UNCOVERING RNA-BINDING PROTEINS IMPLICATED IN HUMAN CANCERS BY INTEGRATING GENOMICS WITH NETWORK-BASED APPROACHES

Bobak Kechavarzi, Sarath Janga

Indiana University School of Informatics, Indianapolis, Indiana

Abstract

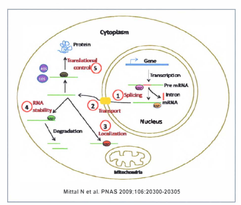
RNA binding proteins (RBPs) are key regulatory mechanism in the cell. Their functions are varied depending on their localization, but many have been identified as essentially in translational regulation and cell proliferation. Their dysfunction has been linked to the development of various disease phenotypes and cancers. Using Human BodyMap v2.0 data and data made available through The Cancer Genome Atlas (TCGA), we proposed to observe the patterns of expression of RBPs in sixteen healthy tissues and across nine cancers, and their altering profiles. Additionally, by incorporating BioGrid protein-protein interaction data and CORUM protein complex information, we explore how network properties of the RNA may infer their dysfunction in cancers. The prognostic effect of RBPs classified by expression and network properties in breast cancer were determined using KM-Plotter. We observed that RBPs as a class are more highly expressed than other factors in the 16 human tissues, and furthermore that they are generally up-expressed in cancers. A smaller subset of RBPs (30) is many-fold higher expressed across a large portion of the observed cancers. Network metrics showed no significant differences, except for shortest path distances between subsets (Wilcox, $p < 2x10^{-16}$). Similarly, complex size and membership did not show any trends or significant differences. The negative prognostic effect seems to be associated with mean path lengths of RBPs and their interaction with a highly dysregulated subset of RBPs.

Introduction and Objectives

RNA binding proteins (RBPs) have been identified as key regulatory components within the cell. Their function is largely dependent on their expression and localization within a cell; they may be involved with processes ranging from alternative splicing to RNA degredation. Investigations in *Saccharomyces cerevisae* have indicated the correlation between expression levels and expression variations of RBPs and their number of connections (Mittal, Roy, Babu, & Janga, 2009).

While cancer is a complex genetic disease, many regulatory factors have been identified as being dysregulated. In particular, RBPs have been shown to lead to a cancer phenotype after their function had been altered from basal expression (Wurth, 2012). Enhanced eIF4E and HuR expression has been implicated in the translation initiation of mRNAs encoding mostly for prooncogenic proteins, or other cancer phenotypes. Additionally Sam68 regulates the alternative splicing of cancer-related mRNAs. Even without directly acting on oncogenic genes, RBPs have been documented as major components in cell cycle and cell proliferation functions.

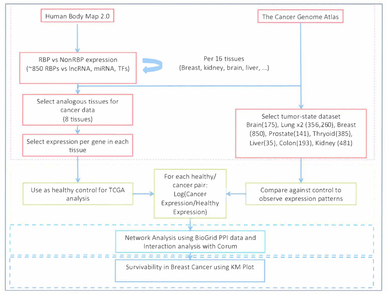
RBP Functions in the cell



Knowing the sensitive nature of RBPs as regulators in cells, and their identification in mammalian systems, we can begin to investigate their profiles of dysregulation in various disease states. The purpose of this study is to observe the expression patterns of RBPs as a class in healthy, human tissues, their fold-change expression in nine human cancers, and network properties that may differentiate the groups leading to altered expression. Finally, after a subset of RBPs has been identified based on their profiles and network metrics, their contribution to survivability of patients with breast cancer will be investigated.

Methods

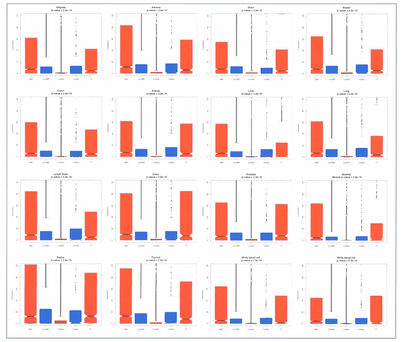
Methodology Workflow



Healthy human transcript expression levels were collected from the Human BodyMap 2.0 data across 16 tissues and 80 individuals (ages 19-80 years old). For this data, expression levels of 850 RBPs identified by (Castello et al., 2012) were compared against all other transcripts, and additionally other regulatory factors in the tissues (miRNA, lncRNA, and TFs). Once this profile was observed, patient RNAseq data for nine cancers was downloaded from The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga) for the eight analogous tissues: breast, brain, kidney, prostate, liver, lung, and thyroid. Then, the expressions were compared between the two states to classify a subset of RBPs that are strongly up-regulated.

