

USTAR Center for
Genetic Discovery

Tracking cancer evolution across time and space

Gabor T. Marth, D.Sc., Aaron Quinlan, Ph.D.

Department of Human Genetics

USTAR Center for Genetic Discovery

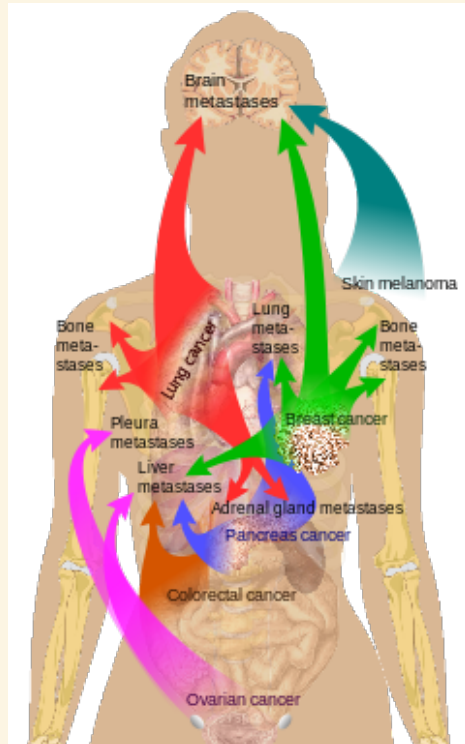
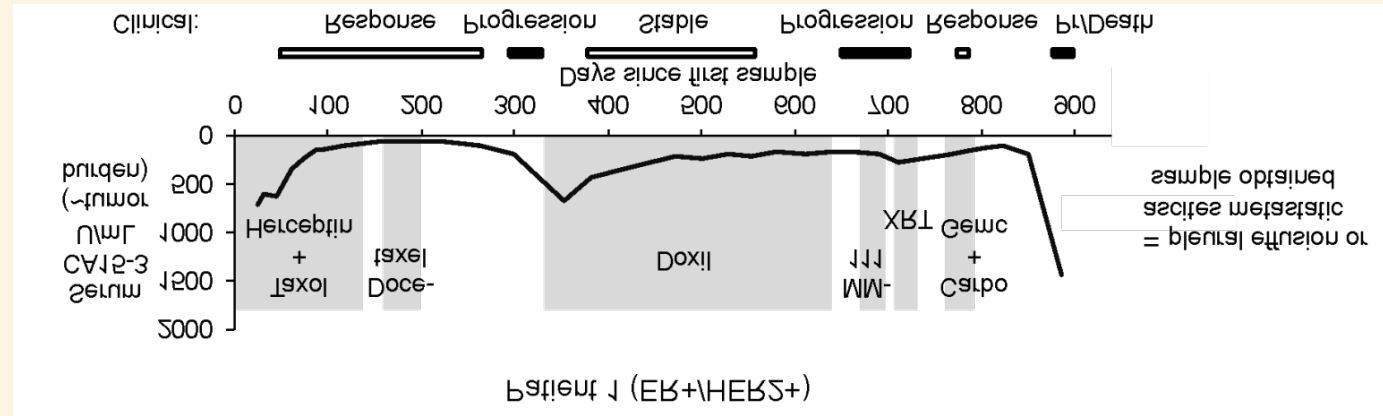
University of Utah

ITCR 2017 Annual Meeting, Santa Cruz, CA

May 31 – June 1, 2017

We want to understand how a patient's tumor evolves...

... during disease progression and relapse at a single tumor site, across multiple courses of chemotherapy

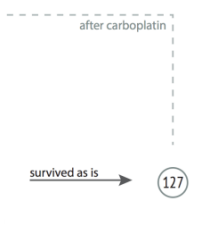
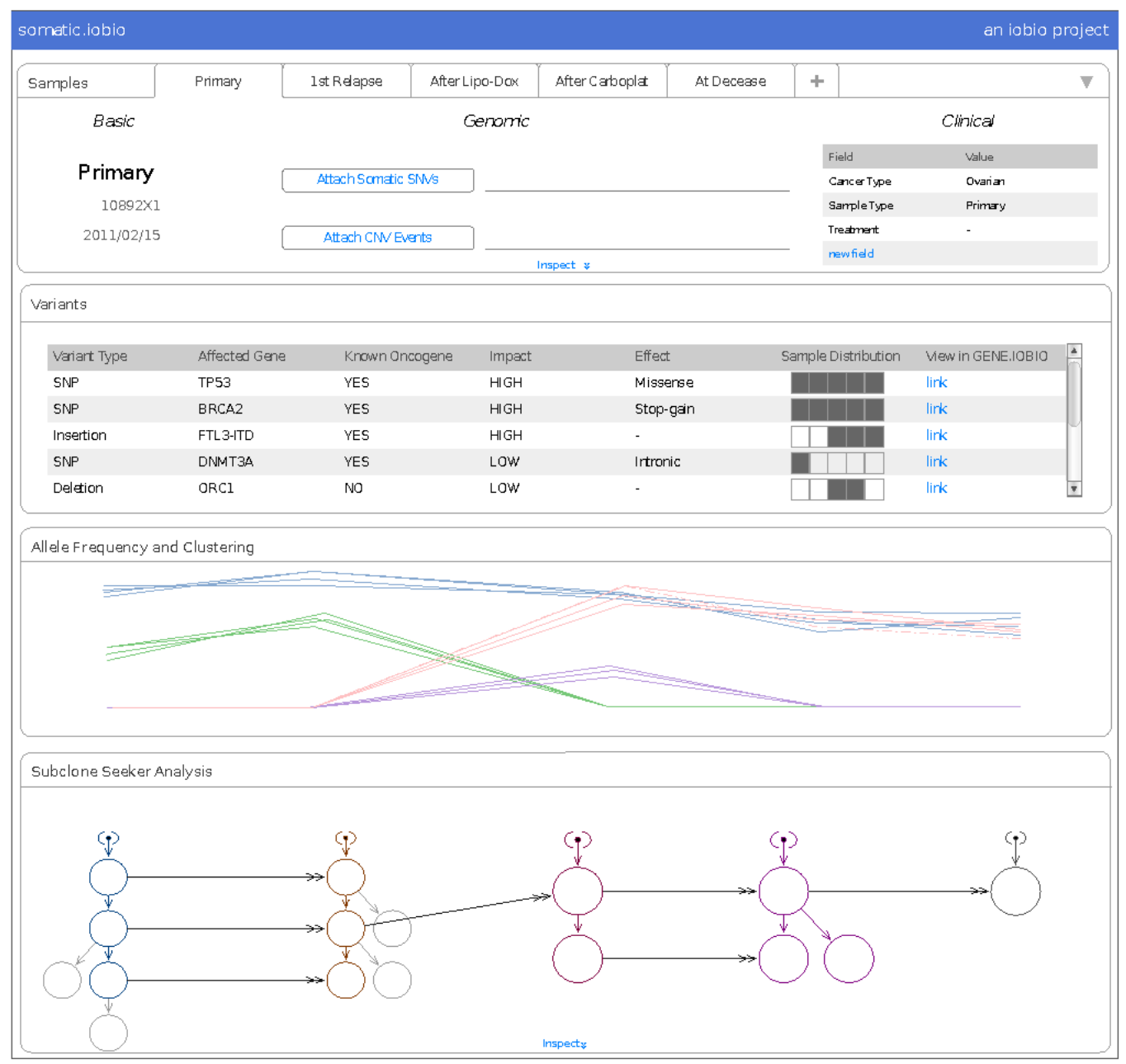


... and across metastasis, while colonizing distant organs and establishing multiple metastatic sites.

