Elucidation of Pathways in Cancer Associated with Survival

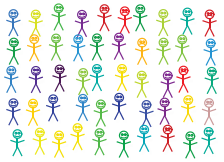
Andrew M. Gross, Janusz Dutkowski, Trey Ideker

Bioinformatics and Systems Biology Program, Department of Medicine, University of California- San Diego, CA

Overview

When seeking to isolate the oncogenic drivers from the spectrum of genomic alterations in cancer, current methods achieve power by citing the presence of a selective advantage in tumor cells and show that these driver events are over-represented verses a background model [1,2]. The inherent limitation of such methods is their inability to annotate rare genomic alterations. We have developed a series of methods for the characterization of such genomic biomarkers not based on prevalence of mutations, as has been done previously, but by their associations with clinical and molecular phenotypes. As opposed to selectivity based methods, these are catered towards elucidating which events differentiate a cancer cohort rather than those events responsible for the original tumorigenesis.

We first address two major confounding factors with the analysis of such associations, namely 1) the presence of global signals such as patient age and mutation rate and 2) the sparse nature of genomic alterations across a cancer cohort limiting the power to make such associations. These issues are addressed by incorporating these global signals as covariates in our models and merging genes onto pathways to boost the sample size (and power) for the association tests. We have systematically applied this method to all available TCGA datasets [3] using mSigDB pathways [4], uncovering associations between somatic mutation status and patient event-free-survival. The results of this analysis recapitulate known hallmarks of cancer and implicate a number of novel pathways. We then show initial work in the generalization and extension of these methods to diverse types of clinical and molecular data.

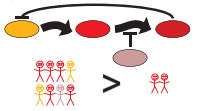


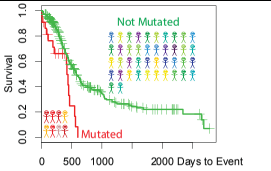
Question What events differentiate people with cancer?

Problem Global variables such as age and mutation rate confound analysis of genes and pathways. **Solution** Modeling covariates along with genomic events acounts for local and global contributions to phenotype.

While two patients may harbor the same mutation, its effect is often dependent on global factors such as age.

Problem Events are sparse across cohort. **Solution** Expanding from a gene-centric to a pathway-centric prospective improves sample size and power.

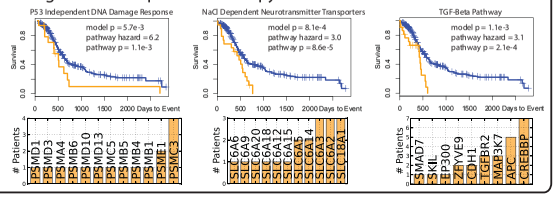




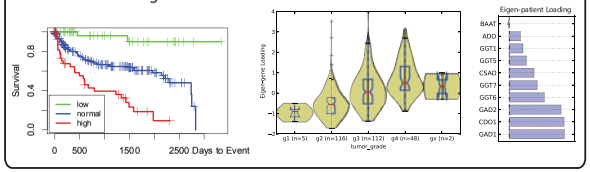
Result Pathway level events can be associated with phenotypes such as patient survival.

Many Pathways have Survival Associations

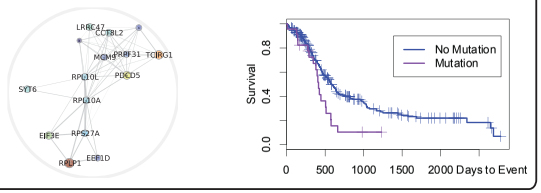
In the **Ovarian** cohort, no single gene can be associated with event-free-survival below a FDR threshold of 20%. Expanding to pathways uncovers a number of candidates we hypothesize to have an impact on tumor growth or response to therapy.



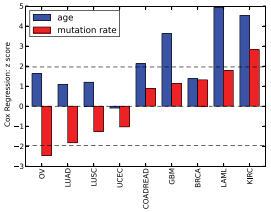
In current work, we have generalized these methods to real valued measurements such as gene expression and methylation. Measurements are mapped to genes and integrated on pathways by taking the first principal-component. Shown here is the association of methylation levels in the KEGG **Taurine and Hypotaurine metabolism pathway** with both patient survival ($p < 10e-20$) as well as tumor-grade in the **KIRC** cohort.



