Alternatives after Patients are past Standard of Care and have run out of Clinical Trials: p1RCC Hackathon Report

RareKidneyCancer.org

W.G. Paseman
bill@RareKidneyCancer.org

20240415

Agenda

- ***Medical Research Issues from the Rare Disease Patient POV**
  - Reproducibility, Reporting, Data Access, Novelty, Urgency, Inclusion
- **Hackathons – Patient led research modeled after Kaggle**
  - One Patient, Many Competing Teams, Unrestricted Rich Data Set
- **Bill Paseman Papillary Kidney Cancer Hackathon**
  - Process, Report and Conclusions
- **Summary**

Patient Journey

- Standard of Care
  - Diagnosis Condition
  - Look up in a Big Book
  - Follow Instructions
  - If doctor don’t follow the book, doctor can lose their license
    - Can’t “Pour purple dust and wave feathers over a broken bone”
- Clinical Trials
  - If it passes the IRB, and is double blind, doctor
  - can “Pour purple dust and wave feathers over a broken bone”
- ?? - Wait; Hope; Spend a lot of time on the Internet

Issues: Reproducibility/Reporting/Data

- Now
  - Poor Reproducibility (Bayer) - 25% replication
  - Poor Reporting (alltrials.net)
    - Current problem: e.g. 20240123 Dana Farber retracts 6 studies
  - Limited Data Access – “Data Silos”

Issues: Novelty/Agility

Cancer Research is done with **big teams**

Field	% of Papers with <4 Authors
Economics	85%
Astronomy + Astrophysics	41%
Genetics	21%
Cancer	12%

This surprised me because...

- NY Times: Can Science be too Big?
  - Big Teams confirm findings
  - Small Teams generate new ideas
  - Disruption is inversely proportional to author count
- Small teams are more agile
  - Small enables closely space milestones (speed)
  - Closely spaced milestones keep you on track

https://www.natureindex.com/news-blog/paper-authorship-goes-hyper
https://www.nytimes.com/2019/02/13/science/science-research-psychology.html

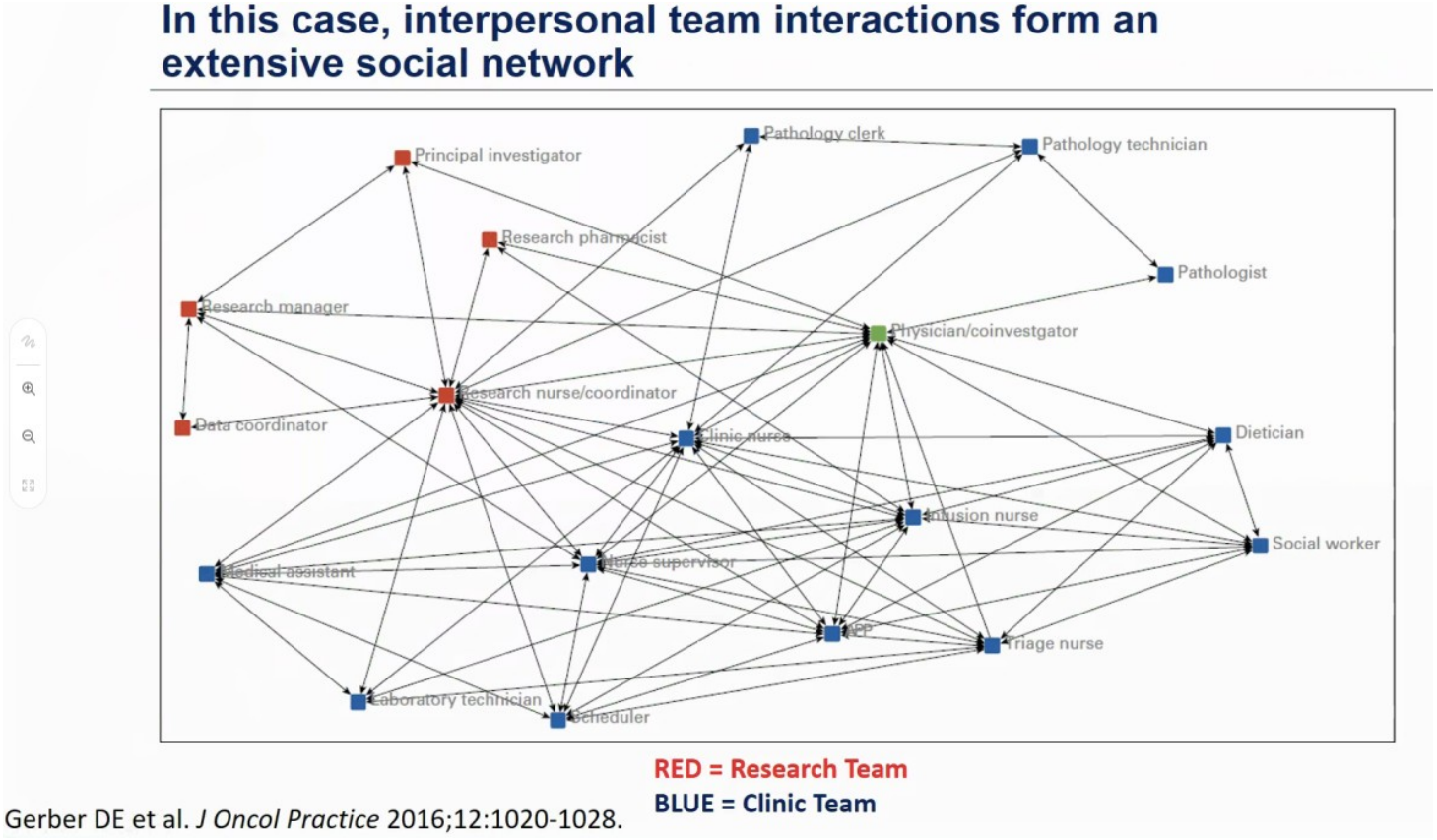
Issues: Urgency

Researchers and Patients view time differently

- Different Units of Measure
  - Researchers time measurement: yearly conferences and tenure tracks
  - Patients time measurement: PFS, OS, months between treatments; months until death
- Time pressure affects Motivation (for all you patients out there)
  - Researchers are motivated to win
  - (Some) Patients are desperate to win
    - 49ers Quarterback Steve Young: Everyone wants to win. Steve was desperate to win
  - Question to patients in the audience: Can you do something if your life depended on it?
    - If “yes” and you are not doing it now, it is a motivation issue, not a training issue
  - Hippocrates
    - It's far more important to know what person the disease has than what disease the person has
    - If you are not your own doctor, you are a fool.

