Translational Science and Real World Evidence

Paul Avillach, MD, PhD

Assistant Professor Paris Descartes University

Postdoc at the Center for Biomedical Informatics – Harvard Medical School

Oct, 16th 2013

HEGP Background

Location

• Paris 15th district

Within AP-HP

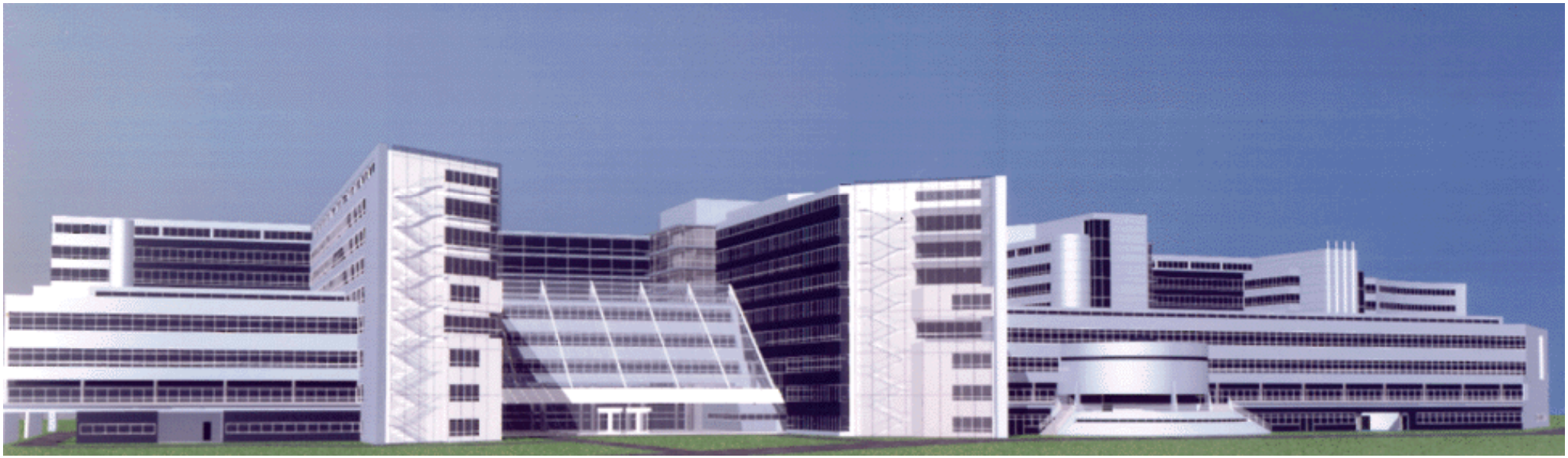
• HEGP is the most recent acute care hospital within the 38 AP-HP hospitals

• 900 beds

• HEGP meets the needs of the 600, 000 inhabitants of the Paris south-west

HEGP background

Opening : July 2000-



Hôpital Laennec (1634)

Hôpital Boucicaut

Hôpital Broussais

Allergie
○ Oui ○ Non

Asthme
○ Oui ○ Non

Allergie à l'iode
○ Oui ○ Non

Données cliniques

Diabète insulino-dépendant
○ Oui ○ Non

Immuno-déprimé
○ Oui ○ Non

Diabète non insulino-dépendant
○ Oui ○ Non

Insuffisance cardiaque
○ Oui ○ Non

Insuffisa. rénale creat>120µm/l
○ Oui ○ Non

Insuffisance hépatique
○ 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

Transplantation
○ Oui ○ Non

HEGP BDW

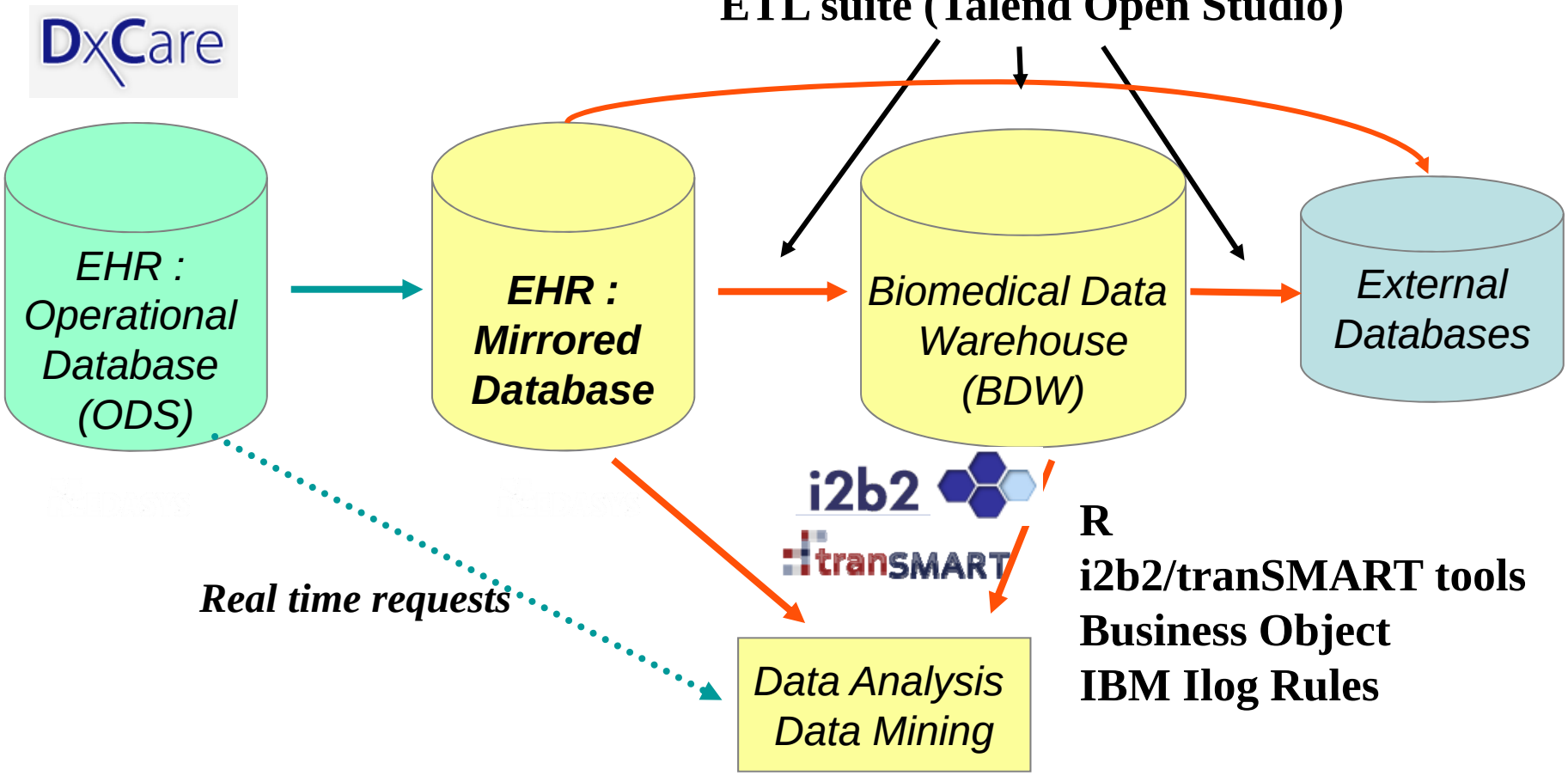
EHR/BDW integration

Pr Patrice Degoulet CIO

Evaluation/Research environment

Production environment

ETL suite (Talend Open Studio)



R
i2b2/tranSMART tools
Business Object
IBM Ilog Rules

PI: Isaac Kohane – Harvard - CBMI

HEGP BDW

Why : Objectives

1. **EHR driven research (EHR data reuse)**
   - Patient selection for CR studies (e.g., EU FP7 EHR4CR)
   - In-silico evaluation of clinical decision rules
   - Phenotypic augmentation (e.g., reuse of EHR data to feed a CR study)
2. **Genotype/Phenotype integration**
   - Biomarker research (e.g., PheWAS studies)
   - Personalized medicine

- Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods Inf Med*. 2009; 48(1): 38-44.
- Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet*. 2011; 12(6): 417-28.
- Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012; 13(6): 395-405.