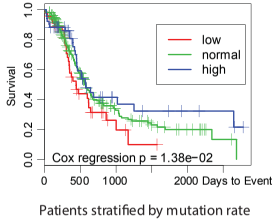
In current work we have extended the framework for pathways developed here to de-novo modules obtained by clustering mutations on a probabilistic gene-by-gene functional network. Shown here is a connected module of associated genes and the corresponding survival phenotype.



Global Variables are Associated with Survival



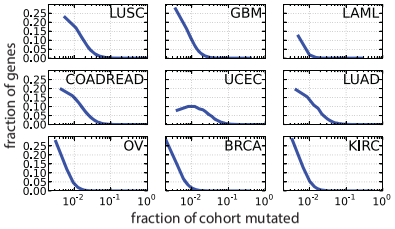
Z-scores for age and mutation-rate association with patient survival assessed by Cox regression.



Patients stratified by mutation rate in the **Ovarian** cohort. Here low and high represent +/- 1 standard deviation from the mean.

Genetic Mutations are Sparse

For Cox PH model it is recomended that the minimum sample size is 5-9 events per variable to keep bias small [5]. Until cohort sizes become considerably larger, this restricts analysis to only highly mutated (and usually well studied) genes. Shown here are density plots for each cohort.

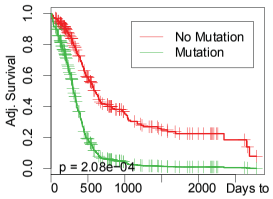


Cox PH Model Adjusts Global Covariates

- Non-parametric survival function adjusted with parametric covariates
- Assumption of time-independence of covariates
- Assumes a proportional increase with increasing values of covariates (linear in log- space)
- Interaction of pathway mutation variable with covariates allows model to explain away 'noisy' mutations in patients with high mutation rates, or mutations accumulated with age

```
model = Surv(time, event) ~ covariates + interactions
```

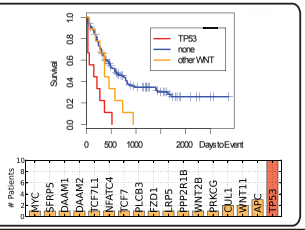
	hazard ratio	p-value
pathway	3.10	0.0002
age	1.01	0.0740
rate	0.67	0.0033
pathway:age	0.99	0.7071
pathway:rate	0.71	0.6077



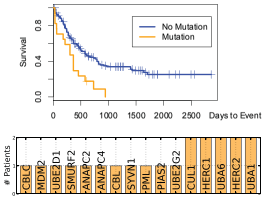
Left, table of covariates and interaction terms that are fit in the Cox model. **Right figure** shows the effect of varying the pathway mutation variable in the fit model while holding the age and rate constant at their median values.

Pathway Level Analysis Can Characterize Sparse Mutations

In the LAML dataset, **TP53** is muated in 10 patients who are associated with poor survival. Expanding to all of the genes in the **WNT pathway** adds an additional 10 patients with a similar phenotype.



Often mutations are very sparse across the pathway but have a common effect. The **ubiquitin mediated proteolysis pathway** contains 138 genes, in the **LAML** cohort 19 of those contain mutations covering 20 patients. While this pathway is definately not under selection, taken together these patients have poor prognosis.



Conclusion

Selectivity is useful for finding genes that cause cancer, but we are also interested in markers for cancer progression and treatment success.
Integration of sparse genetic events over networks and pathways boosts our power to uncover novel targets.
Finding **associations** with clinical phenotypes allows for annotation of new genes and pathways.

While rare genomic alterations elude single gene centered analysis, their characterization through pathway and network based approaches is an important milestone on the path towards patient-tailored treatment of cancer.

[1] Dees, N. D. et al. MuSiC: Identifying mutational significance in cancer genomes. Genome Research 22, 1589–1598 (2012).
[2] Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Research 22, 398–406 (2011).
[3] Bell, D. et al. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615 (2011).
[4] Subramanian, A. From the Cover: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 102, 15545–15550 (2005).
[5] Vittinghoff, E. & McCulloch, C. E. Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. American Journal of Epidemiology 165, 710–718 (2007).