Issues: Inclusion

In a Cancer Researchers' social network...

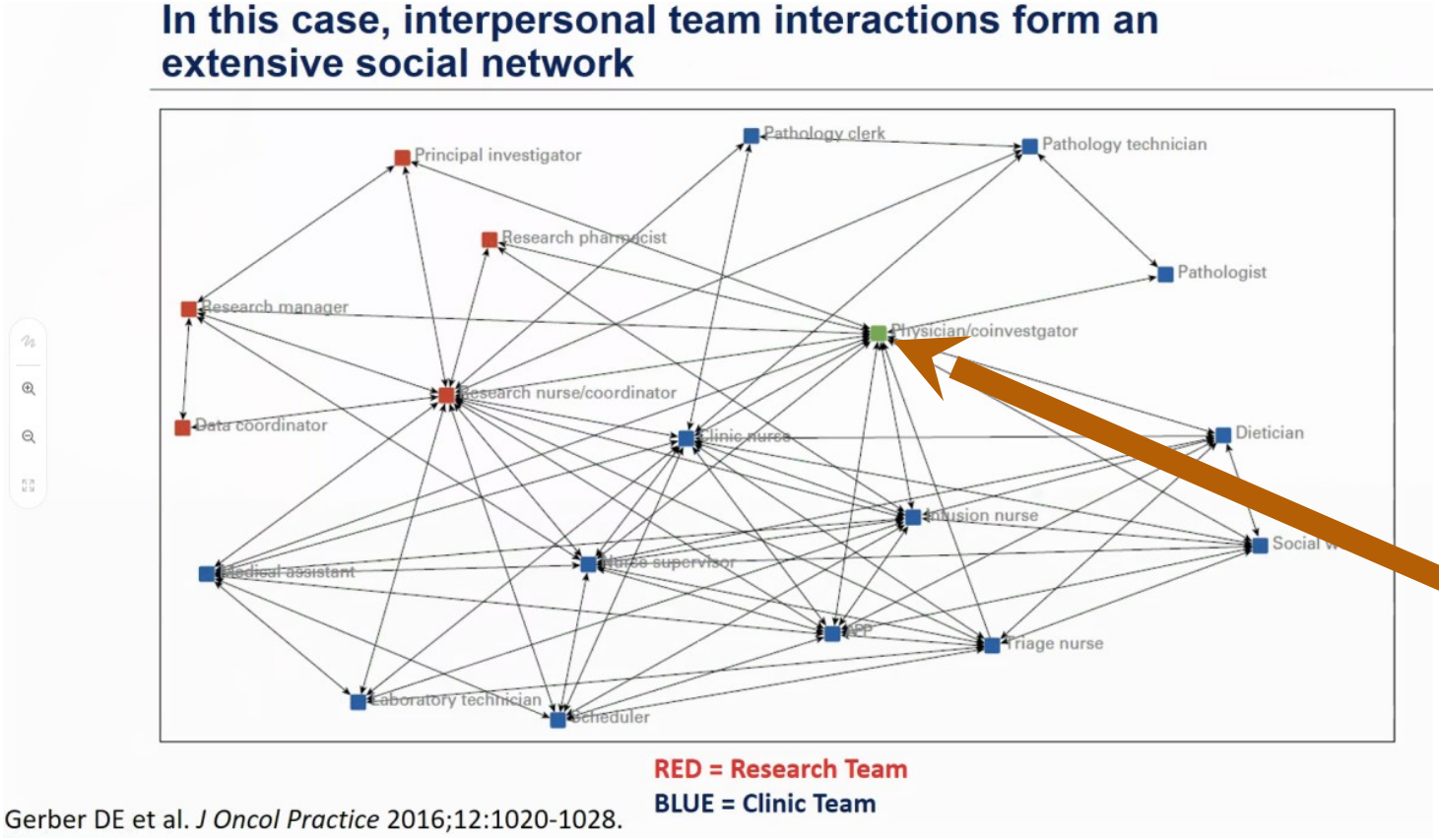


Gerber DE et al. *J Oncol Practice* 2016;12:1020-1028.

Principal investigator
Research Coordinator
Pathologist
Physician
...

Issues: Inclusion

… patients don't exist



Gerber DE et al. *J Oncol Practice* 2016;12:1020-1028.

- Peanuts

Principal investigator
Research Coordinator
Pathologist
Physician
...

PATIENT!!

Patient Advocate (e.g. Smart Patients)

Issues: Inclusion

"A "Seat at the Table" when its not your table

- Proposal: Renal Cancer incidence in Native Americans
- My Comments
  - Record if its a rare cancer
    - "Rate of Renal Cell Carcinoma Subtypes in Different Races"
  - Record Occupation
    - My "lived Experience":
      - Lots of Native Americans in Texas, Oklahoma and the Dakotas
      - Lots of Petrochem jobs in Texas, Oklahoma and the Dakotas
      - Lots of Kidney Cancer in Texas, Oklahoma and the Dakotas
- All points rejected, none recorded.
- T-test comment not acted on.

Advice given to a 14 year old:
If you don’t like other kid’s parties
Throw your own

Agenda

- **Medical Research Issues from the Rare Disease Patient POV**
  - Reproducibility, Reporting, Data Access, Novelty, Urgency, Inclusion
- ***Hackathons – Patient led research modeled after Kaggle**
  - One Patient, Many Competing Teams, Unrestricted Rich Data Set
- **Bill Paseman Papillary Kidney Cancer Hackathon**
  - Process, Report and Conclusions
- **Summary**

Kaggle

The competition host prepares the data and a description of the problem; the host may choose whether it's going to be rewarded with money or be unpaid. Participants experiment with different techniques and compete against each other to produce the best models. Work is shared publicly through Kaggle Kernels to achieve a better benchmark and to inspire new ideas.

For most competitions, submissions are scored immediately (based on their predictive accuracy relative to a hidden solution file) and summarized on a live leaderboard.

Tianqi Chen from the University of Washington also used Kaggle to show the power of XGBoost, which has since replaced Random Forest as one of the main methods used to win Kaggle competitions.

https://en.wikipedia.org/wiki/Kaggle

Hackathons

Baseline Approach:

Big teams doing the same things slowly and often secretly without any patient involvement.

Goal: Improve knowledge, FDA approval

Hackathon Approach

Small teams doing different things quickly and openly with patient involvement.

