AMIRA

Aggregated
Medical **I**nformation for
Research & **A**nalysis

Determining Therapeutic Efficacy via crowd-sourced data and crowd-sourced data science for our collective battle against NF2

presented by
- Saif Ahmed
- Rashid Niaz, MD, MPH
- Eric Schendel, PhD

A Proposal to the SVAI NF2 Genomics Hackathon Panel

Google Launchpad San Francisco, 25 June 2017

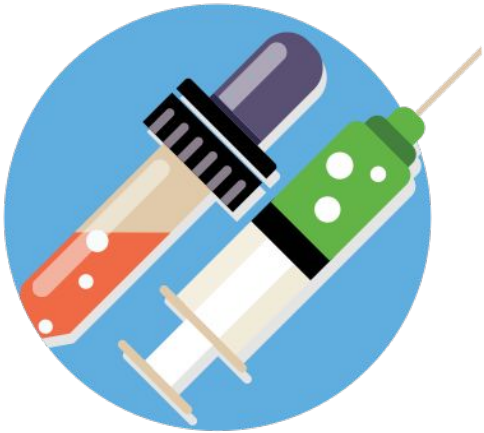
Everyday, selfless individuals donate bone marrow to help those with leukemia and lymphoma.

WOULD INDIVIDUALS DONATE THEIR DATA FOR THE FIGHT AGAINST NF2?

Why Data?

Therapeutic efficacy is not universal, it depends on ***a variety of attributes*** as simple as gender and ethnicity and as complex as various trace levels in blood.

NF2 treatments need to be reviewed so the **best** treatment is given to each patient based on their **individual attributes (Precision Medicine)**. This should be multifaceted -- genetic (as Onno Faber is doing) as well as nutritional, behavioral, geographic, and clinical.



What Data Specifically?

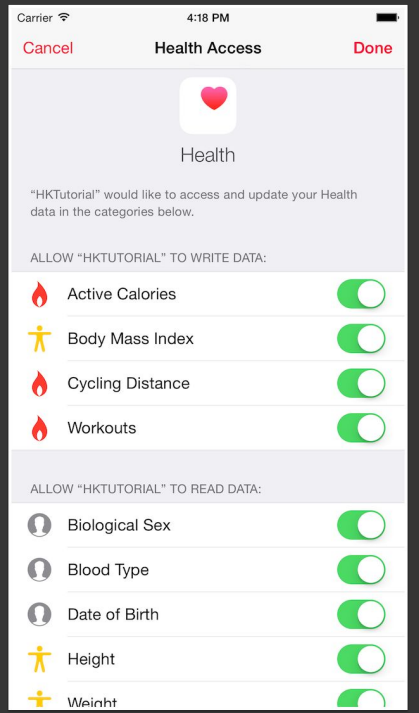
- **Environmental and Lifestyle**
  Geography, local condition indicators, timeline of vitals, activity level, drug abuse, *etc*.
- **Symptoms**
  Hearing loss, ringing in the ears, and balance problems.
- **Nutrition**
  Vitamin and mineral levels will affect response to treatment. For example, B12 is essential for the preservation of the myelin sheath. Vegans, smokers, and drinkers experience decreased B12 levels and poor health outcomes -- *that have nothing to do with the efficacy of medications.* With nutritional data we can segment out such cases, and more subtle ones also.
- **Treatment Course and Outcomes**
  This forms the ground-truth that models will calibrate to and test against.
- **Genetic Information**
  Collection of individual genomic data along with any NF2 mutation variants. This is essential for tumor targeting.

What Data Sources?

- Apple HealthKit
- Google Fit
- Samsung Health
- Nokia/Withings Health Mate
- 23andMe, deCODEme, FamilyTreeDNA via OpenSNP
- MyFitnessPal
- Shimmer
- iHealth
- Personal Genome Project
- PHR/EMR (*e.g.*, Cerner, Epic)

How can we *widen the pool* of data contributors?

- Piggy-back on standard authorization services
- Mobile phone driven initiation
- One-click easy enrollment and contribution
- Ensure trust via independent third party
- CRA/CRM outreach to encourage participation
- Use common growth hacking techniques to ***truly change the world***



How can we *widen the pool* of data scientists?

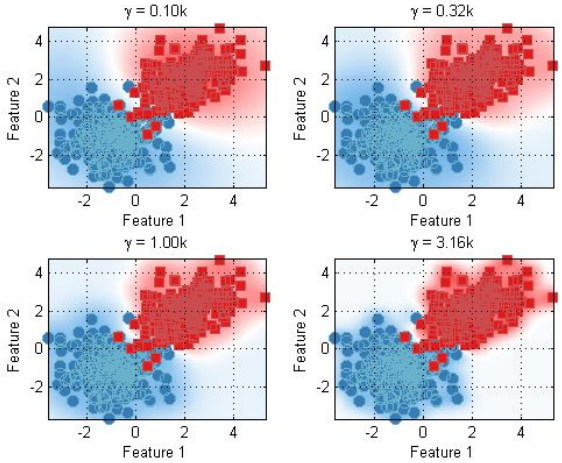
- Data pipelines with auto-downloadable training/test sets
- Ever-expanding training sets on existing pipeline
- Data repackaged into ready-consumable formats (csv, python pickles, numpy, *etc*.)
- Simple blind inference validation
- Use already present community (Kaggle)
- Outreach to C.S. departments to encourage this particular challenge