Cancer is a genetic disease

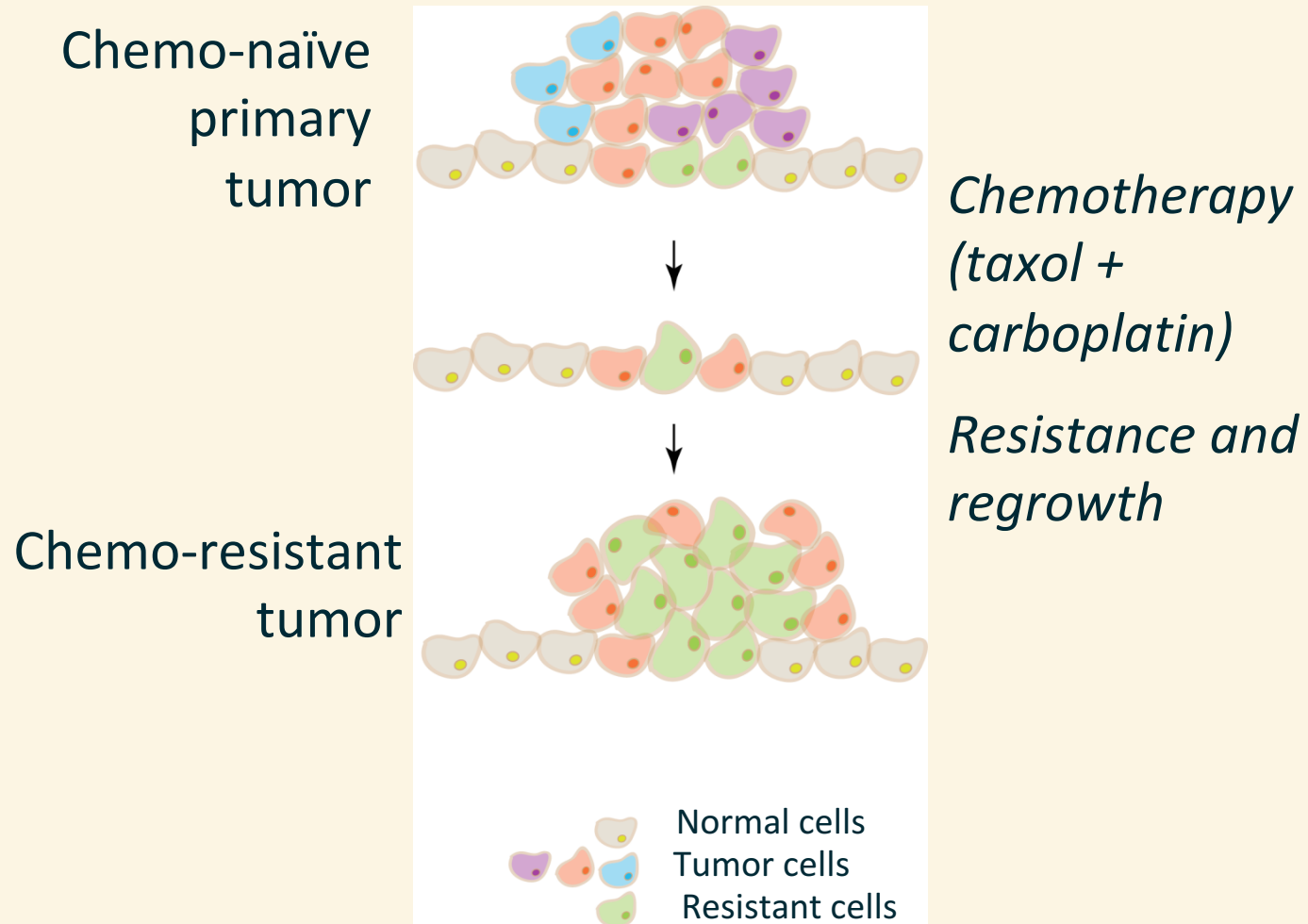
Four distinct areas of tool development

1. Detect all somatic mutations and inherited variants

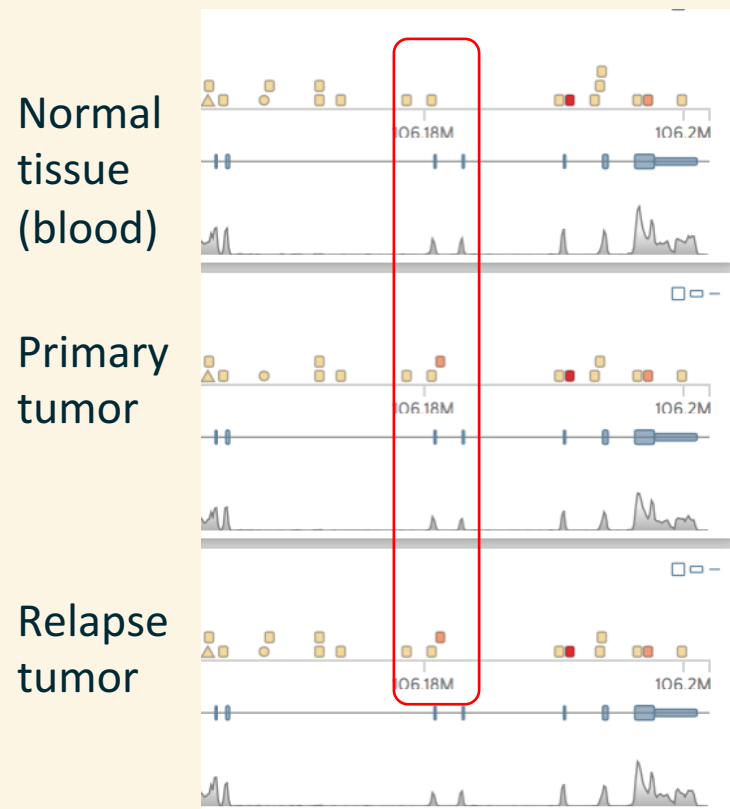


3. ...
mu
int

The main steps of subclone analysis: a *primer*

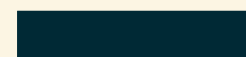
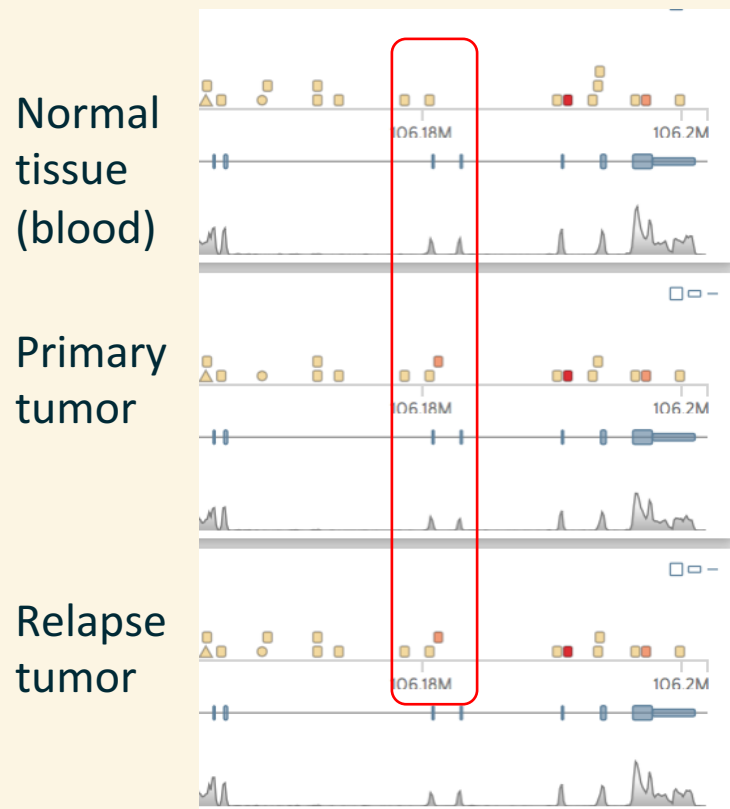


1. Detect somatic mutations

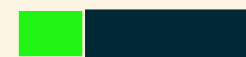


1. Detect somatic mutations

2. Determine mutation allele frequency (AF)



0%



30%

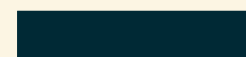
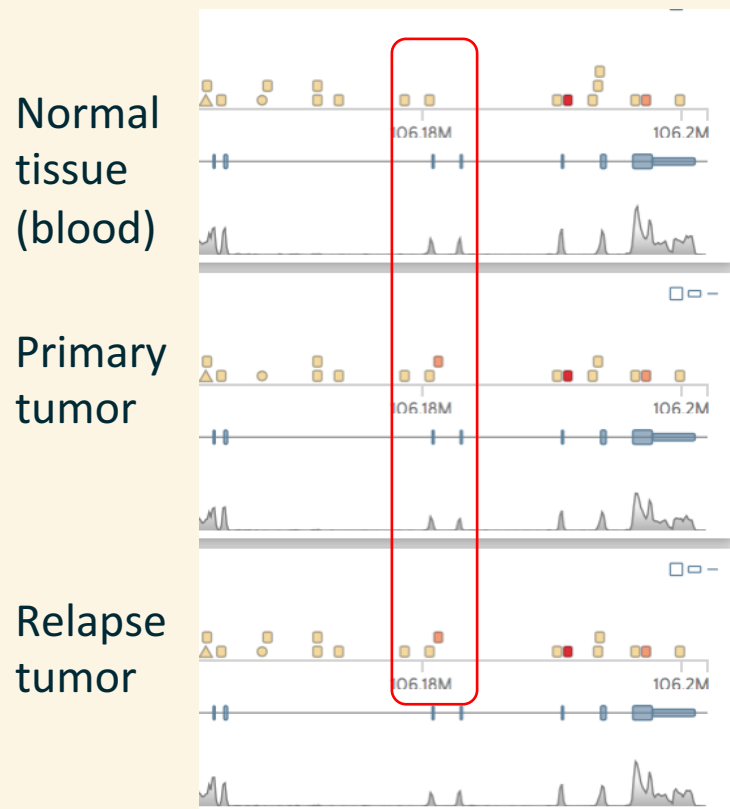


40%

1. Detect somatic mutations

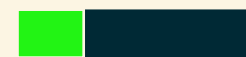
2. Determine allele frequency (AF)

3. Convert to cellular prevalence (CP)



0%

0%



30%

60%

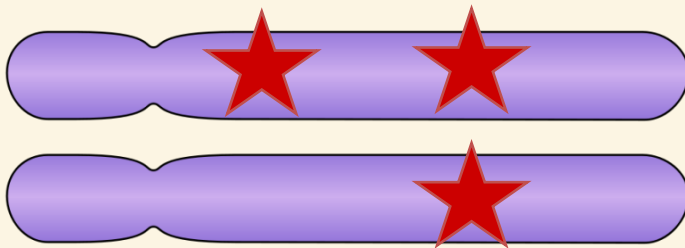


40%

80%

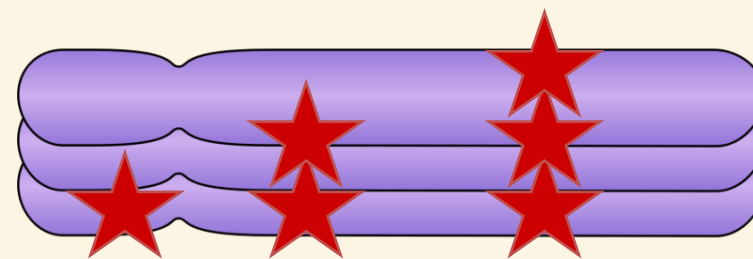
Example: AF=10%

Normal copy number (CN=2)



