MAYO CLINIC

Why is it so Challenging to Share Genomic Data?

AN INFORMATICS PERSPECTIVE

Robert R. Freimuth, PhD

NHGRI Genomic Medicine XV: Genomics and Population Screening
Bethesda, MD
November 9, 2023

OBJECTIVES

- Current state of genomic data sharing
- Examples of challenges: screening data
- Review selected research opportunities

- NOT: focus on standards or technical infrastructure

SHARING GENOMIC DATA

What?

- Observed data
- Lab interpretation
- Derived interpretation

How?

- System capabilities
- Formats (standards)
- Use cases

When? Why? Who?

- Clinical
- Policy

INTEROPERABILITY

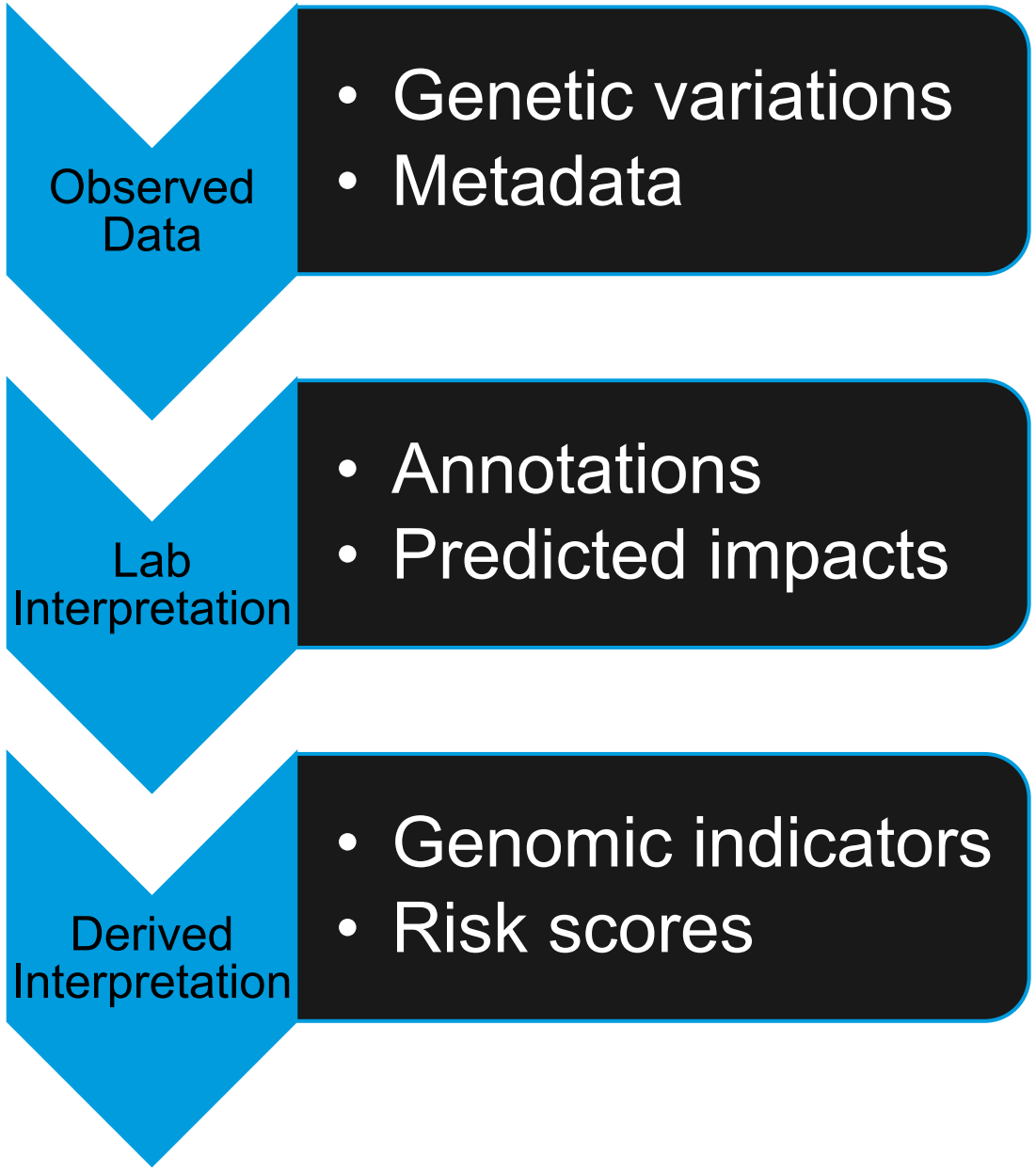
Interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged. (HL7 EHR Interoperability Project Team, 2007)

Semantic

- Ability to interpret and make effective use of the information
- Requires a common understanding of the meaning of the information
- Defines the "things" in a system or data set
  - Names, definitions
  - Relationships

Syntactic/Functional

- Capability to reliably exchange information without error
- Requires the use of a common format to represent the information
- Common platform for data exchange
  - Messaging protocol
  - File format



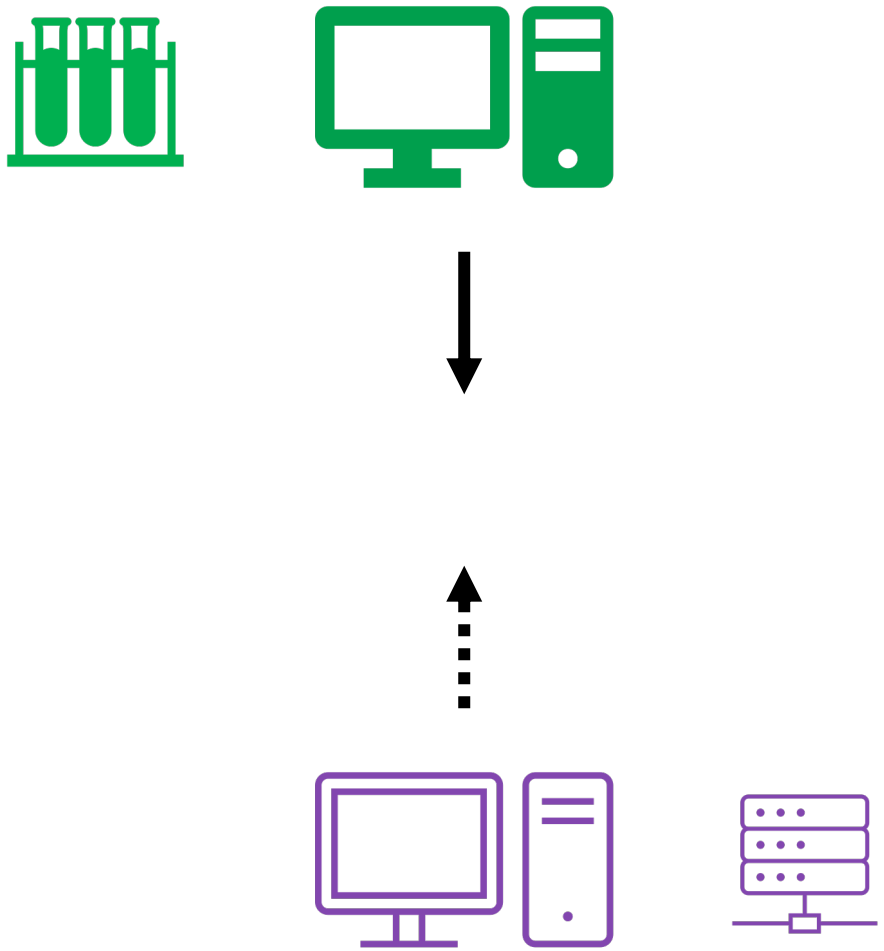
Challenges

- Data representation
  - Complex domain
  - Nuanced semantics
  - Disparate standards
- New data types
- New use cases
- Evolving knowledge