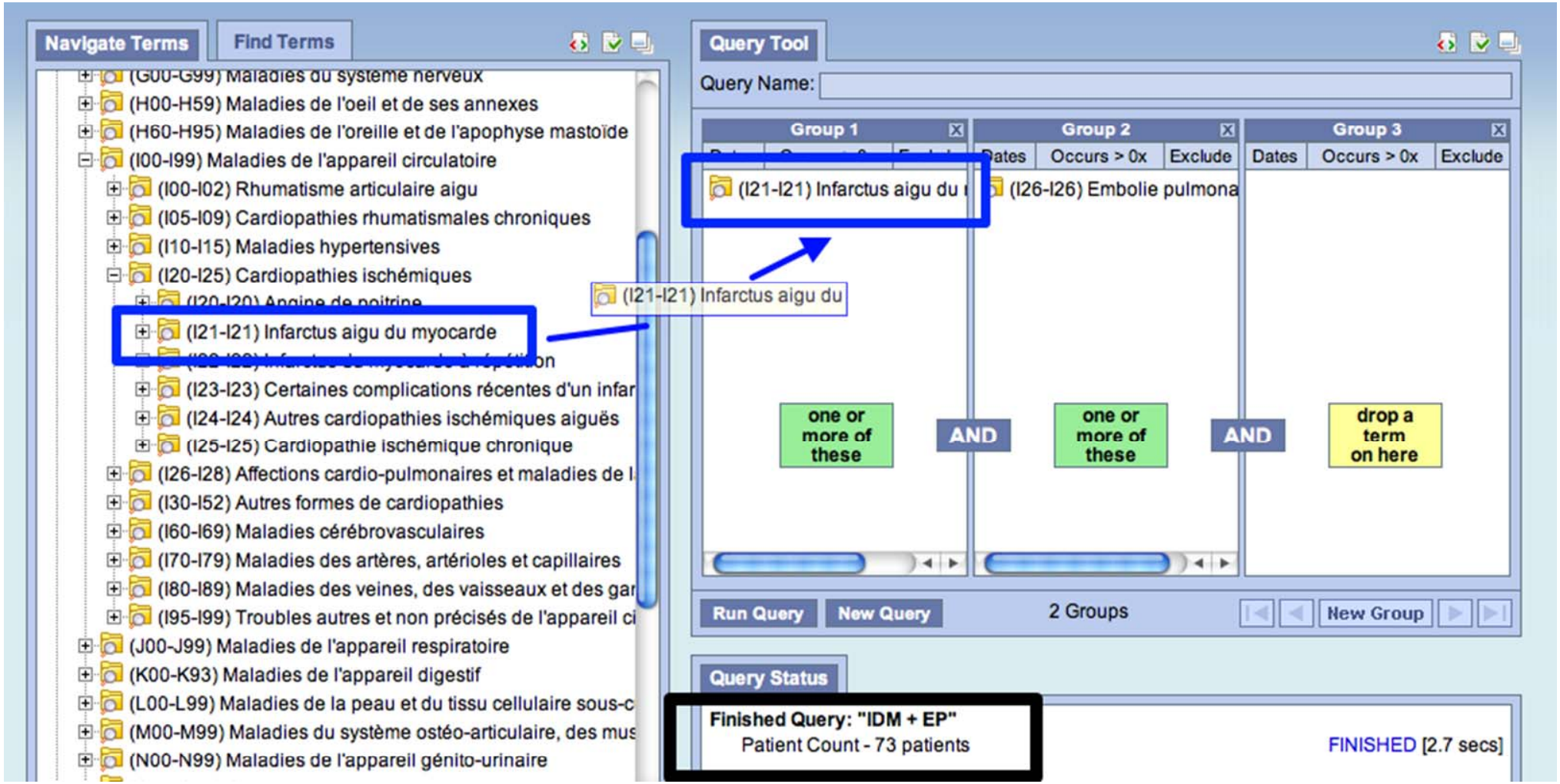
HEGP i2b2

Type of data

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

BDW use

i2b2 CDW queries



BDW use

i2b2 CDW queries (Jan. 2011-april 2013)

- **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**

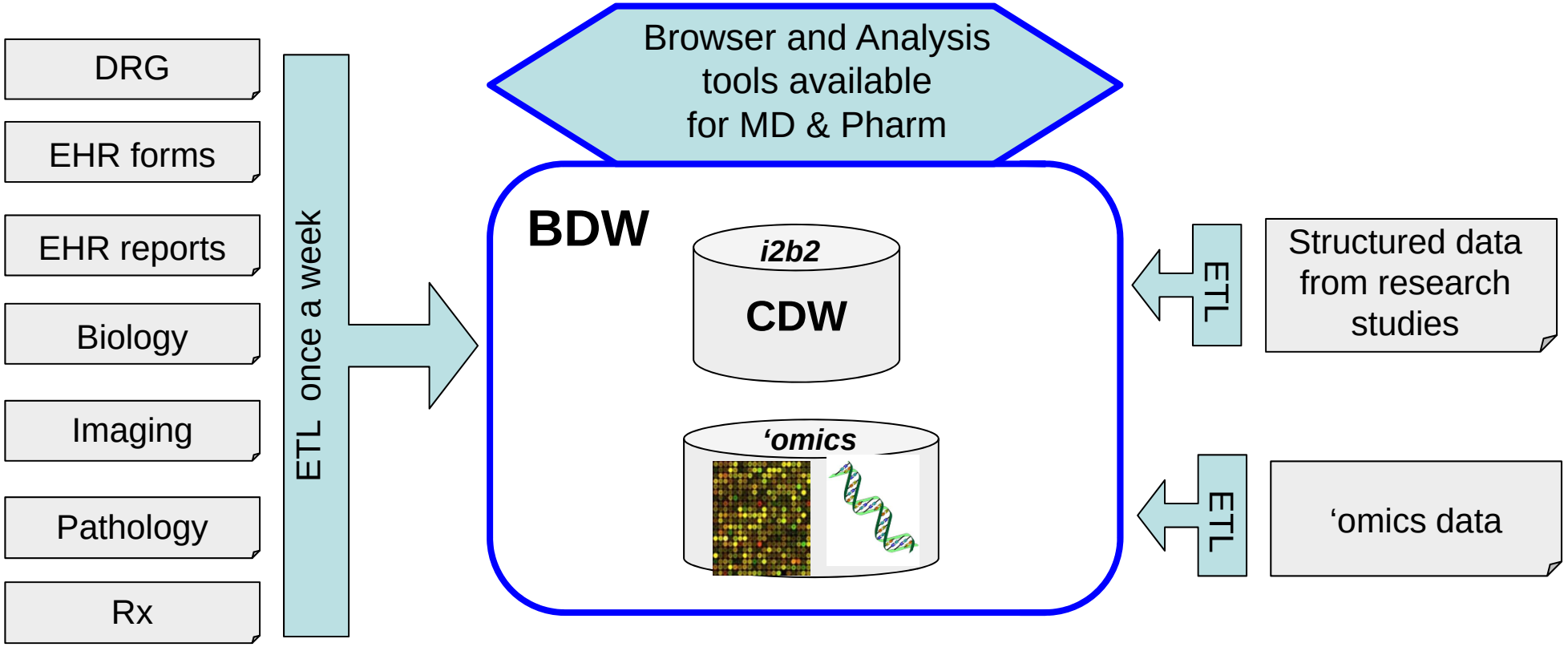
- **Level 2 and 3 studies : access to patient level data**
  - Structured written project
  - Validation by the HEGP ethical/research committee
  - Transmission to the regional IRB committee
  - Level 2 : **anonymized patient data**
  - Level 3 : **de-anonymized patient data**

  - **IRB approval for 32 projects**

HEGP

Health care
Health Information System

Clinical Research



HEGP BDW

tranSMART

- Integrated platform to support translational research
- Initiated by Johnson & Johnson et Recombinant 5 years ago
- Open-source since January, 2012
- Installed at HEGP since June, 2012

- tranSMART [Internet]. [cited 2013 Jan 16]. Available from: http://www.transmartproject.org/
- Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med. 2010; 8:68.

