Creating clinical and 'omics information commons using i2b2 and tranSMART

Paul Avillach, MD, PhD
Harvard Medical School
HEGP Hospital, Paris – Paris Descartes University
INSERM UMRS 872 eq 22
Erasmus MC University
Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease
Report from National academy of science, USA, 2011
HEGP background

Opening: July 2000-

Hôpital Laennec (1634)

Hôpital Boucicaut

Hôpital Broussais
### Allergie
- **Oui**: [ ]
- **Non**: [ ]

### Allergie à l'iode
- **Oui**: [ ]
- **Non**: [ ]

### Données cliniques

#### Diabète insulinodépendant
- **Oui**: [ ]
- **Non**: [ ]

#### Diabète non insulinodépendant
- **Oui**: [ ]
- **Non**: [ ]

#### Insuffisance rénale créatinine > 120µmol/L
- **Oui**: [ ]
- **Non**: [ ]

#### Corps étrangers métalliques/clip
- **Oui**: [ ]
- **Non**: [ ]

### Observations

#### Commentaire

### Traitement suivi au long cours

#### Diabète (Biguanide)
- **Oui**: [ ]
- **Non**: [ ]

#### Anticoagulant
- **Oui**: [ ]
- **Non**: [ ]

#### Chimiothérapie/Radiothérapie
- **Oui**: [ ]
- **Non**: [ ]

### Transplantation
- **Oui**: [ ]
- **Non**: [ ]
HEGP BDW

EHR/BDW integration

Evaluation/Research environment

ETL suite (Talend Open Studio)

Production environment

DxCare

EHR: Operational Database (ODS)

EHR: Mirrored Database

Biomedical Data Warehouse (BDW)

External Databases

R

i2b2/tranSMART tools

Business Object

IBM Ilog Rules

Data Analysis

Data Mining

Real time requests

Pr Patrice Degoulet

Pr Anita Burgun CIO

Boston Children's Hospital

HARVARD MEDICAL SCHOOL

TEACHING HOSPITAL
<table>
<thead>
<tr>
<th>Type of data</th>
<th>H</th>
<th>OV</th>
<th>Start date</th>
<th># unique patients</th>
<th># values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic (age, sex, Hospital vital status)</td>
<td>X</td>
<td>X</td>
<td>1971</td>
<td>606 524</td>
<td></td>
</tr>
<tr>
<td>Vital signs (temperature, blood pressure, weight, ...)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>141 164</td>
<td>14 213 951</td>
</tr>
<tr>
<td>Diagnostic codes (DRG ICD10)</td>
<td>X</td>
<td></td>
<td>1995</td>
<td>305 369</td>
<td>2 626 792</td>
</tr>
<tr>
<td>Procedures (French CCAM codes)</td>
<td>X</td>
<td></td>
<td>2004</td>
<td>241 482</td>
<td>3 200 482</td>
</tr>
<tr>
<td>Clinical data (DxCare questionnaires)</td>
<td>X</td>
<td>X</td>
<td>1971</td>
<td>391 218</td>
<td>46 506 217</td>
</tr>
<tr>
<td>Free text reports*: Hospitalization, Surgery, consultations, ...</td>
<td>X</td>
<td>X</td>
<td>2004</td>
<td>289 614</td>
<td>1 961 985</td>
</tr>
<tr>
<td>Free text reports**: Imaging and pathology</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td></td>
<td>1 000 000</td>
</tr>
<tr>
<td>Pathology codes (ADICAP)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>73 173</td>
<td></td>
</tr>
<tr>
<td>Biology results (without antiograms)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>338 068</td>
<td>88 607 301</td>
</tr>
<tr>
<td>Antibiograms</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>39 040</td>
<td>4 058 842</td>
</tr>
<tr>
<td>Drug prescription (without Chemotherapy)</td>
<td>X</td>
<td>X</td>
<td>1988</td>
<td>88 567</td>
<td>2 612 742</td>
</tr>
<tr>
<td>Validation of Drug prescription by pharmacists</td>
<td>X</td>
<td></td>
<td>2002</td>
<td>67 151</td>
<td>1 691 137</td>
</tr>
</tbody>
</table>
HEGP CDW use

i2b2 CDW queries
• 188 MD + Pharm trained

• IRB
  
  Creation of an HEGP research ethical committee linked to the regional IRB
HEGP CDW

i2b2 CDW queries (Jan. 2011-april 2013)

