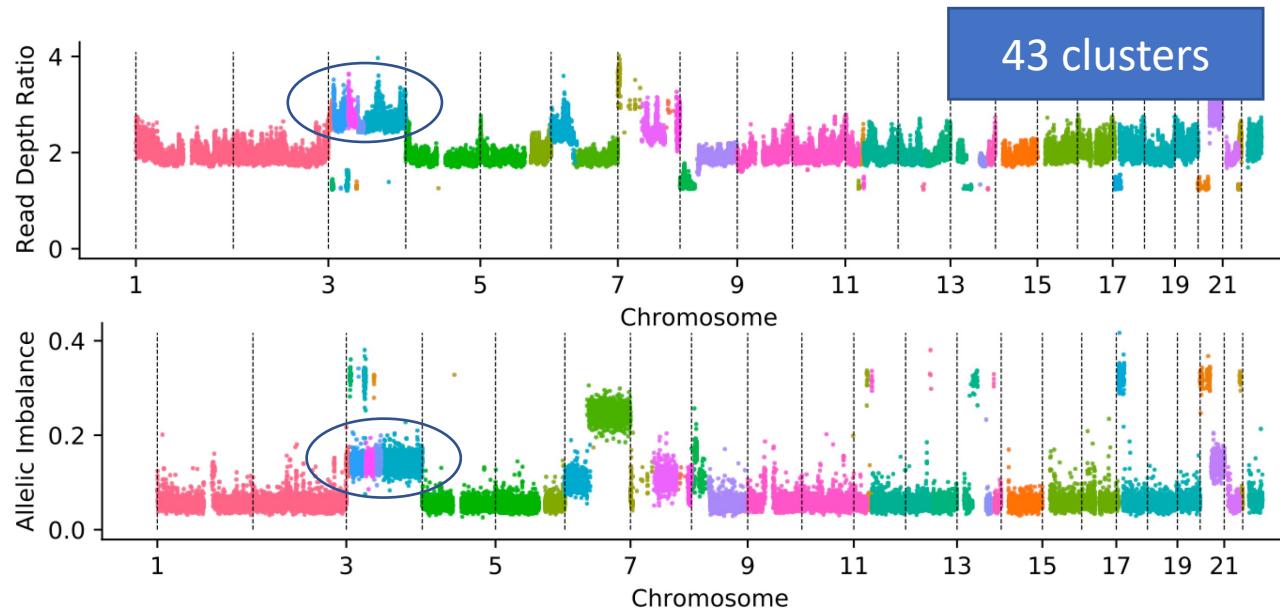
chr	start	end	RDR	BAF
1	1	50000	1.69	0.38
1	50000	100000	1.52	0.36
...
22	51200000	51250000	1.93	0.37

Input Data

Segmentation



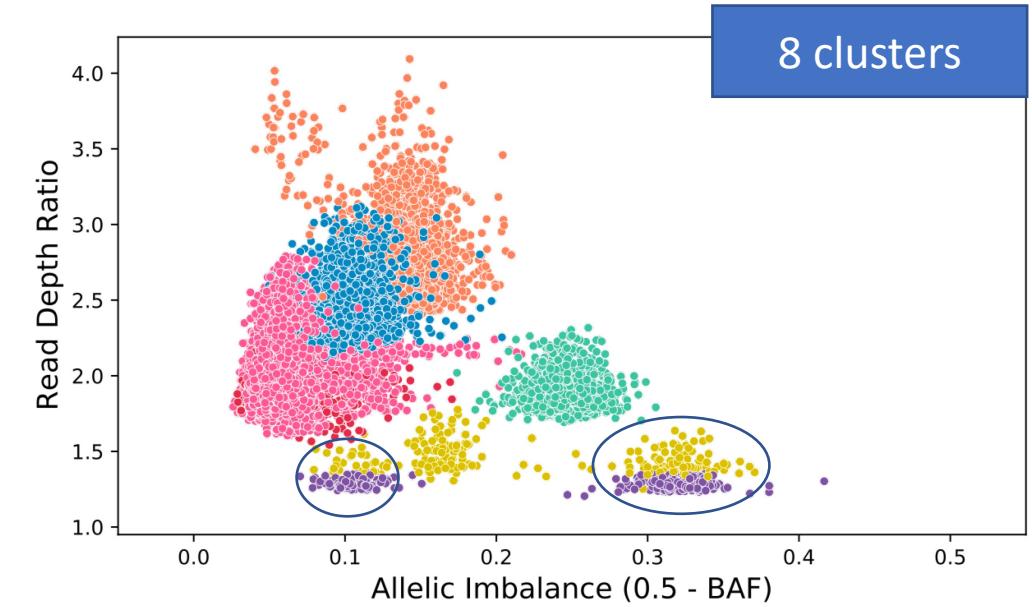
CNA Calling: Pros/Cons to Current Segmentation Approaches



Local approach via ASCAT

[Van Loo et al., PNAS 2010]

- + detects small focal CNAs
- prone to noise and difficult to see trends



Global approach via HATCHet

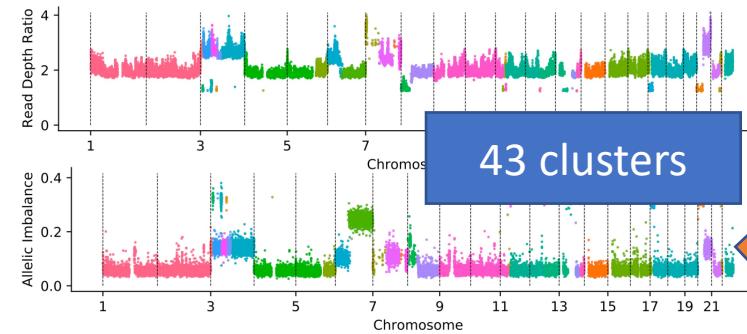
[Zaccaria et al., Nature Communication 2020]

- + aggregation across genome overcomes noise
- focal CNAs might be missed

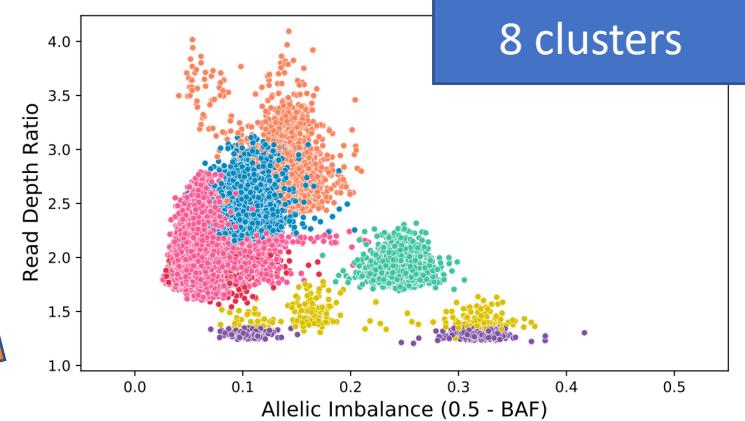
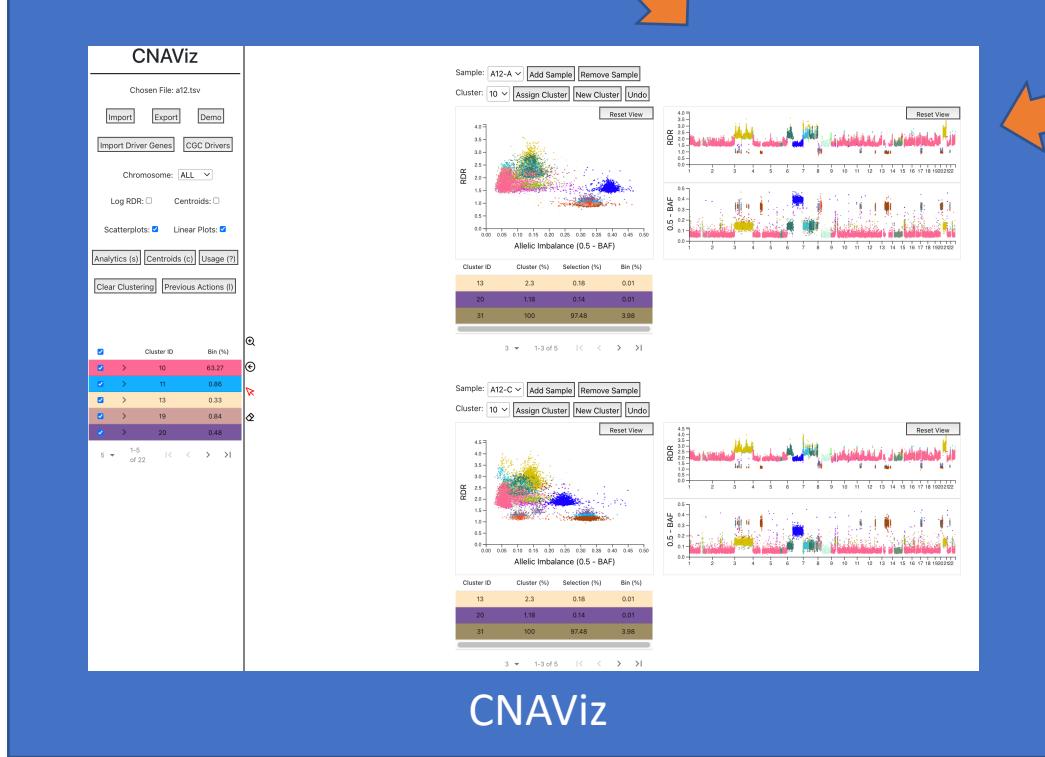
Key Motivation: Current segmentation methods leave room for improvement in copy number calling.