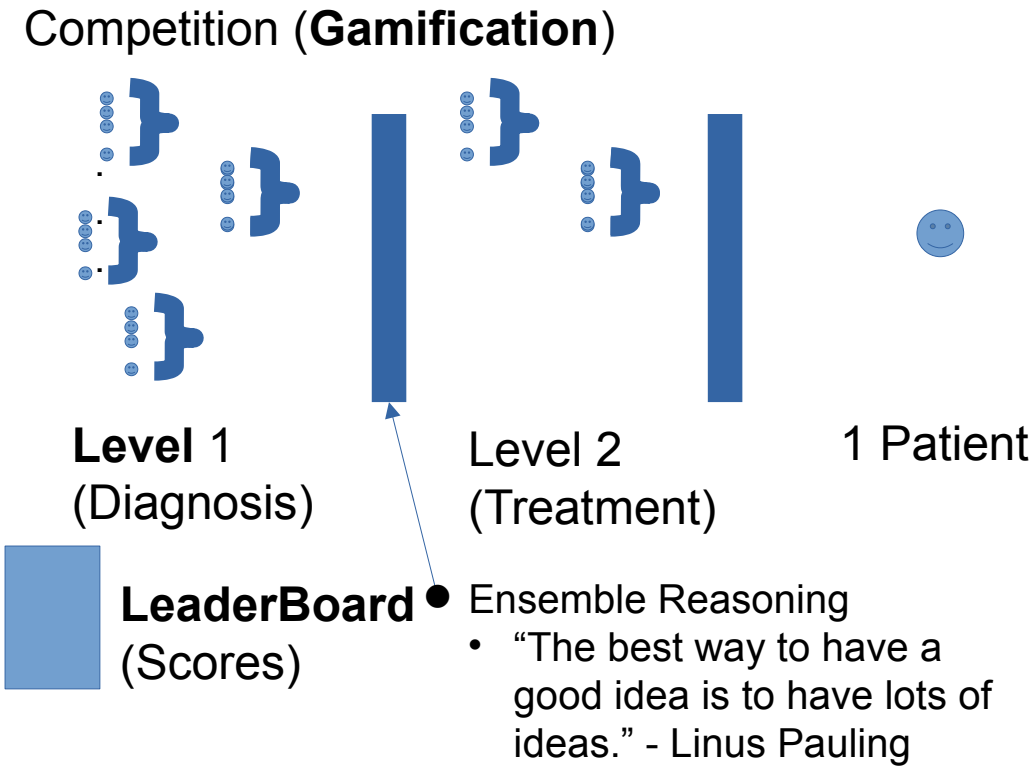
Goal: Improve Health

- If you don’t like someone else’s party, throw your own

p1RCC Hackathon Process

- Organizer: Open Participation
  - Open investigation to Beginners, not just Experts.
- Patient: Opens their own Deep Research Data; TCGA
  - DNA, EHR, etc. is given to everyone
  - RNA-seq is the holdout set
- Researcher: Common Published Deliverable(s)
  - “Genes of interest” is the primary common deliverable which enables objective comparison.
- Everyone: Objective Evaluation
  - Everyone can participate, but scoring mechanisms tell us who to listen to.
  - These scoring mechanism are -objective-, they don’t depend on interpretation by particular persons,
  - and -automated-, which lets computer programs to help scale up the process.

P1RCC Hackathon Process



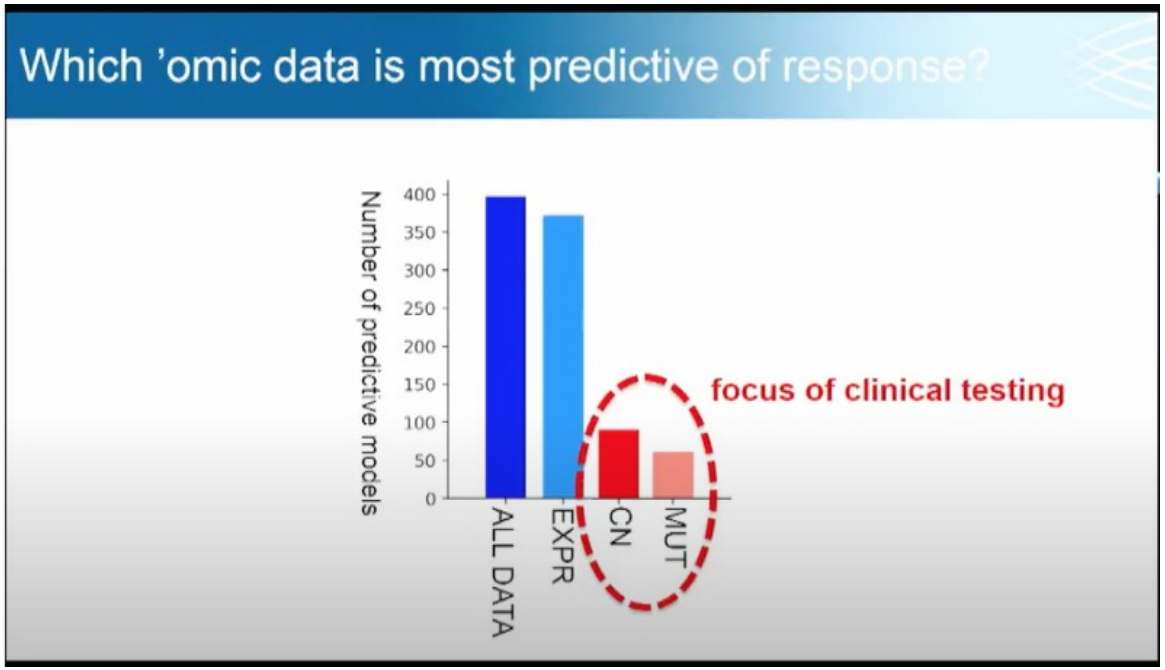
Agenda

- **Medical Research Issues from the Rare Disease Patient POV**
  - Reproducibility, Reporting, Data Access, Novelty, Urgency, Inclusion
- **Hackathons – Patient led research modeled after Kaggle**
  - One Patient, Many Competing Teams, Unrestricted Rich Data Set
- ***Bill Paseman Papillary Kidney Cancer Hackathon**
  - Process, Report and Conclusions
- **Summary**

2018 p1RCC Hackathon Teams

80 People (some Remote) formed 17 Teams (50 pictured)

Clinical vs. Research Data



James Watson: targeting metabolism is a more promising avenue in current cancer research than gene-centered approaches. 20160515 NYT

- Genetics- brother has thyroid cancer
- Genomics - TCGA Thyroid cancer clusters close to p1RCC
- Metabolomics - High Uric Acid - "Thyroid hormones influence kidney function and thereby might alter serum urate levels, a major risk factor for gouty arthritis."
- Co-morbidities - Bradycardia (Slow Heart Rate): "hypothyroidism results in an insufficient amount of thyroid hormone which leads to a slower heart rate
- Dental Records