Access rights

• Level 1 studies: aggregated data (e.g. potential trial recruitment)
  ➢ Free access for all HEGP health professionals
  ➢ 1 978 requests

• Level 2 and 3 studies: access to patient level data
  ➢ Structured written project
  ➢ Validation by the HEGP ethical/research committee
  ➢ Transmission to the regional IRB committee
  ➢ Level 2: anonymized patient data
  ➢ Level 3: de-anonymized patient data

➢ IRB approval for 32 projects
HEGP

Health care
Health Information System

Clinical Research

Browser tools available for Researchers

ETL once a week

i2b2

CDW

‘omics

Analysis tools

Structured data from research studies

‘omics data

DRG
EHR forms
EHR reports
Biology
Imaging
Pathology
Rx

ETL

Boston Children’s Hospital
Until every child is well

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL
• Integrated platform to support translational research
• Initiated by Johnson & Johnson et Recombinant 5 years ago
• Open-source since January, 2012
• Installed at HEGP since May, 2012

• Today, driven and maintained by the tranSMART Foundation & community

http://transmartfoundation.org
Objectives:

1. **Integration** of clinical, biological and ‘omics data in one place – hypothesis free –
2. Generation of **hypothesis** by Clinicians / Researchers
“Omics” data integration

http://www.transmartproject.org
Analysis of PTEN, BRAF, and EGFR Status in Determining Benefit From Cetuximab Therapy in Wild-Type KRAS Metastatic Colon Cancer