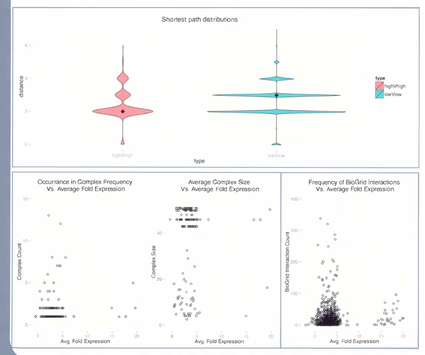
The network properties of these classified RBPs were compared, using BioGrid data to inform their interactions in humans, and their interaction in protein complexes was identified using Corum data (Ruepp et al., 2008). Subsets of RBPs were defined based on their interactors and path lengths for determining their effects on survivability using KM-Plotter (Győrffy et al., 2010).

Results



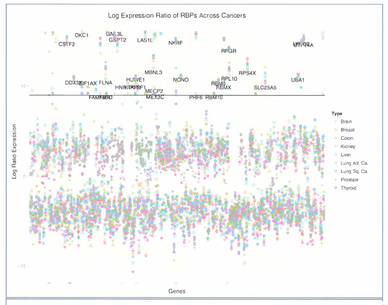
By plotting the log-ratio of cancer to healthy median expression across the nine cancers, we can observe the trends of dysregulation of factors in humans. It becomes apparent that there are different levels of up and down-regulation in the changes of state. There exists a subset considered highly upregulated where their log-fold ratio expression is greater than 9 across 6 or more of the cancers. In RBPs there are a larger proportion of highly-upregulated genes (right, top) compared to nonRBPs (right,botom) (hypergeometeric test, P=0.03).

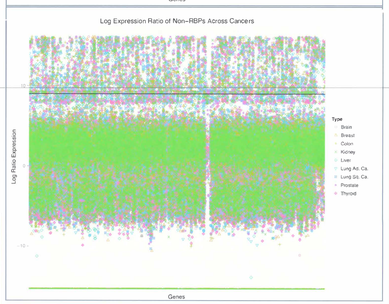
Of these classified RBPs we note several that are implicated in complex genetic disorders and cancer. RBPs such as DDX3X are part of the DDX family that has been documented to impact *Wnt* pathways leading to the developments of cancers (Cruciat et al., 2013). LAS1L was identified as being an interactant with PELP1 that is implicated in pancreatic cancers (Kashiwaya et al., 2010). HUWE1 is a member of the HECT family of E3 ubiquitin ligases that has been identified as being overexpressed in breast, lung and colorectal cancers (Bernassola et al., 2008).



Observing the expression levels of RNA binding proteins across all 16 tissues (left) suggested that RBP expression was significantly higher (Wilcox, $p<2x10^{-16}$) than non-RBPs. Visual observation of the trends also shows that in some tissues, such as testes, lymph, and ovary had particularly higher expression when compared to nonRBP transcripts.

To determine regulatory impact of RBPs their expression were compared against similar factors in the human tissues. The trend remains the same that RBPs are significantly higher in these 16 tissues as compared to the other factors (Wilcox, $p<2x10^{-16}$). Only transcription factors had similar levels of expression in select tissues: ovary, prostate, testes, and lymph. This suggests a need to closely regulate these particular tissues.

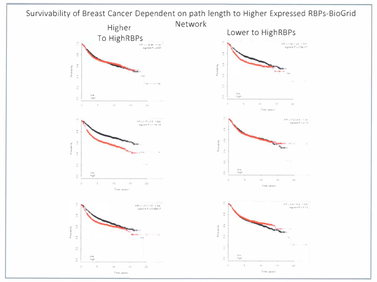




BioGrid PPI data was utilized to generate a static, undirected PPI network. Network interactions (closeness, betweeness, clustering coeff.) were observed and were not signficant. Shortest paths to the classifications of RBPs (high,low upregulation) were found to be significantly different. Highly upregulated genes were found to have smaller shortest paths (left, top).

Interaction data from CORUM and BioGrid were used to observe RBP membership in complexes, and their average size (left, bottom). Examining these characteristics between the two groups did not identify any correlation to expression levels, nor significant differences in distributions.