2018 p1RCC DNA Hackathon Process

TCGA Data

Bill Data (DNA)

Ensemble of 17 weak classifiers

119 Recommended Genes

cancer-genome-workbench								
causalnucleotidenetwork								
RecausalNucleotideNetworks								
Aizheng	AKR1B10	BASP1P1	CLEC2B	CYP4F11	LINC00621	PLEKHO1	PLEKHO2	...
BioMarkers.ai	DMRT2	FHL1	KNG1	PTGER3	UMOD			
DamTheRiver	AC139425.3	ACSM2A	ANO9	AQP12B	GRIN3B	HEXB	HIVEP3	...
GEViz	NRF2-ARE							
HelloKidney	ITGAM	TNFSF4						
KidneyBean	TUBB8							
studentec	AMPD2	DPP6	FLG2	FTMT	ST6GALNAC5			
trimericOGs	AGBL4	ARIDA1	CUL-2	HPSE2	LAMC-1	SK3	TRABD2B	...
DeeperDrugs	**BARD1**	APOB	CDK9	TTRAP				
GNOME	**BARD1**	**PDE4DIP**	AHNAK	ANAPC1	BCLAF1	DNAJ27	PABPC1	...
HelloKidney2		**PDE4DIP**	FOLH1	GDNF	MTHFR	PFKP	PSMA	...
codeomics						**MTOR**	**PIK3CA**	...
HSIEH	**SETD2**	**NF2**	**BAP1**	**KDM6A**	**PBRM1**	**MTOR**	**PIK3CA**	...
ExpressForce	**SETD2**	**NF2**	**BAP1**	**KDM6A**	**PBRM1**	**FGFR1**	ARID1A	...
HIF1AIsNotAnOncogene						**FGFR1**	CDK4	...

10 Up Weighted Classifications (Genes)

- BARD1
- PDE4DP
- SETD2
- NF2
- BAP1
- KDM6A
- PBRM1
- MTOR
- PIK3CA
- FGFR1

SCIENTIFIC REPORTS

OPEN

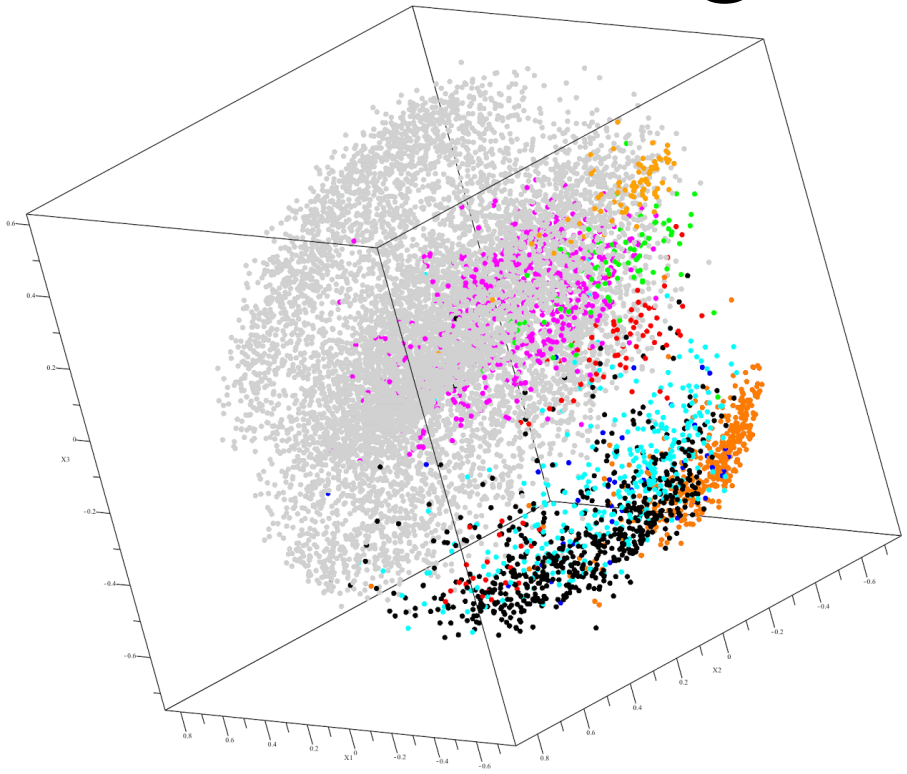
Received: 25 October 2018
Accepted: 28 January 2019
Published online: 27 February 2019

Linking Binary Gene Relationships to Drivers of Renal Cell Carcinoma Reveals Convergent Function in Alternate Tumor Progression Paths

William L. Poehlman[1], James J. Hsieh[2] & F. Alex Feltus[1]

Renal cell carcinoma (RCC) subtypes are characterized by distinct molecular profiles. Using RNA expression profiles from 1,009 RCC samples, we constructed a condition-annotated gene coexpression network (GCN). The RCC GCN contains binary gene coexpression relationships (edges) specific to conditions including RCC subtype and tumor stage. As an application of this resource, we discovered RCC GCN edges and modules that were associated with genetic lesions in known RCC driver genes, including VHL, a common initiating clear cell RCC (ccRCC) genetic lesion, and PBRM1 and BAP1 which are early genetic lesions in the Braided Cancer River Model (BCRM). Since ccRCC tumors with PBRM1 mutations respond to targeted therapy differently than tumors with BAP1 mutations, we focused on ccRCC-specific edges associated with tumors that exhibit alternate mutation profiles: VHL-PBRM1 or VHL-BAP1. We found specific blends molecular functions associated with these two mutation paths. Despite these mutation-associated edges having unique genes, they were enriched for the same immunological functions suggesting a convergent functional role for alternate gene sets consistent with the BCRM. The condition annotated RCC GCN described herein is a novel data mining resource for the assignment of polygenic biomarkers and their relationships to RCC tumors with specific molecular and mutational profiles.