Pierre Laurent-Puig, Anne Cayre, Gilles Manceau, Emmanuel Buc, Jean-Baptiste Bachet, Thierry Lecomte, Philippe Rougier, Astrid Lievre, Bruno Landi, Valérie Boige, Michel Ducreux, Marc Ychou, Frédéric Bibeau, Olivier Bouché, Julia Reid, Steven Stone, and Frédérique Penault-Llorca
<p>| N\textregistered\textsuperscript{e} | Inter | Age | ex | BILANKRAS | Mutat | BRAF | NRAS | MUT | MLN | Toxicit\textsuperscript{e} | Meilleure | Progression | Dur\textsuperscript{e} reg | N&amp;&amp;C\textsuperscript{e} | N&amp;&amp;l\textsuperscript{e} survie | OMS | EGFR copy | Score de HIRSCH | Mutation | PTEN:CYTO | PTEN:MB | PTEN:NX |
| 1 | 71 | M | NM | NM | NM | 1 | 2 | 1 | 1 | 1 | 58 | 14 | oui | 2.5 | 1 | 1 | 10 | pos | NM | 80 | 20 | 20 |
| 5 | 71 | M | NM | NM | NM | 6 | 2 | 2 | 1 | 4 | 65 | oui | 21.6 | 1 | 3 | neg | NM | 200 | 10 | 140 |
| 6 | 44 | F | NM | NM | NM | 4 | 2 | 2 | 1 | 67 | 14 | oui | 48.13 | 0 | 3 | neg | M | 60 | 30 | 0 |
| 7 | 72 | M | M | NM | NM | 4 | 1 | 2 | 0 | 48 | non | 40.4 | 1 | 2.1 | neg | NM | 160 | 0 | 80 |
| 8 | 48 | M | NM | NM | NM | 6 | 2 | 2 | 1 | 34 | 43 | oui | 13.87 | 1 | 11 | pos | NM | 0 | 0 | 60 |
| 9 | 55 | F | NM | NM | NM | 3 | 1 | 2 | 1 | 32 | oui | 15.07 | 2 | 3.4 | neg | NM | 10 | 40 | 0 |
| 10 | 64 | F | NM | NM | NM | 2 | 1 | 2 | 0 | 17 | 1 | oui | 20.03 | 0 | 2.5 | neg | NM | 260 | 20 | 150 |
| 11 | 62 | M | NM | NM | NM | 3 | 2 | 2 | 1 | 52 | oui | 24.23 | 1 | 2.8 | neg | NM | 200 | 0 | 0 |
| 12 | 50 | M | NM | NM | NM | 3 | 2 | 3 | 1 | 14 | 71 | oui | 9.5 | 0 | 2.9 | neg | NM | 130 | 0 | 100 |
| 13 | 54 | M | M | NM | NM | 2 | 2 | 3 | 1 | 20 | oui | 6.93 | 2 | neg | NM | 230 | 0 | 110 |
| 14 | 73 | F | M | NM | NM | 3 | 1 | 3 | 1 | 19.29 | oui | 20.03 | 0 | neg | M | 50 | 0 | 0 |
| 15 | 71 | M | NM | NM | NM | 2 | 1 | 3 | 1 | 16 | oui | 15.47 | 0 | neg | M | 50 | 0 | 0 |
| 16 | 53 | F | M | NM | NM | 3 | 2 | 3 | 1 | 20 | oui | 10.73 | 1 | neg | NM | 10 | 30 | 0 |
| 17 | 78 | M | M | NM | NM | 2 | 1 | 4 | 1 | 11 | 14 | oui | 16 | 33 | 0 | 2.4 | neg | M | 200 | 0 | 50 |
| 18 | 51 | F | M | NM | NM | 2 | 1 | 4 | 1 | 4.43 | oui | 1.3 | 3 | neg | M | 80 | 0 | 120 |
| 19 | 75 | F | NM | NM | NM | 3 | 0 | 4 | 1 | 7.86 | oui | 6 | 0 | 2.3 | neg | M | 60 | 0 | 80 |
| 20 | 69 | M | M | NM | NM | 2 | 2 | 4 | 1 | 5.57 | oui | 10 | 7 | 1 | neg | NM | 10 | 0 | 40 |
| 21 | 72 | M | M | NM | NM | 4 | 0 | 4 | 1 | 6.14 | oui | 2.67 | 1 | neg | M | 190 | 0 | 50 |
| 22 | 61 | F | NM | NM | NM | 2 | 2 | 4 | 1 | 12 | oui | 10 | 33 | 1 | 2.3 | neg | M | 75 | 50 | 10 |
| 23 | 53 | M | M | NM | NM | 5 | 2 | 4 | 1 | 9 | oui | 9.8 | 1 | neg | NM | 0 | 0 | 0 |
| 24 | 59 | M | M | NM | NM | 2 | 1 | 4 | 1 | 8.57 | oui | 3.57 | 0 | neg | NM | 240 | 0 | 120 |
| 25 | 75 | M | M | NM | NM | 2 | 1 | 4 | 1 | 8 | oui | 6.4 | 1 | neg | NM | 0 | 0 | 120 |
| 26 | 58 | F | M | NM | NM | 6 | 2 | 4 | 1 | 8 | oui | 8.93 | 0 | pos | NM | 10 | 50 | 0 |
| 27 | 47 | F | M | NM | NM | 3 | 3 | 4 | 1 | 8 | oui | 5.65 | 0 | 3.3 | neg | NM | 10 | 0 | 0 |
| 28 | 60 | F | M | NM | NM | 3 | 1 | 4 | 1 | 8 | oui | 3.77 | 1 | neg | NM | 5 | 0 | 0 |
| 29 | 58 | F | NM | NM | NM | 4 | 2 | 3 | 1 | 17.57 | oui | 7.2 | 2 | 3 | neg | NM | 40 | 0 | 60 |
| 30 | 58 | M | NM | NM | NM | 2 | 3 | 2 | 0 | 33 | non | 26.97 | 1 | 1 | NM | 0 | 0 | 0 |
