

iSITE: a consolidated view for efficient research

Nicole Villaverde
October 5, 2017



**Mount
Sinai**

iSITE: Integrated System for IBD daTa Exchange

- ▶ LabKey Server-based platform for consolidation of research data
- ▶ Research focus is Inflammatory Bowel Disease (IBD)
 - IBD describes conditions characterized by chronic inflammation of the gastrointestinal tract
 - Two types: Crohn's disease (CD) and ulcerative colitis (UC)
 - In 2015, an estimated 1.3% of US adults (3 million) reported being diagnosed with IBD
 - In 1932, Dr. Burril B Crohn, Dr. Leon Ginzburg, and Dr. Gordon Oppenheimer from Mount Sinai Hospital in New York described the regional ileitis condition that became known as Crohn's disease.
- ▶ Many contributors and users: researchers, clinicians, clinical research coordinators, biostatisticians, lab technicians, etc.
- ▶ Goal is to integrate data with minimal disruption to current study workflow

IBD research at Mount Sinai

Road to Prevention
High risk families

General Genetics
General IBD
recruitment

Legacy
Archival IBD
recruitment

MSCCR
Mount Sinai Crohn's
& Colitis Registry

MECONIUM
Disease transmission
in utero

- 
- Duplication
 - Compatibility
 - Accessibility
 - Data Integration

BioMe
IBD patients from
MS IPM BioBank

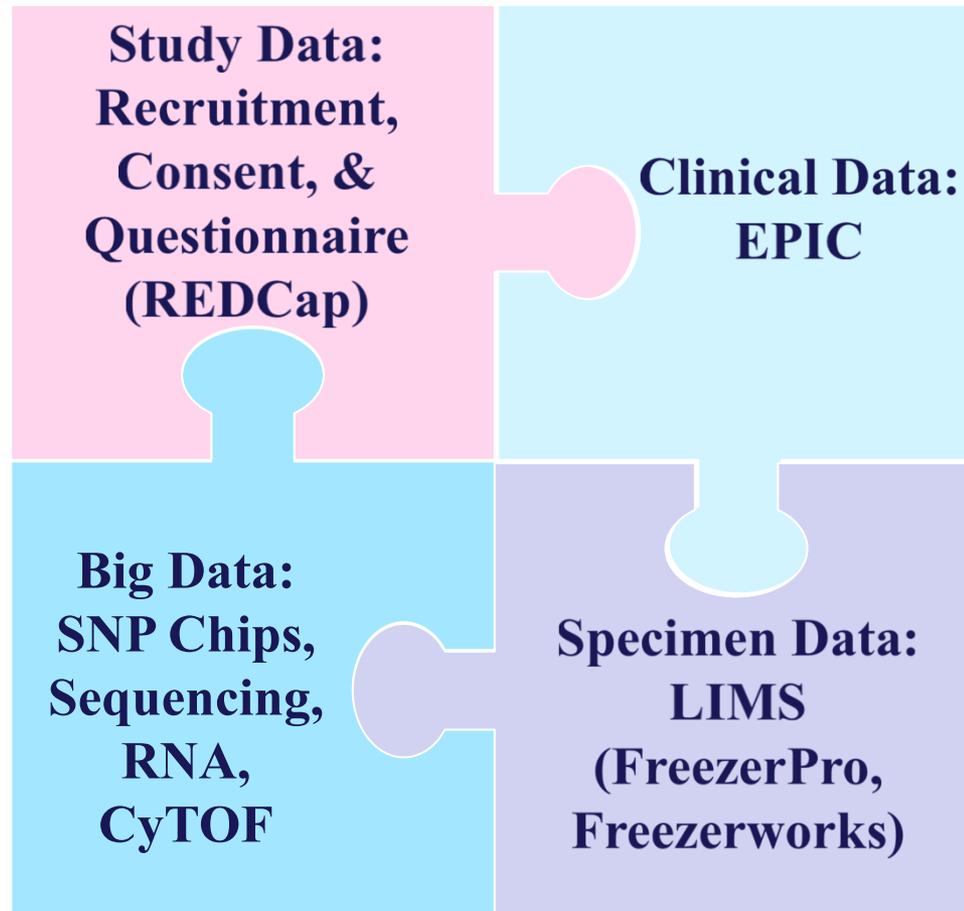
UC Location
Ulcerative colitis
demarcation