20% 10%

Amplification (CN > 2)

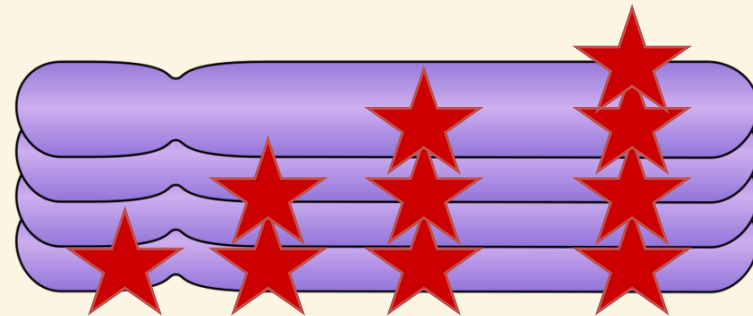


30% 15% 10%

Deletion (CN < 2)



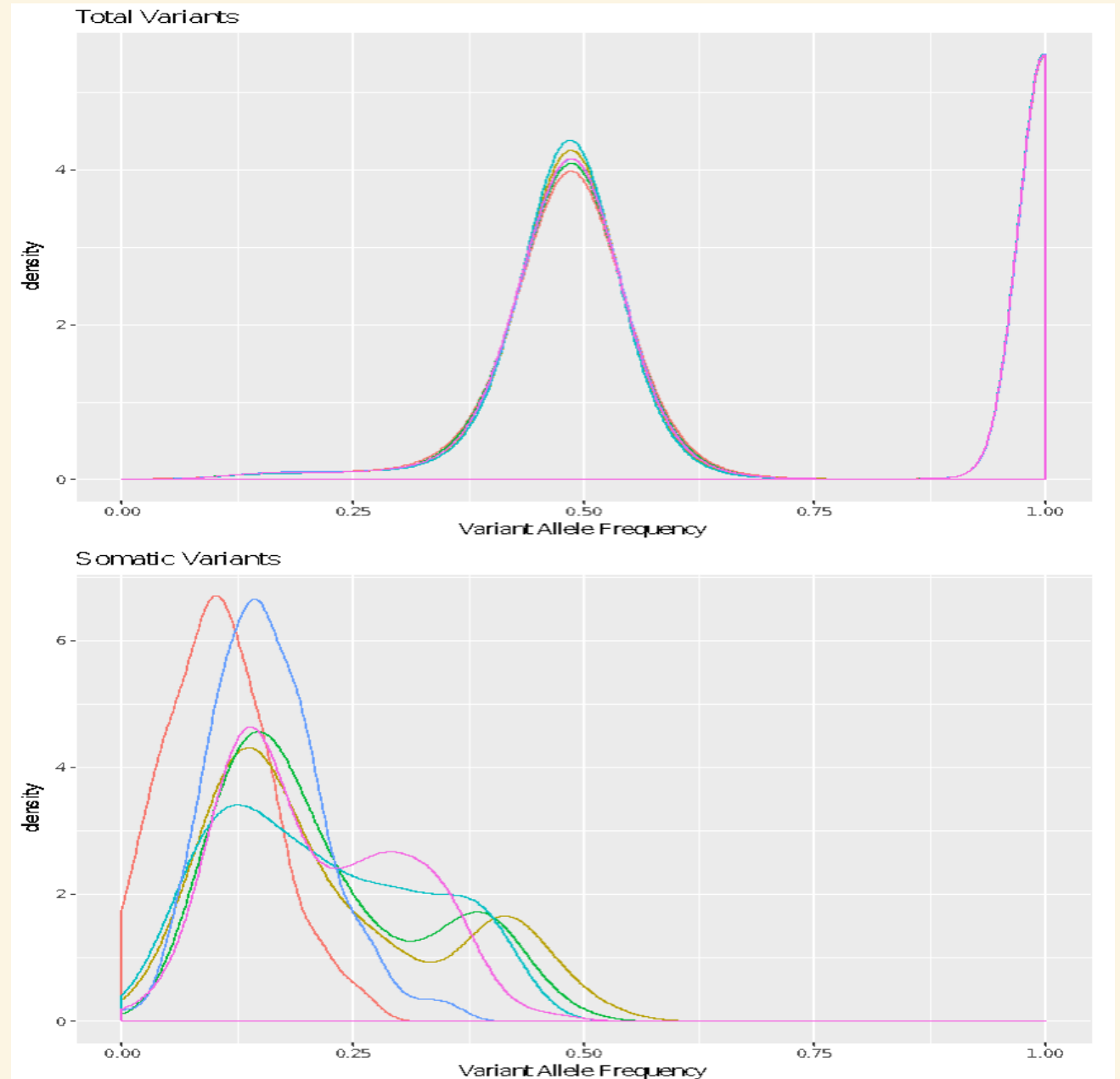
10%



40% 20% 13.3% 10%

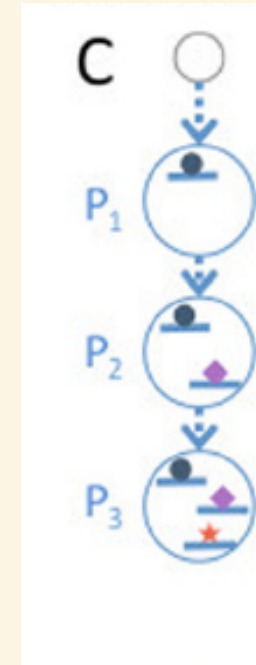
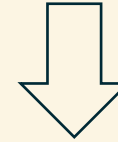
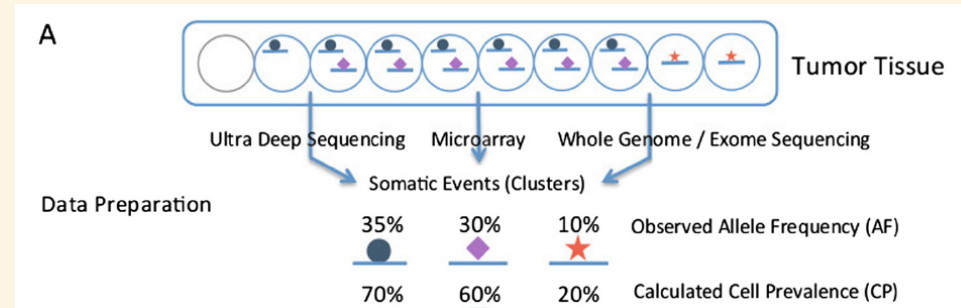
3. Mutation clustering