CNA Calling: Pros/Cons to Current Segmentation Approaches

Local approach via ASCAT [Van Loo et al., PNAS 2010]



- + detects small focal CNAs
- prone to noise and difficult to see trends



Global approach via HATCHet
[Zaccaria et al., Nature Communications 2020]

- + aggregation across genome overcomes noise
- focal CNAs might be missed

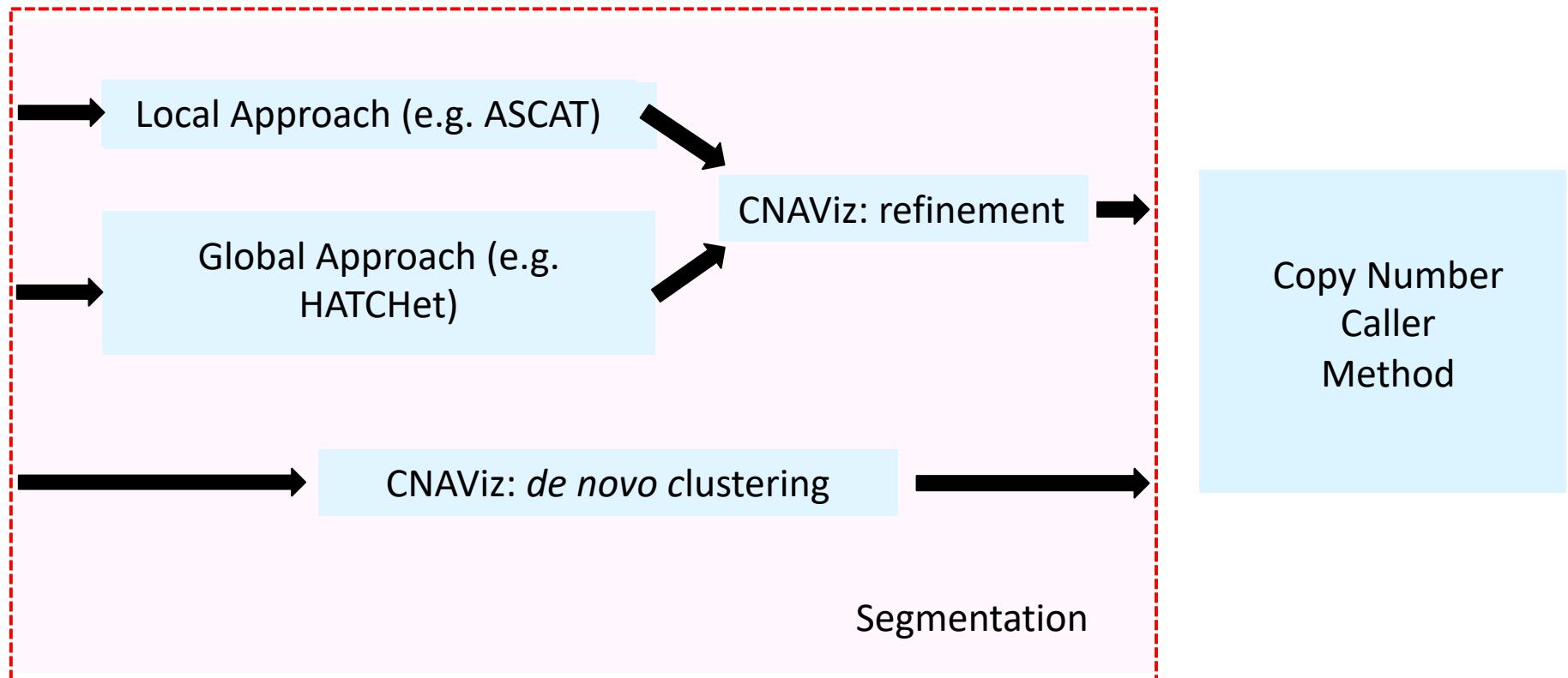
CNAViz is a user-guided interactive tool to perform segmentation. By using CNAViz to combine local and global segmentation, we can perform more accurate copy-number aberrations.

CNAViz: User-guided Unification of Local and Global Insights

By using CNAViz to combine local and global segmentation, we can perform more accurate copy-number aberrations.

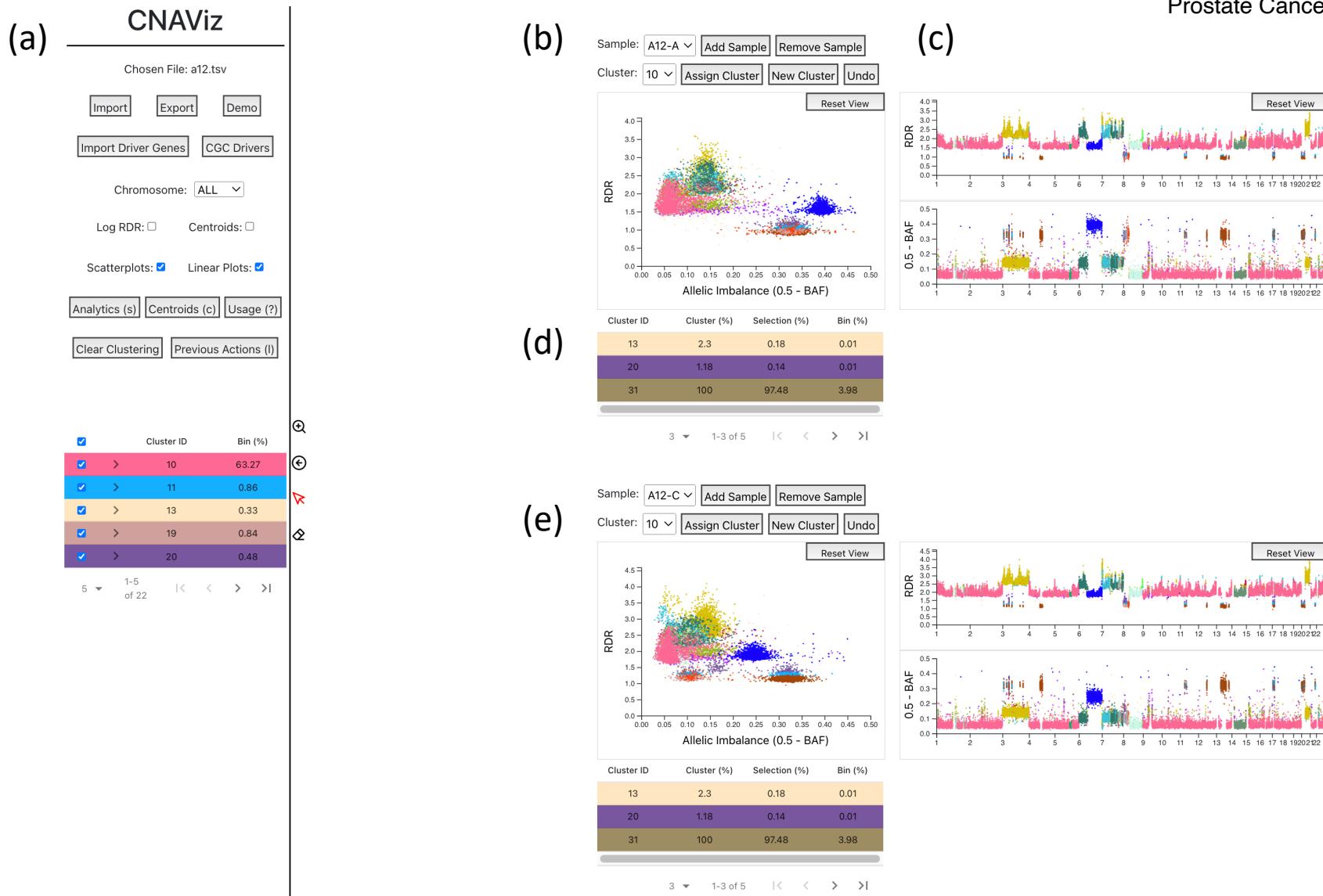
chr	start	end	RDR	BAF
1	1	50000	1.69	0.38
1	50000	100000	1.52	0.36
...
22	51200000	51250000	1.93	0.37

