Development of Cancer GMR Software Package: Personalized Cancer Gene Therapy

Nneka Ede and Dumitru A. Iacobas, Ph.D. (Faculty Advisor)

Personalized Genomics Laboratory - Center for Computational Systems Biology College of Engineering, Prairie View A&M University, Prairie View, TX

Abstract

Gene Master Regulators (GMR) are defined as genes whose highly protected expression levels by the cellular homeostatic mechanisms control major functional pathways. As such, their silencing selectively kills the cells they command.

This report presents the friendly-user software to identify the GMRs that reduces the calculation time of the huge amount of genomic data from few days to ten minutes. The Graphic interface was developed using Python. The user needs only to upload individual gene expression levels in four biological replicas (format .csv). The Cancer GMR Software is expected to be an invaluable tool for both basic science research and clinical oncology.

Introduction

A very rich literature compared gene sequences and expression profiles in tissues collected from healthy and cancer donors to identify the biomarkers to be used for both diagnostic and therapeutic purposes. However, after decades of intense research, thousands of papers in major journals and billions of dollars spent no viable solution resulted. Therefore, a novel, revolutionary approach of cancer gene therapy is much needed.

In 2016, Dr. Iacobas introduced the Gene Master Regulator (GMR) concept and developed the experimental protocol and an advanced mathematical algorithm to identify the GMRs from gene expression data in cancer nuclei of surgically removed solid tumors or in blood samples of leukemia patients. The composite metric termed Gene Commanding Height (GCH) was introduced to establish the gene hierarchy in each cell phenotype of the profiled tissue, with GMRs topping the GCH-based hierarchy. GCH combines the relative estimate of the transcript abundance control and the estimate of the expression coordination. Previous publications proved that cancer nuclei and surrounding quasi-normal tissues are governed by different GCH hierarchies and that manipulating the expression of a gene has larger effects in cells where that gene has higher GCH.

Methods

All programs were developed using the Anaconda distribution of Python 3 and Jupyter Notebooks as Integrated Development Environment. packages such as SciPy, NumPy, and Pandas were used for data manipulation, computation and to boost run time speed, and the graphical user interface Tkinter to package the software.

Weighted Pathway Regulation Software

This software is able to identify and quantify the significantly altered functional pathways in the cancer phenotype.

$$WPR_{\Gamma}^{(cancer)} = \left\langle \mu_i^{(normal)} \left(\left| x_i^{(cancer)} \right| - CUT_i^{(cancer)} \right) \left(1 - p_i^{(cancer)} \right) \right\rangle_{all\ i \in \Gamma} \quad , \quad where:$$

$$\left| x_i \right| > CUT_i = 1 + \sqrt{2\left(\left(REV_i^{(cancer)} \right)^2 + \left(REV_i^{(normal)} \right)^2 \right)}$$

$$\left| x_i \right| = \begin{cases} \dfrac{\mu_i^{(cancer)}}{\mu_i^{(normal)}} & , \text{ if } \mu_i^{(cancer)} > \mu_i^{(normal)} \\ -\dfrac{\mu_i^{(normal)}}{\mu_i^{(cancer)}} & , \text{ if } \mu_i^{(cancer)} < \mu_i^{(normal)} \end{cases}$$

$$REV_i(\varepsilon) = \frac{1}{2}\left(\sqrt{\frac{r_i}{\chi^2\left(r_i; 1-\varepsilon/2\right)}} + \sqrt{\frac{r_i}{\chi^2\left(r_i; \varepsilon/2\right)}} \right) \sqrt{\sum_{k=1}^{R_i} \left(\frac{s_{ik}}{\mu_{ik}} \right)^2}$$

s_{ki} = standard deviation of the expression level of gene i probed by spot k

μ_{ik} = average expression level of gene i probed by spot k

$r_i = \lambda R_i - 1 =$ number of degrees of freedom

λ = number of biological replicas ($\lambda \geq 4$)

R_i = number of microarray spots probing redundantly transcript i

Gene Commanding Height Software

The software establishes the GCH hierarchy of the genes and identifies the GMRs of cancer nuclei.

$$GCH_i = \exp\left(WES_i + CPT_i \right) \quad , \quad where:$$

$$WES_i(\varepsilon) = \ln\left(\frac{\langle REV \rangle}{REV_i(\varepsilon)} \right) \quad , \quad \langle REV \rangle = \text{median of all transcripts}$$

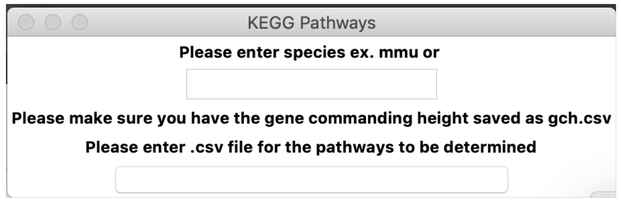
$$CPT_i^{(all)} = \frac{N \sum_{j=1, j \neq i}^{N} \rho_{ij}^2}{\sum_{k=1}^{N} \left(\sum_{j=1, j \neq i}^{N} \rho_{kj}^2 \right)} \quad , \quad \rho_{ij} \text{ is the pair-wise Pearson correlation}$$

coefficient of the expression levels of genes i and j in biological replicates

Results

The software is composed of several subprograms that can be run either together or independently for intermediate analyses. Application users need only to upload (.csv) extension files with gene symbolism and the biological replicas conditions.

