

The repertoire of copy number alteration signatures in human cancer

Xue-Song Liu^{1, ID}
liuxs@shanghaitech.edu.cn

Ziyu Tao^{1-3*} Shixiang Wang^{1-3*} Chenxu Wu^{1-3*} Huimin Li¹ Tao Wu¹
Xiangyu Zhao¹ Wei Ning¹ Guangshuai Wang¹

¹ School of Life Science and Technology, ShanghaiTech University, Shanghai 201203, China
² Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China
³ University of Chinese Academy of Sciences, Beijing, China

Abstract

DNA alteration signatures are recurring patterns that are the imprints of mutagenic processes accumulated during the evolution of cancer cell. Despite the importance of copy number alteration (CNA) in cancer progression, comprehensive understanding about the mutational processes and signatures of CNA is still lacking. Here we developed a method to categorize CNA based on various fragment properties, which reflect the consequences of mutagenic processes and can be extracted from different types of data, including whole genome sequencing (WGS), whole exome sequencing (WES) and SNP array. The signature of CNA has been extracted from 2778 pan-cancer analysis of whole genomes (PCAWG) WGS samples, and further validated using 10851 the cancer genome atlas (TCGA) SNP array dataset. Novel copy number signatures associated with haploid chromosome have been identified. The activities of some copy number signatures consistently predict cancer patients' prognosis. This study provides a repertoire for understanding the signatures and mutational processes of CNA.

