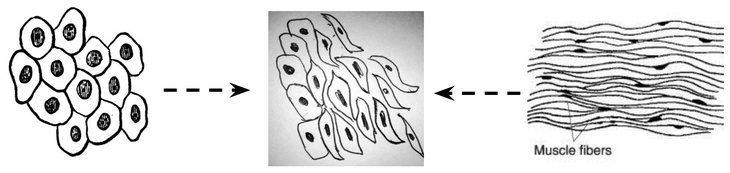
Development of gene signature for enabling early detection and treatment of sarcomatoid tumor in Kidney

Mridul Chaudhary[1], Malini Manoharan[1], Rohit Gupta[1], Amit Chaudhuri[2], Ravi Gupta[1]

[1] Research & Development Bioinformatics Group, MedGenome Labs Ltd, Bangalore, Karnataka, 560099, India [2] MedGenome Inc., Foster City, California, 94404, USA. Email: ravig@medgenome.com

INTRODUCTION

- Cancer of epithelial cells is termed carcinoma whereas cancer of mesenchymal cells is termed sarcoma
- One in six cases of advanced kidney renal clear cell carcinoma (KIRC) express sarcoma-like morphology, a condition termed as Sarcomatoid Renal Cell Carcinoma (SRCC), where a subset of epithelioid-tumor cells transform to sarcomatoid phenotype
- SRCC is resistant to therapy, has poor prognosis and high mortality rate
- Both types of cells co-exist in SRCC as shown



- SRCC is currently identified by histopathological examination
- We developed a gene signature expression based method to detect the presence of SRCC in a renal cancer sample

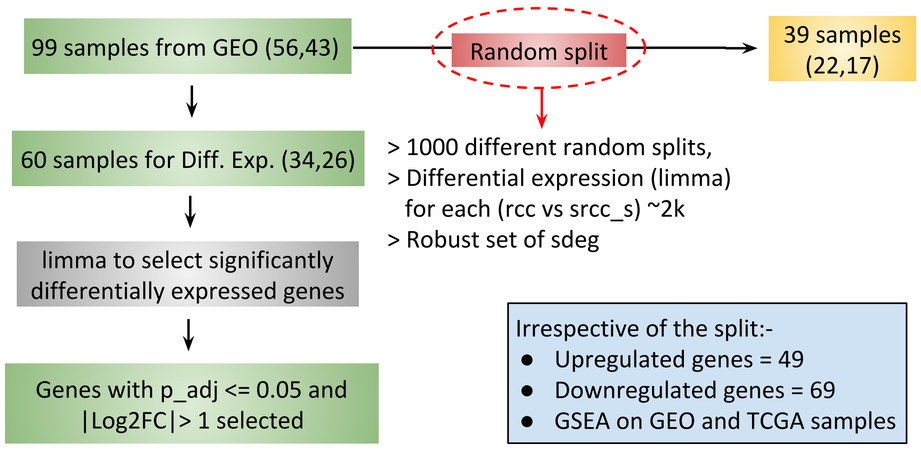
DATASETS

GSE59266: Illumina beadchip gene expression array for 142 KIRC tissue samples. 56 Non-sarcomatoid RCC samples, 43 Epithelioid (SRCC_E) and 43 sarcomatoid (SRCC_S) paired samples of SRCC

GSE59066: RNASeq data for some of the above samples (7 SRCC_S and 5 RCC) provided by the same group

TCGA-KIRC: RNASeq data for 530 KIRC cases provided by TCGA

METHODOLOGY

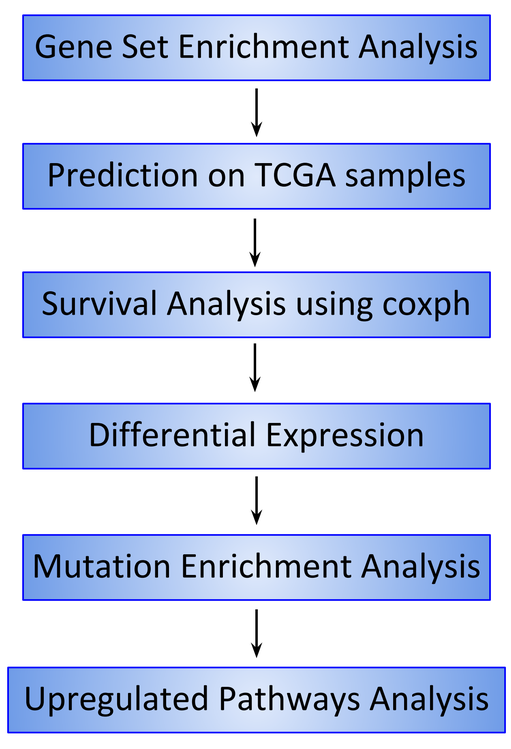


Data processing and filtering

Taking only the protein coding genes measured for all the three datasets, we had 15044 genes for each sample