HEGP BDW

tranSMART

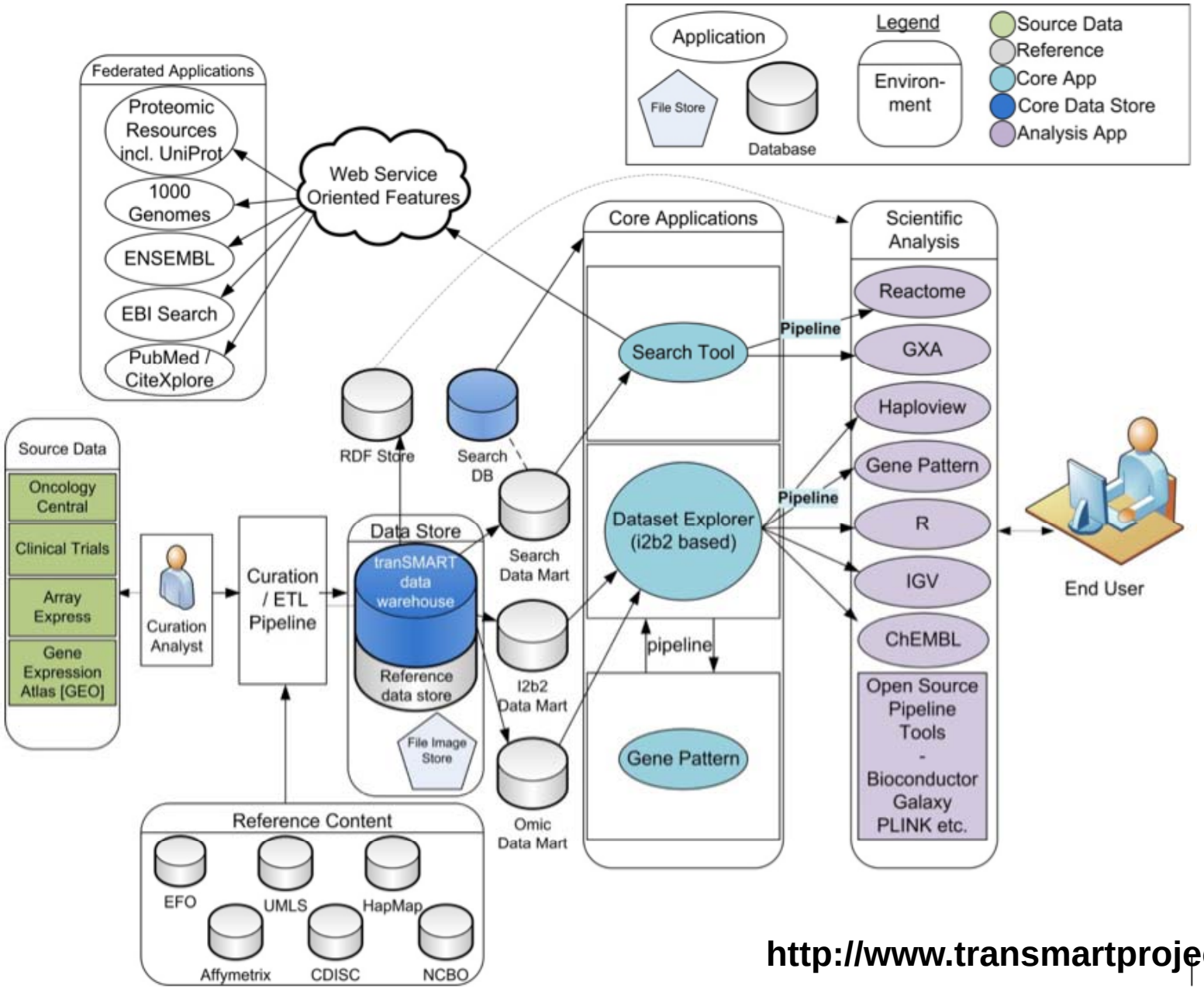
- Objectives :
  1. Integration of clinical, biological and ‘omics data in one place – hypothesis free –
  2. Generation of hypothesis by Clinicians / Researchers

- tranSMART [Internet]. [cited 2013 Jan 16]. Available from: http://www.transmartproject.org/
- Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med. 2010; 8:68.

BDW content

“Omics” data integration

tranSMART



http://www.transmartproject.org

BDW content

Pilot Study

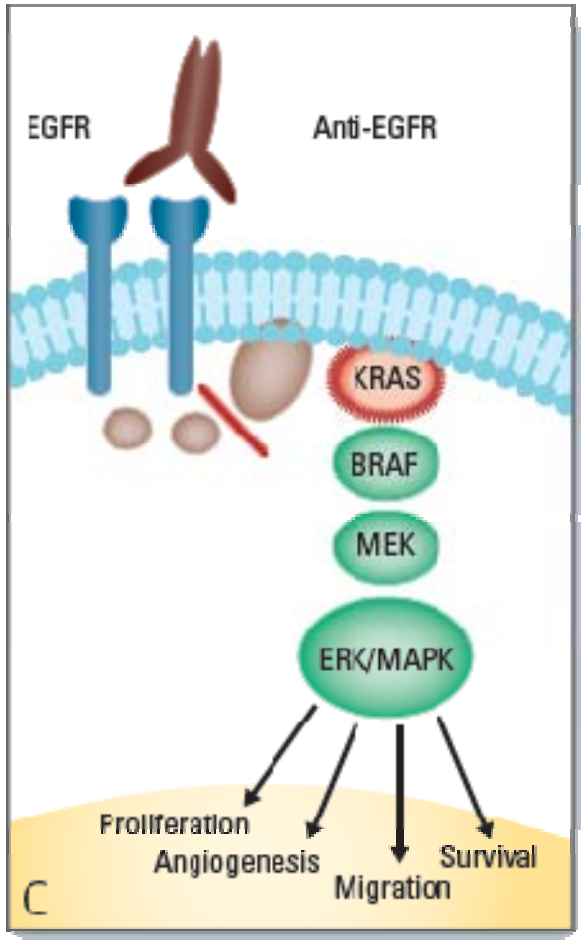
VOLUME 27 · NUMBER 35 · DECEMBER 10 2009

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

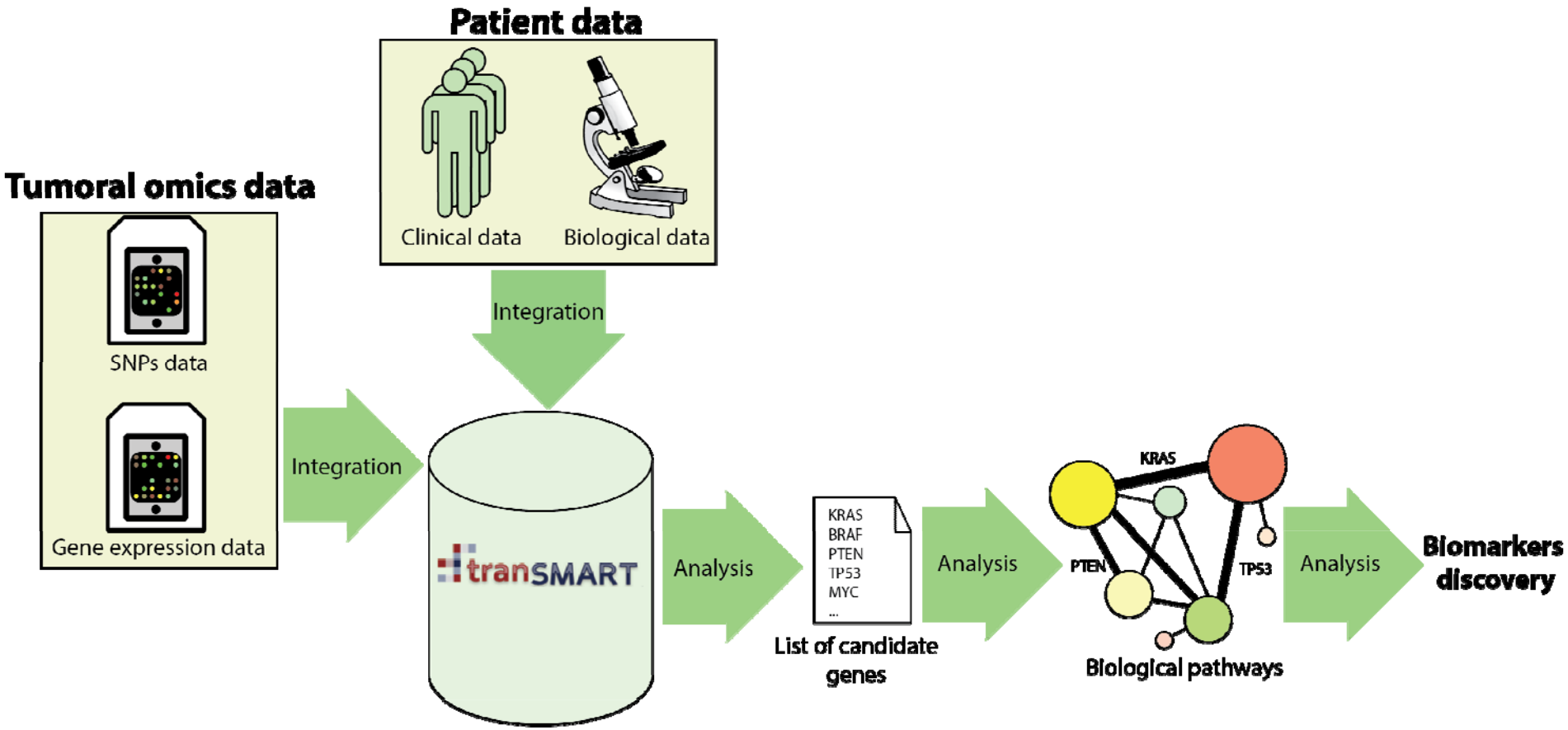
Analysis of *PTEN*, *BRAF*, and *EGFR* Status in Determining Benefit From Cetuximab Therapy in Wild-Type *KRAS* Metastatic Colon Cancer

