Sciences Economiques & Sociales de la Santé
& Traitement de l'Information Médicale

www.sesstim-orspaca.org

**Paul AVILLACH**
Assistant Professor
Harvard Medical School

Boston - USA

# *Plateformes de recherche translationnelle intégrant des données cliniques et omiques*

janvier 2015

Cliquez ici pour voir l'intégralité des ressources associées à ce document

# Toward Precision Medicine: Building a patient centric information commons on common and rare diseases with I2b2/tranSMART

## Application to Autism and Phelan McDermid Syndrome

**Paul Avillach, MD, PhD**

Assistant Professor - Harvard Medical School

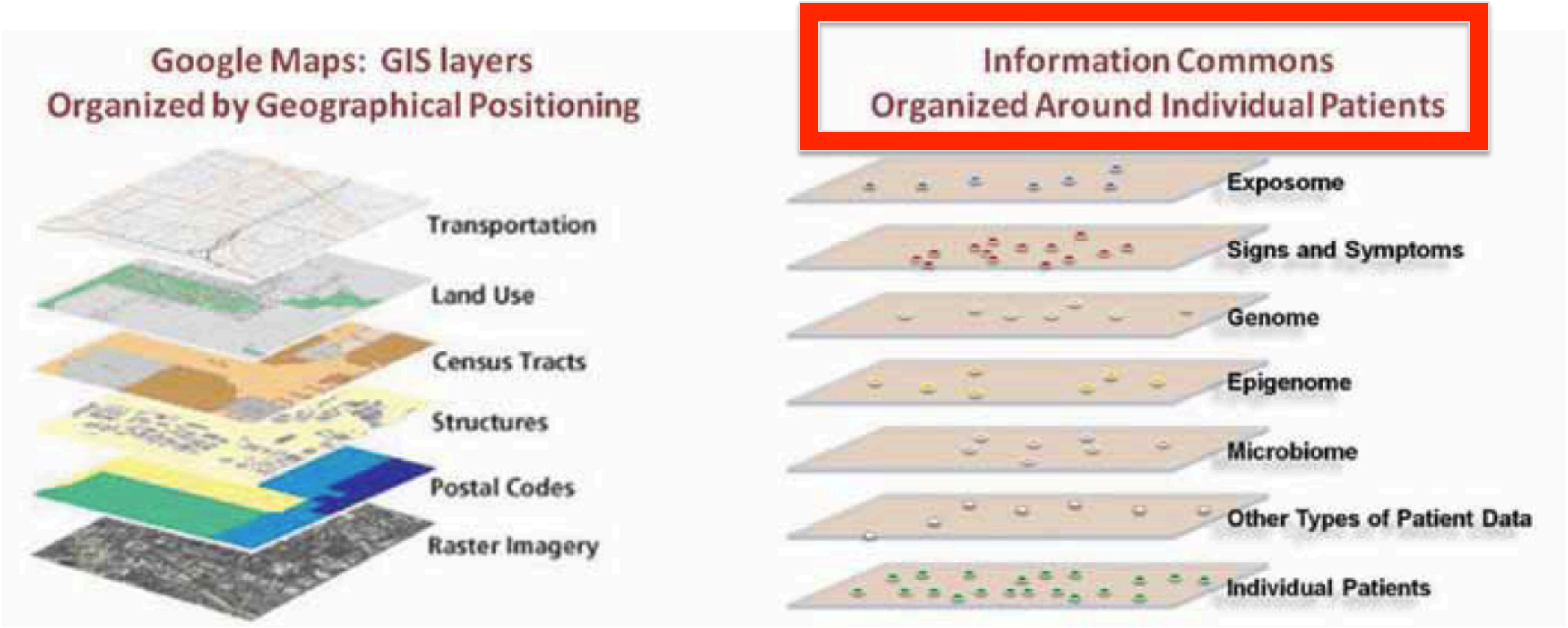
Center of Biomedical Informatics

Research Connection – Boston Children Hospital

Medical-informatics - Erasmus MC University, Rotterdam, The Netherlands
INSERM UMRS 872 eq 22, Paris, France

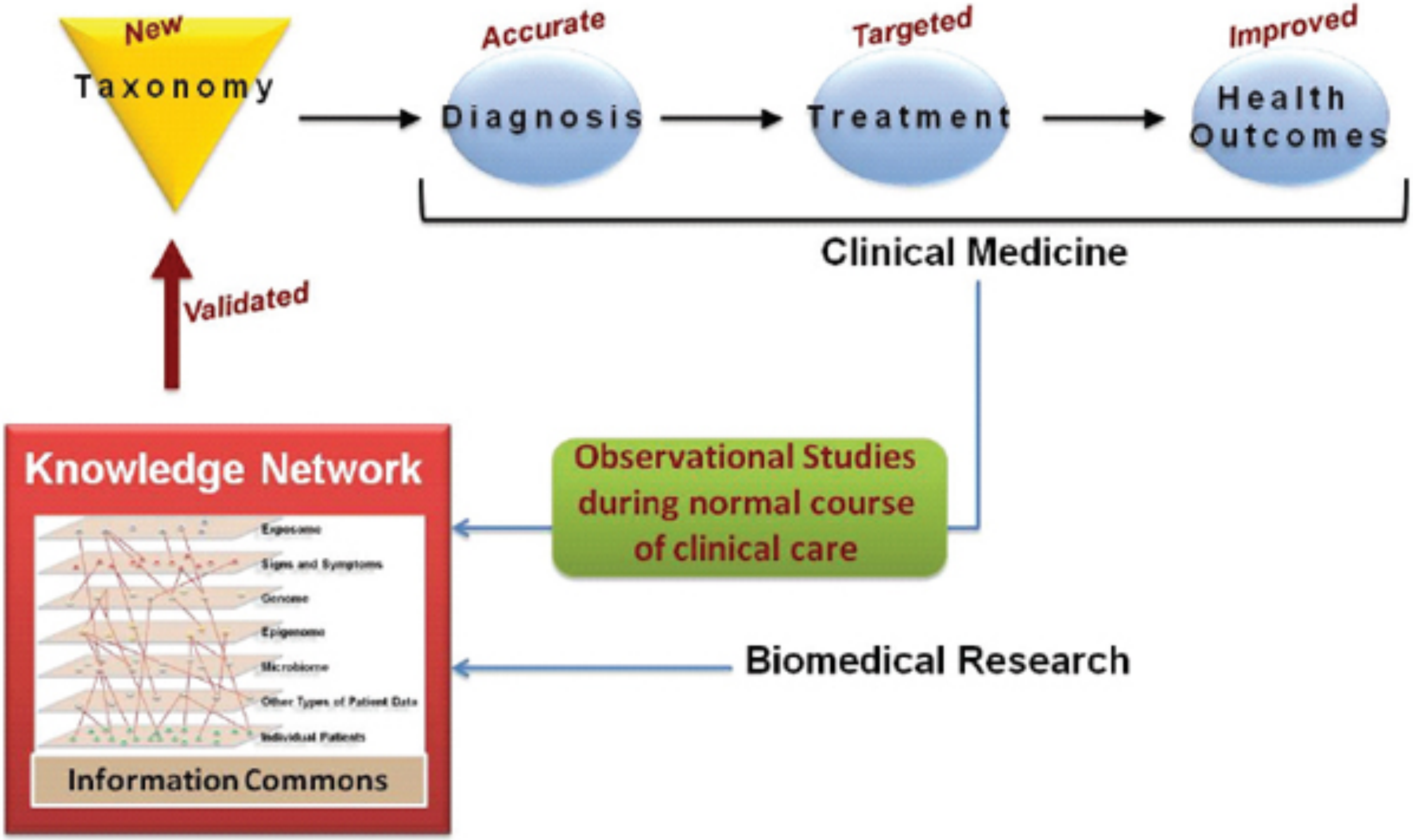
# Disclosure:

- Consultant for CLARITAS GENOMICS


Google Maps: GIS layers
Organized by Geographical Positioning
Transportation
Land Use
Census Tracts
Structures
Postal Codes
Raster Imagery
Information Commons
Organized Around Individual Patients
Exposome
Signs and Symptoms
Genome
Epigenome
Microbiome
Other Types of Patient Data
Individual Patients

Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

Report from National academy of science, USA, 2011



New Taxonomy → Accurate Diagnosis → Targeted Treatment → Improved Health Outcomes

Clinical Medicine

Validated

Knowledge Network

Observational Studies during normal course of clinical care

Biomedical Research

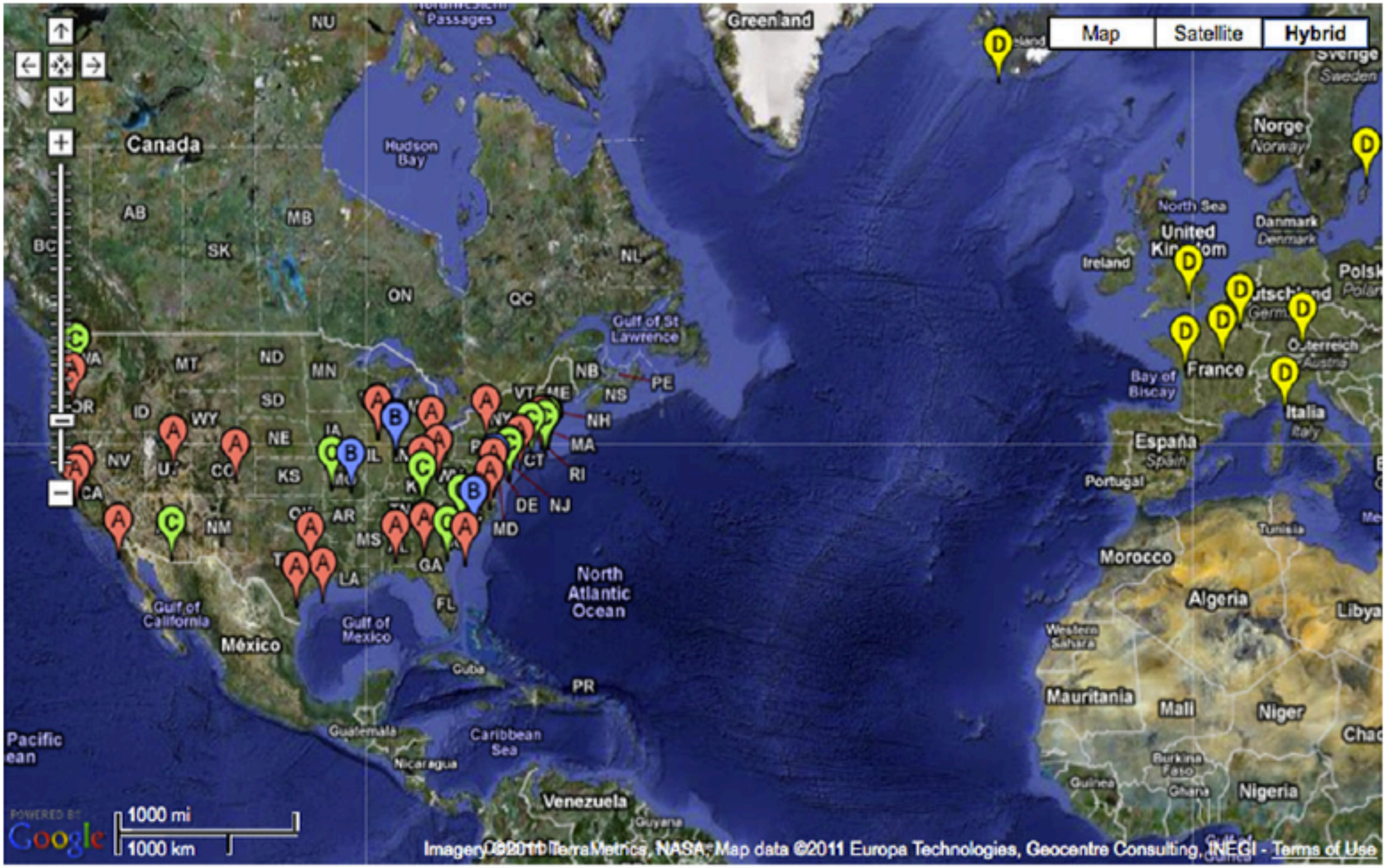
Information Commons

Exposome
Signs and Symptoms
Genome
Epigenome
Microbiome
Other Types of Patient Data
Individual Patients

Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

Report from National academy of science, USA, 2011 i2b2

# PI: Isaac Kohane – Harvard - CBMI


Map
Satellite
Hybrid
Greenland
NU
Canada
Hudson Bay
AB
BC
SK
MB
ON
QC
NL
Gulf of St Lawrence
NB
NS
PE
WA
OR
ID
MT
WY
ND
SD
MN
NE
IA
IL
KS
MO
NV
UT
CO
CA
NM
AR
MS
LA
GA
FL
VT
ME
NH
NY
MA
CT
RI
PA
NJ
DE
MD
México
Gulf of California
Gulf of Mexico
Cuba
PR
Caribbean Sea
Guatemala
Nicaragua
Venezuela
Guyana
Pacific Ocean
North Atlantic Ocean
Sverige
Sweden
Norge
Norway
North Sea
Danmark
Denmark
United Kingdom
Ireland
Polska
Poland
Deutschland
Germany
Österreich
Austria
France
Bay of Biscay
Italia
Italy
España
Spain
Portugal
Tunisia
Morocco
Algeria
Libya
Western Sahara
Mauritania
Mali
Niger
Chad
Burkina Faso
Guinea
Ghana
Nigeria
A
B
C
D
POWERED BY Google
1000 mi
1000 km
Imagery ©2011 TerraMetrics, NASA, Map data ©2011 Europa Technologies, Geocentre Consulting, INEGI - Terms of Use

Boston Children's Hospital
Until every child is well™

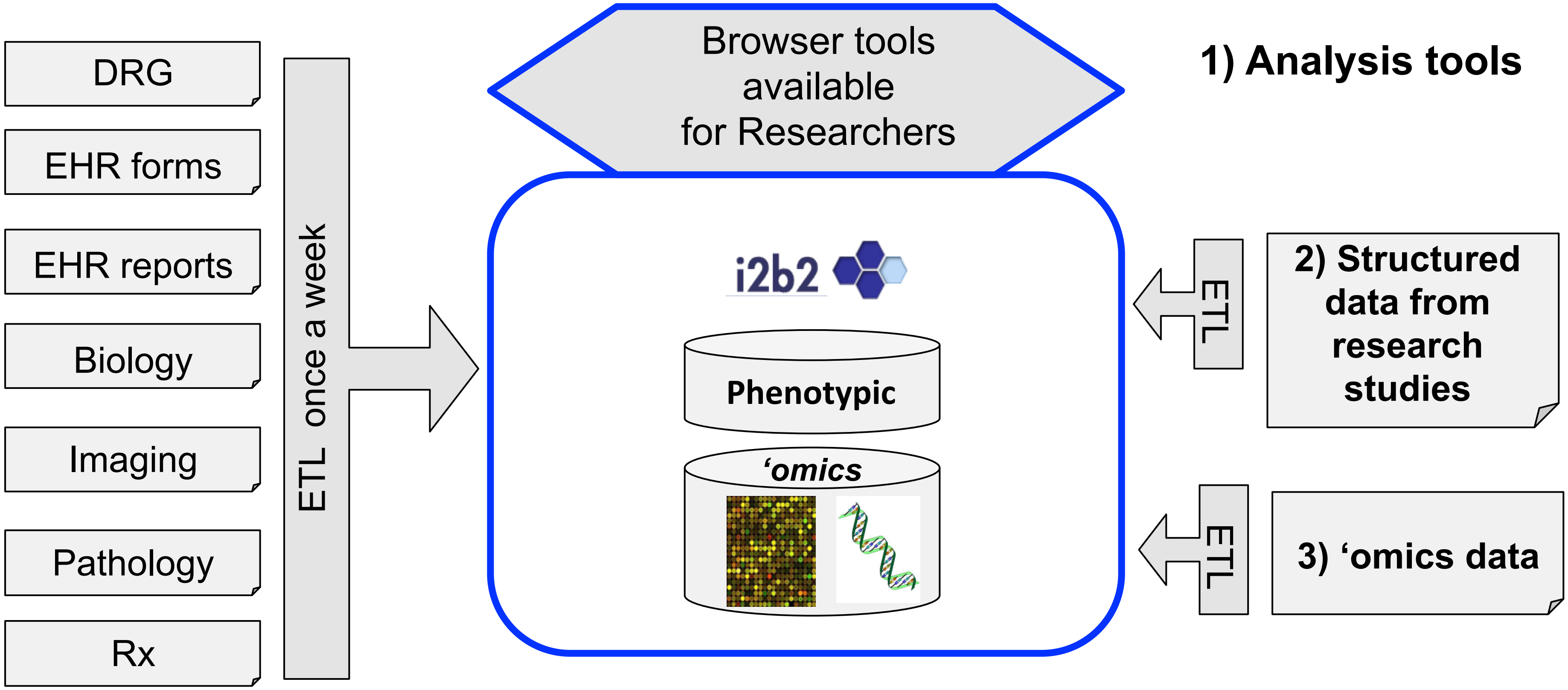
HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

i2b2

tranSMART

## Health care
### Health Information System

## Clinical Research


DRG
EHR forms
EHR reports
Biology
Imaging
Pathology
Rx
ETL once a week
Browser tools
available
for Researchers
1) Analysis tools
i2b2
Phenotypic
'omics
ETL
2) Structured data from research studies
ETL
3) 'omics data

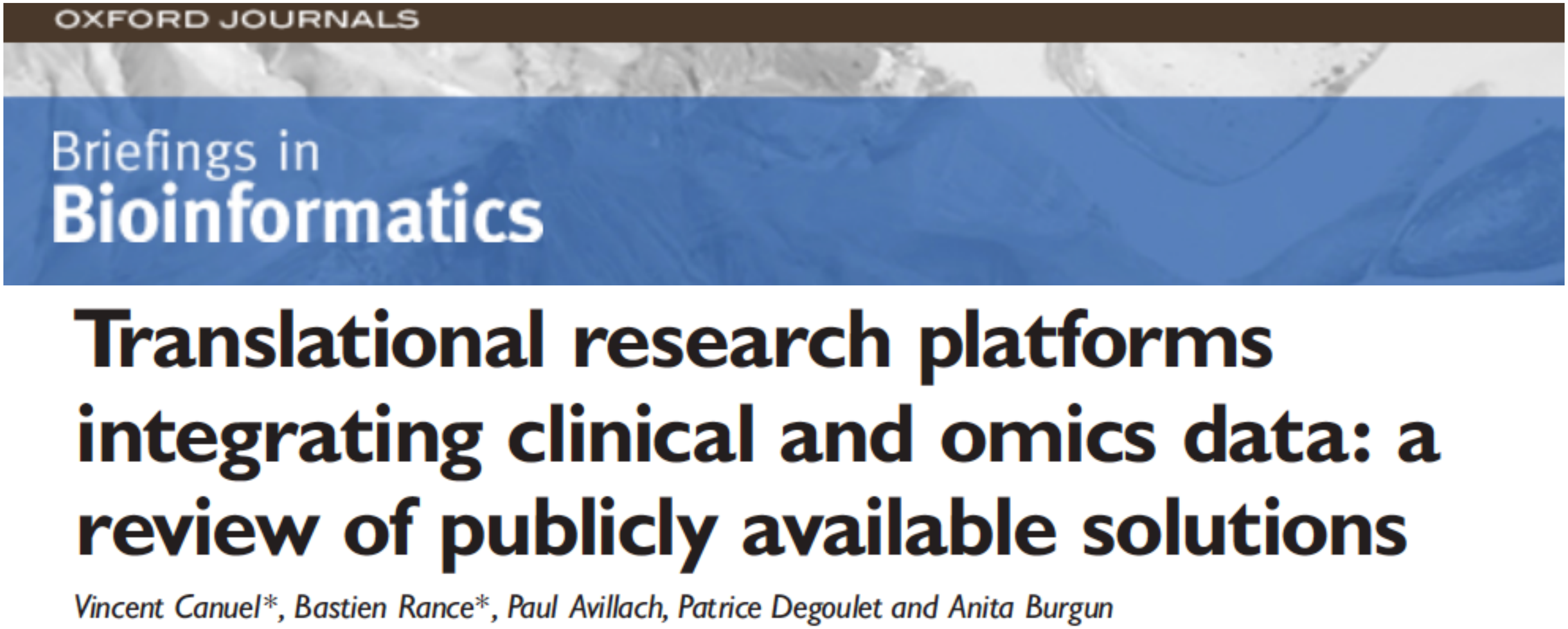
tranSMART

- Integrated platform to support translational research
- Initiated by Johnson & Johnson et Recombinant 6 years ago
  - PI: Eric Perakslis

- Open-source since January 24th , 2012
- Installed at HEGP Hospital, Paris since May, 2012
- Today, driven and maintained by the tranSMART

tranSMART FOUNDATION

http://transmartfoundation.org



OXFORD JOURNALS

Briefings in Bioinformatics

# Translational research platforms integrating clinical and omics data: a review of publicly available solutions

*Vincent Canuel*, Bastien Rance*, Paul Avillach, Patrice Degoulet and Anita Burgun*

BRISK,
caTRIP
cBio Cancer Portal
G-DOC

iCOD
iDASH
**tranSMART (i2b2)**

# tranSMART main installations

- International Research Initiatives
  - IMI – eTRIKS, EMIF
  - CTMM – TraIT
- Pharma & Biotech
  - Sanofi, Millennium, Pfizer, JNJ, Roche
- Government aligned Institutions
  - FDA
- Non-Profits
  - 1Mind4Research, Orion Bionetworks
- Hospitals / Academics
  - U Michigan, John's Hopkins, St. Jude, HEGP, Harvard/Boston Children Hospital
- Service Providers
  - Thomson Reuters, Recombinant(Deloitte), theHyve, Rancho Biosciences, BTGS

| Start | Organization | Type | Stage |
|---|---|---|---|
| 2008 | Johnson & Johnson | Pharma | Production |
| 2008 | Recombinant by Deloitte | Services | Multiple |
| 2010 | Sage Bionetworks | Non profit | Production |
| 2010 | Thomson Reuters | Services | Support |
| 2010 | U-BIOPRED | Consortium | Production |
| 2011 | SAFE-T | Consortium | Pilot |
| 2011 | University of Michigan, Comprehensive Cancer Center (UMCCC) | Academic | Production |
| **2012** | **APHP-HEGP Paris France** | **Academic** | **Production** |
| 2012 | BT Cure | Consortium | Pilot |
| 2012 | CTMM/TraIT | Consortium | Development |
| 2012 | FDA | Government | Development |
| 2012 | IMI/eTRIKS | Consortium | Development |
| 2012 | Merck | Pharma | Pilot |
| 2012 | Millennium Pharmaceuticals | Pharma | Production |
| 2012 | One Mind for Research (1M4R) | Non profit | Production |
| 2012 | Pfizer | Pharma | Production |
| 2012 | Roche | Pharma | Evaluation |
| 2012 | Sanofi-Aventis | Pharma | Development |
| 2012 | St. Jude | Non profit medical center | Development |
| 2012 | University of Michigan, Department of Computational Medicine & Bioinformatics (DCM&B) | Academic | Development, Pilot |
| 2013 | Agios | Biotechnology | Evaluation |
| **2013** | **CARPEM – Cancer personalized medicine** | **Academic French grant** | **Development** |
| **2013** | **Harvard Medical School / Boston Children Hospital** | **Academic** | **Autism Pilot** |
| 2013 | Boehringer Ingelheim | Pharma | Pilot |
| 2013 | Bristol Myers Squibb | Pharma | Evaluation |
| 2013 | BT Global Services | Services | Pilot |
| 2013 | Accelerated Cure Project for MS | Non profit | Development |
| **2014** | **Personalized medicine and colorectal cancers** | **Academic French grant** | **Development** |
| **2014** | **PCORI PRRN Phelan-McDermid Syndrome Data Network** | **Academic US grant** | **Development** |

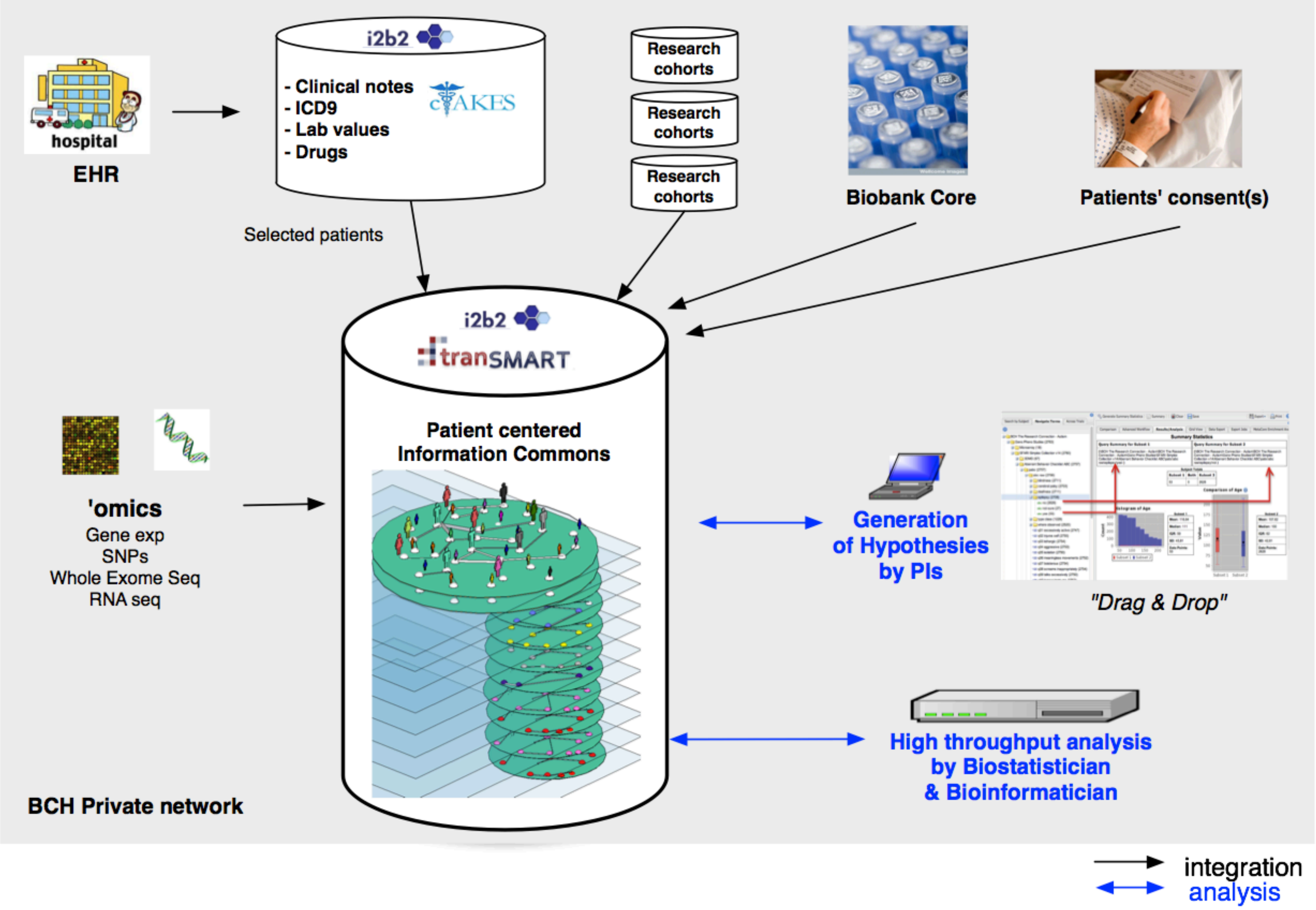
tranSMART

Objectives :

