1. RNA Seq – Rice

In this study, a novel GRAS transcription factor gene named OsGRAS23, which is located in a drought-resistant QTL interval on chromosome 4 of rice, was isolated. The expression of OsGRAS23 was induced by drought, NaCl, and jasmonic acid treatments. The OsGRAS23-GFP fused protein was localized in the nucleus of tobacco epidermal cells. A trans-activation assay in yeast cells demonstrated that the OsGRAS23 protein possessed a strong transcriptional activation activity. OsGRAS23-overexpressing rice plants showed improved drought resistance and oxidative stress tolerance as well as less H2O2 accumulation compared with the wild-type plants. Furthermore, microarray analysis showed that several anti-oxidation related genes were up-regulated in the OsGRAS23-overexpressing rice plants. The yeast one hybrid test indicated that OsGRAS23 could bind to the promoters of its potential target genes.

https://scilifelab.github.io/courses/rnaseq/labs/CuffDiff

1. RNA Seq – Brain

Public Glioblastoma Datasets > RNA Seq> pathway analysis

1. Cell cycle checkpoint
2. Immune pathways – upregulated in female
3. Na-K Transport Pathways
4. MicroRNA – Tulane

**HYPOTHESIS AND SPECIFIC AIMS:**

**...**

**Aim #3: Develop miR biomarkers for prostate cancer diagnosis.**

**...**

**RESEARCH STRATEGY**

**Significance:**

**Innovation:**

**Approach:**

**Aim #3. Develop miR biomarkers for prostate cancer diagnosis.**

**Introduction:** Prostate cancer is the most common noncutaneous malignancy among men in the United States and the second most common cause of cancer mortality.

...

**Justification, feasibility and preliminary data:** We have identified a several significant miRs in previous studies.

...

**Research Design:** The general strategy of Aim #3 is to identify deferentially expressed (DE) miRs, map them to corresponding targeted genes, and map those genes to relevant biological processes and pathways. Using DE miRs build a miR signature that will accurately predict prostate cancer grade (Figure 1). The developed signature will be validated using previously published data. The entire analysis will be perform using BGL PLATFORM devolved at Digicon.

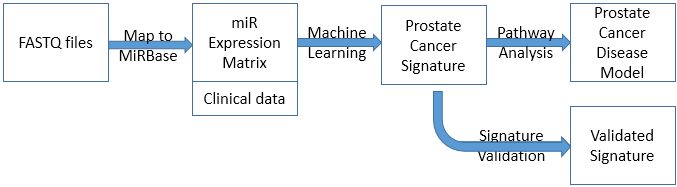


Figure 1. Prostate Cancer Signature analysis pipeline

1.1 miSeq sequencing and QC: We will run a several QC steps to evaluate quality of FASTQ file statistics such as number of total reads, number of mapped reads, and miSeq platform technical parameters. Specifically, we will run Principal Component Analysis and outlier detector using Local Outlier Factor algorithm (12).

1.2 Map FASTQ reads to existing miRs: We will use miRDeep2 tool (1) to calculate total reads counts for each known miR presented in the current version of MiRBase (2) in each sample. We will keep reads count for miR with miRDeep2 score zero or higher (3). Reads counts will be aggregated in one matrix using the 75th percentile normalization (4).

1.3 Build miR prostate cancer signature: We will use Random Forest algorithm (5) to build prostate cancer signature. Random forest (RF) is one of the most robust and accurate machine learning method. It can be used to rank the importance of variables in classification problem in a natural way. We will use RF Gini impurity score (6) to select a list of miRs that are significant for discriminating low and high grade of prostate tumor. Selected miRs will be used to build a Random forest classifier. Outcomes of this step will be several type of prediction quality estimations: cross validation matrix, area under the curve (AUC), principal component analysis (PCA), prediction error, specificity, and sensitivity.

1.4 Mapping signature miRs to biological pathways: Direct effects of miRs on gene pathways are not studies extensively. The most common way to identity human biological processes and molecular pathways that can be effected by miRs is done by identifying genes that targeted by those miRs. miRWalk2.0 is a comprehensive archive, supplying the largest available collection of predicted and experimentally verified microRNA (miRNA)-target interactions (7). We will query the database with the list of the signature miRs. A result list of targeted genes will be sorted by number of each gene occurs in the result list. Gene Set Enrichment Analysis (GSEA) and DIANE tools will be used to reveal relevant biological processes and molecular pathways (8,12).

1.5 Signature validation: The signature form 1.3 will be tested using a data set that has not been used for developing the miR signature. The validation set will be compiled from … Outcome of this step will be several type of prediction power estimations: cross validation matrix, area under the curve (AUC), prediction error, specificity, and sensitivity.

**Anticipated Results, potential pitfalls, and alternative approaches:**

**REFERENCES**

1. **miRDeep2 Tutorial.** <https://wiki.hpcc.msu.edu/display/Bioinfo/miRDeep2+tutorial>
2. **miRBase: the microRNA database.** <http://www.mirbase.org/index.shtml>
3. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N., **miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades**. *Nucleic Acids Res. 2012 Jan;40(1):37-52.*
4. Garmire LX, Subramaniam S., **Evaluation of normalization methods in mammalian microRNA-Seq data. *RNA.*** *2012 Jun;18(6):1279-88.*
5. Kam HY., [**The Random Subspace Method for Constructing Decision Forests"**](http://ect.bell-labs.com/who/tkh/publications/papers/df.pdf)**,***IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998; **20** (8): 832–844.
6. **Decision tree learning.** <https://en.wikipedia.org/wiki/Decision_tree_learning>
7. **miRWalk 2.0: a comprehensive atlas of predicted and validated miRNA-target interactions.** <http://zmf.umm.uni-heidelberg.de/mirwalk2>
8. **Gene Set Enrichment Analysis.** <http://www.broad.mit.edu/gsea/>
9. Bild A, Yao G, Chang J, Wang Q, Potti A, Chasse D, Joshi M-B, Harpole D, Lancaster J, Berchuck A, Olson J, Marks J, Dressman H, West M, Nevins J. **Oncogenic pathway signatures in human cancers as a guide to targeted therapies**. *Nature*. 2006;**439**(7074):353-7.
10. Su J, Yoon B-J, Dougherty E. **Accurate and reliable cancer classification based on probabilistic inference of pathway activity**. *PloS one*. 2009;**4**(12).
11. **DIANE Tools.** <http://diana.imis.athena-innovation.gr/DianaTools/>
12. Breunig MM, Kriegel HP., Ng RT., and Sander J., [**LOF: identifying density-based local outliers.**](http://www.dbs.ifi.lmu.de/Publikationen/Papers/LOF.pdf) *Proc. ACM SIGMOD, 2000*
13. Metagenomics – FDA
14. Tutorial – Frederick CC
15. Installation of Tools – Kraken
16. Liquid Biopsies
17. Single Cell Sequencing