Methodology

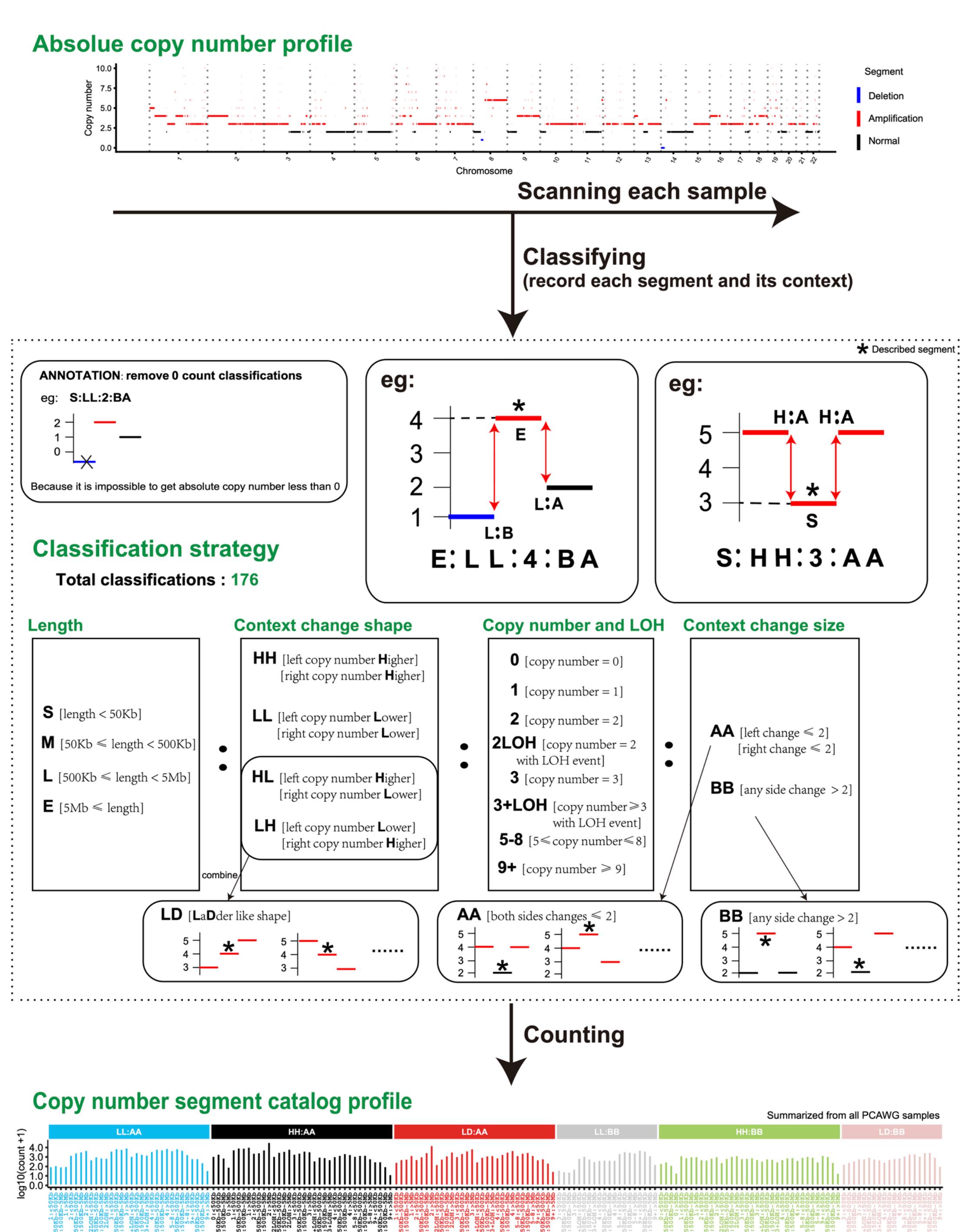


Figure 1: CNA classification strategy for signature analysis.

For each CNA segment, the following features have been considered: 1, segment context, including segment shape and copy number change number; 2, Absolute copy number; 3, LOH status; 4, segment size. In total 176 types of CNA segments have been defined accordingly.