| 33 | 67 | M | M | NM | NM | 3 | 2 | 3 | 1 | 29.71 | oui | 9 | 2 | NM | 100 | 0 | 30 |
| 34 | 68 | M | NM | NM | NM | 2 | 2 | 2 | 1 | 33.14 | oui | 72.17 | 1 | NM | 0 | 0 | 0 |
| 35 | 59 | F | NM | NM | NM | 2 | 1 | 2 | 0 | 31.9 | oui | 11.83 | 1 | NM | 0 | 0 | 0 |
| 36 | 61 | M | M | NM | NM | 4 | 0 | 4 | 1 | 3 | oui | 2.8 | 0 | NM | 0 | 0 | 0 |
| 37 | 57 | M | M | NM | NM | 3 | 1 | 3 | 1 | 17.43 | oui | 5.1 | 1 | NM | 0 | 0 | 0 |
| 38 | 71 | F | M | NM | NM | 3 | 0 | 3 | 1 | 21.57 | oui | 7.83 | 1 | NM | 0 | 0 | 0 |
| 39 | 63 | M | M | NM | NM | 4 | 0 | 4 | 1 | 6.14 | oui | 5.57 | 1 | NM | 0 | 0 | 0 |
| 42 | 60 | F | M | NM | NM | 2 | 1 | 4 | 1 | 12 | oui | 5.1 | 1 | neg | NM | 150 | 0 | 200 |
| 43 | 60 | F | M | NM | NM | 2 | 2 | 3 | 0 | 24 | oui | 16.77 | 0 | neg | NM | 40 | 0 | 160 |
| 44 | 59 | F | NM | NM | NM | 2 | 2 | 3 | 1 | 34 | oui | 7.93 | 1 | 2.4 | neg | M | 90 | 0 | 120 |</p>
<table>
<thead>
<tr>
<th>Filename</th>
<th>Category Code</th>
<th>Column Number</th>
<th>Data Label</th>
<th>Data Label Source</th>
<th>Controlled Vocab Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Demographics</td>
<td>1</td>
<td>SUBJ_ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Demographics</td>
<td>2</td>
<td>AGE</td>
<td></td>
<td>424144002</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Demographics</td>
<td>3</td>
<td>SEX</td>
<td></td>
<td>263495000</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Mutation_Detection</td>
<td>4</td>
<td>KRAS Mutation</td>
<td></td>
<td>190070</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Mutation_Detection</td>
<td>5</td>
<td>BRAF Mutation</td>
<td></td>
<td>164757</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Mutation_Detection</td>
<td>6</td>
<td>NRAS Mutation</td>
<td></td>
<td>164790</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Treatment+Chemotherapy</td>
<td>7</td>
<td>Number of lines</td>
<td></td>
<td>399042005</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td></td>
<td>8</td>
<td>OMIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td></td>
<td>9</td>
<td>OMIT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Outcome</td>
<td>10</td>
<td>Progression</td>
<td></td>
<td>419835002</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Outcome</td>
<td>11</td>
<td>Duration of Response</td>
<td></td>
<td>445397003</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Outcome</td>
<td>12</td>
<td>Death</td>
<td></td>
<td>419620001</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Outcome</td>
<td>13</td>
<td>Overall Survival</td>
<td></td>
<td>445320007</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Clinical_Data+Outcome</td>
<td>14</td>
<td>OMS Score</td>
<td></td>
<td>373802001</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>15</td>
<td>EGFR Copy Number</td>
<td></td>
<td>5006</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>16</td>
<td>HIRSC Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>17</td>
<td>PIK3CA Mutation</td>
<td></td>
<td>171834</td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGP0001_data.csv</td>
<td>Biomarker_Data+Non_Omics+Immunological</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• R module in tranSMART