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

**CENTER FOR BIOMEDICAL INFORMATICS**
HARVARD MEDICAL SCHOOL

**Boston Children's Hospital**
Research Connection


hospital

**EHR**

i2b2

- **Clinical notes**
- **ICD9**
- **Lab values**
- **Drugs**

cTAKES

Selected patients

**Research cohorts**

**Research cohorts**

**Research cohorts**

**Biobank Core**

**Patients' consent(s)**

**'omics**

Gene exp

SNPs

Whole Exome Seq

RNA seq

i2b2

tranSMART

**Patient centered Information Commons**

**Generation of Hypothesies by PIs**

*"Drag & Drop"*

**High throughput analysis by Biostatistician & Bioinformatician**

**BCH Private network**

integration

analysis

# Autism cohorts

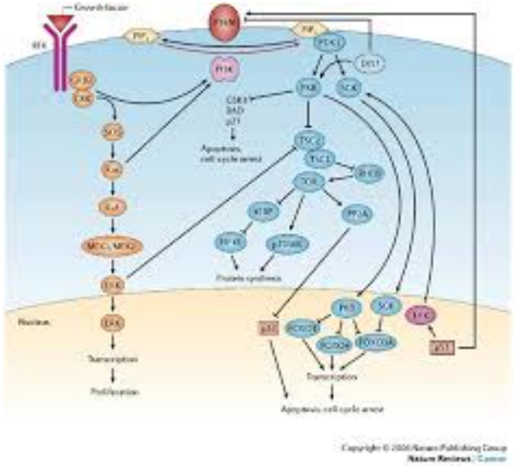
## Phenotype data

| | |
|---|---|
| Simons Simplex Collection (SSC) | 2,760 |
| AGRE | 3,300 |
| Autism Consortium (AC) | 525 |

- Gene-Pheno studies – Lou Kunkel - HMS
  - Pre AC
  - AC
  - SSC

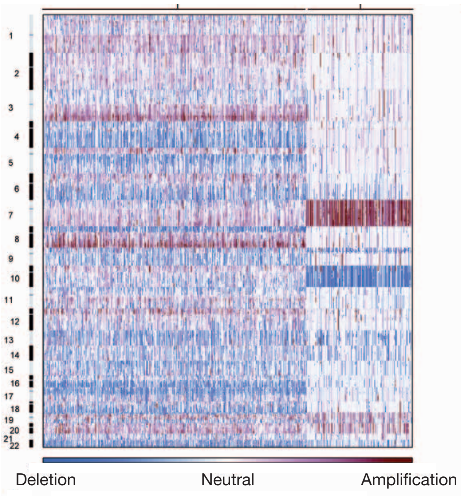
# BIG data

[Bioinformatics & Integrative Genomics]

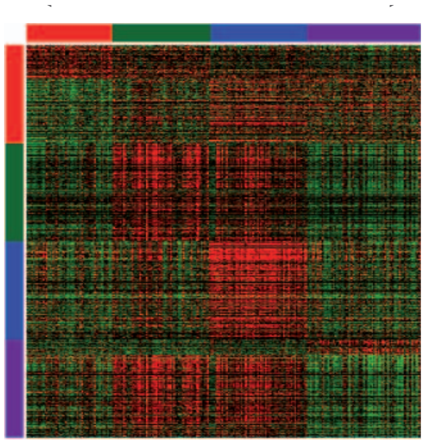

Copyright © 2006 Nature Publishing Group
Nature Reviews | Cancer

Known relationships between genes

Sequence variation



Copy number variation



Gene expression

+ many more

# Raw data

# RNA

| | |
|---|---|
| SSC (**928**) | |
| Blood (Kunkel-Kohane) | |
| – Affymetrix Gene ST 1.0 | 316 |
| – Affymetrix U133+2 | 19 |
| – Illumina HiSeq | 154 |
| LCL (Geschwind) | |
| - Illumina REF-8 3.0 | 439 |
| AC (**166**) | |
| Blood (Kunkel-Kohane) | |
| – Affymetrix Gene ST 1.0 | 117 |
| – Affymetrix U133+2 | 21 |
| – Illumina HiSeq | 28 |
| BCH (**386**) | |
| Blood (Kunkel-Kohane) | |
| – Affymetrix Gene ST 1.0 | 186 |
| – Affymetrix U133+2 | 168 |
| – Illumina HiSeq | 32 |
| AGRE (**1,048**) | |
| LCL (Geschwind) | |
| - Illumina REF-8 3.0 | 1,048 |

# DNA

- Static genomic predisposition
  - Goal: enable streamlined analysis of genomic variation at **any *functional unit* resolution**
    - Single variant / variant type
    - Single gene
    - Gene set / pathway
    - Regulatory module
    - Cellular system
    - Genomic location / context

    …
- Measurement types

# Raw genotyping array data by cohort

**DNA**

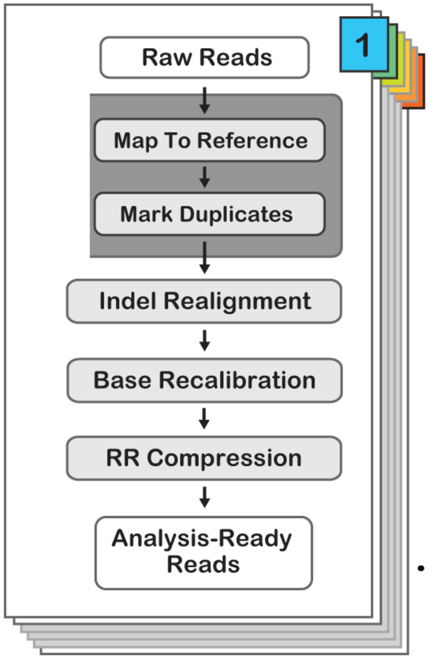
- SSC (**3,184**)
  - Ilumina Infinium1M
  - Illumina 1M Duo
- AC (**60**)
  - Affymetrix SNP 6.0
- AGRE (**3,832**)
  - Affymetrix 10K
  - Affymetrix 500K
  - Illumina HumanHap550
  - Illumina Infinium 1M

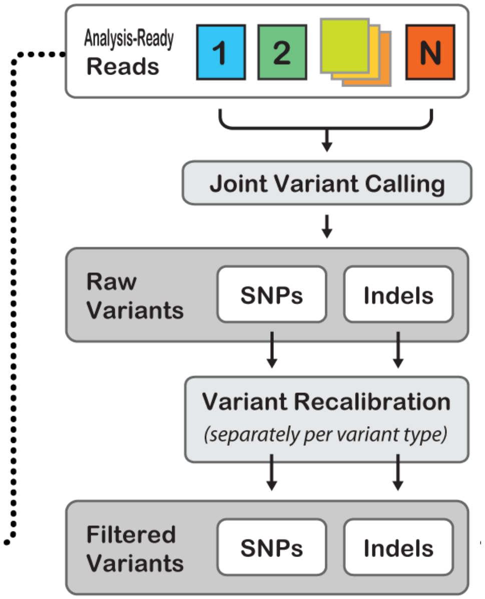
# WHOLE EXOME SEQUENCE data by cohort

**DNA**

- SSC (**2,963**)
  - State **914**
    - Illumina GAIIx
    - Illumina HiSeq
  - Eichler **676**
    - Illumina GAIIx
    - Illumina HiSeq
  - Wigler **1373**
    - Illumina HiSeq
- AC (**381**)
  - Daly **381**
    - Illumina HiSeq
- AGRE (**1672**)
  - Walsh **750**
    - Illumina HiSeq
  - BI-BCM **922**
    - Illumina HiSeq
    - ABI SOLiD

# Exome sequence data processing > Variant calling > Variant annotation > tranSMART input


1
Raw Reads
Map To Reference
Mark Duplicates
Indel Realignment
Base Recalibration
RR Compression
Analysis-Ready Reads


Analysis-Ready Reads
1
2
N
Joint Variant Calling
Raw Variants
SNPs
Indels
Variant Recalibration
(separately per variant type)
Filtered Variants
SNPs
Indels

gatk

**Physical location**
e.g. Chr:start-end
Cytoband
…

**Gene**
e.g. Gene name
Variant function
…

**Gene set**
e.g. Pathway
Molecular proces
…

**Predicted variant impact**
e.g. SIFT
PolyPhen
…

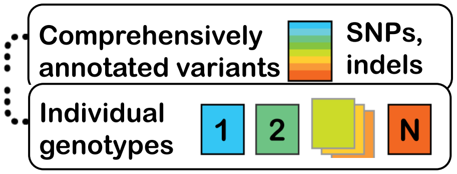
**Conservation**
e.g. GERP
PhyloP
…

**Population frequency**
e.g. 1000 Genomes
ESP 6500
…

**Clinical significance**
e.g. ClinVar
OMIM
…

**Expression patterns**
e.g. GTEx
BrainSpan
…

**Transcriptional regulatio**
e.g. ENCODE TFBS
Histone modifications
…


Comprehensively annotated variants
SNPs, indels
Individual genotypes
1
2
N

**ANNOVAR**

tranSMART

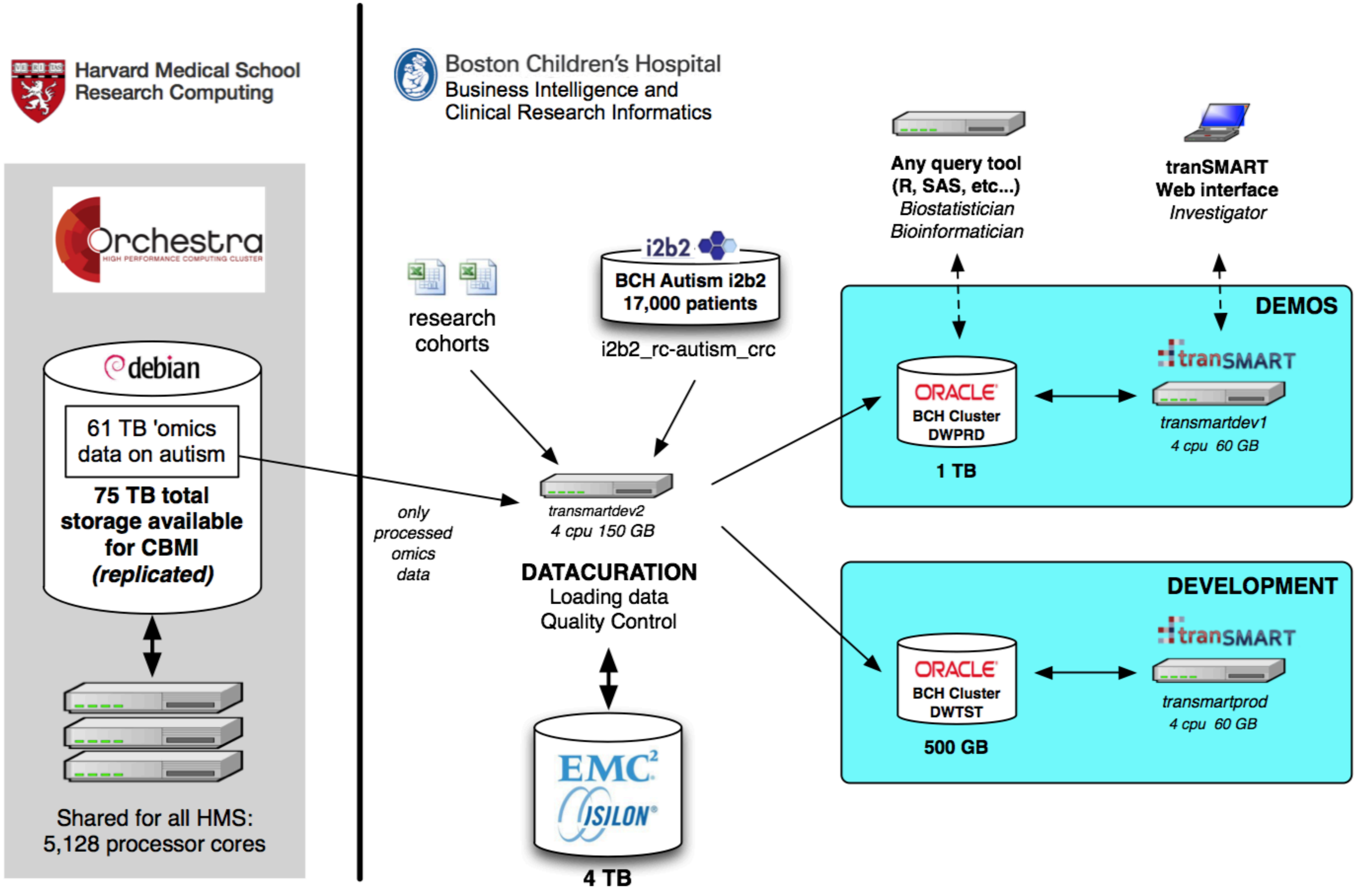
# Infrastructure

*February 7th 2014*

*Paul Avillach, Ally Eran, Michael McDuffie*

Boston Children's Hospital Research Connection

CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL



Harvard Medical School Research Computing

Orchestra
HIGH PERFORMANCE COMPUTING CLUSTER

debian

61 TB 'omics data on autism

**75 TB total storage available for CBMI *(replicated)***

Shared for all HMS:
5,128 processor cores

*only processed omics data*

Boston Children's Hospital
Business Intelligence and
Clinical Research Informatics

research cohorts

i2b2

**BCH Autism i2b2**
**17,000 patients**

i2b2_rc-autism_crc

*transmartdev2*
*4 cpu 150 GB*

**DATACURATION**
Loading data
Quality Control

EMC² ISILON

**4 TB**

**Any query tool**
**(R, SAS, etc...)**
*Biostatistician*
*Bioinformatician*

**tranSMART**
**Web interface**
*Investigator*

**DEMOS**

ORACLE
**BCH Cluster**
**DWPRD**

**1 TB**

tranSMART
*transmartdev1*
*4 cpu 60 GB*

**DEVELOPMENT**

ORACLE
**BCH Cluster**
**DWTST**

**500 GB**

tranSMART
*transmartprod*
*4 cpu 60 GB*

# Live DEMO

https://www.youtube.com/watch?v=rUFH697a2n4

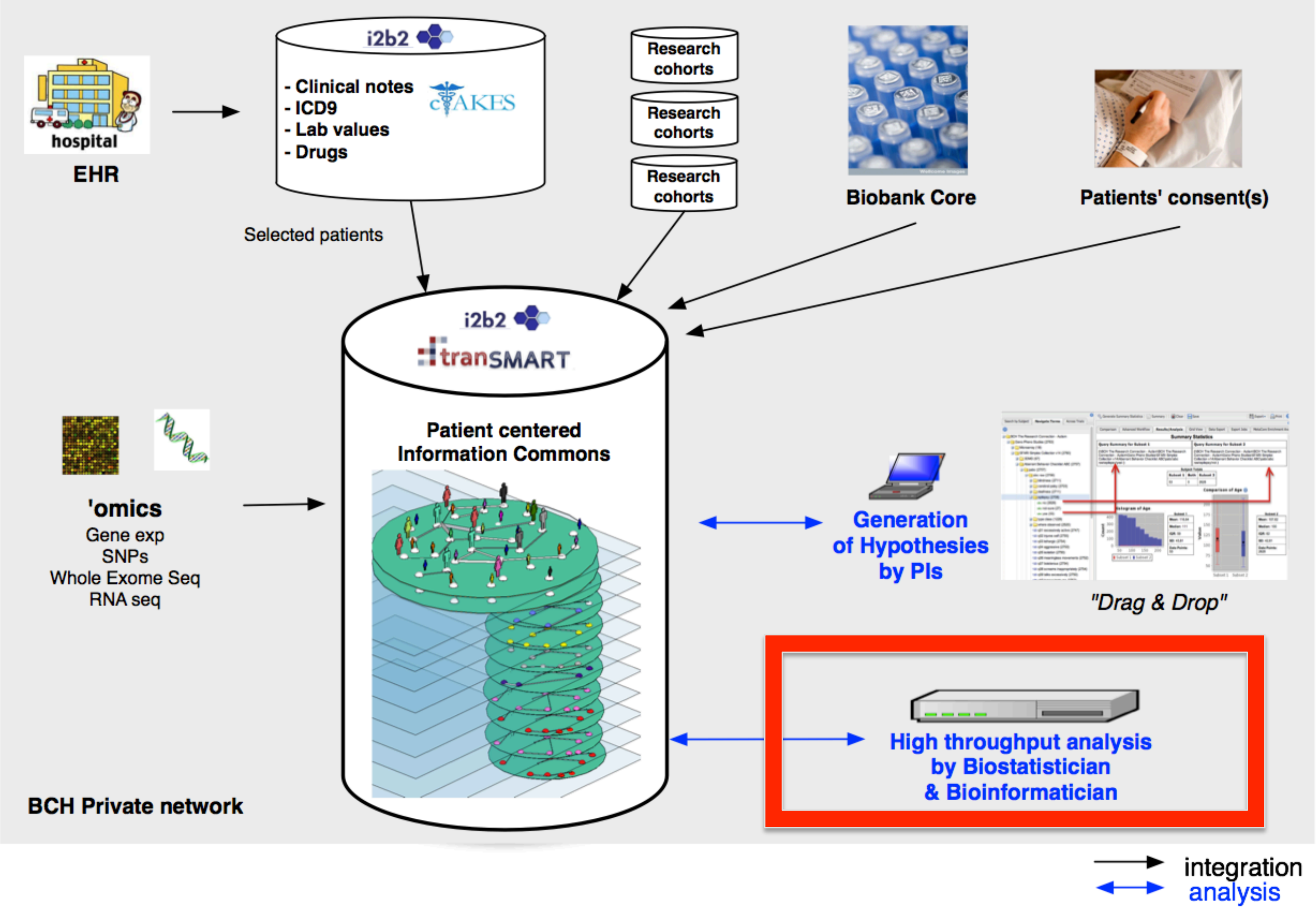
tranSMART

Boston Children's Hospital
Research Connection

CENTER FOR
BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

Boston Children's Hospital
Research Connection



hospital

EHR

i2b2

- Clinical notes
- ICD9
- Lab values
- Drugs

cTAKES

Research cohorts

Research cohorts

Research cohorts

Biobank Core

Patients' consent(s)

Selected patients

i2b2

tranSMART

Patient centered Information Commons

'omics

Gene exp

SNPs

Whole Exome Seq

RNA seq

Generation of Hypothesies by PIs

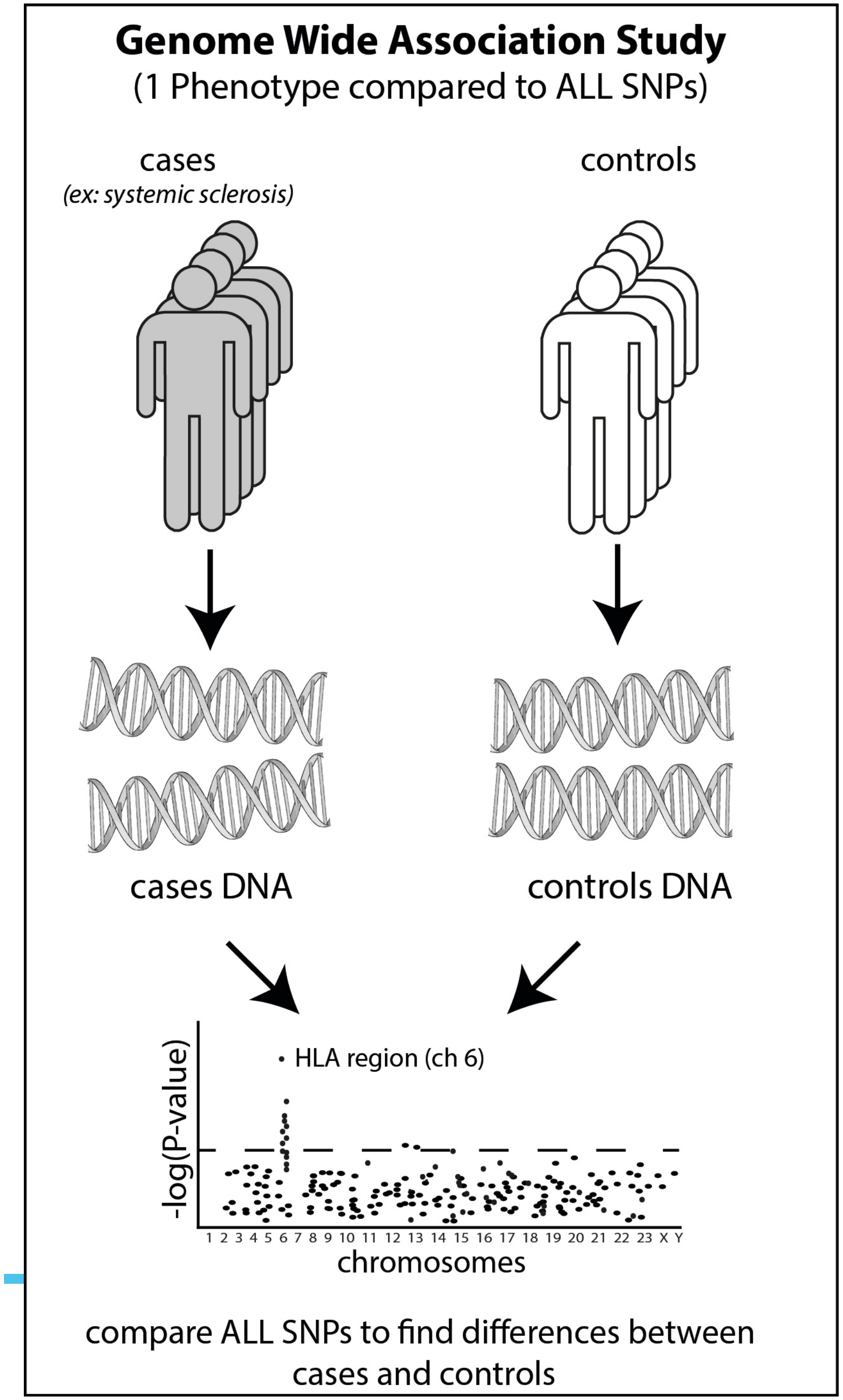
*"Drag & Drop"*

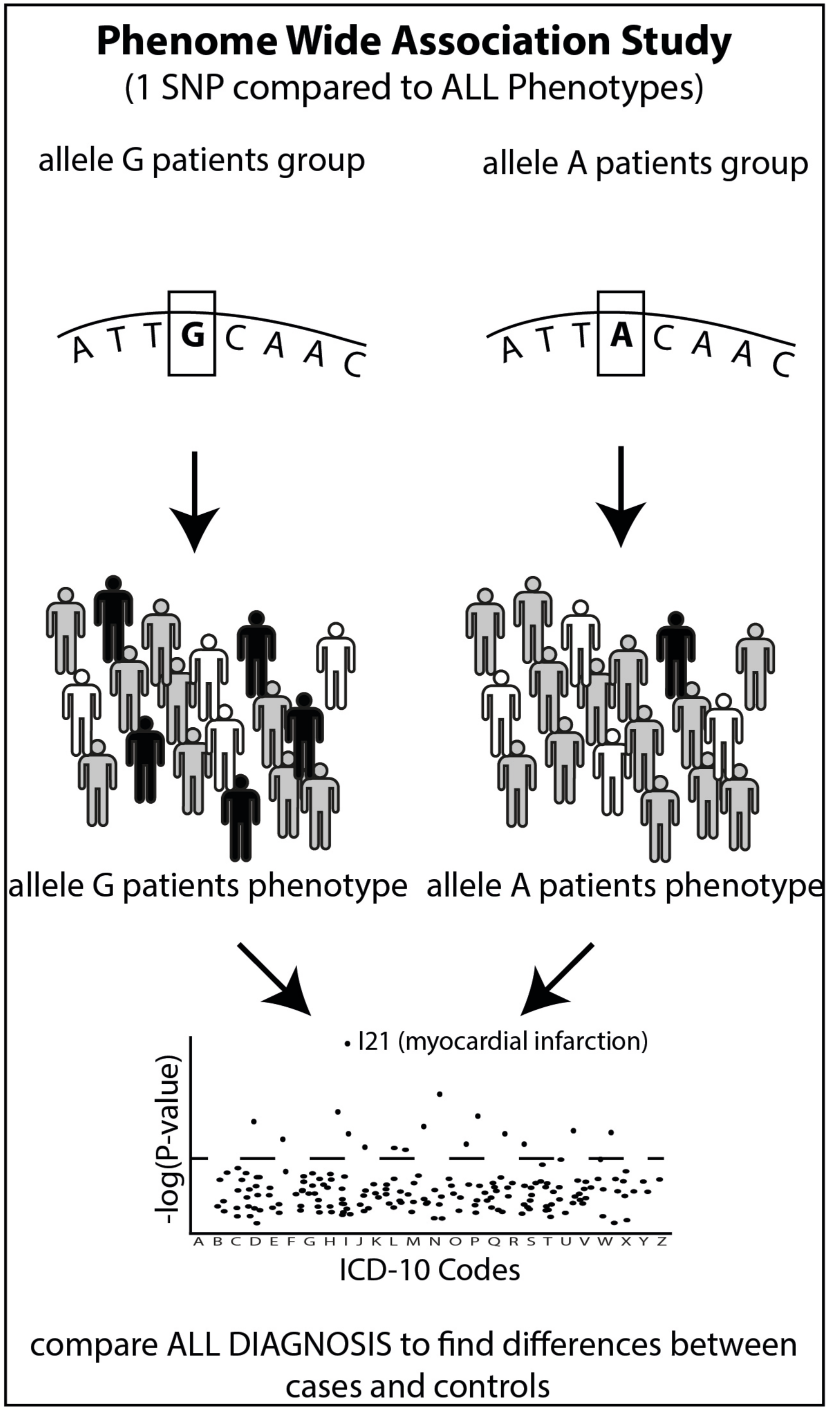
High throughput analysis by Biostatistician & Bioinformatician

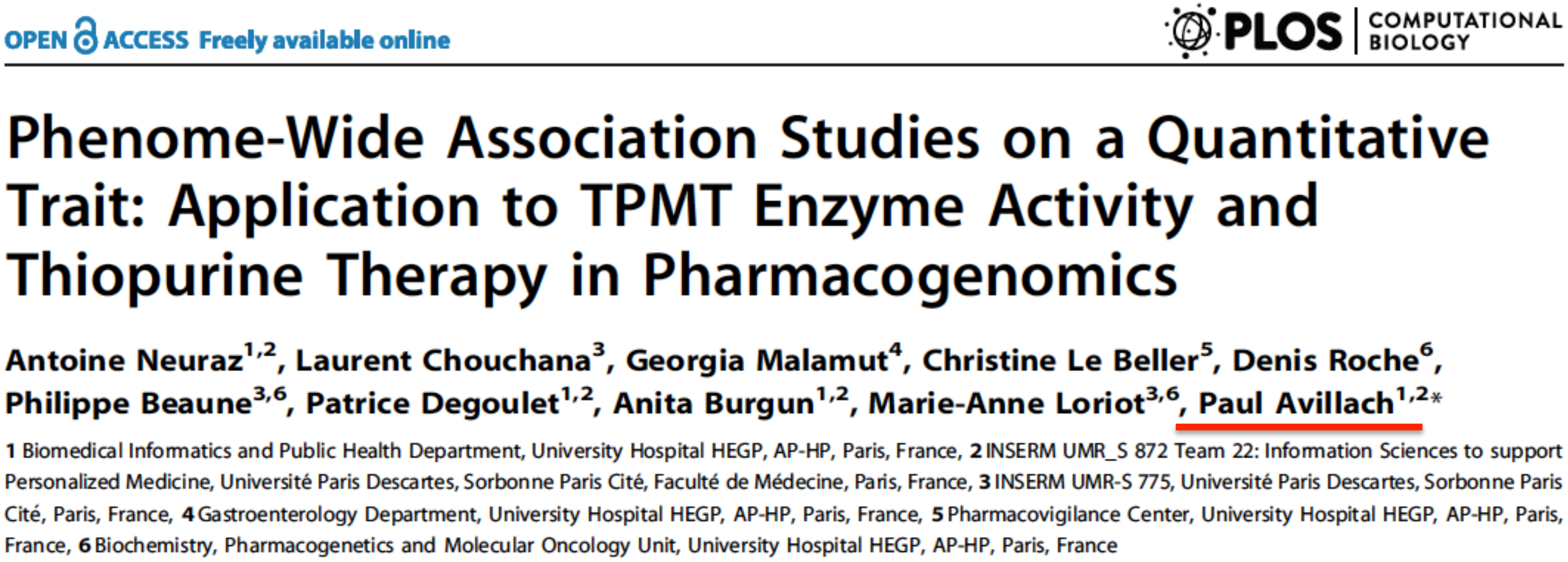
BCH Private network

integration

analysis


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-value)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 X Y
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
A T T G C A A C
A T T A C A A C
allele G patients phenotype
allele A patients phenotype
I21 (myocardial infarction)
-log(P-value)
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
ICD-10 Codes
compare ALL DIAGNOSIS to find differences between cases and controls



OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

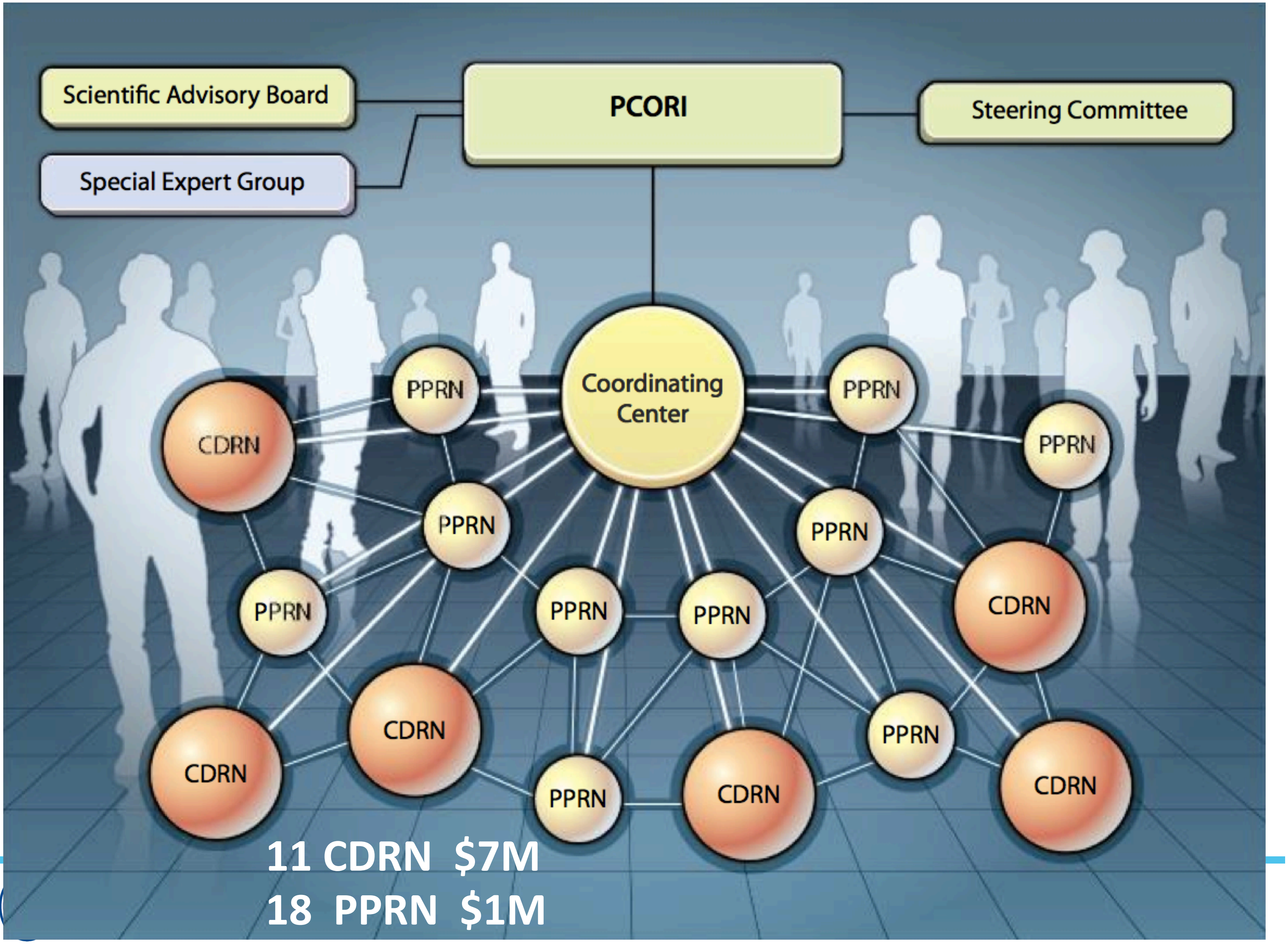
# Phenome-Wide Association Studies on a Quantitative Trait: Application to TPMT Enzyme Activity and Thiopurine Therapy in Pharmacogenomics

**Antoine Neuraz[1,2], Laurent Chouchana[3], Georgia Malamut[4], Christine Le Beller[5], Denis Roche[6], Philippe Beaune[3,6], Patrice Degoulet[1,2], Anita Burgun[1,2], Marie-Anne Loriot[3,6], Paul Avillach[1,2]***

**1** Biomedical Informatics and Public Health Department, University Hospital HEGP, AP-HP, Paris, France, **2** INSERM UMR_S 872 Team 22: Information Sciences to support Personalized Medicine, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris, France, **3** INSERM UMR-S 775, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, **4** Gastroenterology Department, University Hospital HEGP, AP-HP, Paris, France, **5** Pharmacovigilance Center, University Hospital HEGP, AP-HP, Paris, France, **6** Biochemistry, Pharmacogenetics and Molecular Oncology Unit, University Hospital HEGP, AP-HP, Paris, France

pcori | Patient-Centered Outcomes Research Institute

**PCORnet $95 million Start date Jan 2014 18 months**


Scientific Advisory Board
Special Expert Group
PCORI
Steering Committee
Coordinating Center
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
PPRN
CDRN
CDRN
CDRN
CDRN
CDRN
CDRN

11 CDRN $7M

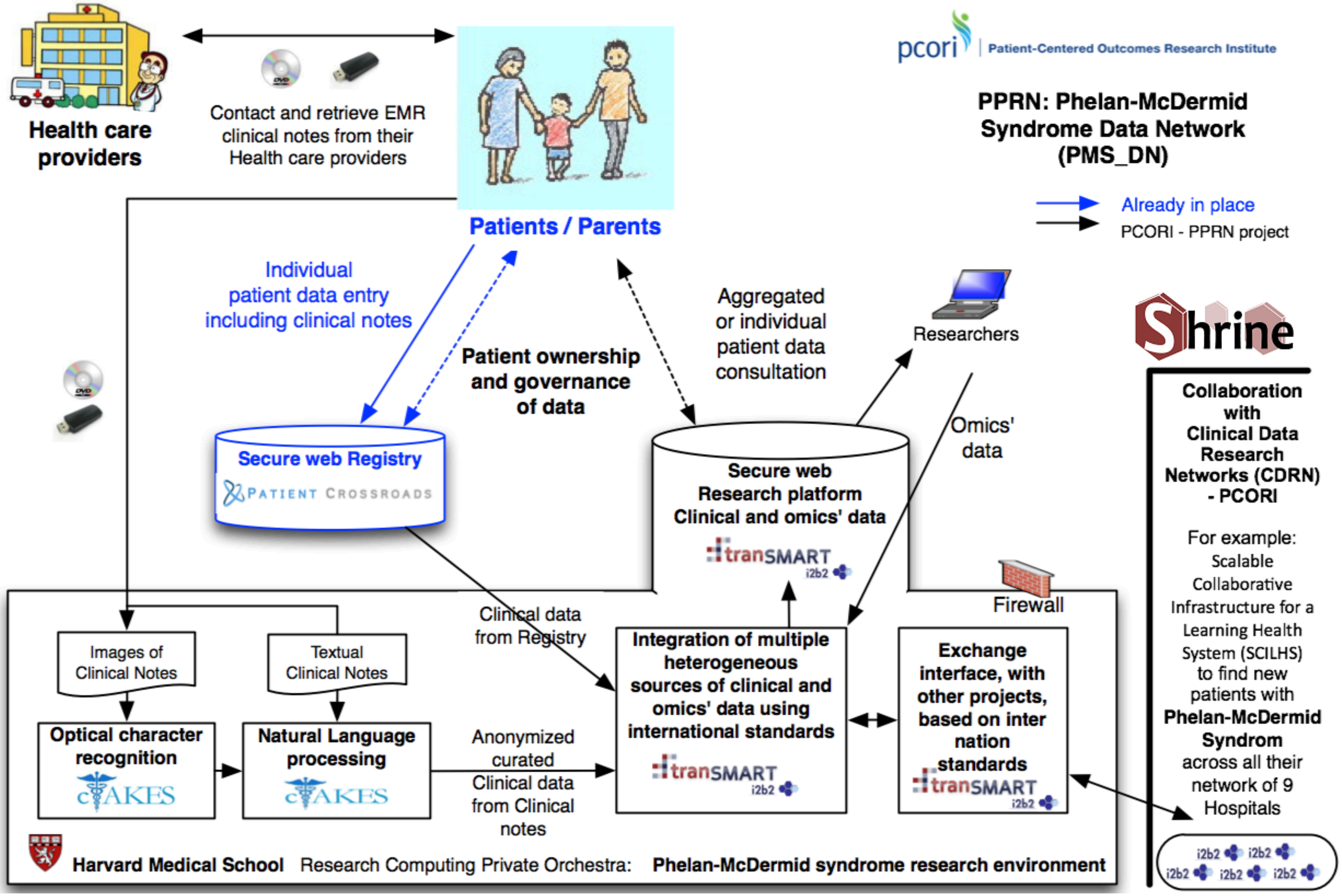
18 PPRN $1M

# PCORI PPRN Grant

- Phelan-McDermid Syndrome Data Network
  - PI: Megan O’Boyle, Mother of PMS patient
  - Co-PI: Paul Avillach, MD, PhD
- Total: $1M
- 18 months
- To collect **all available patient data** from Phelan-McDermid Syndrome (PMS) patients to make meaningful, well-annotated clinical data available to researchers and to share insights with members of the PCORI network

pcori | Patient-Centered Outcomes Research Institute

# PPRN: Phelan-McDermid Syndrome Data Network (PMS_DN)



Already in place

PCORI - PPRN project

**Health care providers**

Contact and retrieve EMR clinical notes from their Health care providers

**Patients / Parents**

Individual patient data entry including clinical notes

**Patient ownership and governance of data**

Aggregated or individual patient data consultation

Researchers

Omics' data

**Secure web Registry**

PATIENT CROSSROADS

**Secure web Research platform Clinical and omics' data**

tranSMART i2b2

Firewall

Shrine

**Collaboration with Clinical Data Research Networks (CDRN) - PCORI**

For example:
Scalable Collaborative Infrastructure for a Learning Health System (SCILHS) to find new patients with **Phelan-McDermid Syndrom** across all their network of 9 Hospitals

i2b2 i2b2 i2b2 i2b2 i2b2

Images of Clinical Notes

Textual Clinical Notes

**Optical character recognition**

cTAKES

**Natural Language processing**

cTAKES

Clinical data from Registry

Anonymized curated Clinical data from Clinical notes

**Integration of multiple heterogeneous sources of clinical and omics' data using international standards**

tranSMART i2b2

**Exchange interface, with other projects, based on inter nation standards**

tranSMART i2b2

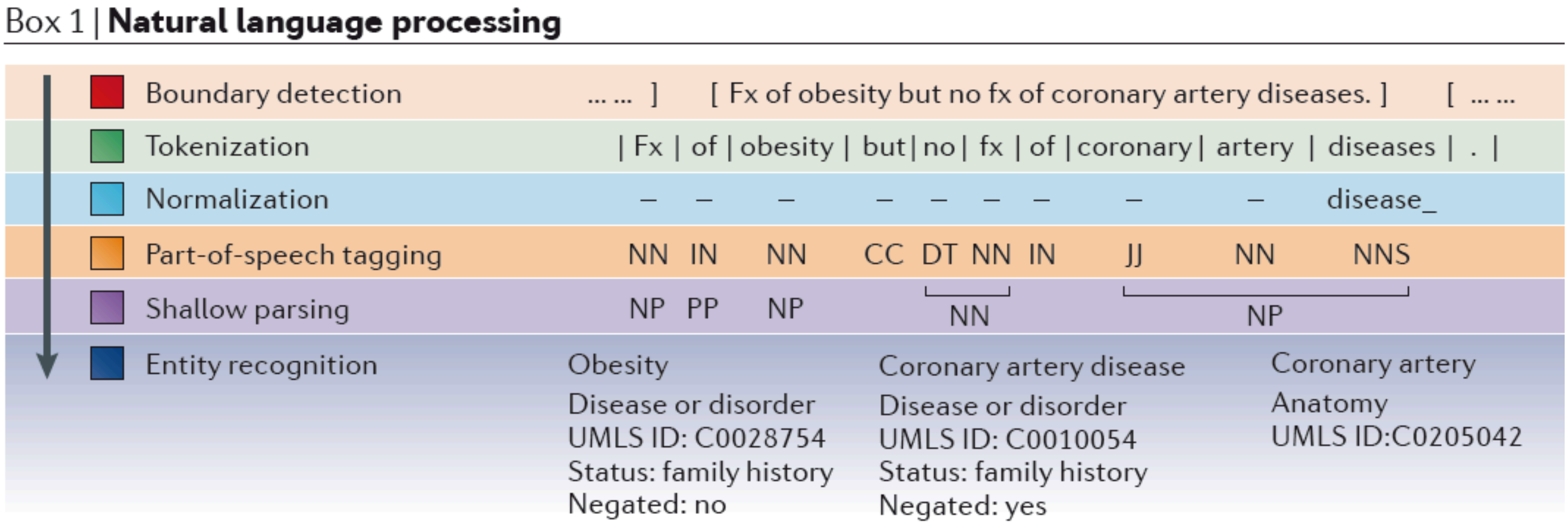
**Harvard Medical School** Research Computing Private Orchestra: **Phelan-McDermid syndrome research environment**

Boston Children's Hospital
Until every child is well

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

cTAKES™

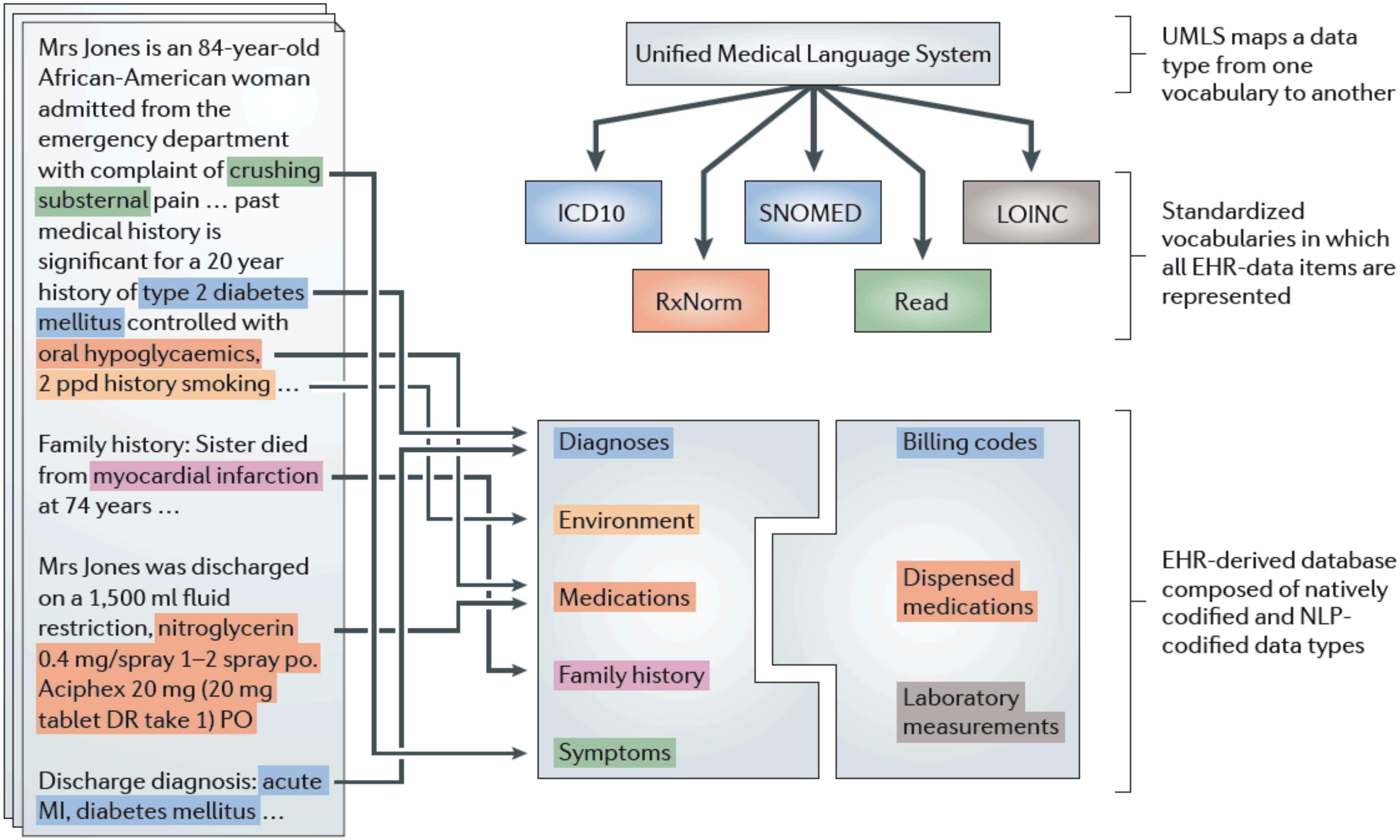
The Apache Software Foundation
http://www.apache.org/



Box 1 | **Natural language processing**

| | |
|---|---|
| 🟥 Boundary detection | …… ] [ Fx of obesity but no fx of coronary artery diseases. ] [ …… |
| 🟩 Tokenization | \| Fx \| of \| obesity \| but \| no \| fx \| of \| coronary \| artery \| diseases \| . \| |
| 🟦 Normalization | – – – – – – – – – disease_ |
| 🟧 Part-of-speech tagging | NN IN NN CC DT NN IN JJ NN NNS |
| 🟪 Shallow parsing | NP PP NP NN NP |
| 🟦 Entity recognition | Obesity<br>Disease or disorder<br>UMLS ID: C0028754<br>Status: family history<br>Negated: no |
| | Coronary artery disease<br>Disease or disorder<br>UMLS ID: C0010054<br>Status: family history<br>Negated: yes |
| | Coronary artery<br>Anatomy<br>UMLS ID:C0205042 |

Peter B. Jensen, Lars J. Jensen and Søren Brunak, Nat Rev Genet. 2012

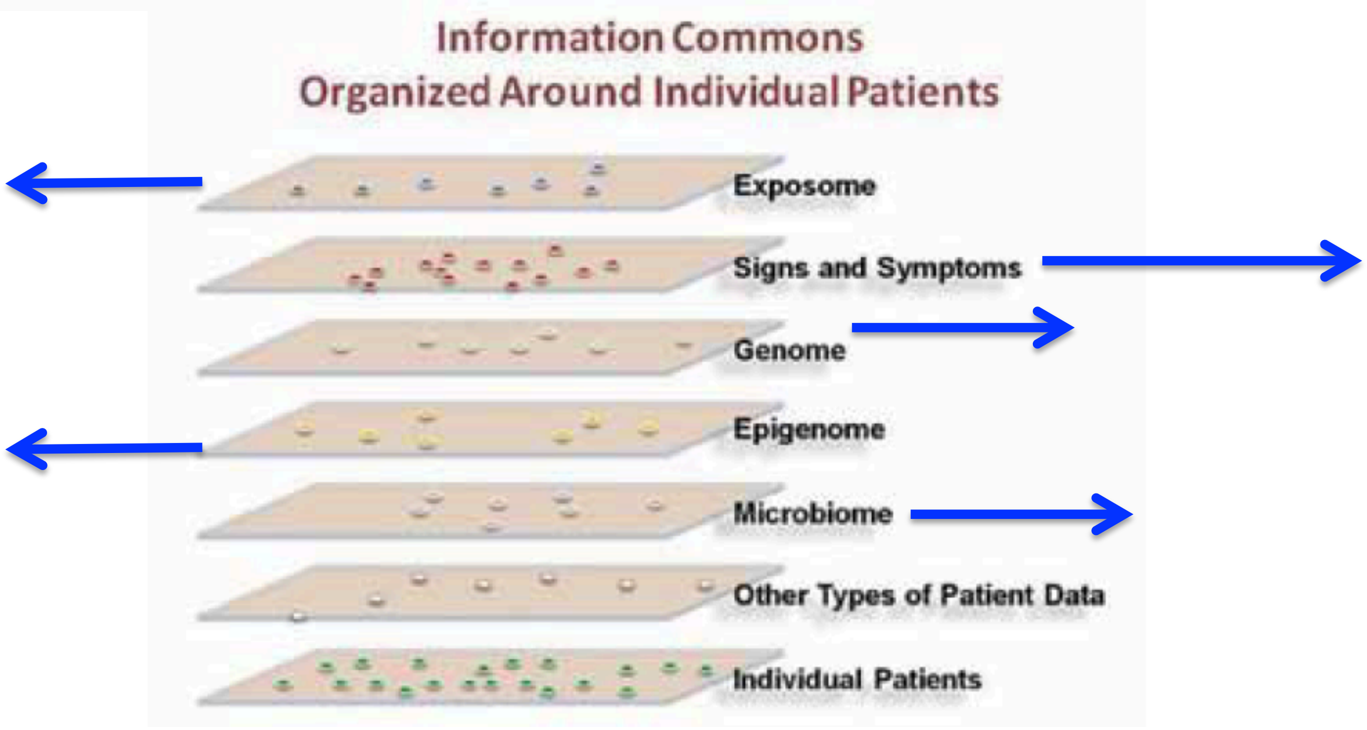

Mrs Jones is an 84-year-old African-American woman admitted from the emergency department with complaint of crushing substernal pain … past medical history is significant for a 20 year history of type 2 diabetes mellitus controlled with oral hypoglycaemics, 2 ppd history smoking …
Family history: Sister died from myocardial infarction at 74 years …
Mrs Jones was discharged on a 1,500 ml fluid restriction, nitroglycerin 0.4 mg/spray 1–2 spray po. Aciphex 20 mg (20 mg tablet DR take 1) PO
Discharge diagnosis: acute MI, diabetes mellitus …
Unified Medical Language System
ICD10
SNOMED
LOINC
RxNorm
Read
UMLS maps a data type from one vocabulary to another
Standardized vocabularies in which all EHR-data items are represented
Diagnoses
Environment
Medications
Family history
Symptoms
Billing codes
Dispensed medications
Laboratory measurements
EHR-derived database composed of natively codified and NLP-codified data types

Kohane I, Nature Review Gen. 2011

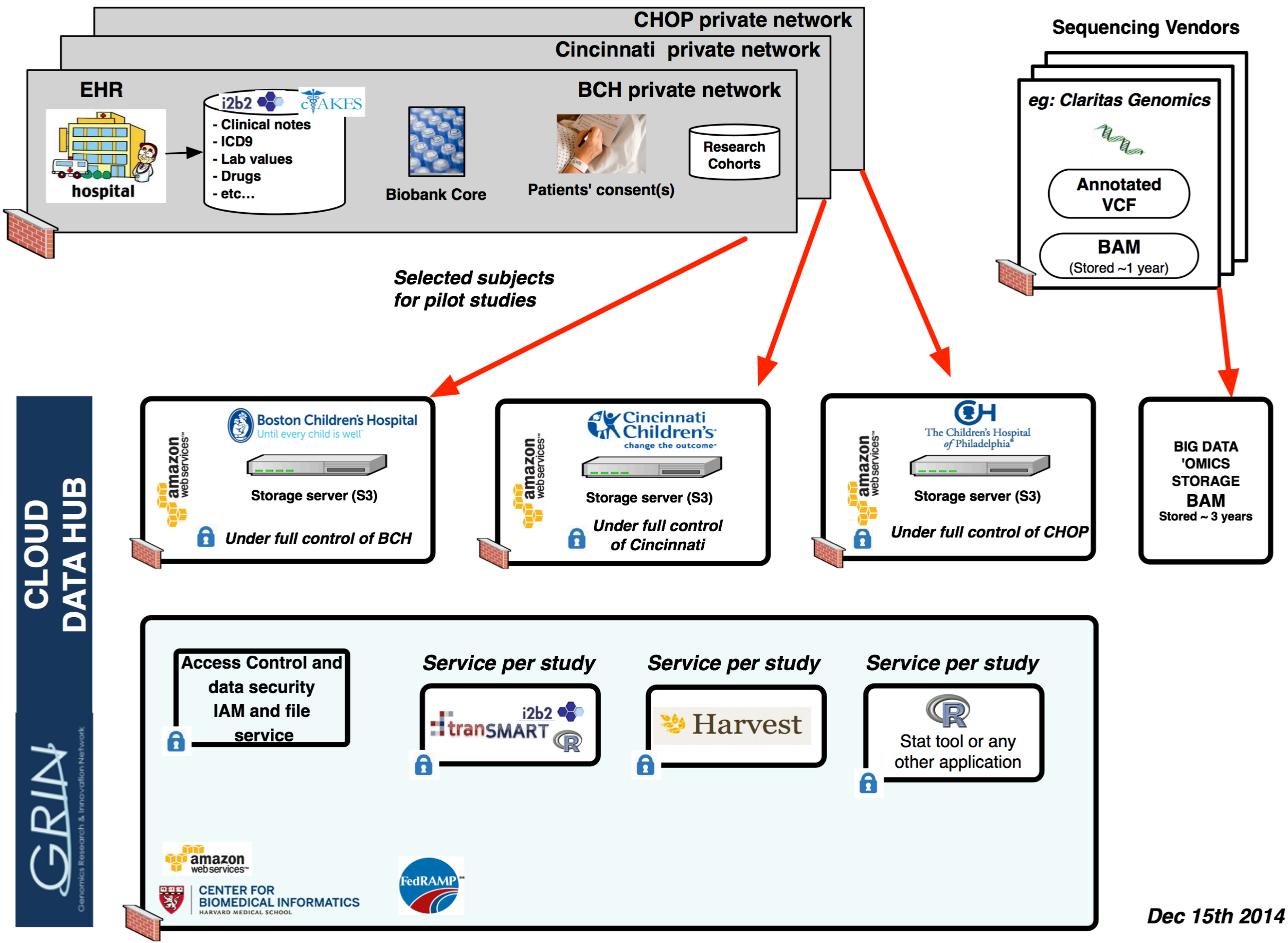
# Live Demo PMS_DN

NIH Big Data to Knowledge (BD2K)

# Patient Centric Information Commons (PIC)
PI: Isaac Kohane


Information Commons
Organized Around Individual Patients

Exposome
Signs and Symptoms
Genome
Epigenome
Microbiome
Other Types of Patient Data
Individual Patients



**CHOP private network**

**Cincinnati private network**

**BCH private network**

**EHR**

**hospital**

i2b2 cTAKES

- Clinical notes
- ICD9
- Lab values
- Drugs
- etc…

**Biobank Core**

**Patients' consent(s)**

**Research Cohorts**

**Sequencing Vendors**

***eg: Claritas Genomics***

**Annotated VCF**

**BAM**
(Stored ~1 year)

***Selected subjects for pilot studies***

**CLOUD DATA HUB**

GRIN Genomics Research & Innovation Network

amazon web services™

Boston Children's Hospital
Until every child is well

**Storage server (S3)**

***Under full control of BCH***

amazon web services™

Cincinnati Children's
change the outcome

**Storage server (S3)**

***Under full control of Cincinnati***

amazon web services™

CH The Children's Hospital of Philadelphia

**Storage server (S3)**

***Under full control of CHOP***

**BIG DATA 'OMICS STORAGE**
**BAM**
**Stored ~ 3 years**

**Access Control and data security IAM and file service**

***Service per study***

i2b2 tranSMART R

***Service per study***

Harvest

***Service per study***

R
Stat tool or any other application

amazon web services™

CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

FedRAMP℠

***Dec 15th 2014***

Boston Children's Hospital
Research Connection

CENTER FOR
BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

# Autism Cohort

**Division of Developmental Medicine**

Leonard Rappaport, MD, MS
Ellen Hanson, PhD

**BCH Division of Genetics & Genomics**

Timothy Yu, MD, PhD
Ingrid Holm, MD, MPH
Stephanie Brewster, MS, CGC
Joanna Reinwald, MS, GC
Frank Jackson

