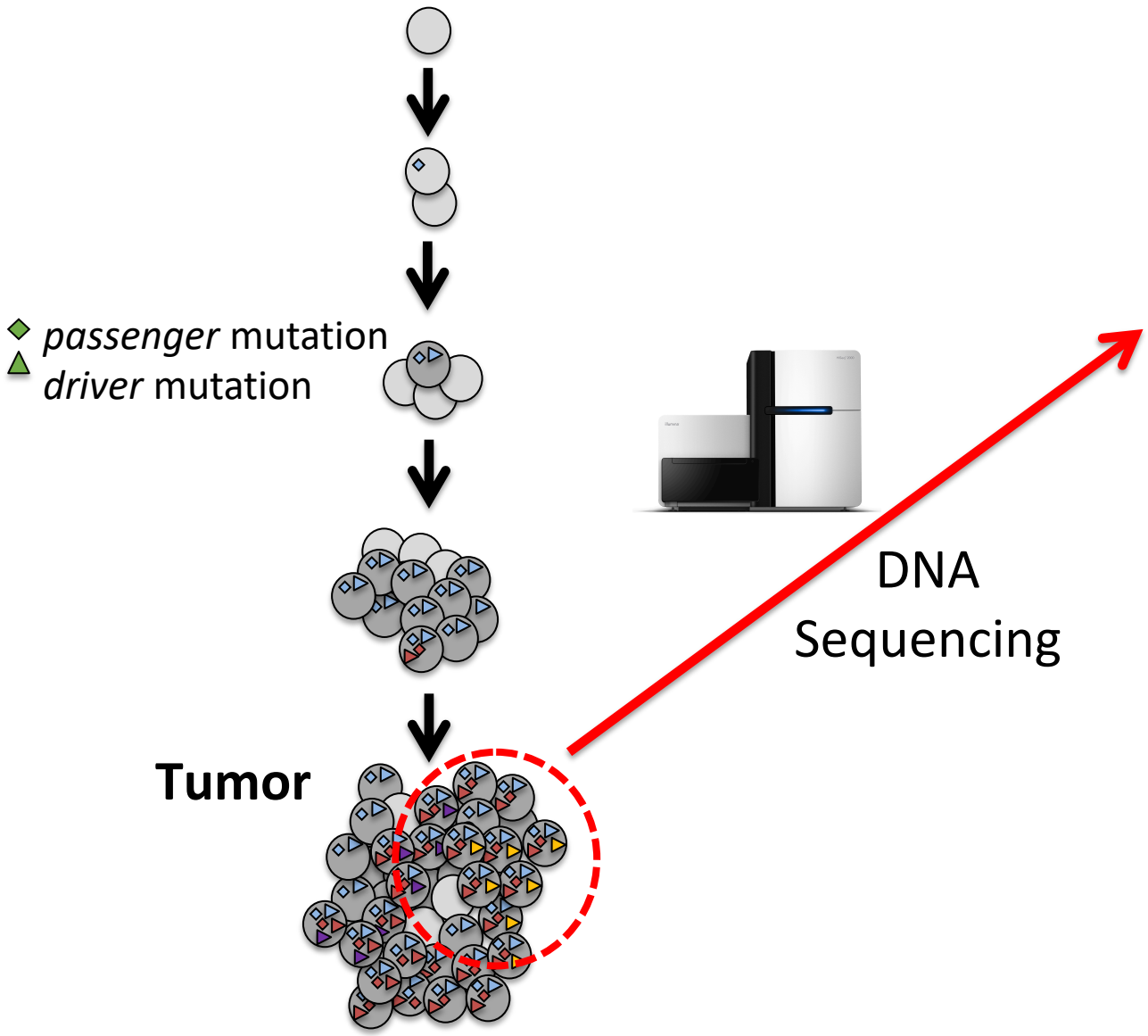
ITCR U24
Comprehensive and Robust Tools for Analysis of Tumor Heterogeneity and Evolution

PI: Ben Raphael

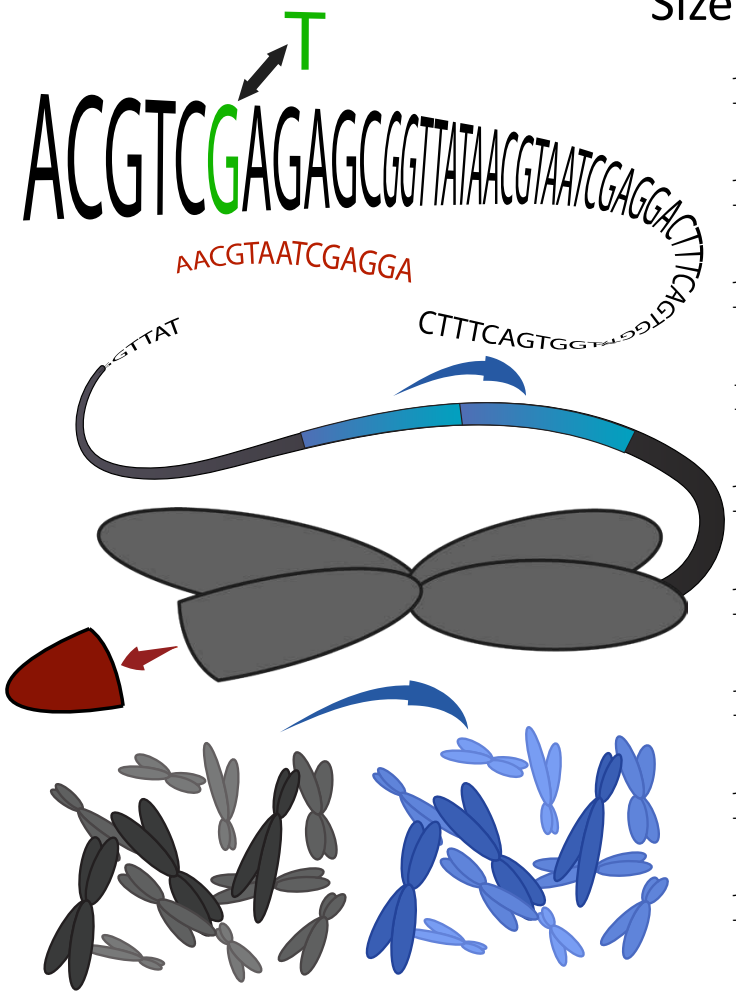
ITCR Meeting 3/5/21

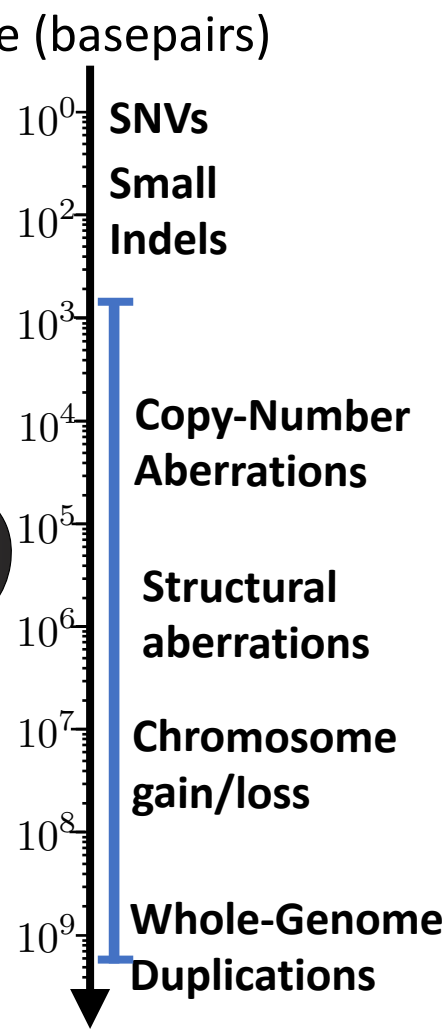
PRINCETON
UNIVERSITY

Cancer is an ***evolutionary*** process driven by **somatic** mutations over **all** genomic *scales*

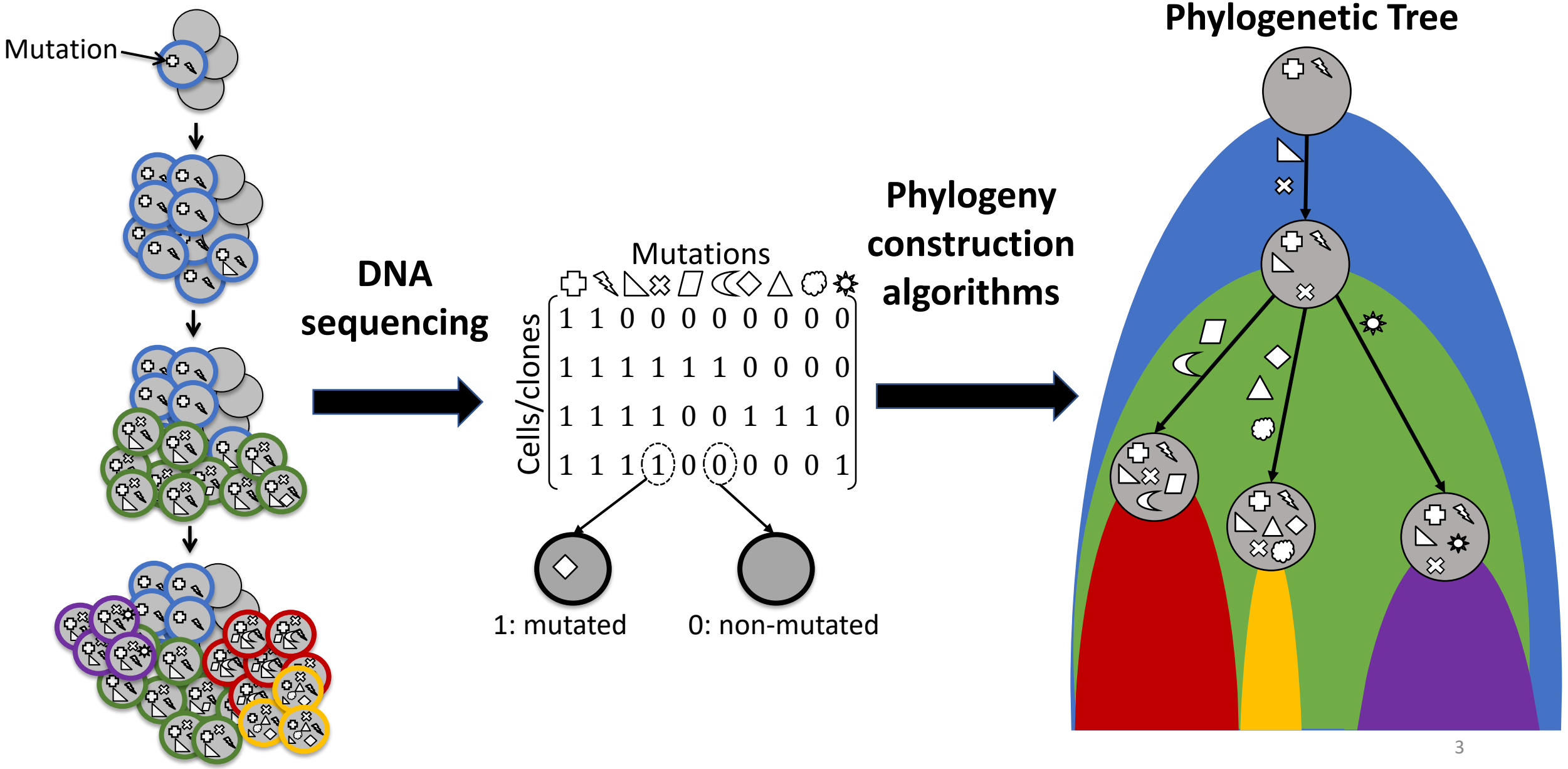


DNA Sequencing



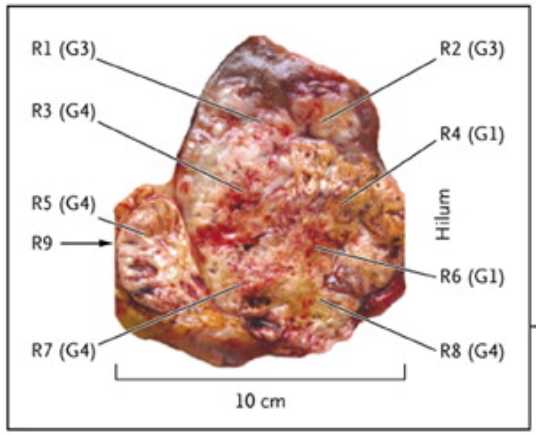


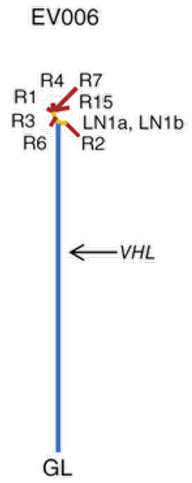
Cancer evolution is described by phylogenetic tree



Tumor Evolution is Key to Understanding and Treating Cancer

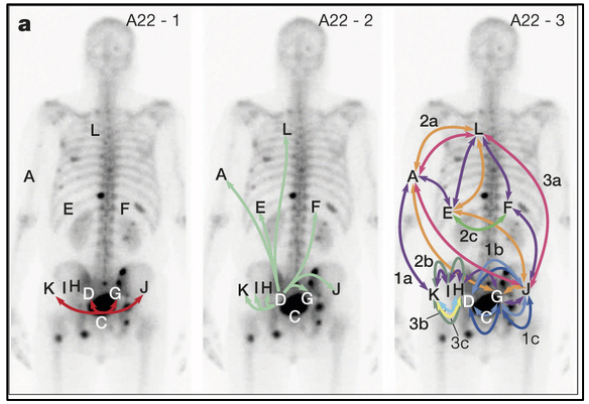
Regional intratumor heterogeneity





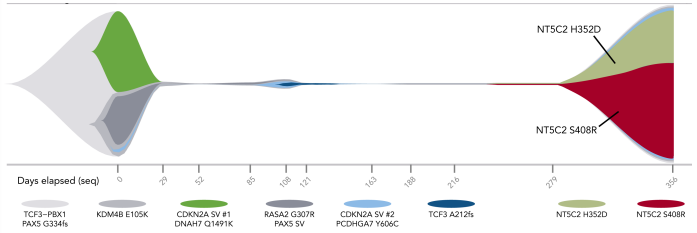
Gerlinger M et al. NEJM (2012).
Renal cell carcinoma

Metastasis

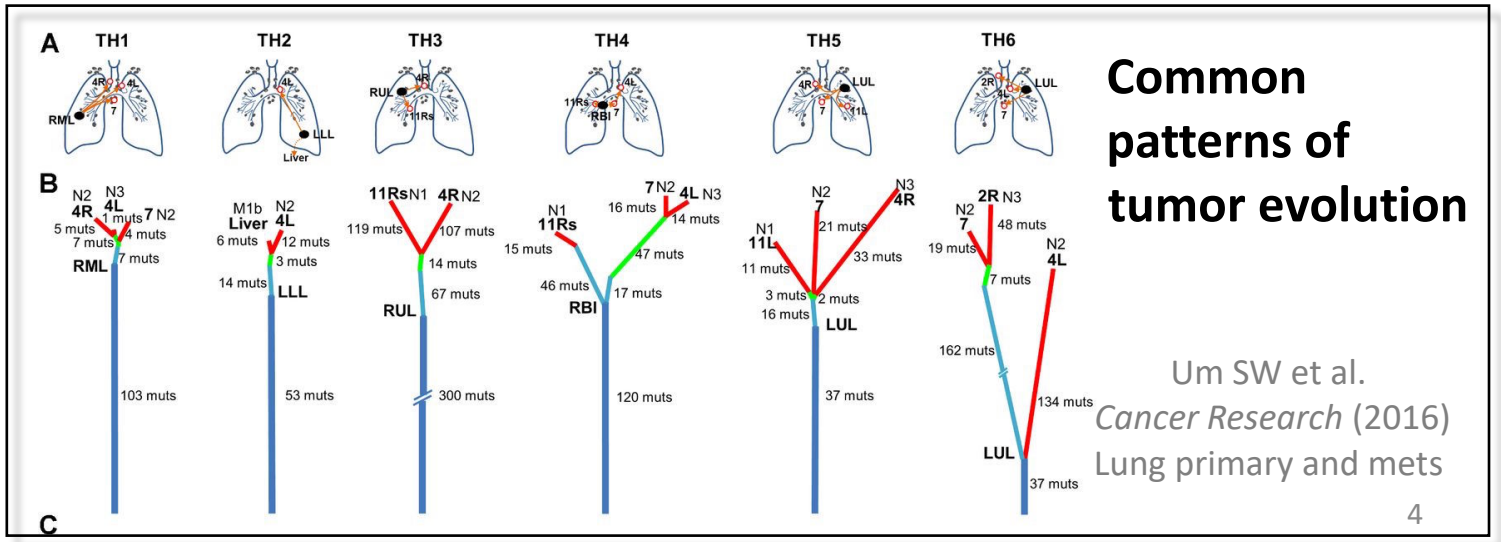


Gundem et al., *Nature* 2015

Longitudinal sequencing treatment response



Li et al., *Blood* 2020

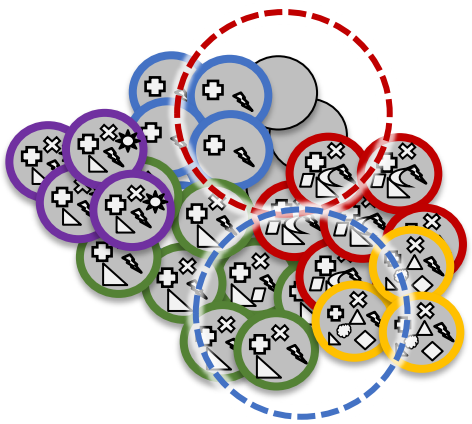


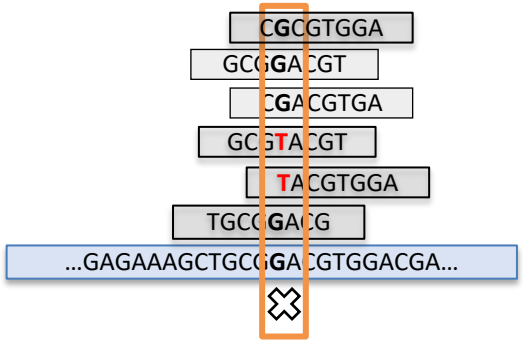
Common patterns of tumor evolution

Um SW et al.
Cancer Research (2016)
Lung primary and mets

Deriving **accurate** tumor phylogenies is **difficult**

1. Bulk sequencing: mixture of mutations





Variant allele frequency (VAF)

