**Abstract**

*Cancer driver genes* are mutated genes that promote cancer by causing or contributing to unusual cell proliferation. Driver mutations are accompanied by many passenger mutations, which do not effect cancer. Distinguishing between driver and passenger genes is of high importance for diagnostics and potential therapies. Mutations in driver genes are caused by genomic alterations. Those include single nucleotide changes, copy number changes, segmental translocations, etc.

*Personalized detection* of driver genes can be obtained by the integration of genomic and transcriptomic data from a single patient. We have previously shown how to detect and rank driver genes in an individual based on single nucleotide variants (SNVs) and differentially expressed genes. The integration of multiple types of genomic alterations, including large scale alterations, can reveal individual deregulated mechanisms more accurately.

**Previous Work**

PRODIGY ([Dinstag and Shamir, Bioinformatics 2019](#)) ranks genes with single nucleotide variations by their chance to be the drivers of cancer in an individual. Ranking is based on the gene’s effect on *pathways deregulation*.

**Algorithm Outline:**
- Find differentially expressed genes (DEGs).
- Define deregulated pathways that are enriched with DEGs.
- Build a weighted deregulation network for each gene and pathway p.
- Compute influence scores of each gene g on each pathway p, with the PCST (Prize Collecting Steiner Tree) method.
- Rank the genes by their overall influence score.

**Copy number variations:**
We use copy number segments as an additional data, and specifically focus on genomic segments that went through homozygous deletions or high-level copy number amplifications. The patient specific process is:
- Find the most significant CNV segments.
- Extract genes which reside in them.
- Run PRODIGY with the extracted genes as driver candidates, in addition to the SNV genes.

**Prize Collecting Steiner Tree for a set of genes:**
We examine three approaches to compute the influence score of a set of genes originating from the same CNV segment on each pathway:
- Score each gene separately and compute the average.
- Score each gene separately and choose the maximum.
- Score all genes in the set with a Prize Collecting Steiner Forest formulation rather than a tree.

**Results and Future Directions**

In 5 cancer types, we noticed that about 60% of the CNV segments contain only 1 gene (we call those solo CNV genes), and about 20% contain 2 - 10 genes.

We ran PRODIGY with the solo CNV genes, and we observed an improvement in the performance in comparison to using only SNVs.

**Expansion to Large Scale Variations**

SNVs compose ~40% of the experimentally validated drivers to date in the COSMIC catalogue. This strongly motivates us to extend the analysis.

**Summary Table**

<table>
<thead>
<tr>
<th>Variation</th>
<th>SNV</th>
<th>CNV</th>
<th>Translocations</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Drivers</td>
<td>248 (37%)</td>
<td>65 (10%)</td>
<td>314 (47%)</td>
<td>37 (6%)</td>
</tr>
</tbody>
</table>

SNV – single nucleotide variation, CNV – copy number variation

This study was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University.