Role of metabolic and glycolytic gene copy number alterations in oncogenesis

Jessica Zhou, Nicholas A. Graham

USC Viterbi School of Engineering

The Mork Family Department of Chemical Engineering and Materials Science

2017 Undergraduate Research Symposium for Scholarly and Creative Work

Introduction: Modern technological advances have generated massive amounts of data, consequently necessitating meaningful interpretation of these data. This project takes a bioinformatics approach to omics data to make sense of the genomic and metabolic drivers of cancer and inform development personalized medical treatments, as well as further experimental studies.

***Abstract*:** Tumor cells must generate a significant amount of energy to support their constant and aggressive growth. Consequently, they exhibit distinctively active metabolic profiles compared to healthy cells. This phenotype is a hallmark of cancer, and has inspired research to characterize the link between altered metabolism and oncogenesis. One driver of cancer is the acquisition of genetic mutations and copy number alterations (CNAs). A previous study profiling the proteomic response to CNAs in breast cancer revealed a number of significant positive correlations (cis) between CNA-mRNA and CNA-protein pairs, of which a large part of genes correlated on both CNA-mRNA and CNA-protein were identified as cancer-relevant genes, suggesting cis-regulatory role for these CNAs on transcriptional and translational levels. This project takes a similar approach to metabolic and glycolytic genes in order to characterize the underlying mechanisms by which metabolic dysregulation drives oncogenesis. Here, we correlate TCGA genomic data with proteomic and phosphoproteomic data to identify and characterize regulatory behaviors of metabolic and glycolytic genes.

Methods: Breast and ovarian tumor **proteomic** and **phosphoproteomic data** was used from Mertins *et al*[1] and Zhang *et al*[2], respectively. **CNA data** was obtained from The Cancer Genome Atlas (TCGA), and we observed correlations between genes and proteins for which CNA and proteomic data were both available. A list of **genes in the human genome** annotated to include known roles of the gene (e.g. oncogene, tumor suppressor gene (TSG), glycolytic, metabolic, etc.) was obtained from Unigene.

1. TCGA samples with **PCA scores** based on DNA copy numbers mapped to samples for proteogenomic data was provided in the aforementioned papers to ensure random distribution of samples for an unbiased study (Fig. 1A-B).
2. Correlations between CNA and proteomic data were calculated per gene per sample, for each breast and ovarian tumors. Correlations were measured with **Spearman's rho** (ρ), and associated p-values were calculated to quantify significance of correlations. A floor requirement of ≥ 20 data points was implemented to ensure that ρ-values are only computed for genes with a minimum number of samples with valid values, as small sample sizes will yield inflated p-values.
3. Spearman's rho p-value was corrected for false discovery error using the Benjamini-Hochberg procedure, and p-values were normalized by signed logarithmic transformation.
4. Data filtered to include only correlation data for genes found in both breast and ovarian tumors and visualized by plotting (breast, ovarian) for log signed p-values for each gene (Fig. 2).
5. Established baseline p-value significance threshold to be all genes with positive correlations for which $p < 0.05$ in both breast and ovarian tumors.
6. Fisher's Exact Test applied at different FDR thresholds to detect enrichment of genes of a different functions in the subset of genes meeting the defined significance thresholds (Table 1).

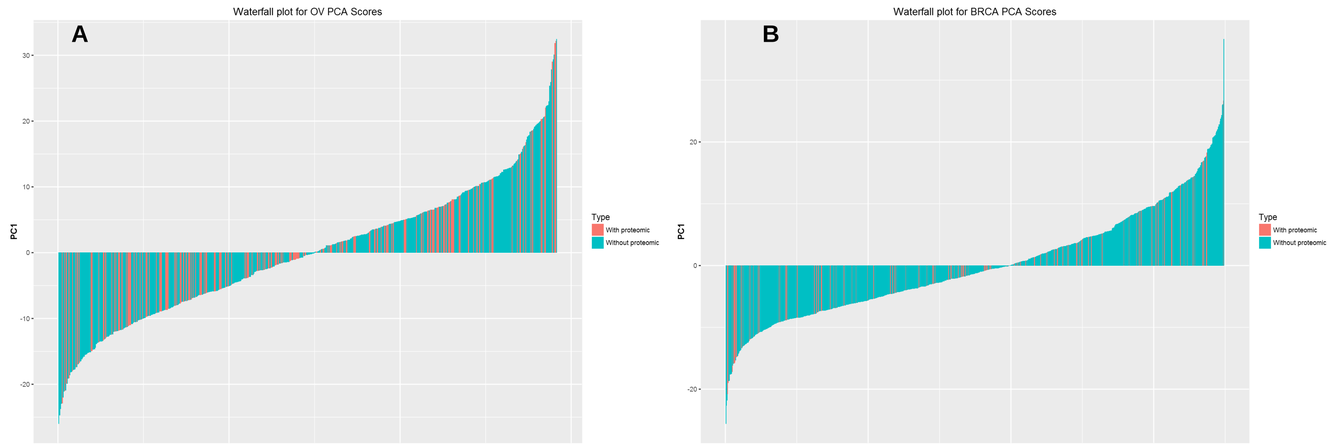

Figure 1. Mapping TCGA tumors for which CNA data is available to tumors with proteomic and phosphoproteomic data (red) in A) ovarian, and B) breast, confirmed a random distribution., which was quantified by a Kolmogorov-Smirnov test.

Gene function	FDR < 0.05	FDR < 0.01
Oncogenes and TSG	0.558184	0.758787
Metabolic genes	0.056216	0.045055
Glycolytic genes	0.223906	0.222733

Table 1. Fisher's Exact Test p-values by gene function. Genes falling below the threshold value did so in each ovarian and breast tumors. Significant enrichment of metabolic genes was observed across both tumor types.