Mutation Frequencies

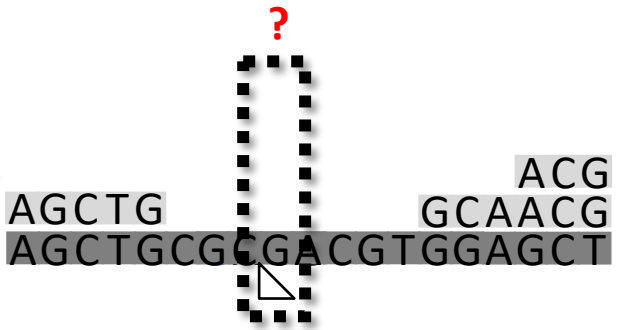
Mutations

Samples

$$\begin{bmatrix} 1 & 1 & .5 & .5 & .2 & .2 & 0 & 0 & 0 & .3 \\ 1 & 1 & .7 & .7 & .3 & .3 & .4 & .4 & .4 & 0 \end{bmatrix}$$

2. Single cell sequencing: high dropout / error rates

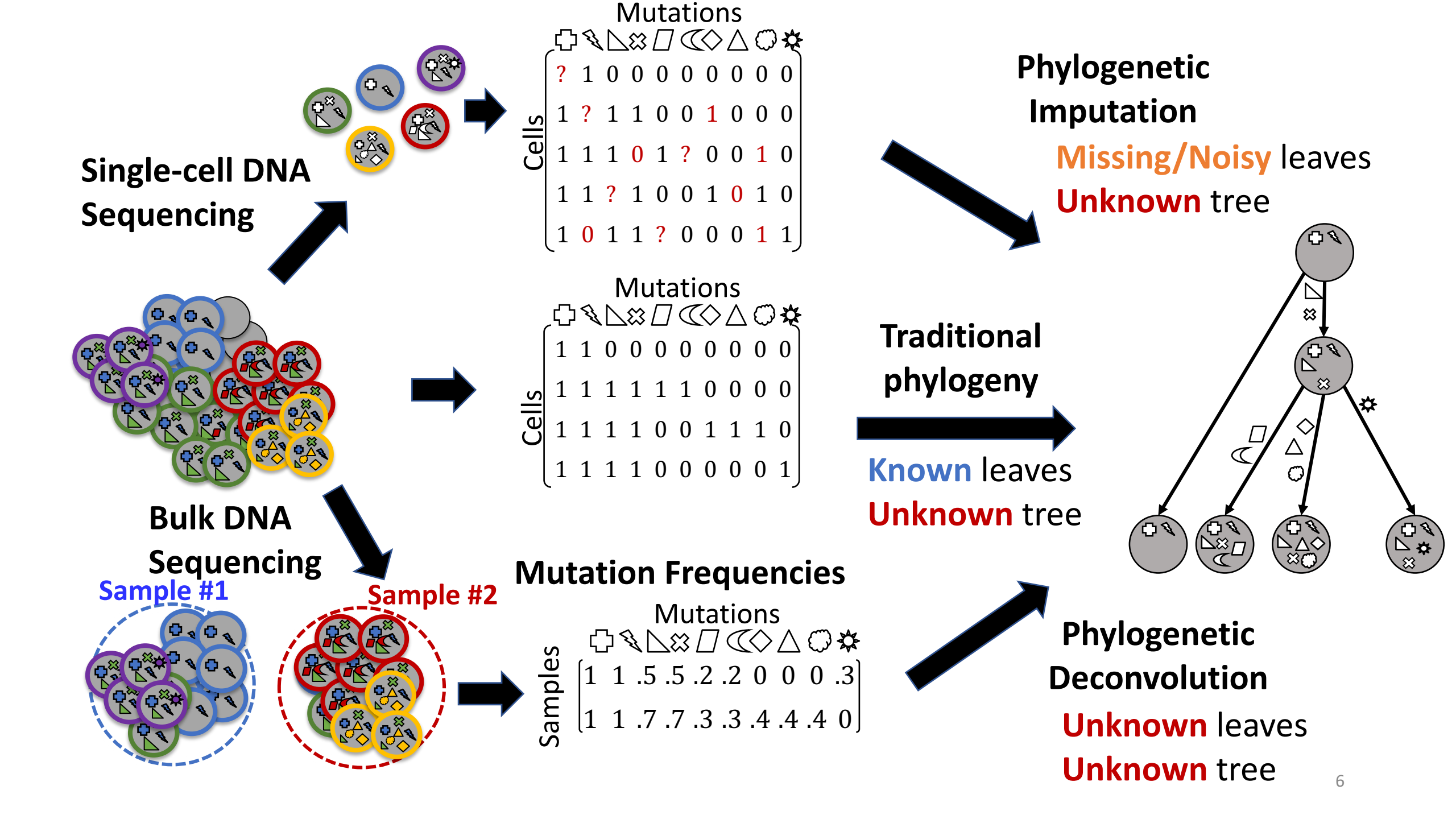
Amplification



Mutations

Cells

$$\begin{bmatrix} ? & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & ? & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & ? & 0 & 0 & 1 & 0 \\ 1 & 1 & ? & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & ? & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$


Single-cell DNA Sequencing
Mutations
Cells
? 1 0 0 0 0 0 0 0 0
1 ? 1 1 0 0 1 0 0 0
1 1 1 0 1 ? 0 0 1 0
1 1 ? 1 0 0 1 0 1 0
1 0 1 1 ? 0 0 0 1 1
Phylogenetic Imputation
Missing/Noisy leaves
Unknown tree
Bulk DNA Sequencing
Mutations
Cells
1 1 0 0 0 0 0 0 0 0
1 1 1 1 1 1 0 0 0 0
1 1 1 1 0 0 1 1 1 0
1 1 1 1 0 0 0 0 0 1
Traditional phylogeny
Known leaves
Unknown tree
Sample #1
Sample #2
Mutation Frequencies
Mutations
Samples
1 1 .5 .5 .2 .2 0 0 0 .3
1 1 .7 .7 .3 .3 .4 .4 .4 0
Phylogenetic Deconvolution
Unknown leaves
Unknown tree

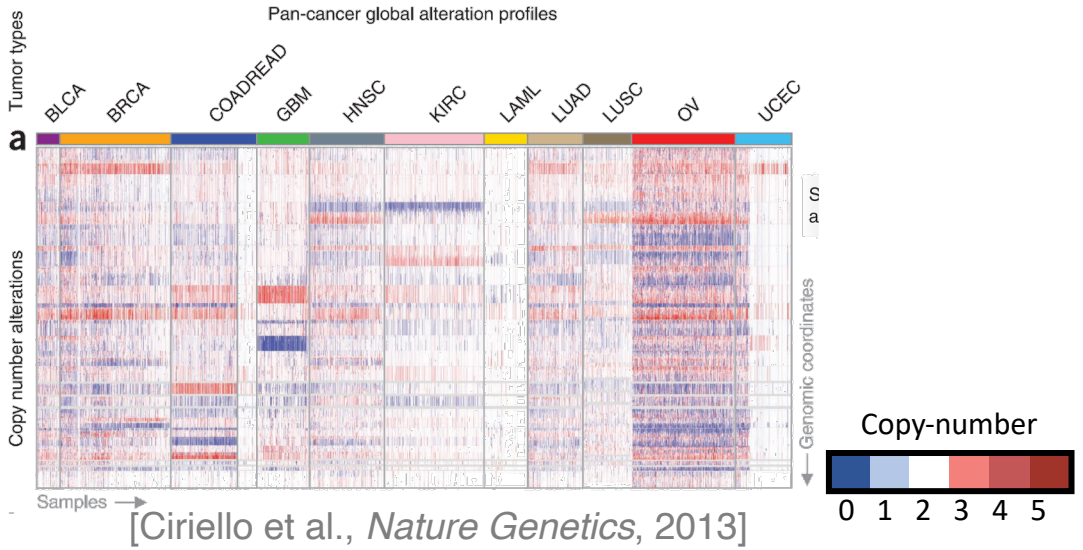
Deriving **accurate** tumor phylogenies is **difficult**

3. Aneuploid genomes

TCGA data

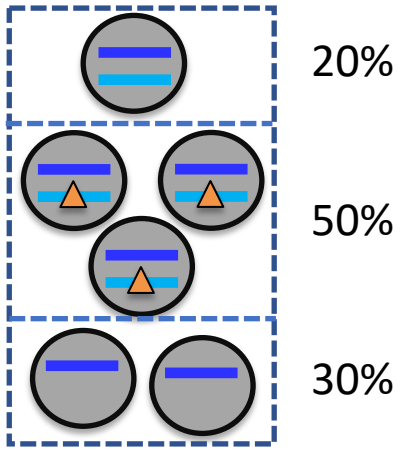
- ~23 focal CNAs (15Mb avg.) per sample
- ~9 arm-level CNAs per sample
- ~37% of tumors with WGDs

[Zack et al., *Nature Genetics*, 2013]



[Ciriello et al., *Nature Genetics*, 2013]

Copy number states and proportions



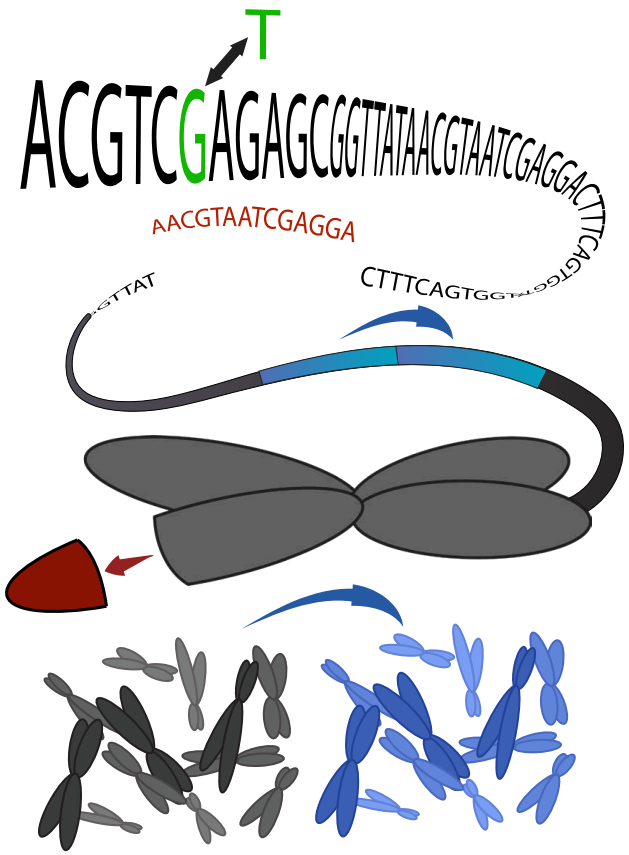
Evolutionary model

State 0 1 0?

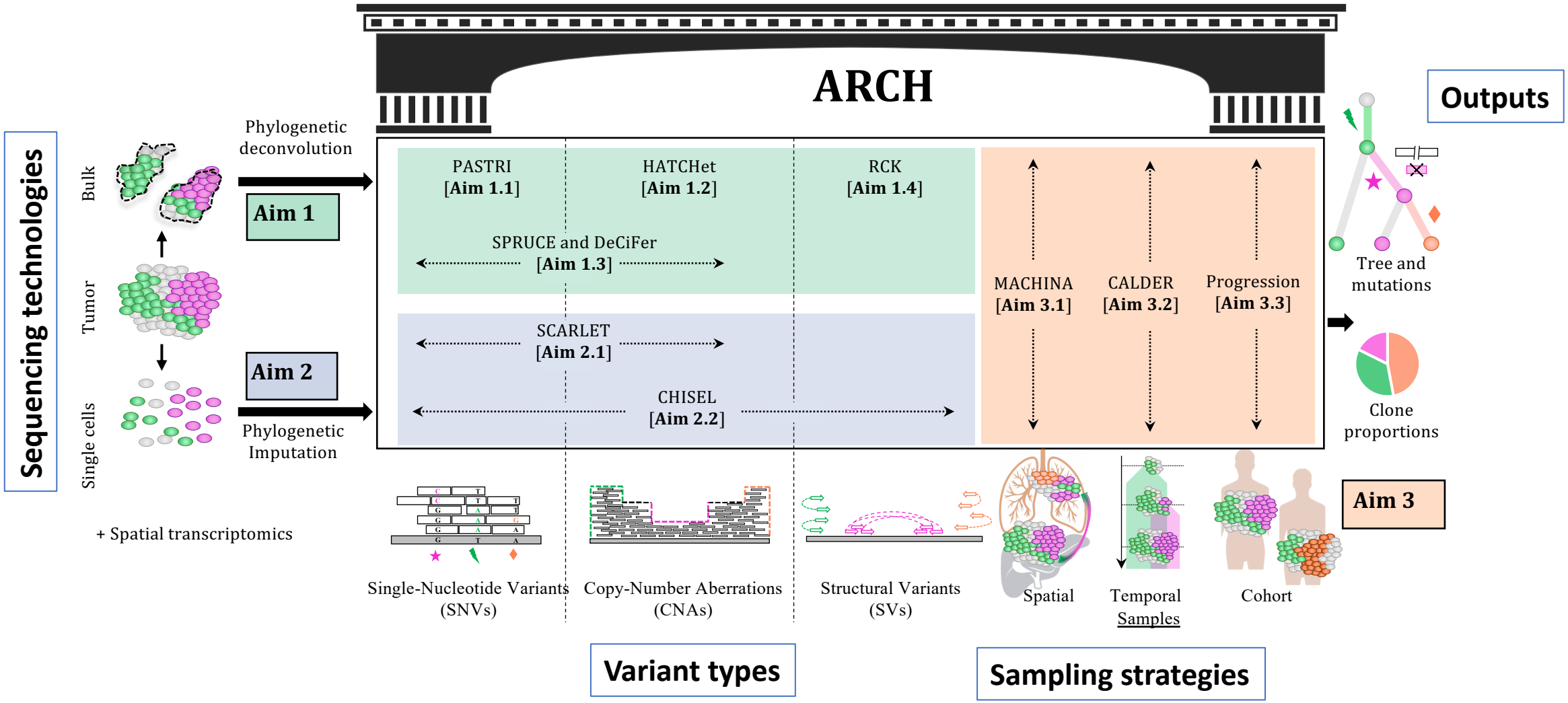
Motivation and Goal

- ***Unique challenges of cancer sequencing data*** (bulk vs. single-cell sequencing, copy number aberrations, etc.) are often not adequately modeled in standard phylogenetics software.
- Numerous ***specialized*** algorithms and tools have been developed recently to address ***specific challenges*** in tumor heterogeneity/evolution but tools are scattered and not always accessible

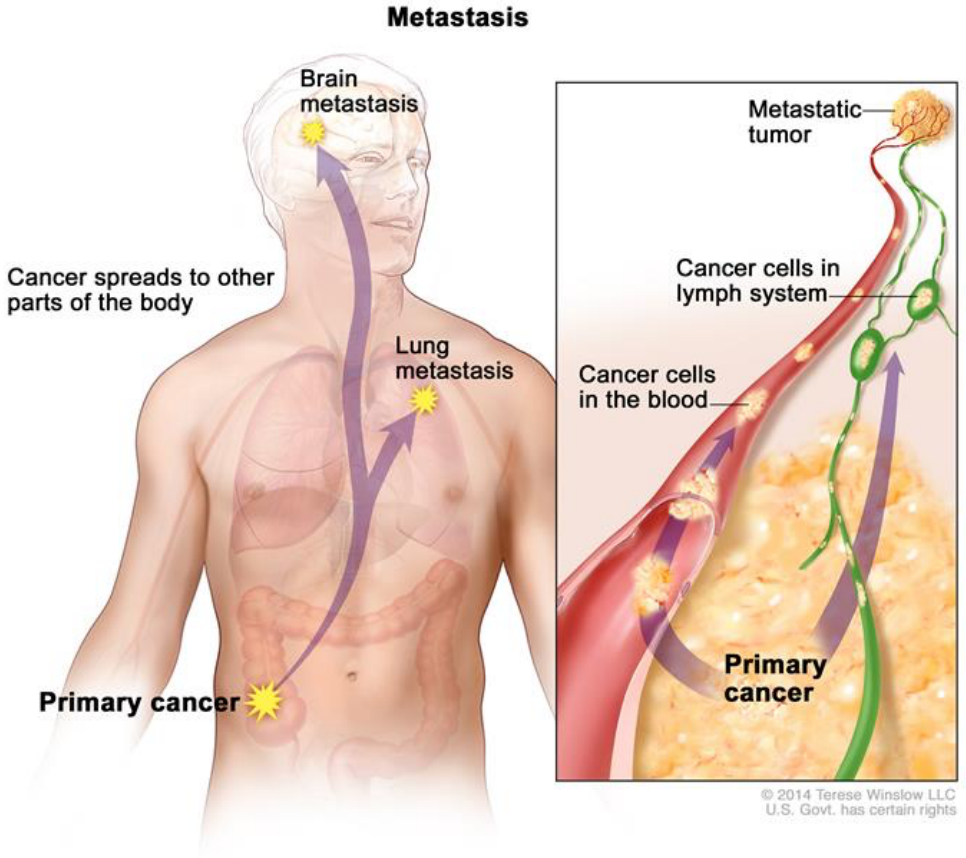
Goal: Develop a ***comprehensive, robust and user-friendly software package*** for tumor heterogeneity and cancer evolution studies that integrates analyses across diverse sequencing technologies, sampling strategies, genomic scales, etc.