- Coupling of data and use cases
- Knowledge management
- Provenance and metadata

COMPLEX USE CASES FOR GENOMICS TESTING

- Structural variation/rearrangements
- Trio testing, family studies
  - Multiple WES/WGS (>2 subjects)
  - Secondary findings/analysis
- Tumor:Normal testing
  - Serial tumor sequencing
  - Same “normal” reference
- Cascade/reflex testing
  - Observed results produced by different assay technologies
- “Negative” tests are still informative
  - Need to know regions assayed and technologies used even if no variants are reported
- Copy number variation
- Variant reinterpretation
  - New knowledge, regardless of report or assay
  - New test result, in context of multiple other variants from previous tests
  - Need access to non-reported results
- Result reconciliation
  - Screening => diagnostic
  - Targeted <=> WES/WGS
- Risk score calculation
  - Pharmacogenomics: drug selection or dosing algorithms
  - Polygenic risk scores



MAKING THE CONNECTION

1. Change one to fit the other
2. Adopt a common standard
3. Use an adapter

TYPES OF STANDARDS

- Data standards
  - Information models
  - Data elements (observation, result)
- Process standards
  - Method performed
  - Result interpretation process (translation)
- Terminologies/ontologies/etc
  - Coded results
  - LOINC, RxNorm, ICD, SNOMED
- Message/interface standards
  - HL7 V2, FHIR
  - BAM, (g)VCF, BED

Genetic data (variation)
Test metadata

Haplotype inference
Phenotype translation

Molecular phenotype
Treatments (drugs, procedures)

EHR interfaces
Bioinformatics tooling

GENOMIC DATA STANDARDS:

REQUIREMENTS TO SUPPORT CLINICAL IMPLEMENTATION OF TRANSLATIONAL RESEARCH

Clinical

- Clinical test results
- Variety of data representations
  - Nomenclatures
  - Free text
- Usable within clinical systems (e.g., CDS)

Research

- Research results
- Tightly defined data elements
  - Atomic, discrete
  - Value constraints
- Usable with research tooling and databases

- Supports data normalization
- Extensible for new data types and conventions
- Supports a unified approach to genomic data management

COMBINING THE BEST OF BOTH WORLDS: WORK IN PROGRESS

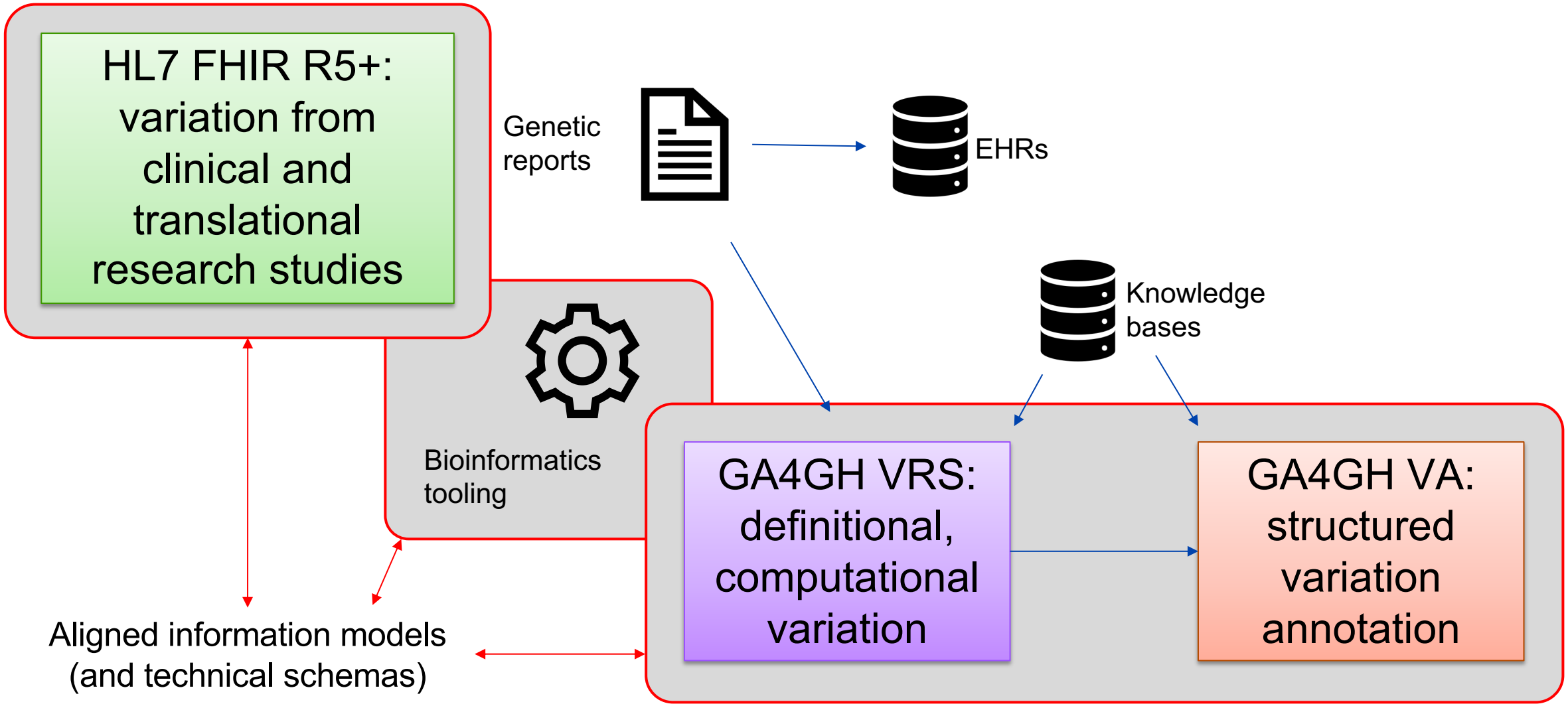
HL7 FHIR

- Must support unstructured text
- Very complex but extensible
- Defines clinical context and use
- Native to clinical systems
- Clinical decision support rules

GA4GH GKS

- Computable representations
- Minimalistic, no optional fields
- Agnostic of use case
- Near-native to research software
- Public knowledge bases