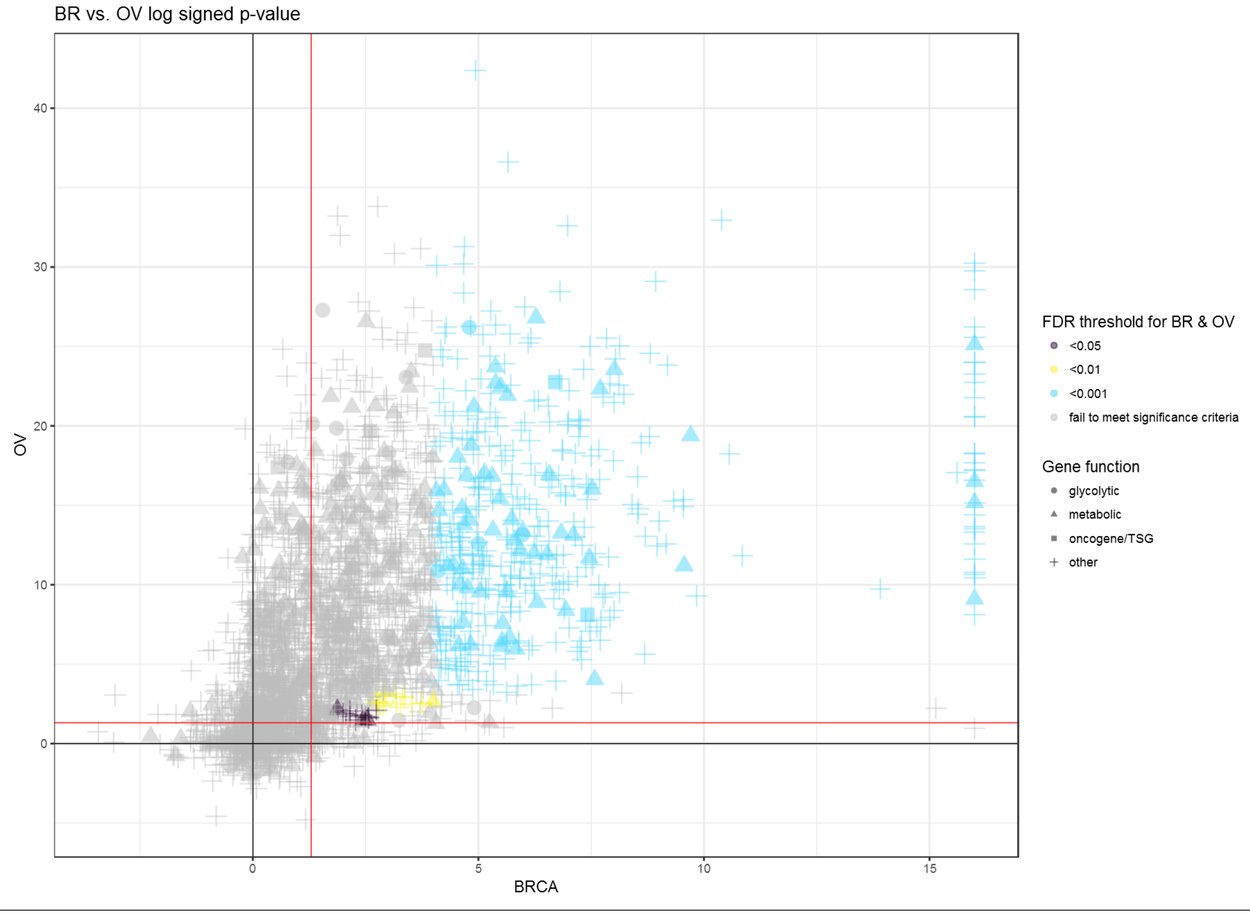

Figure 2. Scatterplot of log signed p-values for CNA-protein correlations in ovarian tumors versus breast tumors, with normalized p-value thresholds denoted by red axes and genes with FDRs below threshold in both ovarian and breast tumors colored based on confidence of statistical significance. Point shapes denote the primary classes of gene functions observed in this study.

Results & Conclusions: We found enrichment of genes with metabolic function within genes that exhibited strongly positive CNA-protein correlations. There was a vast number of genes whose functions were not catalogued at the initial execution of the analysis; thus, further annotation of our current set of genes correlated across both ovarian and breast tumors could yield to findings of more enriched genes of a particular function. These data so far indicate that metabolic genes play a cis-regulatory role, and that dysregulation of their copy numbers drives cancer development across both breast and ovarian tumors, suggesting a potential pan-cancer effect. Their specific role could be further characterized by performing a gene set enrichment of these metabolic genes to determine whether they belong to a particular metabolic pathway. Additionally, we could increase the resolution of our Fisher's Exact Tests by reducing our groups of genes to more specific functions (i.e. examining genes involved in glycolysis-gluconeogenesis separately from all glycolytic genes). Ultimately, we hope that our findings can be applied to the design of personalized medicine for treating cancer.

Acknowledgements

Funding provided by the Rose Hills Foundation Science and Engineering Fellowship, the University of Southern California Provost's Fellowship and Women in Science and Engineering Research Fellowship.

References

(1) Mertins, P., et al. (2016). Proteogenomics Connects Somatic Mutations to Signalling in Breast Cancer. Nature 534(7605): 55–62. (2) Zhang, H., et al. (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell, 166(3), 755-765.

wise · The Rose Hills Foundation · USC University of Southern California