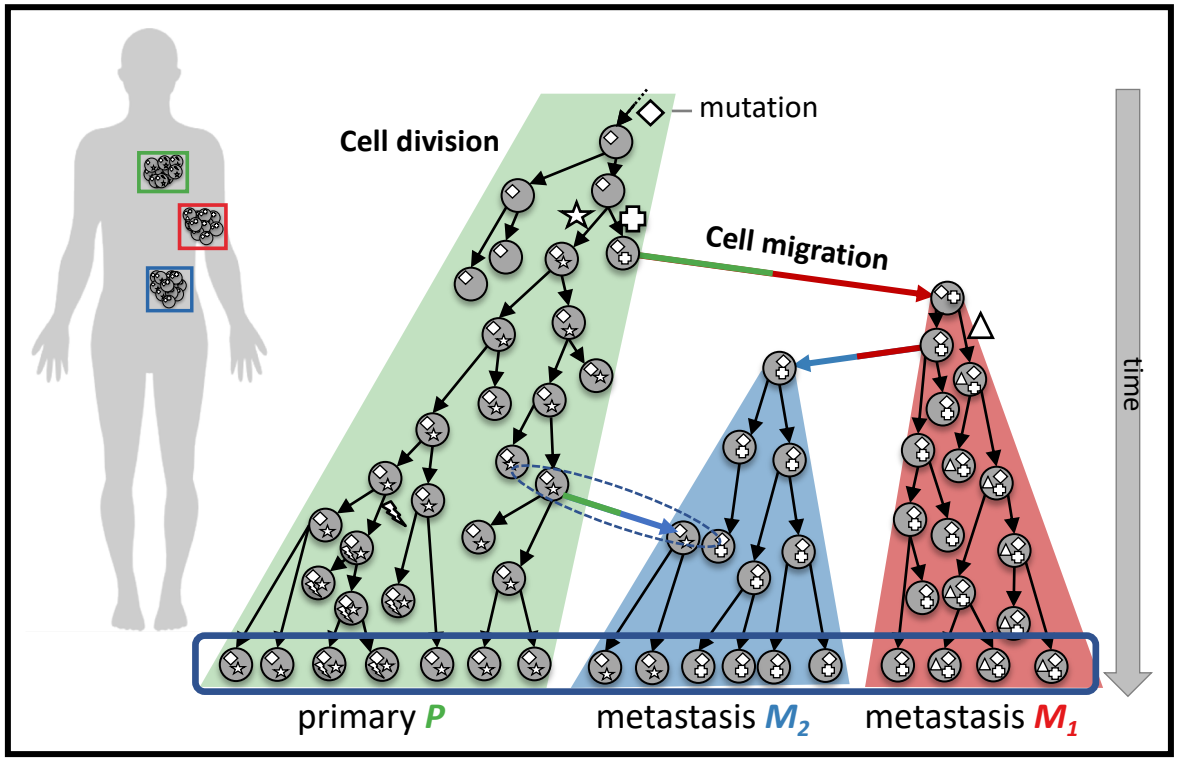


ARCH: A robust toolkit for tumor heterogeneity and evolution



How do cancer cells metastasize to seed tumors at distant locations?





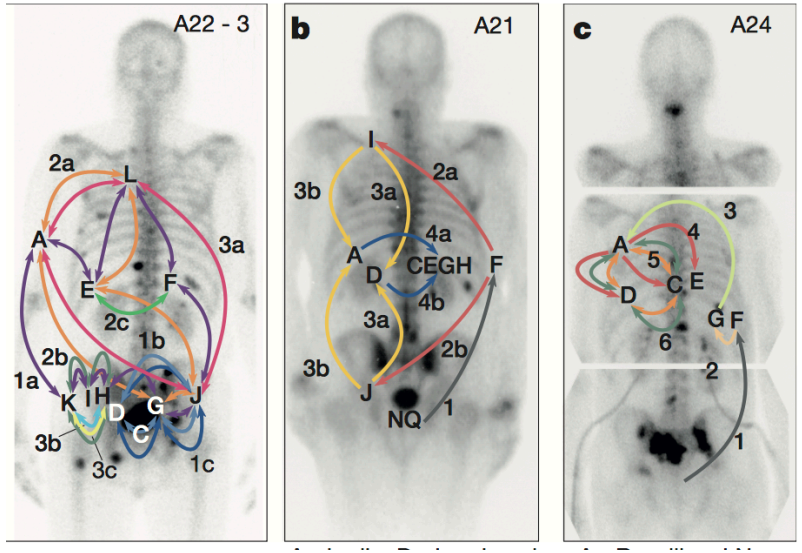
Traditional view: Each metastasis results from a single migration event (**monoclonal**)

Recent studies: Multiple cellular migrations between anatomic sites (**polyclonal**)

Polyclonal/Multiclonal/Re- Seeding of Metastases?

Prostate Cancer

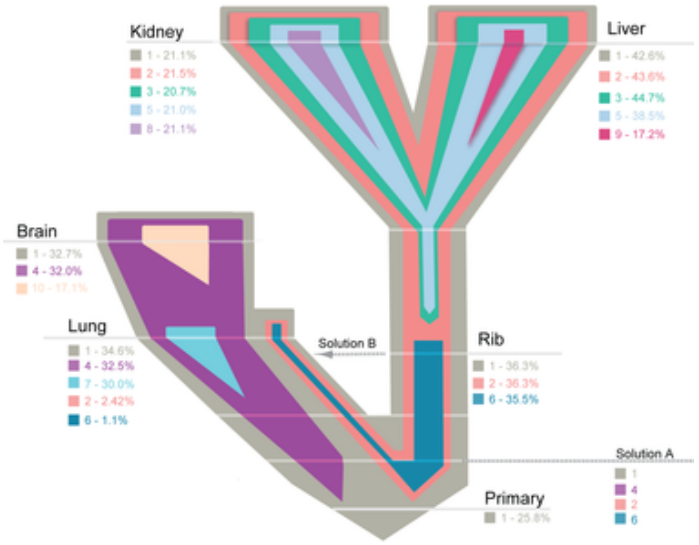
[Gundem et al. (2015) *Nature*]



*"..in all five patients with **poly-clonal seeding**, subclones…were found to have **re-seeded** multiple sites."*

Breast Cancer

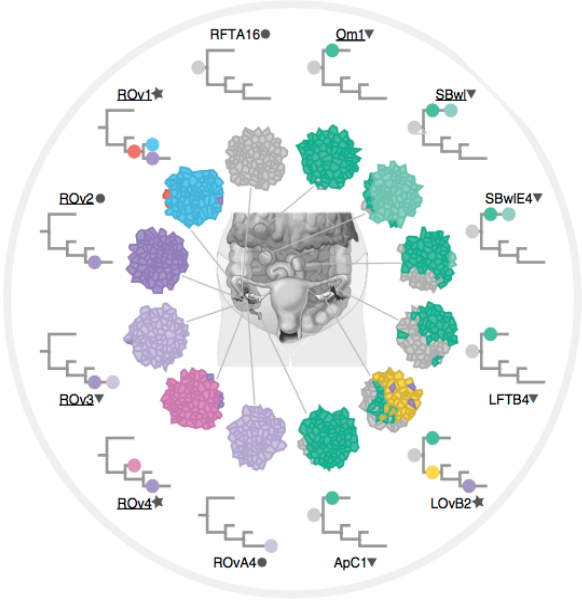
[Hoadley KA, et al. (2016) *PLOS Medicine*]



*"**multiclonal seeding** from the primary tumor to the metastases can occur…"*

Ovarian Cancer

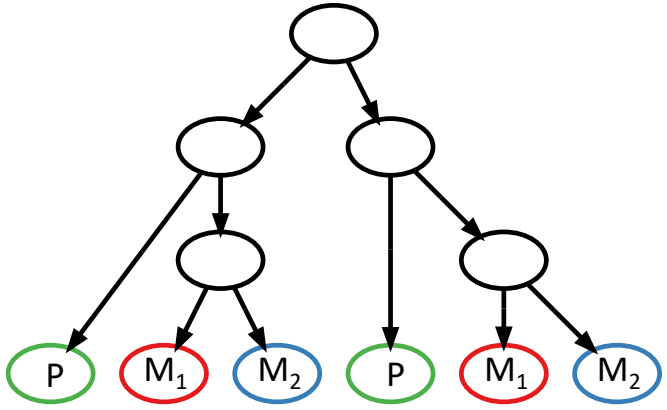
[McPherson et al. (2016) *Nature Genetics*]



*"multiple samples composed of divergent lineages is concordant with **polyclonal reseeding** or **polyclonal migration**…"*

Migration Patterns in Metastasis

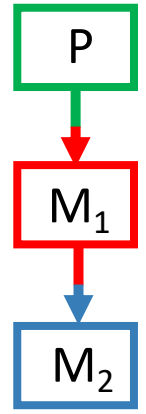
Cell/clone tree *T*



MACHINA

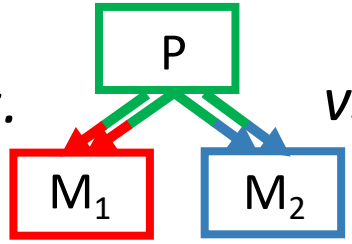
Anatomical locations of ancestral cells is **unknown**

Migration graph *S*



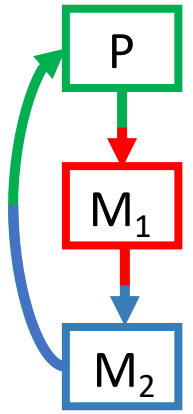
Monoclonal

vs.



Polyclonal

vs.



Reseeding

1. Cell/clone tree ***T*** does ***not*** uniquely determine migration pattern ***S***.
2. Uncertainty in tree ***T*** for bulk sequencing→ ***substantial*** uncertainty in migration pattern ***S***.

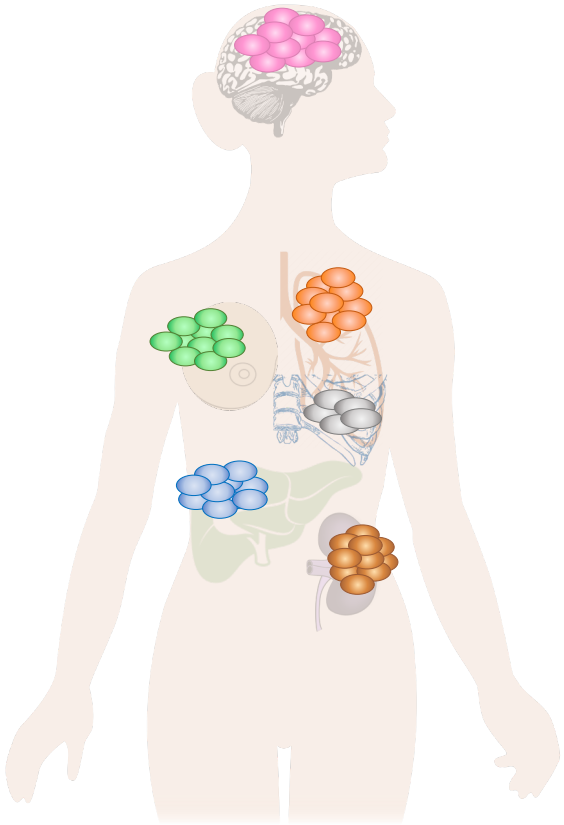
Mohammed El-Kebir

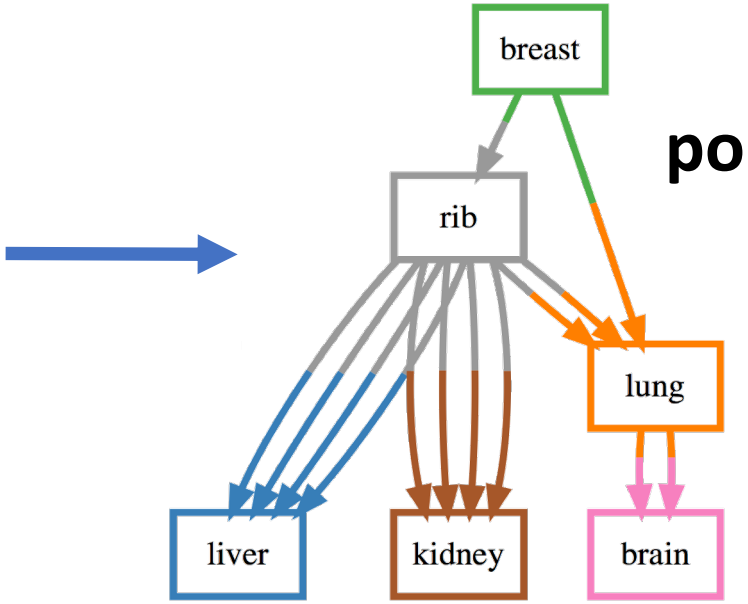
Gryte Satas

[El-Kebir et al. *Nature Genetics* 2018]

Inferring Migration Patterns in Metastasis

Metastatic Breast Cancer



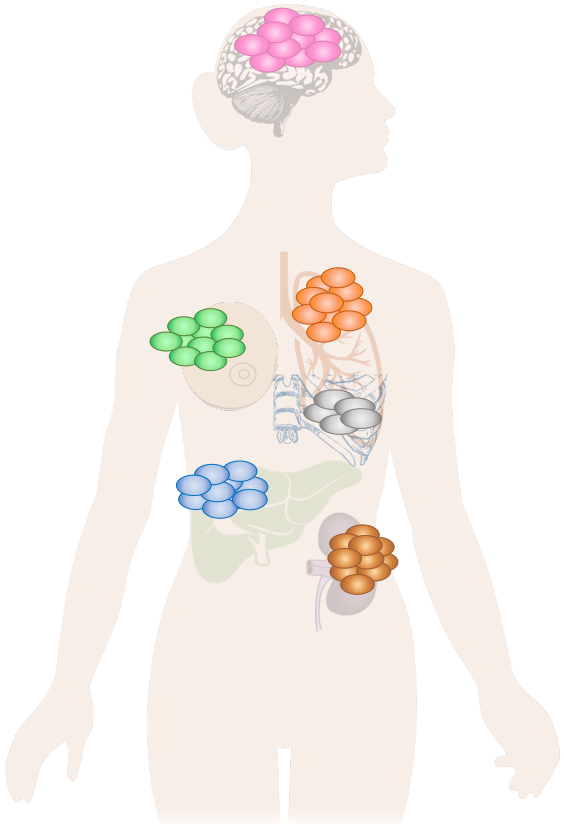


14 migrations
polyclonal seeding

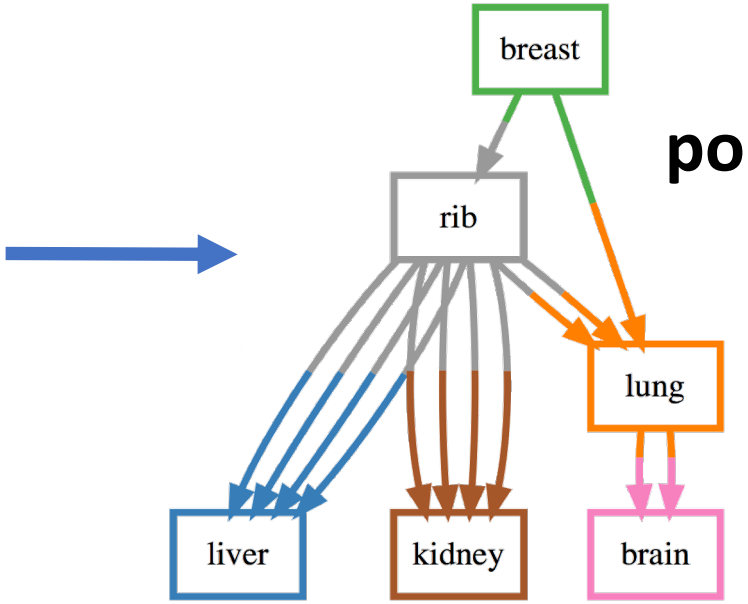
Hoadley KA,, et al.
PLOS Medicine (2016)