Sarcomatoid Gene Signature

From a set of 99 samples (56 RCC and 43 SRCC), 60 samples were selected randomly (34 RCC and 26 SRCC) and differential gene expression analysis was performed using limma. The above process was iterated 1000 times with a different random set of samples. Based on the samples that got selected, it resulted in 400 to 1900 upregulated genes in SRCC and 400 to 2200 downregulated genes in SRCC w.r.t RCC (p.adj <=0.05 and |log2FoldChange|> 1). In order to make the sarcomatoid gene signature robust, we selected only the genes which were significantly upregulated (n=49) or downregulated (n=69) in all the 1000 iterations.



Single sample GSEA for array and RNAseq data using the two genesets found in above analysis

ssGSEA on TCGA KIRC samples and selecting 25 samples as likely sarcomatoid based on score criteria

Cox proportional hazards survival analysis was performed taking into account the age, sex, grade & stage

DESeq2 on likely sarcomatoid and likely epithelioid samples to find significantly differentially expressed genes

To find genes for which relatively higher number of samples are mutated in sarcomatoid group

Using the results from DESeq2, we find pathways upregulated in likely sarcomatoid group

RESULTS

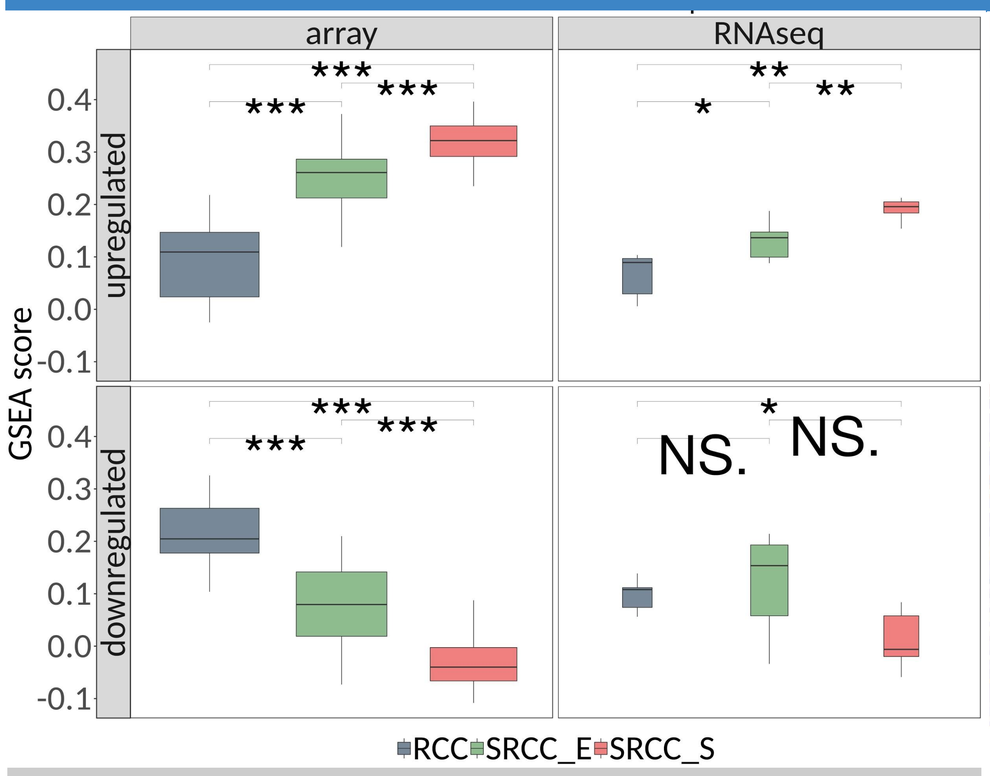


Figure 1: **Gene Set Enrichment Analysis Score:** GSEA scores for the upregulated and downregulated gene signatures were calculated for array samples as well as RNASeq samples in GEO datasets. Although, genesets were derived using only RCC and SRCC_S samples, SRCC_E (epithelial cells of a SRCC tumor) samples appear exactly in the middle in array samples. Signatures remain valid for RNASeq samples as well