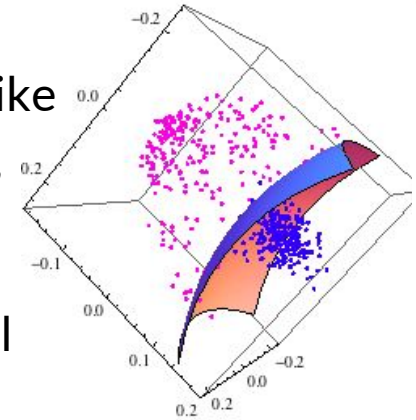
The Objective

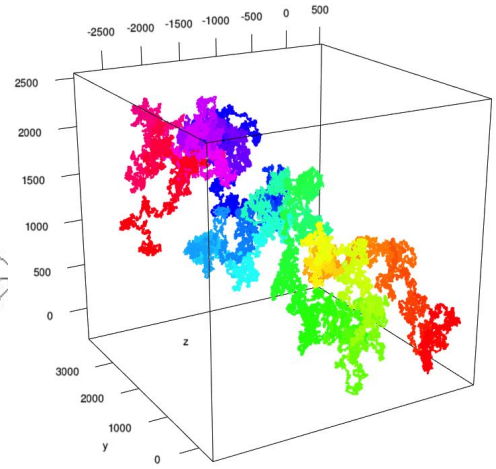
Our objective is to classify patients into segments which will or will-not benefit from particular experimental treatments. Further, we aim to find a minimal set of attributes which drive these segments.

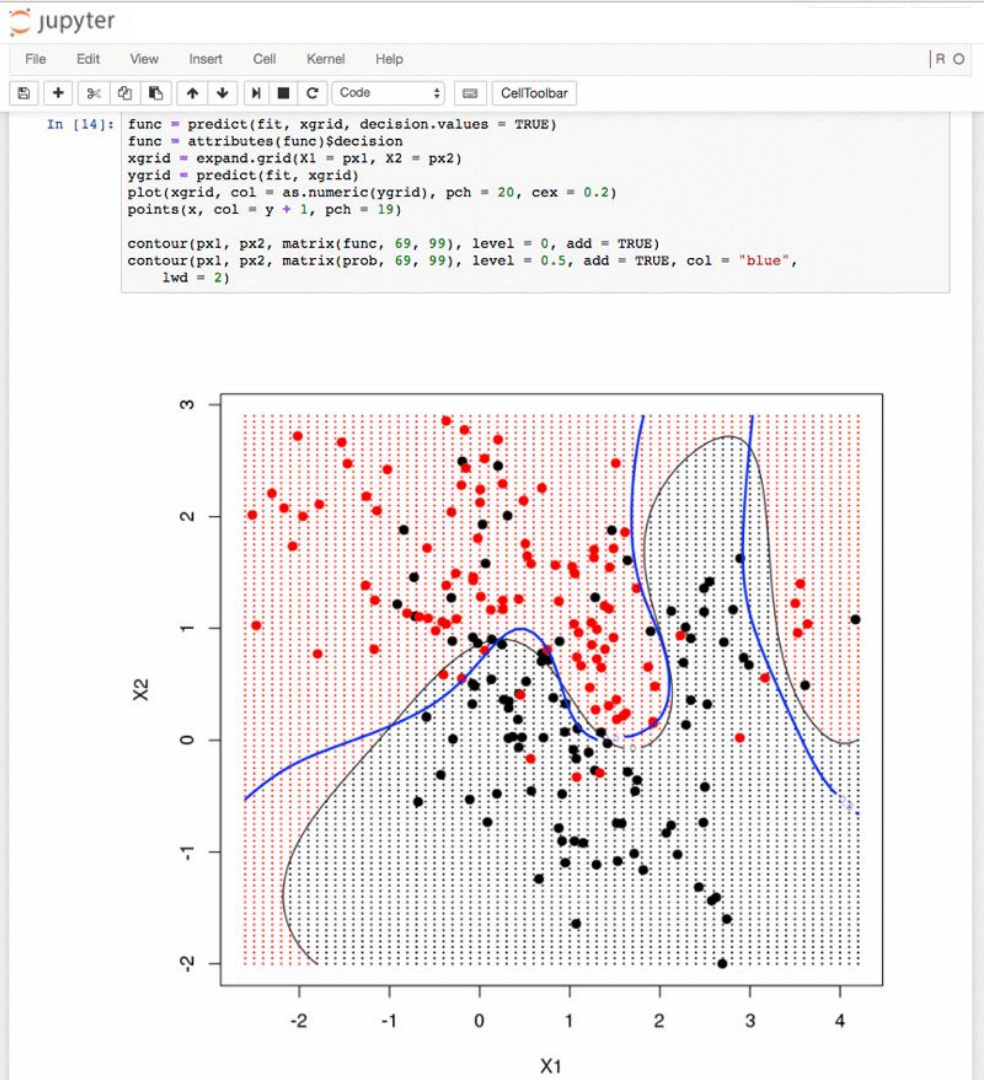
We will achieve this using machine learning: segmentation and classification techniques like logistic regression, support vector machines, random forest decision trees, etc.