STANDARDS ALIGNMENT: VISION



CHALLENGES:

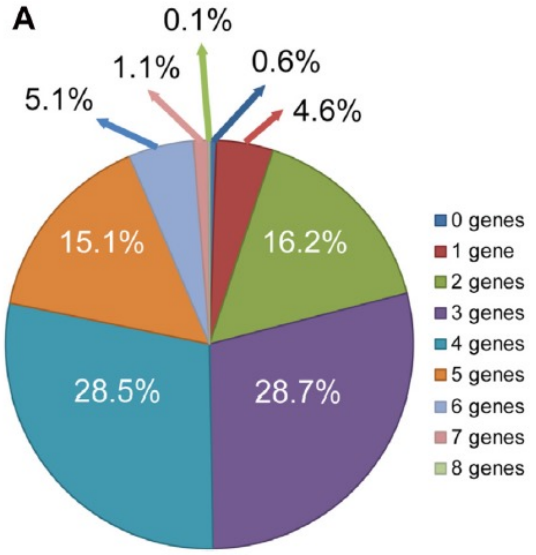
DERIVED DATA PROVENANCE METADATA

Table 1 Drug–gene pair alerts implemented in the Mayo Clinic EHR by year of implementation

Drug	Gene(s)	Year implemented
Abacavir	*HLA-B*57:01*	2013
Azathioprine	*TPMT* and *NUDT15*[a]	2013
Carbamazepine	*HLA-B*15:02* and *HLA-A*31:01*[b]	2013
Codeine	*CYP2D6*	2013
Mercaptopurine	*TPMT* and *NUDT15*[a]	2013
Tamoxifen	*CYP2D6*	2013
Thioguanine	*TPMT* and *NUDT15*[a]	2013
Tramadol	*CYP2D6*	2013
Allopurinol	*HLA-B*58:01*	2014
Clopidogrel	*CYP2C19*	2014
Simvastatin	*SLCO1B1*	2014
Warfarin	*CYP2C9* and *VKORC1*	2014
Citalopram	*CYP2C19*	2015
Escitalopram	*CYP2C19*	2015
Fluvoxamine	*CYP2D6*	2015
Fluoxetine	*CYP2D6*	2015
Paroxetine	*CYP2D6*	2015
Venlafaxine	*CYP2D6*	2015
Tacrolimus	*CYP3A5*	2016
Capecitabine	*DPYD*	2017
Fluorouracil	*DPYD*	2017

```
CYP2C19 Genotype, B
     CYP2C19 Phenotype       Poor metabolizer
     CYP2C19 Star Alleles    2/2
```

THE MAYO-BAYLOR RIGHT 10K STUDY



B

Number of Genes	Number of Subjects	% of Subjects
0	59	0.6%
1	465	4.6%
2	1636	16.2%
3	2893	28.7%
4	2869	28.5%
5	1521	15.1%
6	514	5.1%
7	111	1.1%
8	9	0.1%

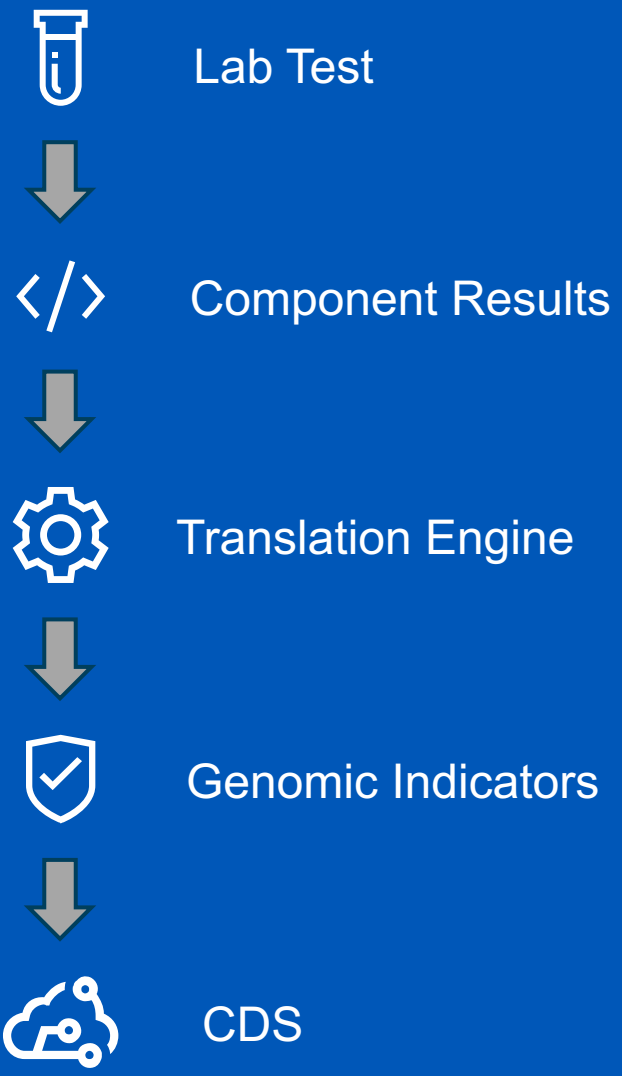
Figure 1 **Percentage of study subjects harboring clinically actionable PGx variants.** The figure shows the number of genes that contained clinically actionable genomic variants for the 13 genes included in the drug–gene pair alerts listed in Table 1 that were observed in each of the 10,077 RIGHT 10K Study subjects and the percentage of study subjects included in each group. A. The pie chart shows these data graphically, whereas the table in (B.) lists the information upon which the pie chart is based.

>10,000 participants
77 pharmacogenes
Pre-emptive sequencing

Drug-based CDS
Education for providers and patients

Implementation of preemptive DNA sequence–based pharmacogenomics testing across a large academic medical center: The Mayo-Baylor RIGHT 10K Study
Wang L, *et al.* Genetics in Medicine 2022