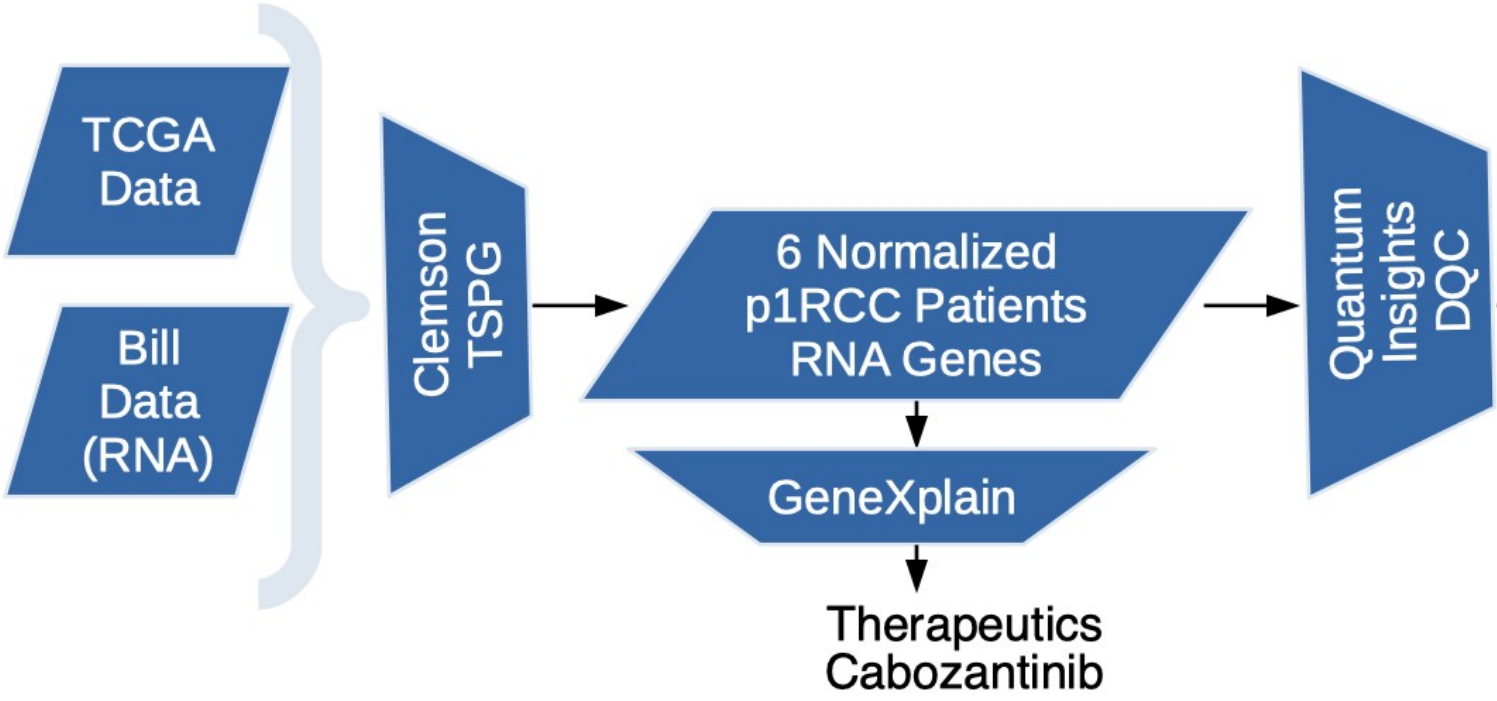
Parents,Siblings,Cohort Genetics



KIRP	Cyan
KIRC	Black
KICH	Red
LIHC	Coral
THCA	Magenta
CHOL	Blue
UVM	Orange
ACC	Lime
Others	Light Gray

2018 QuantumInsights.io DQC

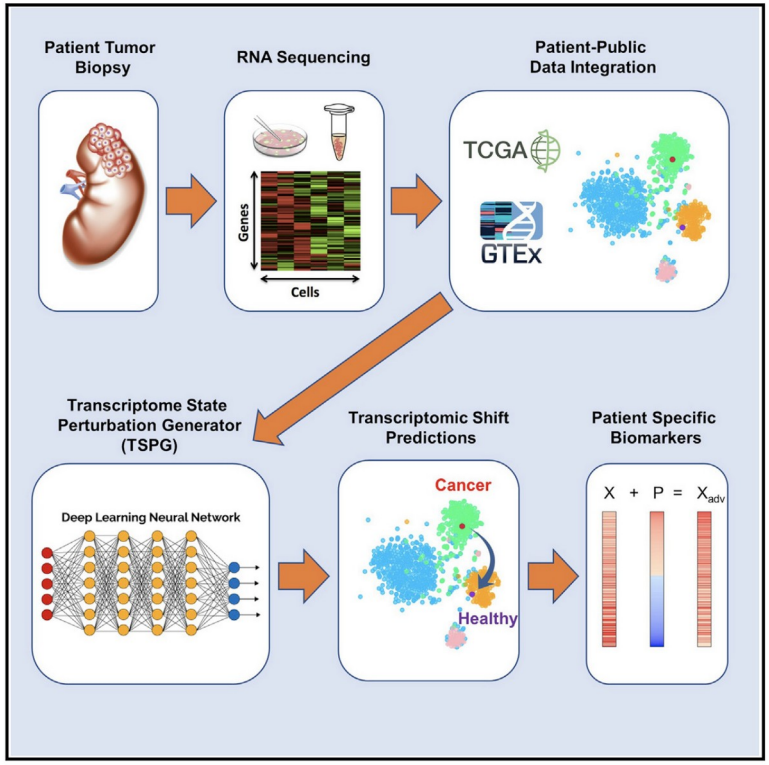
2020 p1RCC RNA Hackathon Process



- Three Teams
  - Clemson
  - GeneXplain
  - QuantumInsights.io

Cellular State Transformations Using Deep Learning for Precision Medicine Applications

Graphical Abstract



Highlights

- We present the Transcriptome State Perturbation Generator (TSPG) application

Authors

Colin Targonski, M. Reed Bender, Benjamin T. Shealy, Benafsh Husain, Bill Paseman, Melissa C. Smith, F. Alex Feltus

Correspondence

ffeltus@clemson.edu

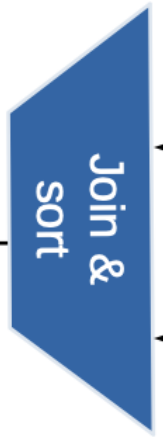
In Brief

Cells express genes in unique combinations that enable distinct functions. Using machine learning, we present an algorithm that takes a source gene expression snapshot and learns how to change it to mimic a target snapshot. We applied the Transcriptome State Perturbation Generator algorithm to learn which genes have changed in a single patient's tumor relative to a normal tissue sample. By knowing which gene expression changes are required to leave a normal state in a single person, it is possible to design therapeutic strategies tailored for that patient.