COMPASS
Recently diagnosed
individuals

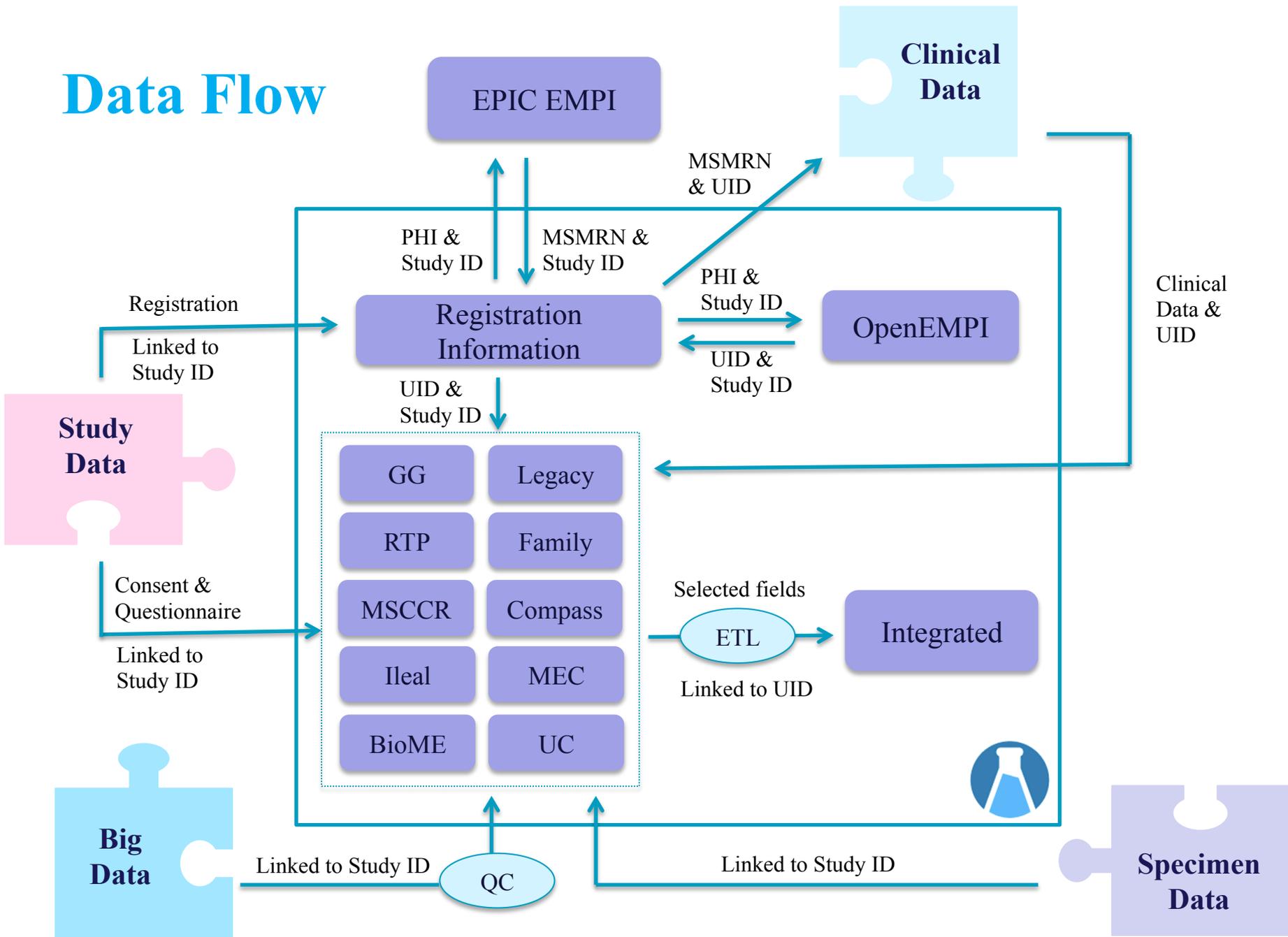
Family
Families with IBD
affected members

