iSITE: a consolidated view for efficient research

Nicole Villaverde
October 5, 2017
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

BioMe
IBD patients from MS IPM BioBank

COMPASS
Recently diagnosed individuals

UC Location
Ulcerative colitis demarcation

Family
Families with IBD affected members

Ileal Resection
Post resection recurrence

• Duplication
• Compatibility
• Accessibility
• Data Integration
Elements of the integrated database

Study Data: Recruitment, Consent, & Questionnaire (REDCap)

Big Data: SNP Chips, Sequencing, RNA, CyTOF

Clinical Data: EPIC

Specimen Data: LIMS (FreezerPro, Freezerworks)
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

<table>
<thead>
<tr>
<th>Family</th>
<th>Legacy</th>
<th>RTP</th>
<th>Meconium</th>
<th>MSCCR</th>
<th>BioMe</th>
<th>GG</th>
<th>Ileal</th>
<th>UC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>628</td>
<td>53</td>
<td>25</td>
<td>1</td>
<td>29</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Legacy</td>
<td>53</td>
<td>1780</td>
<td>10</td>
<td>4</td>
<td>201</td>
<td>7</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>RTP</td>
<td>25</td>
<td>10</td>
<td>375</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Meconium</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>695</td>
<td>18</td>
<td>9</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>MSCCR</td>
<td>29</td>
<td>201</td>
<td>7</td>
<td>18</td>
<td>1950</td>
<td>138</td>
<td>56</td>
<td>15</td>
</tr>
<tr>
<td>BioMe</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td>138</td>
<td>672</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>GG</td>
<td>8</td>
<td>27</td>
<td>10</td>
<td>5</td>
<td>56</td>
<td>11</td>
<td>757</td>
<td>8</td>
</tr>
<tr>
<td>Ileal</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>4</td>
<td>8</td>
<td>117</td>
</tr>
<tr>
<td>UC</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Number of IBD Studies in which Research Participants are Registered

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants</td>
<td>5931</td>
<td>451</td>
<td>68</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Number of Fields per Study</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>92</td>
</tr>
<tr>
<td>Legacy</td>
<td>389</td>
</tr>
<tr>
<td>RTP</td>
<td>4414</td>
</tr>
<tr>
<td>UC</td>
<td>467</td>
</tr>
<tr>
<td>Meconium</td>
<td>1296</td>
</tr>
<tr>
<td>BioMe</td>
<td>155</td>
</tr>
<tr>
<td>GG</td>
<td>315</td>
</tr>
<tr>
<td>MSCCR</td>
<td>412</td>
</tr>
<tr>
<td>Ileal</td>
<td>516</td>
</tr>
<tr>
<td>COMPASS</td>
<td>182</td>
</tr>
</tbody>
</table>
Questionnaire

An overview:

**Diagnosis**
N=6457
- CD 46%
- UC 28%
- IBD 2%
- Unaffected 24%

**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:

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

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

Study Data: Recruitment, Consent, & Questionnaire (REDCap)

Clinical Data: (MSDW)

Big Data: SNP Chips, Sequencing, RNA, CyTOF

Specimen Data: LIMS (FreezerPro, Freezerworks)
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