Interface



Genomic Data

The software was tested on transcriptomic data from samples of papillary and anaplastic thyroid cancer, metastatic clear cell renal cell carcinoma, (Fig) and prostate cancer obtained by Iacobas Lab.

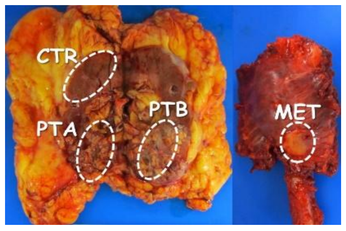


Figure 1: Metastatic CCRCC Fuhrman grade 3. CTR = normal, PTA, PTB = primary cancer nuclei in the right kidney, MET = metastasis.

GENE	DESCRIPTION	CTR	PTA	PTB	MET
DAPK3	death-associated protein kinase 3	**30.31**	4.73	1.15	2.52
PMPCA	peptidase	**28.35**	6.82	3.24	4.26
COA1	cytochrome c oxidase assembly factor 1 homolog	**22.40**	4.83	3.94	1.42
FAM208A	family with sequence similarity 208, member A	3.08	**63.97**	1.59	5.40
BCR	breakpoint cluster region	1.15	**57.43**	1.14	1.22
C2orf81	chromosome 2 open reading frame 81	2.24	**51.24**	3.19	1.84
FAM27C	family with sequence similarity 27, member C	1.75	6.03	**57.19**	3.73
GTPBP3	GTP binding protein 3	2.07	29.80	**40.06**	14.01
CASC10	cancer susceptibility candidate 10	2.57	5.55	**31.14**	4.06
ALG13	ALG13, UDP-N-acetylglucosaminyltransferase subunit	3.64	9.97	2.12	**82.95**
NUDT18	nudix	1.64	2.69	1.89	**48.40**
RAD54B	RAD54 homolog B	0.96	6.10	4.09	**40.02**

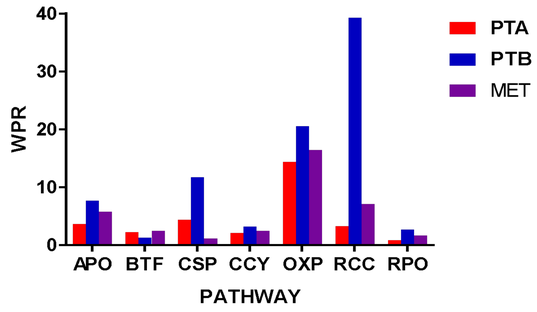


Figure 2: Weighted pathway regulation in the cancer regions (PTA, PTB and MET) with respect to the normal tissue (CTR).

Summary

This poster presents the algorithms and methods used to develop the software of the Gene Master Regulators (GMR) approach for the personalized and time-sensitive cancer gene therapy. The software reduces the calculation time by over 90% compared of doing all the calculations in Office Excel.

Pending on adequate advertising, Cancer GMR Software may become an essential IT tool for oncology hospital facilities.

References

Iacobas DA. (2016) The Genomic Fabric Perspective on the Transcriptome between Universal Quantifiers and Personalized Genomic Medicine. *Biological Theory* 11(3): 123-137.

Iacobas DA, Iacobas S (2017) Towards a personalized cancer gene therapy: a case of clear cell renal cell carcinoma. *Cancer & Oncol Res* 5(3):45:52.

Iacobas DA, Tuli N, Iacobas S, Rasamny JK, Moscatello A, Geliebter J, Tiwari RK (2018) Gene master regulators of papillary and anaplastic thyroid cancer phenotypes. *Oncotarget* 9(2), 2410-2424.

Iacobas S, Ede N, Iacobas DA (submitted). The Gene Master Regulators Approach in Cancer Gene Therapy.

Acknowledgments

This research was funded by the Undergraduate Research Funding Program From the Office of the Vice President for Research, Innovation, and Sponsored Programs.

This research work was supported in part by the US National Science Foundation (NSF) award 1736196 and by the Texas A&M University System Chancellor's Research Initiative (CRI).