Parsimonious Migration Patterns in Metastasis

Metastatic Breast Cancer



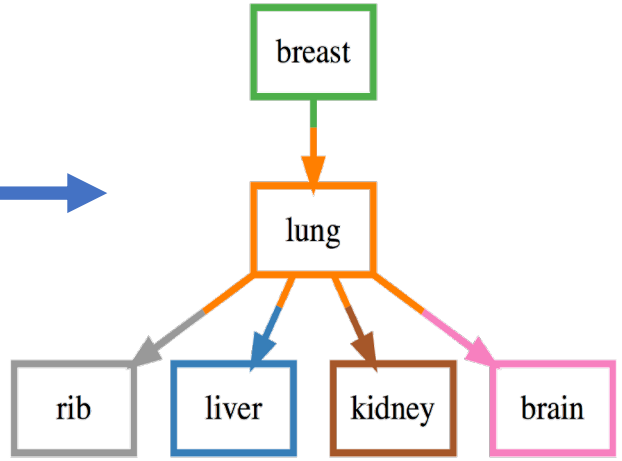
Hoadley KA,, et al.
PLOS Medicine (2016)



14 migrations
polyclonal seeding

MACHINA

5 migrations
monoclonal seeding

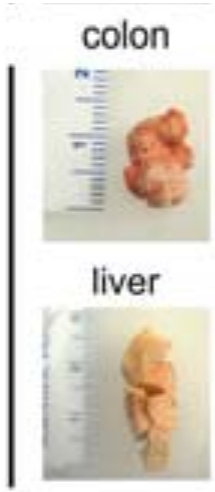


Mohammed El-Kebir

Gryte Satas

[El-Kebir, et al. *Nature Genetics* 2018]

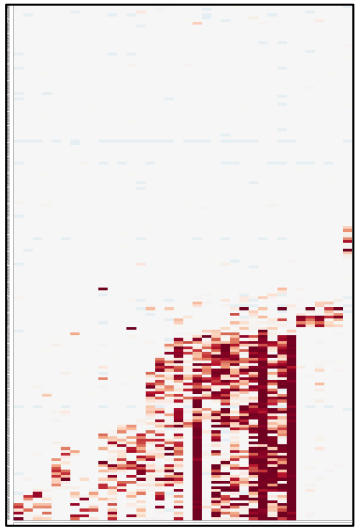
Single-cell sequencing to the rescue?



WES of 186 cells from metastatic CRC patient (matched primary + liver met)

[Leung et al. *Genome Research* 2017]

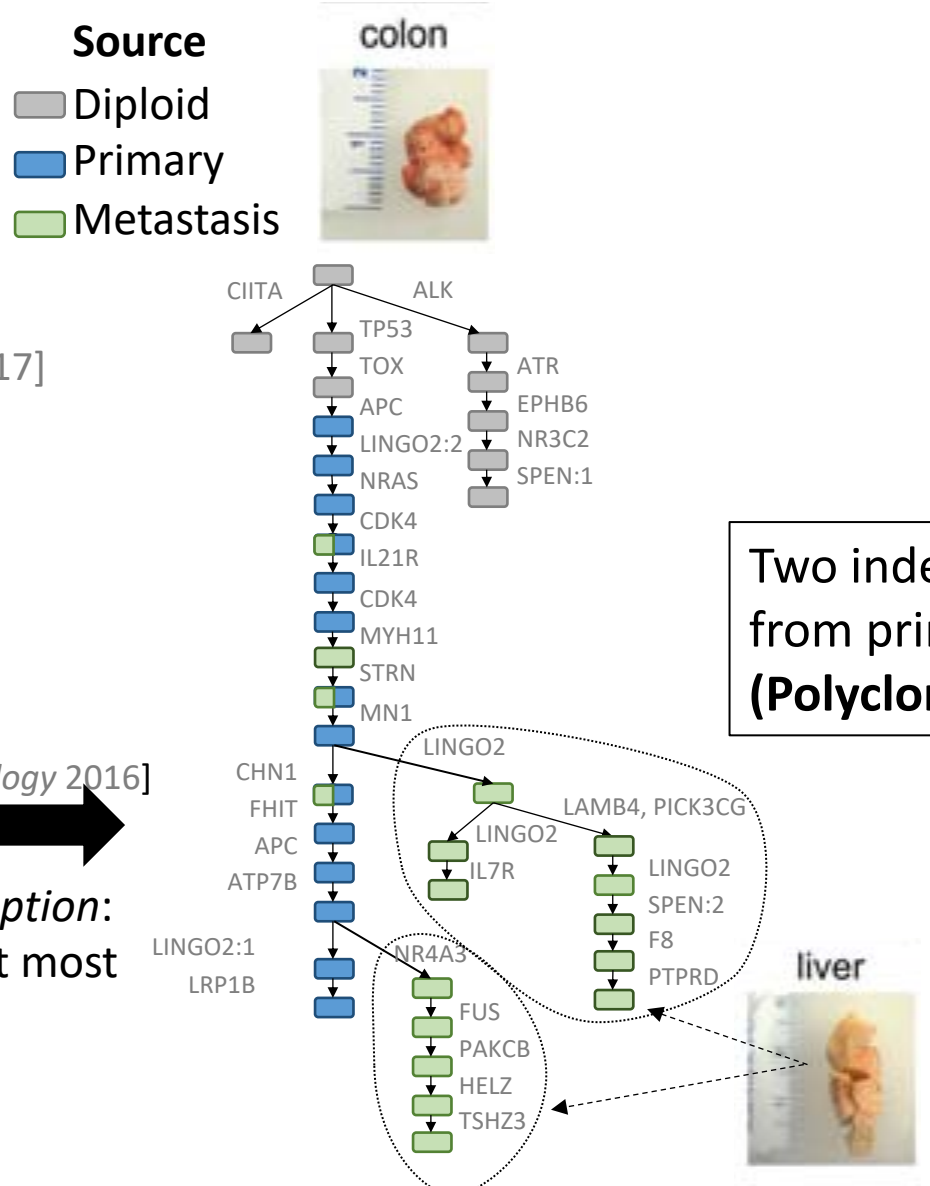
Single nucleotide mutations



SCITE

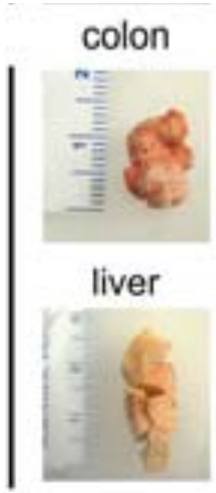
[Jahn et al., *Genome Biology* 2016]

Infinite sites assumption: Mutation occurs at most once



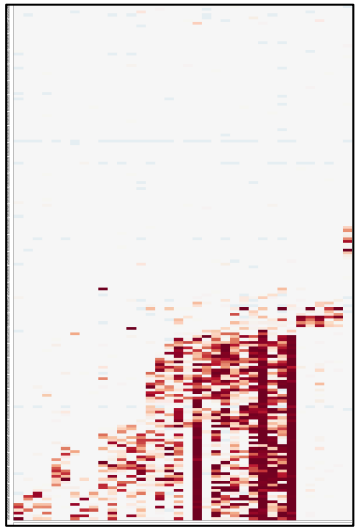
Two independent cellular migrations from primary tumor to metastasis!
(Polyclonal seeding)

Copy number data inconsistent with SNV-based phylogeny



WES of 186 cells from metastatic CRC patient (matched primary + liver met)
[Leung et al. *Genome Research* 2017]

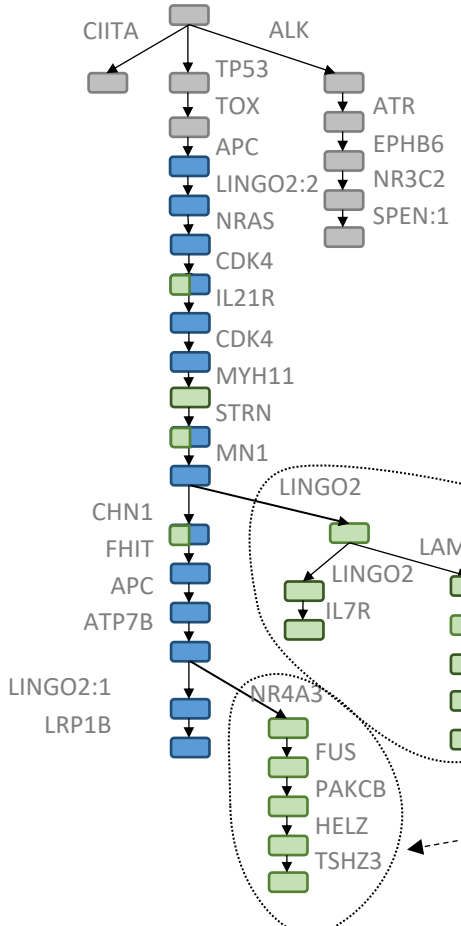
Single nucleotide mutations



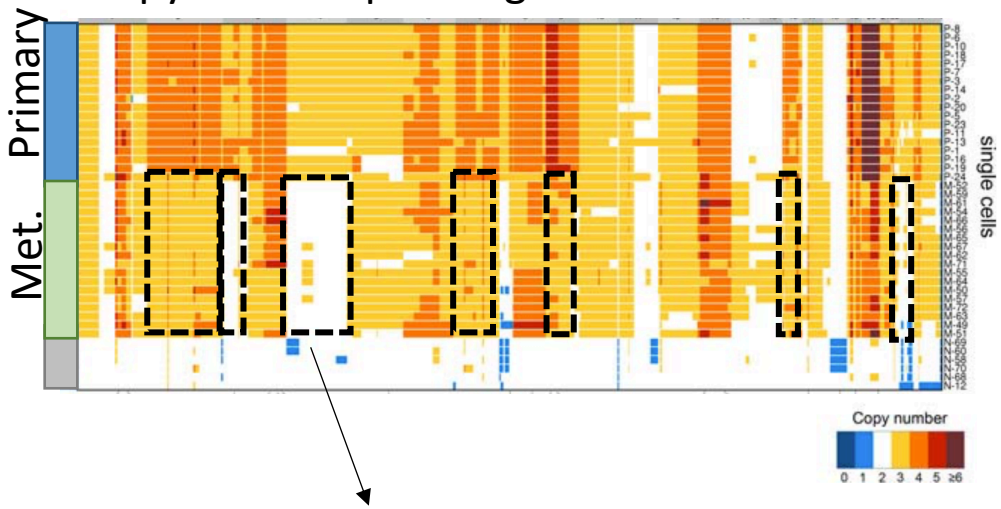
SCITE

Infinite sites assumption: Mutation occurs at most once

Source
Diploid
Primary
Metastasis

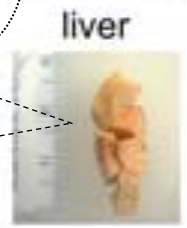


Copy-number profiling from WGS of 42 cells



Multiple metastasis-specific CNAs
All occurred twice independently?

Two independent cellular migrations from primary tumor to metastasis!
(Polyclonal seeding)



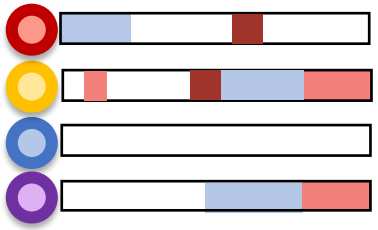
SCARLET: Construct Tumor Evolution across Genomic Scale

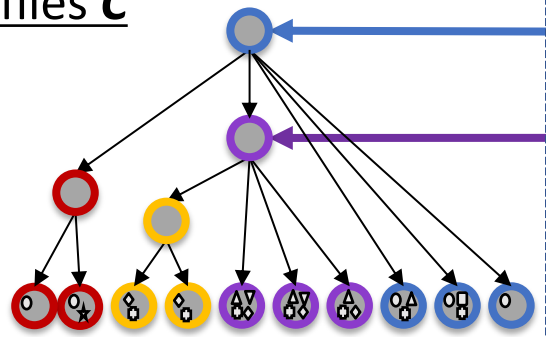
(Single-Cell Analysis for Reconstructing Loss-supported Evolution of Tumors

Input:

Copy-number tree ***S***

Copy-number profiles ***C***

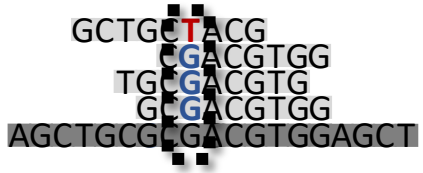




Single-cell targeted sequencing/WES/WGS

Mutation Matrix ***B***

Probabilistic model for sequencing data



Pr(1 | # Variant Reads, # Total Reads)

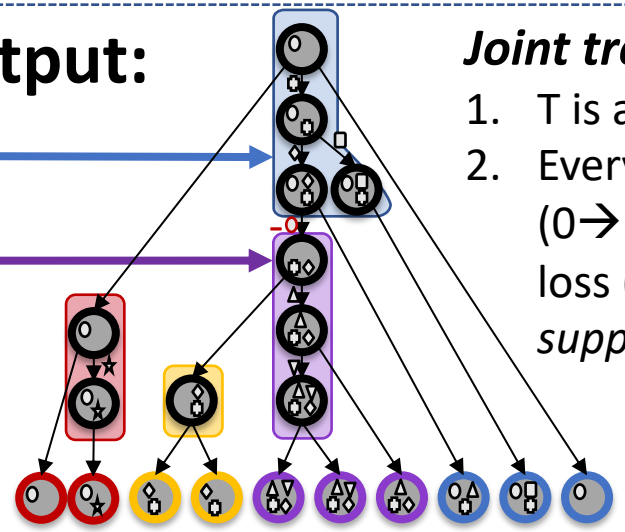
SCARLET

Mutations

Cells

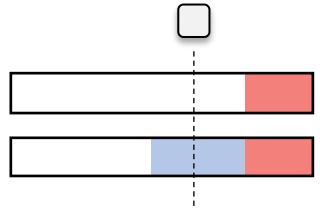
1	0	0	0	0	?	0
0	1	0	0	0	0	0
1	1	?	0	0	0	0
1	0	0	1	?	0	0
1	0	0	1	0	0	0
0	0	1	0	1	0	?
0	1	0	0	1	0	0
?	1	1	0	1	1	0
?	1	0	0	0	0	0
0	1	0	0	1	0	1

Output:



Joint tree T satisfying:

1. T is a *refinement* of S
2. Every mutation: gain (0→1) *at most once* and loss (1→ 0) only when *supported*



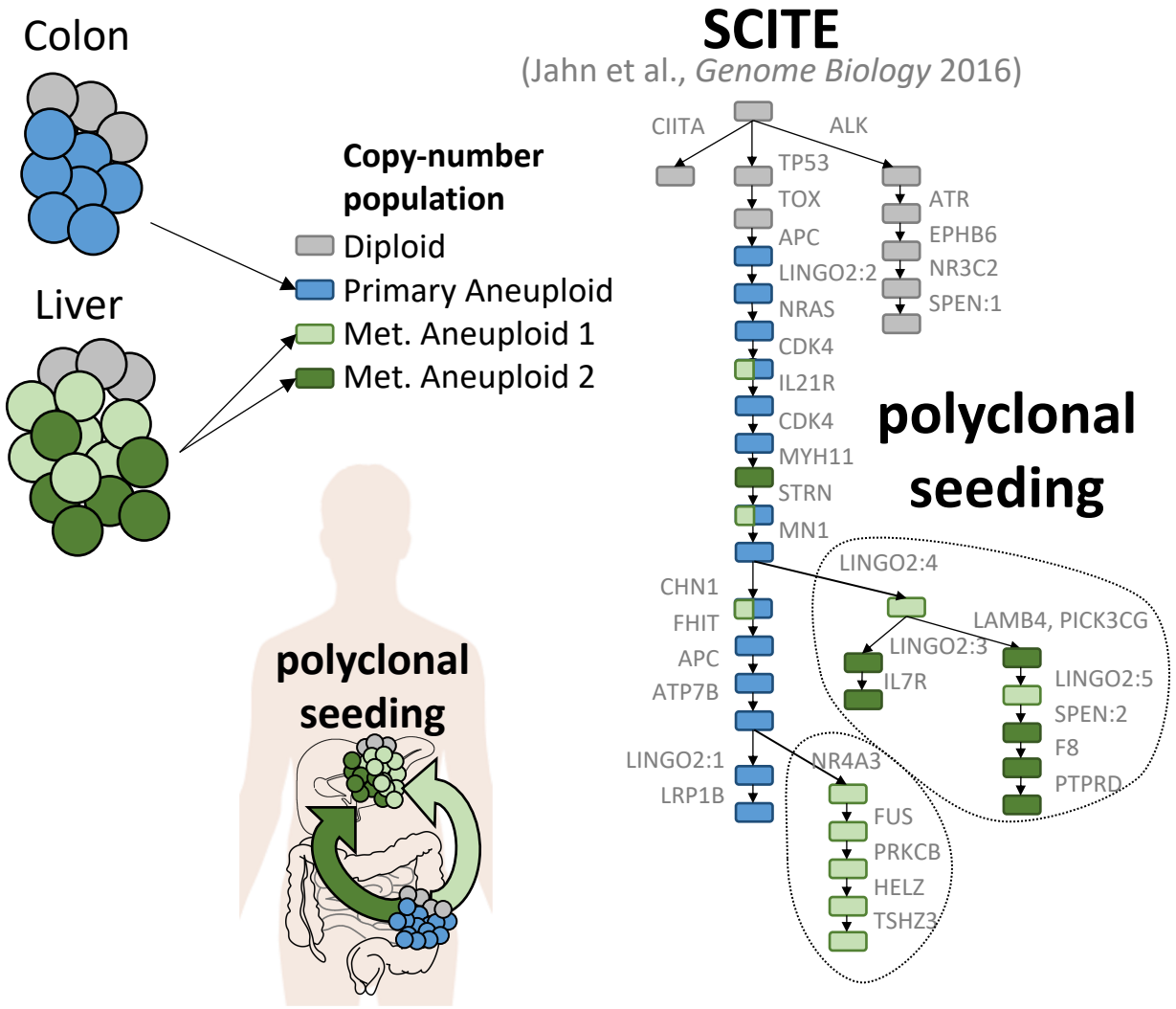
1	0	0	0	0	**0**	0
1	1	0	0	0	0	0
1	1	**1**	0	0	0	0
1	0	0	**1**	0	0	0
1	0	0	1	0	0	0
0	**1**	1	0	1	0	**0**
0	1	**1**	0	1	**1**	0
0	1	1	0	1	1	0
0	1	0	0	**1**	0	0
0	1	0	0	1	0	1

Gryte Satas

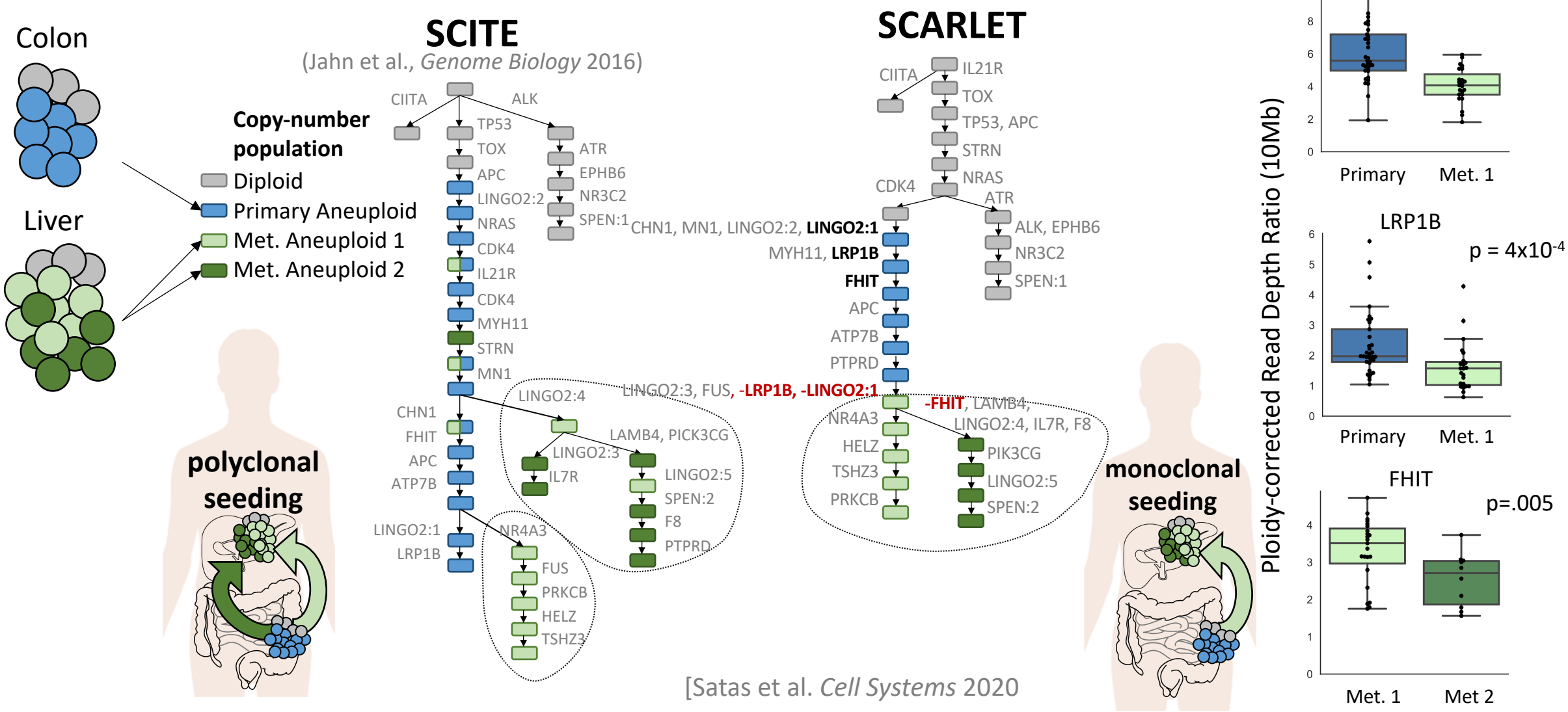
[Satas et al. *Cell Systems* 2020, RECOMB 2020]

Single-nucleotide mutations → *polyclonal seeding* of liver metastasis

(Leung et al., *Genome Research*, 2017)



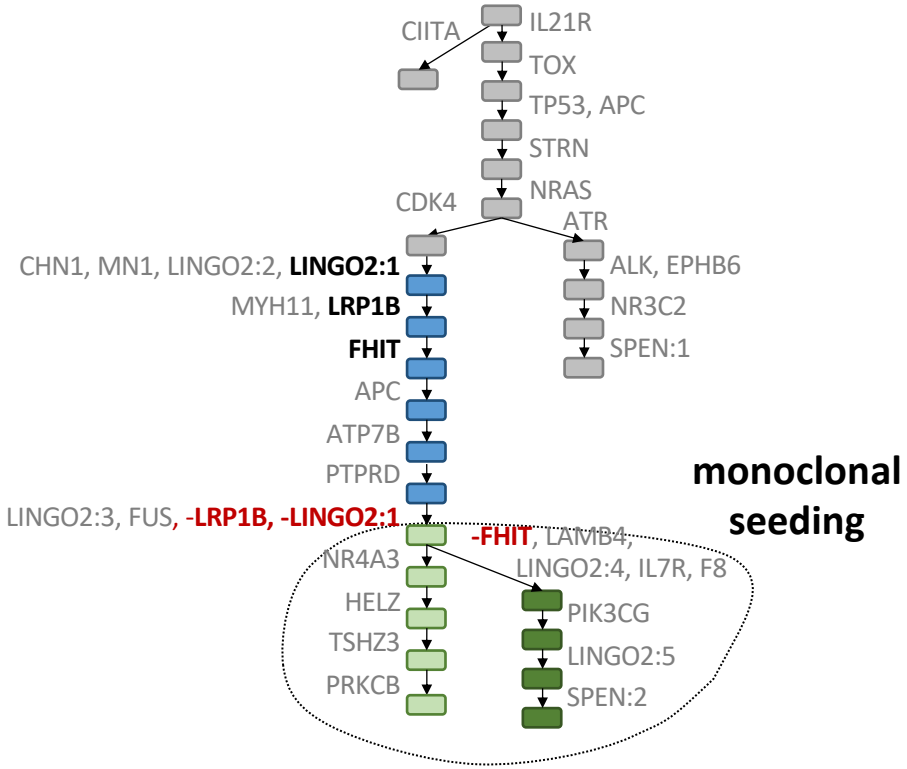
SCARLET: single-nucleotide + copy number mutations → *monoclonal seeding* of metastasis

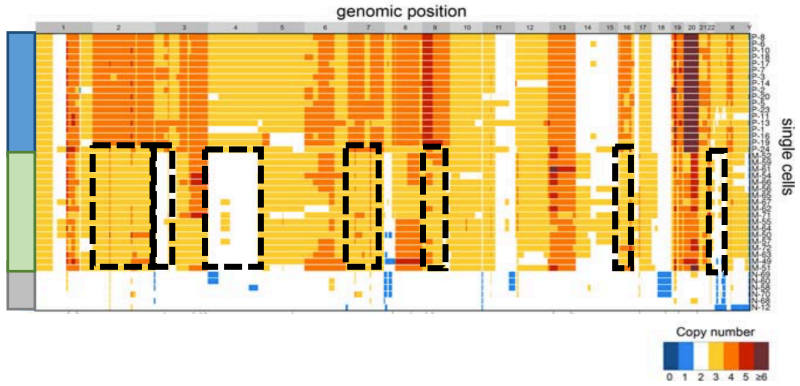


Summary

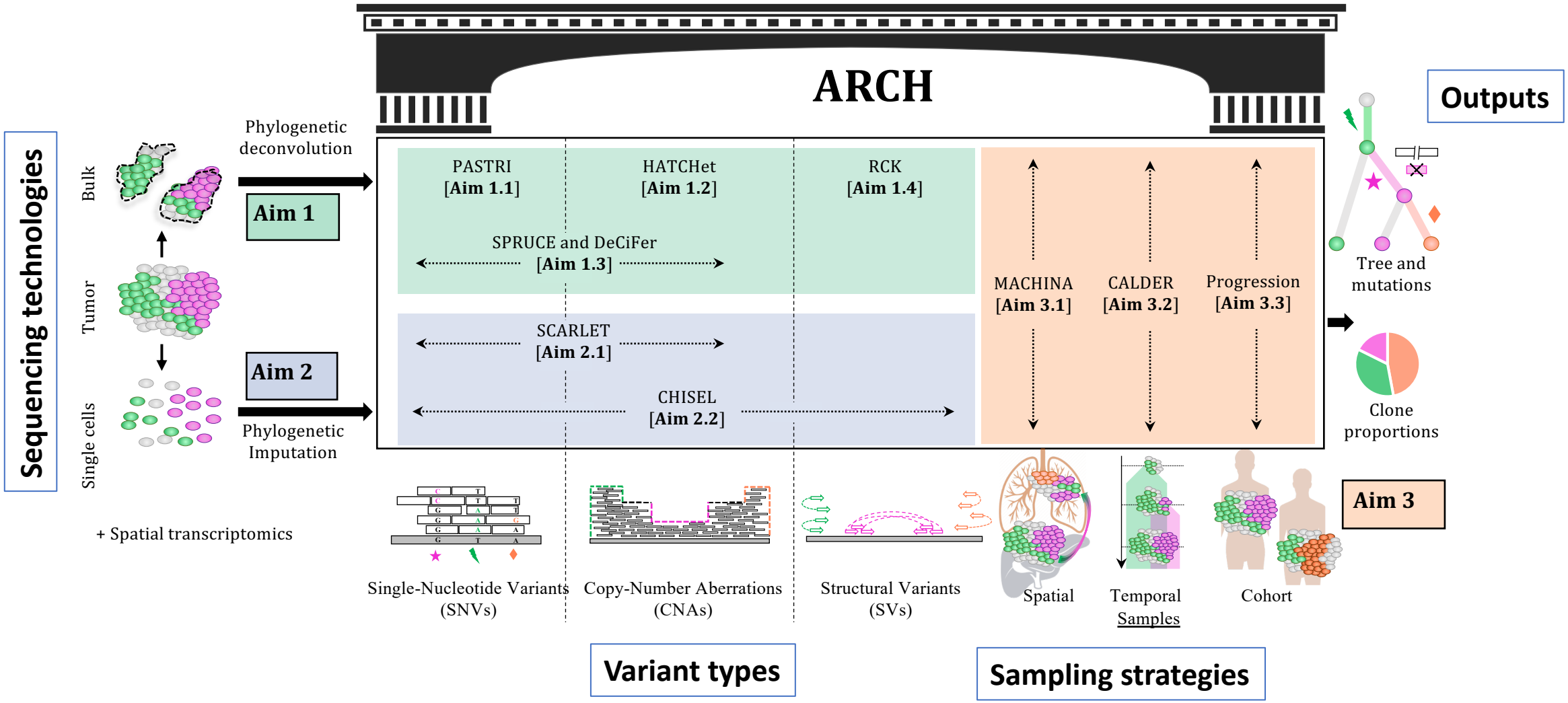
Inference of metastatic seeding patterns (monoclonal vs. polyclonal vs. reseeding of metastases) requires:

1. Deriving reliable phylogenies from bulk or single-cell sequencing data, each with unique challenges
2. Analyzing mutations across genomic scale; e.g. interactions between single-nucleotide mutations and copy number aberrations





A robust toolkit for tumor heterogeneity and evolution

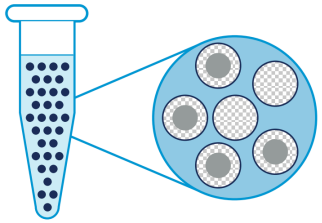


Ultra low-coverage whole-genome single-cell sequencing

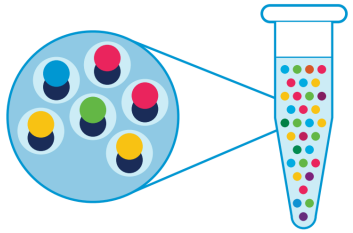
Chromium Single Cell CNV Solution

10x GENOMICS

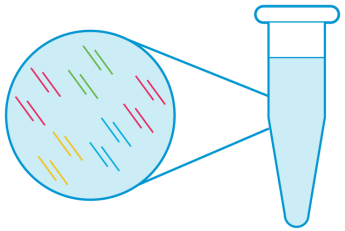
Single cells



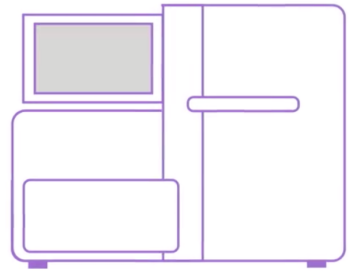
Barcoded cells



Barcoded library

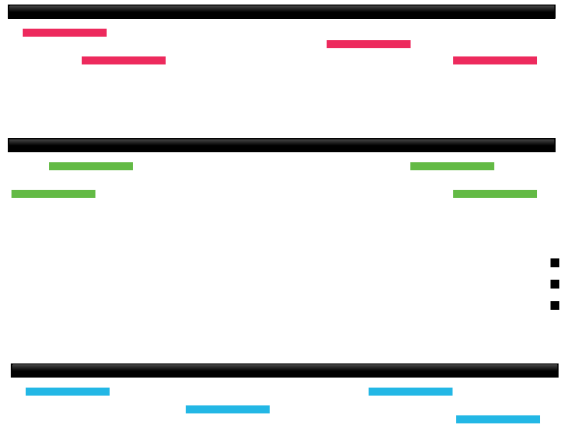


DNA Sequencing



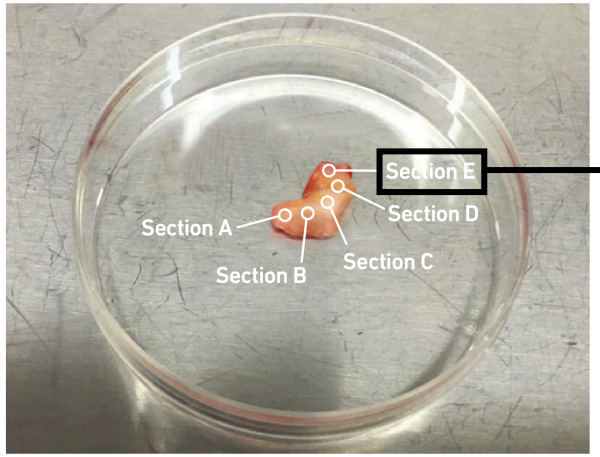
Ultra-low (~.03x) coverage per single cell from ~2000 cells

