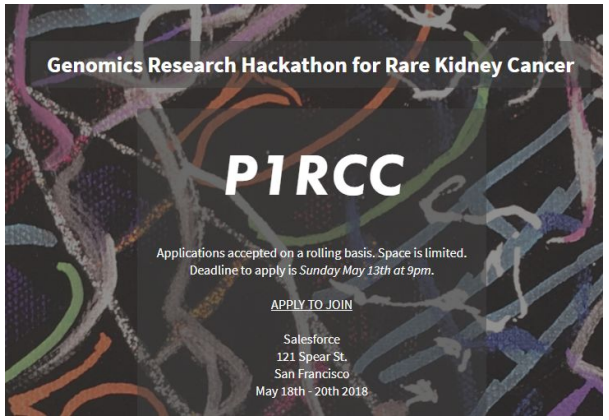


Candidate p1RCC biomarkers and environmental factors influencing their expression levels

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.

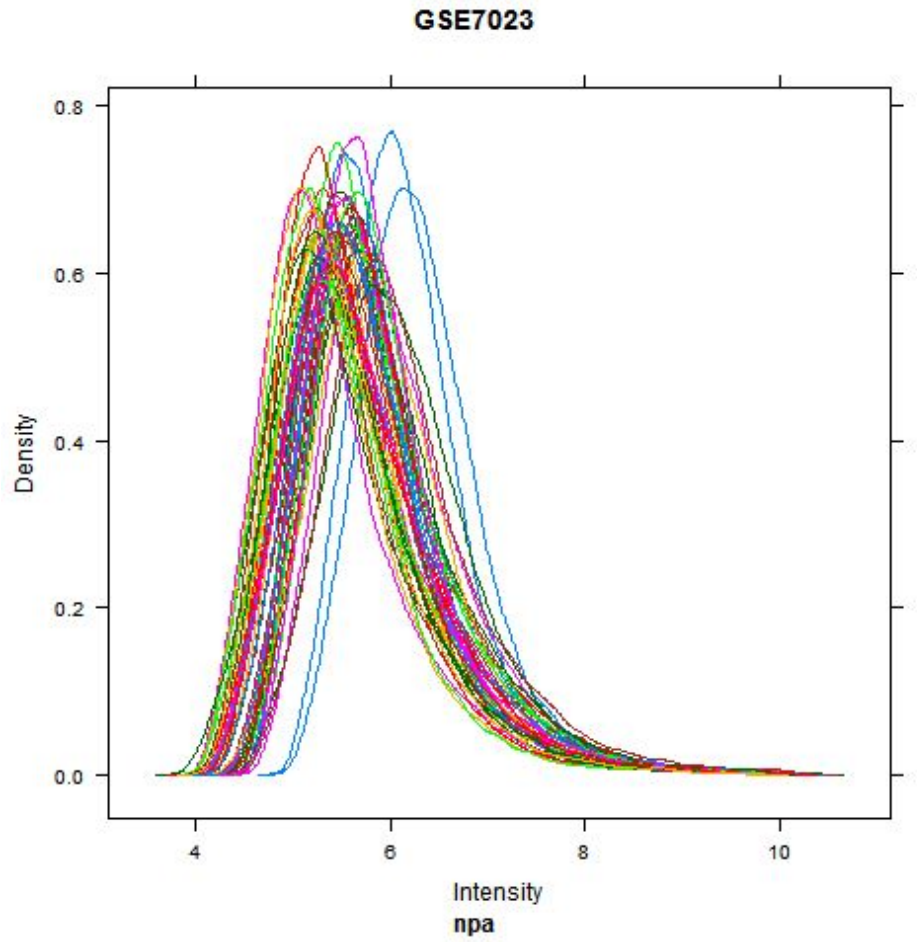
Microarray gene expression data was downloaded for p1RCC and normal kidney tissue

- 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)

Initial view using no normalization or background correction shows data consistency

Raw data (CEL files) from 47 microarrays were processed with:

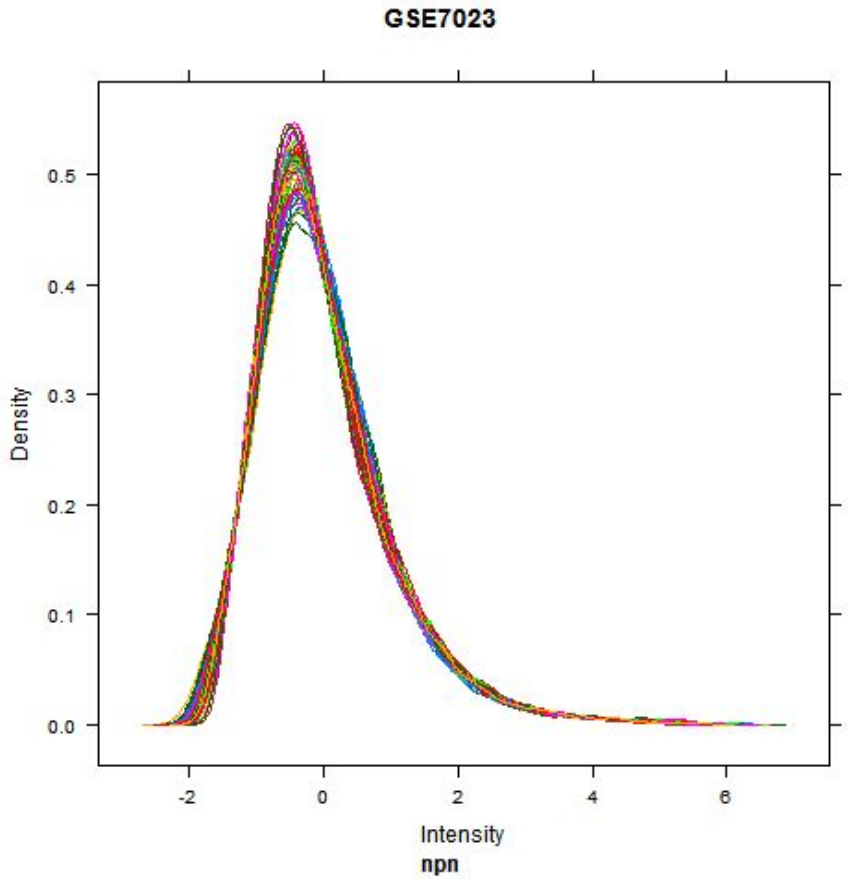
- Only using perfect match (PM) probes with median summarization



Using normalization within each array but not between arrays

Raw data (CEL files) from all 47 microarrays were processed with:

- No background correction
- Only perfect match (PM)
- Median summarization
- **Within-array normalization**
- No between-array normalization

