CNSigs: An R package for the identification of copy number signatures

David Wellman, Shawn Striker, Sinclair Stockard, Zachary Weber, Sarah Asad, Katharine Collier, Elizabeth Adams, Jerome Bey, Daniel G. Stover
Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, OH.

The James
The Ohio State University
Comprehensive Cancer Center

Background

- Copy number aberrations (CNAs) are gains and losses of large genomic segments present across most cancer types and are a hallmark of cancer for genomic alterations. However, the processes underlying CNAs and characteristic patterns of CNAs are poorly understood.
- Using single nucleotide variant (SNV) data, bioinformatic advances have identified underlying mutational signatures resulting from distinct mutational processes. The development of algorithms able to uncover similar signatures for CNAs, rather than SNVs, is still in its infancy.
- Here we present an analysis package for the R programming language called CNSigs that allows for the robust and reproducible derivation of copy number signatures and apply

CNSigs: An R package for the identification of copy number signatures

Hypothesis

We hypothesize that there will be a set of copy number signatures that are both unique to or shared across cancer types and that these signatures will provide insight into underlying molecular mechanisms.

Methods

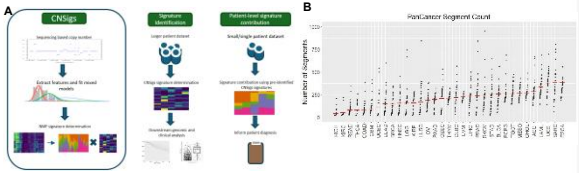


Figure 1A: Using segmented data files from DNA sequencing, six copy number features are extracted by CNSigs for signature determination: segment size, the number of breakpoints per 10 megabase bin, number of copy number oscillation events, average size of changepoints, average copy number, and number of breakpoints per chromosome arm, along with ploidy. Based on extracted copy number features, mixed model approaches and non-negative matrix factorization are utilized to derive CNA signatures across cancer types.
Figure 1B: Somatic copy number variations are a predominant form of DNA alteration in several cancer types.

Results

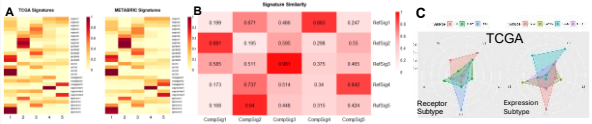


Figure 2A: To verify the reproducibility of the signatures, we derived five signatures from TCGA and METABRIC datasets in BRCA. **Figure 2B:** These signatures were shown to be consistent between datasets with an average cosine similarity of 0.89. **Figure 2C:** The identified signatures maintained a separate biological identification from known breast cancer receptor-based or expression-based subtypes.

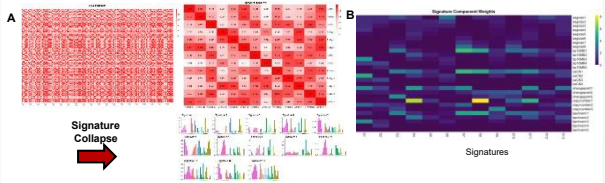


Figure 3A: To assess the robustness of the algorithm in a pan-cancer approach, the pipeline was applied to 11 different cancer types from TCGA dataset. 13 signatures were derived by merging the highest similarity score iteratively until no cosine similarity was >0.9. **Figure 3B:** The 13 derived signatures are visualized to show the unique component contributions.

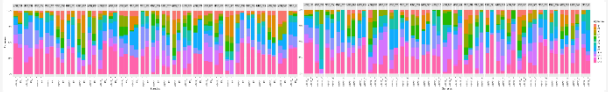


Figure 4: To investigate potential application to ctDNA, we evaluated patients with metastatic ER+/HER2- paired tumor and ctDNA sequencing acquired at the same time (n=24 pairs), demonstrating that CNSigs are detectable via WES ctDNA.

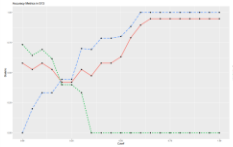


Figure 5: To identify what level of signature exposure is active or inactive, comparison statistics were generated from ctDNA using paired whole tumor samples as truth. The signature activation threshold of 0.15 was chosen to balance sensitivity with specificity.

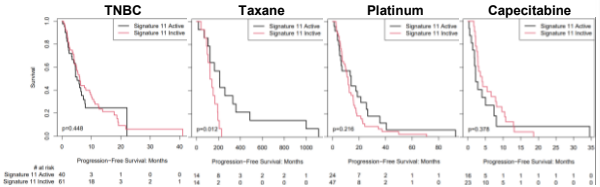


Figure 6: Exploratory application of CNSigs in a large cohort of mTNBC patients revealed association of progression-free survival for the presence of one specific signature, CNSig 11, with extended response to taxane chemotherapy.

Conclusions

The CNSigs R package allows researchers to easily analyze their own samples to derive copy number signatures and evaluate clinical associations. These signatures are replicable across datasets and provide a different biological classification beyond known biological subtypes. We demonstrate potential application in ctDNA and association with treatment response. The development of this package allows further investigation of underlying processes that may be responsible for these CNA fingerprints.

Funding and Acknowledgments

susan g. komen.