PHARMACOGENOMICS IMPLEMENTATION



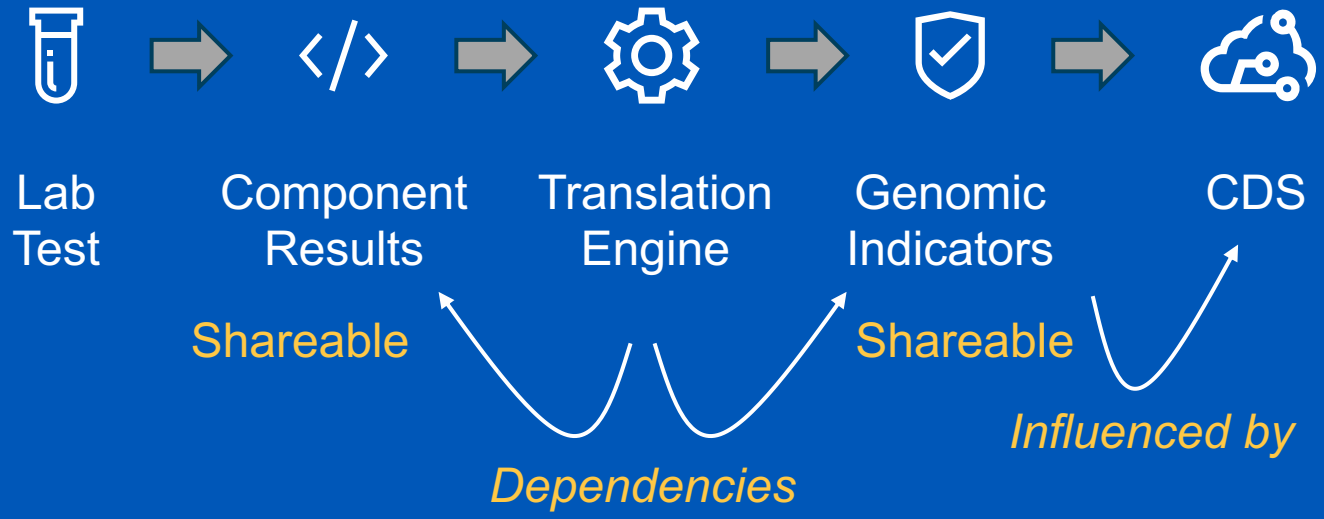
- 61 lab tests supported
- 256 component results configured
- 148 genomic indicators defined
- >400k GIs on >38k patients' charts
- 234 PGx-related CDS rules live

Table 1 Drug–gene pair alerts implemented in the Mayo Clinic EHR by year of implementation

Drug	Gene(s)	Year implemented
Abacavir	*HLA-B*57:01*	2013
Azathioprine	*TPMT* and *NUDT15*[a]	2013
Carbamazepine	*HLA-B*15:02* and *HLA-A*31:01*[b]	2013
Codeine	*CYP2D6*	2013
Mercaptopurine	*TPMT* and *NUDT15*[a]	2013
Tamoxifen	*CYP2D6*	2013
Thioguanine	*TPMT* and *NUDT15*[a]	2013
Tramadol	*CYP2D6*	2013
Allopurinol	*HLA-B*58:01*	2014
Clopidogrel	*CYP2C19*	2014
Simvastatin	*SLCO1B1*	2014
Warfarin	*CYP2C9* and *VKORC1*	2014
Citalopram	*CYP2C19*	2015
Escitalopram	*CYP2C19*	2015
Fluvoxamine	*CYP2D6*	2015
Fluoxetine	*CYP2D6*	2015
Paroxetine	*CYP2D6*	2015
Venlafaxine	*CYP2D6*	2015
Tacrolimus	*CYP3A5*	2016
Capecitabine	*DPYD*	2017
Fluorouracil	*DPYD*	2017

```
CYP2C19 Genotype, B
     CYP2C19 Phenotype       Poor metabolizer
     CYP2C19 Star Alleles    2/2
```