Ileal Resection
Post resection
recurrence

Elements of the integrated database

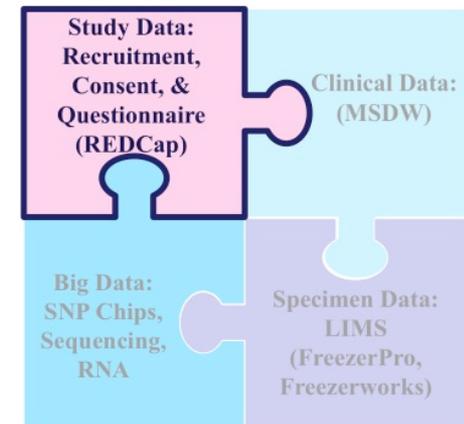


Data Flow



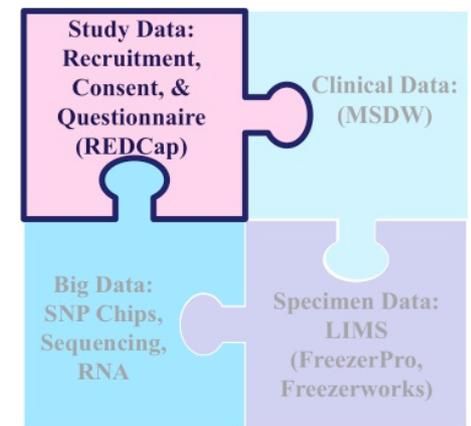
Recruitment

- ▶ 10 data sources - BioMe, Family, General Genetics, Ileal Resection, Legacy, Meconium, MSCCR, RTP, UC Location, and COMPASS
- ▶ Create unique identifiers across studies through Enterprise Master Patient Index (EMPI)
- ▶ Relevant fields for unique ID assignment: First Name, Middle Name, Last Name, Date of Birth, Sex, Street Address, City, State, Country, Zip Code, Telephone, Email, Race, MSMRN
- ▶ Recruitment template form on REDCap for easy integration with EMPI tool



Recruitment

- ▶ Total study participants: 7066
- ▶ Total *unique* study participants: 6457



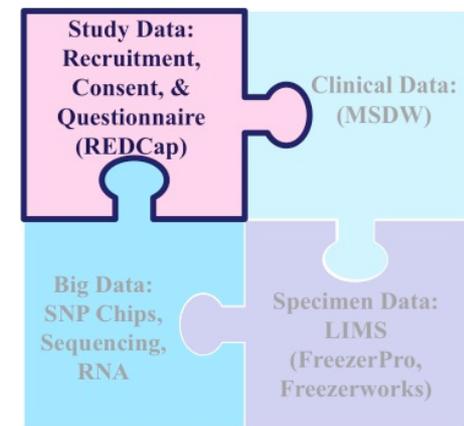
	Family	Legacy	RTP	Meconium	MSCCR	BioMe	GG	Ileal	UC
Family	628	53	25	1	29	6	8	0	0
Legacy	53	1780	10	4	201	7	27	6	5
RTP	25	10	375	1	7	0	10	0	1
Meconium	1	4	1	695	18	9	5	2	0
MSCCR	29	201	7	18	1950	138	56	15	4
BioMe	6	7	0	9	138	672	11	4	5
GG	8	27	10	5	56	11	757	8	5
Ileal	0	6	0	2	15	4	8	117	0
UC	0	5	1	0	4	5	5	0	92