**Laboratory of cognitive neuroscience**

Charles Nelson, PhD
Vanessa Vogel-Farley
Nicole Coman

**CBMI / ResCon tranSMART team**

Paul Avillach, MD, PhD
Michael McDuffie, MS
Ally Eran, PhD

**The Research Connection**

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

**CBMI**

Eric D Perakslis, PhD
Alexa T. McCray, PhD
Dennis Wall, PhD
Nathan Palmer, PhD
Sek Won Kong, MD
Finale Doshi-Velez, PhD

**i2b2 / Partners**

Shawn Murphy, MD, PhD
Lori Phillips, Ms
Michael Mendis

**Principal Investigators**

Isaac Kohane, MD, PhD
Louis Kunkel, PhD
David Margulies, MD
Jonathan Bickel, MD, MS
Paul Avillach, MD, PhD

**Business Intelligence and Clinical Research Informatics**

Mohamad Daniar
Nandan Patibandla
Rick Agrella
Paul OByrne
Lynne N. Alley
Gina Bianco

**Clinical NLP**

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

# PMS_DN team

- **Megan O'Boyle, PI & Mom of Shannon**
- Paul Avillach, MD, PhD , Co-PI, Harvard Medical School
- Liz Horn, PhD, Co-PI, Network Director

PMSF Research director

- Geraldine Bliss, MSc & Mom of

Project Manager

- Andria Cornell Mann

LGC  Data Network Specialist

- Rebecca Davis

Family Engagement Specialist

- Jackie Malasky

Harvard Medical School : CBMI

- Sushma Hanawal
- Michael McDuffie, MSc
- Isaac Kohane, MD, PhD
- Eric Perakslis, PhD

Boston Children's Hospital: cTAKES NLP

- Guergana Savova, PhD
- Pei Chen

Harvard Medical School : IT Infrastructure Support

- Christopher Botka
- David Hummel
- Daniel Lewis

# Additional slides

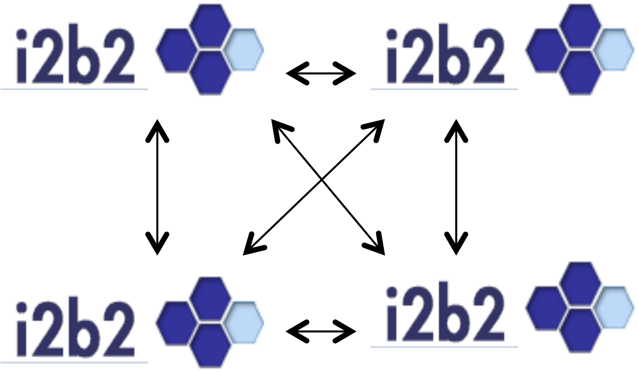
OPEN ACCESS Freely available online

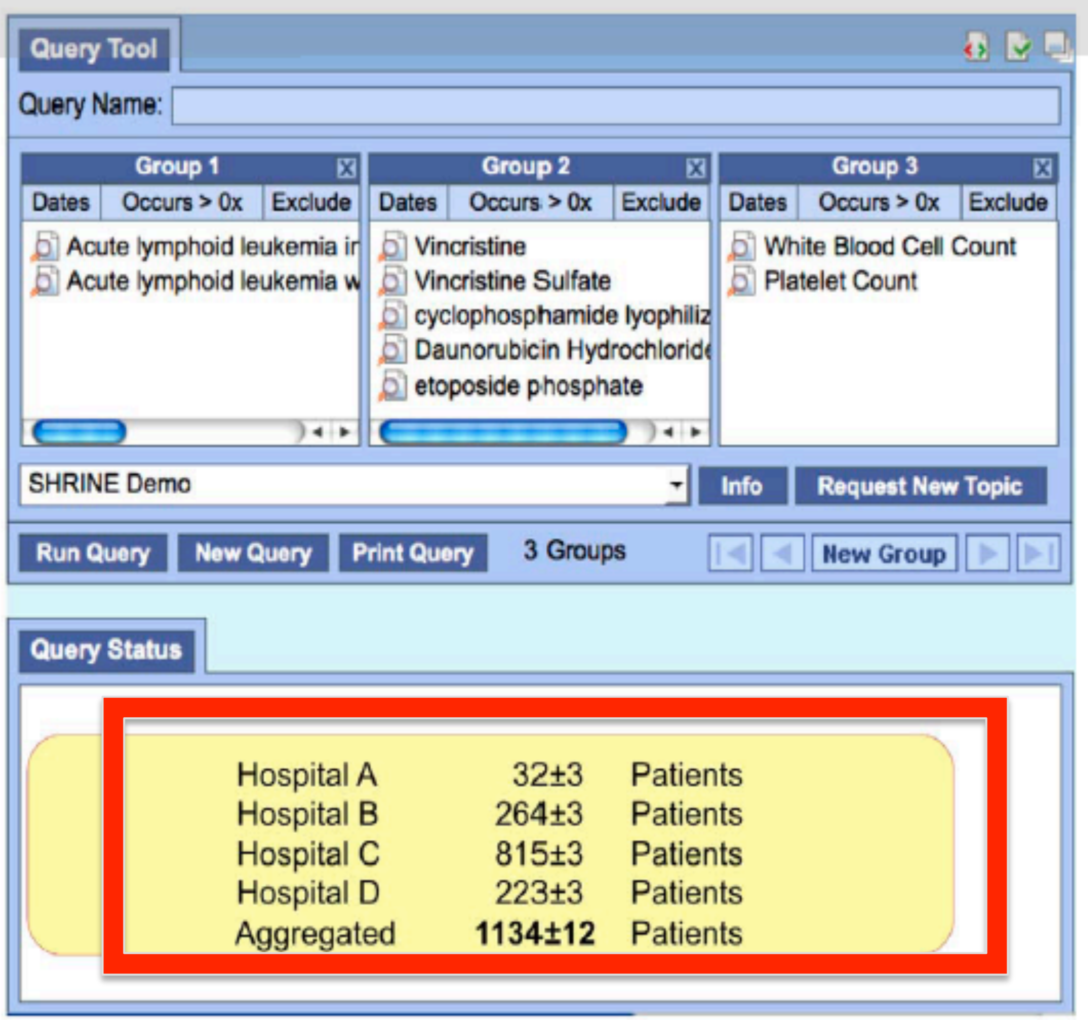
PLOS | ONE

# SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies

Andrew J. McMurry[1,2,3,4*], Shawn N. Murphy[3,5,6], Douglas MacFadden[1], Griffin Weber[3,7], William W. Simons[1], John Orechia[8], Jonathan Bickel[2,9], Nich Wattanasin[5], Clint Gilbert[1], Philip Trevvett[1], Susanne Churchill[3,5], Isaac S. Kohane[1,2,3]

Shrine


i2b2
i2b2
i2b2
i2b2


Query Tool
Query Name:
Group 1
Dates
Occurs > 0x
Exclude
Group 2
Dates
Occurs > 0x
Exclude
Group 3
Dates
Occurs > 0x
Exclude
Info
Request New Topic
Run Query
New Query
Print Query
3 Groups
New Group
Query Status

Group 1:
- Acute lymphoid leukemia ir
- Acute lymphoid leukemia w

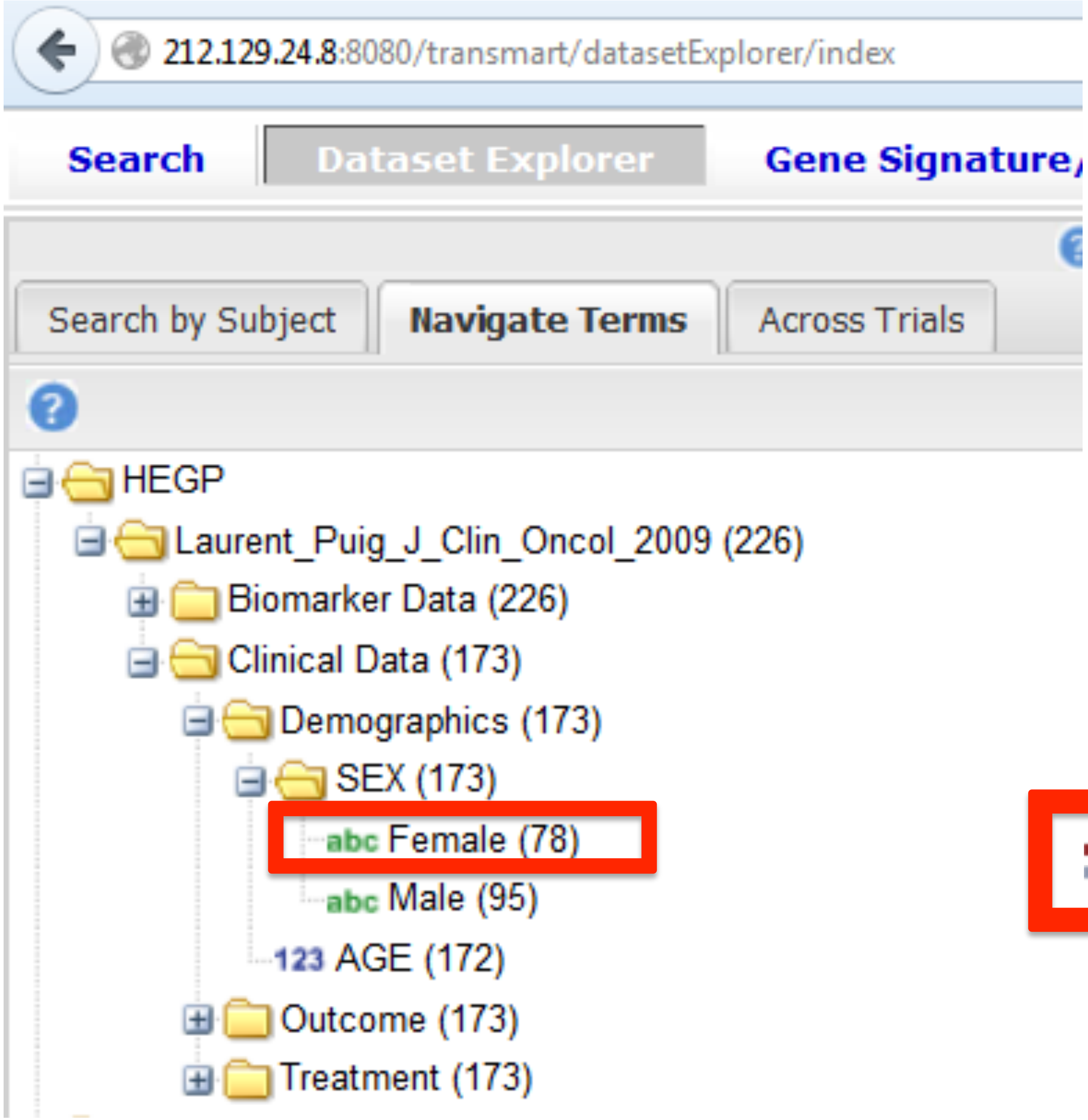
Group 2:
- Vincristine
- Vincristine Sulfate
- cyclophosphamide lyophiliz
- Daunorubicin Hydrochlorid
- etoposide phosphate

Group 3:
- White Blood Cell Count
- Platelet Count

SHRINE Demo

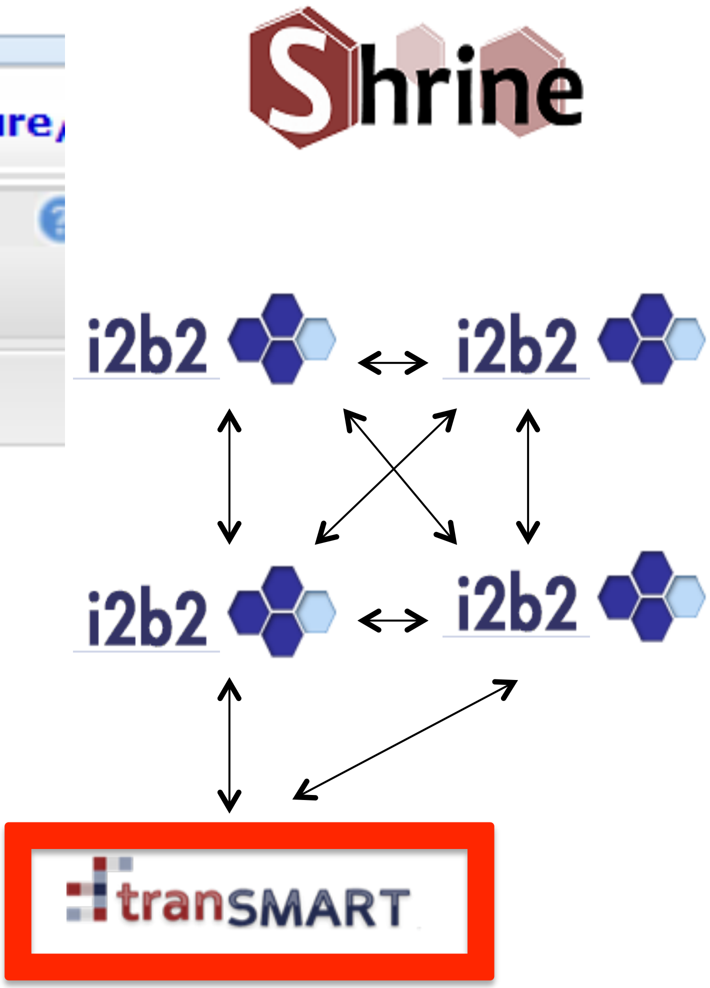
| | | |
|---|---|---|
| Hospital A | 32±3 | Patients |
| Hospital B | 264±3 | Patients |
| Hospital C | 815±3 | Patients |
| Hospital D | 223±3 | Patients |
| Aggregated | **1134±12** | Patients |

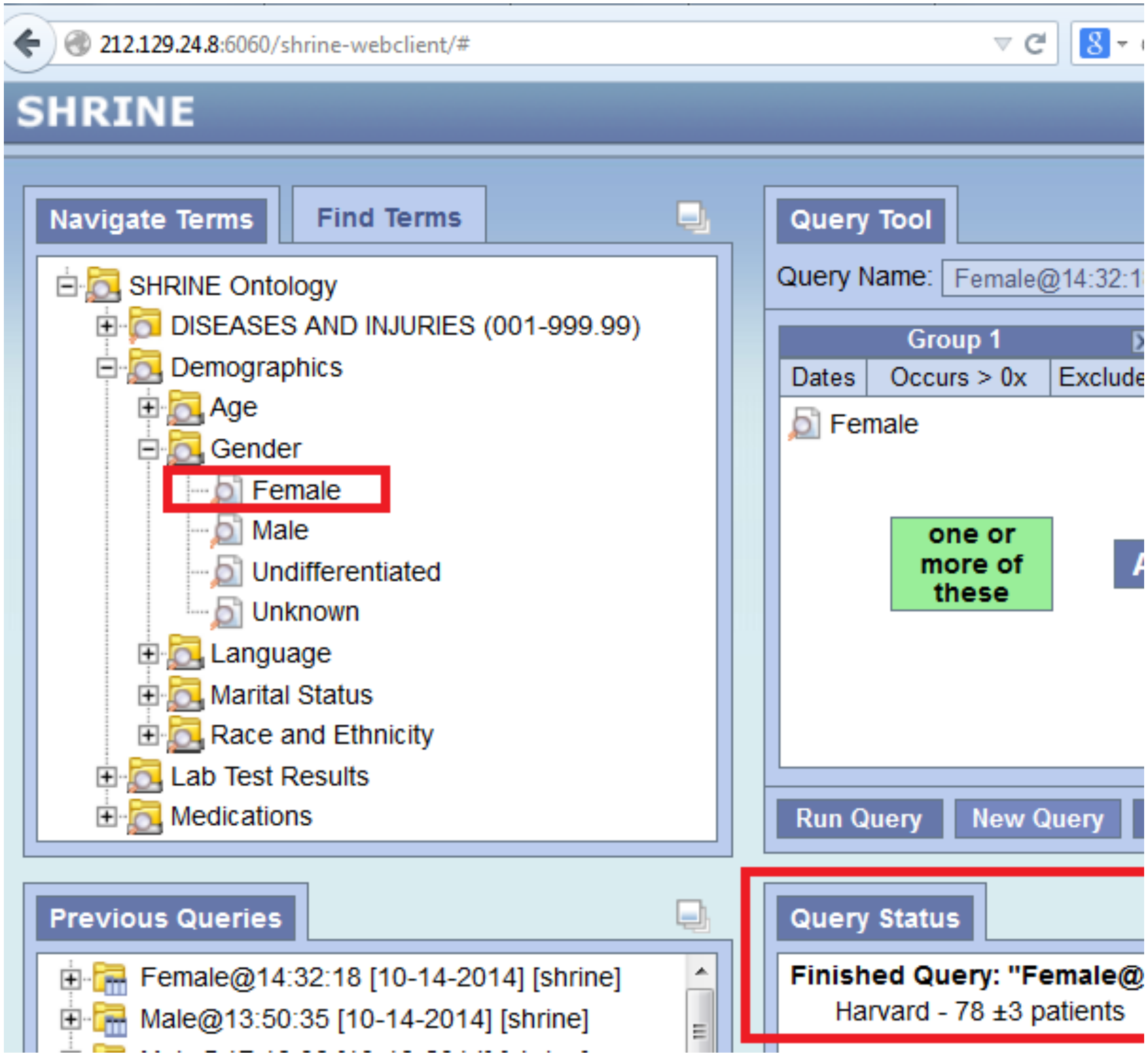
Boston Children's Hospital
Until every child is well™



212.129.24.8:8080/transmart/datasetExplorer/index

Search
Dataset Explorer
Gene Signature
Search by Subject
Navigate Terms
Across Trials
HEGP
Laurent_Puig_J_Clin_Oncol_2009 (226)
Biomarker Data (226)
Clinical Data (173)
Demographics (173)
SEX (173)
abc Female (78)
abc Male (95)
123 AGE (172)
Outcome (173)
Treatment (173)


Shrine
i2b2
i2b2
i2b2
i2b2
tranSMART


212.129.24.8:6060/shrine-webclient/#

SHRINE

Navigate Terms
Find Terms
Query Tool
Run Query
New Query
Previous Queries
Query Status

- SHRINE Ontology
  - DISEASES AND INJURIES (001-999.99)
  - Demographics
    - Age
    - Gender
      - Female
      - Male
      - Undifferentiated
      - Unknown
    - Language
    - Marital Status
    - Race and Ethnicity
  - Lab Test Results
  - Medications

Query Name: Female@14:32:1

Group 1

Dates | Occurs > 0x | Exclude

Female

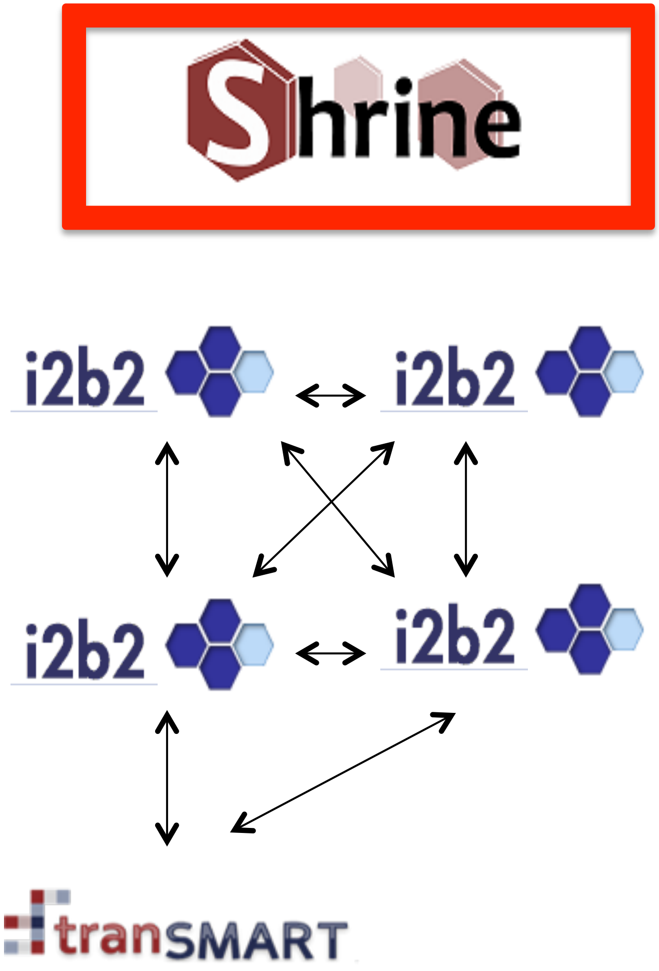
one or more of these

Female@14:32:18 [10-14-2014] [shrine]

Male@13:50:35 [10-14-2014] [shrine]

**Finished Query: "Female@**

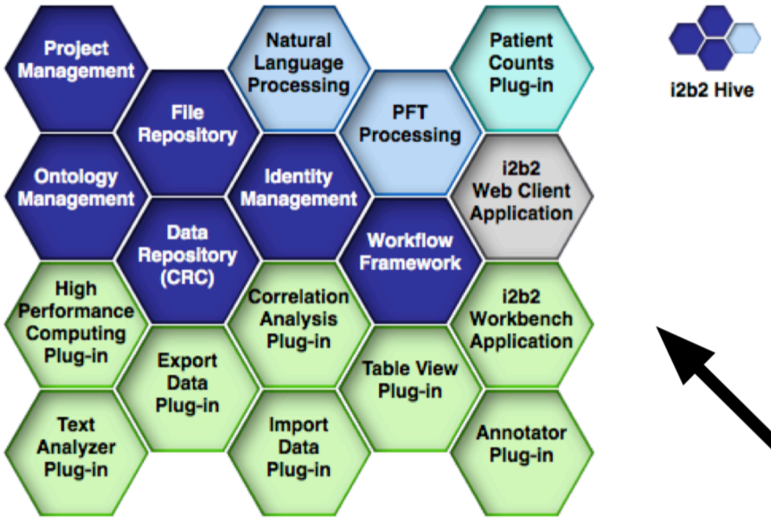
Harvard - 78 ±3 patients


Shrine
i2b2
i2b2
i2b2
i2b2
tranSMART

# i2b2 World

Federated Advance cohort selection

Advance cohort selection


Project Management
File Repository
Natural Language Processing
PFT Processing
Patient Counts Plug-in
Ontology Management
Identity Management
i2b2 Web Client Application
Data Repository (CRC)
Workflow Framework
High Performance Computing Plug-in
Correlation Analysis Plug-in
i2b2 Workbench Application
Export Data Plug-in
Table View Plug-in
Text Analyzer Plug-in
Import Data Plug-in
Annotator Plug-in
i2b2 Hive

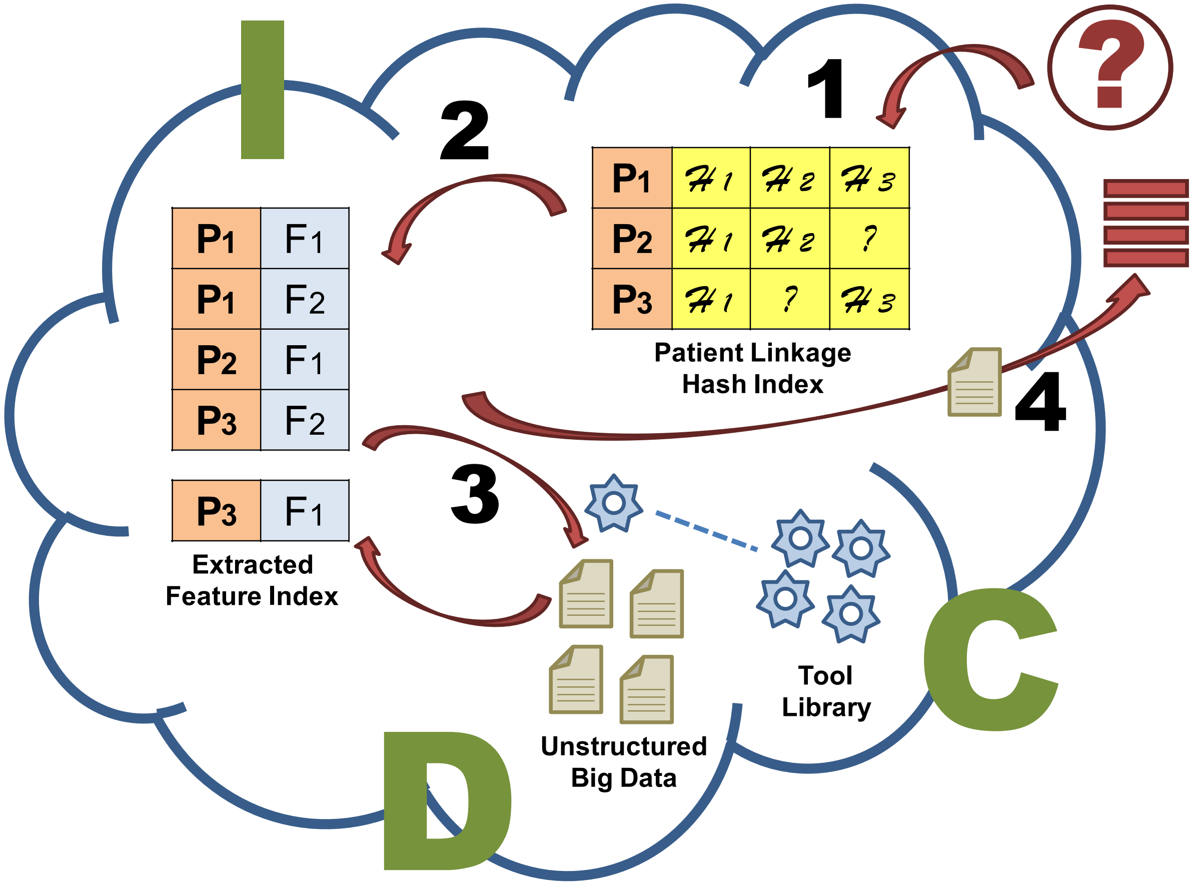
Advance statistical tools
Biobank explorer
Variant explorer

Patient level data lookup
Interoperable tools

NIH Big Data to Knowledge (BD2K)

