PS3-04-24: Common and rare germline variants together with somatic mutations alter the integrity of cancer hallmark regulatory networks

Jiawei Dai, PhD[1] ; Mate Posta, MD[2,3] ; Michal Marczyk, PhD[4,1] ; Tao Qing, PhD[1] ; Balazs Gyorffy, MD, PhD[2,3,5] ; Lajos Pusztai, MD, DPhil[1]

[1]Yale School of Medicine, [2]Semmelweis University, [3]HUN-REN Research Centre for Natural Sciences, [4]Silesian University of Technology, [5]University of Pecs

Yale University School of Medicine

SAN ANTONIO BREAST CANCER SYMPOSIUM®

UT Health San Antonio Mays Cancer Center

AACR American Association for Cancer Research

Background

- Each cancer harbors hundreds of somatic DNA alterations that have accumulated against the background of thousands of germline variants.
- Emerging evidence suggests that both common and rare germline variants, together with recurrent and rare "long-tail" somatic mutations, can collectively disrupt cancer-related pathways in a highly individualized manner.
- **Our goal was to develop a computational tool, named Cancer Systems integrity score (CanSys), that quantifies the contributions of germline and somatic alterations to pathway-level disturbances, and to apply it to public databases to uncover disruptions of key biological processes involved in cancer biology in each individual cancer sample.**

Methods and Materials

CanSys architecture

Input: a VCF file , optionally with mRNA expression data.

Calculate gene-level impact scores: for each gene g with N variants, combine the CADD deleteriousness score[1,2] with the DepMap importance score[3].

$$geneImpact\,score_g = \frac{\sum_{n=1}^{N} CADD\,score_n}{100} \cdot geneDepMap\,score_g$$

Expression data refines scores by assigning 0 to non-expressed genes.

Quantify pathway-level disturbances: for each pathway S, map gene-level impact scores to its gene set and compute an enrichment score (ES) using a modified GSEA.

$$P_{Hit}(S,i) = \sum_{gene_j \in S, j \le i} \frac{impact\,score_j}{impact\,score_{total}}$$

$$P_{Miss}(S,i) = \sum_{gene_j \notin S, j \le i} \frac{1}{N - N_S}$$

$$ES(S) = max_{i=1}^{N}(runningsum\,for\,ES_i) = max_{i=1}^{N}(|P_{Hit}(S,j) - P_{Miss}(S,j)|)$$

Compute the CanSys disturbance metric. NES denotes the normalized ES.

$$CanSys\,score(S) = NES(S) \cdot \frac{\sum_{gene_j \in S} impact\,score_j}{N_S}$$

Visualization: build a public-facing interface to display individual sample-level results.

Clinical datasets

We applied CanSys to The Cancer Genome Atlas (TCGA; n=9,384)[4] samples and used the Breast Cancer Genome Guided Therapy Study (BEAUTY; n=111)[5] as validation.

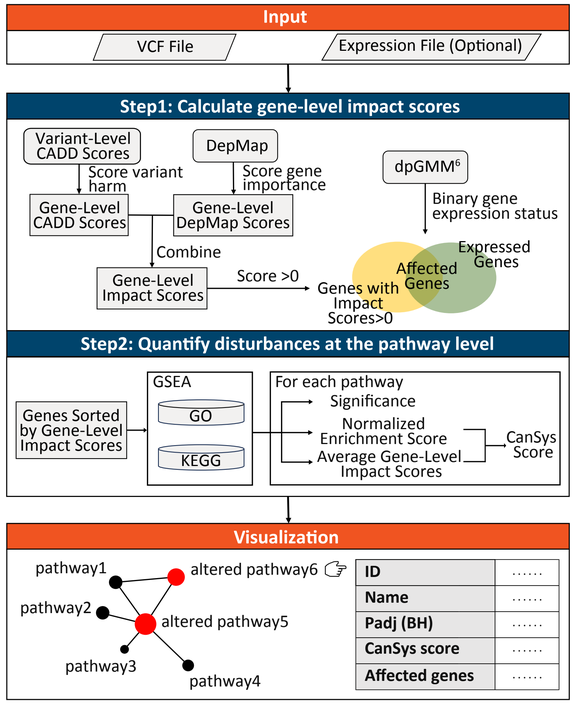


Figure 1. Overview of the CanSys architecture.