Inherited variants



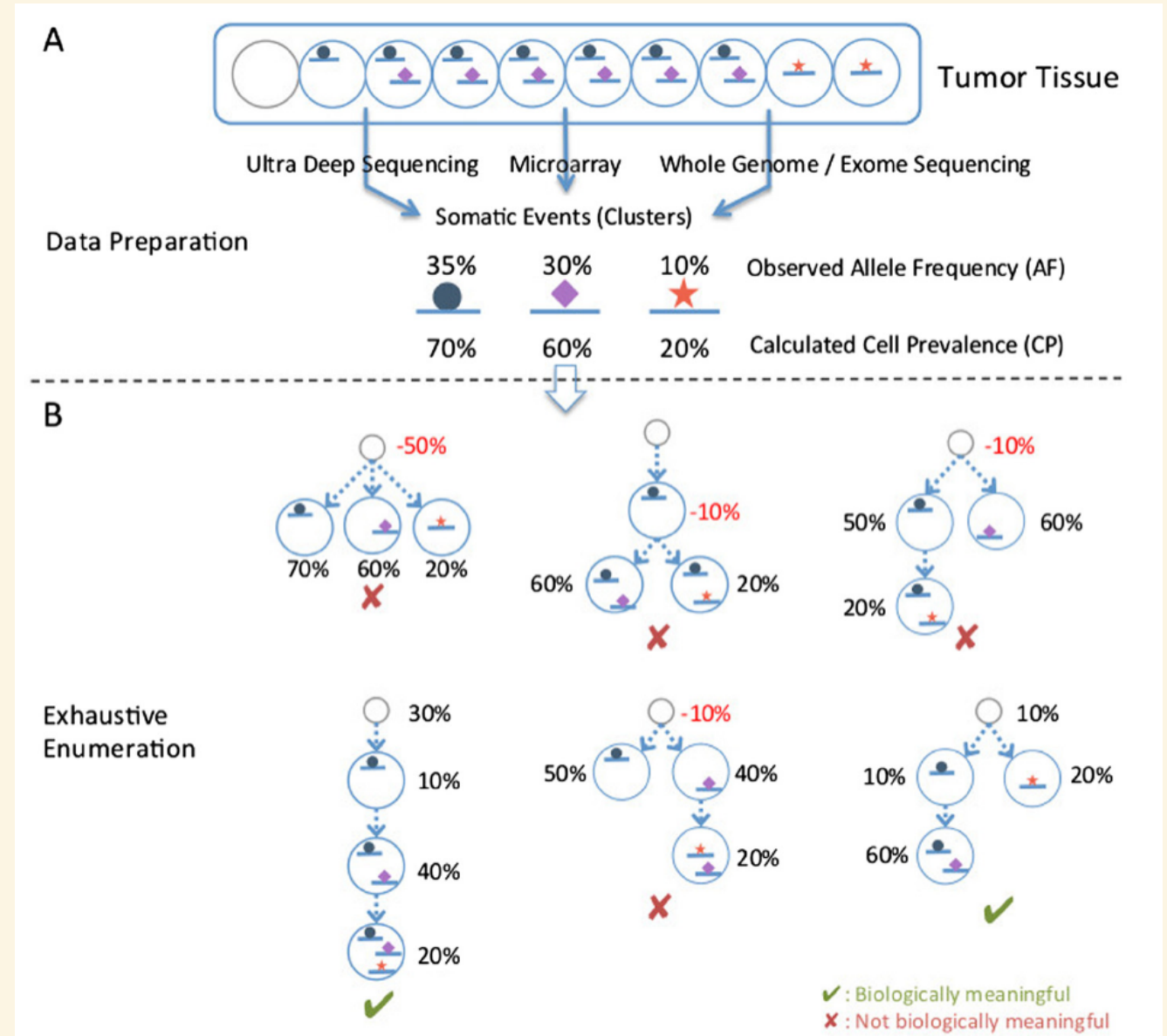
Somatic mutations

5. Subclone inference (reconstruction)

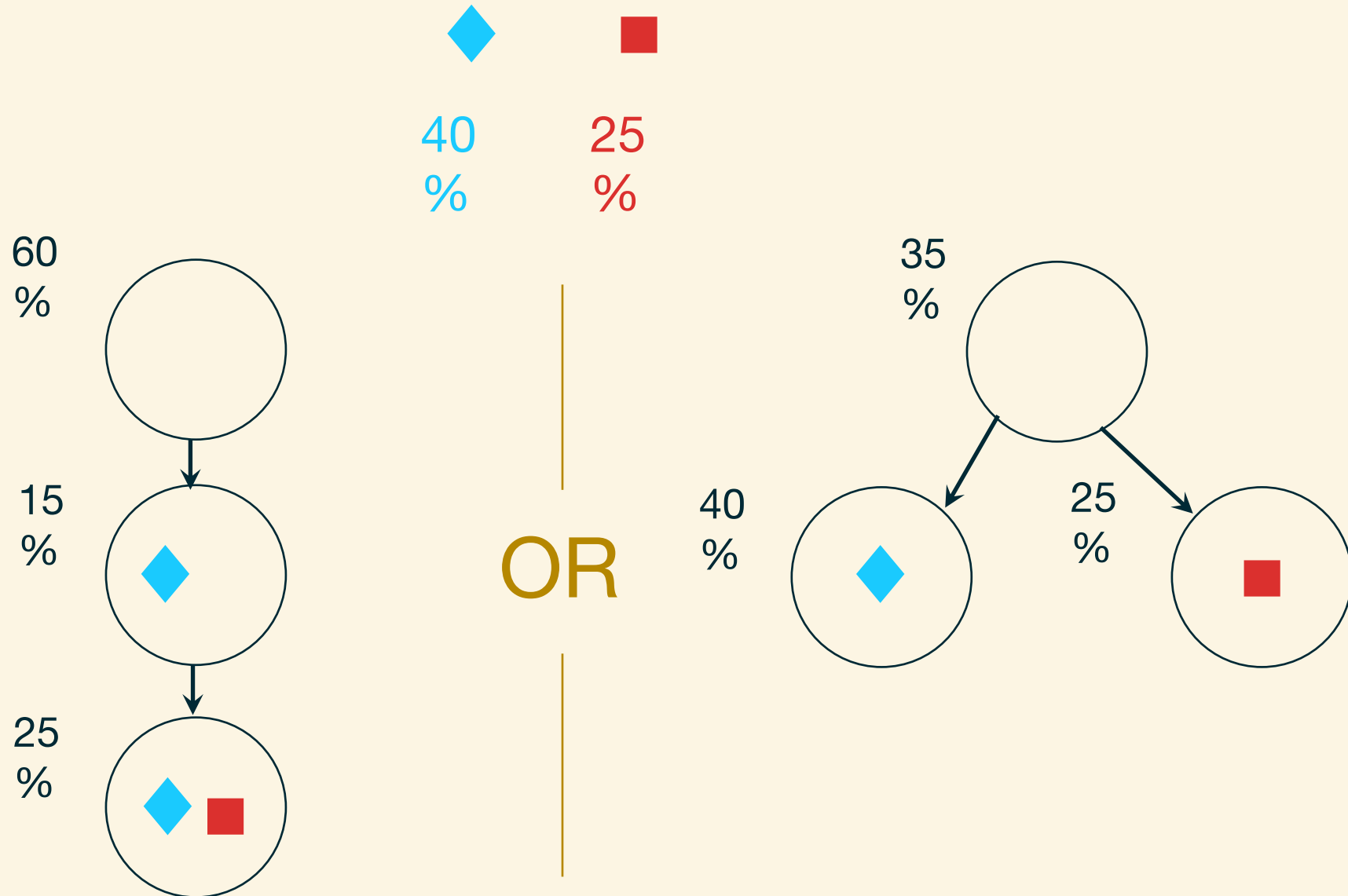


Our SubcloneSeeker method identifies all subclone evolutionary “trajectories” consistent with the bulk mutation AF data

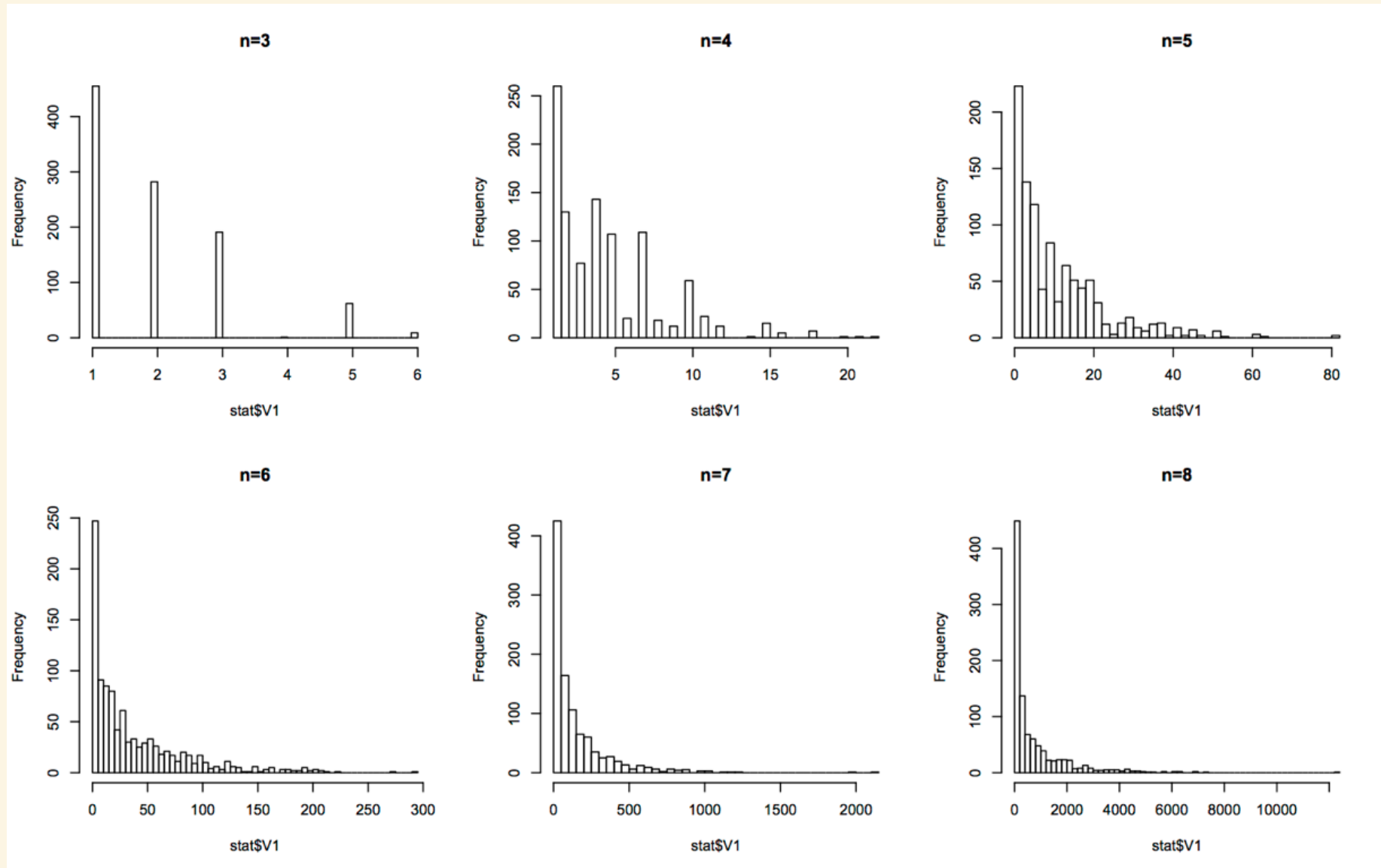
Qiao et al., **SubcloneSeeker**,
Genome Biology, 2014