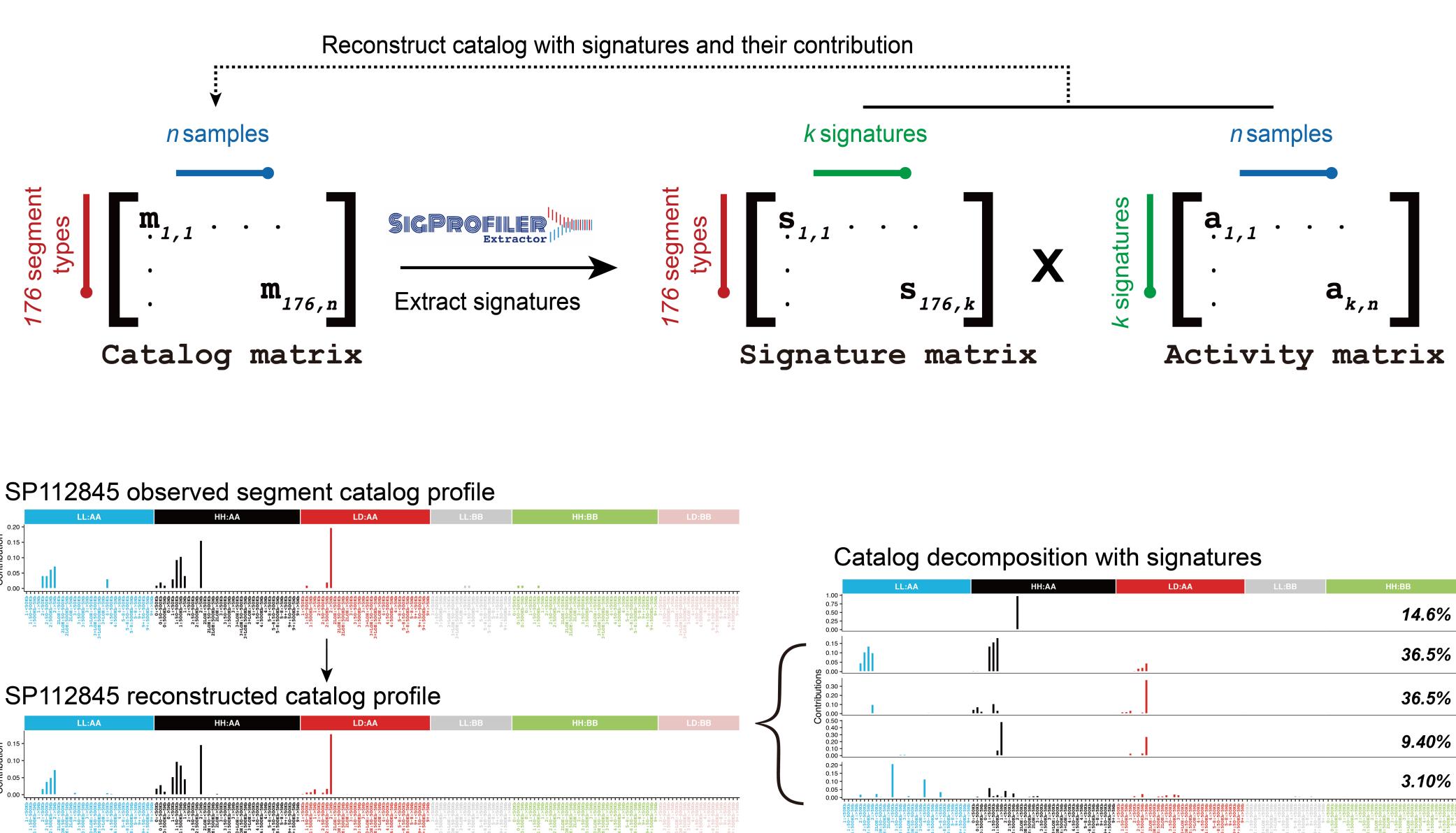


Figure 2: De novo CNA signature extraction with sigprofiler.

We selected the signature extraction solution as the maximum signature number which meets the following criteria: 1. No over fit. 2. Stability should be at least local maximal. 3. Mean cosine distance should be as small as possible. Based on the rules, we selected 14 signatures for PCAWG copy number data.