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

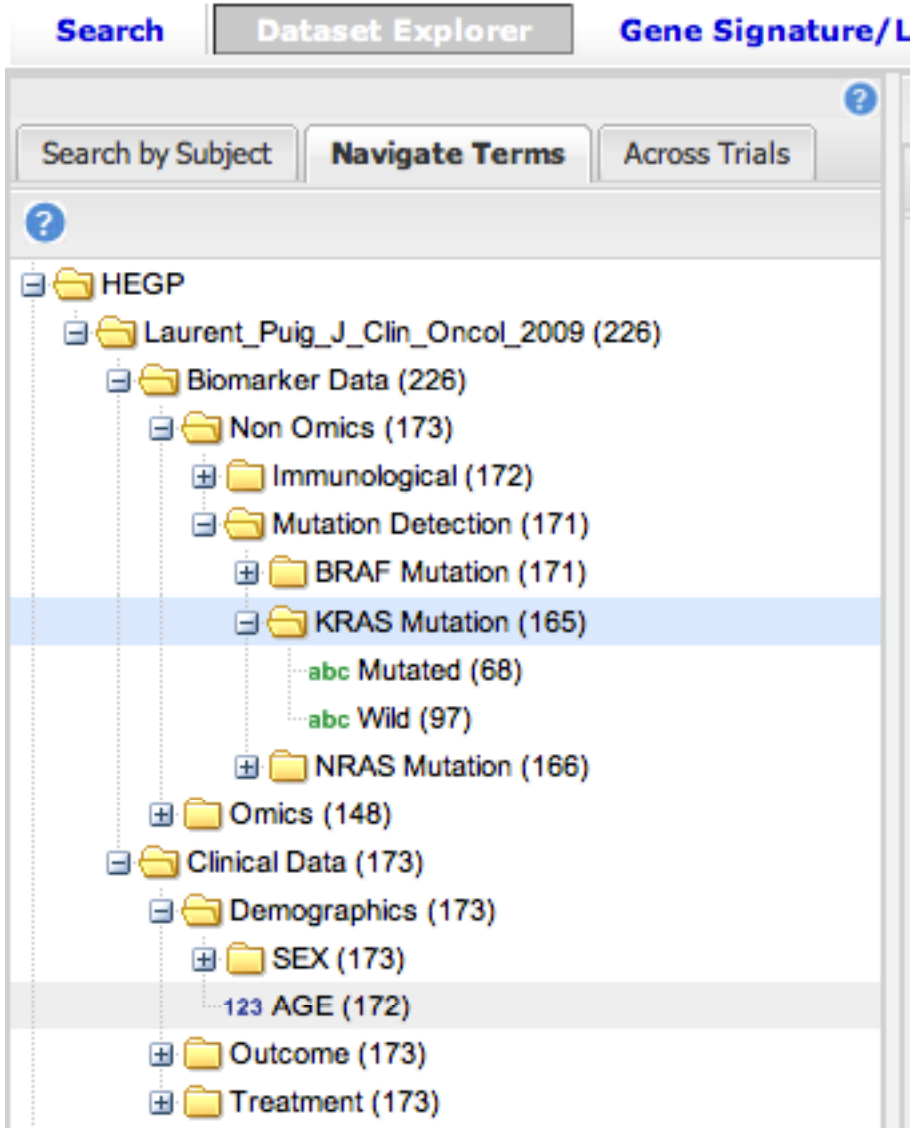


- KRAS, BRAF testing in Colorectal Cancer - Clarient [Internet]. [cited 2013 Jan 29]. Available from: http://www.clarientinc.com/clarient-home/kras-and-braf.aspx