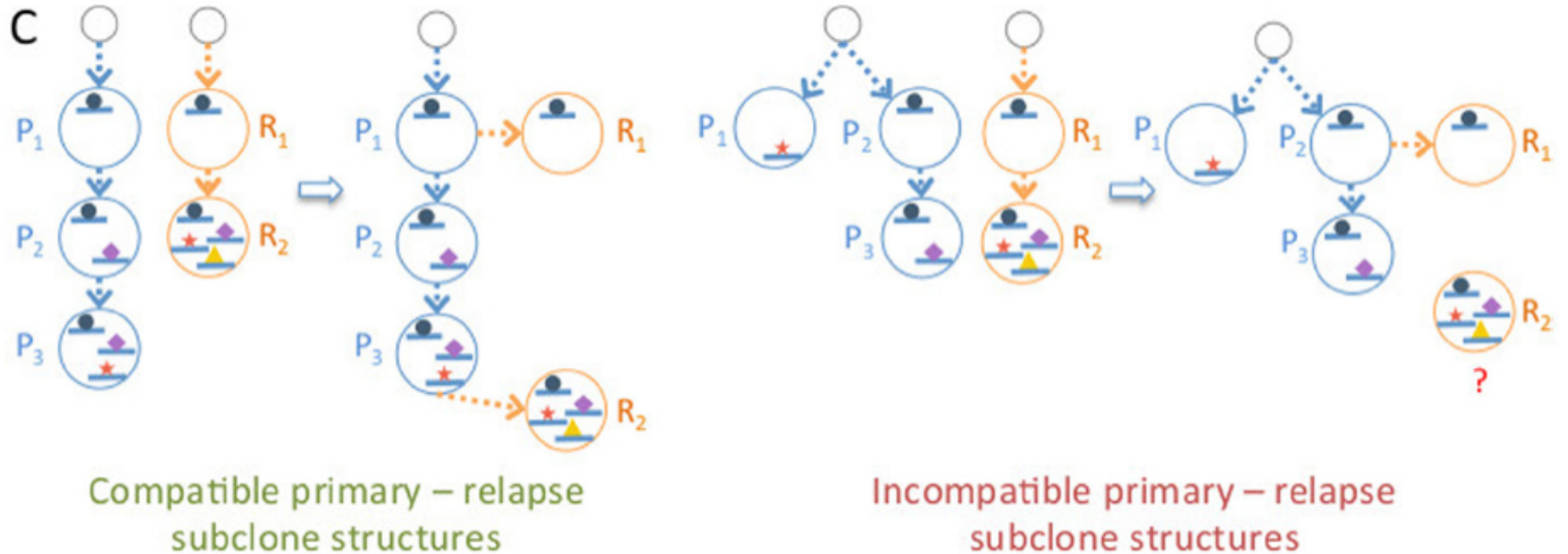
Multiple subclone structures may account for bulk data



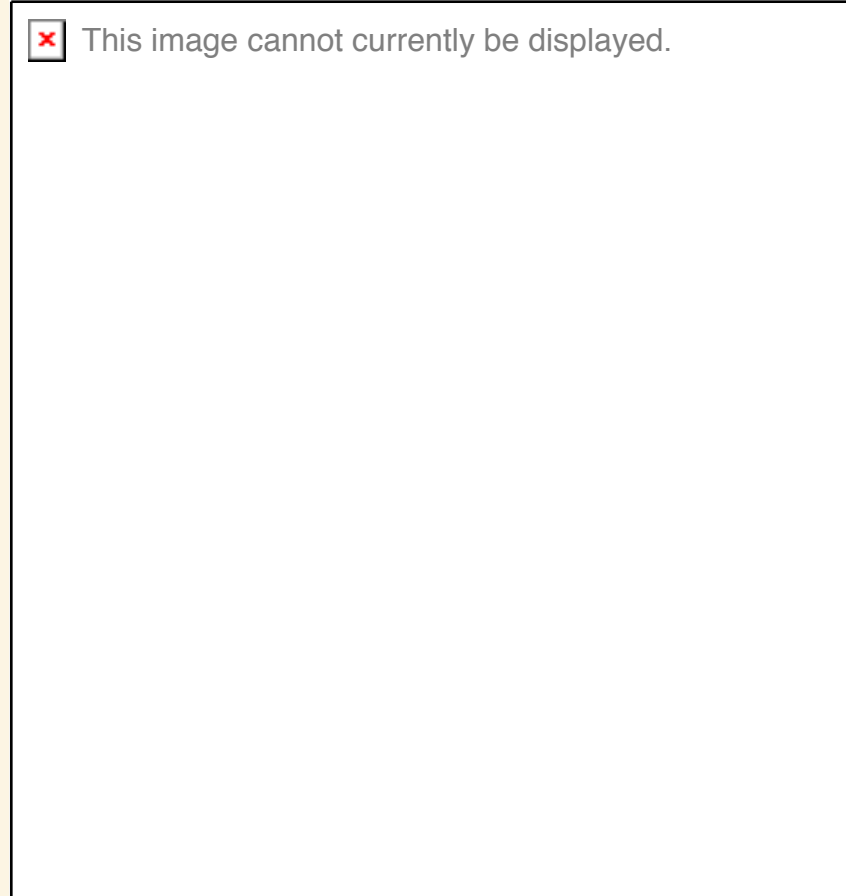
Number of “consistent” subclone structures is small



Further reduction of the “solution space” is possible with multiple biopsies from the same patient



Back to our primary chemo-naïve / resistant relapse ovarian example:





This image cannot currently be displayed.



This image cannot currently be displayed.



This image cannot currently be displayed.

Extending our tools to enable the analysis of **longitudinal tumor evolution**

The progression of a refractory breast cancer patient across three courses of treatment over 2.5 years, whose tumor was sequenced at 5 distinct time points, until just before death.

Mutation calling and clustering in longitudinal tumor sequencing data

B2

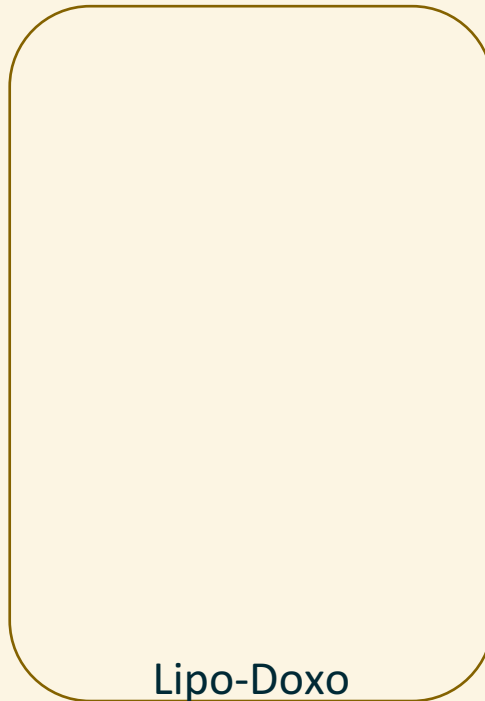
- Mutations are clustered according to their allele frequencies (cellular prevalence) at **all** sequenced time points
- All variants had to be **jointly** called across the normal control sample and the five tumor biopsies

Subclone evolution reconstructed by running SubcloneSeeker over consecutive pairs of time points

Multiple “solutions” exist

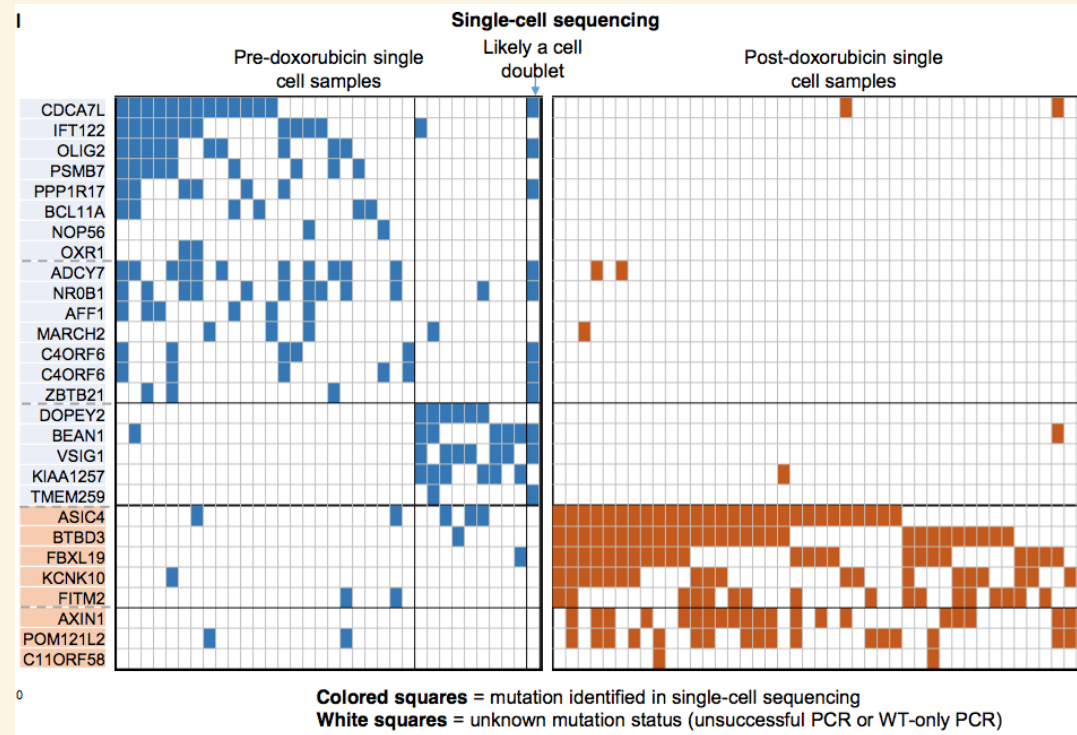
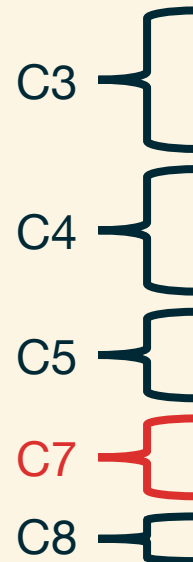
← “Consensus”

Confirming computational subclone structure prediction with single-cell genotyping