Results

Web-based application

CanSys enables the analysis and visualization of custom VCF files and map sample-level alterations onto GO or KEGG networks .

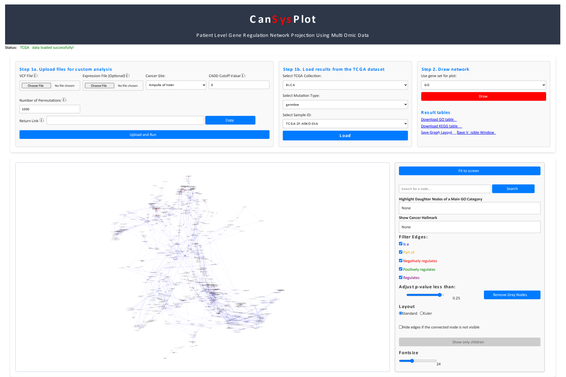


Figure 2. An example of using the interactive network visualization tool.

Overall patterns of pathway disturbances in TCGA samples

Many biological processes are altered at varying frequencies across cancers, contributing to each cancer's distinct clinical course.

Table 1. Number of significantly affected GO and KEGG pathways (Padj<0.25) across TCGA samples, grouped by % of samples affected.

Affected Type	<25%	≥25%, <50%	≥50%, <75%	≥75%
GO, Somatic	14806	404	61	12
GO, Germline	14841	175	91	176
GO, Overlap (Both)	15156	95	29	3
KEGG, Somatic	157	19	0	0
KEGG, Germline	153	6	7	10
KEGG, Overlap (Both)	174	2	0	0

Cancer hallmark-linked pathway disruptions

All cancers carry common and rare **germline** alterations affecting cell cycle, telomere maintenance, and DNA repair pathways. Pathway alterations from **somatic** mutations differ markedly across cancer types.

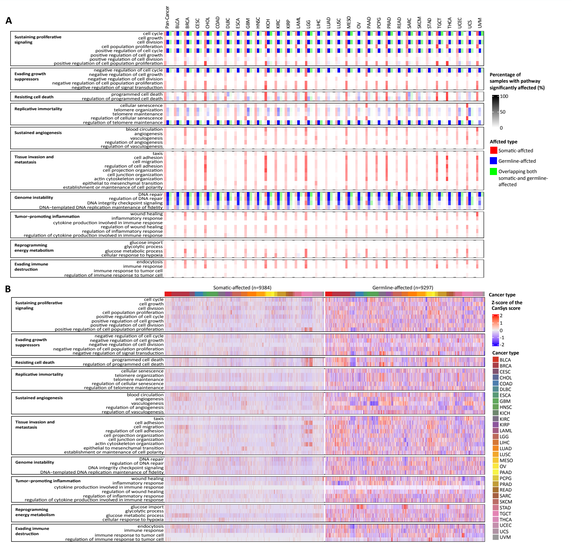


Figure 3. Alteration prevalence of 52 hallmark-associated GO terms in TCGA. A: Percentage of samples significantly affected. B: CanSys scores.

Conclusions

- We develop a tool to quantify germline and somatic pathway-level disturbances in each individual cancer, with a public facing interface (**https://cansysplot.com**).
- Germline variants mainly affect cell cycle, telomere maintenance, and DNA repair, while somatic mutations disrupt adhesion, motility, metabolism, and signaling pathways.
- The unique combination of pathway alterations might explain the unique behavior of each cancer.

Contact

Jiawei Dai
Pusztai Laboratory
Yale Cancer Center, Yale School of Medicine
Email: jiawei.dai@yale.edu

References

1.Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014, 46:310-315.
2.Schubach M, Maass T, Nazaretyan L, Röner S, Kircher M: CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. Nucleic Acids Res 2024, 52:D1143-d1154.
3.Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM: Defining a cancer dependency map. Cell 2017, 170:564-576. e516.
4.TCGA Research Network: https://www.cancer.gov/tcga.
5.https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001050.v1.p1.
6.Polanski A, Marczyk M, Pietrowska M, Widlak P, Polanska J: Initialising the EM algorithm for univariate Gaussian, multi-component, heteroscedastic mixture models by dynamic programming partitions. International Journal of Computational Methods 2018, 15:1850012.

This presentation is the intellectual property of the author/presenter. Contact them at jiawei.dai@yale.edu for permission to reprint and/or distribute.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com