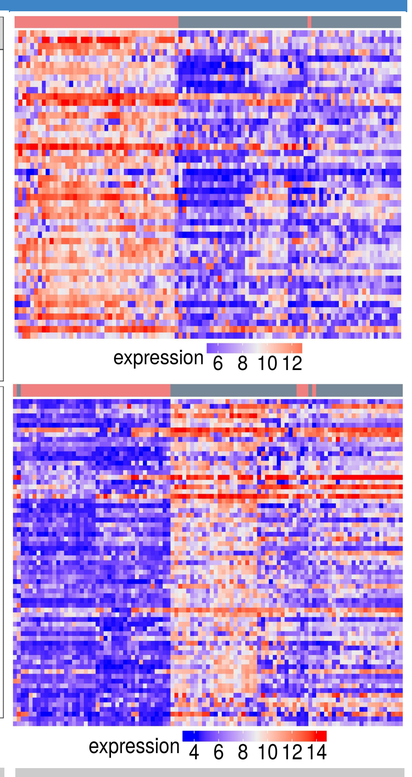


Figure 2: **Heatmap for RCC & SRCC_S samples:** Heatmaps showing expression of signature genesets in RCC and SRCC_S samples for array data

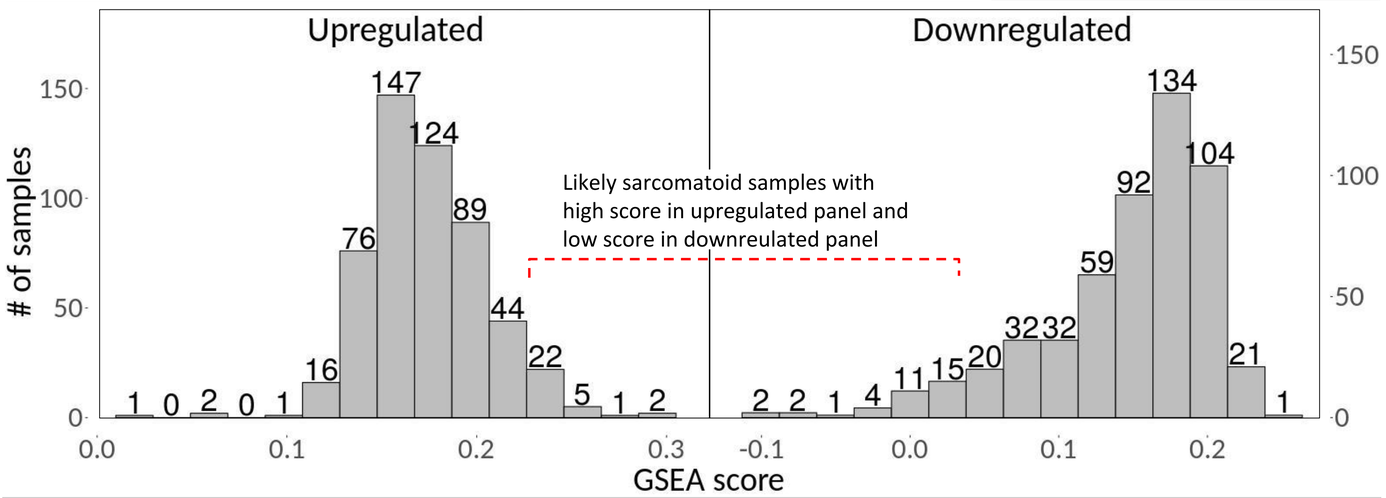


Figure 3: **Distribution of GSEA scores for TCGA-KIRC:** GSEA scores for upregulated and downregulated gene signatures were calculated for TCGA KIRC samples. From the distribution shown above, intersection of samples with high up-signature score and low down-signature score were termed as likely sarcomatoid