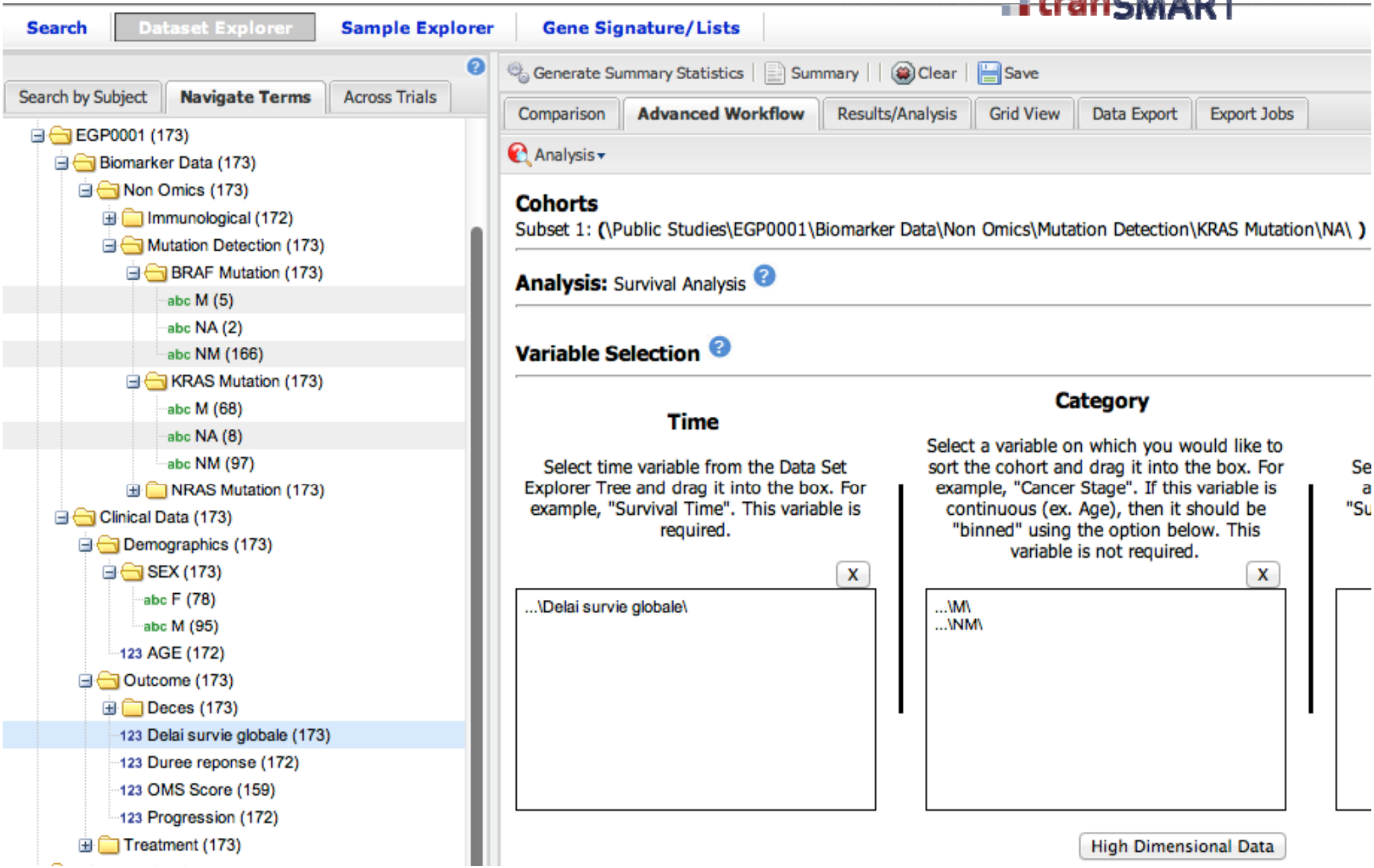
HEGP tranSMART



HEGP tranSMART

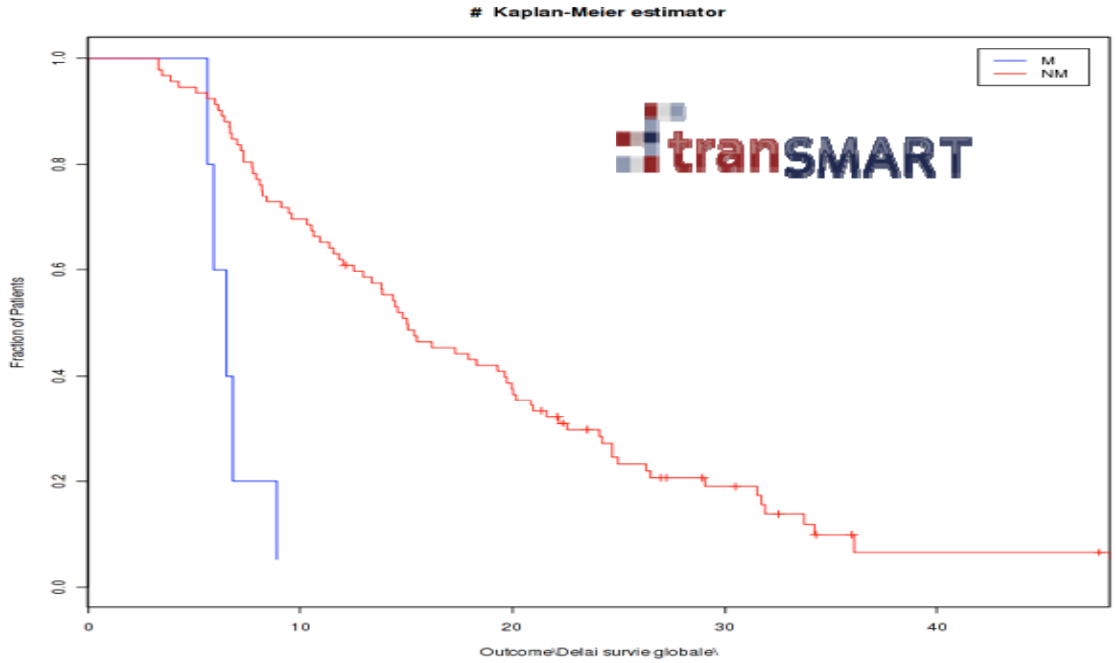


HEGP tranSMART

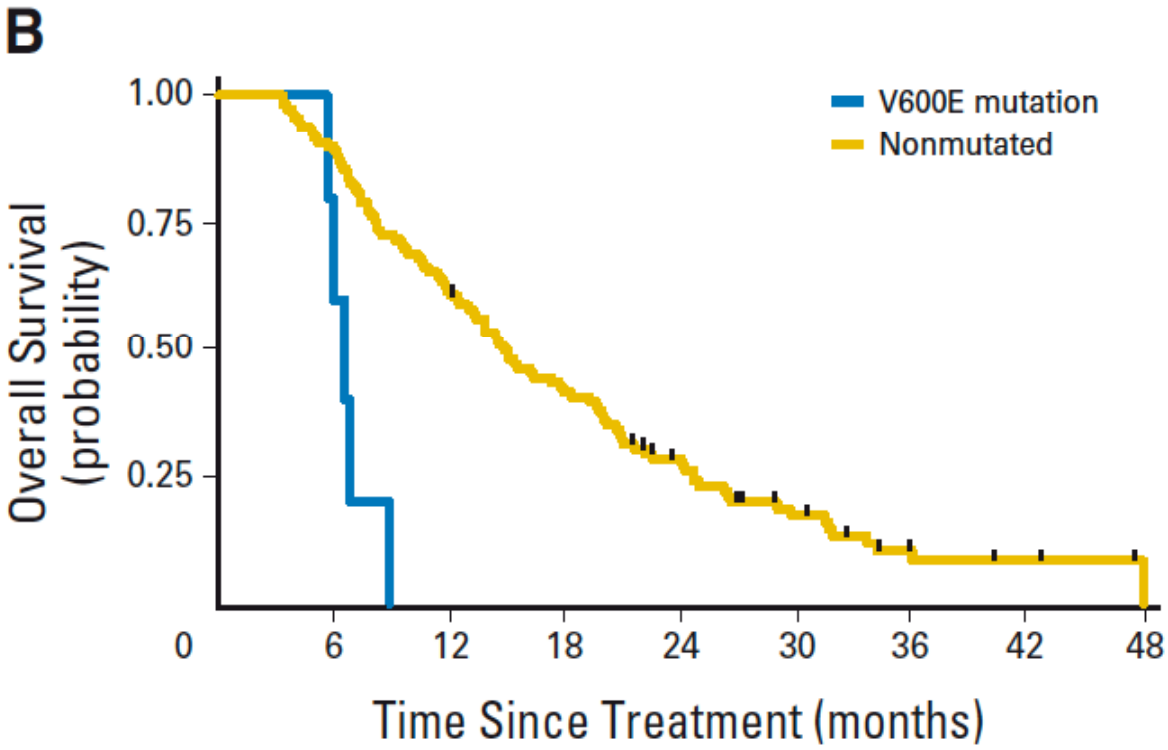


HEGP tranSMART

- R module in tranSMART

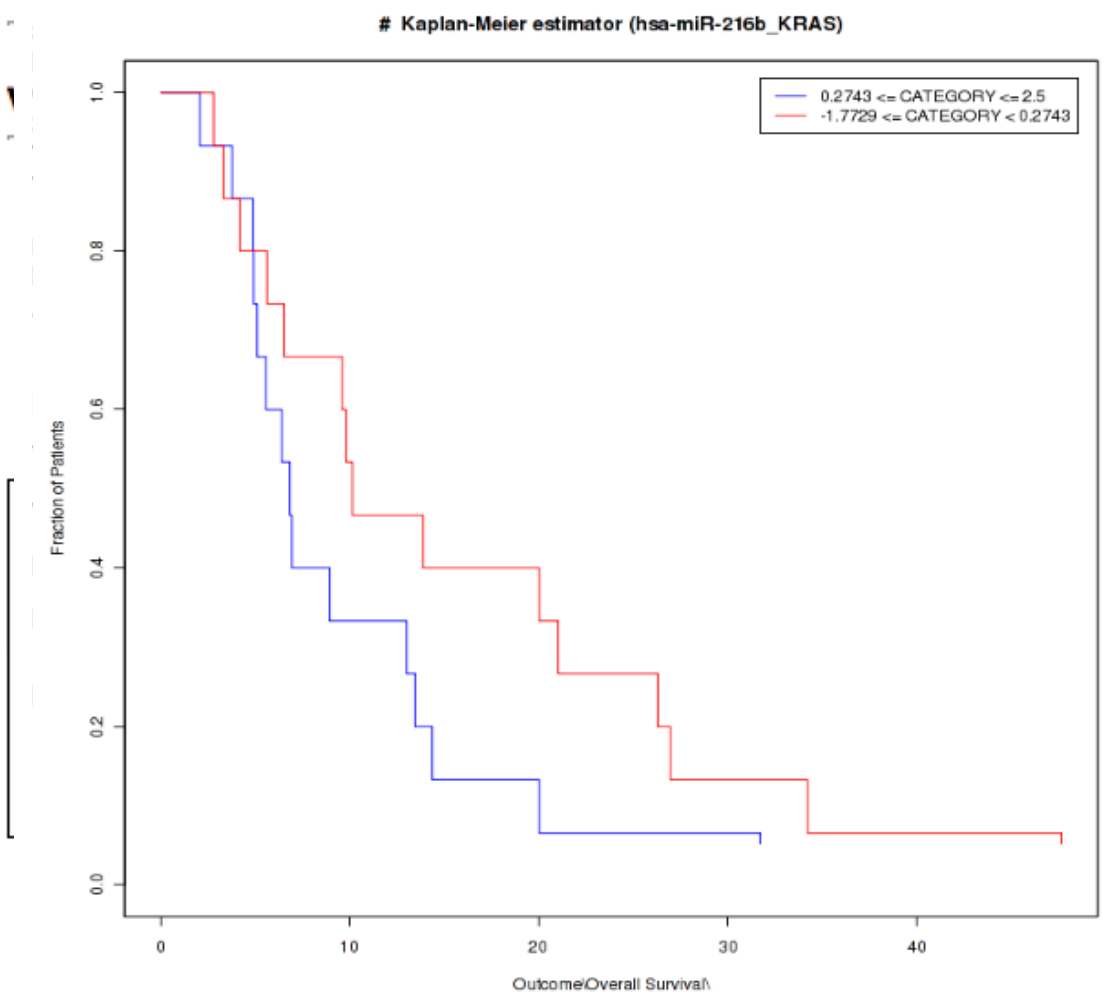


- Published figure in JCO



- Laurent_Puig_J_Clin_Oncol_2009 (226)
  - Biomarker Data (226)
    - Non Omics (173)
    - Omics (148)
      - Gene Expression (85)
        - HG-U133 Plus 2 (85)
          - Tumoral Tissue (85)
      - miRNA Expression (78)
        - Illumina Human V2 (78)
          - Tumoral Tissue (78)
  - Clinical Data (173)
    - Demographics (173)
      - SEX (173)
      - 123 AGE (172)
    - Outcome (173)
      - Death (173)
      - 123 Duration of Response (172)
      - 123 OMS Score (159)
      - 123 Overall Survival (173)
      - 123 Progression (172)
    - Treatment (173)