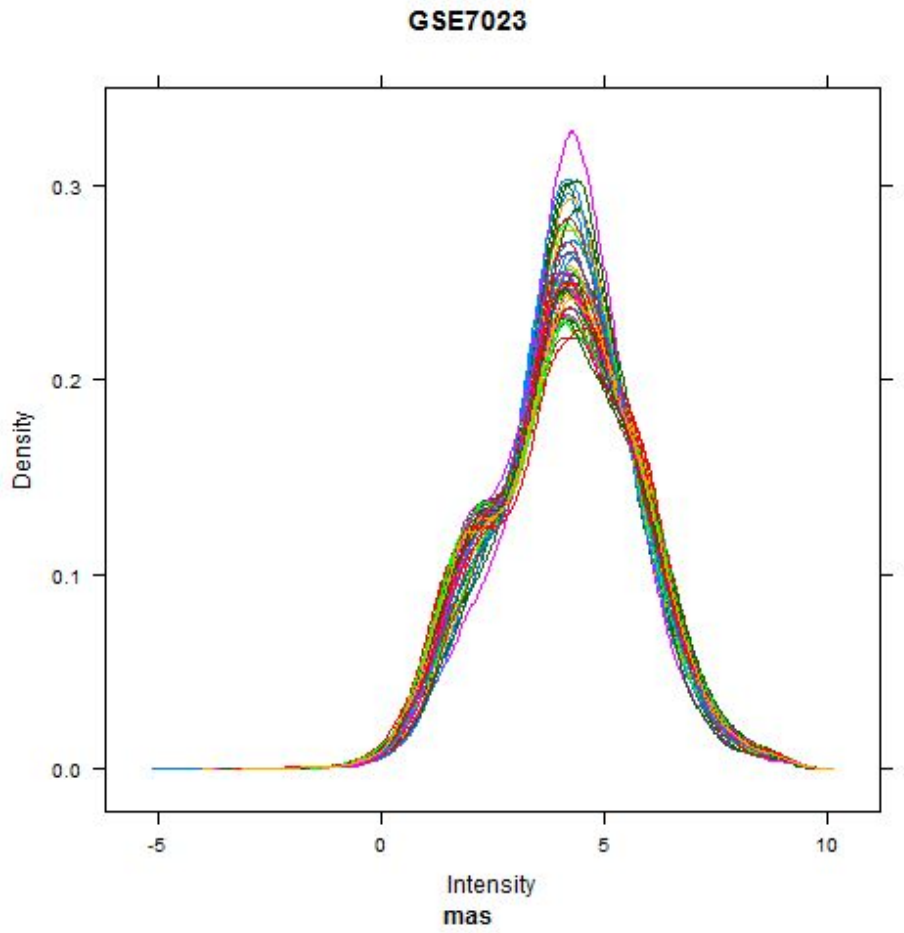


Adding between-array normalization (MASS method) was not as useful

Raw data (CEL files) from 47 microarrays were processed and **normalized** using the **MAS5** method.

Using MAS5 or any between-array normalization methods can introduce artifacts. Please see the supplementary document:

A Case Against Microarray Data Normalization.pdf



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**.

Reliability Test: Ratio of overlap should not be affected by different pre-processing methods

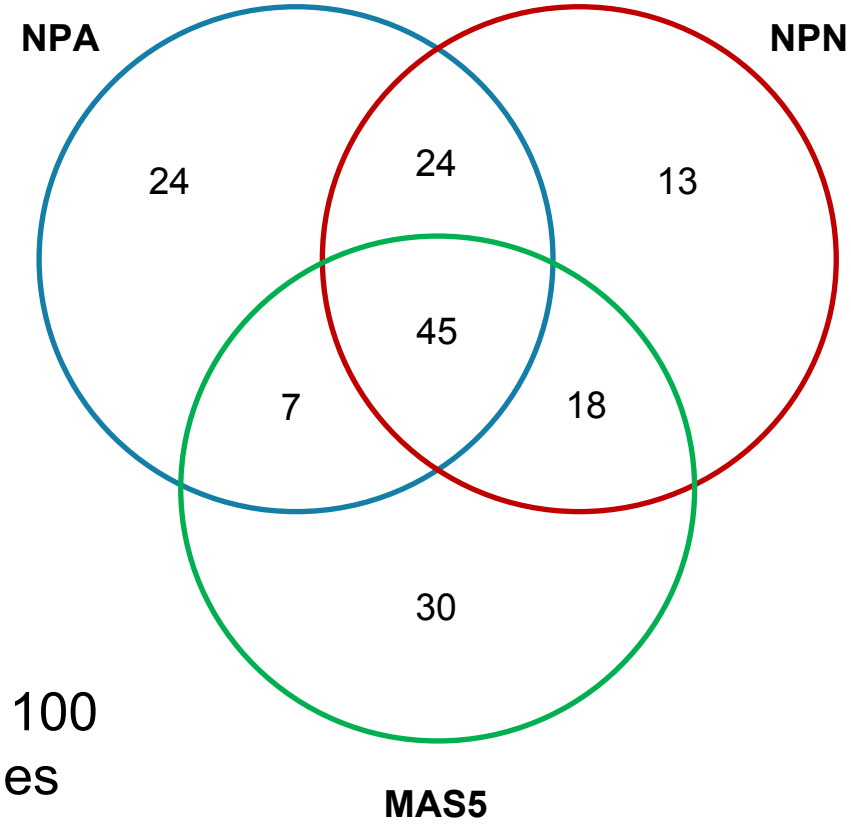
There is a relatively significant overlap between Top 100 genes from three different preprocessing methods.

