



How Clinicians Access and Benefit from Precision Health Tools

Erin O'Brien Kaleba
Brahmajee Nallamothu



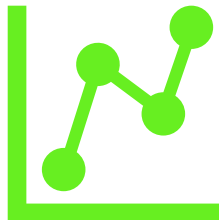
Overview

Data



EMR
Genetic
Wearables
SES
Geolocation

Compute



Analytic
Platform

Facilitation



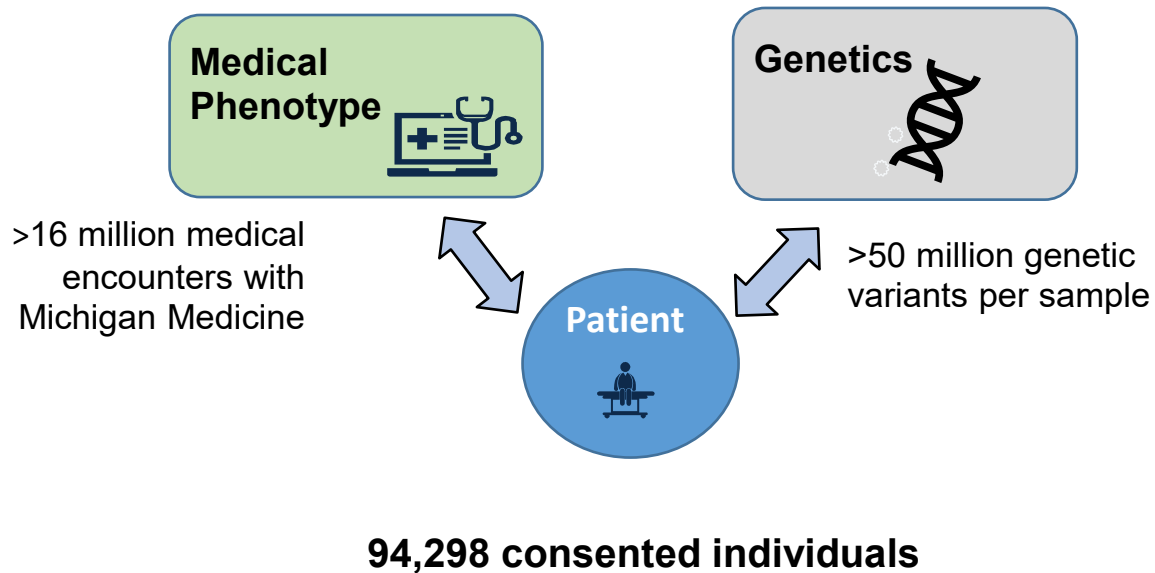
1:1 Consults
with Research
Facilitators

Examples

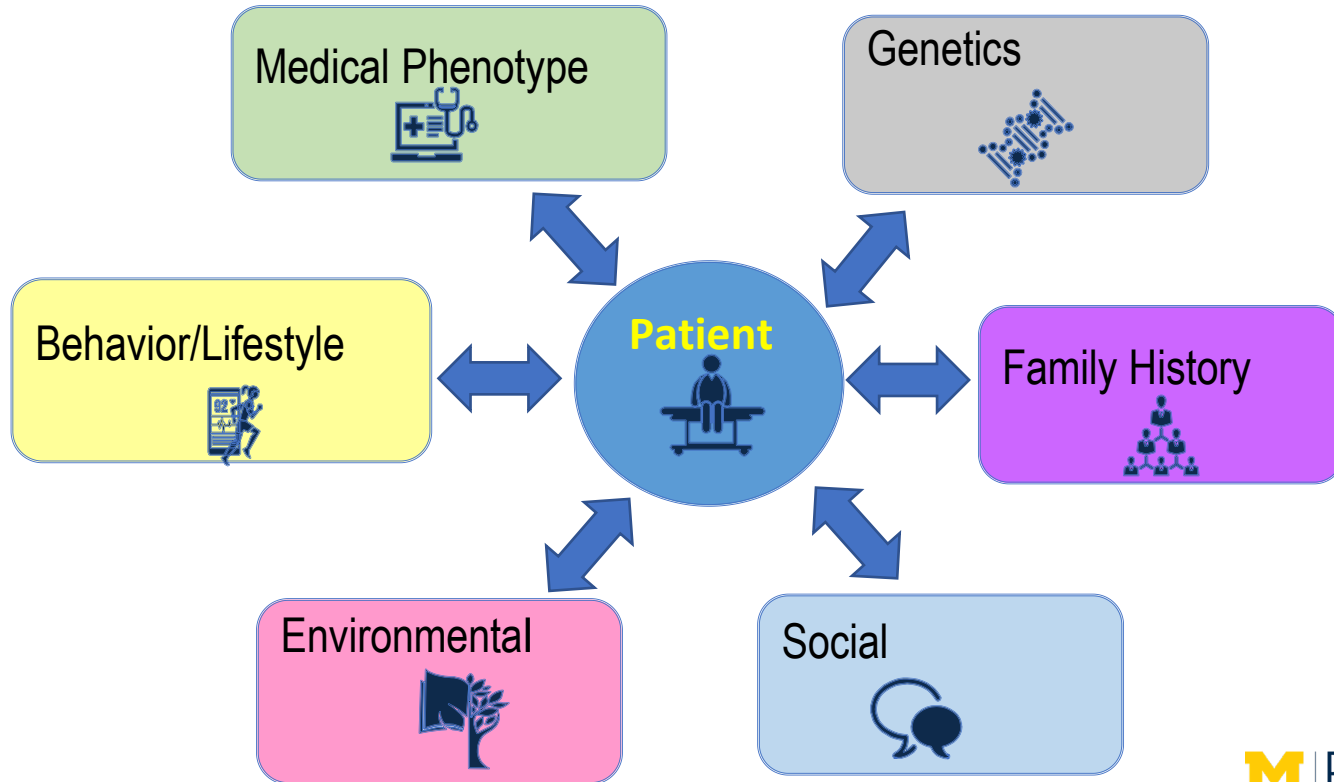


3 Recent
Success
Stories

Data: How it started



Data: How it's going



Data: What is Available?