Input Data



- Background & Motivation
- **CNAViz Toolkit**
 - Overview
 - Coordinate Plots
 - Linear Plots
- Exploration of Simulated Data
- Results on Biological Data
- Concluding Remarks

CNAViz Toolkit: Overview



CNAViz Toolkit: Global View

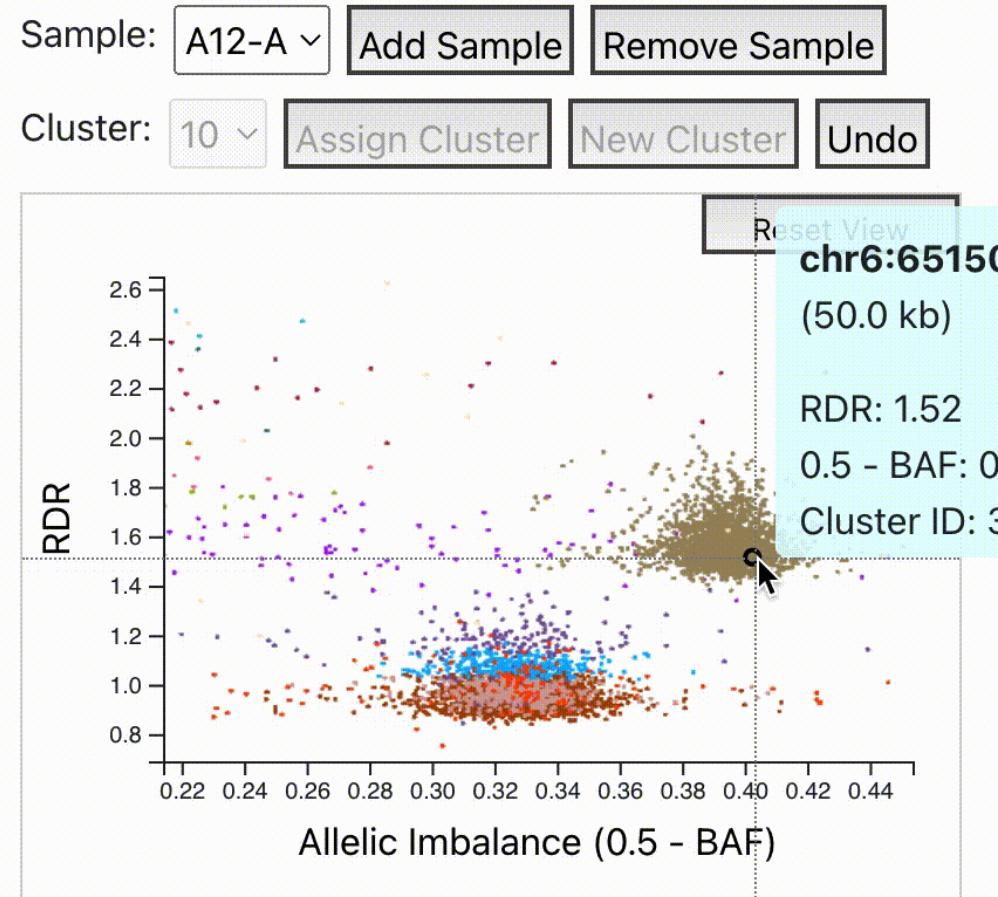
Illustrated Abilities:

- Zooming and Panning
- Hovering
- Selecting/De-selecting
- Assigning to Clusters

Motivation:

- The ability to split and merge clusters across chromosomes

Prostate Cancer Patient A12, Gundem et al., *Nature* 2015



CNAViz Toolkit: Local View

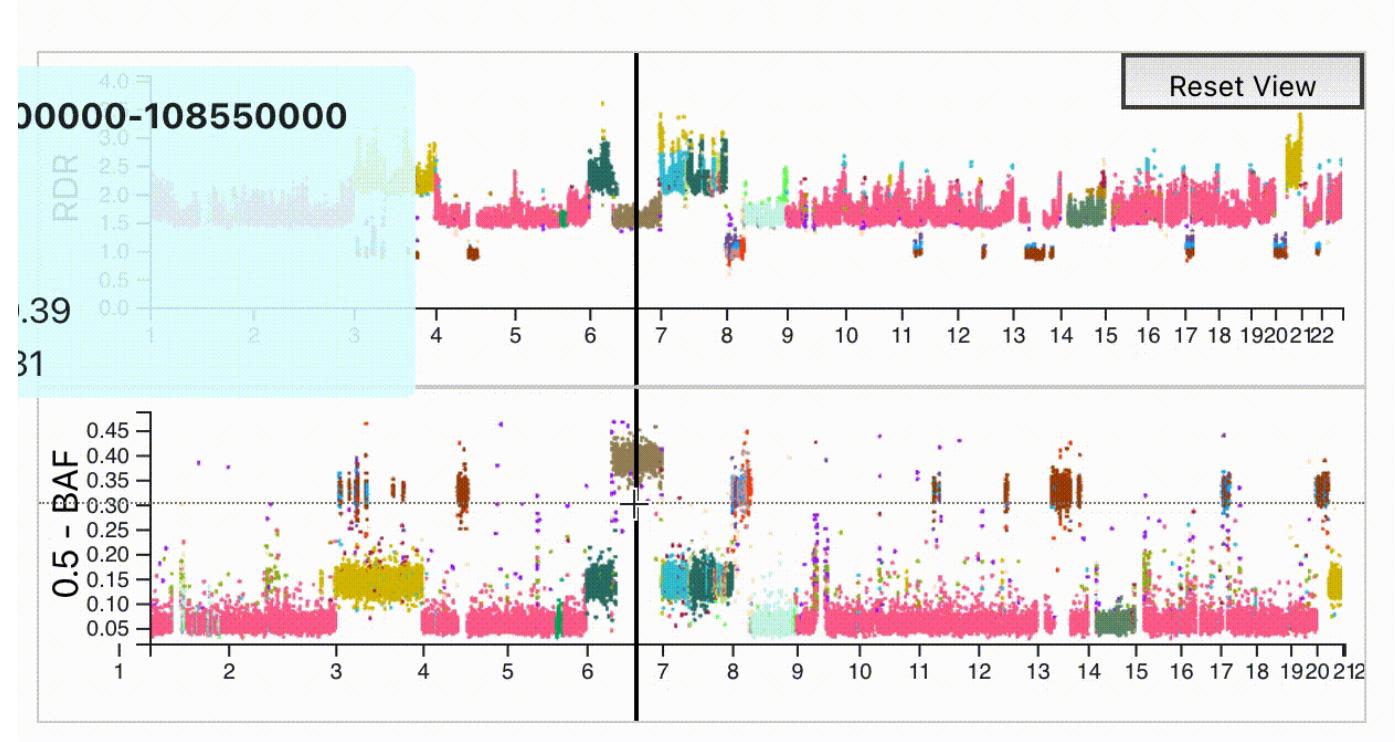
Prostate Cancer Patient A12, Gundem et al., *Nature* 2015

Illustrated Abilities:

- Zooming and Panning
- Hovering
- Selecting/De-selecting
- Assigning to Clusters