Number of IBD Studies in which Research Participants are Registered

	1	2	3	4	5
Number of Participants	5931	451	68	6	1

Questionnaire

- ▶ Raw data import into study specific folders
- ▶ ETL (Extract Transform Load) of questionnaire fields identified as being of common interest for uniform querying across studies
- ▶ 154 fields to query questionnaire data
- ▶ Field topics include: diagnosis, age of diagnosis, disease activity, family history, medications, smoking, race, ethnicity, etc.
- ▶ REDCap questionnaire template forms for easy integration

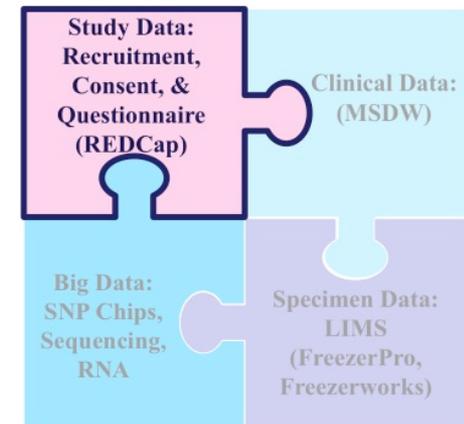


Number of Fields per Study

Family	92
Legacy	389
RTP	4414
UC	467
Meconium	1296
BioMe	155
GG	315
MSCCR	412
Ileal	516
COMPASS	182

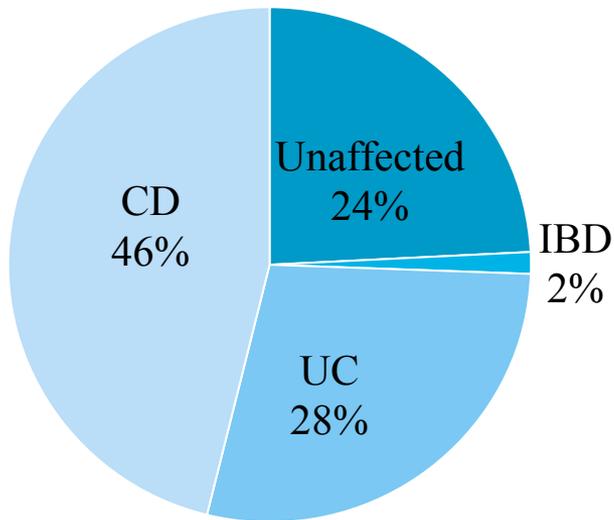
Questionnaire

► An overview:



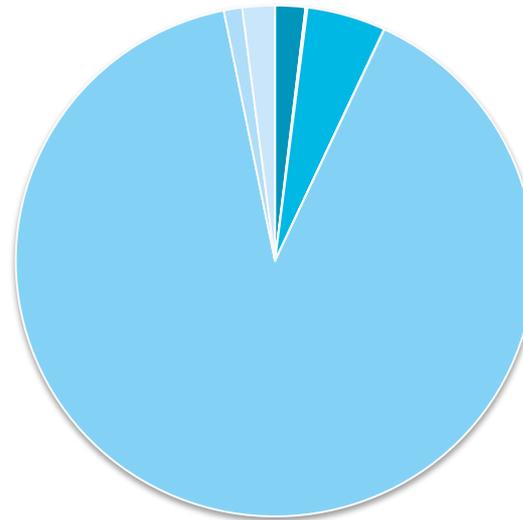
Diagnosis

N=6457



Race

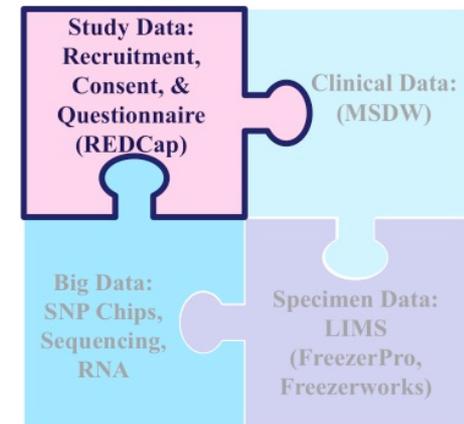
N=5679



- Asian 2%
- American Indian <1%
- Black 5%
- Pacific Islander <1%
- White 90%
- Other/mixed 1%
- Unknown 2%