Medical Phenotype

- Vitals
- Medications
- Labs
- Encounters
- Diagnoses
- Procedures
- Orders

4+ Million Patients

Behavior/Lifestyle

Genetics

Environmental

Family History

Social

Data: What is Available?

Environmental

- Latitude/Longitude
- Census Block ID

>100 SES data elements
(eg, affluence, household
income, education, parks)

Medical Phenotype

Behavior/Lifestyle

Social

Genetics

Family History

Data: What is Available?

Behavior/Lifestyle

- Wearables
- Pain Levels
- Depression/Anxiety
- Physical Activity
- Sleep

Medical Phenotype

Environmental

Social

Genetics

Family History

Data: What is Available?

Genetics

- GWAS data on 60K patients
- Whole Exome
- Targeted Sequencing
- Self-serve tool called *Encore*

Medical Phenotype

Behavior/Lifestyle

Social

Environmental

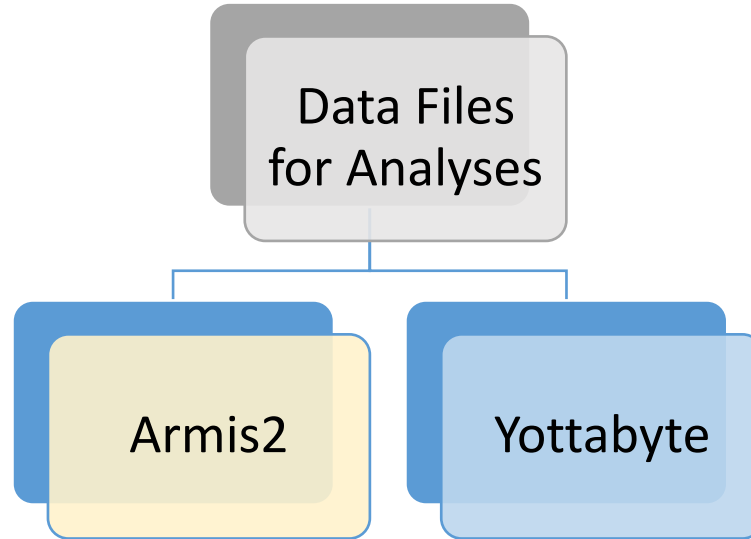
Family History

Data: What is Available?

Data Type	Description	#Participants
Genome-wide genotypes	~600K variants directly assayed by genotyping array and genotype imputed to > 51M variants with the Trans-Omics for Precision Medicine reference panel or > 32M variants with the Haplotype Reference Consortium panel. All currently available genotypes were assayed on the Infinium CoreExome array . In the future: Infinium Global Screening Array	60,215
Whole exome sequences	Sequence data covering protein coding gene regions (~2% of genome) as captured by the Roche/Nimblegen SeqCap EZ v2.0 or Agilent SureSelect V5-post systems	561
Targeted sequences	Sequence data covering 151 targeted gene regions	963
HLA gene allele and amino acid inferences	Inferences for human leukocyte antigen genes HLA-A, -B, -C, -DQA1, -DQB1, -DRB1, -DPA1, and -DPB1	60,215
Pharmacogenomic star allele inferences	Inferences for 51 distinct pharmacogenes with polymorphic alleles, including CYP2C9, CYP2B6, CYP2C19, CYP3A5, NUDT15, TPMT, SLCO1B1, UGT1A1, DPYD, and CYP2D6*	60,215

Analytic Platform

- High-Performance Computing Environment



- HIPAA aligned
- **Linux based** 2&4-socket nodes
- 24 cores, 156 GB memory
- 12 V100 GPUs

- HIPAA aligned
- **Windows based**
- Private cloud environment
- High-performance, secure, and flexible computing environments

Facilitation

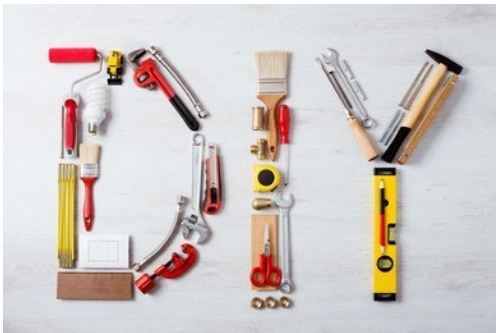
- **Research Scientific Facilitators (2.0 FTE)**

- Regulatory (IRB) Expertise – *why type of application do I need?*
- Data Expertise – *do you have ejection fraction?*
- Compliance Expertise – *can I store this on DropBox?*
- Technical Expertise – *how many GPUs are available for my analyses?*
- Data Sharing Expertise – *can I share these data with Duke? With Industry?*

Multiple Pathways for Access

- **Self Serve Access**

- DataDirect – structured data
- Direct DB access – SQL Server
- EMERSE – free text
- Encore - association tests with large-scale genetic sequencing data



shutterstock.com · 371937541

- **Custom Data Extract**

- Expert SQL programmers
- One-time or ongoing data feeds
- Linkage across disparate data sets
- \$60/hourly rate



What's Next

1. Computable Phenotypes
2. Radiology Images for ML
3. Claims data (CMS, BCBS)
4. Natural Language Processing
5. Industry – preferred relationships
6. What else?



EXAMPLES