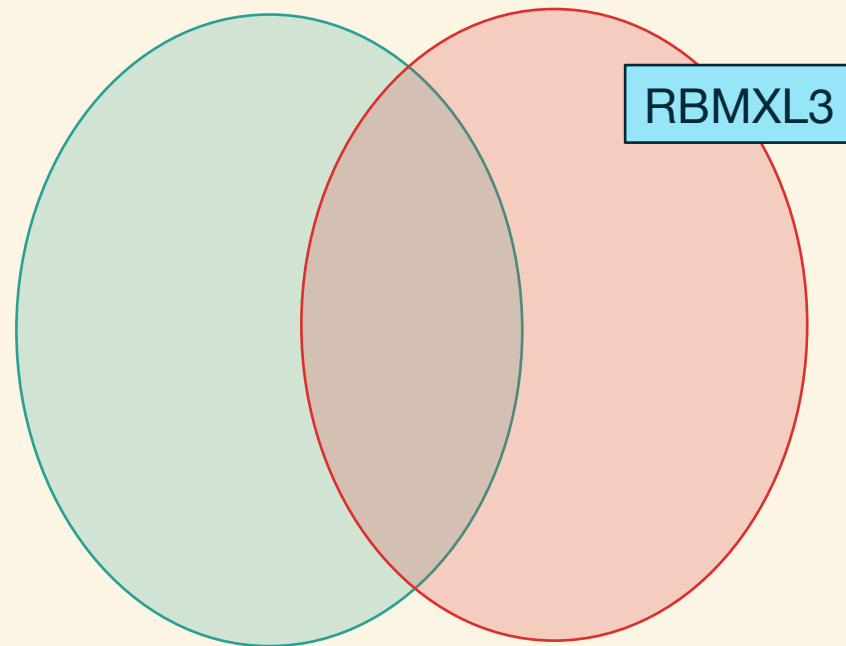


Single-cell sequencing data from **pre-Doxo** and **post-Doxo** cells, by genotyping groups of mutations that define the variant clusters, and the subclones

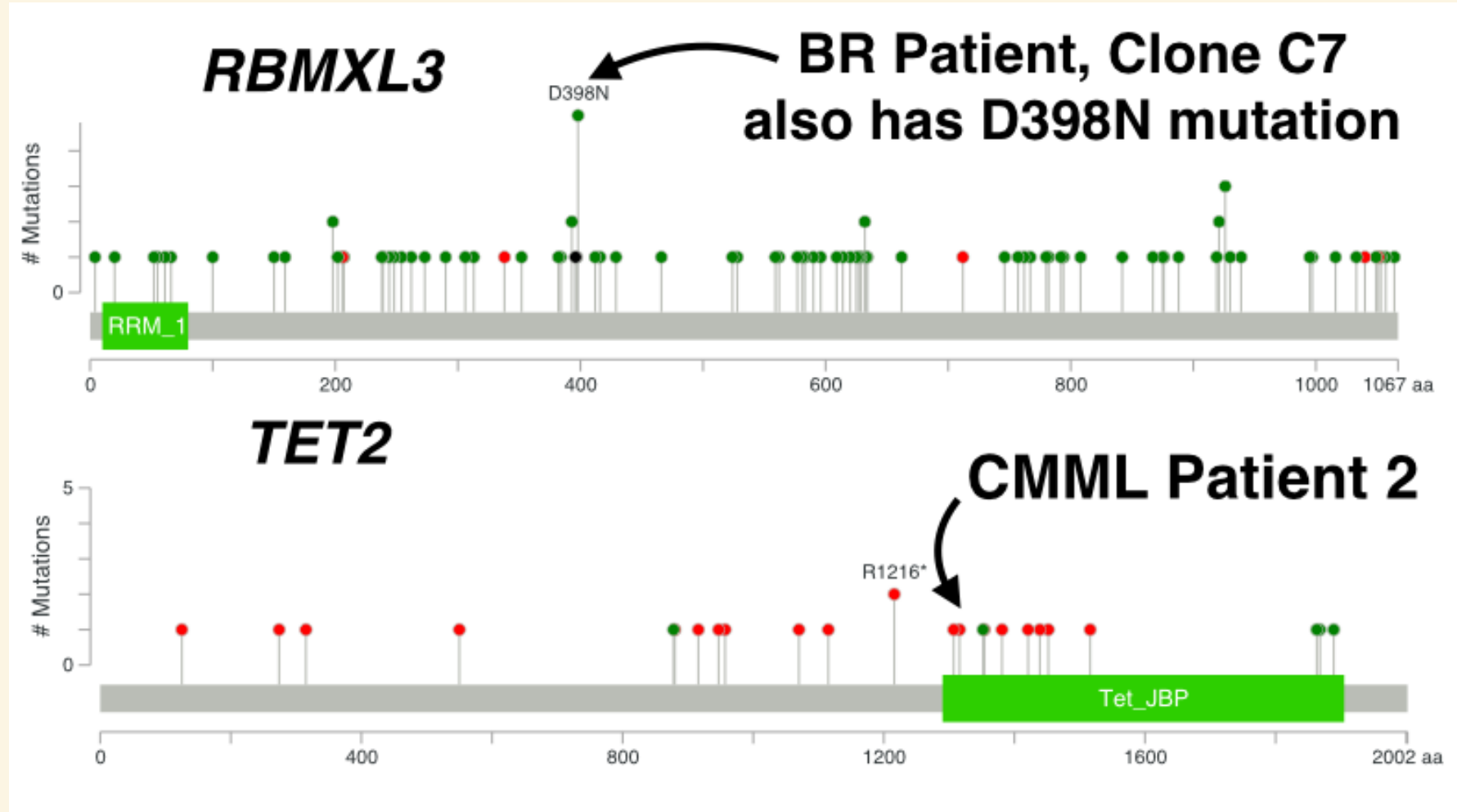
Mutations that
define the subclones



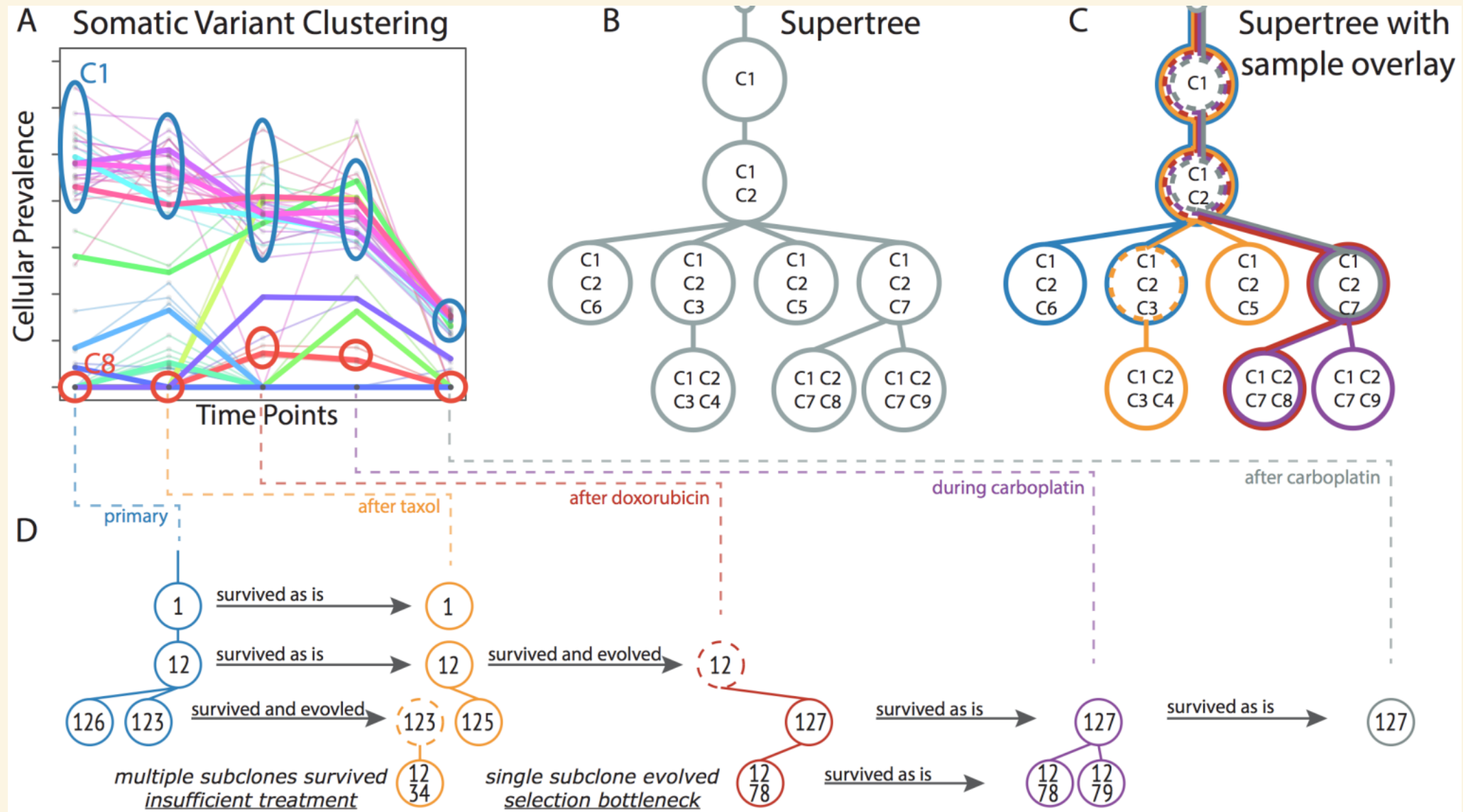
Interpreting subclone evolution in the context of disease progression