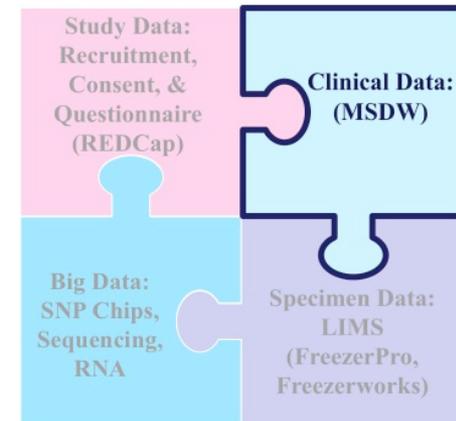
Consents

- ▶ The integrated database adheres to all conditions participants agreed to in their study consent through flags indicating sharing permissions.
- ▶ 25 different consents reviewed with varying language and questions.
- ▶ Strong need for omnibus consent/protocol and electronic consent/signature



Clinical Data

- ▶ For those consented to accessing EHR, clinical data will be pulled into iSITE
- ▶ 24 tables covering topics such as diagnosis, medication, immunization, encounters, orders, order results, family history, medical history, surgical history, vitals, social history, OB history
- ▶ Majority of participants consented to EHR access and about 3,000 have clinical encounters at Mount Sinai

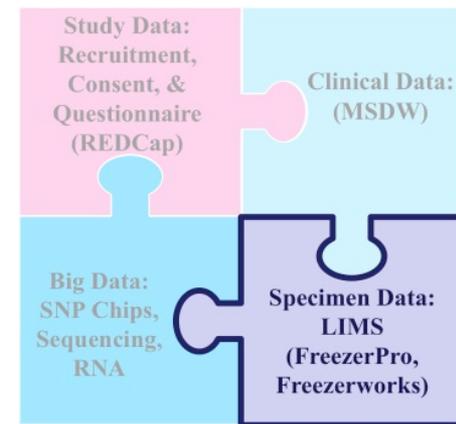


Specimen Data

- ▶ Wide variety of samples:

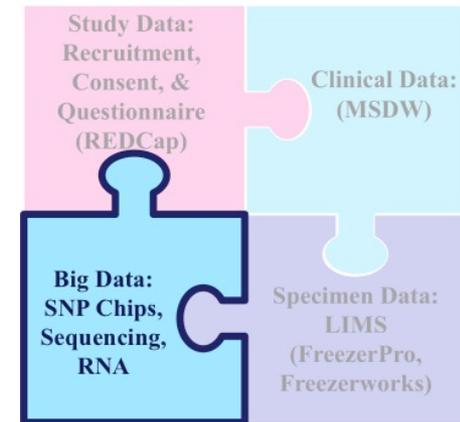
Sample Type	Count
Placenta	425
Amniotic membrane	81
Umbilical Cord	49
Stool	5594
Meconium	703
Saliva	913
Buccal Brush	206
Vaginal Swab	164
Umbilical cord blood	502
Breast Milk	176
Blood	1575
Intestinal tissue	1793
DNA	8088
Serum	3156
Plasma	5554
Total	28979

- ▶ Integrated system will provide number of samples per patient with description and freezer location



Big Data

- ▶ View which assays have been performed on samples from study participants
- ▶ Examples: Exome chip, Exome seq, GWAS MEGAEX, immunochip, RNAseq, CyTOF, Single Cell Sequencing
- ▶ Standardize QC and way of querying by variant
- ▶ Next projects in pipeline:
 - GSA chip array with custom content of AJ-enriched variants
 - Whole exome seq (~800 samples)