# Patient Centric Information Commons (PIC)

## PI: Isaac Kohane


I
2
1
?
4
3
C
D
Unstructured Big Data
Tool Library

|  |  |  |  |
|---|---|---|---|
| P1 | H1 | H2 | H3 |
| P2 | H1 | H2 | ? |
| P3 | H1 | ? | H3 |

Patient Linkage Hash Index

|  |  |
|---|---|
| P1 | F1 |
| P1 | F2 |
| P2 | F1 |
| P3 | F2 |

|  |  |
|---|---|
| P3 | F1 |

Extracted Feature Index

Boston Children's Hospital
Until every child is well

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

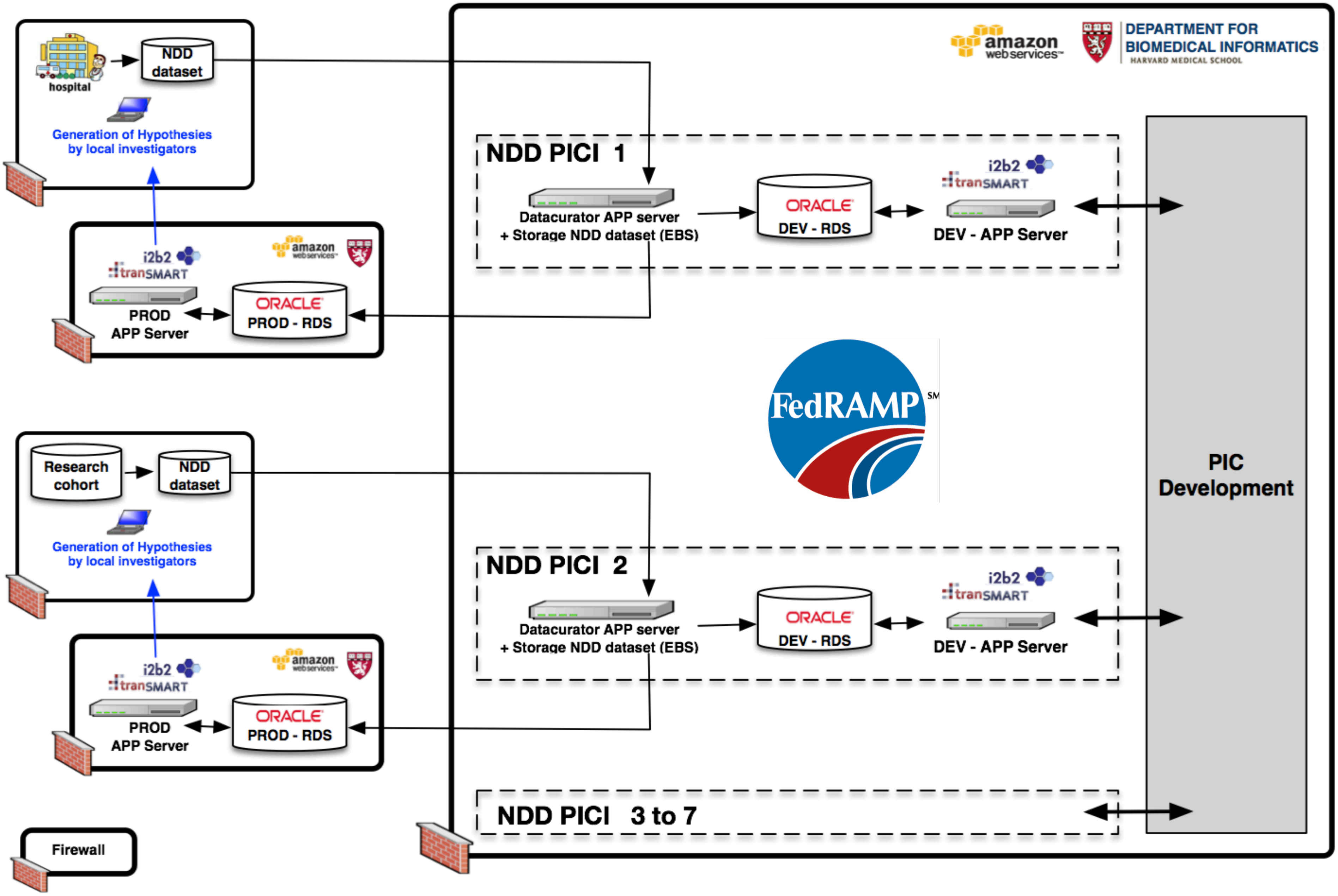
NIH Big Data to Knowledge (BD2K)

# Patient Centric Information Commons (PIC)

## Sandbox: Neurodevelopment Disorders PICIs (NDD PICIs)

— DRAFT —

Oct 3rd 2014



hospital

NDD dataset

Generation of Hypotheses by local investigators

i2b2

tranSMART

PROD APP Server

ORACLE PROD - RDS

amazon web services

Research cohort

NDD dataset

Generation of Hypotheses by local investigators

i2b2

tranSMART

PROD APP Server

ORACLE PROD - RDS

Firewall

amazon web services

DEPARTMENT FOR BIOMEDICAL INFORMATICS

HARVARD MEDICAL SCHOOL

**NDD PICI 1**

Datacurator APP server + Storage NDD dataset (EBS)

ORACLE DEV - RDS

i2b2

tranSMART

DEV - APP Server

FedRAMP℠

**NDD PICI 2**

Datacurator APP server + Storage NDD dataset (EBS)

ORACLE DEV - RDS

i2b2

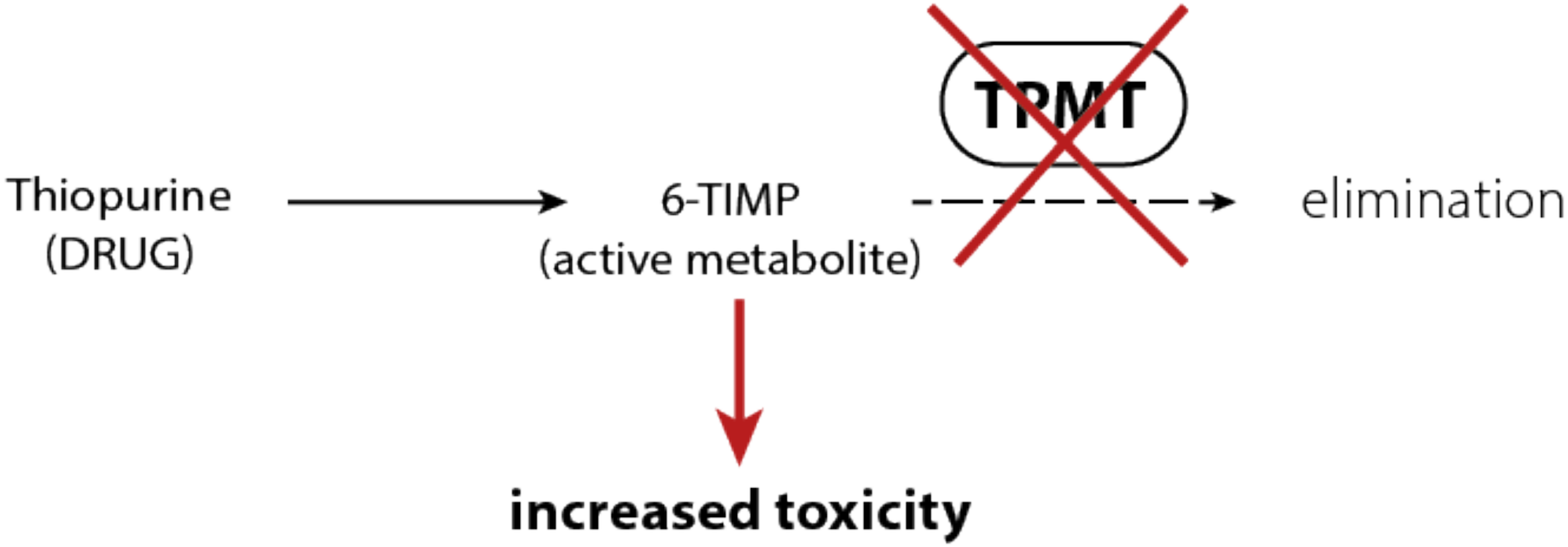
tranSMART

DEV - APP Server

**NDD PICI 3 to 7**

PIC Development

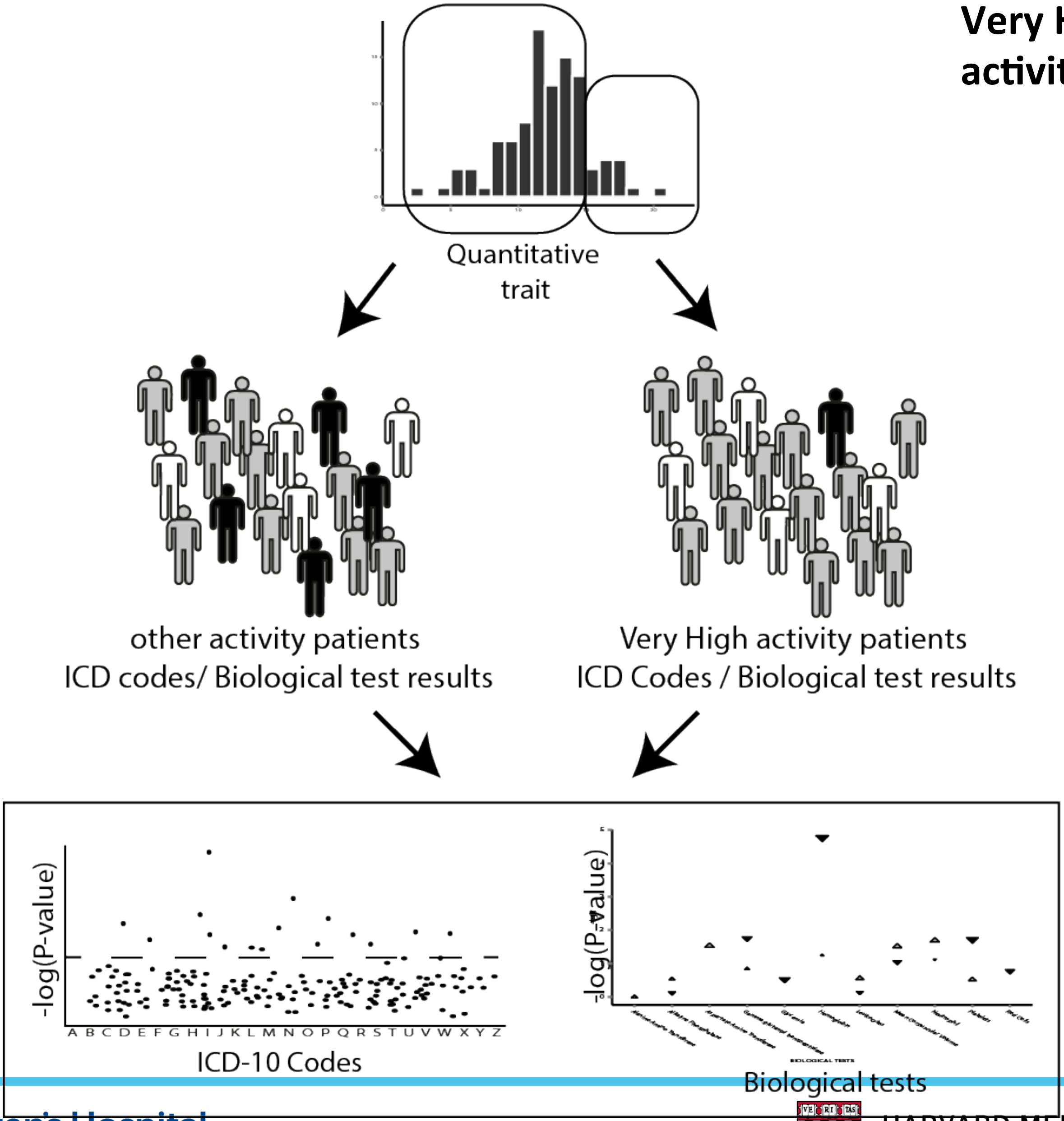
# Methodes: Selection of trait: enzymatic Activity TPMT


Thiopurine
(DRUG)
6-TIMP
(active metabolite)
TPMT
elimination
increased toxicity

| | | | | | |
|---|---|---|---|---|---|
| FDA & EMA recommendations | **Phenotype** | **Low activity** | **Intermediate activity** | **Normal Activity** | ? |
| | **Thiopurine dose** | **10 % dose** | **30 – 70 % dose** | **100 % dose** | |

# TPMT

**Very High TPMT activity *vs* others**



Quantitative trait

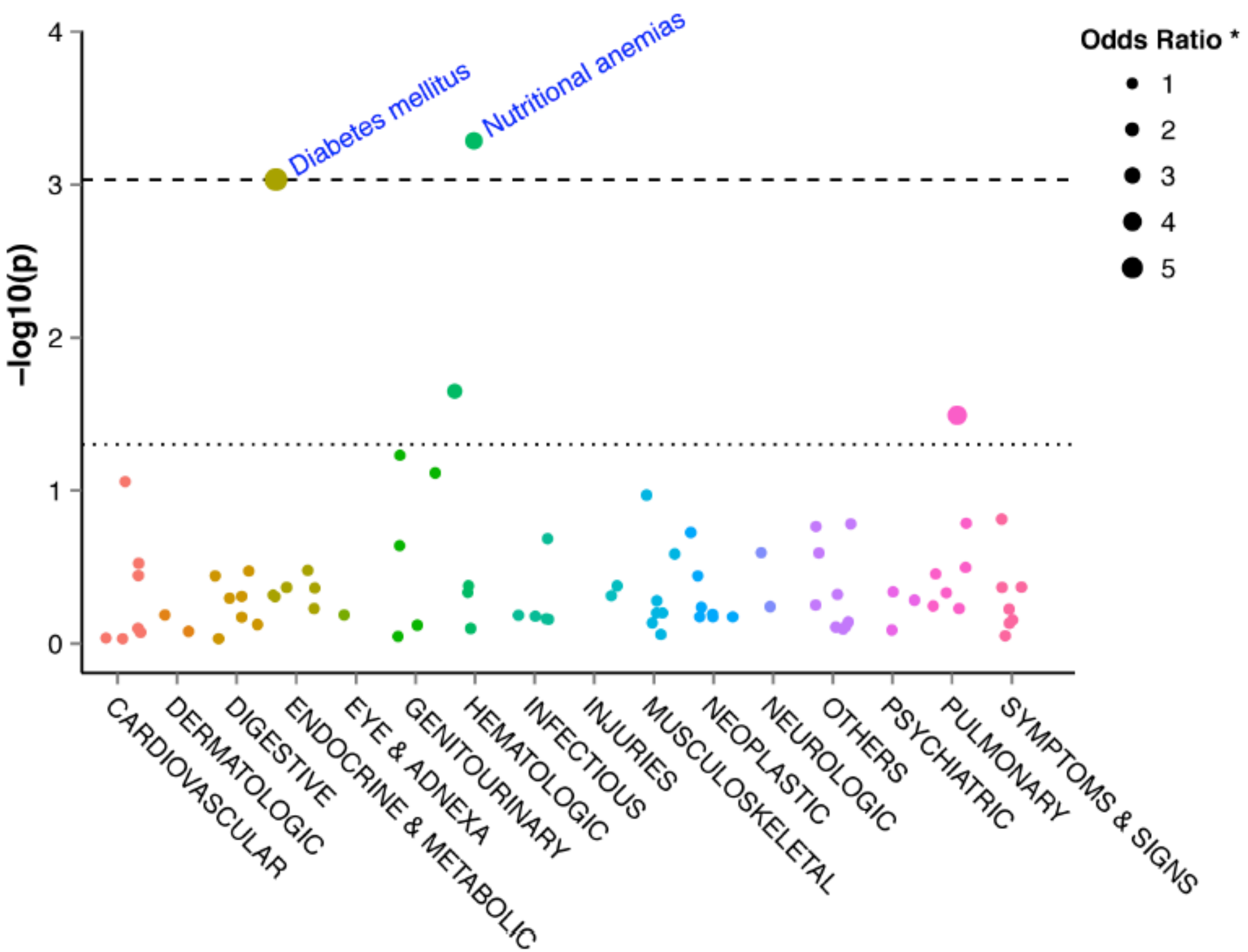
other activity patients
ICD codes/ Biological test results

Very High activity patients
ICD Codes / Biological test results

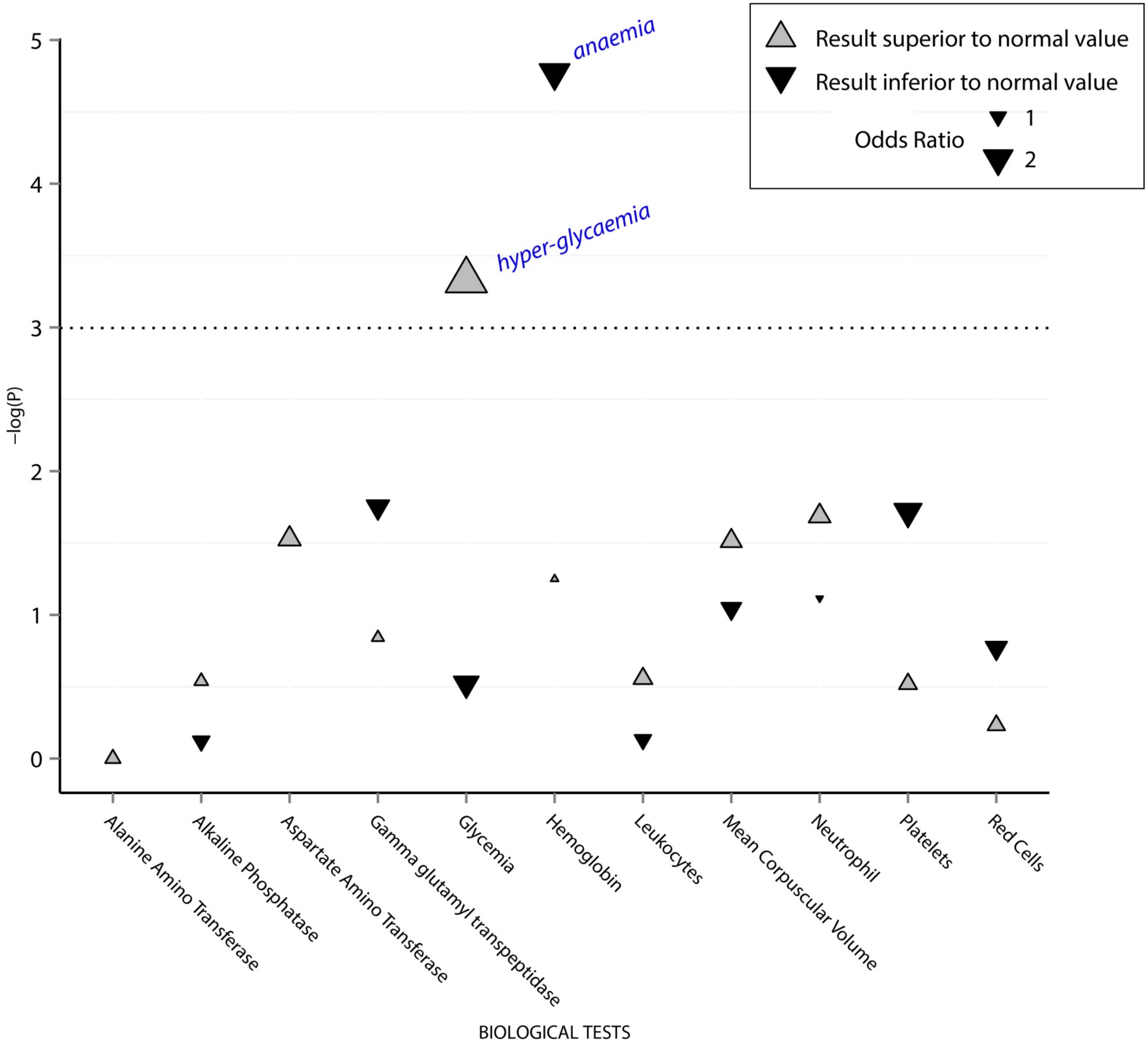
-log(P-value)
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
ICD-10 Codes
-log(P-value)
Biological tests

# Very High TPMT activity *vs* others

**ICD10**

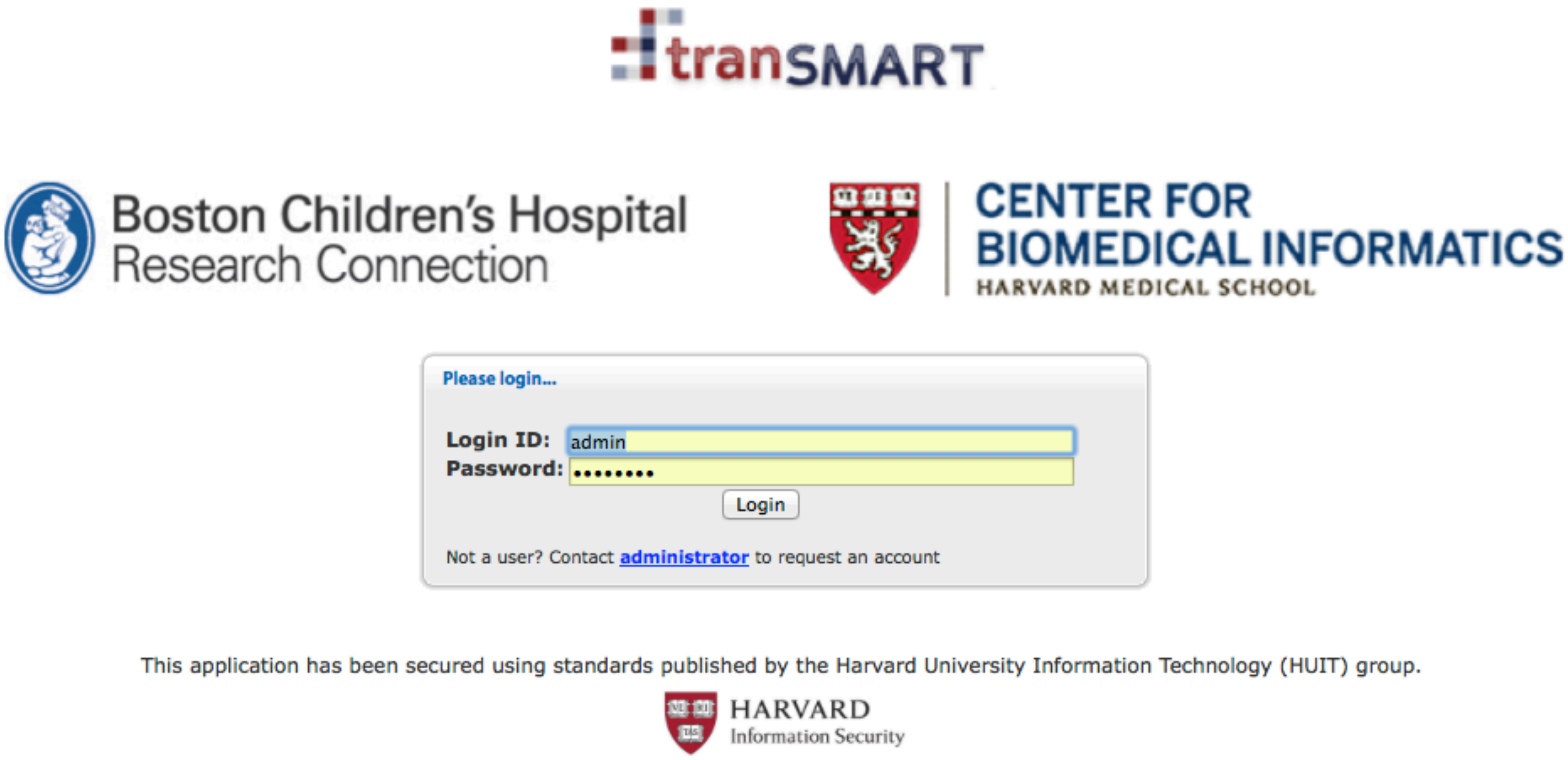


**257 ICD-10 based aggregated**



Result superior to normal value
Result inferior to normal value
Odds Ratio
1
2
anaemia
hyper-glycaemia
5
4
3
2
1
0
−log(P)
Alanine Amino Transferase
Alkaline Phosphatase
Aspartate Amino Transferase
Gamma glutamyl transpeptidase
Glycemia
Hemoglobin
Leukocytes
Mean Corpuscular Volume
Neutrophil
Platelets
Red Cells
BIOLOGICAL TESTS

OOL

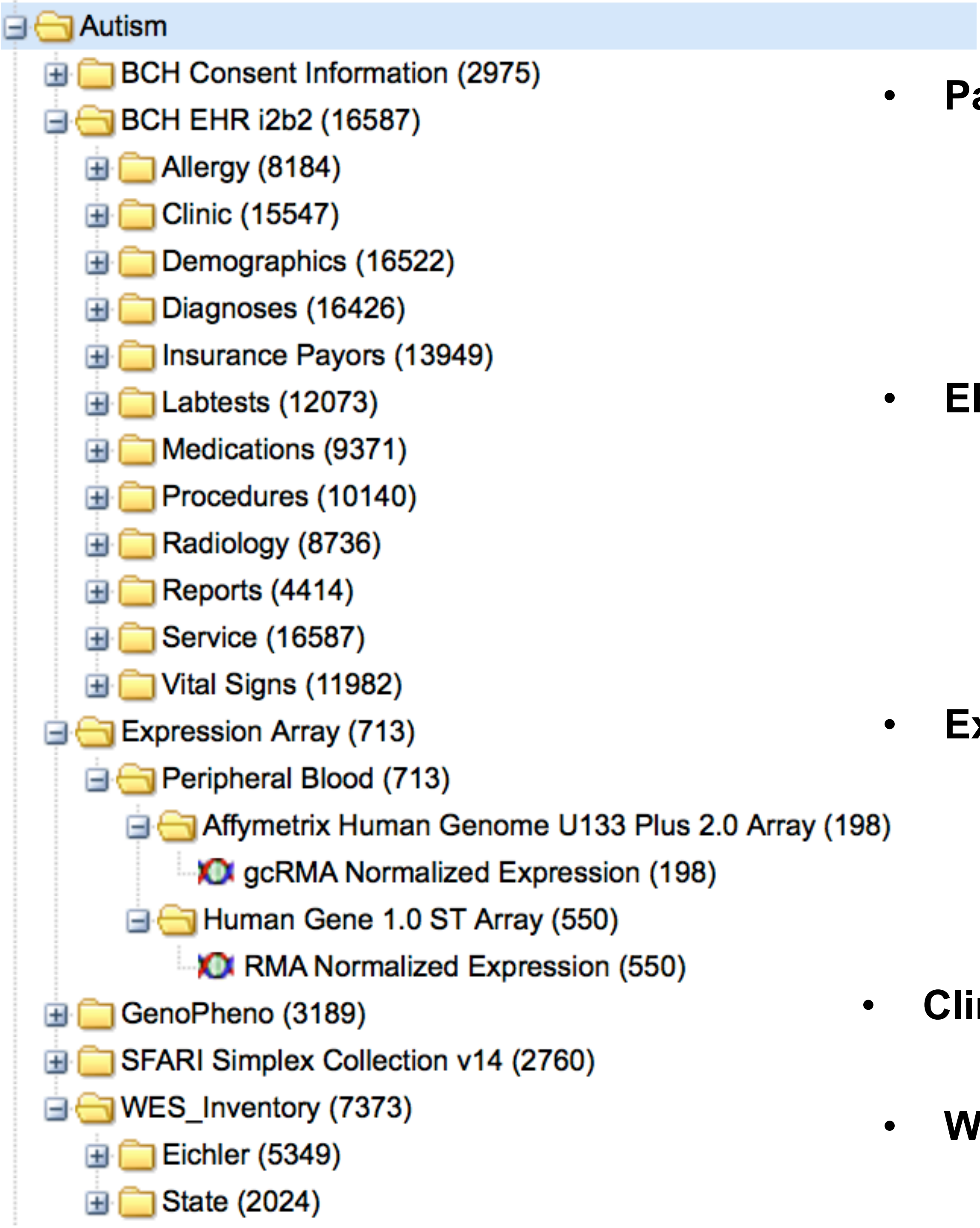

tranSMART
Boston Children's Hospital
Research Connection
CENTER FOR
BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL
Login
HARVARD
Information Security

Please login...