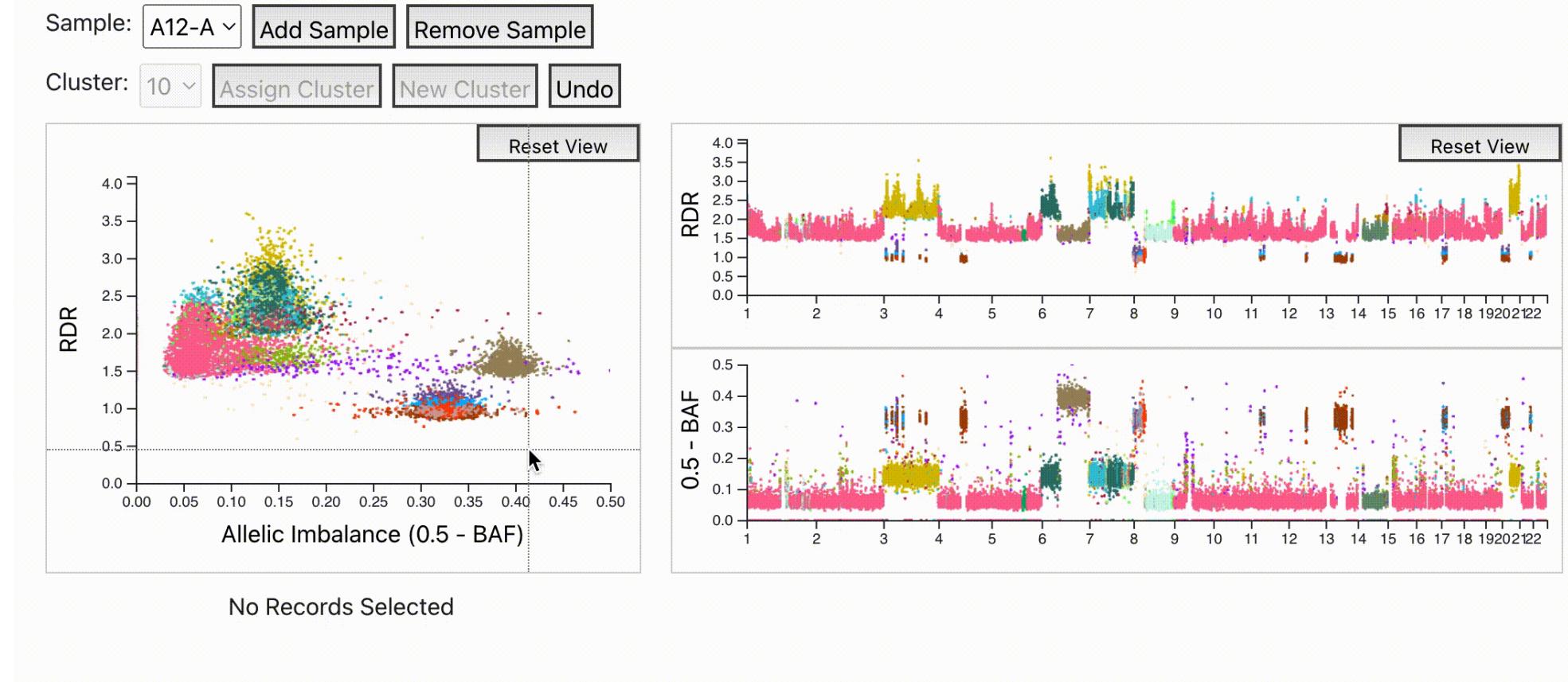
Motivation:

- The ability to split and merge clusters contiguously across the chromosomes



CNAViz Toolkit: Linear and Global Synced

Prostate Cancer Patient A12, Gundem et al., *Nature* 2015



Motivation:

- The ability to split and merge clusters considering local chromosome and global aggregate information

CNAViz Toolkit: Synced Samples

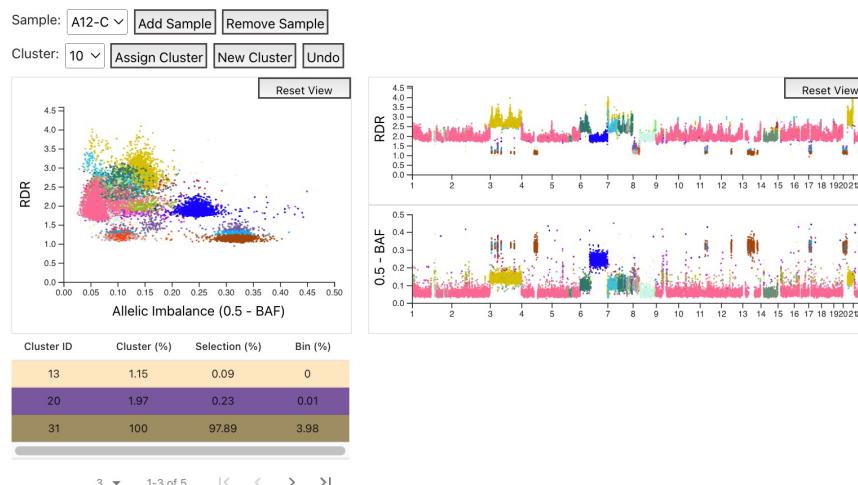
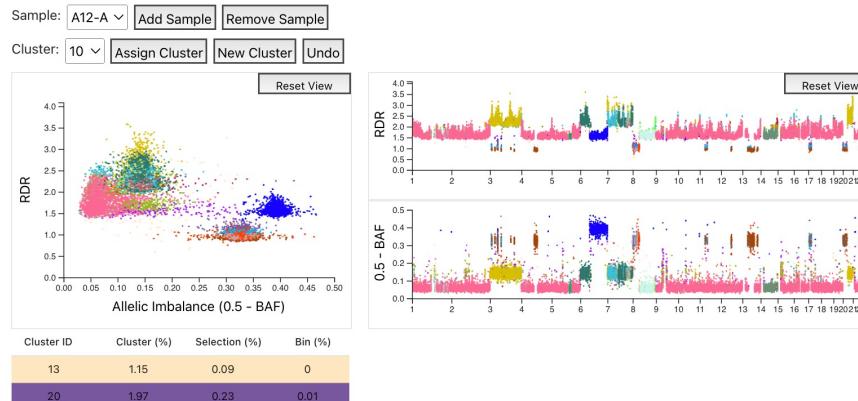
Prostate Cancer Patient A12, Gundem et al., *Nature* 2015

Illustrated Abilities:

- Zooming and Panning
- Hovering
- Selecting/De-selecting
- Assigning to Clusters
- Synced across plots and samples

Motivation:

- The ability to share information across tumor samples



CNAViz Toolkit: Gene Annotations

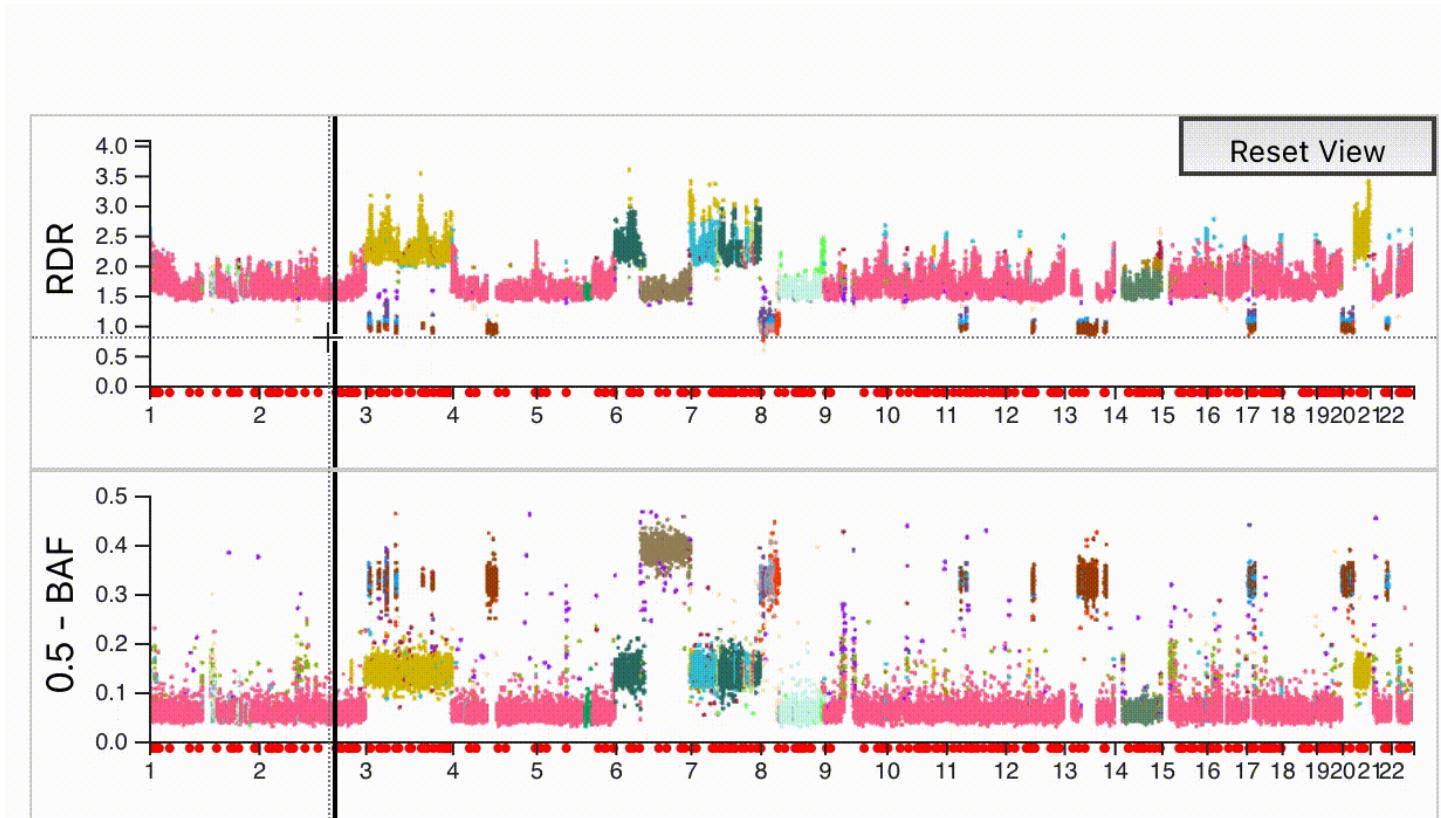
Prostate Cancer Patient A12, Gundem et al., *Nature* 2015