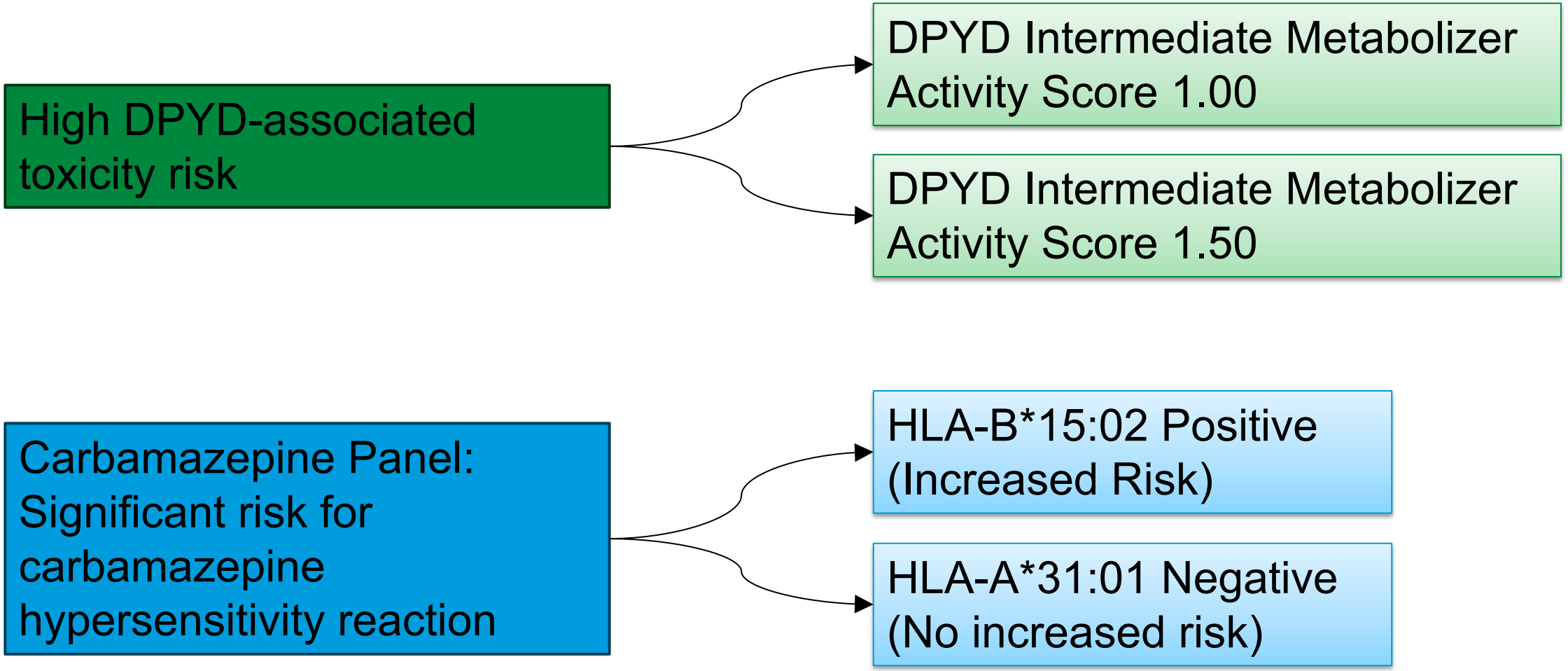
PHARMACOGENOMICS IMPLEMENTATION



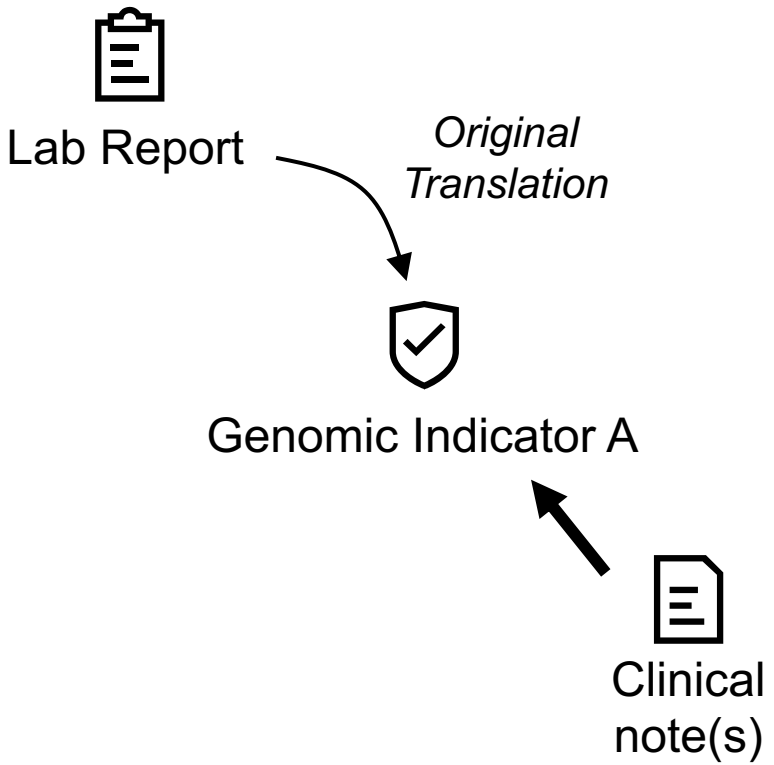
Localization complicates sharing

Implementation of preemptive DNA sequence–based pharmacogenomics testing across a large academic medical center: The Mayo-Baylor RIGHT 10K Study
Wang L, *et al.* Genetics in Medicine 2022