GAN
Generation

Merging 2018 and 2020 Results

Team - 2018	Gene	BP-Tumor -2020
studentec	FLG2	-0.569807
BioMarkers.ai	FHL1	-0.370446
HelloKidney2	TAS2R19	-0.363179
ExpressForce	TERT	-0.358329
HelloKidney2	TYMS	-0.287382
...	...	
trimericOGs	HPSE2	0.567236
BioMarkers.ai	PTGER3	0.59603
BioMarkers.ai	DMRT2	0.621588
BioMarkers.ai	UMOD	0.657959
BioMarkers.ai	KNG1	0.668831



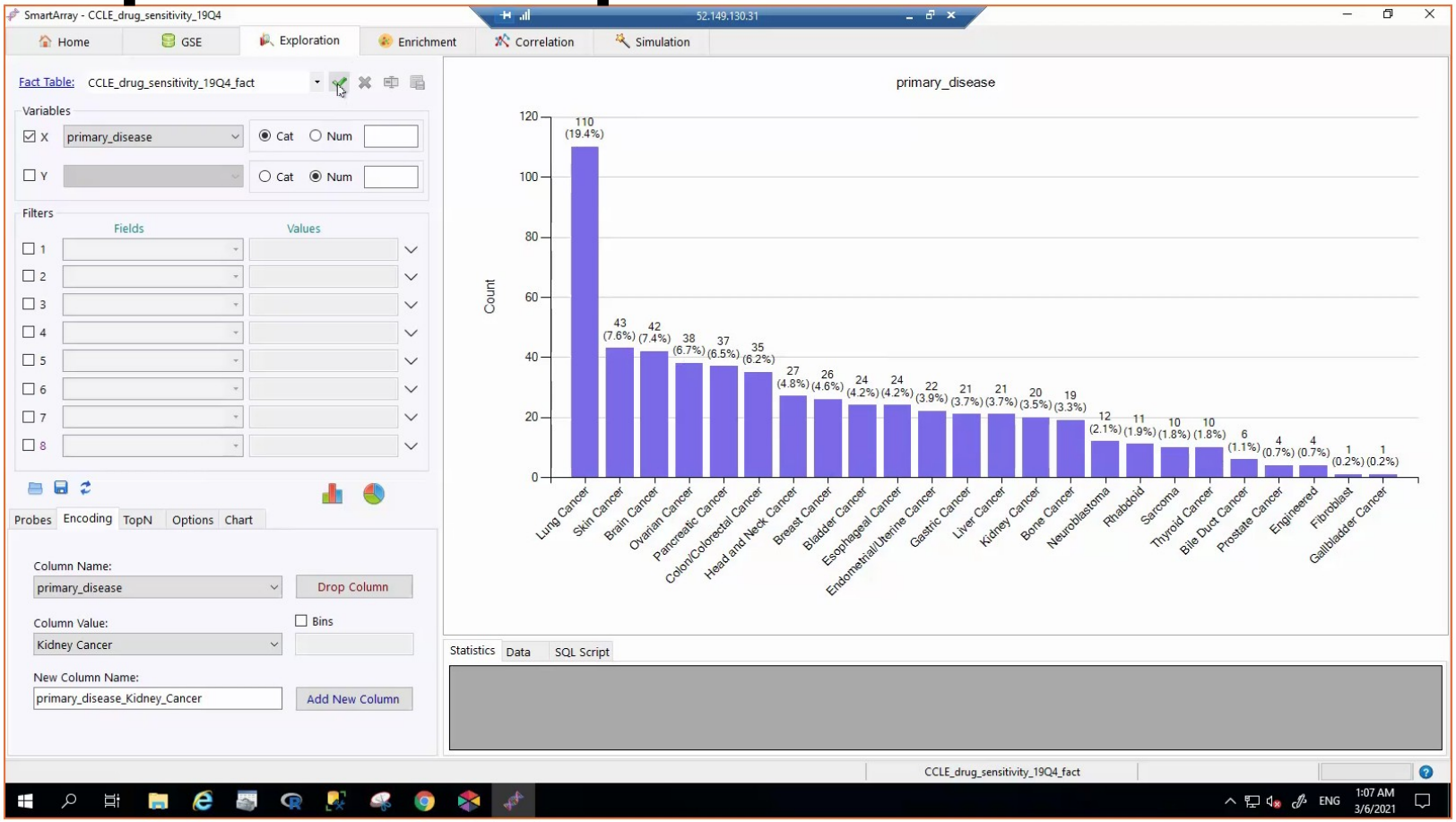
- 2018 p1RCC
  - 119 recommended Genes
- 2020 p1RCC
  - 6 Normalized
  - p1RCC patient's RNA genes

- BioMarkers.ai sorted to either end of the chart.
  - Perhaps diagnostic
  - Likely not therapeutic

Leaderboard Openly

Team - 2018	Gene	BP-Tumor -2020	Approach
studentec	FLG2	-0.569807	https://github.com/SVAI/studentec
BioMarkers.ai	FHL1	-0.370446	https://github.com/SVAI/Biomarkers.AI
HelloKidney2	TAS2R19	-0.363179	https://github.com/SVAI/HelloKidney2
ExpressForce	TERT	-0.358329	https://github.com/SVAI/ExpressForce
HelloKidney2	TYMS	-0.287382	https://github.com/SVAI/HelloKidney2
...	...		
trimericOGs	HPSE2	0.567236	https://github.com/SVAI/trimericOGs
BioMarkers.ai	PTGER3	0.59603	https://github.com/SVAI/Biomarkers.AI
BioMarkers.ai	DMRT2	0.621588	https://github.com/SVAI/Biomarkers.AI
BioMarkers.ai	UMOD	0.657959	https://github.com/SVAI/Biomarkers.AI
BioMarkers.ai	KNG1	0.668831	https://github.com/SVAI/Biomarkers.AI

Therapeutic Options



TBD: Wetlab

- Travera
  - 20 wells on a tray
  - Each with fresh tumor
  - And a different Treatment in each well
- Rare Cancer Research Foundation
- https://www.arctoris.com/
  - Cell Line Labs