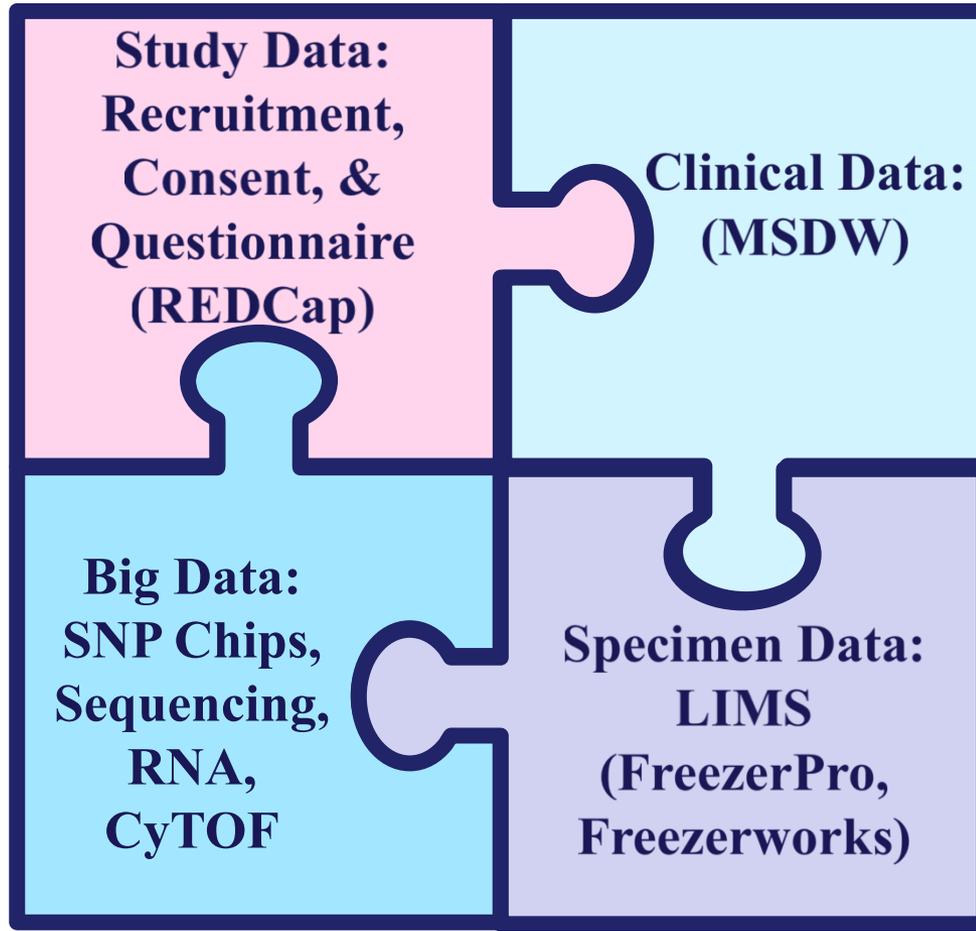
What we learned

- ▶ Type and amount of available resources
- ▶ Need for change in regulatory
 - Omnibus consent
 - Guidelines for protocol and workflow
 - Electronic consents
- ▶ Templates are necessary
 - REDCap templates for registration, consent and questionnaire
- ▶ High demand for collaboration platform

Future steps

- ▶ OpenEMPI
 - Patient lookup function
- ▶ Assay Integration
 - Genetic Data - chip and sequencing
 - CyTOF
 - Single Cell Sequencing
- ▶ Natural language processing (NLP) and Images
 - Pathology
 - Radiology
 - Discharge summary

Elements of the integrated database



Thank you

Judy Cho, MD

Inga Peter, PhD

Yashoda Sharma, PhD

Drew Helmus

Marla Dubinsky, MD

Bruce Sands, MD

Joana Torres, MD

Ryan Ungaro, MD

Steve Ellis

Tom Kaszemacher

Rajiv Nadukuru

Amanda Merkelson

Ernie Chen

Caroline Eisele

Maria Mucci

Jason Rogers

Amanda Hurley

Anabella Castillo, RN

Roman Kosoy, PhD

Mamta Giri

Erica Krasny

Elaine Martinez

Katrina Svoboda

Bernie Lee

Adam Rauch