• Published figure in JCO

HEGP: Canuel V, Avillach P
Phenotypic augmentation
Phenotypic augmentation
tranSMART + i2b2 = Phenotypic augmentation

Type of Data

<table>
<thead>
<tr>
<th>Vital Status</th>
<th>Clinical Data</th>
<th>Biology</th>
<th>Pathology</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head and Neck Cancer Study</td>
<td>Cancer recurrence</td>
<td>Study specific samples</td>
<td>Re-intervention</td>
<td></td>
</tr>
</tbody>
</table>

HEGP: Joubert F, Avillach P

Electronic Health Record
Head and Neck Study
Badoual Cancer Res. 2013

HEGP: Joubert F, Avillach P

p = 0.04
p = 0.03

AVG: 10 months of additional follow up with i2b2 data

HEGP: Joubert F, Avillach P
PheWAS study on i2b2/tranSMART


Phenome-wide association studies on a quantitative trait: Application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics.

Genome Wide Association Study
(1 Phenotype compared to ALL SNPs)

cases
(ex: systemic sclerosis)
cases DNA

controls
controls DNA

-log(P-value) vs. chromosomes

- HLA region (ch 6)

compare ALL SNPs to find differences between cases and controls
Genome Wide Association Study
(1 Phenotype compared to ALL SNPs)

- cases (ex: systemic sclerosis)
- controls

- cases DNA
- controls DNA

- HLA region (ch 6)

- \(-\log(P\text{-value})\) chromosomes

compare ALL SNPs to find differences between cases and controls

Phenome Wide Association Study
(1 SNP compared to ALL Phenotypes)

- allele G patients group
- allele A patients group

- allele G patients phenotype
- allele A patients phenotype

- I21 (myocardial infarction)

- \(-\log(P\text{-value})\) ICD-10 Codes

compare ALL DIAGNOSIS to find differences between cases and controls
Workflow

1. Sélection du trait
2. Extraction données
   - Agregations codes
   - Projections sur ICD9-CM
   - Restrictions temporelles (extraction directe)
3. Constitution groupes
   - Discrétisation: - globale - fréquentielle
   - Restrictions temporelles (extraction textuelle)
4. Analyse
5. Graphs
6. Hypothèses

\( f \): fonctions R réutilisables
Methodes: Selection of trait: enzymatic Activity TPMT

Thiopurine (DRUG) → 6-TIMP (active metabolite) → elimination

increased toxicity

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Low activity</th>
<th>Intermediate activity</th>
<th>Normal Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiopurine dose</td>
<td>10% dose</td>
<td>30 – 70% dose</td>
<td>100% dose</td>
</tr>
</tbody>
</table>

FDA & EMA recommendations

?
TPMT

Quantitative trait

other activity patients
ICD codes / Biological test results

Very High activity patients
ICD Codes / Biological test results

log(P-value)

ICD-10 Codes

log(P-value)

Biological tests
Very High TPMT activity VS others

ICD9-CM Mapping

Secondary hypertension
Angina pectoris
Diabetes Mellitus
Iron deficiency anemias

HEGP: Neuraz A,..., Avillach P
Acknowledgments

HEGP

Informatics & Public Health Dept:
• Eric Zapletal,
• Vincent Canuel,
• Antoine Neuraz,
• Fabien Joubert

Prof Patrice Degoulet, past CIO
Prof Anita Burgun, CIO

Contact:

paul.avillach@egp.aphp.fr
Anita.burgun@egp.aphp.fr

www.i2b2.org
www.transmartproject.org
European Medical Information Framework
EMIF project: European Medical Information Framework

- 58 partners
- 58 Million euros
- Started 1st Jan 2013
- 5 year project
- PI: Bart Vannieuwenhuyse (Janssen) & Prof. Simon Lovestone (KCL)
- PI Platform: Prof. Johan van der Lei
- Detect new biomarkers:
  - predisposition Alzheimer's disease
  - Metabolic complication diabetes
EMIF project: European Medical Information Framework

3 projects, 3 datasources:

• EHR derived patient data: **52 millions patient records**: observational studies (**Jerboa tool**)

• Cohort clinical + ‘omics data from: **tranSMART option**
  – AD
  – Metabolic:
    ~10 000 patients

• **WP11**: Semantic harmonization

• Same concepts across all database: ETL process
  Jerboa & tranSMART
Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease
Report from National academy of science, USA, 2011
Autism cohorts

- Gene-Pheno studies – Lou Kunkel
  - Pre AC
  - AC
  - SSC
- SSC
- AC
- AGRE
Autism information commons

From SSC National cohort to BCH i2b2

- SSC Cohort - Clinical data n=2725 patients
- SSC Cohort - SNPs = n = 2147 + Families
- BCH EHR data (Clinical Notes (CTAKES), Diagnosis ICD9) n=420 patients
- BCH EHR data (Vital Signs, Lab tests) n= 360 patients
- BCH EHR data (Prodecures, Medications, Allergy) n= 220 patients
- BCH EHR data (Insurance Payors) n = 375 patients
- Gene Expression = 375
- Exoms / RNA seq ?
- Environment / Exposure from EHR?
- Biobank data (Aliquots lefts)
- Patient consent (Recontact? Willingness to share data? etc..)
## ASDs  SSC Cohort to i2b2 BCH