Results

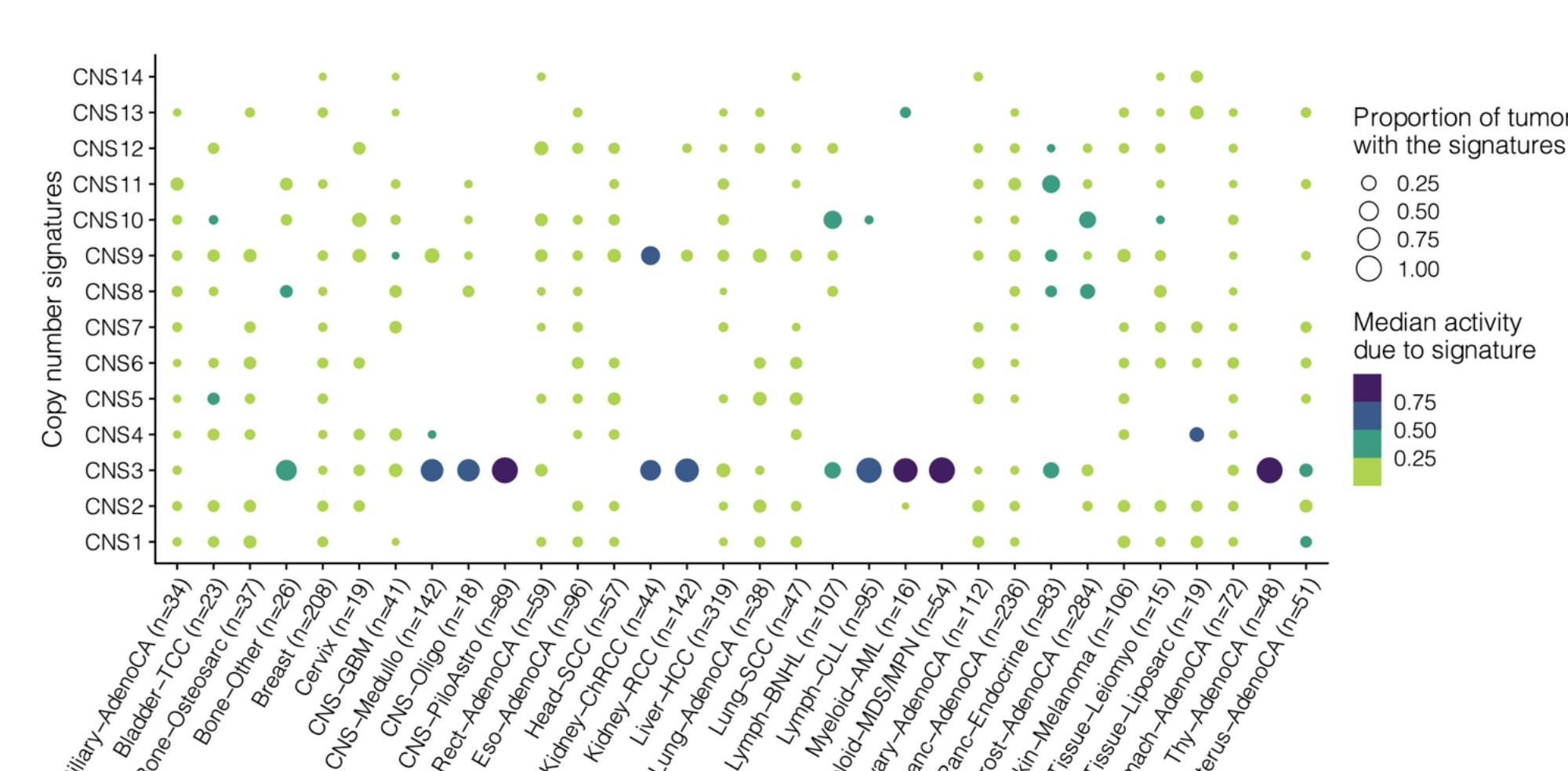


Figure 3: Proportion of tumors with the signature and the median activity of the signature are shown for 32 PCAWG cancer types. For each individual tumor, only signatures that contribute to >=5% of the total are counted.