Illustrated Abilities:

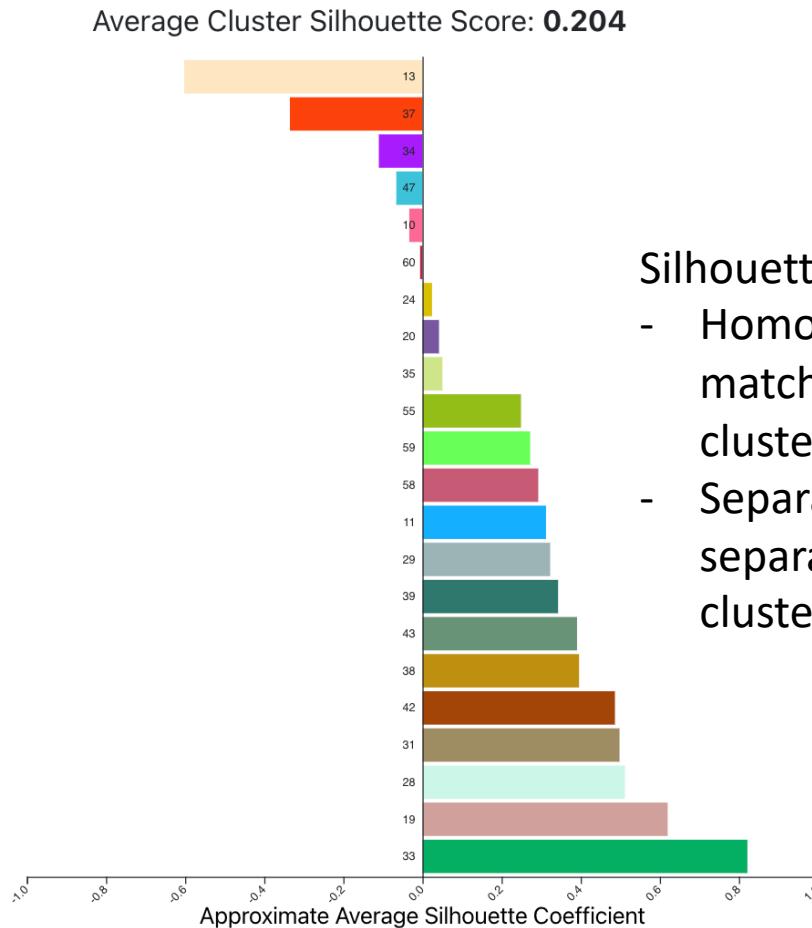
- Zooming and Panning
- Hovering
- Visualizing driver genes

Motivation:

- The ability to investigate whether driver genes are deleted or amplified

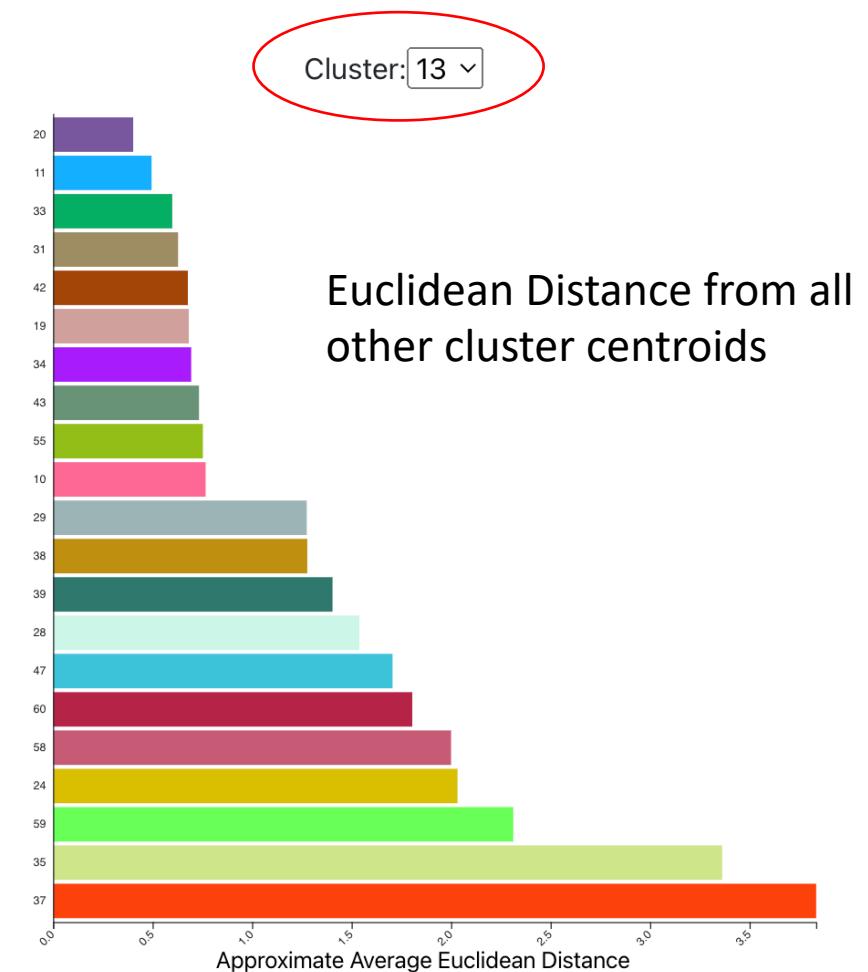


CNAViz Toolkit: Real-time Clustering Analysis



Silhouette Score [-1, 1]

- Homogeneity: how well bins match other bins in the same cluster
- Separation: how well bins separate from bins in other clusters