Aligned reads

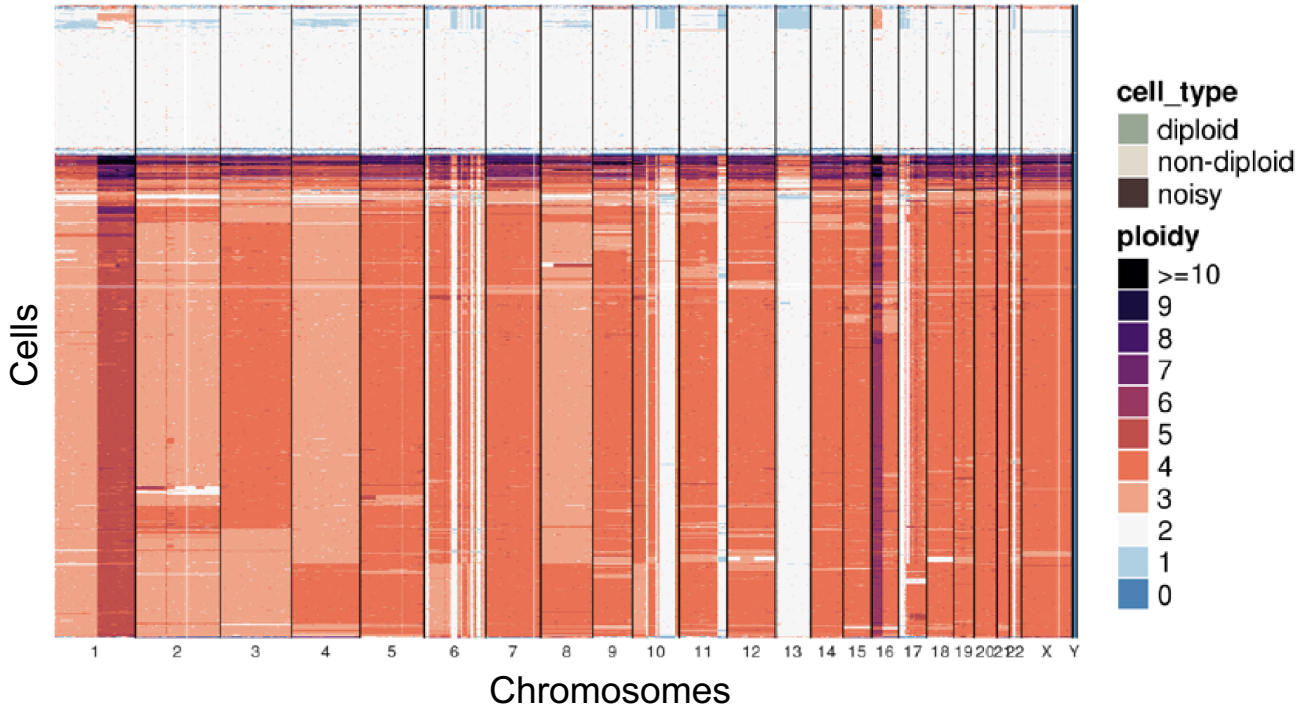


Other technologies: **DOP-PCR** [Navin et al. Nature 2011] and **DLP/DLP+** [Zahn et al. Nat. Methods 2017]

Single-cell copy-number analysis in breast cancer



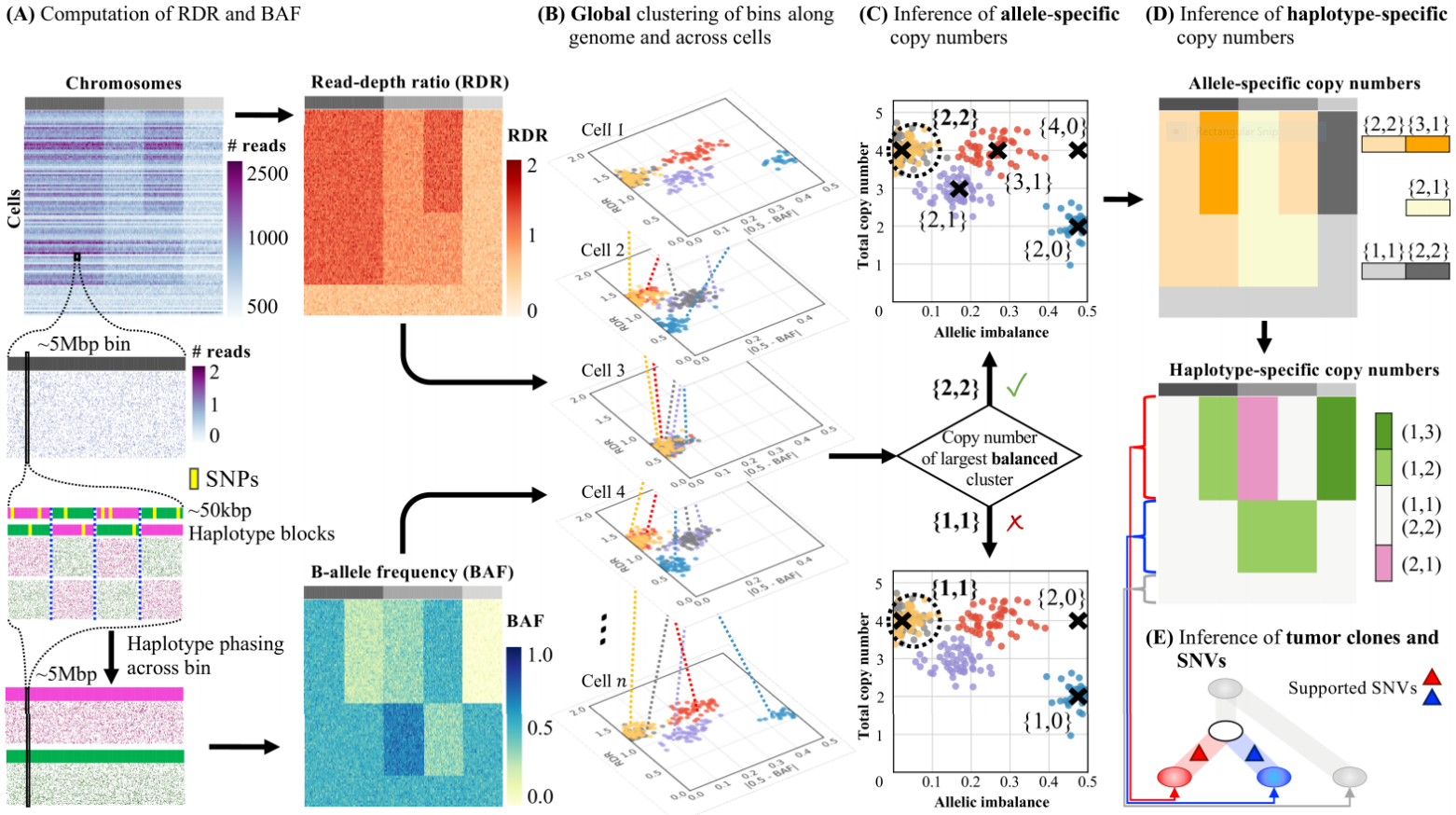
[*Assessing Tumor Heterogeneity with Single Cell CNV* www.10xgenomics.com/solutions/single-cell-cnv/]



- ~2000 cells from 5 sections of a breast tumor sequenced with 10X CNV Solution
- Cell Ranger DNA inferred **copy numbers** and identified multiple clusters of cells

CHISEL: Copy-number Haplotype Inference in Single-cells using Evolutionary Links

First method to infer **allele-specific copy numbers** from **ultra low-coverage DNA sequencing WGS**



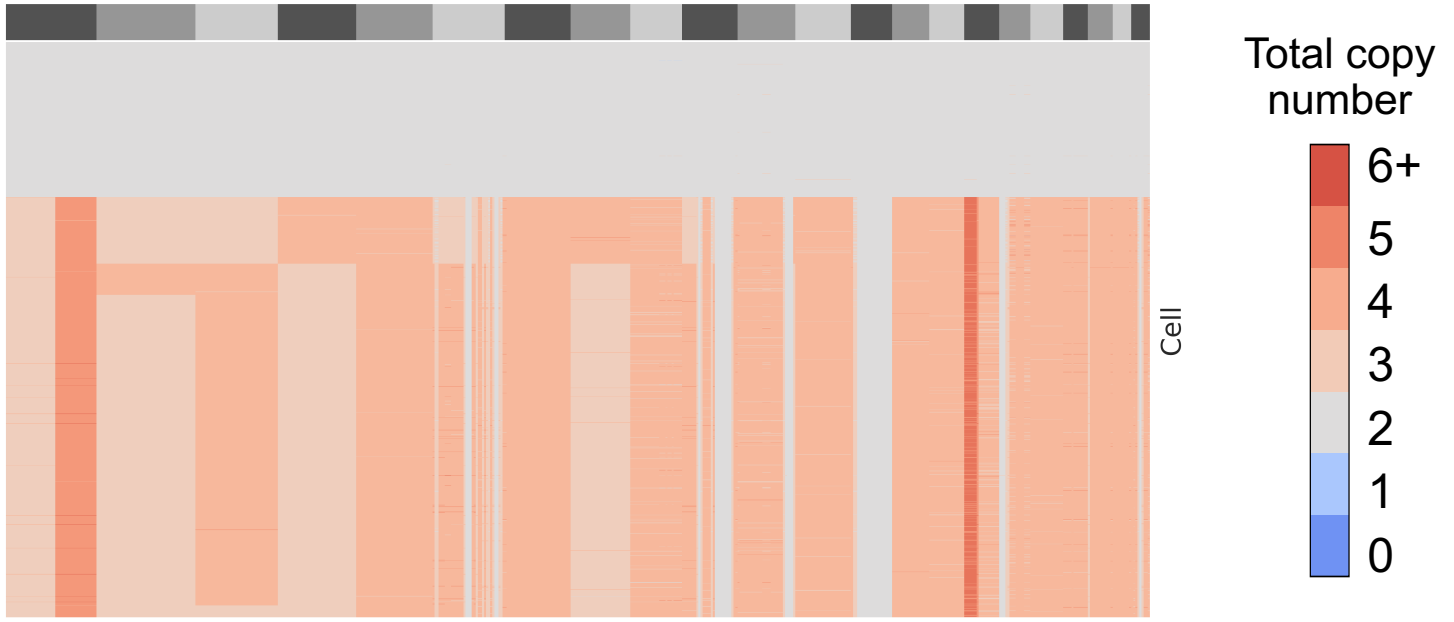
[S. Zaccaria and B. Raphael, *Nature Biotechnology* (2020)]

Simone Zaccaria

CHISEL reveals events **invisible to** total **copy number analysis**

Total copy number

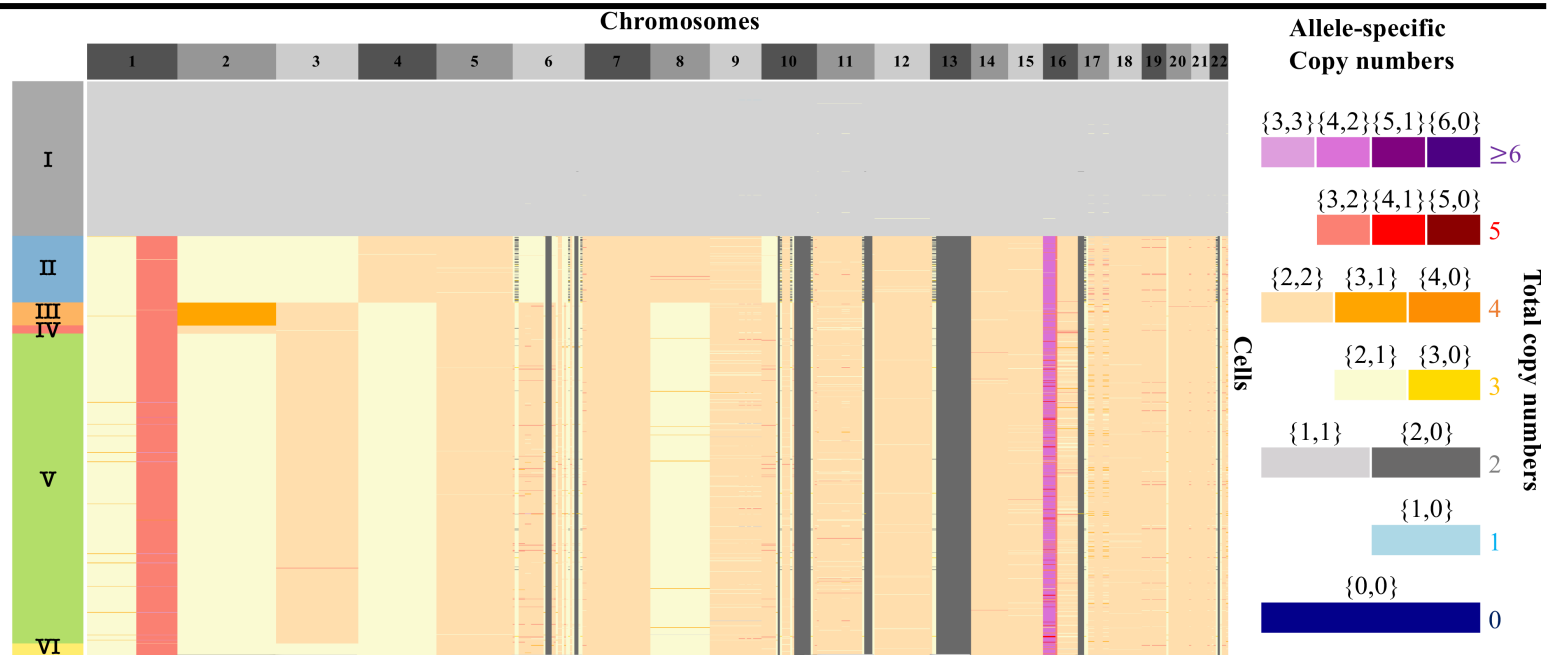
$$c = \hat{c} + \check{c}$$



Tumor section E from breast cancer patient S0:

Allele-specific copy numbers

$$\{\hat{c}, \check{c}\}$$



Copy neutral loss of heterozygosity (CN-LOH)

Tumor section E from breast cancer patient S0:

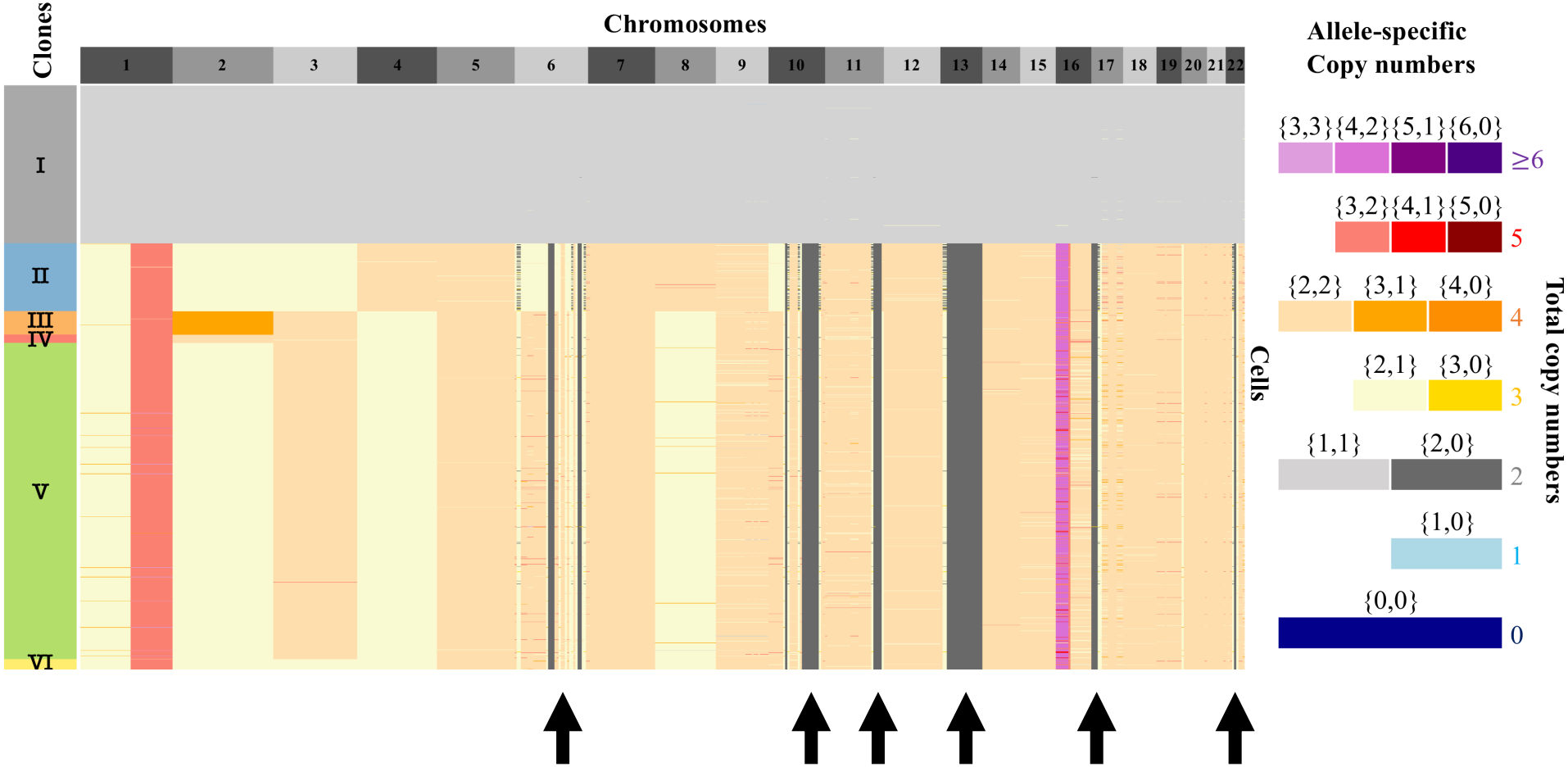
- Allele-specific copy numbers $\{\hat{c}, \check{c}\}$
- Total copy number $c = \hat{c} + \check{c}$