Signatures	Representative sample profile	Prominent features	Potential mechanism
CNS1	SP117605	(A) CN amplification (3-4 copies) with different size flanked by higher copy (5-8 copies) segment; (B) Adjacent CN change is irregular; (C) Localized amplification with various fold.	Likely Chromoanaphysis due to MMBIR
CNS2	SP123894	(A) Localized CN amplification(3 copies >5Mb); (B) Adjacent CN change is within 2; (C) Oscillating CN with 3-4-3 pattern.	Unknown
CNS3	SP105708	(A) Long DNA segment (>5Mb) without copy number alteration (CN=2).	Genome without CNA
CNS4	SP116529	(A) Many localized high CN amplification (copy<2, size<5Mb); (B) Adjacent CN change is >2 copies; (C) Poor overall survival.	EcDNA or neochromosome
CNS5	SP117809	(A) Mostly large-sized (>50Mb) DNA segment with amplification in different levels(3-8 copies); (B) Adjacent CN change is mostly regular; (C) Unequal CN amplification in the terminal of chromosome.	Likely BFB
CNS6	SP117443	(A) Many DNA segments of various size (>5Mb, or 50Mb-5Mb); (B) Adjacent CN change is irregular.	LOH and amplification
CNS7	SP116903	(A) CN duplication (1-2 copies) of 50Mb-5Mb flanked by neutral segment (copy number=2 with size); (B) Adjacent CN change is with 2 copies; (C) Localized oscillating CN with 2-1-2 pattern.	Chromosome fragmentation and amplification
CNS8	SP116839	(A) Shallow CN change (>5Mb) flanked by neutral segment with similar size; (B) Adjacent CN change is within 2 copies; (C) Localized oscillating CN with 2-1-2 pattern.	Chromothripsis
CNS9	SP118048	(A) Large sized (>5Mb) CN amplification (mostly 4 copies); (B) Majority of adjacent CN change is less than 2; (C) The signature is associated with WGD.	WGD
CNS10	SP112887	(A) DNA segment of small to medium size(<5Mb) or (50Mb-1 Mb) interspersed with large sized (>5Mb) neutral segment(CN=2); (B) Adjacent CN change is within 2.	Homozygous deletion
CNS11	SP113532	(A) Large sized (>5Mb) shallow deletion along with large sized (>5Mb) neutral segment; (B) Adjacent CN change is within 2;	Haploid chromosome
CNS12	SP117636	(A) Large sized (>5Mb) DNA segment (CN=3) along with large sized (>5Mb) neutral segment harboring LOH; (B) Adjacent CN change is within 2; (C) Poor overall survival.	Triplid chromosome
CNS13	SP113664	(A) Localized high number of amplification (CN=4-8) and shallow deletions; (B) Adjacent CN change is mostly >2 copies.	Unknown
CNS14	SP116450	(A) Evenly distributed small sized(<50Mb) segment (CN=1-4) flanked by DNA segment with small CN change; (B) Adjacent CN change is within 2 copies; (C) The signature is strongly associated with HRD.	HRD

NOTE: To better visualize the signature features, only some chromosomes are plotted for some samples.
Abbreviations: CN: copy number; MMBIR: microhomology-mediated break induced replication; BFB: breakage-fusion-bridge; HRD: homologous recombination deficiency

Figure 4: Representative CNA profile, prominent features and potential mechanisms for each identified CNA signatures extracted in PCAWG dataset.

An important purpose of CNA signature analysis is to identify the underlying mutational processes for CNA, and classify CNA based on mutational processes. Potential mutational processes for CNA include intrinsic inducers and extrinsic inducers. Intrinsic CNA inducers include: double-strand break repair defect (HRD etc.); cell cycle defect; DNA replication defects; telomere loss, etc.