**Login ID:** admin
**Password:** •••••••••

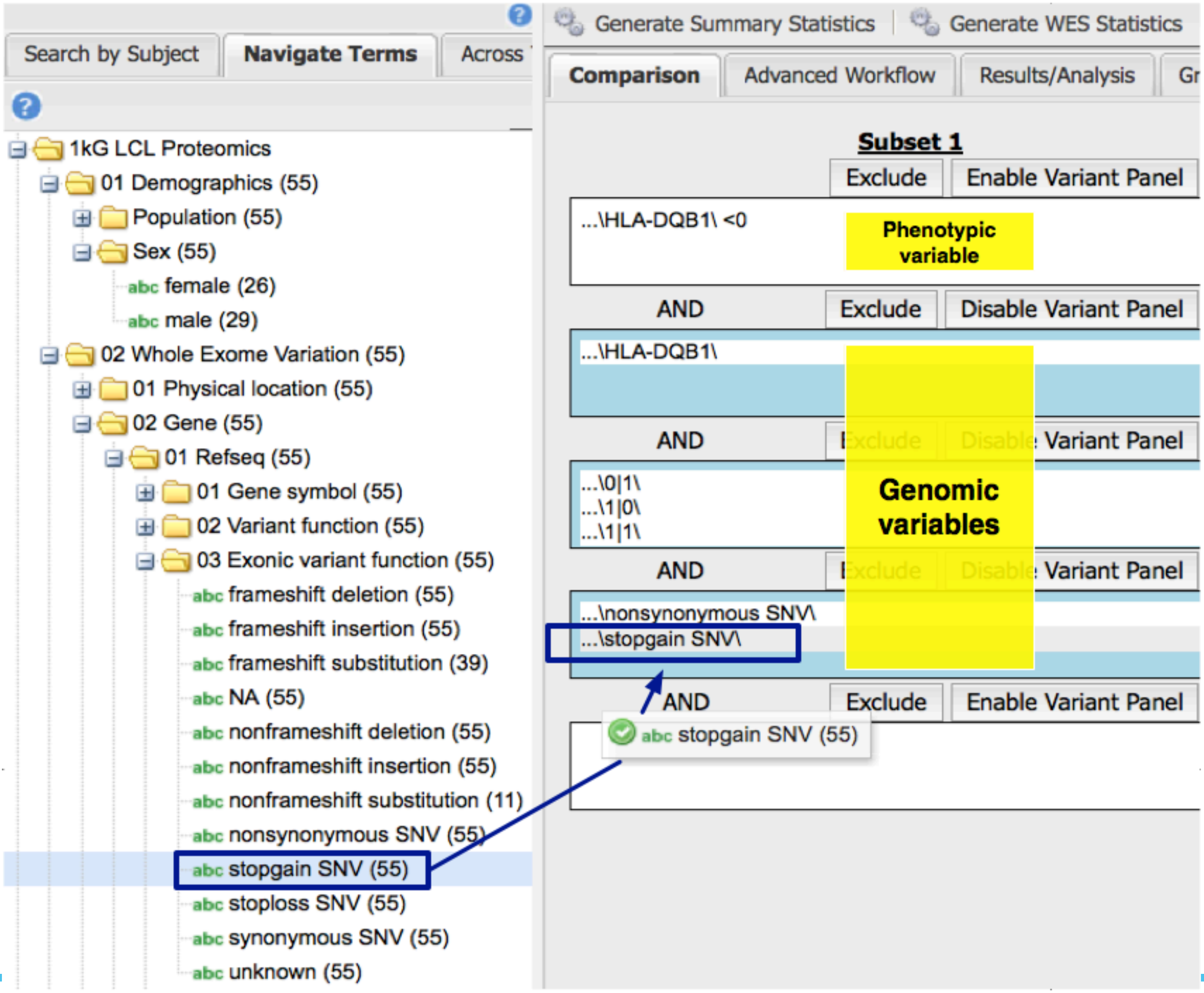
Not a user? Contact **administrator** to request an account

This application has been secured using standards published by the Harvard University Information Technology (HUIT) group.



- Autism
  - BCH Consent Information (2975)
  - BCH EHR i2b2 (16587)
    - Allergy (8184)
    - Clinic (15547)
    - Demographics (16522)
    - Diagnoses (16426)
    - Insurance Payors (13949)
    - Labtests (12073)
    - Medications (9371)
    - Procedures (10140)
    - Radiology (8736)
    - Reports (4414)
    - Service (16587)
    - Vital Signs (11982)
  - Expression Array (713)
    - Peripheral Blood (713)
      - Affymetrix Human Genome U133 Plus 2.0 Array (198)
        - gcRMA Normalized Expression (198)
      - Human Gene 1.0 ST Array (550)
        - RMA Normalized Expression (550)
  - GenoPheno (3189)
  - SFARI Simplex Collection v14 (2760)
  - WES_Inventory (7373)
    - Eichler (5349)
    - State (2024)

- **Patient consent(s)**
- **EHR longitudinal data**
- **Expression Arrays**
- **Clinical Cohorts**
- **WES data**


Search by Subject
Navigate Terms
Across
Generate Summary Statistics
Generate WES Statistics
Comparison
Advanced Workflow
Results/Analysis
Gr

- 1kG LCL Proteomics
  - 01 Demographics (55)
    - Population (55)
    - Sex (55)
      - abc female (26)
      - abc male (29)
  - 02 Whole Exome Variation (55)
    - 01 Physical location (55)
    - 02 Gene (55)
      - 01 Refseq (55)
        - 01 Gene symbol (55)
        - 02 Variant function (55)
        - 03 Exonic variant function (55)
          - abc frameshift deletion (55)
          - abc frameshift insertion (55)
          - abc frameshift substitution (39)
          - abc NA (55)
          - abc nonframeshift deletion (55)
          - abc nonframeshift insertion (55)
          - abc nonframeshift substitution (11)
          - abc nonsynonymous SNV (55)
          - abc stopgain SNV (55)
          - abc stoploss SNV (55)
          - abc synonymous SNV (55)
          - abc unknown (55)

**Subset 1**

Exclude | Enable Variant Panel

...\HLA-DQB1\ <0

**Phenotypic variable**

AND | Exclude | Disable Variant Panel

...\HLA-DQB1\

AND | Exclude | Disable Variant Panel

...\0|1\
...\1|0\
...\1|1\

AND | Exclude | Disable Variant Panel

...\nonsynonymous SNV\
...\stopgain SNV\

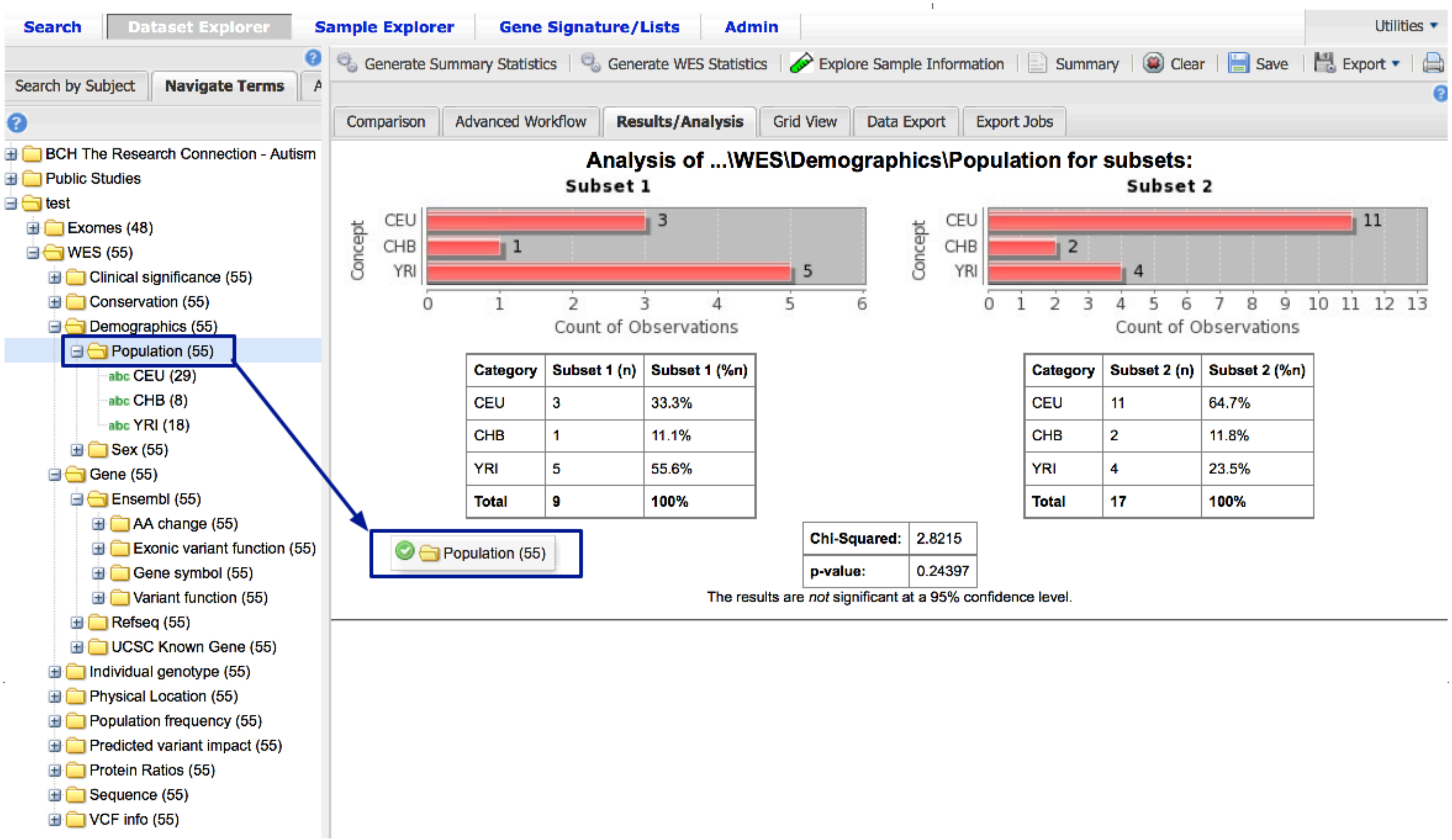
AND | Exclude | Enable Variant Panel

abc stopgain SNV (55)

**Genomic variables**

Boston Children's Hospital
Until every child is well™

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL


Search
Dataset Explorer
Sample Explorer
Gene Signature/Lists
Admin
Utilities
Generate Summary Statistics
Generate WES Statistics
Explore Sample Information
Summary
Clear
Save
Export
Search by Subject
Navigate Terms
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
BCH The Research Connection - Autism
Public Studies
test
Exomes (48)
WES (55)
Clinical significance (55)
Conservation (55)
Demographics (55)
Population (55)
abc CEU (29)
abc CHB (8)
abc YRI (18)
Sex (55)
Gene (55)
Ensembl (55)
AA change (55)
Exonic variant function (55)
Gene symbol (55)
Variant function (55)
Refseq (55)
UCSC Known Gene (55)
Individual genotype (55)
Physical Location (55)
Population frequency (55)
Predicted variant impact (55)
Protein Ratios (55)
Sequence (55)
VCF info (55)
Population (55)
Subset 1
Concept
CEU
CHB
YRI
3
1
5
0
1
2
3
4
5
6
Count of Observations
Subset 2
Concept
CEU
CHB
YRI
11
2
4
0
1
2
3
4
5
6
7
8
9
10
11
12
13
Count of Observations

**Analysis of ...\WES\Demographics\Population for subsets:**

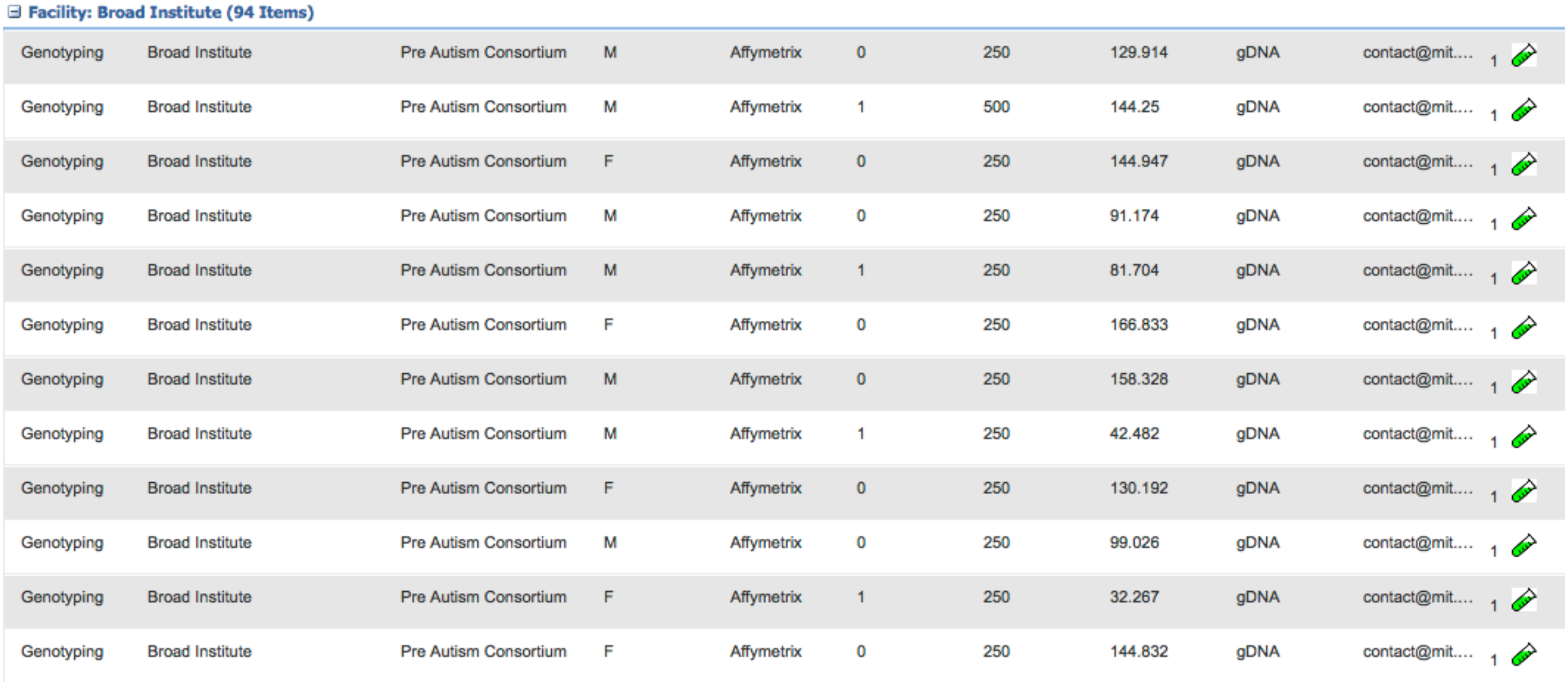
| Category | Subset 1 (n) | Subset 1 (%n) |
|---|---|---|
| CEU | 3 | 33.3% |
| CHB | 1 | 11.1% |
| YRI | 5 | 55.6% |
| **Total** | **9** | **100%** |

| Category | Subset 2 (n) | Subset 2 (%n) |
|---|---|---|
| CEU | 11 | 64.7% |
| CHB | 2 | 11.8% |
| YRI | 4 | 23.5% |
| **Total** | **17** | **100%** |

| | |
|---|---|
| **Chi-Squared:** | 2.8215 |
| **p-value:** | 0.24397 |

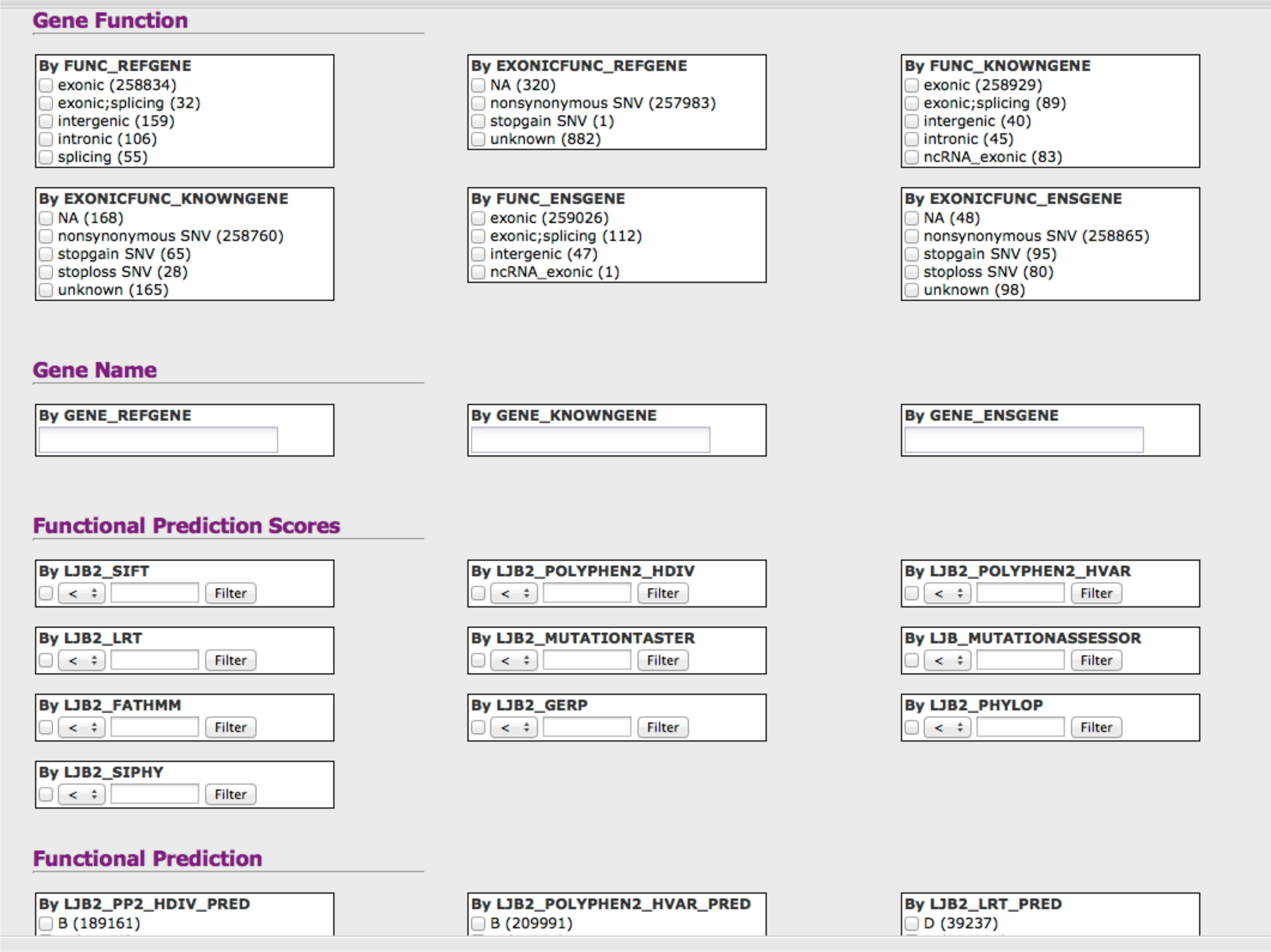
The results are *not* significant at a 95% confidence level.

# Biobank Explorer



**Facility: Broad Institute (94 Items)**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 0 | 250 | 129.914 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 1 | 500 | 144.25 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | F | Affymetrix | 0 | 250 | 144.947 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 0 | 250 | 91.174 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 1 | 250 | 81.704 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | F | Affymetrix | 0 | 250 | 166.833 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 0 | 250 | 158.328 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 1 | 250 | 42.482 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | F | Affymetrix | 0 | 250 | 130.192 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | M | Affymetrix | 0 | 250 | 99.026 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | F | Affymetrix | 1 | 250 | 32.267 | gDNA | contact@mit.... | 1 |
| Genotyping | Broad Institute | Pre Autism Consortium | F | Affymetrix | 0 | 250 | 144.832 | gDNA | contact@mit.... | 1 |



## Gene Function

**By FUNC_REFGENE**
- ☐ exonic (258834)
- ☐ exonic;splicing (32)
- ☐ intergenic (159)
- ☐ intronic (106)
- ☐ splicing (55)

**By EXONICFUNC_REFGENE**
- ☐ NA (320)
- ☐ nonsynonymous SNV (257983)
- ☐ stopgain SNV (1)
- ☐ unknown (882)

**By FUNC_KNOWNGENE**
- ☐ exonic (258929)
- ☐ exonic;splicing (89)
- ☐ intergenic (40)
- ☐ intronic (45)
- ☐ ncRNA_exonic (83)

**By EXONICFUNC_KNOWNGENE**
- ☐ NA (168)
- ☐ nonsynonymous SNV (258760)
- ☐ stopgain SNV (65)
- ☐ stoploss SNV (28)
- ☐ unknown (165)

**By FUNC_ENSGENE**
- ☐ exonic (259026)
- ☐ exonic;splicing (112)
- ☐ intergenic (47)
- ☐ ncRNA_exonic (1)

**By EXONICFUNC_ENSGENE**
- ☐ NA (48)
- ☐ nonsynonymous SNV (258865)
- ☐ stopgain SNV (95)
- ☐ stoploss SNV (80)
- ☐ unknown (98)

## Gene Name

**By GENE_REFGENE**

**By GENE_KNOWNGENE**

**By GENE_ENSGENE**

## Functional Prediction Scores

**By LJB2_SIFT**
☐ < Filter

**By LJB2_POLYPHEN2_HDIV**
☐ < Filter

**By LJB2_POLYPHEN2_HVAR**
☐ < Filter

**By LJB2_LRT**
☐ < Filter

**By LJB2_MUTATIONTASTER**
☐ < Filter

**By LJB_MUTATIONASSESSOR**
☐ < Filter

**By LJB2_FATHMM**
☐ < Filter

**By LJB2_GERP**
☐ < Filter

**By LJB2_PHYLOP**
☐ < Filter

**By LJB2_SIPHY**
☐ < Filter

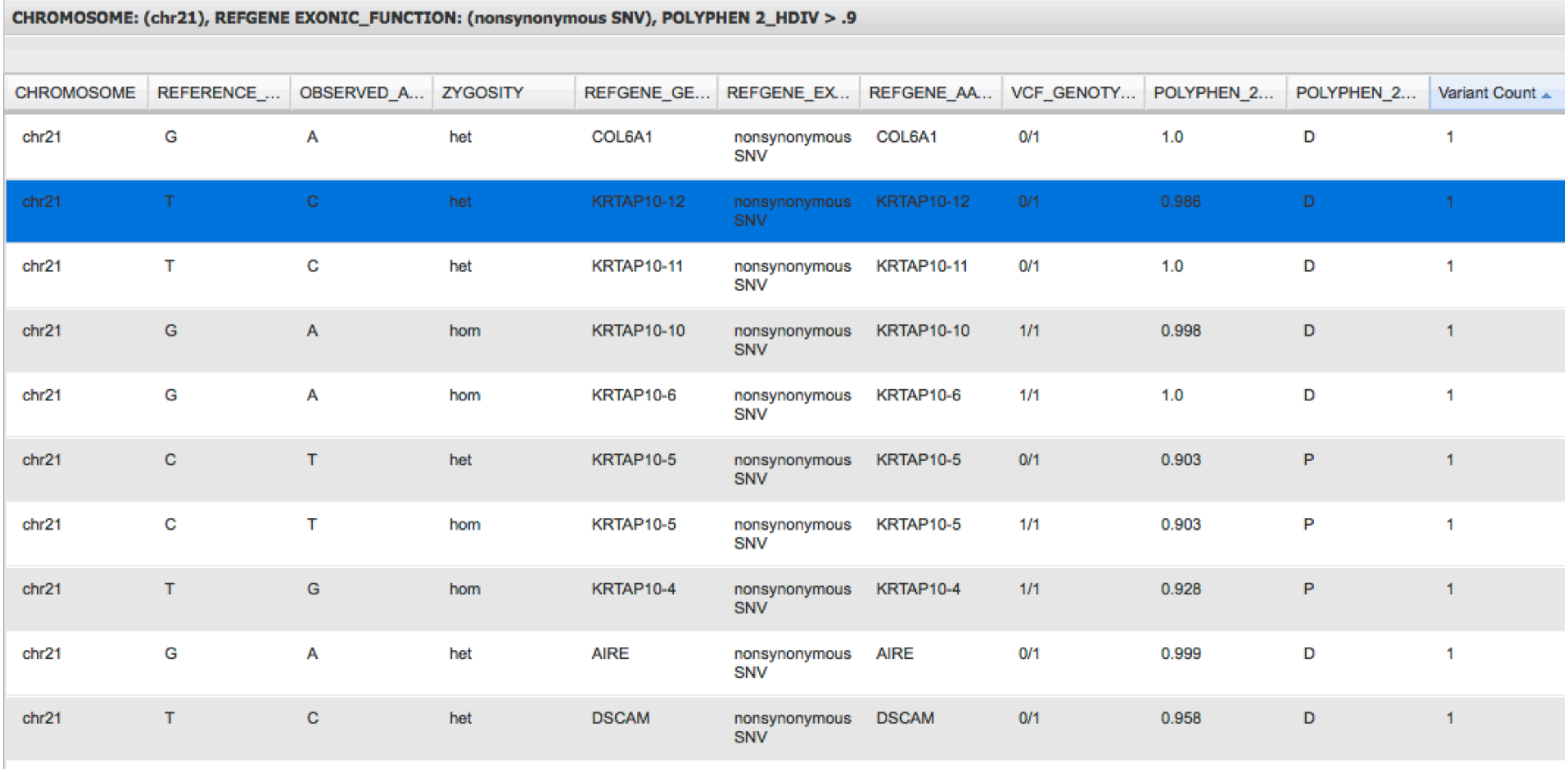
## Functional Prediction

**By LJB2_PP2_HDIV_PRED**
- ☐ B (189161)

**By LJB2_POLYPHEN2_HVAR_PRED**
- ☐ B (209991)

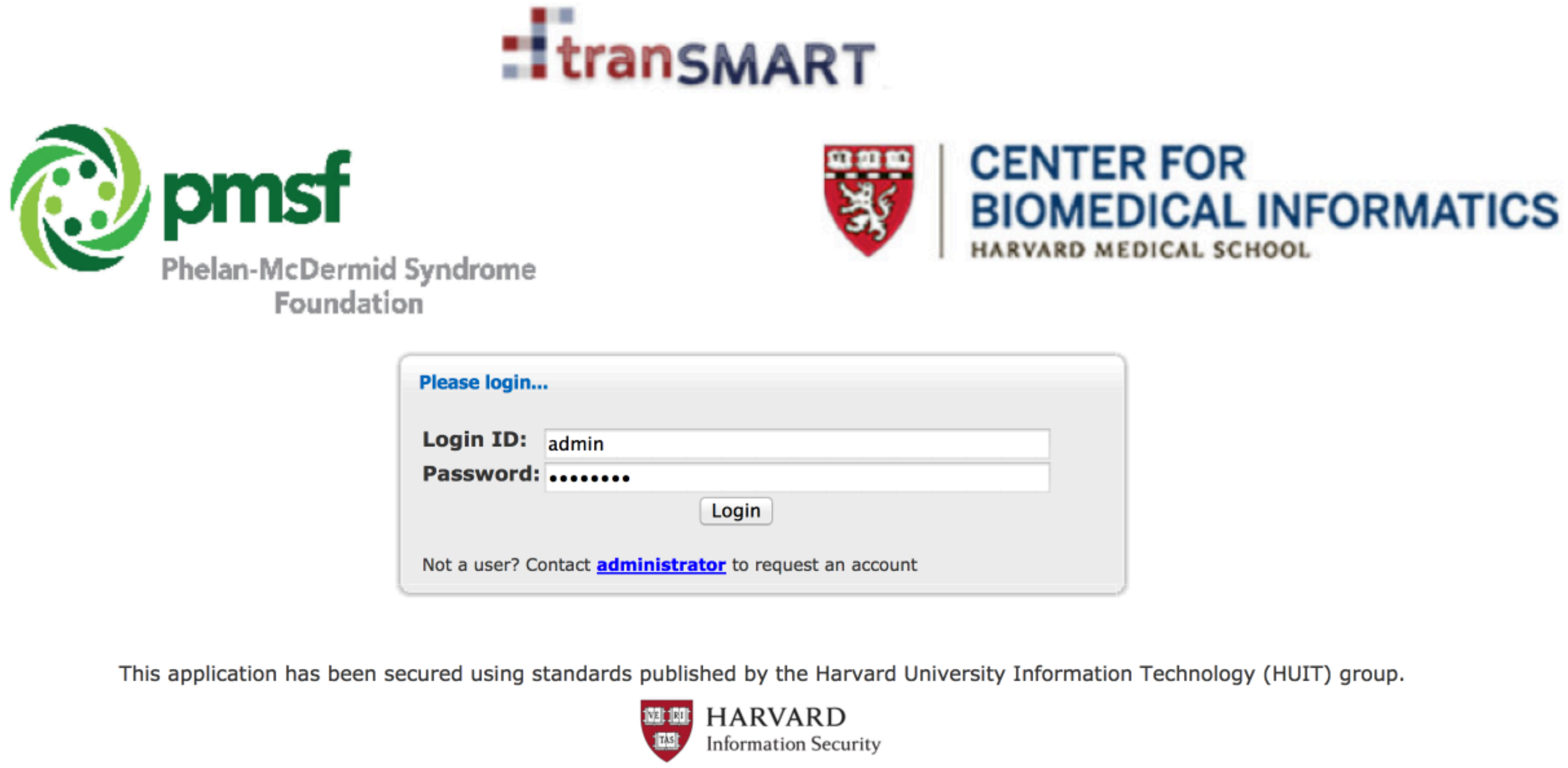
**By LJB2_LRT_PRED**
- ☐ D (39237)