Variant interpretation

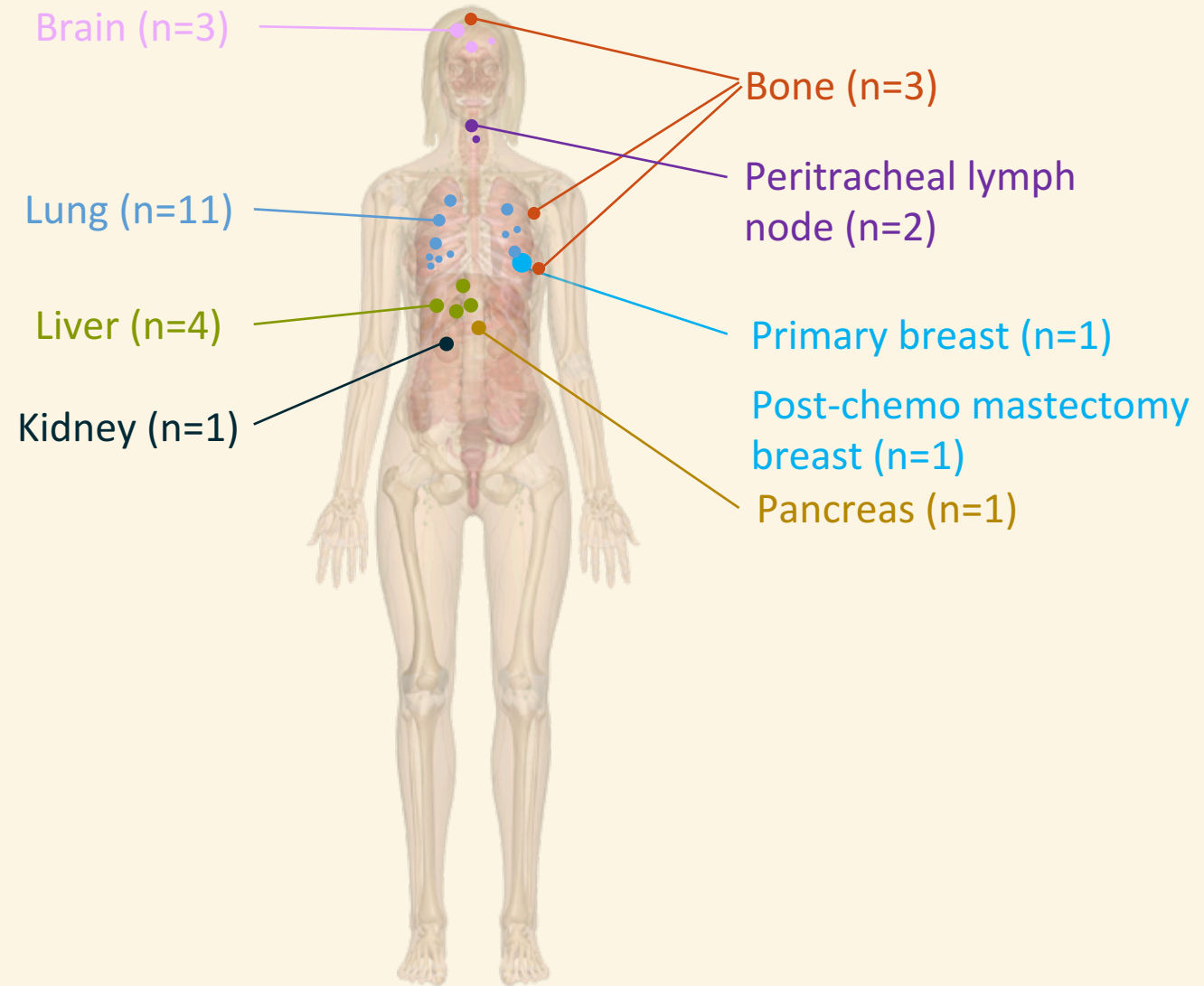


Current focus is on developing a general algorithmic solution for reconstructing subclone structure in a longitudinal setting (SuperSeeker)



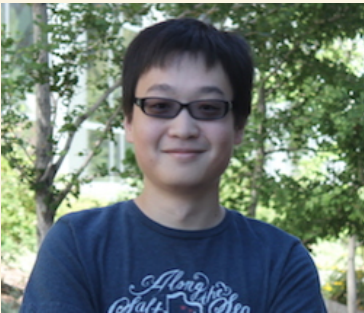
Our “driver” dataset is from a “rapid autopsy” metastatic breast cancer patient

- Triple-negative disease
 - Primary biopsy at diagnosis
 - Mastectomy at recurrence
 - 26 metastases from various organs removed during *rapid autopsy*, hours after death
 - 2 normal control samples (skin)
 - **30 biopsies across 8 organs**
- 60X WGS + WES Illumina sequencing data collected at Wash U for all biopsies + 2 normal controls from skin + 2 primary tumors
- Inherited variants in BRCA2, not in BRCA1
- All tumor sites share somatic homozygous TP53 missense SNV and RB loss



Please visit our posters to see the current state of this analysis

Yi Qiao



Computational toolkit for reconstructing metastatic expansion at subclone level

Yi Qiao¹, Xiaomeng Huang¹, Samuel W. Brady², Andrea H. Bild², Gabor T. Marth¹

¹USTAR Center for Genetic Discovery, Eccles Institute of Human Genetics, University of Utah
²Department of Pharmacology & Toxicology, University of Utah
Contact: yi.qiao@genetics.utah.edu



INTRODUCTION: In most cancers, metastasis is the major cause of treatment failure and patient death. Understanding metastatic tumor evolution at a subclonal level is likely to offer vital insight into mechanism. The identification of aggressive subclones responsible for metastatic colonization into distant organs offers the possibility to preferentially target these subclones, rather than more benign groups of cells within the tumor.

METHOD: Here we present the SeederSeeker computational toolkit, built on top of our published SubcloneSeeker¹ algorithm, that examines many types of somatic variants including CNVs, LOHs, and SNVs to reconstruct the phylogenetic relationships among metastatic tumor samples. The entire workflow is broken into three steps: 1) Using large scale chromosomal events such as CNVs and LOHs to partition samples into groups of metastatic waves. 2) Using WGS SNVs on CNV and LOH neutral regions to reconstruct sample level phylogeny. And 3) Using SNV variant allele frequencies to reconstruct subclone level phylogeny.

RESULT: We applied the method to a triple negative breast cancer rapid-autopsy dataset consists of 26 metastatic sites as well as primary diagnostic tumor. The result suggests that the primary breast cancer sample initially invaded lung, and then spread in four metastatic waves, each involving a different group of organs (wave 1 invaded abdominal organs such as liver and pancreas, wave 2 and 3 invaded lymph nodes, and wave 4 invaded bones and brain tissues). In each wave lung was always an early invasion target, suggesting that lung could play an important role for breast cancer evolution and adaption in our case.

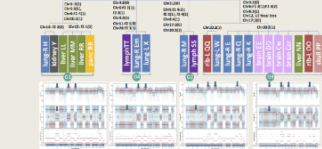
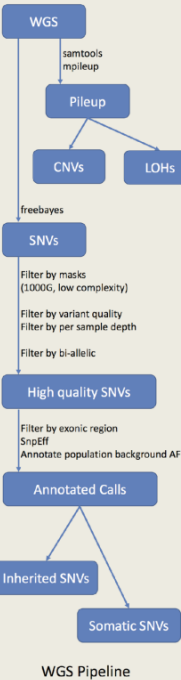
CONCLUSION:

- We have previously developed a subclone deconvolution method, which we extended to perform multi-site metastatic phylogeny reconstruction.
- We applied the method to an index patient, and discovered how metastatic waves occurred with associated evolution trajectories.
- The method can be used to identify seeding patterns across metastatic sites, potentially revealing cancer origin, metastatic mechanisms and guide treatment.

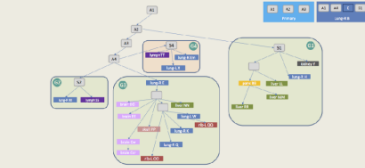
References

1. Qiao, Y. et al. "SubcloneSeeker: a computational framework for reconstructing tumor clone structure for cancer variant interpretation and prioritization." *Genome Biol* 15 (2014): 443.
2. Shen, R. et al. "ACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing." *Nucleic acids research*, 44(16): e131.

<https://github.com/yiq/stk>



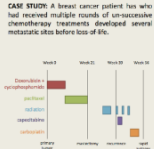
Sample grouping by CNV and LOH states. Through an expert driven process, WGS sequencing data is analyzed by CNV and LOH calling algorithms (e.g. FACETS¹), and samples grouped by shared chromosomal events, such as chr17/3(0), to establish coarse-grain sample relationships. This corresponds well with SNV based clustering (data not shown) and suggests metastatic waves.



Sample phylogeny reconstruction. Although CNV and LOH events can guide sample clustering, the resolution they provide is limited. We can, however, utilize the state of shared vs unique somatic SNV events to resolve sample phylogeny in much finer detail. This can only be done in CNV and LOH neutral regions of the genome with high certainty.



Subclonal phylogeny reconstruction. Sample level phylogeny gives a finer picture on how each sample is related to each other, but provides no insight into intra-sample clonal evolution. Using variant allele frequencies (as detailed in SubcloneSeeker paper 1), one can further resolve how samples are related on a subclonal level, providing even more insight into tumor evolution (e.g. sample B appears to be a "purity place" for several clones of different origin).



