# Algorithmic Deconstruction of Somatic Copy Number Alterations for Therapeutic Target Discovery in Cancer

**Ellie Rahm Kim[1,2], Shahab Sarmashghi[2,3,4], Rameen Beroukhim[2,3,4]**

1) Department of Neuroscience, The University of Texas at Austin; 2) Cancer Program, Broad Institute of MIT and Harvard; 3) Department of Cancer Biology, Dana-Farber Cancer Institute; 4) Department of Medicine, Harvard Medical School

## Introduction

- Somatic copy number alterations (SCNAs) are changes in the number of copies of specific DNA segments in cells, which can lead to genetic abnormalities.
- SCNAs are among the most prevalent mutational events in cancer, driving tumor growth by modulating the dosage of key genes.[1]
- Despite their significance, the wide-ranging effects of SCNAs often make it difficult to identify their targets, presenting a major challenge for developing cancer therapies.[2]
- Our project aims to address this challenge by creating an advanced computer algorithm that leverages biological mechanisms behind SCNA formation.

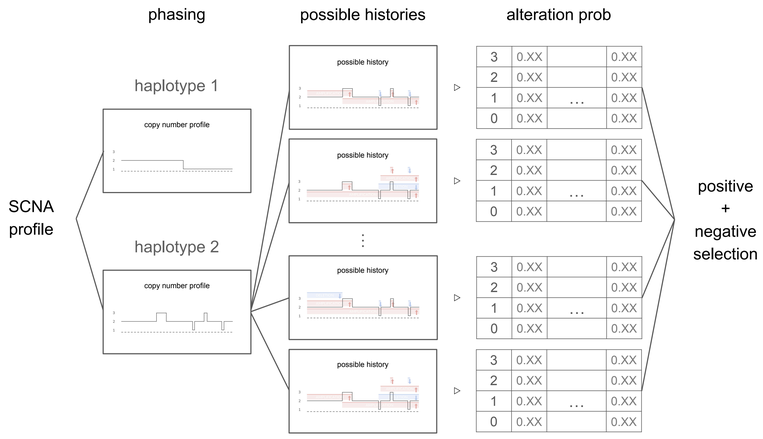
## Hypotheses

- We hypothesize that our algorithm will pinpoint genes whose lack or abundance is pivotal to cancer, providing valuable insights for drug discovery.

## Methods

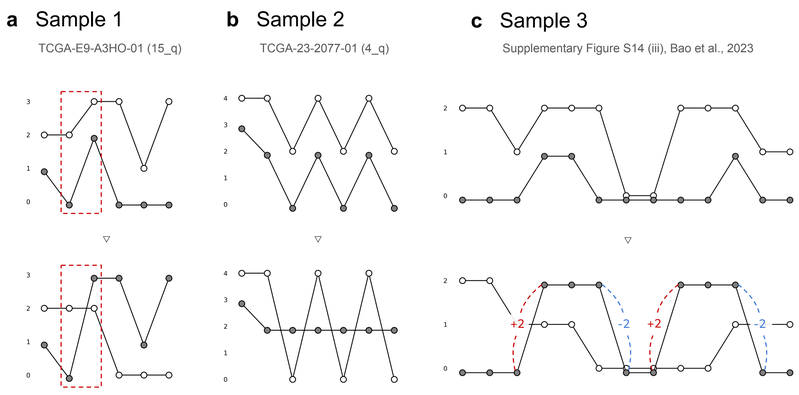
- We leverage a large-scale dataset of tumor samples curated from public consortiums, including The Cancer Genome Atlas (TCGA) and the Genomics England 100 000 Genomes Project.[3,4]
- Haplotype-specific SCNA patterns are inferred using a phasing algorithm based on experimental evidence from previous studies (**Fig. 2**).[5]
- We then apply a mathematical strategy to reconstruct all possible histories of duplications and deletions that could have resulted in a given SCNA profile (**Fig. 3**).
- An iterative algorithm evaluates the likelihood of different histories, which informs the downstream analysis of potential genetic targets.
- Through simulation of SCNA generation, we assess genomic regions that exhibit disproportionate alterations than expected under neutral conditions.
- We introduce the concept of relative fitness to quantify the impact of these alterations on cell survival (**Fig. 7**).

**Fig. 1 | Schematic overview of the framework**



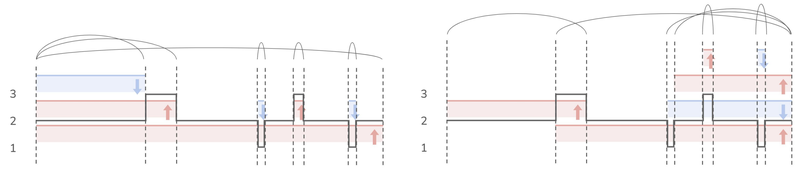

Genomic regions that promote (positive selection) or inhibit (negative selection) cancer growth are identified by integrating results from all possible histories, haplotypes, and tumor samples.