Results Continued



For survivability analyses we utilized the KM-Plotter utility. RBPs of the separate classifications were ranked based on their mean path lengths to all other members in the BioGrid PPI data. The shortest and longest path lengths for a gene in each subset were selected, in addition to a random select of five median-length genes. This led to an observation of a trend for a gene to have more of a negative prognostic impact the longer its mean path length.

The path lengths for a gene in one group to another (i.e. high to high, high to low, etc.) was then determined and similarly ranked. In this case, the trend observed was that, for highly upregulated RBPs, the longer path lengths still had a negative impact on prognosis. However, for the other classification of RBPs, shorter path lengths to the higher upregulated RBPs indicated a negative prognostic impact compared to those with longer path lengths. This suggests that these highly upregulated RBPs may impact prognostic impact if they are one of few interactors with a lesser dysregulated gene.

Conclusions

We observed the expression patterns of identified RBPs in 16 healthy, human tissues. We found that as a class, RBPs were significantly more expressed than non-RBPs. Furthermore, we determined that RBPS were significantly higher expressed than other regulatory elements (TFs, lncRNAs, miRNAs). In some cases, transcription factors were also highly expressed in certain tissues. This illustrates a need for tight regulatory control in those sites.

The patterns of dysregulation from healthy to cancer state was examined in nine human cancers. After calculating the log-fold change in median expressions we were able to identify RBPs in different dysregulated states. A subset of 30 RBPs were identified as having greater than 9 log-fold change in expression in 6 or more cancers. Some RBPs in this class have information supporting our observation of this dysregulated state in cancers. RBPs also have a larger proportion of highly upregulated genes in comparion to nonRBPs. This illustrates their potential sensitivity to dysregulation in their interactants, or their own dysregulation perpetuating the cancer phenotype.

The network properties of the subgroups of upregulated RBPs generally show no significant difference. This suggests that the proportionality or counts of interactions between these types do not inform the reason behind their dysregulation or expression levels. However, by observing their network properties, we can see a difference in path lengths within these groups. When incorporating path length information in determining negative prognostic impact we observe that in general, longer path lengths lead to more negative prognoses. However, when paths are shorter to the highly dysregulated RBPs, prognostic impact decreases, suggesting some effect of dysregulation of these factors leading to cancer.

References

Bernassola, F., et al. (2008). "The HECT Family of E3 Ubiquitin Ligases: Multiple Players in Cancer Development." Cancer Cell 14(1): 10-21.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, Benedikt M., Strein, C., . . . Hentze, Matthias W. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, *149*(6), 1393-1406. doi: http://dx.doi.org/10.1016/j.cell.2012.04.031

Cruciat, C.-M., Dolde, C., de Groot, R. E. A., Ohkawara, B., Reinhard, C., Korswagen, H. C., & Niehrs, C. (2013). RNA Helicase DDX3 Is a Regulatory Subunit of Casein Kinase 1 in Wnt-β-Catenin Signaling. *Science*, *339*(6126), 1436-1441. doi: 10.1126/science.1231499

Győrffy, B., Lanczky, A., Eklund, A., Denkert, C., Budczies, J., Li, Q., & Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment*, *123*(3), 725-731. doi: 10.1007/s10549-009-0674-9

Kashiwaya, K., et al. (2010). "Involvement of the Tubulin Tyrosine Ligase-Like Family Member 4 Polyglutamylase in PELP1 Polyglutamylation and Chromatin Remodeling in Pancreatic Cancer Cells." Cancer Research **70**(10): 4024-4033.Cancer Res; 70(10); 4024-33. ©2010 AACR.

Lukong, K. E., Chang, K.-w., Khandjian, E. W., & Richard, S. (2008). RNA-binding proteins in human genetic disease. *Trends in Genetics*, *24*(8), 416-425. doi: http://dx.doi.org/10.1016/j.tig.2008.05.004

Mittal, N., Roy, N., Babu, M. M., & Janga, S. C. (2009). Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks. *Proceedings of the National Academy of Sciences*, *106*(48), 20300-20305. doi: 10.1073/pnas.0906940106

Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., . . . Mewes, H. W. (2008). CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Research*, *36*(suppl 1), D646-D650. doi: 10.1093/nar/gkm936

Wurth, L. (2012). Versatility of RNA-Binding Proteins in Cancer. *Comparative and Functional Genomics*, *2012*, 11. doi: 10.1155/2012/178525

Contact Information

Bobak Kechavarzi: bkechava@indiana.edu
Dr. Sarath Janga: scjanga@iupui.edu

IUPUI