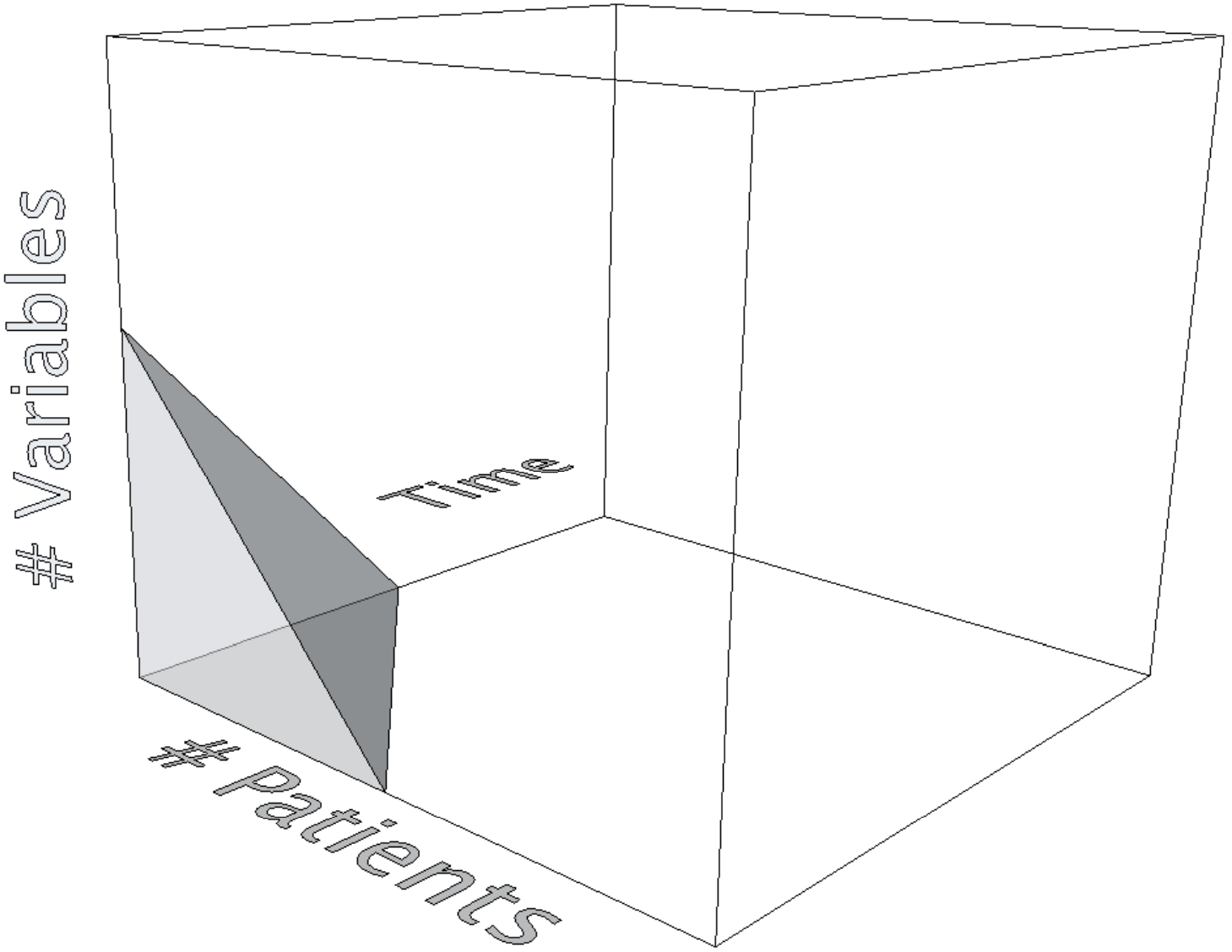
Survival Curve



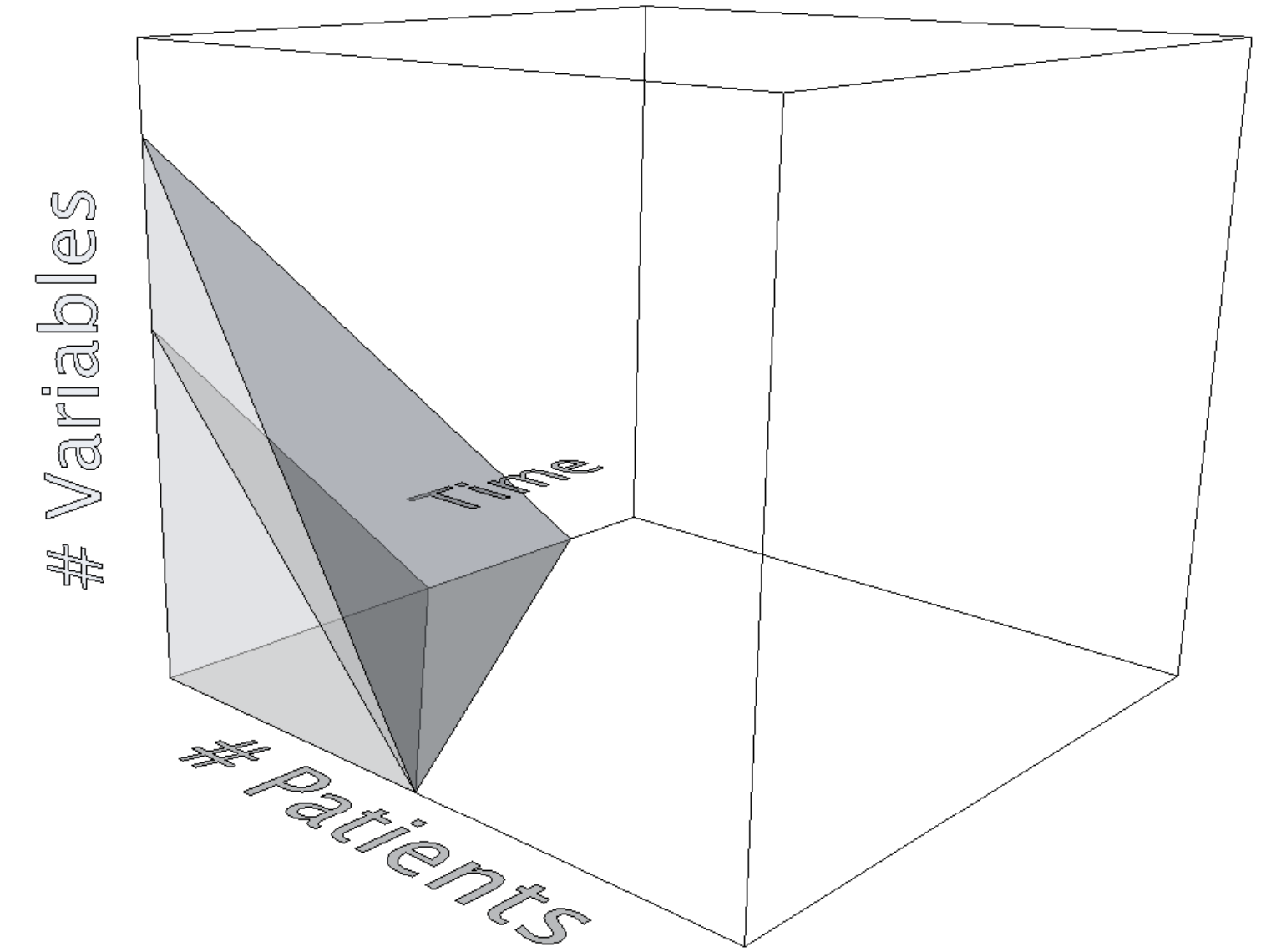
Run

nal Data

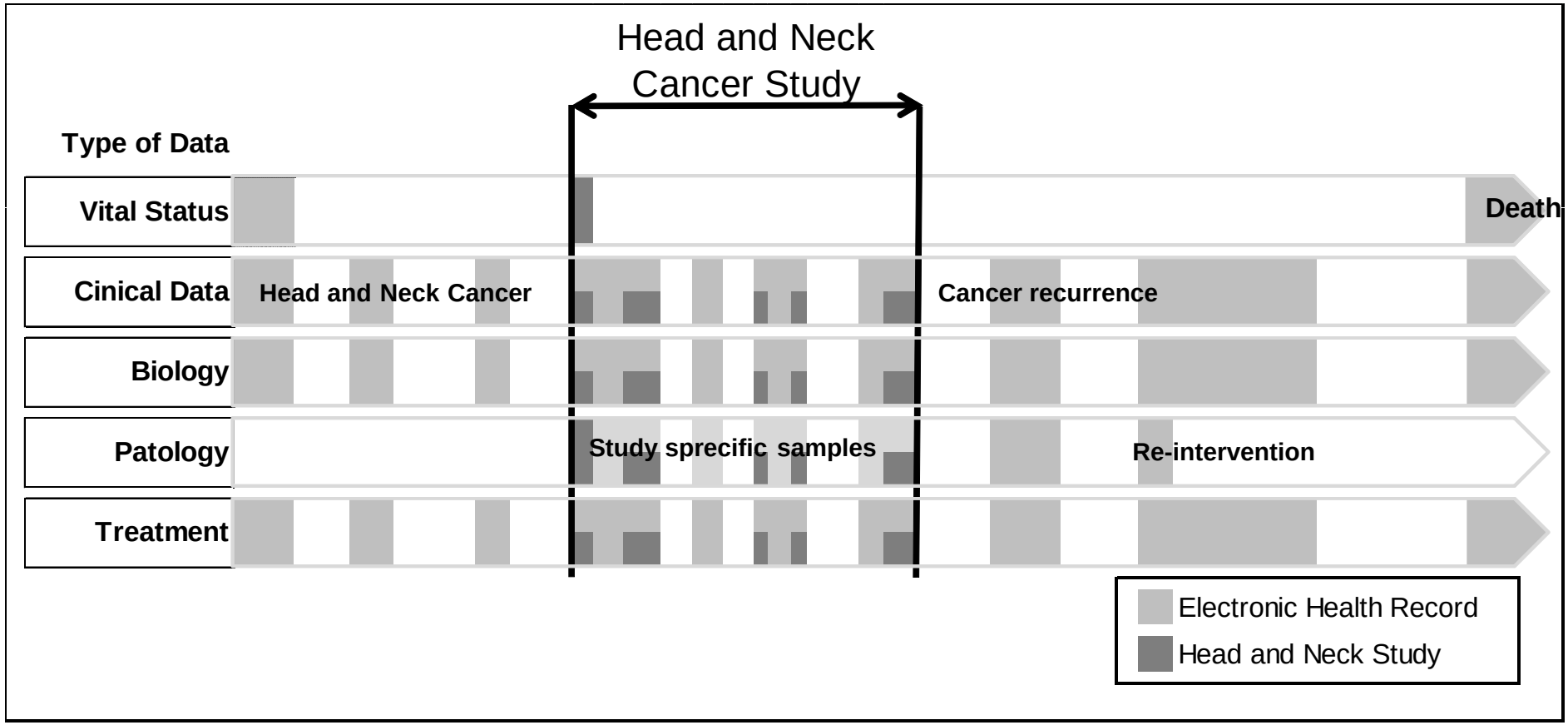
Phenotypic augmentation

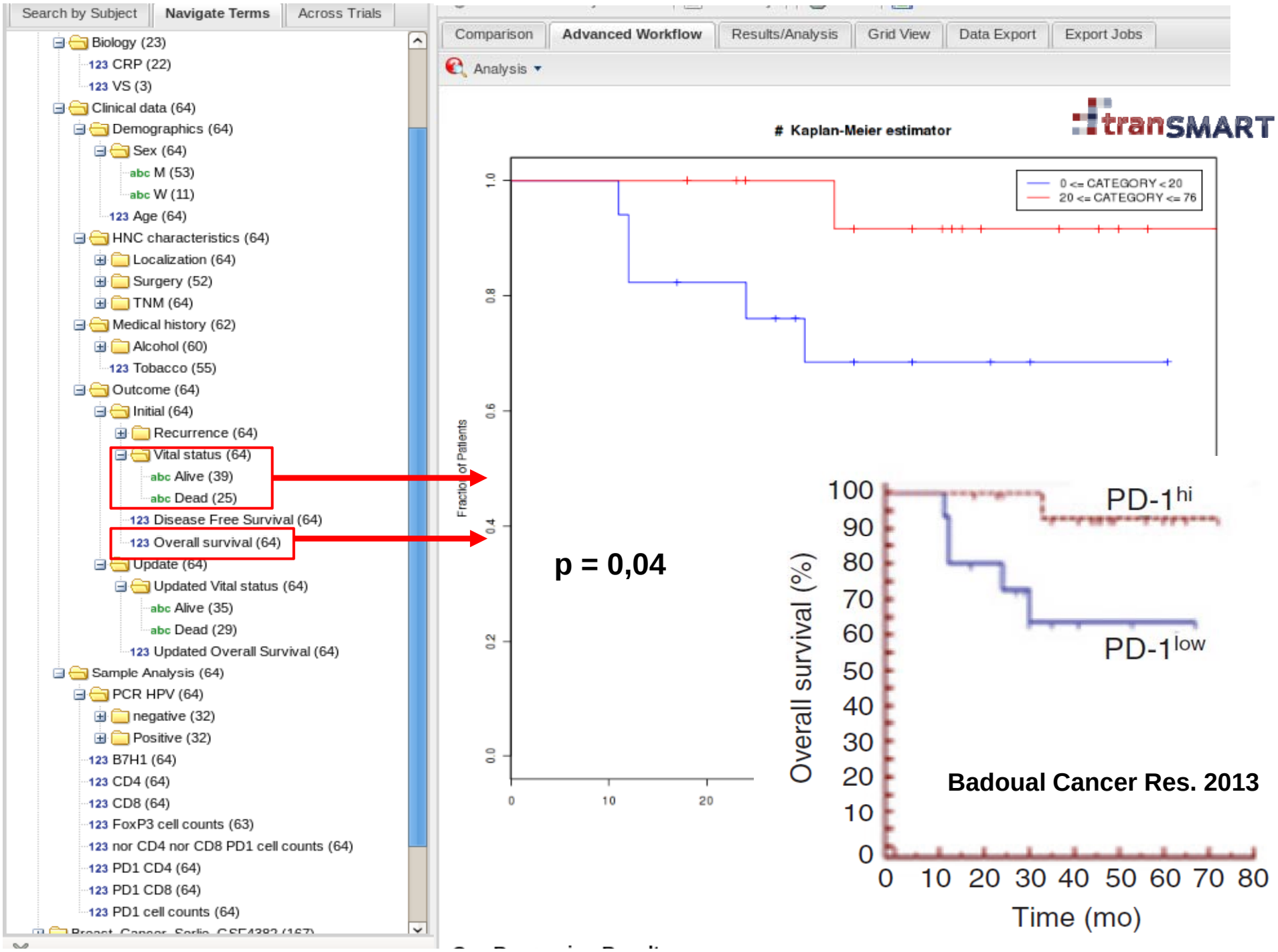


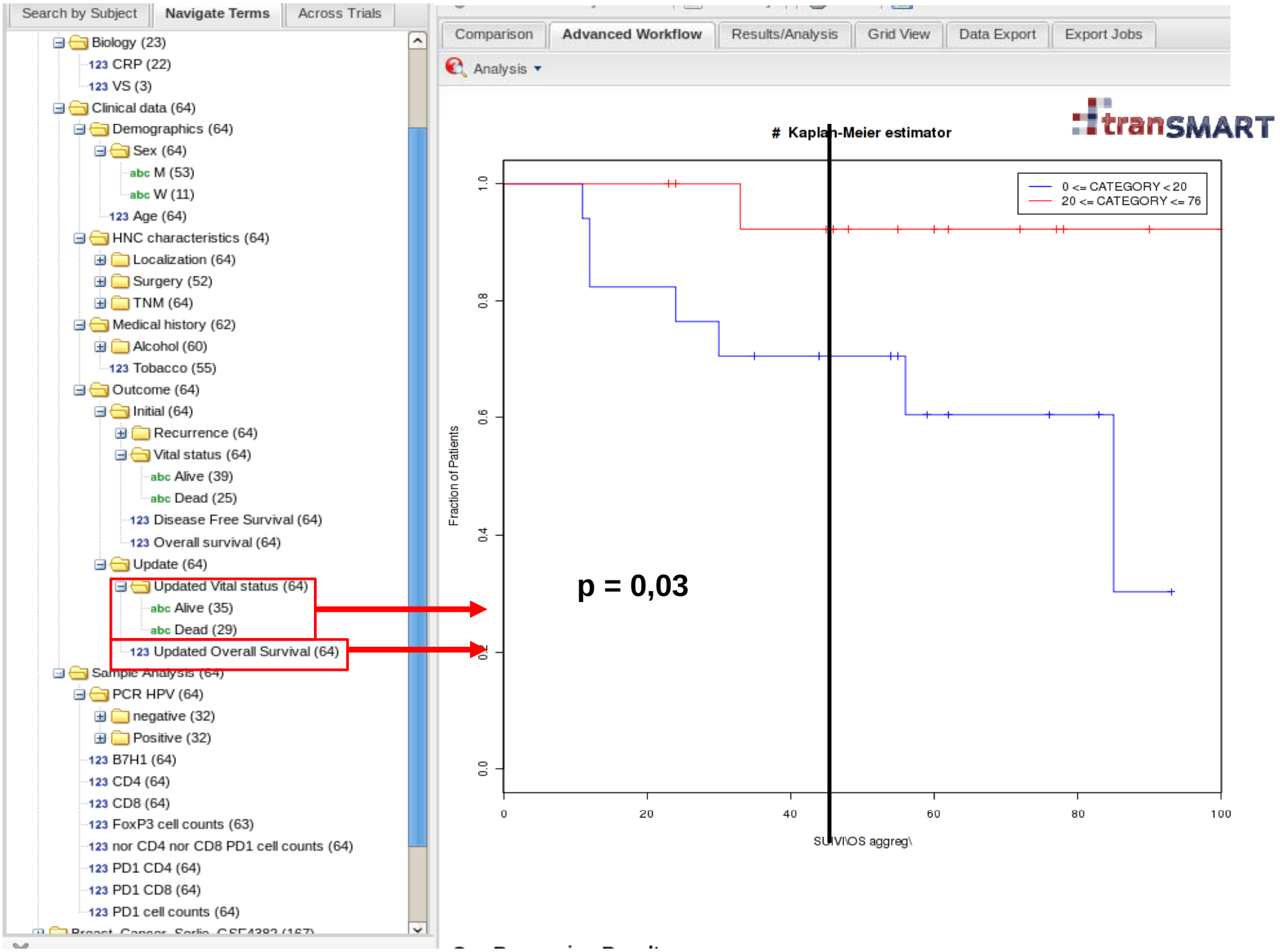
Phenotypic augmentation

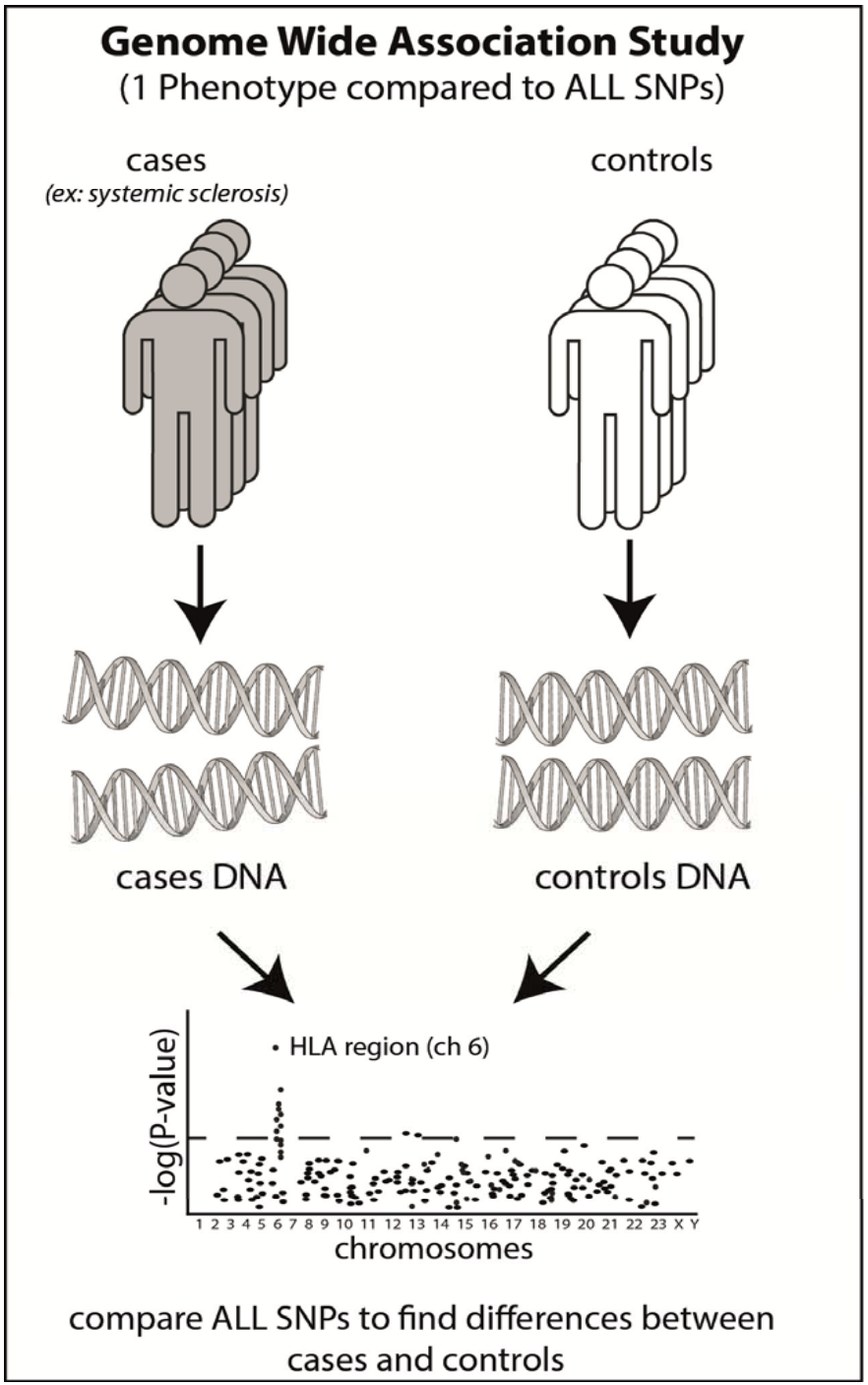


tranSMART + i2b2 = Phenotypic augmentation




Search by Subject
Navigate Terms
Across Trials
Biology (23)
CRP (22)
VS (3)
Clinical data (64)
Demographics (64)
Sex (64)
M (53)
W (11)
Age (64)
HNC characteristics (64)
Localization (64)
Surgery (52)
TNM (64)
Medical history (62)
Alcohol (60)
Tobacco (55)
Outcome (64)
Initial (64)
Recurrence (64)
Vital status (64)
Alive (39)
Dead (25)
Disease Free Survival (64)
Overall survival (64)
Update (64)
Updated Vital status (64)
Alive (35)
Dead (29)
Updated Overall Survival (64)
Sample Analysis (64)
PCR HPV (64)