**Fig. 2 | Strategies of haplotype inference**



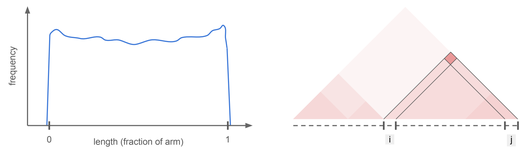

The top panels display the raw data, while the bottom panels show the inferred haplotypes. Experimental observations include minimal breakpoints in copy number (**a**), multiple changes likely accumulating within a single haplotype (**b**), and copy level changes often occurring in pairs (**c**).[5]

**Fig. 3 | Reconstruction of underlying events**



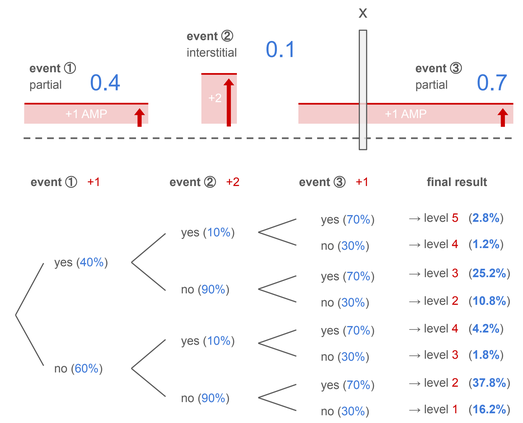
Every possible set of duplication and deletion events is generated by permuting breakpoint pairings within a given profile.

**Fig. 4 | Length distribution of events**



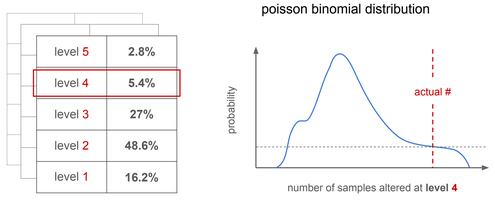

Different event types (**a**, **b**) show distinct probability distributions of extending to specific lengths.[2,6]

**Fig. 5 | Expected alteration levels**




Alteration probabilities are determined by simulating different combinations of events extending to location *x*.

**Fig. 6 | Integration**




Results across tumor samples are integrated using a poisson binomial distribution.

**Fig. 7 | Relative fitness**

$$\frac{\textbf{level } \textit{n} \text{ reality} \,/\, \textbf{level } \textit{n} \text{ expectation}}{\textbf{level } \textit{ref} \text{ reality} \,/\, \textbf{level } \textit{ref} \text{ expectation}}$$

The fitness advantage of level *n* alteration is assessed by comparing observed survival rates to expected ones, relative to the reference level.

## Results

- Our work introduces four novel algorithms, each tackling a central component of the methodology, forming a comprehensive and sophisticated workflow.
- This method overcomes the limitations of previous tools (e.g., GISTIC and BISCUT) by incorporating information from all SCNA event types and detecting regions under both positive and negative selection.[2,7]

## Conclusion

- Preliminary analyses show that these algorithms are effective in identifying candidate genomic regions of interest.
- We anticipate this work to uncover previously undiscovered targets of SCNAs, contributing to the advancement of personalized therapies that are tailored to the unique genetic composition of each patient's cancer.
- Furthermore, it will serve as a vital resource for researchers studying the underlying causes of different tumors.

## References

1) Beroukhim, Rameen, et al. "The landscape of somatic copy-number alteration across human cancers." Nature 463.7283 (2010): 899-905.
2) Shih, Juliann, et al. "Cancer aneuploidies are shaped primarily by effects on tumour fitness." Nature 619.7971 (2023): 793-800.
3) Weinstein, John N., et al. "The cancer genome atlas pan-cancer analysis project." Nature genetics 45.10 (2013): 1113-1120.
4) Turnbull, Clare. "Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000 Genomes Project." Annals of Oncology 29.4 (2018): 784-787.
5) Bao, Chunyang, et al. "Genomic signatures of past and present chromosomal instability in Barrett's esophagus and early esophageal adenocarcinoma." Nature communications 14.1 (2023): 6203.
6) Fudenberg, Geoffrey, et al. "High-order chromatin architecture determines the landscape of chromosomal alterations in cancer." Nature Precedings (2011): 1-1.
7) Beroukhim, Rameen, et al. "Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma." Proceedings of the National Academy of Sciences 104.50 (2007): 20007-20012.

## Acknowledgements

I extend my gratitude to the members of the Beroukhim lab for their immense support and expertise throughout this project.

ERK is supported by the Pediatric Oncology Student Training Grant from Alex's Lemonade Stand Foundation for Childhood Cancer.