Copy neutral LOH:

{2,0}

Diploid region:

{1,1}



CN-LOH supported by BAF in clones

Tumor section E from breast cancer patient S0:

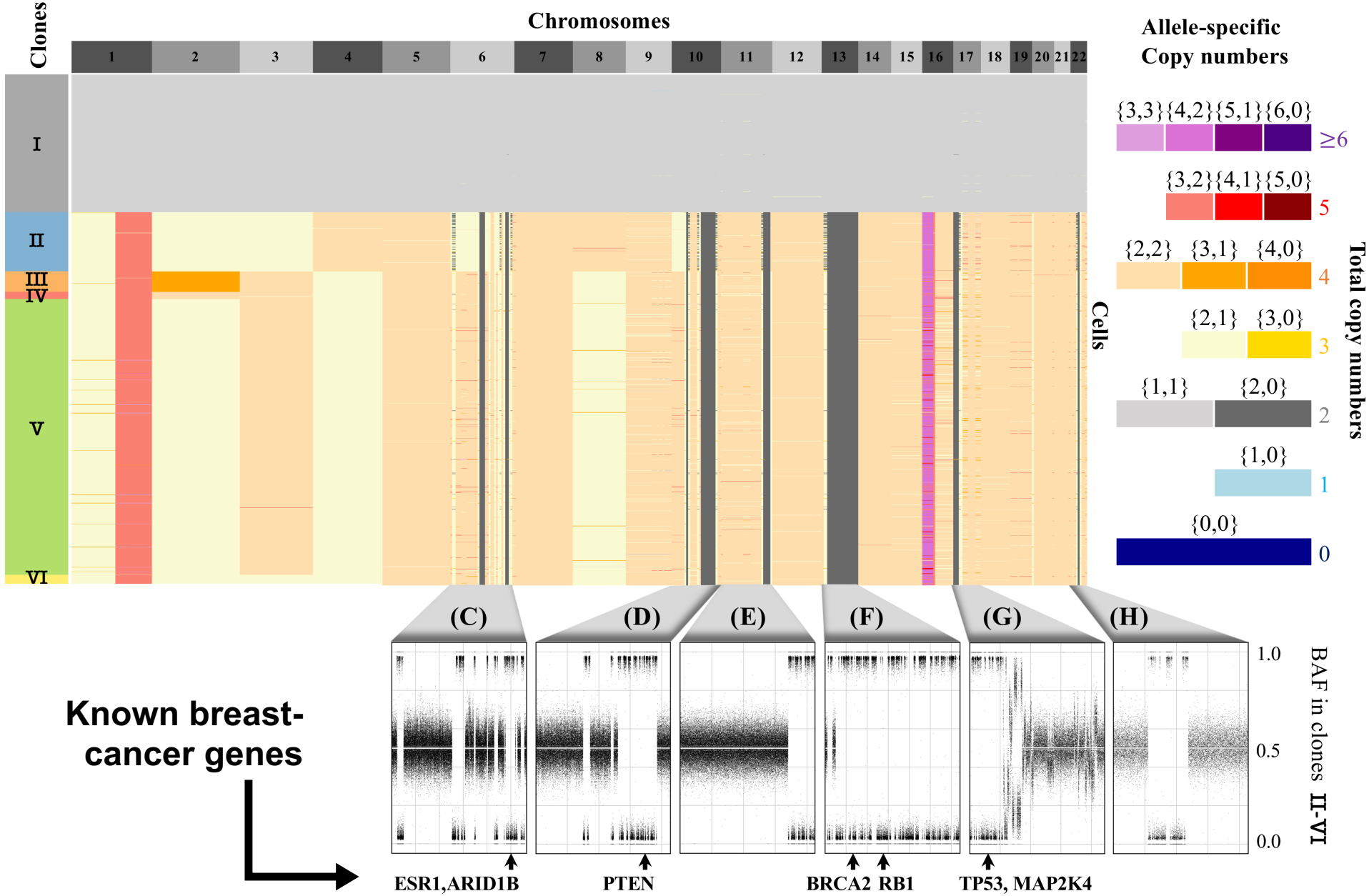
- Allele-specific copy numbers $\{\hat{c}, \check{c}\}$
- Total copy number $c = \hat{c} + \check{c}$

Copy neutral LOH:

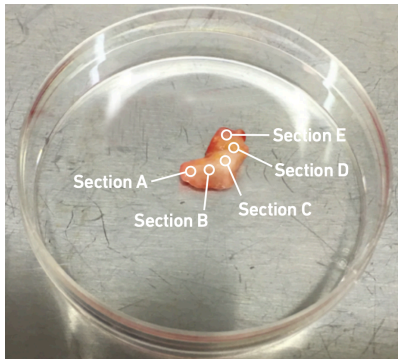
{2,0}

Diploid region:

{1,1}



Tumor evolution across all cells in five tumor sections



10,202 cells

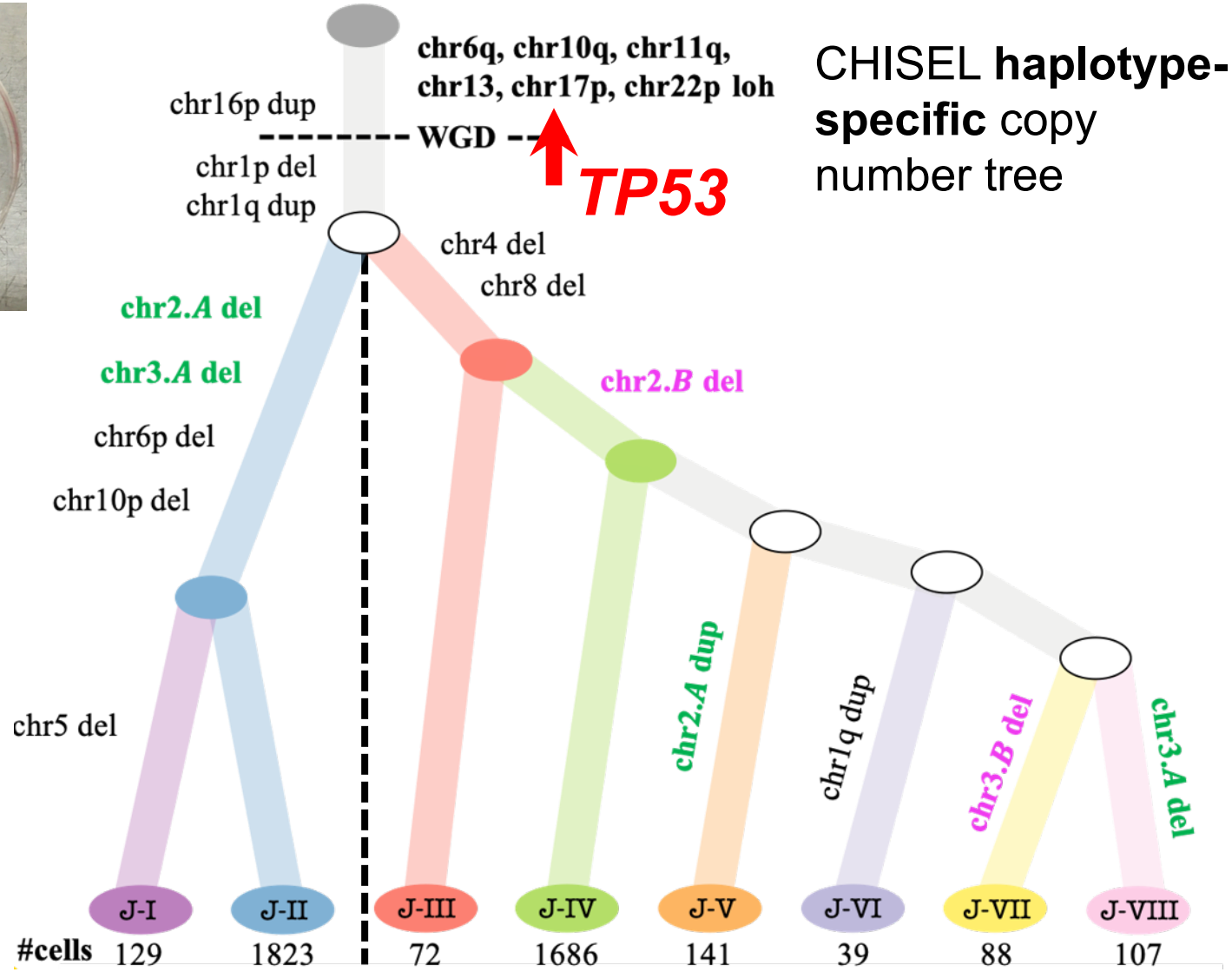
Findings:

1. LOH affects *TP53* before WGD
   ➔ Evidence for ***TP53*** **inactivation preceding WGD**

   [Bielski et al., *Nature Genetics*, 2018]

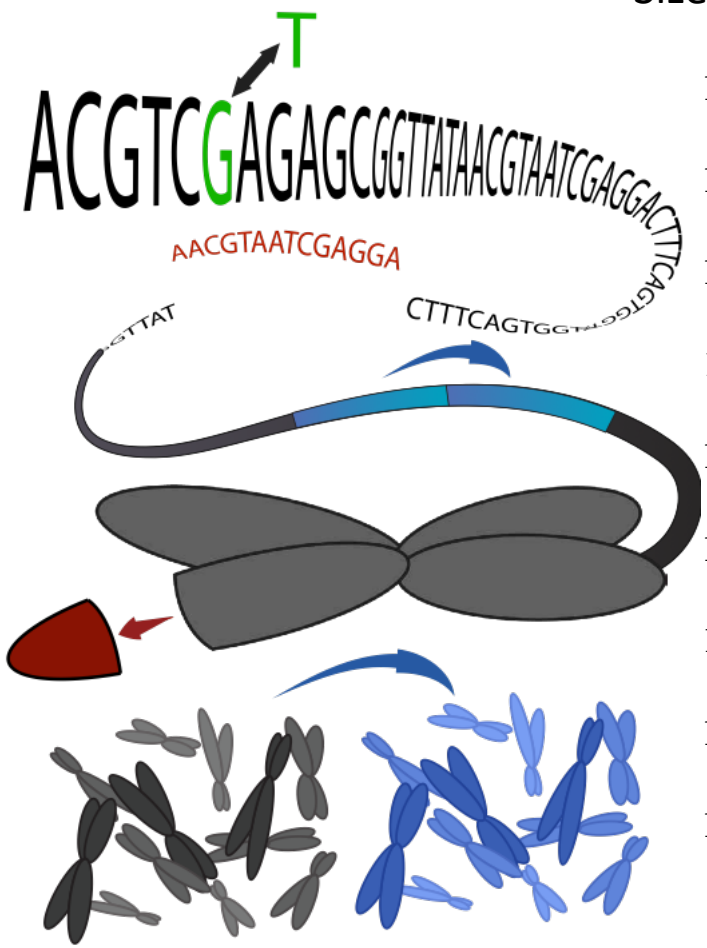
2. Mirrored-subclonal CNAs separate two branches and affect breast cancer tumor suppressor genes
   ➔ **Convergent evolution**

   [Jamal-Hanjani et al., NEJM, 2017]

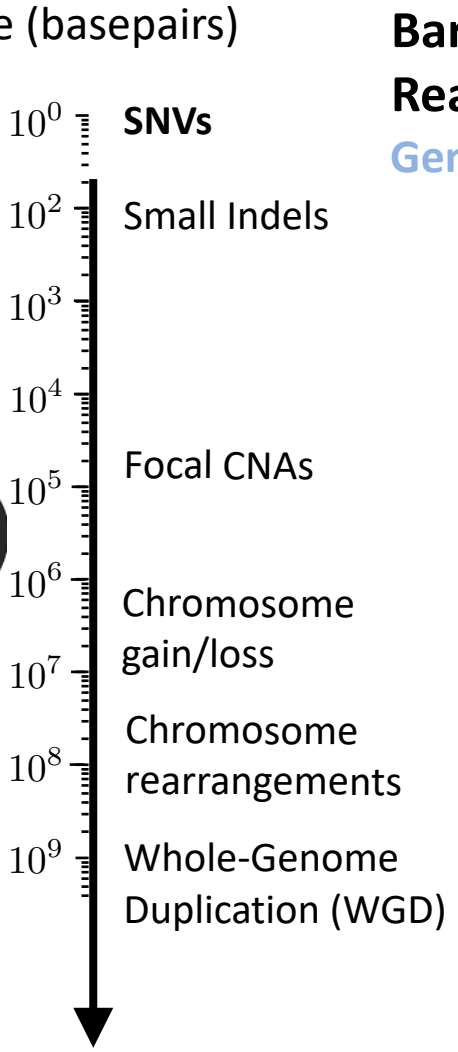


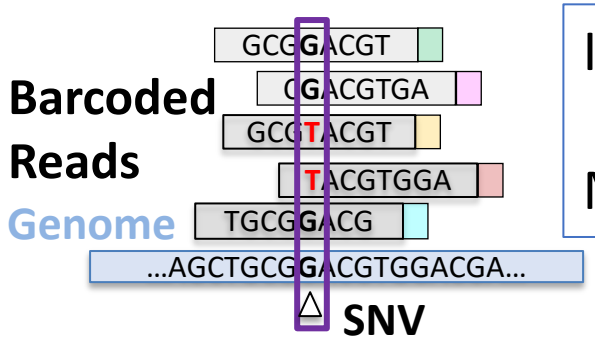
CHISEL **haplotype-specific** copy number tree

Orthogonal validation using somatic SNVs



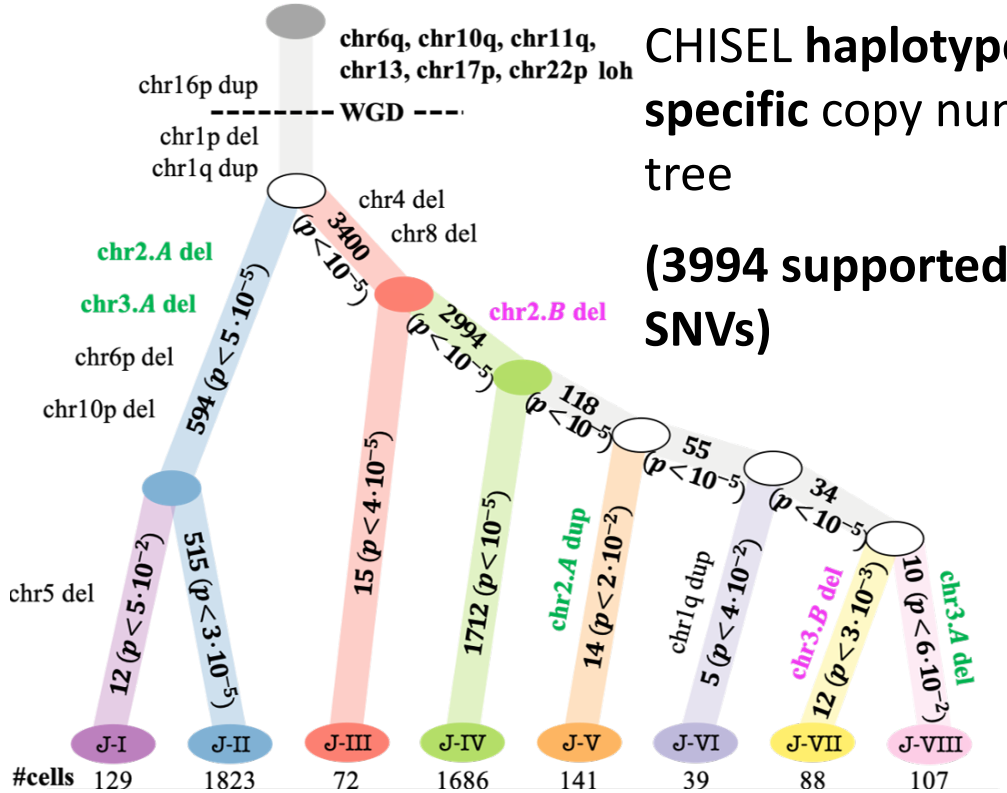
[Figure: Gryte Satas]





Identify from pseudo-bulk sample

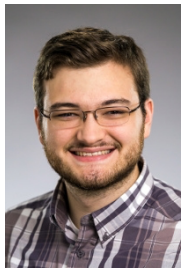
Most cells have 0 covering reads!