# Genome Variant Explorer



**CHROMOSOME: (chr21), REFGENE EXONIC_FUNCTION: (nonsynonymous SNV), POLYPHEN 2_HDIV > .9**

| CHROMOSOME | REFERENCE_... | OBSERVED_A... | ZYGOSITY | REFGENE_GE... | REFGENE_EX... | REFGENE_AA... | VCF_GENOTY... | POLYPHEN_2... | POLYPHEN_2... | Variant Count ▲ |
|---|---|---|---|---|---|---|---|---|---|---|
| chr21 | G | A | het | COL6A1 | nonsynonymous SNV | COL6A1 | 0/1 | 1.0 | D | 1 |
| chr21 | T | C | het | KRTAP10-12 | nonsynonymous SNV | KRTAP10-12 | 0/1 | 0.986 | D | 1 |
| chr21 | T | C | het | KRTAP10-11 | nonsynonymous SNV | KRTAP10-11 | 0/1 | 1.0 | D | 1 |
| chr21 | G | A | hom | KRTAP10-10 | nonsynonymous SNV | KRTAP10-10 | 1/1 | 0.998 | D | 1 |
| chr21 | G | A | hom | KRTAP10-6 | nonsynonymous SNV | KRTAP10-6 | 1/1 | 1.0 | D | 1 |
| chr21 | C | T | het | KRTAP10-5 | nonsynonymous SNV | KRTAP10-5 | 0/1 | 0.903 | P | 1 |
| chr21 | C | T | hom | KRTAP10-5 | nonsynonymous SNV | KRTAP10-5 | 1/1 | 0.903 | P | 1 |
| chr21 | T | G | hom | KRTAP10-4 | nonsynonymous SNV | KRTAP10-4 | 1/1 | 0.928 | P | 1 |
| chr21 | G | A | het | AIRE | nonsynonymous SNV | AIRE | 0/1 | 0.999 | D | 1 |
| chr21 | T | C | het | DSCAM | nonsynonymous SNV | DSCAM | 0/1 | 0.958 | D | 1 |



tranSMART
pmsf
Phelan-McDermid Syndrome Foundation
CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

**Please login...**

**Login ID:** admin

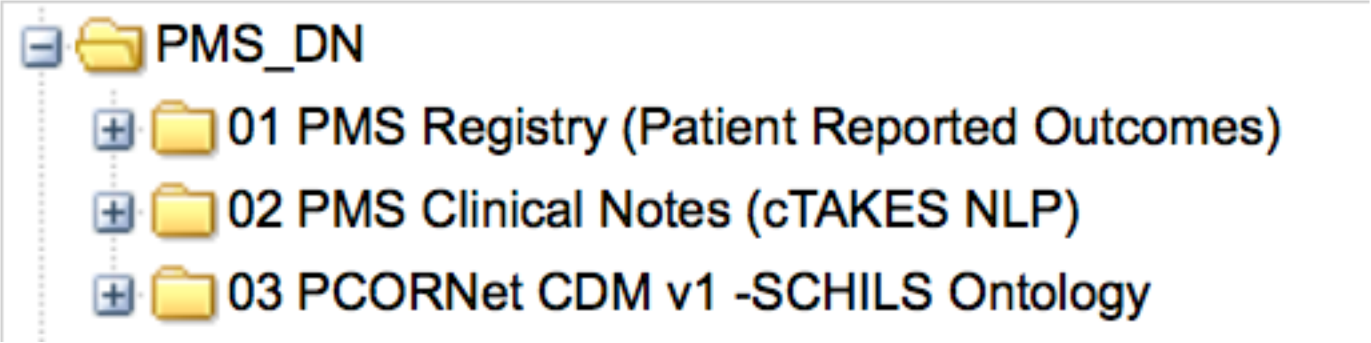
**Password:** ••••••••

Login

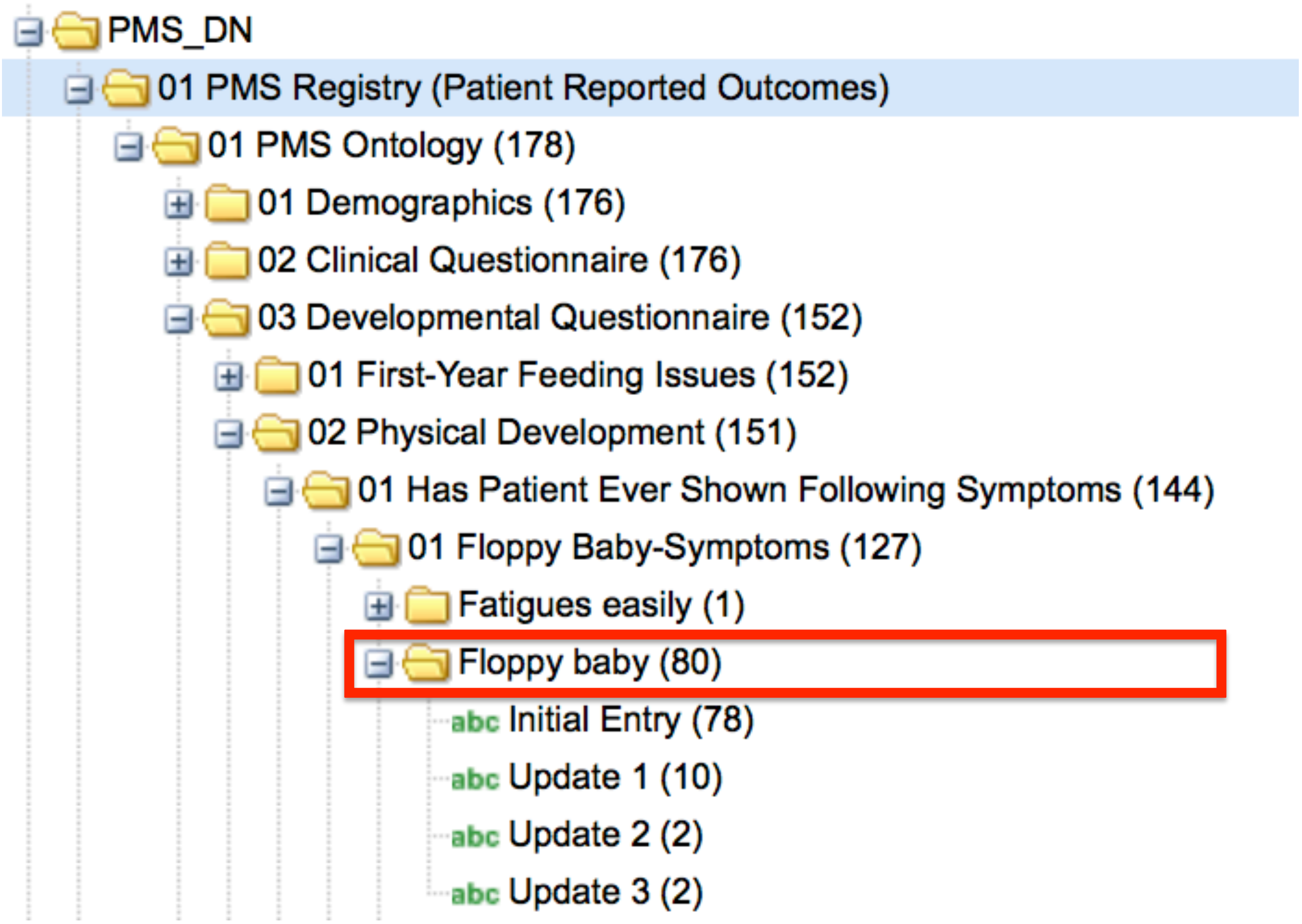
Not a user? Contact **administrator** to request an account

This application has been secured using standards published by the Harvard University Information Technology (HUIT) group.

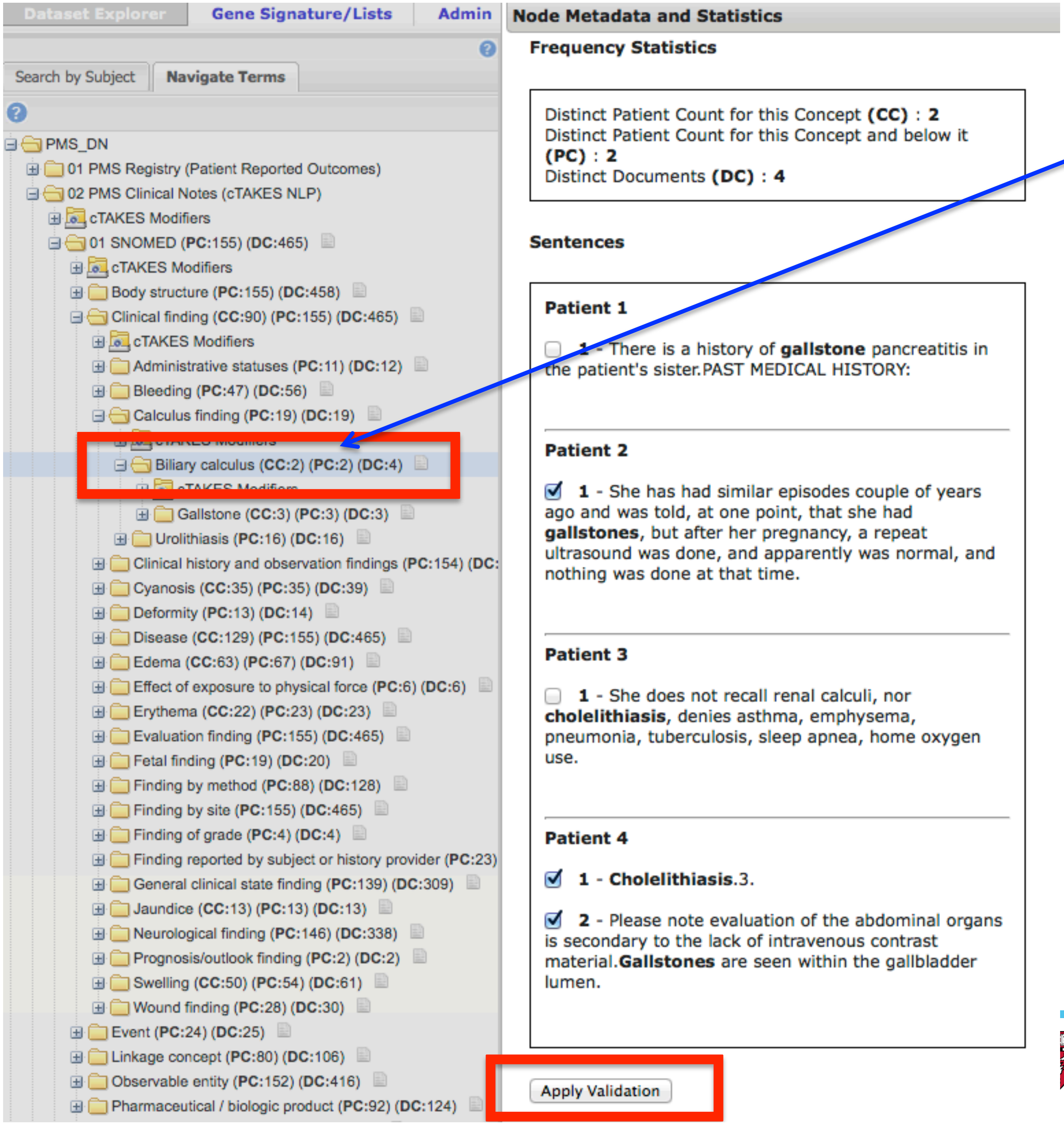
HARVARD
Information Security



- PMS_DN
  - 01 PMS Registry (Patient Reported Outcomes)
  - 02 PMS Clinical Notes (cTAKES NLP)
  - 03 PCORNet CDM v1 -SCHILS Ontology



- PMS_DN
  - 01 PMS Registry (Patient Reported Outcomes)
    - 01 PMS Ontology (178)
      - 01 Demographics (176)
      - 02 Clinical Questionnaire (176)
      - 03 Developmental Questionnaire (152)
        - 01 First-Year Feeding Issues (152)
        - 02 Physical Development (151)
          - 01 Has Patient Ever Shown Following Symptoms (144)
            - 01 Floppy Baby-Symptoms (127)
              - Fatigues easily (1)
              - Floppy baby (80)
                - abc Initial Entry (78)
                - abc Update 1 (10)
                - abc Update 2 (2)
                - abc Update 3 (2)



- PMS_DN
  - 01 PMS Registry (Patient Reported Outcomes)
  - 02 PMS Clinical Notes (cTAKES NLP)
    - cTAKES Modifiers
    - 01 SNOMED (**PC**:155) (**DC**:465)
      - cTAKES Modifiers
      - Body structure (**PC**:155) (**DC**:458)
      - Clinical finding (**CC**:90) (**PC**:155) (**DC**:465)
        - cTAKES Modifiers
        - Administrative statuses (**PC**:11) (**DC**:12)
        - Bleeding (**PC**:47) (**DC**:56)
        - Calculus finding (**PC**:19) (**DC**:19)
          - cTAKES Modifiers
          - Biliary calculus (**CC**:2) (**PC**:2) (**DC**:4)
            - cTAKES Modifiers
            - Gallstone (**CC**:3) (**PC**:3) (**DC**:3)
          - Urolithiasis (**PC**:16) (**DC**:16)
        - Clinical history and observation findings (**PC**:154) (**DC**:
        - Cyanosis (**CC**:35) (**PC**:35) (**DC**:39)
        - Deformity (**PC**:13) (**DC**:14)
        - Disease (**CC**:129) (**PC**:155) (**DC**:465)
        - Edema (**CC**:63) (**PC**:67) (**DC**:91)
        - Effect of exposure to physical force (**PC**:6) (**DC**:6)
        - Erythema (**CC**:22) (**PC**:23) (**DC**:23)
        - Evaluation finding (**PC**:155) (**DC**:465)
        - Fetal finding (**PC**:19) (**DC**:20)
        - Finding by method (**PC**:88) (**DC**:128)
        - Finding by site (**PC**:155) (**DC**:465)
        - Finding of grade (**PC**:4) (**DC**:4)
        - Finding reported by subject or history provider (**PC**:23)
        - General clinical state finding (**PC**:139) (**DC**:309)
        - Jaundice (**CC**:13) (**PC**:13) (**DC**:13)
        - Neurological finding (**PC**:146) (**DC**:338)
        - Prognosis/outlook finding (**PC**:2) (**DC**:2)
        - Swelling (**CC**:50) (**PC**:54) (**DC**:61)
        - Wound finding (**PC**:28) (**DC**:30)
      - Event (**PC**:24) (**DC**:25)
      - Linkage concept (**PC**:80) (**DC**:106)
      - Observable entity (**PC**:152) (**DC**:416)
      - Pharmaceutical / biologic product (**PC**:92) (**DC**:124)

**Node Metadata and Statistics**

**Frequency Statistics**

Distinct Patient Count for this Concept **(CC)** : **2**
Distinct Patient Count for this Concept and below it **(PC)** : **2**
Distinct Documents **(DC)** : **4**

**Sentences**

**Patient 1**

☐ **1** - There is a history of **gallstone** pancreatitis in the patient's sister.PAST MEDICAL HISTORY:

**Patient 2**

☑ **1** - She has had similar episodes couple of years ago and was told, at one point, that she had **gallstones**, but after her pregnancy, a repeat ultrasound was done, and apparently was normal, and nothing was done at that time.

**Patient 3**

☐ **1** - She does not recall renal calculi, nor **cholelithiasis**, denies asthma, emphysema, pneumonia, tuberculosis, sleep apnea, home oxygen use.

**Patient 4**

☑ **1** - **Cholelithiasis**.3.

☑ **2** - Please note evaluation of the abdominal organs is secondary to the lack of intravenous contrast material.**Gallstones** are seen within the gallbladder lumen.

**2 patients with gallstones**

*No: It's Sister's history*

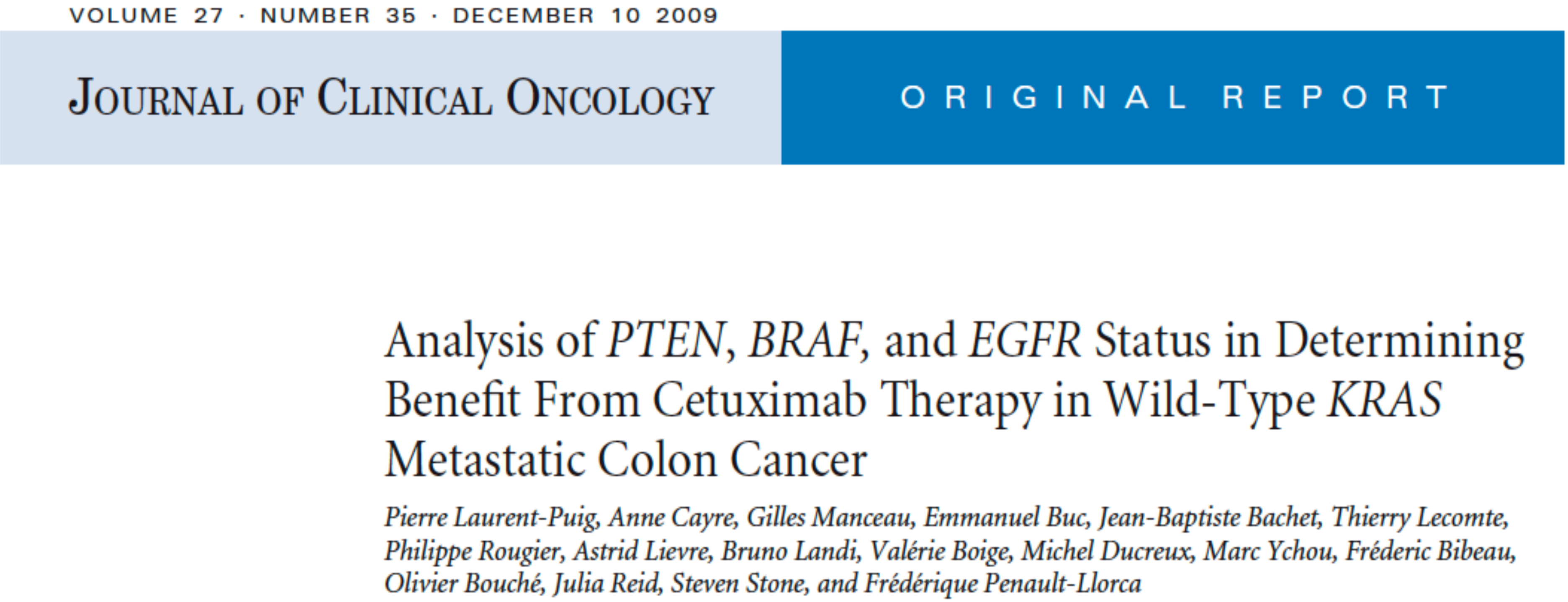
*OK*

*No it's a Negation*

*OK*

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

# Pilot Study
# HEGP - Paris



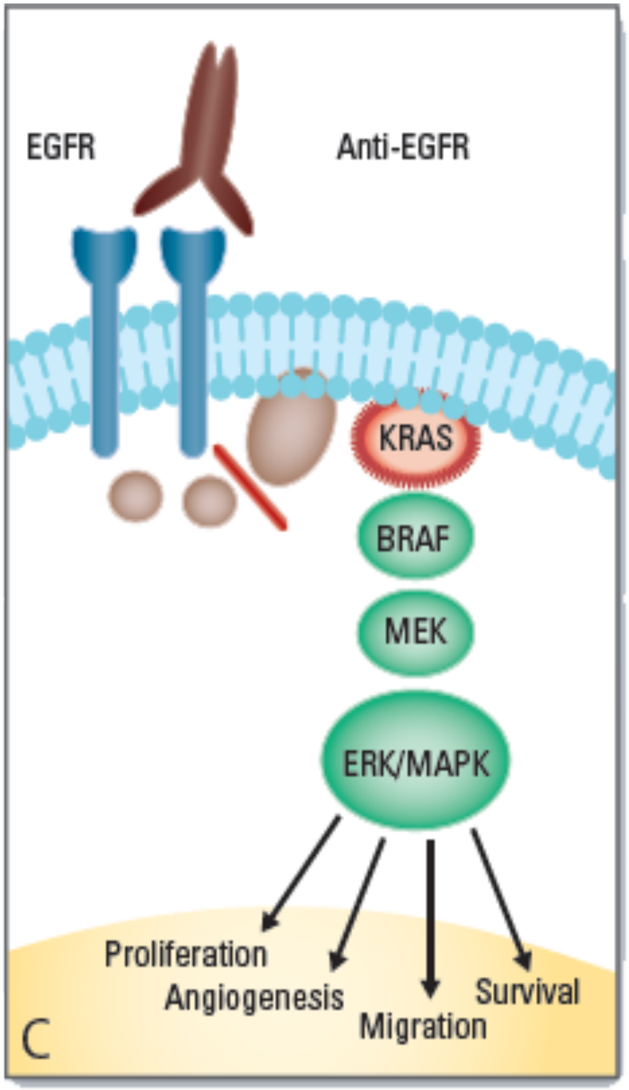
VOLUME 27 · NUMBER 35 · DECEMBER 10 2009

JOURNAL OF CLINICAL ONCOLOGY

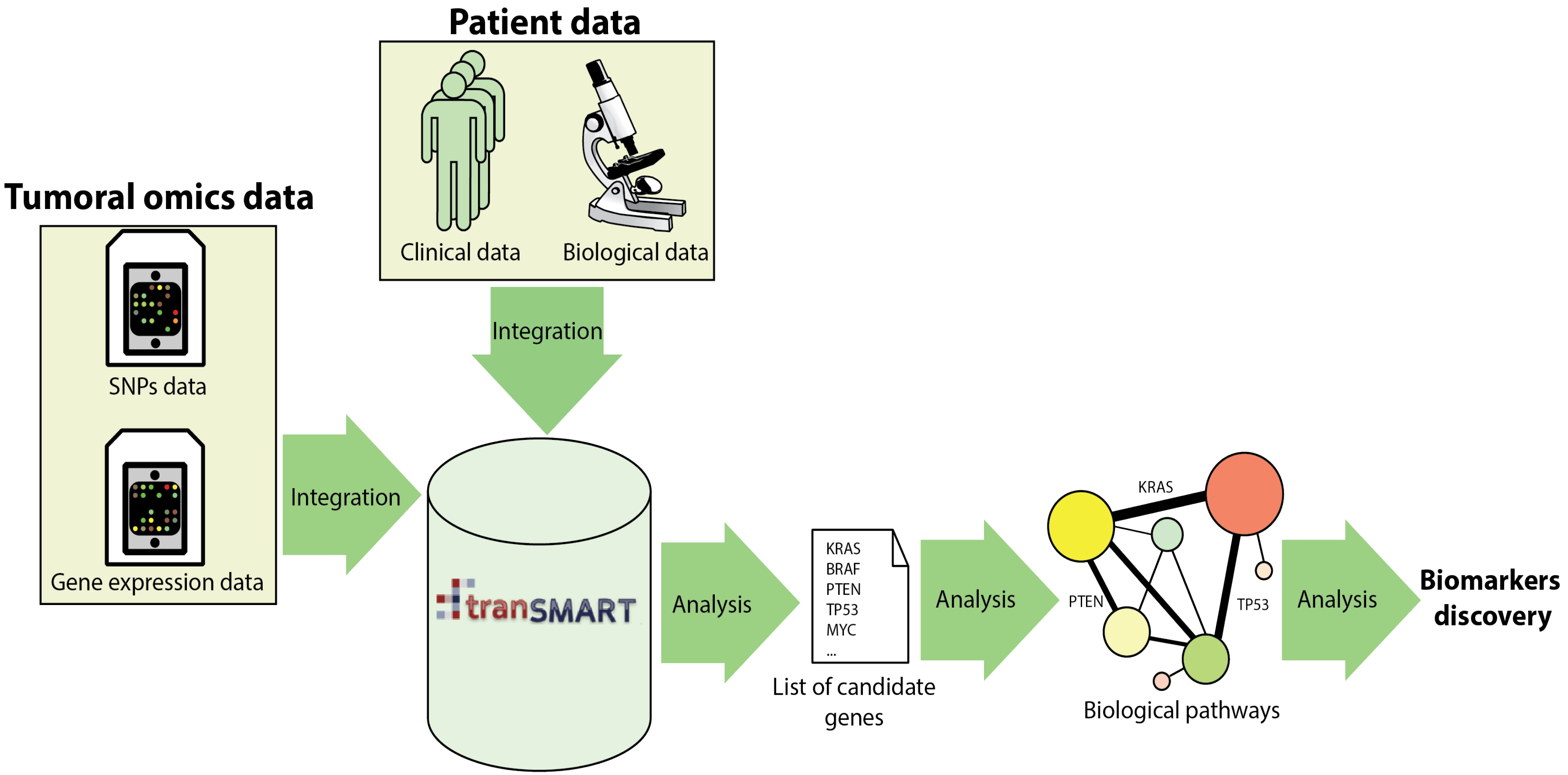
ORIGINAL REPORT

## 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, Fréderic Bibeau, Olivier Bouché, Julia Reid, Steven Stone, and Frédérique Penault-Llorca*


EGFR
Anti-EGFR
KRAS
BRAF
MEK
ERK/MAPK
Proliferation
Angiogenesis
Migration
Survival
C

tranSMART



**Tumoral omics data**

SNPs data

Gene expression data

Integration

**Patient data**

Clinical data

Biological data

Integration

tranSMART

Analysis

KRAS
BRAF
PTEN
TP53
MYC
...