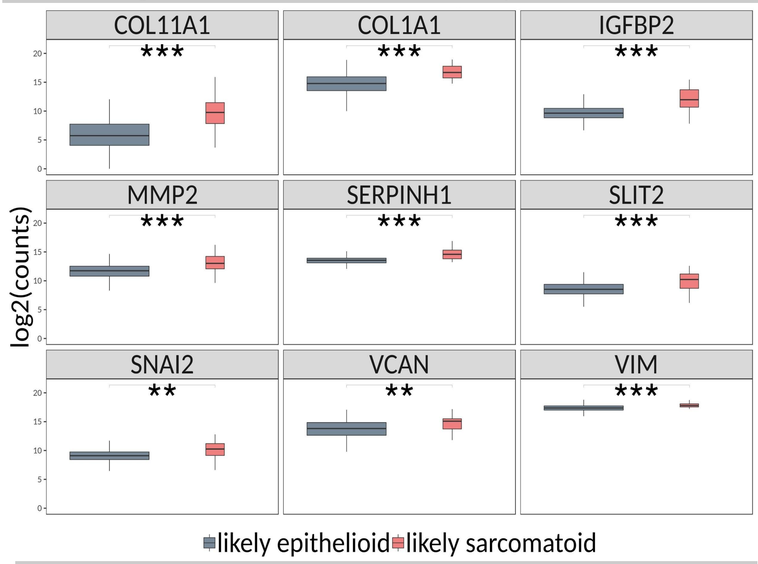


Figure 4: **Expression of EMT associated genes:** EMT genes are overexpressed in likely sarcomatoid samples compared to likely epithelioid samples

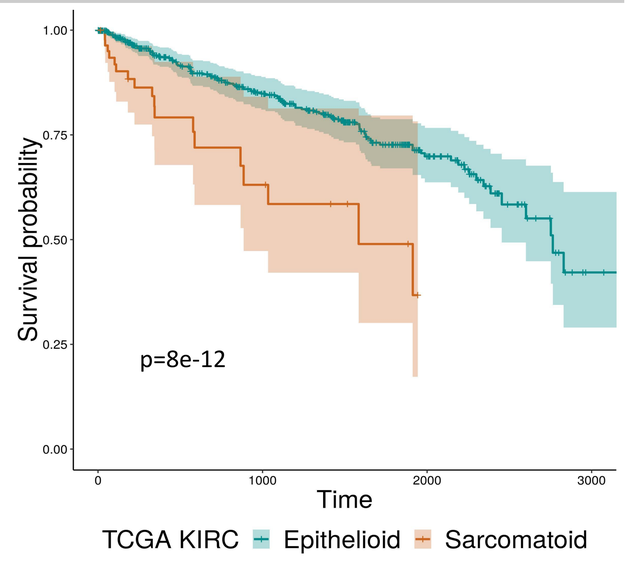


Figure 5: **Coxph Survival Analysis:** Poor survival in likely sarcomatoid group compared to epithelioid. Coxph takes into account age, sex, stage and grade of samples

Table 1: **Mutation Enrichment Analysis (% of samples mutated in a group):** Mutation Enrichment for some genes in likely Sarcomatoid

Mutated Gene	# samples (sarcomatoid)	# samples (epithelioid)	Enrichment Factor	Chi.Sq p
TP53	5	11	8	0.00002716
NF2	3	3	18	0.00005243
RELN	3	5	11	0.000829
APC	3	6	9	0.002096
DNHD1	3	6	9	0.002096
CUL9	3	9	6	0.01362
AHNAK2	4	26	3	0.1001
AHNAK	3	17	3	0.1338

Table 2: **Pathway Analysis** (reactome.org)

Upregulated pathways in likely sarcomatoid	FDR
Extracellular matrix organization	1.01E-06
Assembly of collagen fibrils	3.99E-05
Collagen degradation	5.01E-05
Degradation of the extracellular matrix	5.03E-05
Collagen formation	7.62E-05
Collagen chain trimerization	8.57E-05
Collagen biosynthesis and modifying enzymes	4.58E-04
ECM proteoglycans	8.43E-04
Polo-like kinase mediated events	0.002
Amplification of signal from the kinetochores	0.003

CONCLUSIONS

- Gene signatures (both upregulated and downregulated) can classify renal cancer samples as a sarcomatoid carcinoma based on expression values
- Likely sarcomatoid samples in TCGA-KIRC showed poor survival compared to likely epithelioid after normalizing for tumor stage and grade and age of the patient
- EMT genes responsible for transition from Epithelial type cell type to Mesenchymal type cell type were also upregulated in likely sarcomatoid samples of TCGA-KIRC
- Sarcomatoid samples were enriched in TP53 mutations compared to epithelioid
- Genes differentially upregulated in likely sarcomatoid samples belonged to extracellular matrix reorganization suggesting change in cell morphology