<table>
<thead>
<tr>
<th>Data Type</th>
<th># Patients</th>
<th># Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vital:Weight</td>
<td>367</td>
<td>4 727</td>
</tr>
<tr>
<td>Vital:Height</td>
<td>359</td>
<td>7 220</td>
</tr>
<tr>
<td>SER</td>
<td>421</td>
<td>17 915</td>
</tr>
<tr>
<td>Radiology Procedures</td>
<td>174</td>
<td>1 142</td>
</tr>
<tr>
<td>Procedures (CPT)</td>
<td>209</td>
<td>1 470</td>
</tr>
<tr>
<td>Medications (Prescription)</td>
<td>212</td>
<td>3 171</td>
</tr>
<tr>
<td>Pathology Reports</td>
<td>220</td>
<td>2 255</td>
</tr>
<tr>
<td>Clinical Notes</td>
<td>419</td>
<td>33 978</td>
</tr>
<tr>
<td>Medications</td>
<td>189</td>
<td>19 321</td>
</tr>
<tr>
<td>Lab tests</td>
<td>360</td>
<td>72 433</td>
</tr>
<tr>
<td>Insurance Payors</td>
<td>375</td>
<td>8 869</td>
</tr>
<tr>
<td>Medications (History)</td>
<td>217</td>
<td>835</td>
</tr>
<tr>
<td>Diagnosis (ICD9)</td>
<td>421</td>
<td>26 323</td>
</tr>
<tr>
<td>Clinic</td>
<td>421</td>
<td>14 169</td>
</tr>
<tr>
<td>AMED</td>
<td>164</td>
<td>7 142</td>
</tr>
<tr>
<td>Allergy</td>
<td>246</td>
<td>583</td>
</tr>
</tbody>
</table>

**AVG = 298**  **TOTAL = 221 553**
Epilepsy = YES

- age < 5
- WES: homozygous
- WES: OR4F5
- WES: nonsynonymous SNV
SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies

Andrew J. McMurry¹,²,³,⁴, Shawn N. Murphy³,⁵,⁶, Douglas MacFadden¹, Griffin Weber²,⁷, William W. Simons¹, John Orecchia⁸, Jonathan Bickel²,⁹, Nich Wattanasin⁵, Clint Gilbert¹, Philip Trevvett¹, Susanne Churchill³,⁵, Isaac S. Kohane¹,²,³

---

![SHRINE diagram and table]

**Hospital A** 32±3 Patients
**Hospital B** 264±3 Patients
**Hospital C** 815±3 Patients
**Hospital D** 223±3 Patients
**Aggregated** 1134±12 Patients
Informatics & Public Health Dept:

- Eric Zapletal
- Vincent Canuel
- Antoine Neuraz
- Fabien Joubert
- Anita Burgun
- Patrice Degoulet

Contact:

paul.avillach@egp.aphp.fr

www.i2b2.org

www.transmartproject.org

www.recomdata.com
Autism Cohort

Division of Developmental Medicine
Leonard Rappaport, MD, MS

CH Genomic program
Ellen Hanson, PhD
Ingrid Holm, MD, MPH
Stephanie Brewster, MS, CGC
Joanna Reinwald, MS, GC
Franck Jackson

The Research Connection
Wendy Wolf, PhD
Sarah Savage, MS, CGC
Catherine Clinton, MS, CGC
Tram Tran

Business Intelligence and Clinical Research Informatics
Jonathan Bickel, MD, MS
Mohamad Daniar
Nandan Patibandla
Rick Agrella
Paul OBryne
Lynne N. Alley
Gina Bianco

MassGeneral Hospital for Children
Lurie Center for Autism
Timothy Yu, MD, PhD

Mass General Hospital for Children
Lurie Center for Autism
Timothy Yu, MD, PhD

CBMI
CH Informatics program
Alexa T. McCray, PhD
Dennis Wall, PhD
Nathan Palmer, PhD
Sek Won Kong, MD
Ally Eran, PhD
Finale Doshi-Velez, PhD

Clinical NLP
Guergana Savova, PhD - PI
Chen Lin
Dmitriy Dligach, PhD
Pei Chen
Sameer Pradhan, PhD
Sean Finan
Timothy Miller, PhD

i2b2 / Partners
Shawn Murphy, MD, PhD
Lori Phillips, MS
Michael Mendis

Principal Investigators
Isaac Kohane, MD, PhD
Louis Kunkel, PhD
David Margulies, MD
Paul Avillach, MD, PhD