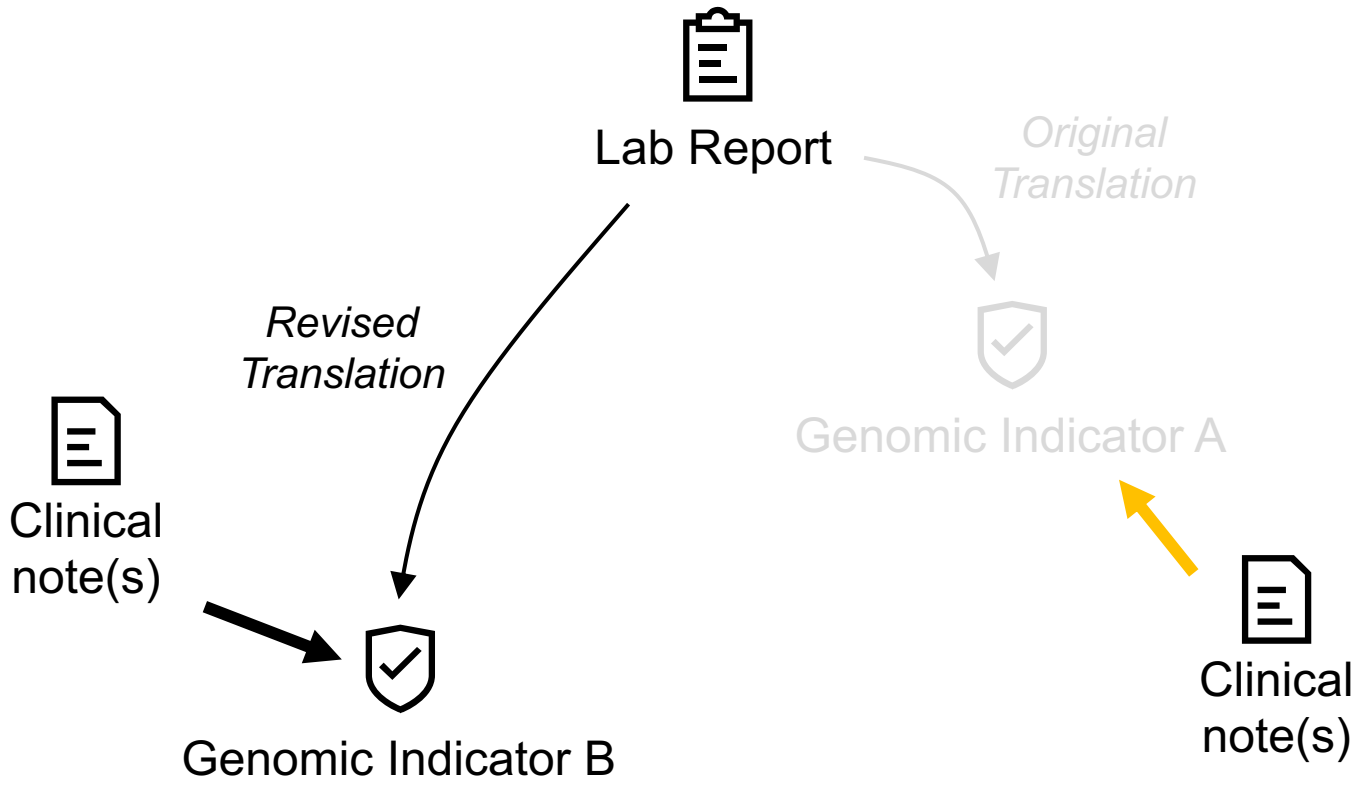
CHANGING KNOWLEDGE AND IMPLEMENTATIONS



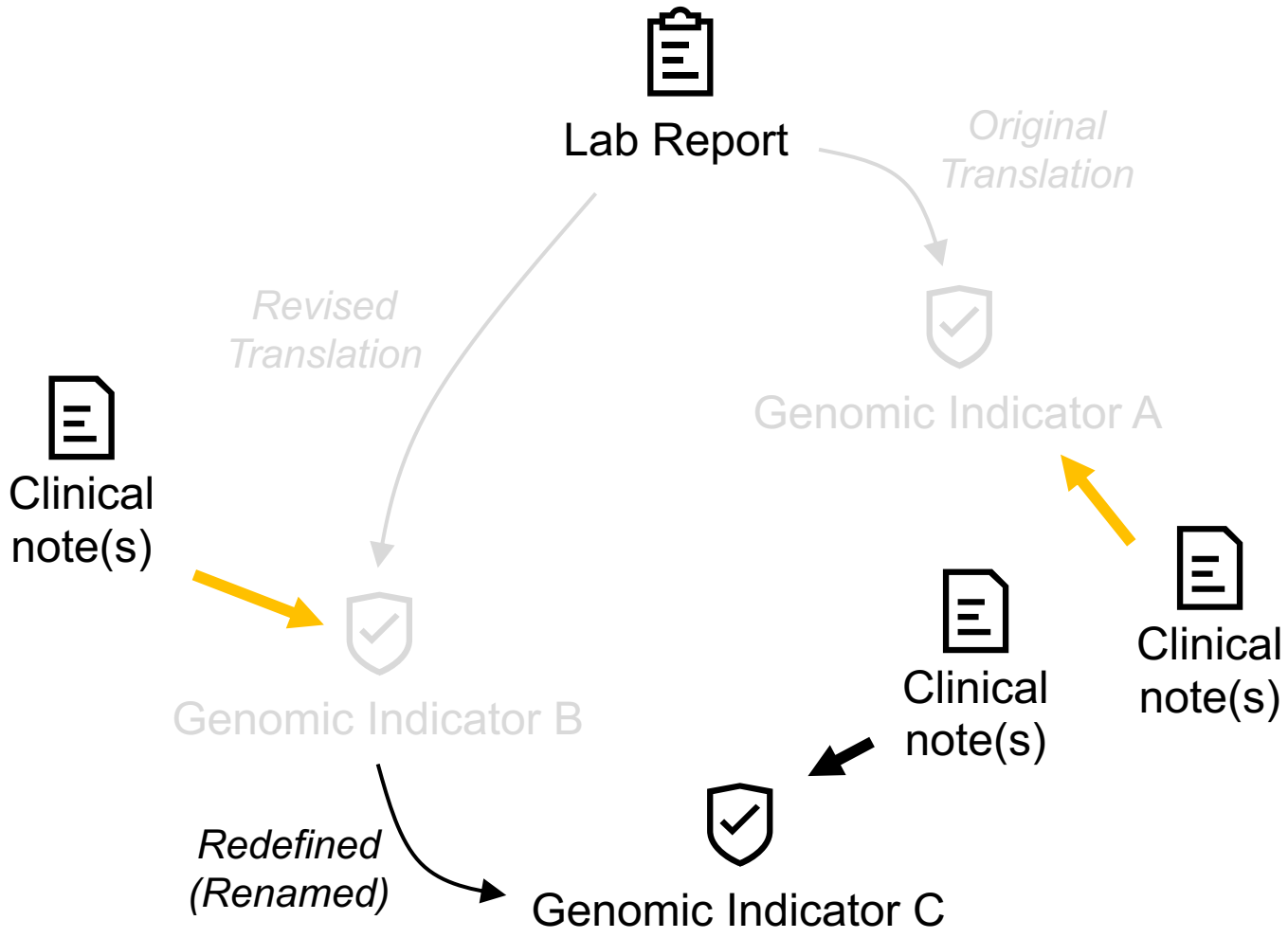
THE NEED FOR PROVENANCE



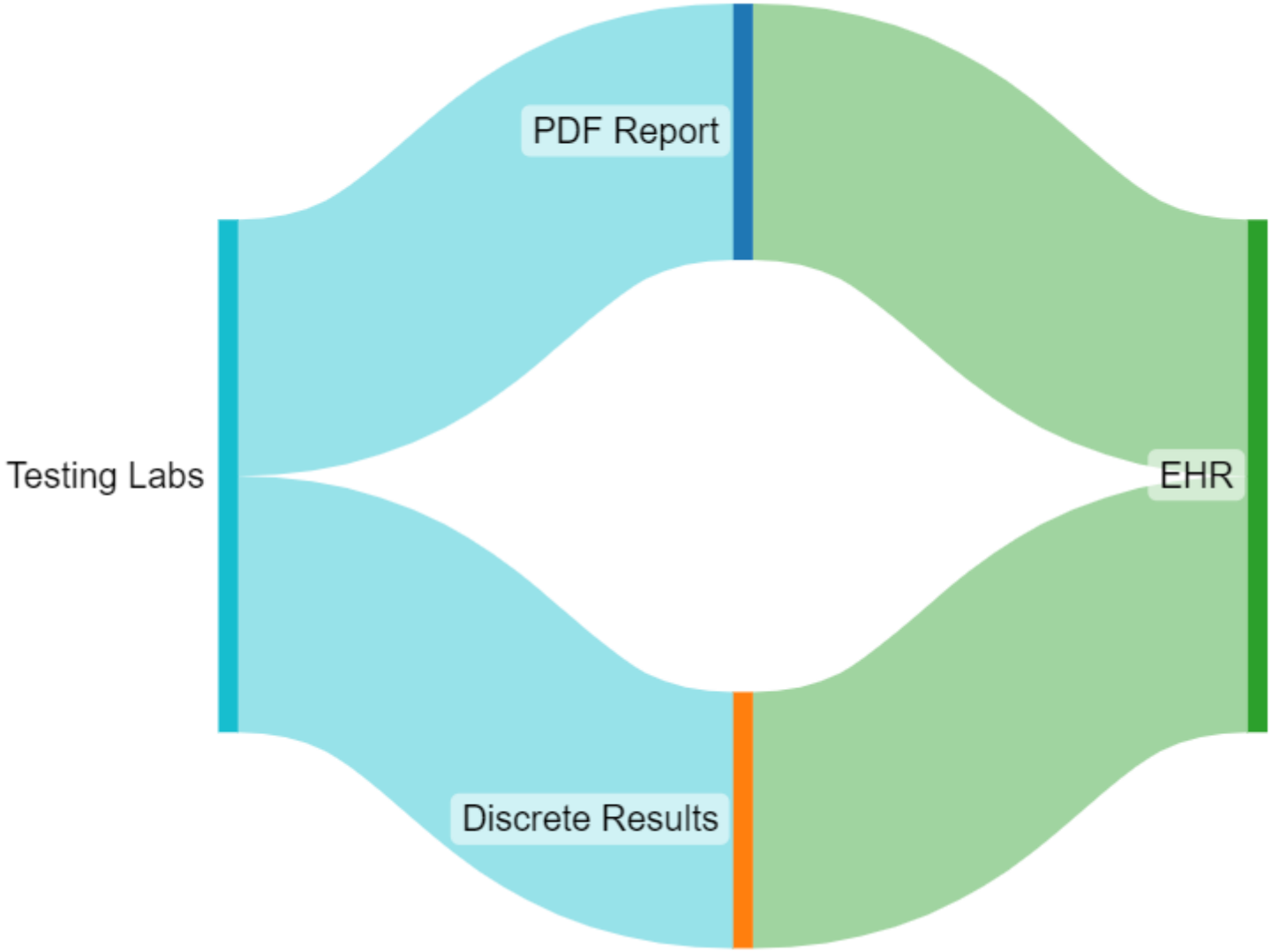
THE NEED FOR PROVENANCE



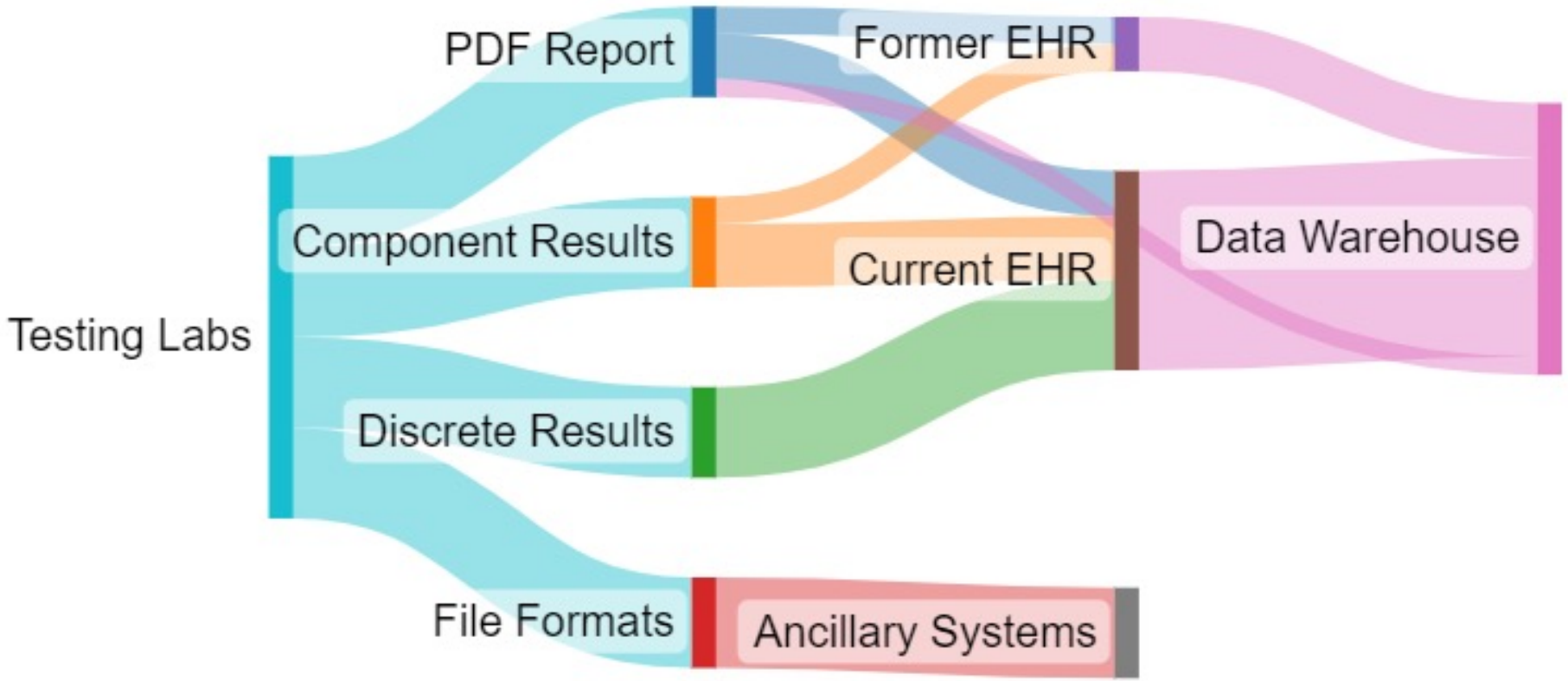
THE NEED FOR PROVENANCE



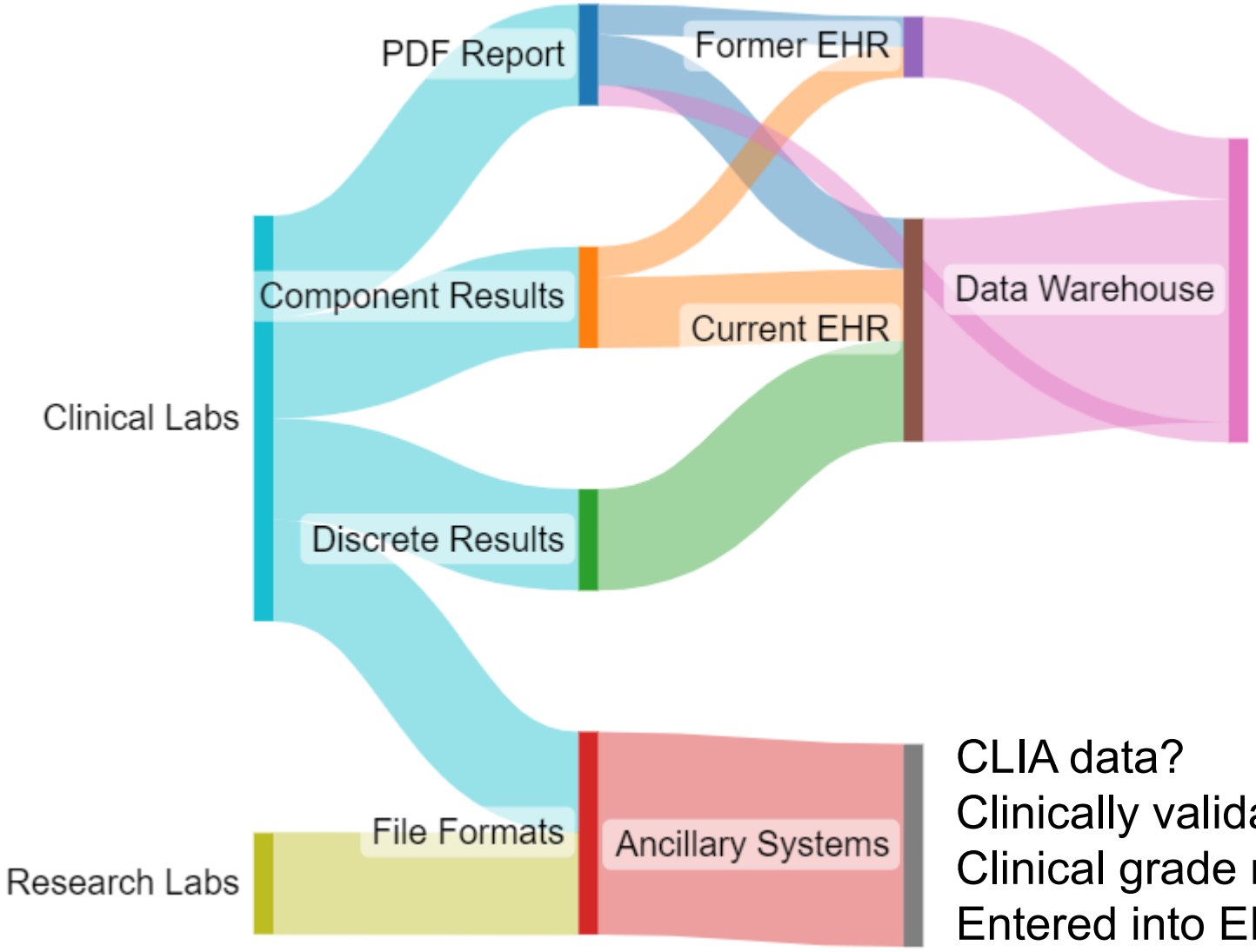
DATA FLOW



DATA FLOW: MORE REALISTIC

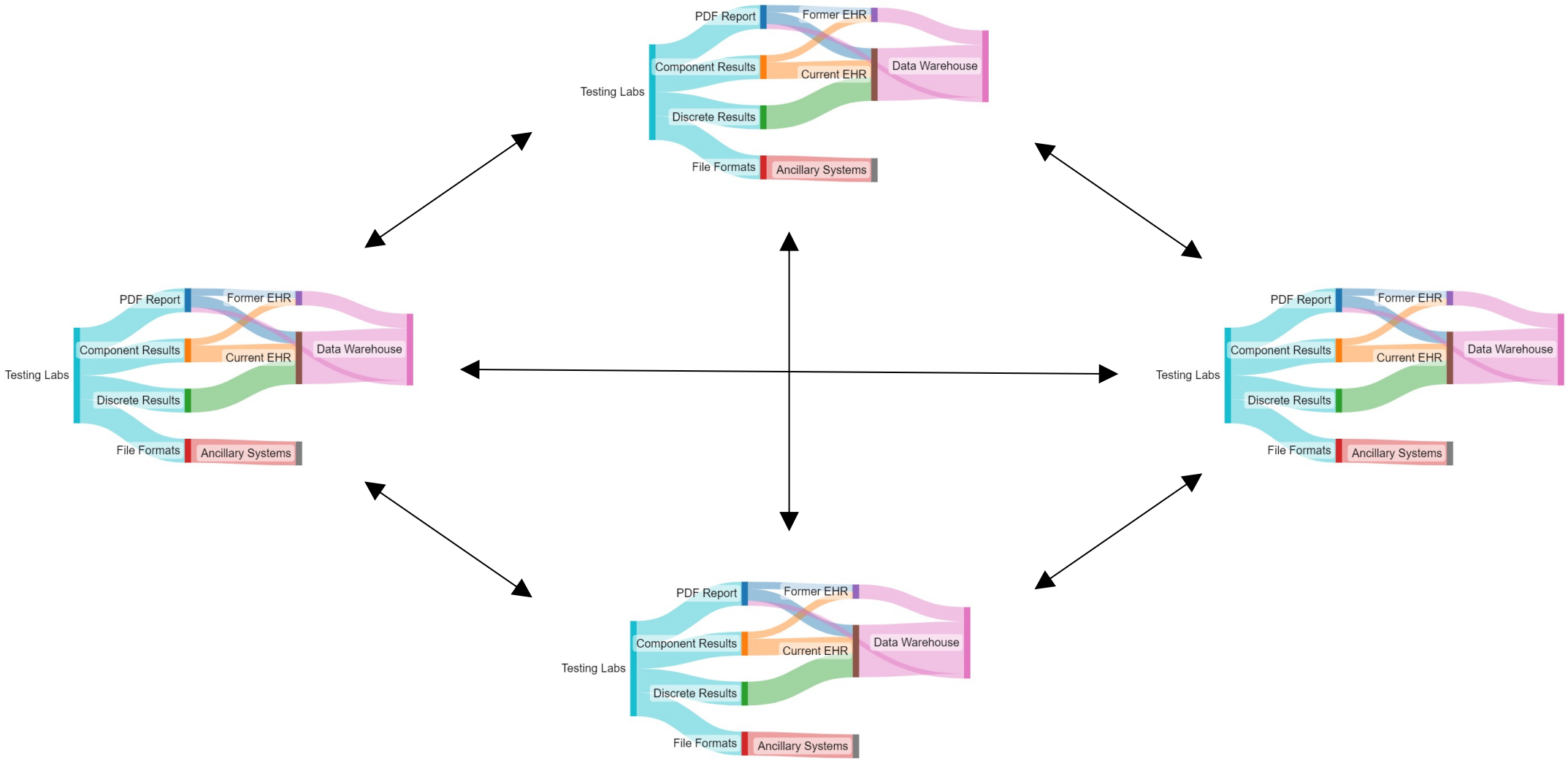


DATA FLOW: MORE REALISTIC, WITH PROVENANCE



CLIA data?
Clinically validated analysis?