CHISEL **haplotype-specific** copy number tree

(3994 supported SNVs)

SBMClone: Infer tumor clones from SNVs in ultra-low coverage single-cell data

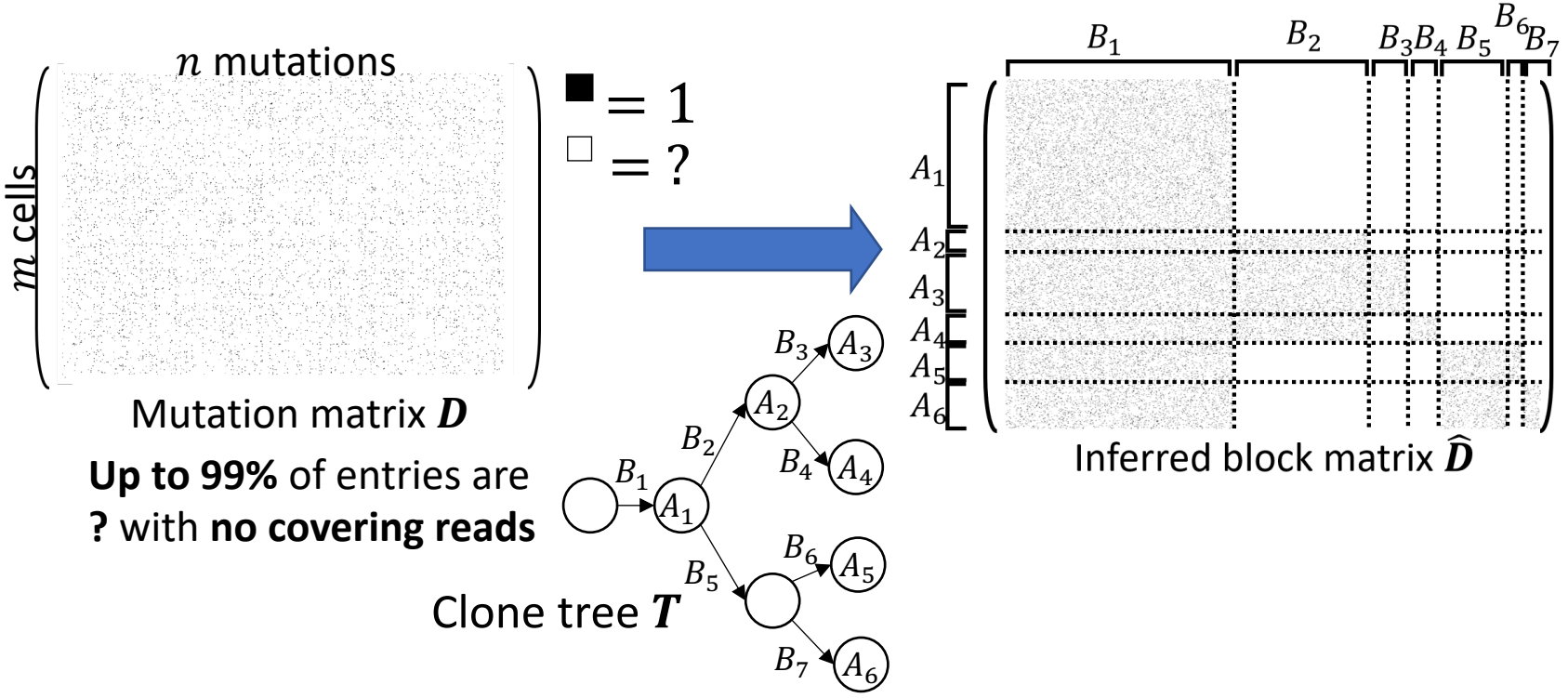


Matt Myers

Simone Zaccaria

Key idea: Clonal structure creates **blocks** in mutation matrix.

Inference: *stochastic block model*



Software: github.com/raphael-group/sbmclone

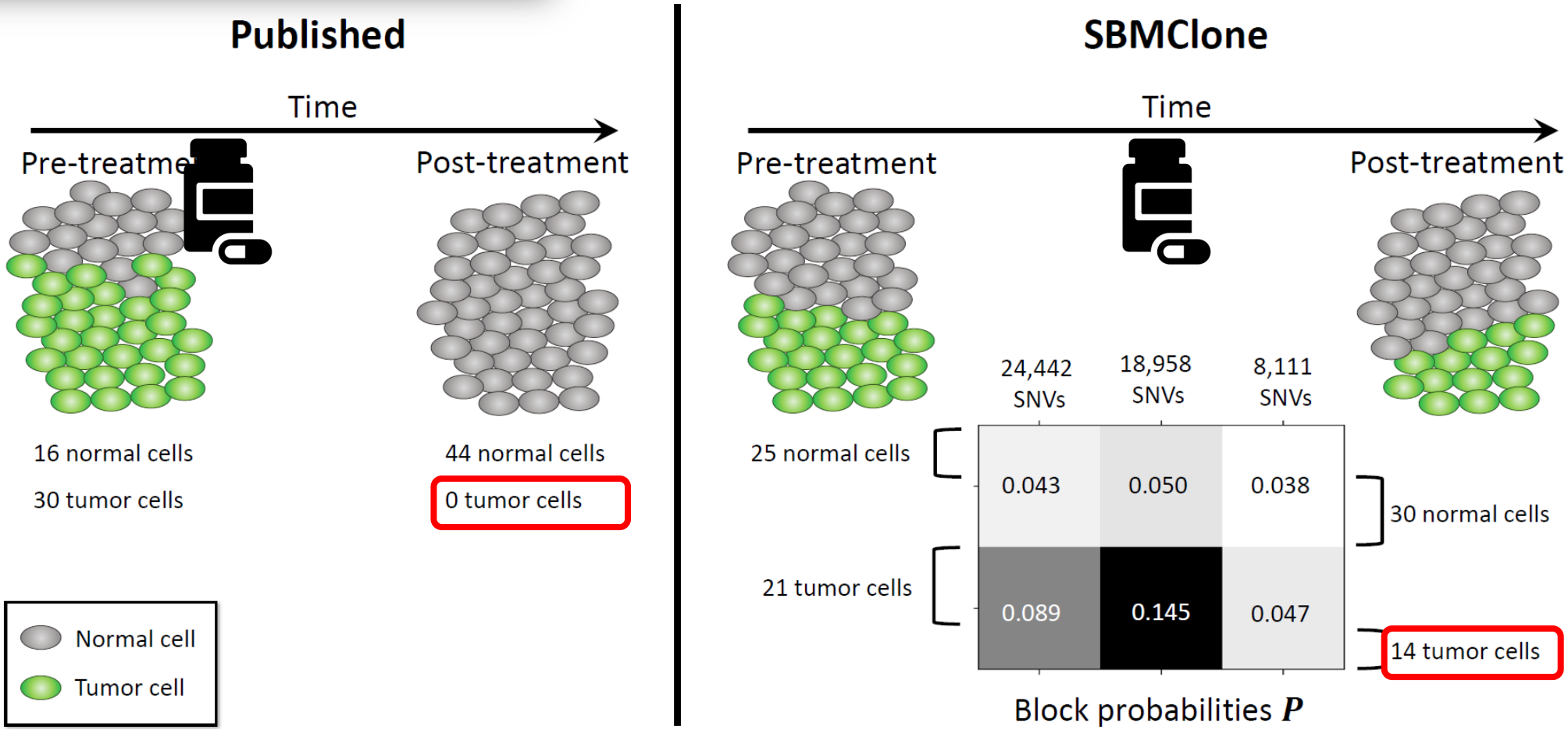
Myers, Zaccaria, R. ISMB/*Bioinformatics* 2020

SBMClone identifies tumor cells in **both** pre- and post-treatment samples → *clonal persistence*

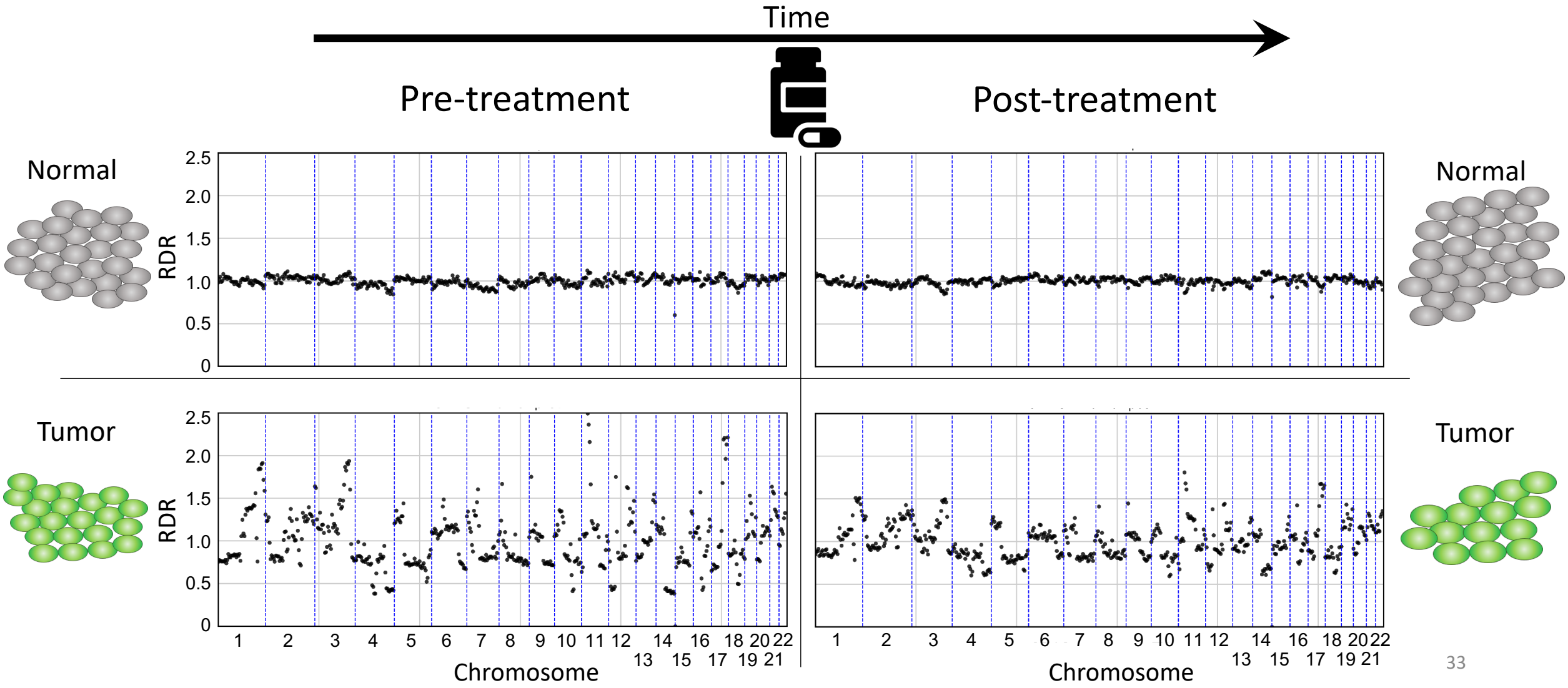
Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing

Charissa Kim,[1,2,6] Ruli Gao,[1,6] Emi Sei,[1] Rachel Brandt,[1] Johan Hartman,[3] Thomas Hatschek,[3] Nicola Crosetto,[4] Theodoros Foukakis,[3,*] and Nicholas E. Navin[1,2,5,7,*]

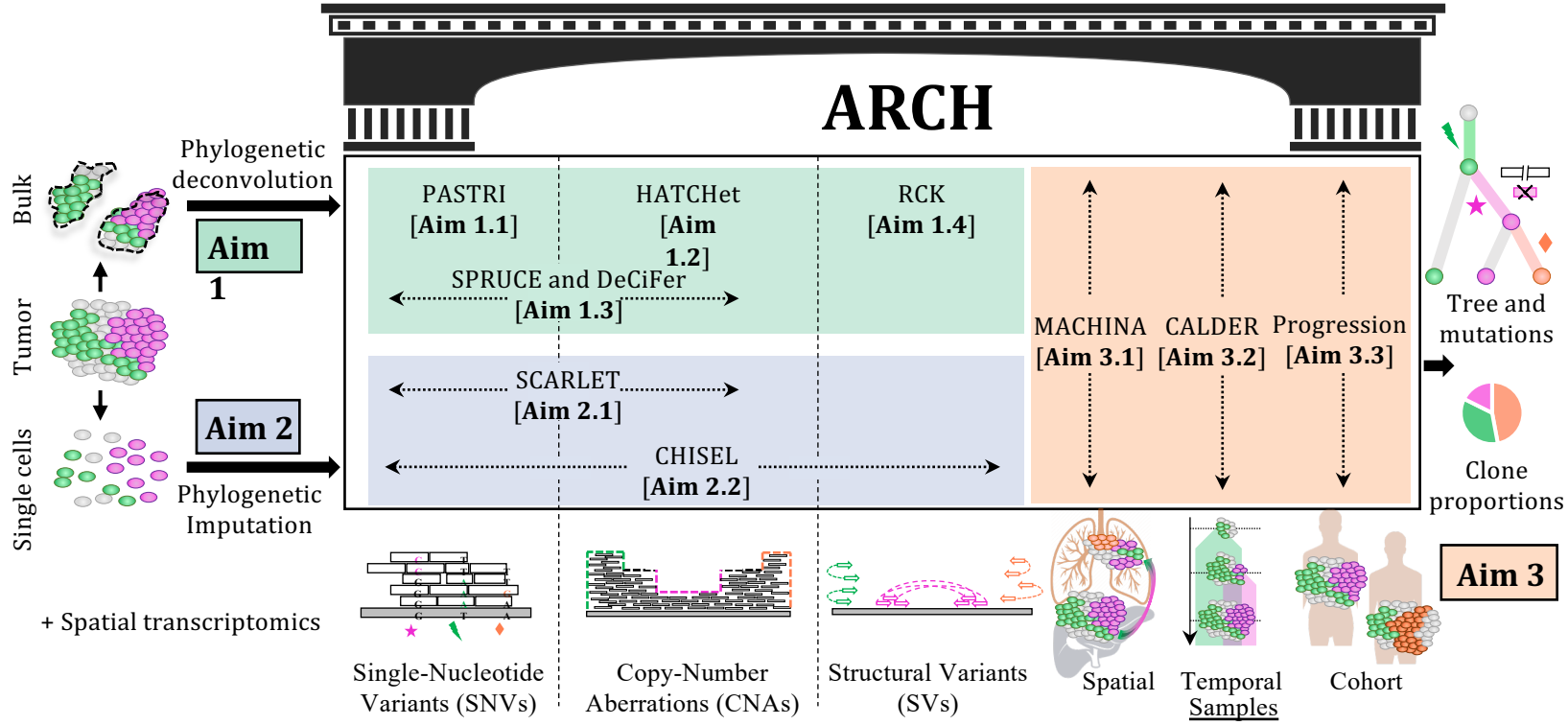
90 cells collected pre- and post-treatment
DOP-PCR sequencing: ≈**0.26X** coverage per cell



Copy-number signal supports tumor cells identified by SBMClone



Goals and Tentative Timeline



	Year 1	Year 2	Year 3	Year 4	Year 5
Software Packaging and Dissemination	**Harden existing packages; installation and Docker**	**R/Bioconductor, Jupyter notebooks, JavaScript viz**		**Modular software library and API, GenePattern notebooks**	**Refinements and bug fixes**
Software and Algorithm Development	Aim 1.1	Aim 1.2	Aim 1.3	Aim 1.4	
	Aim 2.1	Aim 2.2	Aim 3.1	Aim 3.2	Aim 3.3
ITCR Collaborative Activities					
Milestones	**ARCH v0.9**		**ARCH v1.0**	**ARCH v2.0**	**ARCH v2.1**

Questions and Feedback

- Contact: braphael@princeton.edu
- Software: http://github.com/raphael-group
- ARCH: github.com/raphael-group/arch (Under construction!)