List of candidate genes

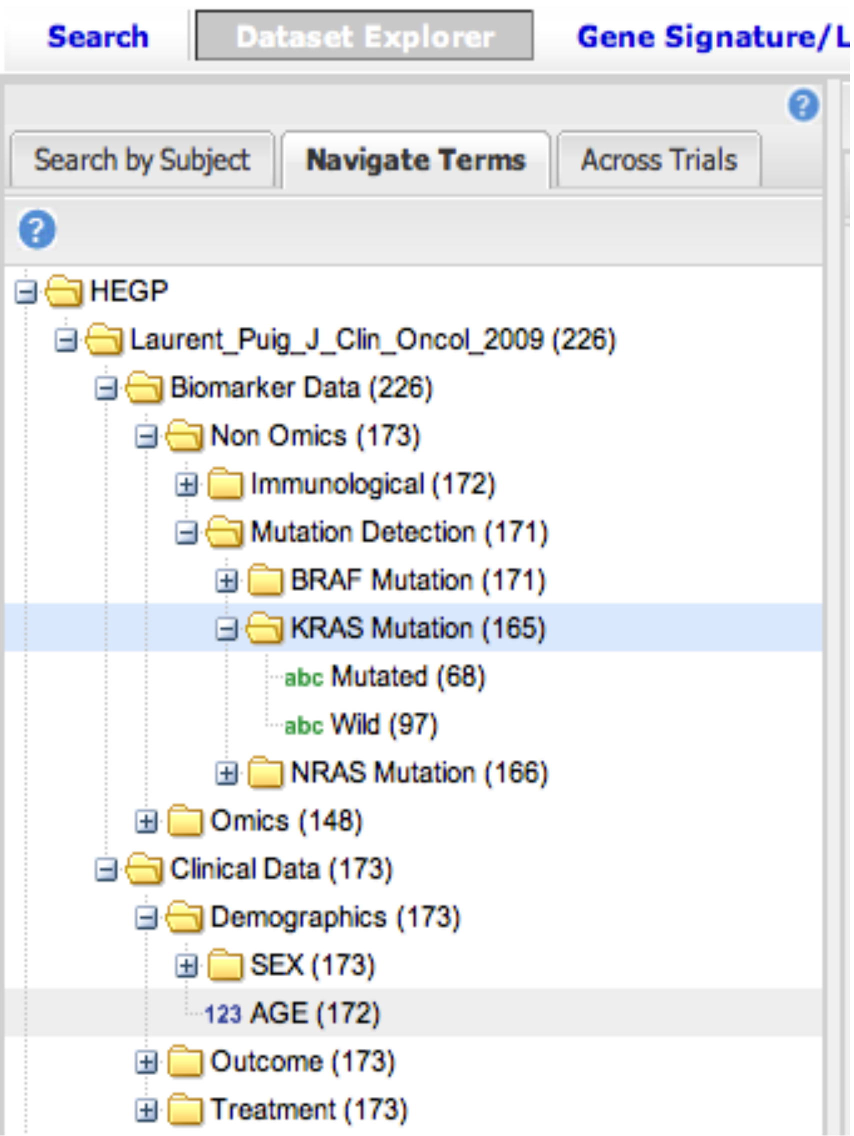
Analysis

Biological pathways

Analysis

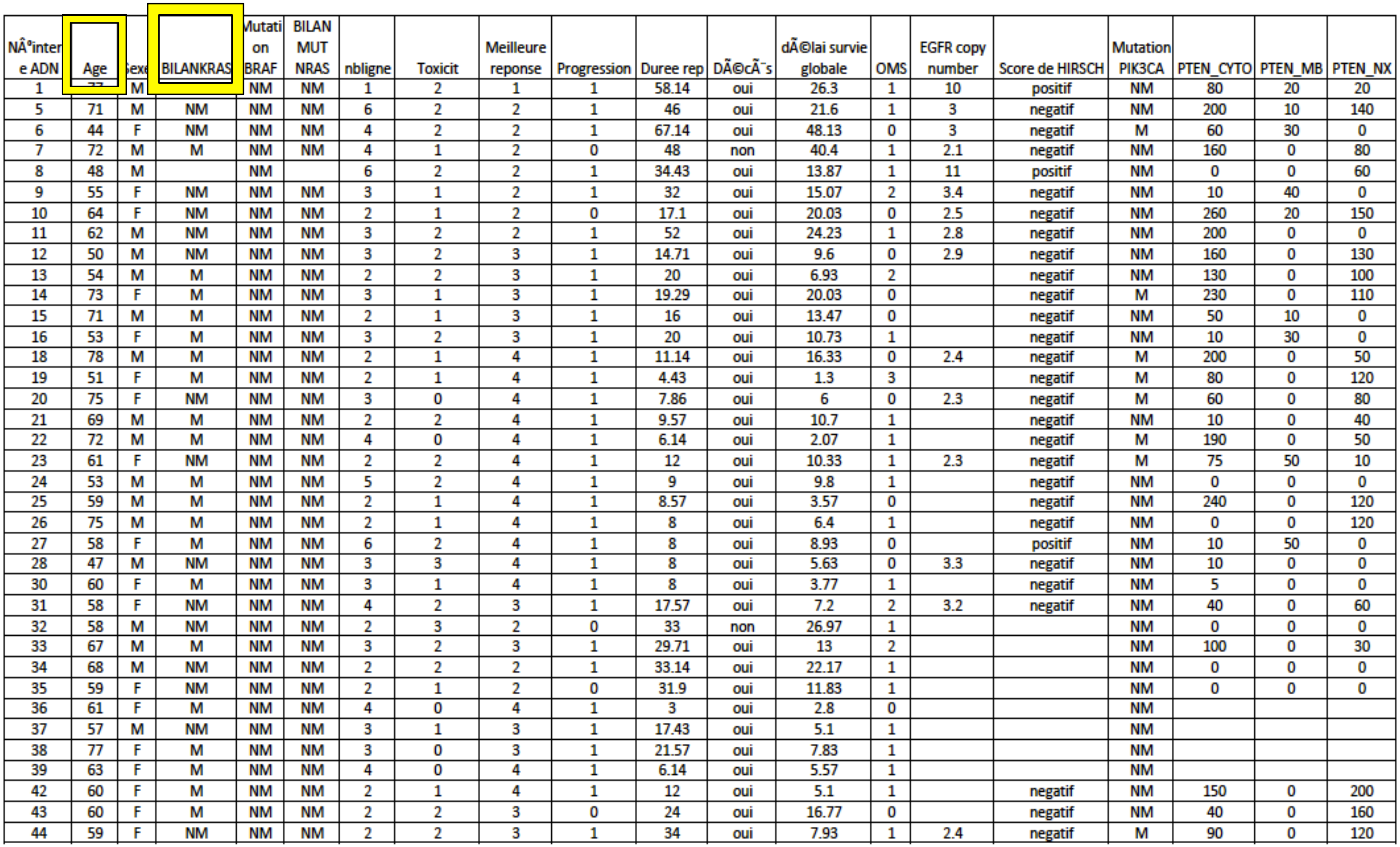
**Biomarkers discovery**

HEGP: Canuel V, Avillach P

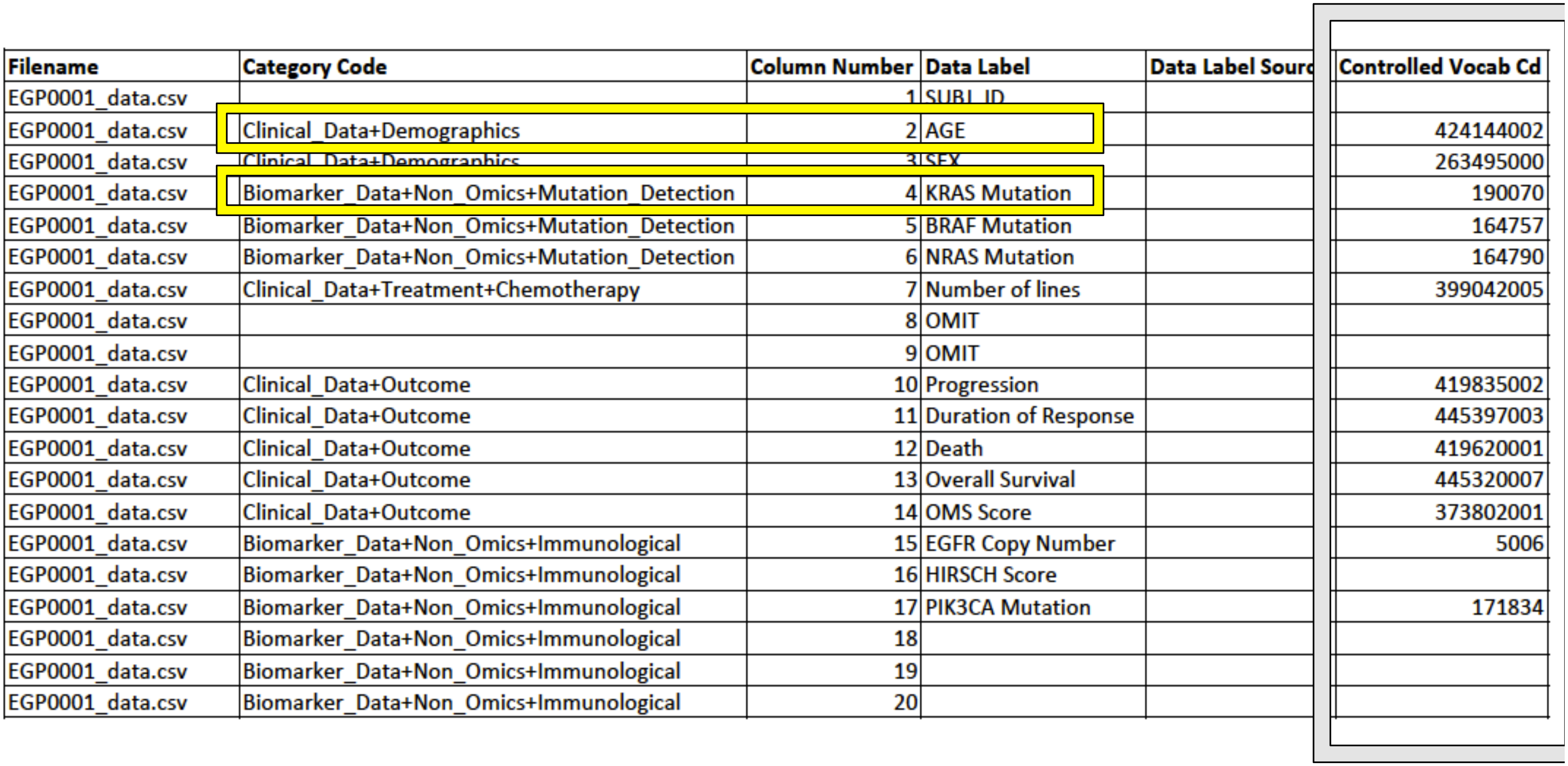

Search
Dataset Explorer
Gene Signature/L
Search by Subject
Navigate Terms
Across Trials

- HEGP
  - Laurent_Puig_J_Clin_Oncol_2009 (226)
    - Biomarker Data (226)
      - Non Omics (173)
        - Immunological (172)
        - Mutation Detection (171)
          - BRAF Mutation (171)
          - KRAS Mutation (165)
            - abc Mutated (68)
            - abc Wild (97)
          - NRAS Mutation (166)
      - Omics (148)
    - Clinical Data (173)
      - Demographics (173)
        - SEX (173)
        - 123 AGE (172)
      - Outcome (173)
      - Treatment (173)

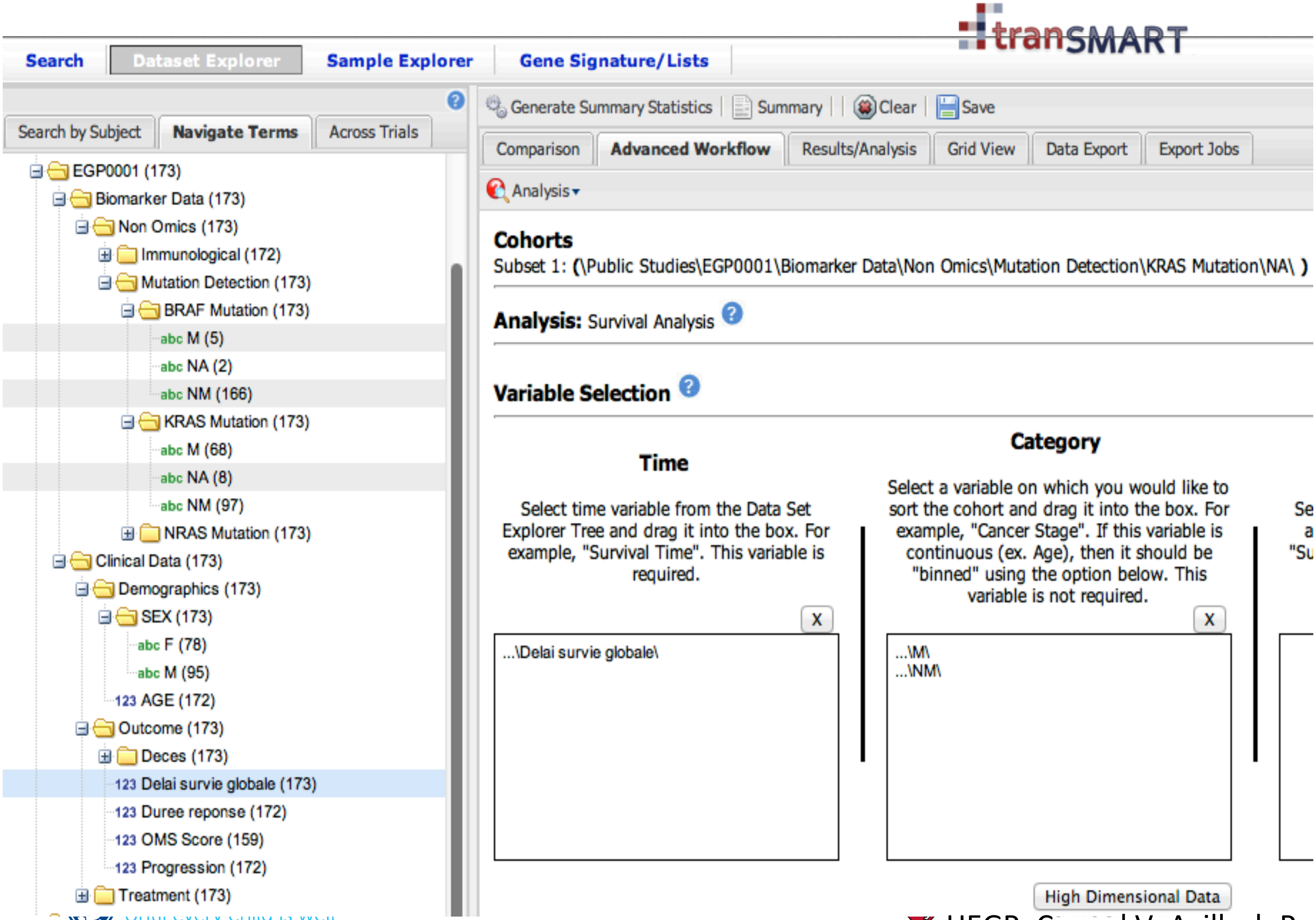
HEGP: Canuel V, Avillach P



| NÂ°interne ADN | Age | exe | BILANKRAS | Mutation BRAF | BILAN MUT NRAS | nbligne | Toxicit | Meilleure reponse | Progression | Duree rep | DÃ©cÃ¨s | dÃ©lai survie globale | OMS | EGFR copy number | Score de HIRSCH | Mutation PIK3CA | PTEN_CYTO | PTEN_MB | PTEN_NX |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | [illegible] | M | [illegible] | NM | NM | 1 | 2 | 1 | 1 | 58.14 | oui | 26.3 | 1 | 10 | positif | NM | 80 | 20 | 20 |
| 5 | 71 | M | NM | NM | NM | 6 | 2 | 2 | 1 | 46 | oui | 21.6 | 1 | 3 | negatif | NM | 200 | 10 | 140 |
| 6 | 44 | F | NM | NM | NM | 4 | 2 | 2 | 1 | 67.14 | oui | 48.13 | 0 | 3 | negatif | M | 60 | 30 | 0 |
| 7 | 72 | M | M | NM | NM | 4 | 1 | 2 | 0 | 48 | non | 40.4 | 1 | 2.1 | negatif | NM | 160 | 0 | 80 |
| 8 | 48 | M |  | NM |  | 6 | 2 | 2 | 1 | 34.43 | oui | 13.87 | 1 | 11 | positif | NM | 0 | 0 | 60 |
| 9 | 55 | F | NM | NM | NM | 3 | 1 | 2 | 1 | 32 | oui | 15.07 | 2 | 3.4 | negatif | NM | 10 | 40 | 0 |
| 10 | 64 | F | NM | NM | NM | 2 | 1 | 2 | 0 | 17.1 | oui | 20.03 | 0 | 2.5 | negatif | NM | 260 | 20 | 150 |
| 11 | 62 | M | NM | NM | NM | 3 | 2 | 2 | 1 | 52 | oui | 24.23 | 1 | 2.8 | negatif | NM | 200 | 0 | 0 |
| 12 | 50 | M | NM | NM | NM | 3 | 2 | 3 | 1 | 14.71 | oui | 9.6 | 0 | 2.9 | negatif | NM | 160 | 0 | 130 |
| 13 | 54 | M | M | NM | NM | 2 | 2 | 3 | 1 | 20 | oui | 6.93 | 2 |  | negatif | NM | 130 | 0 | 100 |
| 14 | 73 | F | M | NM | NM | 3 | 1 | 3 | 1 | 19.29 | oui | 20.03 | 0 |  | negatif | M | 230 | 0 | 110 |
| 15 | 71 | M | M | NM | NM | 2 | 1 | 3 | 1 | 16 | oui | 13.47 | 0 |  | negatif | NM | 50 | 10 | 0 |
| 16 | 53 | F | M | NM | NM | 3 | 2 | 3 | 1 | 20 | oui | 10.73 | 1 |  | negatif | NM | 10 | 30 | 0 |
| 18 | 78 | M | M | NM | NM | 2 | 1 | 4 | 1 | 11.14 | oui | 16.33 | 0 | 2.4 | negatif | M | 200 | 0 | 50 |
| 19 | 51 | F | M | NM | NM | 2 | 1 | 4 | 1 | 4.43 | oui | 1.3 | 3 |  | negatif | M | 80 | 0 | 120 |
| 20 | 75 | F | NM | NM | NM | 3 | 0 | 4 | 1 | 7.86 | oui | 6 | 0 | 2.3 | negatif | M | 60 | 0 | 80 |
| 21 | 69 | M | M | NM | NM | 2 | 2 | 4 | 1 | 9.57 | oui | 10.7 | 1 |  | negatif | NM | 10 | 0 | 40 |
| 22 | 72 | M | M | NM | NM | 4 | 0 | 4 | 1 | 6.14 | oui | 2.07 | 1 |  | negatif | M | 190 | 0 | 50 |
| 23 | 61 | F | NM | NM | NM | 2 | 2 | 4 | 1 | 12 | oui | 10.33 | 1 | 2.3 | negatif | M | 75 | 50 | 10 |
| 24 | 53 | M | M | NM | NM | 5 | 2 | 4 | 1 | 9 | oui | 9.8 | 1 |  | negatif | NM | 0 | 0 | 0 |
| 25 | 59 | M | M | NM | NM | 2 | 1 | 4 | 1 | 8.57 | oui | 3.57 | 0 |  | negatif | NM | 240 | 0 | 120 |
| 26 | 75 | M | M | NM | NM | 2 | 1 | 4 | 1 | 8 | oui | 6.4 | 1 |  | negatif | NM | 0 | 0 | 120 |
| 27 | 58 | F | M | NM | NM | 6 | 2 | 4 | 1 | 8 | oui | 8.93 | 0 |  | positif | NM | 10 | 50 | 0 |
| 28 | 47 | M | NM | NM | NM | 3 | 3 | 4 | 1 | 8 | oui | 5.63 | 0 | 3.3 | negatif | NM | 10 | 0 | 0 |
| 30 | 60 | F | M | NM | NM | 3 | 1 | 4 | 1 | 8 | oui | 3.77 | 1 |  | negatif | NM | 5 | 0 | 0 |
| 31 | 58 | F | NM | NM | NM | 4 | 2 | 3 | 1 | 17.57 | oui | 7.2 | 2 | 3.2 | negatif | NM | 40 | 0 | 60 |
| 32 | 58 | M | NM | NM | NM | 2 | 3 | 2 | 0 | 33 | non | 26.97 | 1 |  |  | NM | 0 | 0 | 0 |
| 33 | 67 | M | M | NM | NM | 3 | 2 | 3 | 1 | 29.71 | oui | 13 | 2 |  |  | NM | 100 | 0 | 30 |
| 34 | 68 | M | NM | NM | NM | 2 | 2 | 2 | 1 | 33.14 | oui | 22.17 | 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 | F | M | NM | NM | 4 | 0 | 4 | 1 | 3 | oui | 2.8 | 0 |  |  | NM |  |  |  |
| 37 | 57 | M | NM | NM | NM | 3 | 1 | 3 | 1 | 17.43 | oui | 5.1 | 1 |  |  | NM |  |  |  |
| 38 | 77 | F | M | NM | NM | 3 | 0 | 3 | 1 | 21.57 | oui | 7.83 | 1 |  |  | NM |  |  |  |
| 39 | 63 | F | M | NM | NM | 4 | 0 | 4 | 1 | 6.14 | oui | 5.57 | 1 |  |  | NM |  |  |  |
| 42 | 60 | F | M | NM | NM | 2 | 1 | 4 | 1 | 12 | oui | 5.1 | 1 |  | negatif | NM | 150 | 0 | 200 |
| 43 | 60 | F | M | NM | NM | 2 | 2 | 3 | 0 | 24 | oui | 16.77 | 0 |  | negatif | NM | 40 | 0 | 160 |
| 44 | 59 | F | NM | NM | NM | 2 | 2 | 3 | 1 | 34 | oui | 7.93 | 1 | 2.4 | negatif | M | 90 | 0 | 120 |



| Filename | Category Code | Column Number | Data Label | Data Label Sourc | Controlled Vocab Cd |
|---|---|---|---|---|---|
| EGP0001_data.csv | | 1 | SUBJ_ID | | |
| EGP0001_data.csv | Clinical_Data+Demographics | 2 | AGE | | 424144002 |
| EGP0001_data.csv | Clinical_Data+Demographics | 3 | SEX | | 263495000 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Mutation_Detection | 4 | KRAS Mutation | | 190070 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Mutation_Detection | 5 | BRAF Mutation | | 164757 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Mutation_Detection | 6 | NRAS Mutation | | 164790 |
| EGP0001_data.csv | Clinical_Data+Treatment+Chemotherapy | 7 | Number of lines | | 399042005 |
| EGP0001_data.csv | | 8 | OMIT | | |
| EGP0001_data.csv | | 9 | OMIT | | |
| EGP0001_data.csv | Clinical_Data+Outcome | 10 | Progression | | 419835002 |
| EGP0001_data.csv | Clinical_Data+Outcome | 11 | Duration of Response | | 445397003 |
| EGP0001_data.csv | Clinical_Data+Outcome | 12 | Death | | 419620001 |
| EGP0001_data.csv | Clinical_Data+Outcome | 13 | Overall Survival | | 445320007 |
| EGP0001_data.csv | Clinical_Data+Outcome | 14 | OMS Score | | 373802001 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 15 | EGFR Copy Number | | 5006 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 16 | HIRSCH Score | | |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 17 | PIK3CA Mutation | | 171834 |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 18 | | | |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 19 | | | |
| EGP0001_data.csv | Biomarker_Data+Non_Omics+Immunological | 20 | | | |


tranSMART
Search
Dataset Explorer
Sample Explorer
Gene Signature/Lists
Search by Subject
Navigate Terms
Across Trials
EGP0001 (173)
Biomarker Data (173)
Non Omics (173)
Immunological (172)
Mutation Detection (173)
BRAF Mutation (173)
abc M (5)
abc NA (2)
abc NM (166)
KRAS Mutation (173)
abc M (68)
abc NA (8)
abc NM (97)
NRAS Mutation (173)
Clinical Data (173)
Demographics (173)
SEX (173)
abc F (78)
abc M (95)
123 AGE (172)
Outcome (173)
Deces (173)
123 Delai survie globale (173)
123 Duree reponse (172)
123 OMS Score (159)
123 Progression (172)
Treatment (173)
Generate Summary Statistics
Summary
Clear
Save
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
Analysis
X
X
High Dimensional Data

**Cohorts**

Subset 1: (\Public Studies\EGP0001\Biomarker Data\Non Omics\Mutation Detection\KRAS Mutation\NA\ )

**Analysis:** Survival Analysis

**Variable Selection**

**Time**

Select time variable from the Data Set Explorer Tree and drag it into the box. For example, "Survival Time". This variable is required.

...\Delai survie globale\

**Category**

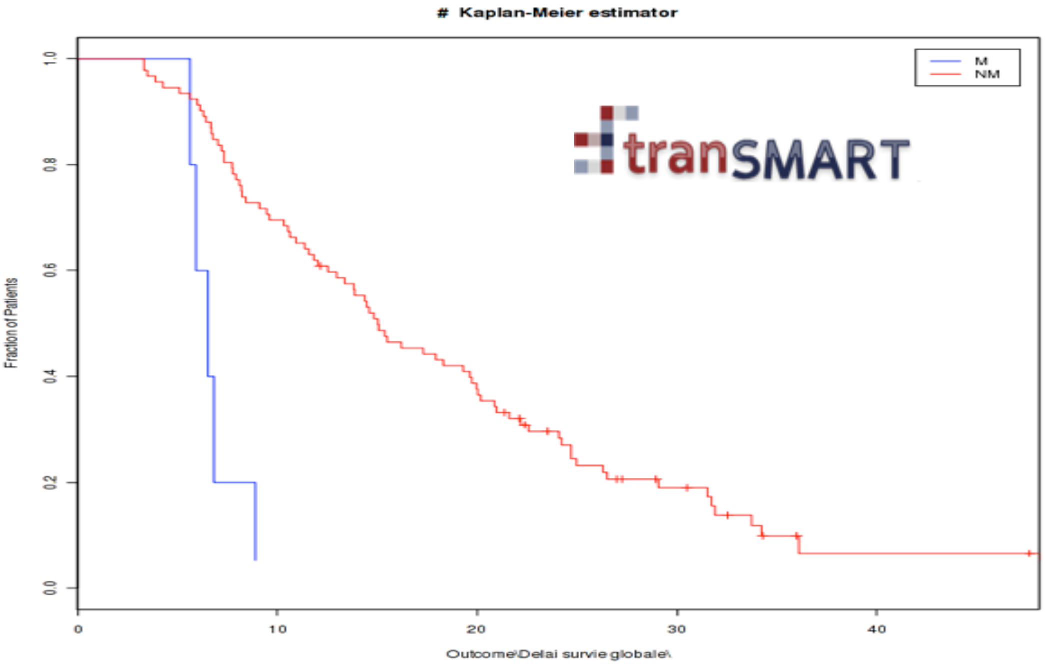
Select a variable on which you would like to sort the cohort and drag it into the box. For example, "Cancer Stage". If this variable is continuous (ex. Age), then it should be "binned" using the option below. This variable is not required.

...\M\
...\NM\

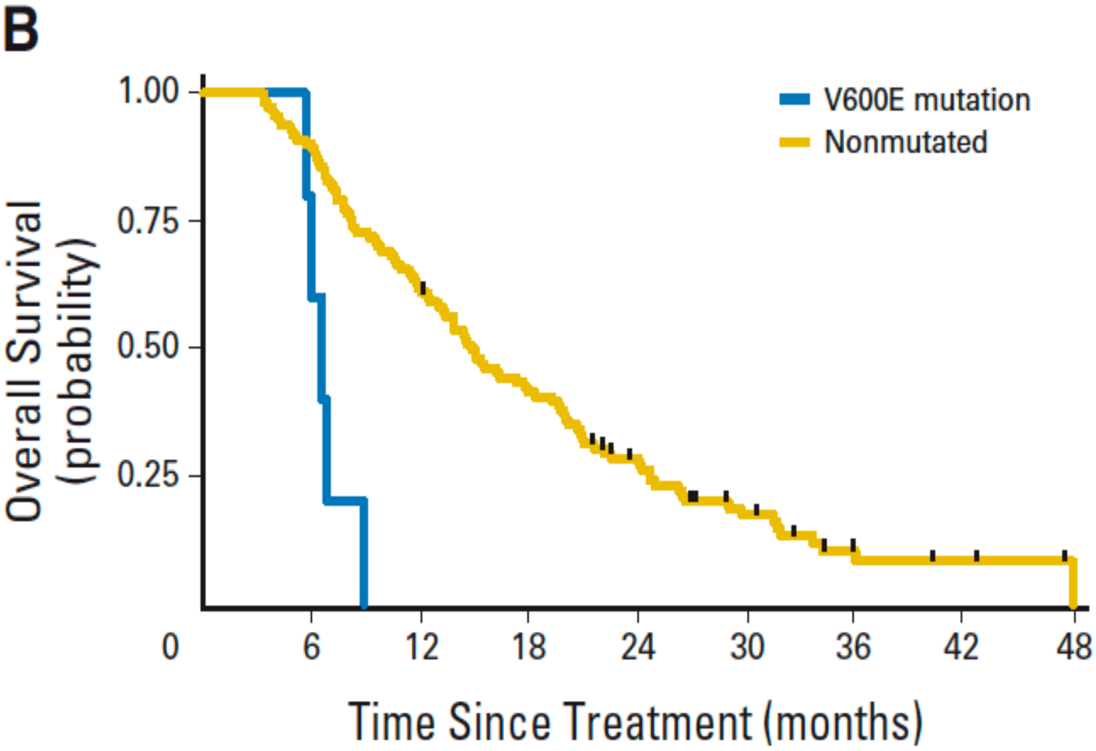
Se
a
"Su

HEGP: Canuel V, Avillach P

- R module in tranSMART



- Published figure in JCO



HEGP: Canuel V, Avillach P