Clinical grade result?
Entered into EHR?

DATA SHARING



REFLECTING ON THE PAST… LOOKING TO THE FUTURE

2023

2013

2003

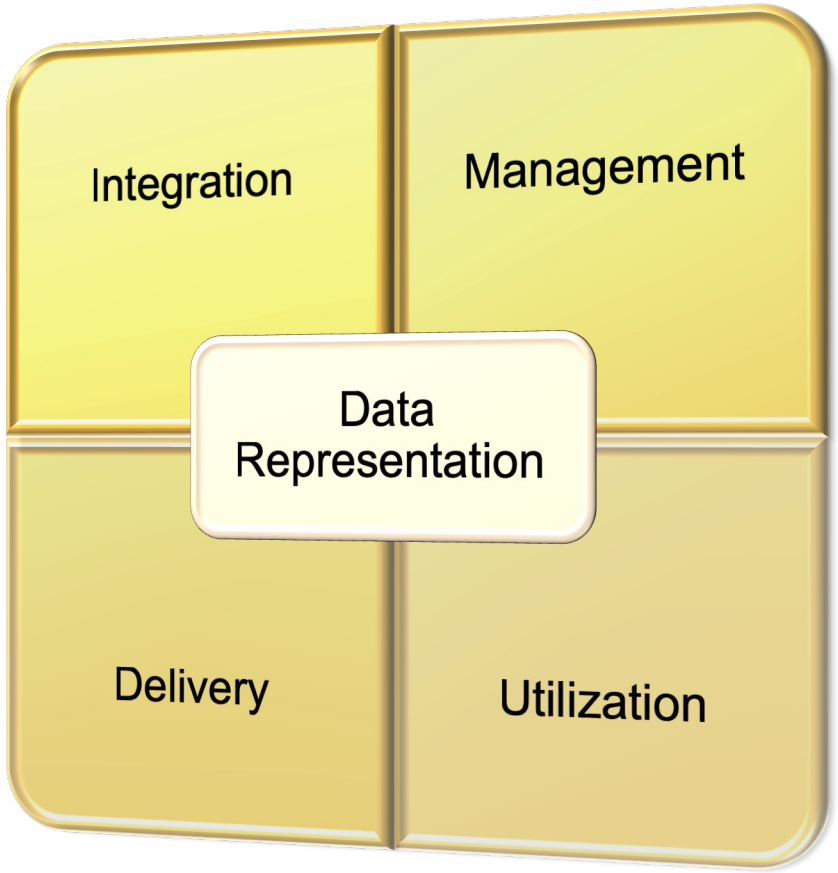
1993

What can we learn from 30 years of genetic testing?

How can we ensure the data that are generated today can be accessed in 2050?

RESEARCH OPPORTUNITIES

- Data and terminology standards
  - Precise yet extensible
  - Generalized yet supports specialization
  - Harmonized across clinical and research
- Define types of derived data
  - Conceptual models inform standards
- Knowledge management
  - Provenance and metadata
  - Human-readable *and* computable



MAYO CLINIC

THANK YOU

- Mayo Clinic Center for Individualized Medicine
  - RIGHT and Tapestry studies
  - ODP and CORE teams
  - Clinical Genomics Advanced Technologies (CGAT) team
- Mayo Clinic eMERGE team
- GA4GH GKS Work Stream
- HL7 Clinical Genomics Work Group
- NIH NHGRI R35HG011899
- NIH NHGRI U24HG006834
- NIH NHGRI U01HG006379