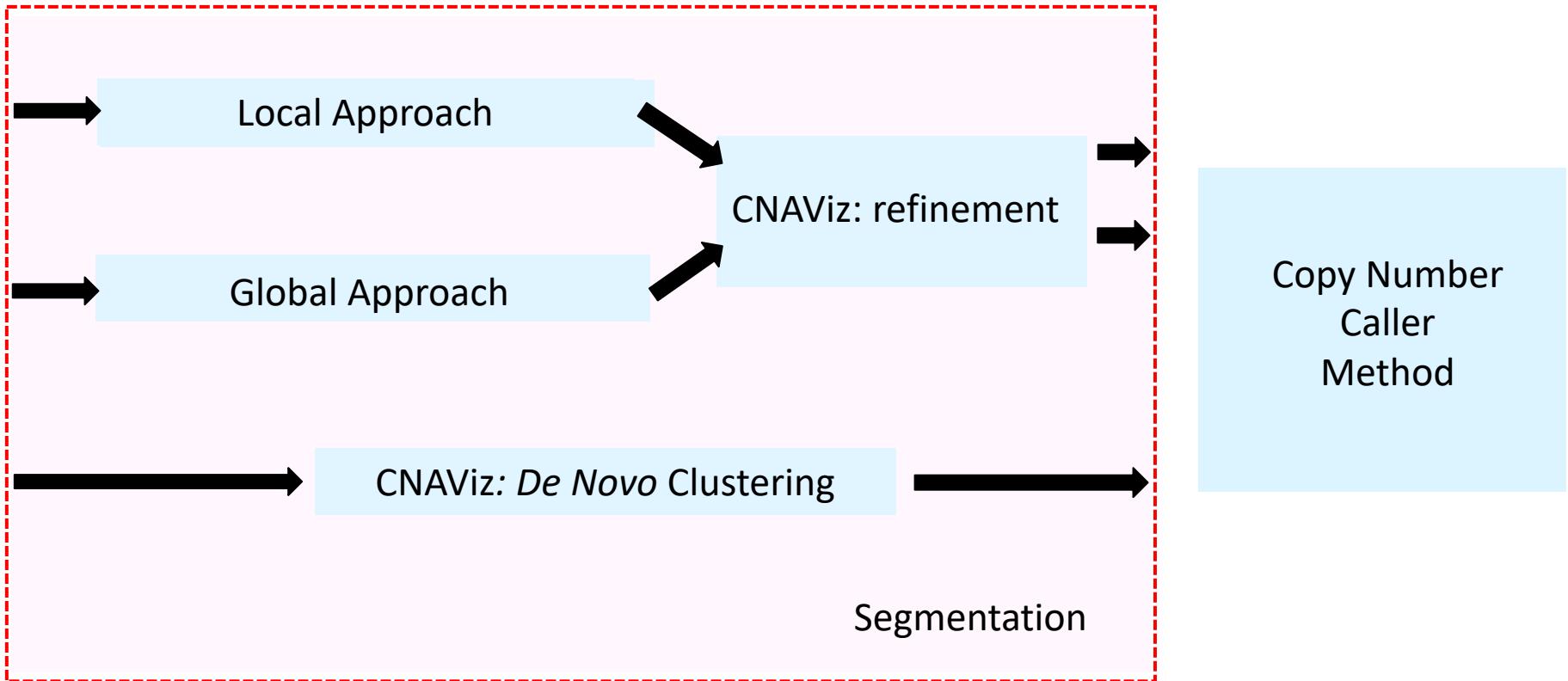


[Rousseeuw et al. 1987]

CNAViz: User-guided Insights

chr	start	end	RDR	BAF
1	1	50000	1.69	0.38
1	50000	100000	1.52	0.36
...
22	51200000	51250000	1.93	0.37

Input Data



By using CNAViz to combine local and global segmentation, we can perform more accurate copy-number calling.

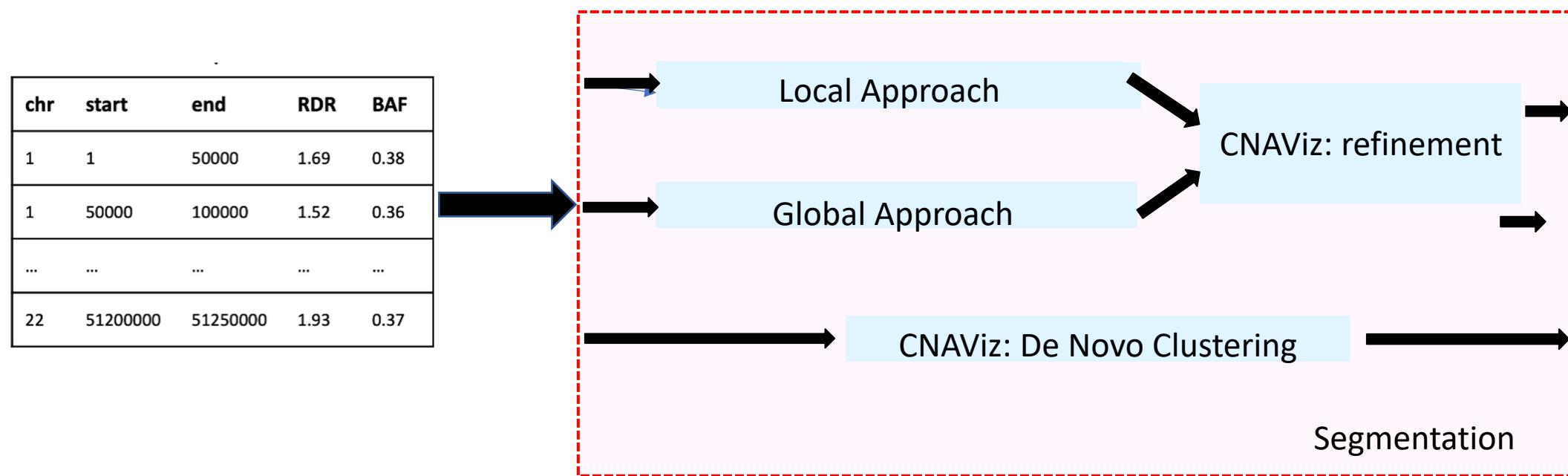
- Background & Motivation
- CNAViz Toolkit
 - Overview
 - Coordinate Plots
 - Linear Plots
- Exploration of Simulated Data
- Results on Biological Data
- Concluding Remarks

Simulated Data: Context

[Zaccaria and Raphael, *Nature* 2020]

Simulated Data (MASCoTE):

- We used 4 bulk sequencing samples, with 2 tumor clones each (ground truth copy numbers)
- Assess our tool in *de novo* mode or with existing local or global segmentation
- Do we improve the segmentation with CNAViz?



Simulated Data: Results

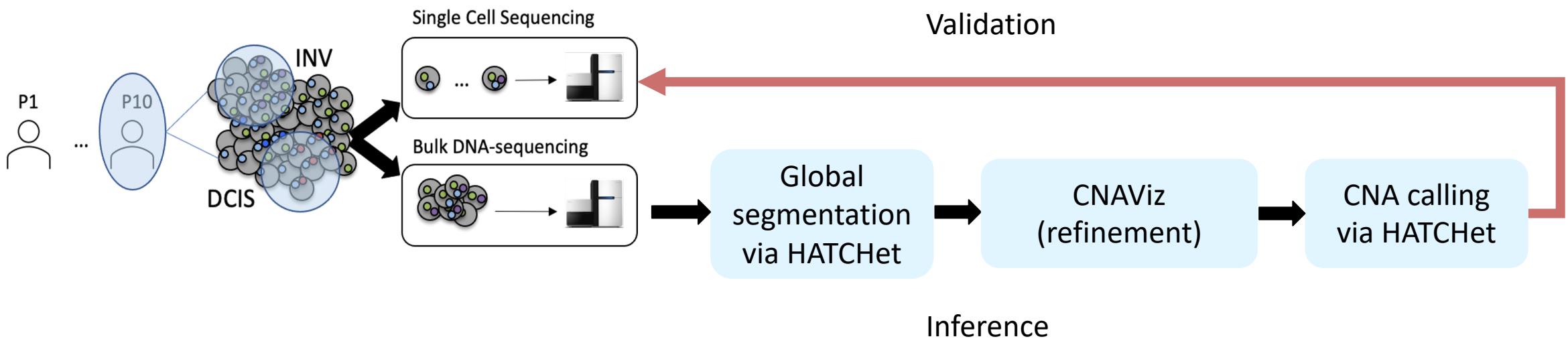
Method	# Clusters	Adjusted Rand Index	V-measure
Ground Truth	22	-	-
CNAViz (<i>de novo</i>)	24	0.99553	0.97048
HATCHet	18	0.99457	0.96303
HATCHet + CNAViz	24	0.99539	0.96983
ASCAT	69	0.07376	0.21984
ASCAT + CNAViz	21	0.99509	0.96804

Using CNAViz in *de novo* mode or to refine existing segmentation improves the clustering by several metrics

- Background & Motivation
- CNAViz Toolkit
 - Overview
 - Coordinate Plots
 - Linear Plots
- Exploration of Simulated Data
- Results on Biological Data
- Concluding Remarks

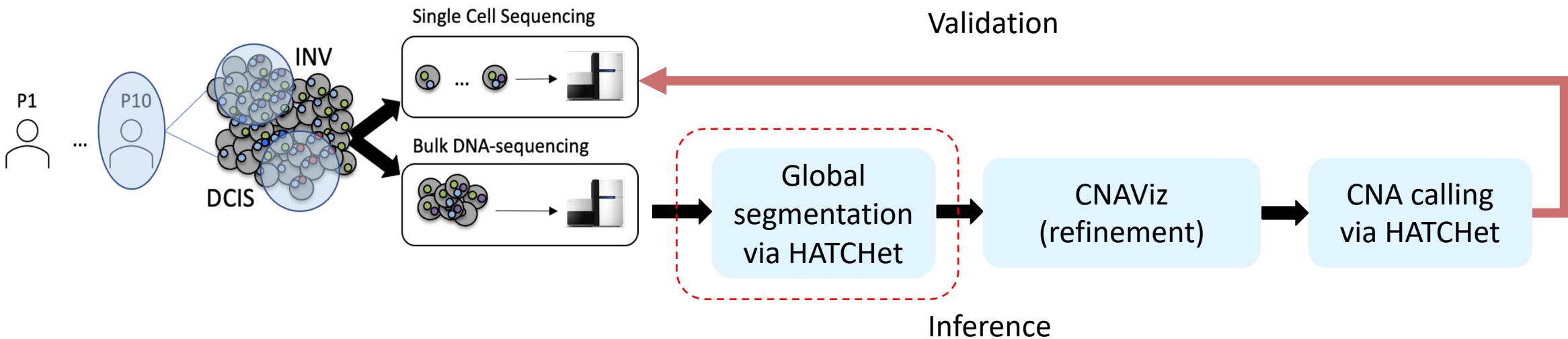
Real Data: Context

[Casasent et al. *Cell* 2018]