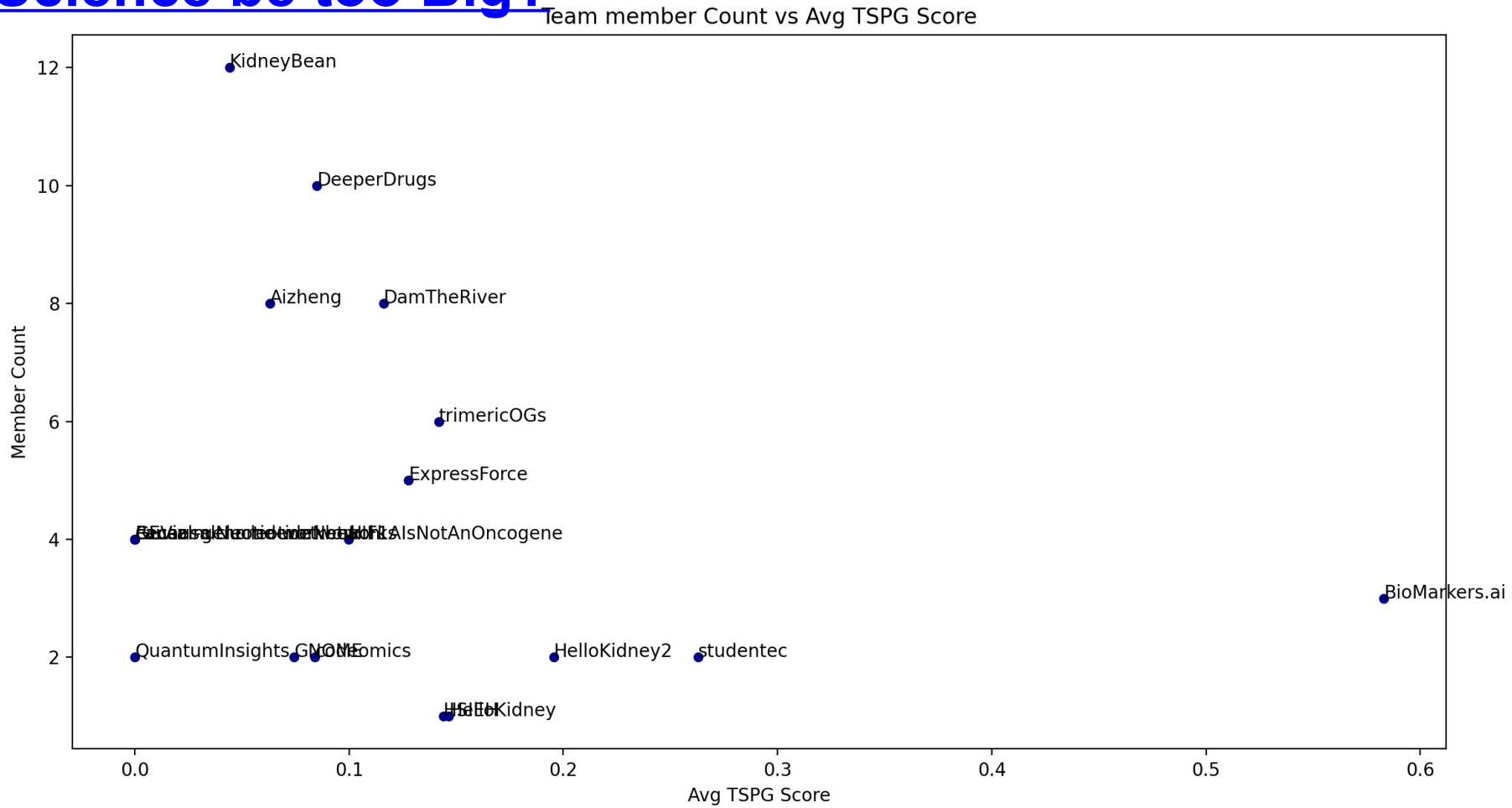
Biomarkers.ai - 2018

- 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. This might explain my medical history.
  - Got warfarin/coumadin for diagnosis of deep vein thrombosis
  - DVT Symptoms returned. Went back and found: 7 cm mass left kidney, cerebral meningioma and spots in lung.
- Uromodulin (encoded by UMOD; also known as Tamm-Horsfall protein) is the most abundant protein in mammalian urine under normal physiological conditions.
  - UMOD can distinguish Normal Tissue from p1RCC with 100% accuracy.
  - Is UMOD also a good urine-based biomarker for p1RCC?
- FHL1 was an indicator for petrochemical exposure. For a time I worked in chemical refineries and on oil rigs. This might be the source of my somatic mutation.
  - Exposure to benzopyrene and several other agents enhances FHL1 expression

Why did BIOada.com do better?

- Saed Sayad came to the hackathon with a set of favorite tools already in place (BIOada.com) which saved analysis time.
- He created a normalized cohort by looking up RNA data on NCBI GEO (Gene Expression) data) using my DNA data as a key. RNA provided a stronger signal than my DNA data, and ultimately matched my RNA-seq data when it became available.
- This stronger signal allowed him to use a simpler data analysis technique (LDA- Linear Discriminant analysis) to get clean data separation and so make better predictions.
- His team was small and focused. Note that a 2019 article entitled "Can Big Science Be Too Big?" posited that papers with few authors tended to report more breakthrough research and papers with many authors tended to confirm existing findings.
- His outsized results are supported by portfolio theory. Dr. Sayad took on a lot of risk (Using one tool, BIOada.com. Abandoning DNA data, using GEO instead. Using one method, LDA. Using a small team, generating fewer new ideas) and so was likely to either get a big win, or go bust.
- In that sense, a hackathon can be viewed as a portfolio of real options, and a hackathon "portfolio" has similar risk/return math to that used in financial portfolio construction.

Can Science be too Big?



Agenda

- **Medical Research Issues from the Rare Disease Patient POV**
  - Reproducibility, Reporting, Data Access, Novelty, Urgency, Inclusion
- **Hackathons – Patient led research modeled after Kaggle**
  - One Patient, Many Competing Teams, Unrestricted Rich Data Set
- **Bill Paseman Papillary Kidney Cancer Hackathon**
  - Process, Report and Conclusions
- ***** **Summary**

Summary

“Patient Centered”

- Patients view themselves as having a “rare disease” that is not served well by cohort analysis. We hope to use sibling and parent genetic data as a "control" in future events.
- Patients themselves host and maintain control of the event and are responsible for providing their own data.
- Data Control allows patients to create a current, longitudinal record over time for each subsequent hackathon as their disease develops.

“Game Elements”

- Hackathon participants are divided up into teams.
- The Game has “levels” which include diagnosis and therapeutic recommendations.
- Team’s results are “scored” which helps the Patient prioritize future research approaches.
- Scores can be posted on a LeaderBoard, which allows sharing of Research Approaches.

