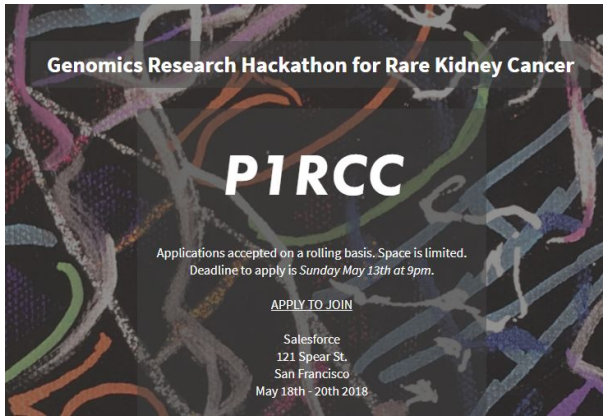

Candidate p1RCC biomarkers and environmental factors influencing their expression levels

SAED SAYAD

PEYMAN MIRTAHERI

SALAR SAYAD

USMAN QAZI

BIOMARKERS.AI TEAM

Problem Definition

- **Papillary renal-cell carcinoma (pRCC)**, comprises 15 - 20% of all kidney cancers. It occurs in the cells lining the small tubules in the kidney that filter waste from the blood and make urine.
- pRCC has two subtypes (**p1RCC** & **p2RCC**) based on histologic, cytogenetic, and gene expression differences.
- **Little is known** about the genetic basis of sporadic papillary renal-cell carcinoma, and no effective forms of therapy for advanced disease exist.
- The goal of this analysis is to find **candidate diagnostic biomarkers** and **treatment regimen** for p1RCC.

Microarray gene expression data was downloaded into for p1RCC and normal kidney tissue

- Data source: NCBI GEO database. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7023
- Platform: Affymetrix GeneChip Human Genome U133 Plus 2.0 Array
- 46 cases
  - 19 p1RCC
  - 12 Normal Tissue
  - (also 16 p2RCC samples that were not used)
- 54,675 probes
- Target: Classify Normal Tissue (0) vs. p1RCC (1)

For meaningful predictive modeling, we selected features capturing the most variation between p1RCC and normal tissue

- Predictive modeling using **54,675 probes** is not practical and even counterproductive (due to large number of collinear probes).
- Subsets of genes capturing the most variation can be located using **Mahalanobis distance (in standard deviations)**, **t-test** and **ANOVA**.
- In each of 10 cross-validation passes of **80% data was randomly assigned to a trainset** and **20% to a testset** (*before* any statistical analysis or predictive modeling).
- For each trainset, **top 100 probes/genes** showing the most variation in expression with the test set were used to find the **most predictive candidate biomarkers**.

Top predictive genes could offer insight on hypothetical biomarker candidates (Comparing p1RCC and Normal Tissue)

NPA	NPN	MAS5
KNG1	KNG1	TMPRSS2
UMOD	UMOD	KNG1
KCNJ1	PTGER3	RALYL
HS6ST2	TFAP2B	FHL1
FHL1	KCNJ1	MUC15
PTGER3	HS6ST2	DMRT2
DMRT2	DMRT2	LRRK2
HSD11B2	KCNJ10	IRX2
EMCN	EMCN	PTGER3
LRRK2	FHL1	UMOD

Can annotations on our top predictive genes offer insights on Bill's history?

Gene	Name	Pathway/Notes
KNG1	kininogen 1	Complement and coagulation cascades
PTGER3	prostaglandin E receptor 3	Calcium signaling pathway
DMRT2	doublesex and mab-3 related transcription factor 2	Sequence-specific DNA binding
UMOD	uromodulin	Most abundant protein in normal urine
FHL1	Four And A Half LIM Domains 1	Tumor suppressor gene on X chromosome. Relates to JAK-STAT pathway

Result of binary classification using the expression of top hundred genes (Normal Tissue vs. p1RCC)

- Binary Classification
  - **Linear Discriminant Analysis**
  - **XGBoost**
  - Both provided **100% accuracy** with 10-fold cross-validation

- **KNG1** provides widest discrimination gap between normal and p1RCC tissue with the highest predictability power.
- Other genes (like PTGER3, UMOD, DMRT2) also show 100% accuracy.

Could decreased KNG1 expression offer an insight into Bill's medical history?

1. Got warfarin/coumadin for diagnosis of deep vein thrombosis.

...

2. Symptoms returned. Went back & found:
   A. 7 cm mass left kidney
   B. Cerebral meningioma
   C. Spots in lung

….

Meningioma hasn't grown.

Chest spots haven't grown.

Gene	Name	Pathway/Notes
KNG1	kininogen 1	Complement and coagulation cascades

KNG1 uses alternative splicing to generate two different proteins: High MWt kininogen (HMWK) and MWt kininogen (LMWK). **HMWK** is **essential for blood coagulation** and assembly of the kallikrein-kinin system.

Chromosome 3 (3q26) Translocations/Deletion as Risk Factors for RCC

Uromodulin should be investigated as a candidate early-diagnosis biomarker for p1RCC

- **UMOD** can distinguish Normal Tissue from p1RCC with 100% accuracy.
- **Uromodulin** (encoded by UMOD; also known as Tamm-Horsfall protein) is the most abundant protein in mammalian urine under physiological conditions.

To explore: Is UMOD also a good **urine-based biomarker** for p1RCC?

Effect of chemicals on expression of top candidate biomarker genes: Exposure to benzopyrene enhances FHL1 expression, increases mutagenesis in PTGER3

Substance	Up	Dn	DMRT2	FHL1	KNG1	PTGER3	UMOD
6-propyl-2-thiouracil	2	2	Dn	Up		Up	Dn
bleomycin A2	1	0		Up			
methoxyacetic acid	1	0		Up			
dexamethasone	3	0		Up	Mi|Up|Mi	Up	
doxorubicin	2	1		Up	Up|Up		Dn
arsenous acid	1	0		Up			
thioacetamide	2	0		Up		Up	
mitoxantrone	1	0		Up			
phenylephrine	1	0		Up			
1-naphthyl isothiocya...	1	0		Up	Mi		
4,4'-diaminodiphenyl...	1	0		Up			
cyclophosphamide	1	0		Up			
benzo[a]pyrene	1	0		Up	Dn|Up	Imu	
carbofuran	1	0		Up			
methylmercury chlori...	1	0		Up			
N-nitrosodimethylami...	1	0		Up			
quercetin	1	1		Up	Dn		
metaproterenol	1	0		Up			
methapyrilene	1	0		Up			
diarsenic trioxide	1	0		Up			
caffeine	1	0		Up			
Monobutylphthalate	1	0		Up			

Thank You!

saed.sayad@utoronto.ca