Our first step will be to set up fully functional notebooks demonstrating this.









Data Science

We don't have non-genetic data to analyze, but we can use random stand-in data which is roughly pre-clustered for demonstration.

Our repo with some concepts demonstrated: https://github.com/SVAI/AMIRA

Key Approaches Proposed: Support Vector Machines, K-Nearest Neighbors, Logistic Regression, Principal Component Analysis, and Deep Learning if sufficient data possible.

Building Bridges

AMIRA is a "**concentrator model**" -- we want to take many of the groups' impressive work we've seen this weekend and offer a **facility to concentrate** their downstream features into an **experimental therapeutic model translating into more precise studies**.

In essence, we're proposing an overarching effort that includes Onno's efforts around genetic profiles fused with other silos.

We hope to promote an adaptable infrastructure for scientists across disciplines (*e.g.*, computer science + medical) to come together.

Getting out of the way

We do not wish to be constrained by tools. We want consumption to be turn-key so anyone can use any tool to contribute data driven insights and drive progress.

However, we can make it easier with templates and starter material to get people going, *e.g.*, **python, scikit-learn, numpy,** and **Jupyter**.

Our A-Team

Saif Ahmed
saif.ahmed@berkeley.edu
+1.212.729.6544
https://www.linkedin.com/in/saifrahmed/

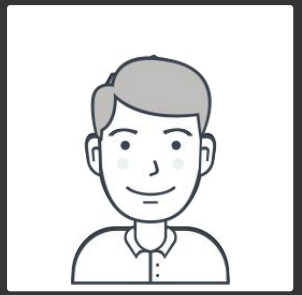
CTO at ML-driven medical imaging diagnostic firm. Ex-Wall St. Quant. U.C. Berkeley graduate student. Health data rebel seeking a cause

Rashid Niaz, MD, MPH
rashid@medeximus.com

Medical doctor and Informaticist with experience in Population Health and Experimental Therapeutics

Eric Schendel, PhD
mabenational@gmail.com

Computer scientist and bioinformatics researcher with extensive large-scale software development experience



What is possible with determined partners?

October 2017

Android & iPhone app with one-click data donation & anonymization from popular health data services

December 2017

Integrations to genomics research outputs into central pipeline

November 2017

Data pipeline and Jupyter-ready machine learning analysis templates: *users hit the ground running*

Beyond

More data sources
More turnkey
More templates
Other conditions (economy of scale)

So what is the ask?

- we *don't* want money
- we need everyone, especially patients to donate data
- we need an independent trustworthy organization to serve as the data concentrator & anonymizer
- we need Kaggle/Google to host the data and blind set verification so the public can participate
- we need enthusiasts like you to build models to segment attributes driving therapeutic efficacy

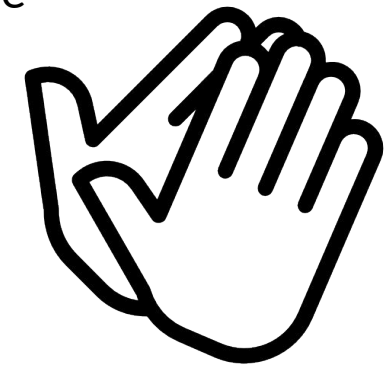
- This is **not that hard** an idea.
- Everything **has already been achieved** in isolation.
- It just needs to be **executed in unison**.

WE CAN ACHIEVE THIS **TOGETHER**

CONTACT US IF YOU CAN HELP

Acknowledgements

- Silicon Valley Artificial Intelligence and Peter Kane
- Onno Faber and NF2 Project
- The Hackathon Mentors and Judges
- Google for the cloud computing credits
- Kaggle for their superb platform (*hint*, *hint*!)
- NF2 Tumor Genomics Hackathon Sponsors:



Google Developers Launchpad Space

Google Cloud

NVIDIA

NF2 Project

RECURSION pharmaceuticals

Google Genomics

SVAI

Accel.AI

powered by NCBI

genlife

Contact Us, *Let's Make it Happen*

Saif Ahmed
saif.ahmed@berkeley.edu
+1.212.729.6544
https://www.linkedin.com/in/saifrahmed/

CTO at ML-driven medical imaging diagnostic firm. Ex-Wall St. Quant. U.C. Berkeley graduate student. Health data rebel seeking a cause

Rashid Niaz, MD, MPH
rashid@medeximus.com

Medical doctor and Informaticist with experience in Population Health and Experimental Therapeutics

Eric Schendel, PhD
mabenational@gmail.com

Computer scientist and bioinformatics researcher with extensive large-scale software development experience