“Treat Research Teams as formal computational objects”

- Apply an “Ensemble Learning” technique called "bucket of models".
- For each model m in the bucket:
- Do c times: (where 'c' is some constant)
- Randomly divide the training dataset into two datasets: A, and B.
- Train m with A; Test m with B
- Select the model that obtains the highest average score
-

Summary

- **There are things to try after Clinical Trials Run out**
- **Contact me if you're interested in participating in future hackathons, or if you want to get involved in improving the process (LLMs)**
  - bill@rarekidneycancer.org
- **I describe improvements to the approach tomorrow**
  - **20240416_1750-1810 - Clinical Trials**
  - **Using gamified tumor boards to accelerate cancer research**

Final Word

An obvious point needs to be made explicitly here. Though I contributed data on the front end of the process and did a few SQL table joins and sorts on the back end of the process, NONE of the biology is my work. It is the work of the many volunteer researchers who spent many hours exploring my data on my behalf. Thank You.

"If you work on frequent cancers, do randomized trials! If you work on rare cancers—find friends!"Olson, TA, Schneider, DT, Brecht, IB, et al. Rare tumors: a different perspective on oncology. In: Schneider, DT, Brecht, IB, Olson, TA, Ferrari, A, eds. Rare Tumors in Children and Adolescents. Berlin: Springer; 2012: 3–15.

Acknowledgements

- Tissue: UCSF's Dr. Max Meng and Tasha Lea
- Sequencing: Yale's Dr. Kaya Bilguvar and Christopher Castaldi and UCLA's Dr. Brian Shuch
- Sequencing Experiment Specification and Validation: Mike D'Amour for specifying the sequencing experiment parameters and fastq Validation Process
- 2018 Venue Donation: Salesforce's Steve Tamm and Lisa Ferrier
- 2018 Hackathon Teams
- 2018 Biomarker.ai Lead: Dr. Saed Sayad
- 2018 and 2020 Hackathon Master of Ceremonies: Ben Busby
- 2020 Hackathon: The TRI-con organizer: Kaitlyn Barago of healthtech
- 2020 Hackathon: Research to the People Organizer: Pete Kane
- 2020 Hackathon: "Clemson's 2020 normalized cohort" creators: Reed Bender, Ben Shealy and Benafsh Hussain from Dr. Alex Feltus' group
- 2020 Hackathon: Therapeutic Recommendations: GeneXplain's Dr. Jeannette Koschmann
- 2018 and 2020 Target Identification: QuantumInsights.io's Bernard Chen and Marvin Weinstein
- 2018 Hackathon: sv.ai volunteers: Ryan Leung, Clayton Melina, Lily Vittayarukskul, Hunter Dunbar, Pete Kane, Bill, Dom Jones, Marguerite, David Schachter, Anabelle Tang. Nina Sardesh, Sean Davis

2018 p1RCC HackathonTeams

Team	Members	Summary
Alzheng	Alex Feltus, Ben Shealy, Colin Targonski, Courtney Shearer, Eddie Weill, Ken Matusow, Sufeng Niu, William Poehlman	Model TCGA-RCC tumors as a “time series” across stage
BioMarkers.ai	Peyman Mirtaheri, Saed Sayad, Usman Qazi	Candidate p1RCC Biomarkers and environmental factors influencing expression
cancer-genome-workbench	Betty, rene lopez, Rui, Sarah	Predict/classify a sample cancer type using genetic data with: Unsupervised clustering, Dimensionality reduction, Somatic SNPs, Data exploration
causalnucleotidenetwork	Arkarachai Fungtammasan, Naina Thangaraj, Ola Zalcman, Steve Osazuwa	Variational Autoencoder and tSNE clustering
codeOmics	Daniel Hornburg, Milena Duerrbaum	Biomarkers to precision drugs
DamTheRiver	Andrew Wallace, Christian Clough, Felix Frayman, Matt Callahan, Nandita Damaraju, Pak Yu, Sebastian Nguyen, William Wright	Identification of neo-antigens present within patient P1RCC sequence data

2018 p1RCC HackathonTeams

DeeperDrugs	Andrew Mills, Biter Bilen, Jeff Lam, Lei Tian, Michael D'Amour, Monika Maleszewska, Prasun Mishra, Tahera Zabuawala, XIAOWEI ZHU	Rigorous variant filtering and target pruning
ExpressForce	Amrit Virdee, Maricris Macabeo, Nikhil Balaji, Sofia Medina Ruiz, Yuri Bendana	Netflix for Genes
geviz	Maytas Monsereenusorn, Natnicha Vanitchanant, Navi Tansaraviput, Thanapat Worasaran	Gene Expression Visualization
GNOME	In-Hee Lee, Sek Won Kong	Prioritizing germline and somatic variants potentially associated with p1RCC
HelloKidney	Terje Norderhaug	Autoimmune Clues to Kidney Cancer
HelloKidney2	Clinton Mielke, Robert Van Spyk	Genetic Markers
HIF1AIsNotAnOncogene	Eric Danziger, Joshua Bloomstein, Stephanie Kinnunen, Wanlin Zheng	A preliminary case study in EGFR

2018 p1RCC HackathonTeams

KidneyBean	Bea Nguy, Eric Kalosa-Kenyon, James (3), Jay (3), Kallen Schwark, Kandy Nachimuthu, Mabel Furutsuki, Maninder Singh, Marcus Strauss, Rahim Hashim, Sam Rapp, Wessam Sonbol	Drug candidates towards personal medicine
RecausalNucleotideNetworks	Andrew Carroll, Jason Chin, Pi-Chuan Chang, Samantha Zarate	How Effective Are Illumina Methods for BGI-SEQ? 20180531 BLOG POST
studentec	Brian Hanley, Rush Tehrani	USING BIGQUERY FOR GENOMIC DATA ANALYSIS
trimericOGs	Christine Kim, Lily Vittayarukskul, Phoebe So, Rohith Krishna, Samson Mataraso, senay yakut	Classifying Tumor Stages based on Structural Variants in Patient Data

Summary

- **There are things to try after Clinical Trials Run out**
- **I describe improvements to this approach tomorrow**
  - **20240416_1750-1810 - Clinical Trials**
  - **Using gamified tumor boards to accelerate cancer research**
    - **Ad hoc tumor boards aided me in my Medical Decisions**
      - Evaluating an adjuvant clinical trial: Participate or not?
      - Evaluating radiation therapy: Proton or Photon?
    - **Hackathons formalize and scale the tumor board process**
      - Focusing 17 Gamified Tumor Boards on one rare disease patient advanced Research
    - **Hackathons can be fully automated**
      - Replacing Patients and Tumor Board members with LLMs (Large Language Models).
- **Contact me if you're interested in participating in future hackathons, or if you want to get involved in improving the hackathon process (LLMs are the current focus)**
  - bill@rarekidneycancer.org