	MAS5	NPA	NPN
MAS5	100	52	63
NPA	52	100	69
NPN	63	69	100

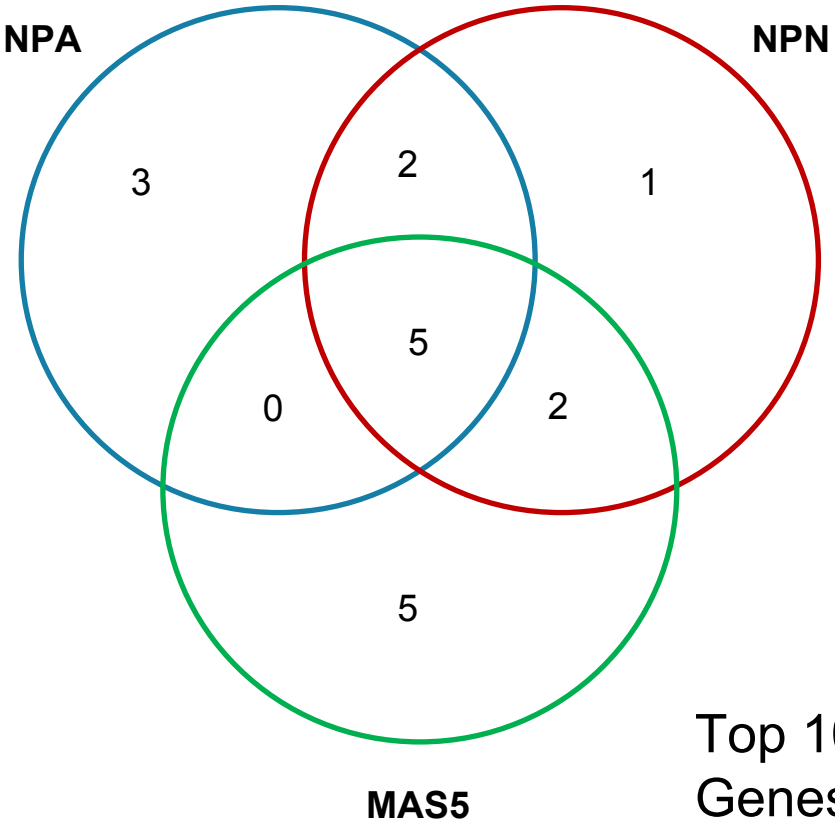
When we use Top 10 genes the ratio of overlap does not change between NPA-NPN and MAS5-NPA but it drops between MAS5-NPN.

	MAS5	NPA	NPN
MAS5	10	5	3
NPA	5	10	7
NPN	3	7	10

Reliability Test: Overlap ratio should not be affected by different pre-processing methods



Top 100 Genes



Top 10 Genes

Top predictive genes and selected pathways 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. Points to JAK-STAT pathway

Modeling
(Normal Tissue & 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?

Tumorigenicity and FHL1

Effect of chemicals on expression of top 10 genes: exposure to benzopyrene and several other agents enhances FHL1 expression

Substance	Up	Dn	FHL1	LRRK2	APBB1IP	TPI
valproic acid	44	11	Ax|Up	Up|Mi	Ax|Mi|Up	
benzo[a]pyrene	25	16	Up	Dn	Dn|Dn|Up	Dn|
trichostatin A	24	7		Up		
aflatoxin B1	22	10	Dm	Dm	Ax	Ax|
cyclosporin A	21	21			Dn	
2,3,7,8-tetrachlorodibenzodioxi...	19	12	Mi|Ax|Dn|Dn	Ax	Dn	Mi|A
all-trans-retinoic acid	18	9	Dn			
phenylmercury acetate	18	11	Dn|Mi			U
17beta-estradiol	17	13	Mi|Dn|Up|Mi|...			Mi|
bisphenol A	16	31	Im|Dn|Dn		Dn|Ax	I
nickel atom	16	15	Dn|Dn	Dn	Up	I
formaldehyde	16	5				I
methyl methanesulfonate	16	14				I
quercetin	15	8	Up			
6-propyl-2-thiouracil	15	11	Up	Dn	Up	
ethyl methanesulfonate	14	7		Up		
butanal	13	15		Dn	Dn	
silicon dioxide	13	13				
nickel sulfate	13	15				
tetrachloromethane	13	5				
N-nitrosodimethylamine	12	1	Up		Up	

Thank You!

saed.sayad@utoronto.ca