Real Data: Context

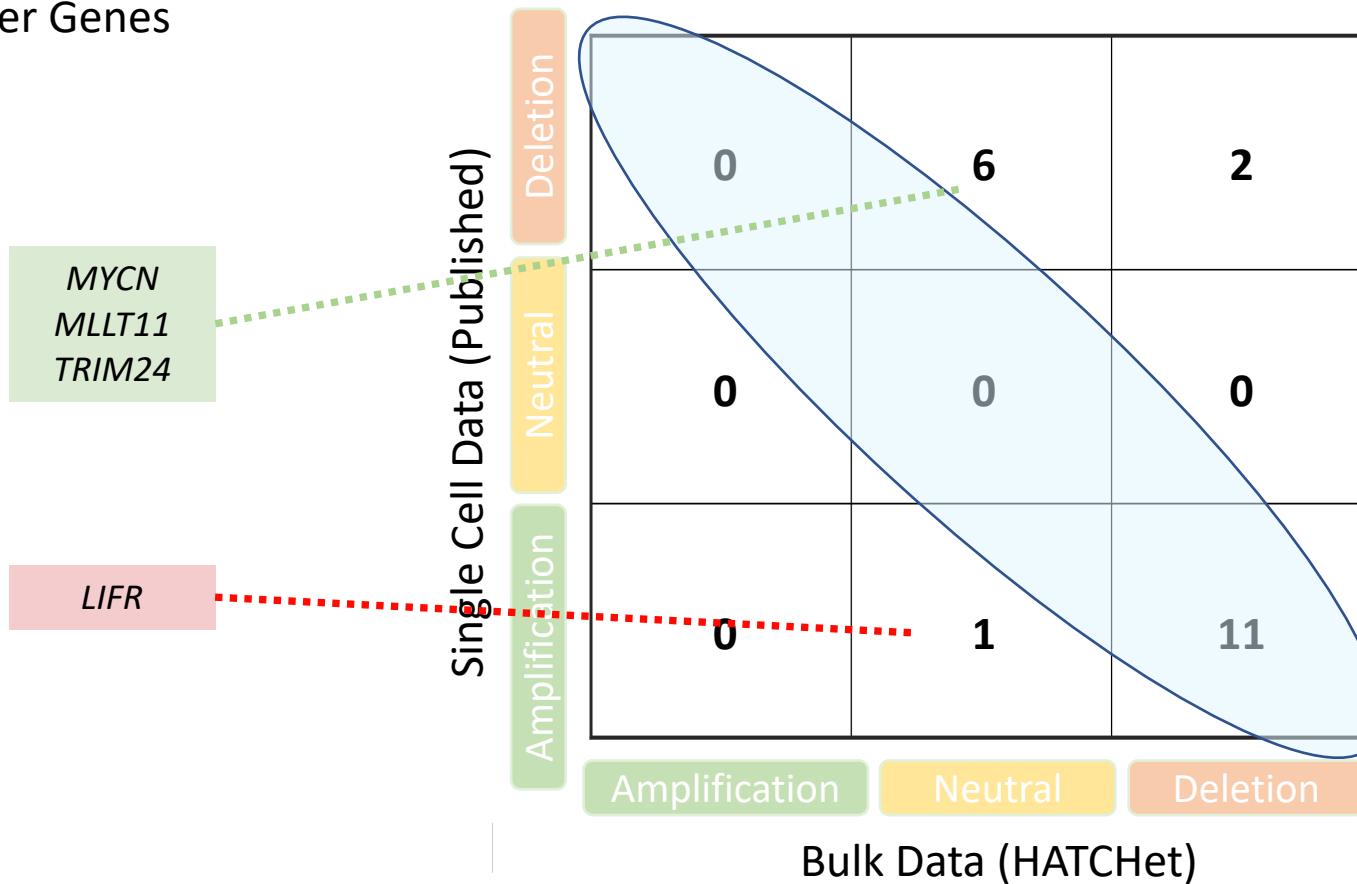
[Casasent et al. *Cell* 2018]



Real Data: Results

[Patient P10, Casasent et al. *Cell* 2018]