Via a rapid-autopsy procedure, we gathered 26 metastatic tumor samples across seven organs hours after the loss of life of the patient. In addition to the primary tumor, lymph and metastatic tumor sample.



Based on the analysis results of the SeederSeeker toolkit, we conclude that the tumor invasion history followed these steps:

1. Primary breast tumor invaded lung (B)
2. As the primary tumor kept on mutating, the lung tumor mutated another lung site (S)
3. The further evolved primary tumor invaded lung again, regarding to a new site (O) and B
4. A subclone in B invaded the abdominal organs (L)
5. Another subclone in B further evolved, and went on invading both lung and pentraheal lymph node (S, O, L)
6. The newly established lung metastasis (S) further evolved, and invaded both rib and skull bone tissues, the brain, one liver site, and several other lung sites (C).

It is interesting to observe that

1. Lung seems to be both the earliest and also the most invaded organ, which may be as a result of all blood eventually flows through the lungs.
2. Metastatic events happens in waves, with each wave involving several lungs, and a set of similar organs / destinations.
3. Some metastatic sites act as a "purity place", attracting tumor subclones not locally evolved to invade and to evolve.

Xiaomeng Huang



Subclonal metastatic expansion in triple negative breast cancer

Xiaomeng Huang¹, Yi Qiao¹, Samuel W. Brady², Andrea H. Bild², Gabor T. Marth¹

¹USTAR Center for Genetic Discovery, ²Department of Pharmacology & Toxicology, University of Utah
Contact: xmd1.huang@gmail.com



INTRODUCTION: Metastatic breast cancer is an advanced-stage disease in which the cancer cells have spread to distant organs, e.g. bones, liver, brain and lung. This type of breast cancer accounts for approximately 6%-10% of all breast cancer diagnoses, with a dramatically lower 5-year survival rate of 22%. The goal of this study is to dissect metastatic tumor expansion at a subclonal level, in order to identify its genomic drivers, as well as the aggressive colonizing subclones seeding new metastatic sites.

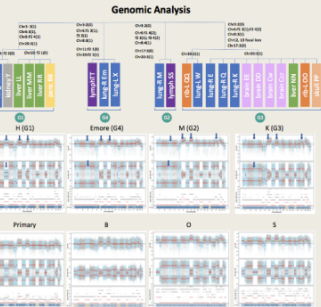
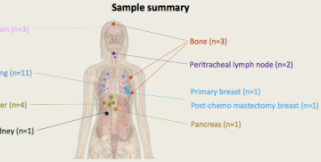
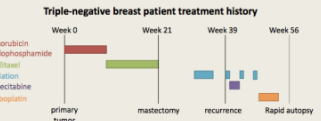
METHOD: As our driving dataset, we have collected two primary tumor biopsies, one at initial diagnosis and one at mastectomy necessitated by the patient's relapse; 26 metastatic tumors across seven organs via a rapid autopsy procedure hours after the patient's decease; as well as two skin biopsies to be utilized as normal control tissues. All samples were subjected to exome-enriched whole genome sequencing with an average genomic coverage of 60X, and higher in exonic regions. FACETS¹ and FreeBayes were used to call copy number variants and SNP variants respectively. We developed SeederSeeker toolkit² to reconstruct the phylogenetic relationships among metastatic samples and tumor evolution at subclone level.

RESULT: Variant calling revealed an inherited missense variant in 5' UTR of BRCA2. Somatically acquired homozygous TP53 missense variants and RB1 loss were present in all tumor samples, explaining the widespread chromosomal aberrations, including both copy number variations (CNVs), and large regions with loss of heterozygosity (LOH). The data suggests the primary breast cancer cell first invaded lung tissue and then migrate to other parts of the body. Samples falling into four groups with distinct CNV and SNP profile indicates that metastatic events occurred in waves: after the initial invading to lung, the tumor invades abdominal organs (liver, pancreas), lymph nodes, and finally, moves to the brain and bones. Interestingly, Subclone analysis reveals that before mastectomy, the cell which seeded samples in group 1 has already migrated. Further more, many inferred ancestor tumor cells were observed in lung metastatic sample B, which may suggest that B was an early metastatic lesion and also an incubator that attracted tumor cells to stay, further evolve and then colonize other sites.

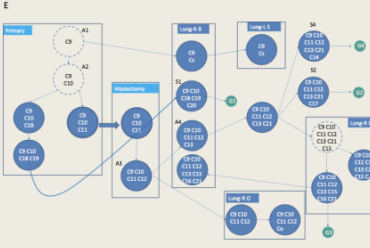
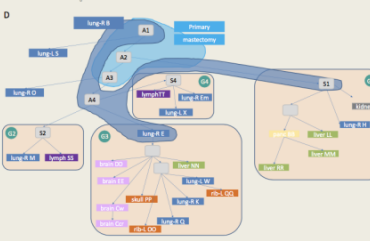
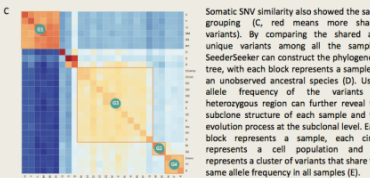
CONCLUSION: Trained on a large (perhaps currently the largest) metastatic biopsy dataset from a single patient, our method provides a novel framework to simultaneously analyze CNV, LOH, and SNV data to reconstruct metastatic tumor expansion at subclonal resolution.

References


1. Shen, R. et al. "ACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing." *Nucleic acids research*, 44(16): e131.
2. Qiao, Y. et al. "SeederSeeker: A computational toolkit for reconstructing metastatic expansion at subclone level." *The Biology of Genomes Annual Meeting 2017*, Cold Spring Harbor Laboratory.



Severe abnormality of chromosomes were observed in all tumor samples in this patients. Despite chromosomal amplifications and deletions, loss of heterozygosity was occurred in most of the chromosomes. Over all, 23 rapid autopsy samples fell into four groups with distinct CNV profile (A). The primary sample and mastectomy sample as well as lung metastatic samples B, S, and O, didn't fall into these four groups, but instead have their own characters (B).



- Novel, reference-free methods for somatic mutation identification

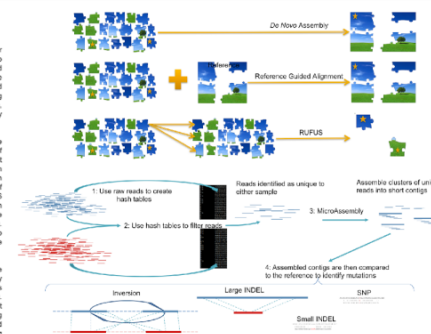


HEALTH
UNIVERSITY OF UTAH

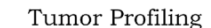
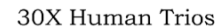
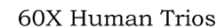


K-mer based variant detection method, RUPUS, shows higher sensitivity as well as lower specificity over mapping based approaches enabling extremely accurate de novo variant detection. RUPUS also shows higher specificity over mapping based variant discovery Read based analysis involves only a tiny fraction of the sequencing genome and is therefore more susceptible to sequencing errors. RUPUS based variant discovery is not disrupted. Unlike the mapping analysis does not require mapping reference information but instead can be carried out in a completely reference-free fashion. RUPUS also identifies variants that prevent the detection of complex variations in highly targeted genomic regions.

plication to the detection of somatic mutation in tumor tissue samples has supported this policy soon in our de novo work and further showcases RUFUS's unique, unbiased, ability to detect mutations of all types and sizes. Of particular interest are insertion deletion events between 50bp and 500bp, lying in this "blind" spot for short-variant detection methods (e.g. GATK, FreeBayes) and structural/CNV detectors (LUMPY, WHAM, etc.). In this data set, RUFUS calls 68 unique calls, of these 95 appear to be true variants missed by mapping methods (FreeBayes and Lumpy). Of the 95 RUFUS-only calls, 84 are indels between 20 and 200bp that no other method is able to detect. This makes RUFUS an ideal method for filling the current hole of medium-length de novo INDEL detection, both in family and tumor sequencing datasets.



- Open to any organism, including ones with no reference sequence
- No reference bias: SNPs and INDELs have an equal chance of being discovered, improves INDEL detection
- Reduced false positive rate due to lack of mapping based errors
- Can identify variations in highly repetitive sequence: telomeres, gene family, centromeres
- All mutation types identified in a single method
- No need for computationally expensive whole genome assembly



Ongoing

- Cancer mutation annotation
 - Set-aside project to integrate with CiViC (Obi and Malachi Griffith)
- Current collaborations
 - Refractory, triple-negative, and metastatic breast and ovarian cancers (Andrea Bild, David Bowtell, Lynn Henry)
 - Hematological malignancies (Michael Deininger, Debbie Stephens, John Byrd)
 - Patient-derived tumor models (Bryan + Alana Welm, K.T. Varley, Jay Gertz)
- Subclone analysis using single-cell sequencing data
 - **Looking for SC collaborations!**

Funding



U01HG006513
R41HG009096
R01HG009000 pending



U24CA209999



UL1TR001067-04S2

National Center
for Advancing
Translational Sciences



THE ERIC S. MARGOLIS
FAMILY FOUNDATION



Team

- PIs: Aaron Quinlan and Gabor Marth (contact)
- Clinical oncology: Theresa Warner (breast, ovarian); Michael Deininger (hematological)
- Experimental/laboratory: Andrea Bild, Samuel Brady
- Computational: Ryan Layer, Andrew Farrell, Dillon Lee (detection); Yi Qiao, Xiaomeng Huang (subclone analysis); Brent Pederson (annotation); Tonya DiSera, Chase Miller (web tools)



Ad

Postdoc positions in computational cancer analysis tool development available.

Email: gmarth@genetics.utah.edu