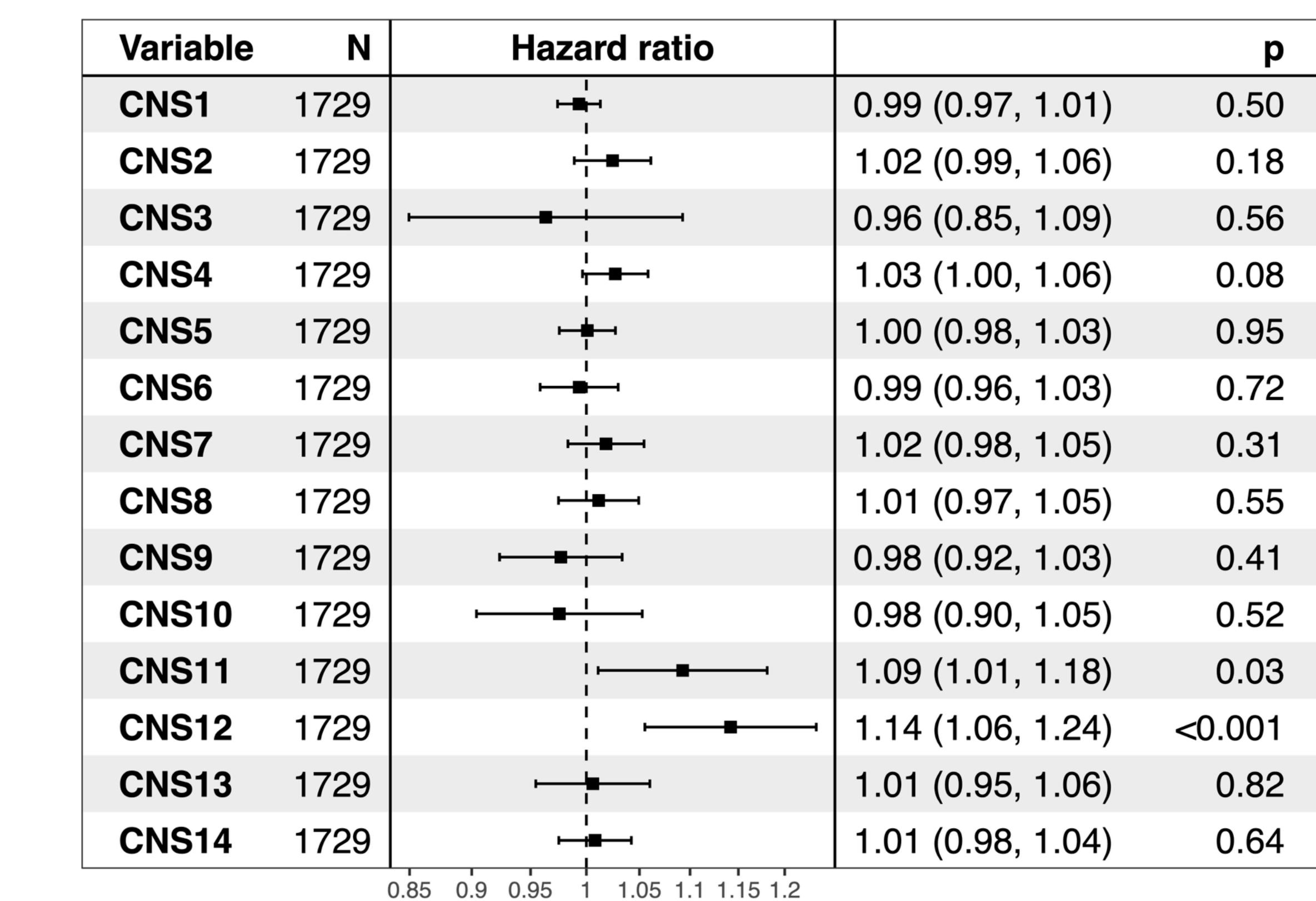


Figure 5: CNA signature activity and cancer patients' prognosis.

The CNA signatures extracted from cancer patients could be cancer prognosis biomarkers. In pan-cancer level, activities of CNS11, CNS12 are significantly associated with poor OS.

Conclusion

- We developed a unified and comprehensive method for copy number signature analysis.
- Our method can be applied in cancer patients with copy number profiles generated with WGS, WES or SNP array data. - Our copy number signature analysis method is based on a novel and comprehensive method to catalog copy number segments.
- copy number signatures could be biomarkers to guide cancer precision medicine.

More

You can scan QR code at bottom center to see online analysis report. All code and related data are published at https://github.com/XSLiuLab/Pan-cancer_CNA_signature.

Acknowledgments

We thank ShanghaiTech University High Performance Computing Public Service Platform for computing services. This work was supported by The National Natural Science Foundation of China (31771373), Shanghai Science and Technology Commission (21ZR1442400), and startup funding from ShanghaiTech University.

Recent works of our lab

- Wang, Shixiang, et al. "Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes." PLoS genetics 17.5 (2021): e1009557.
- Wang, Shixiang, et al. "UCSCXenaShiny: an R/CRAN package for interactive analysis of UCSC xena data." Bioinformatics (2021).
- Wang, Shixiang, et al. "Sigflow: an automated and comprehensive pipeline for cancer genome mutational signature analysis." Bioinformatics 37.11 (2021): 1590-1592.

©Cancer Biology Group 2021

Research group led by Xue-Song Liu in ShanghaiTech University. Lab website is shown in QR code at bottom right.