Copy number calling on
20 Driver Genes

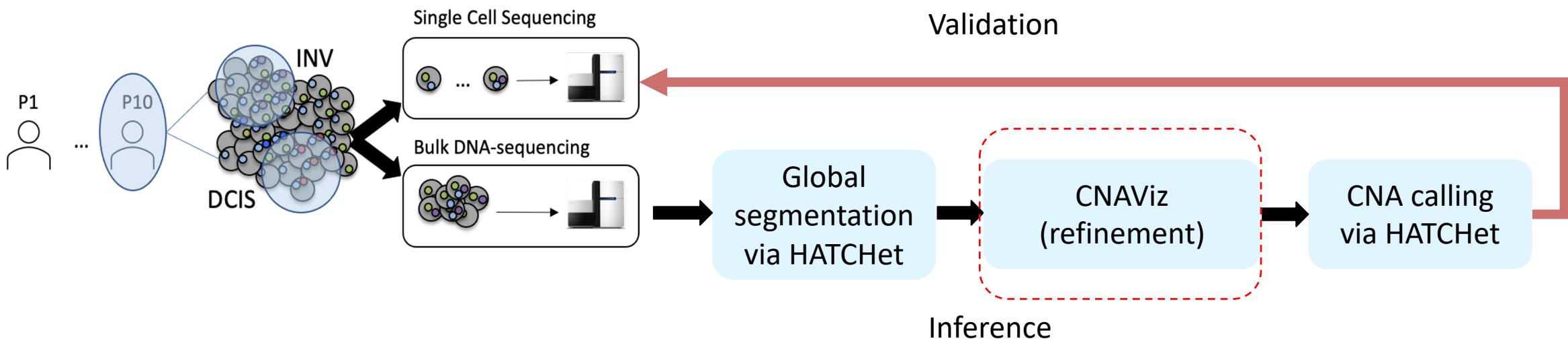


$$\% \text{ Correct} = \frac{11}{20}$$

55% Accuracy

Real Data: Context

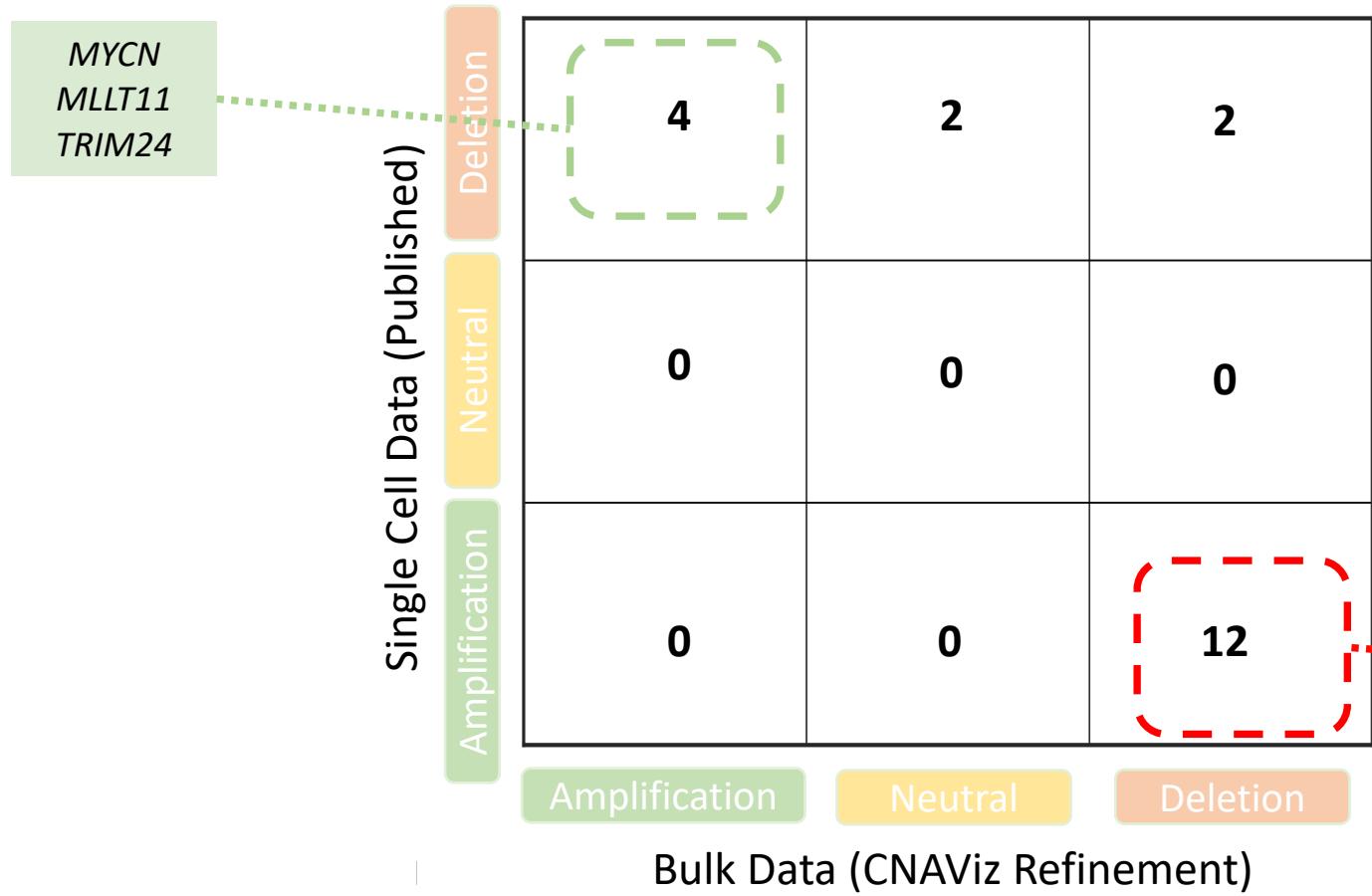
[Casasent et al. *Cell* 2018]



Real Data: Results

[Patient P10, Casasent et al. *Cell* 2018]

Newly Assigned Amplifications



$$\% \text{ Correct} = \frac{16}{20}$$

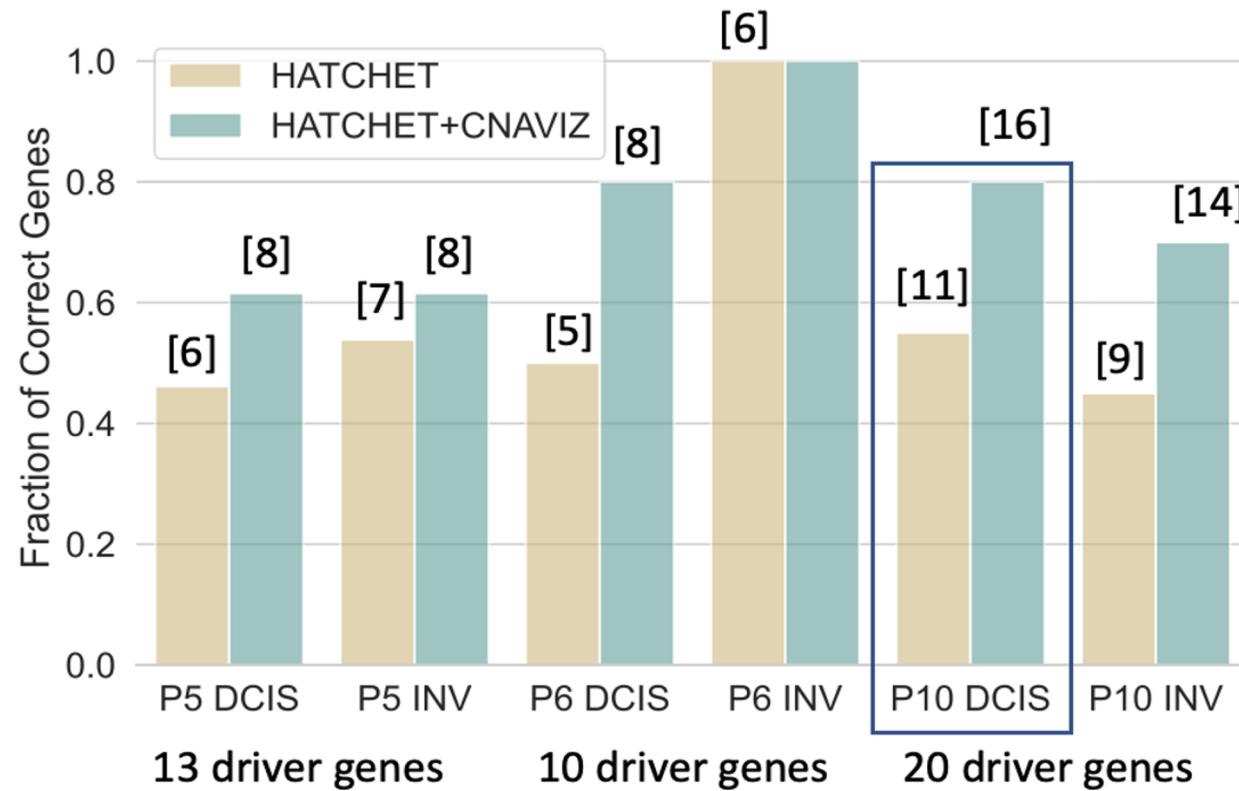
80% Accuracy!

Newly Assigned Deletions

LIFR

Real Data: Results

[Patient P10, Casasent et al. *Cell* 2018]



By using CNAViz to refine global segmentation, we can perform more accurate copy-number aberrations.

Conclusions & Discussion

CNAViz is an interactive user-guided tool to improve copy-number calling on noisy sequencing data.

By using CNAViz to combine local and global segmentation, we can perform more accurate copy-number aberrations.

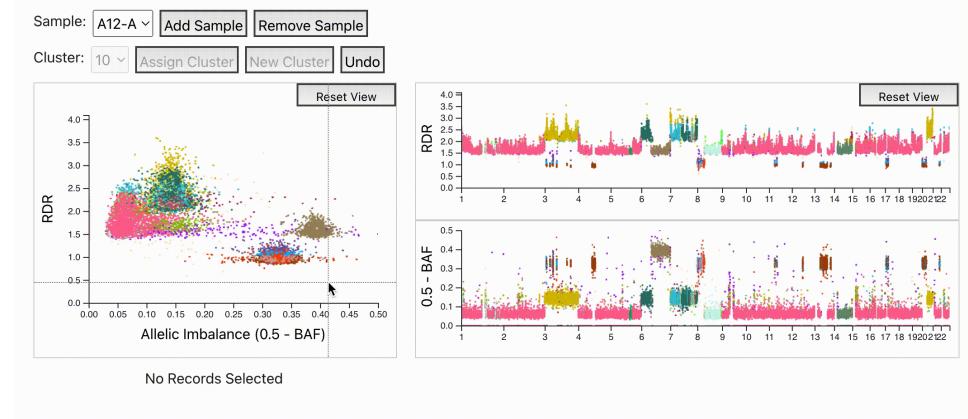
Summary:

- All examples were on bulk but same applies to single cell
- CNAViz can be run in *de novo* mode and refinement mode
- Better clustering by several metrics on simulated data
- Better classification of important driver genes on real data

Future Work:

- Modeling cohort effects
- Real-time suggestions
- Generate generalizable suggestions based on cohort tumor data
- Crowd-sourcing segmentation solutions on public datasets (e.g. TCGA, ICGC)

Availability: <https://elkebir-group.github.io/cnaviz/>



CNAViz

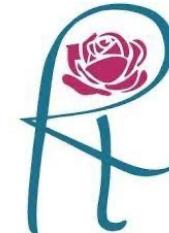
Acknowledgements

Thank you for your attention! Questions?

Thanks to Zubair, Mohammed, Simone and the rest of the El-Kebir lab!



CCF-1850502
CCF-2046488



M917

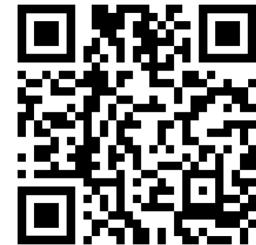
Availability: <https://elkebir-group.github.io/cnaviz/>

Simone is hiring for a **Postdoc in Computational Cancer Genomics** working with single cells!

Details: <https://www.ucl.ac.uk/cancer/zaccaria-lab>

Twitter: @ZaccaSimo

Email: s.zaccaria@ucl.ac.uk



CNAViz