negative (32)
Positive (32)
B7H1 (64)
CD4 (64)
CD8 (64)
FoxP3 cell counts (63)
nor CD4 nor CD8 PD1 cell counts (64)
PD1 CD4 (64)
PD1 CD8 (64)
PD1 cell counts (64)
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
Analysis
Kaplan-Meier estimator
tranSMART
0 <= CATEGORY < 20
20 <= CATEGORY <= 76
Fraction of Patients
p = 0,04
Overall survival (%)
PD-1hi
PD-1low
Badoual Cancer Res. 2013
Time (mo)


Search by Subject
Navigate Terms
Across Trials
Biology (23)
123 CRP (22)
123 VS (3)
Clinical data (64)
Demographics (64)
Sex (64)
abc M (53)
abc W (11)
123 Age (64)
HNC characteristics (64)
Localization (64)
Surgery (52)
TNM (64)
Medical history (62)
Alcohol (60)
123 Tobacco (55)
Outcome (64)
Initial (64)
Recurrence (64)
Vital status (64)
abc Alive (39)
abc Dead (25)
123 Disease Free Survival (64)
123 Overall survival (64)
Update (64)
Updated Vital status (64)
abc Alive (35)
abc Dead (29)
123 Updated Overall Survival (64)
Sample Analysis (64)
PCR HPV (64)
negative (32)
Positive (32)
123 B7H1 (64)
123 CD4 (64)
123 CD8 (64)
123 FoxP3 cell counts (63)
123 nor CD4 nor CD8 PD1 cell counts (64)
123 PD1 CD4 (64)
123 PD1 CD8 (64)
123 PD1 cell counts (64)
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
Analysis
tranSMART
Kaplan-Meier estimator
0 <= CATEGORY < 20
20 <= CATEGORY <= 76
Fraction of Patients
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
SURV\OS aggreg\
p = 0,03


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

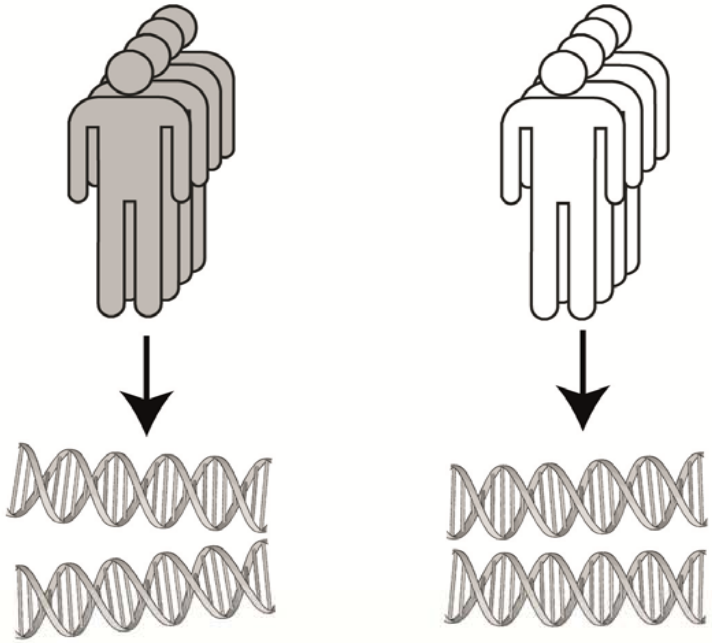
Genome Wide Association Study

(1 Phenotype compared to ALL SNPs)

cases

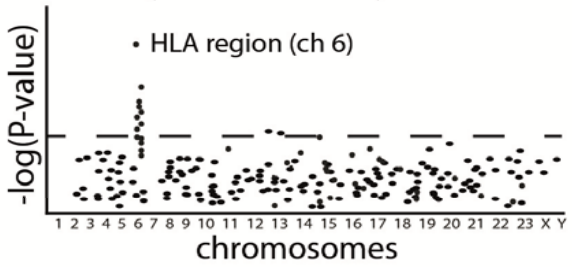
(ex: systemic sclerosis)

controls



cases DNA

controls DNA



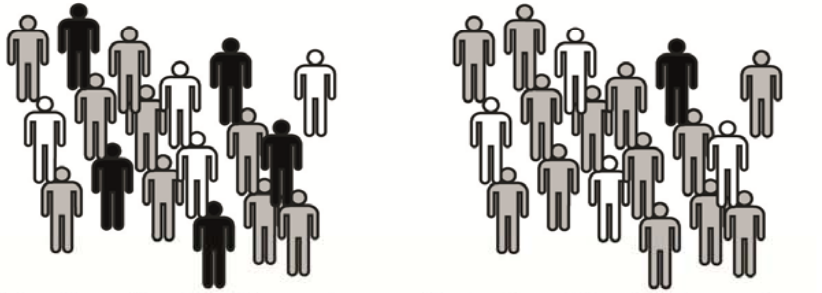
compare ALL SNPs to find differences between cases and controls

Phenome Wide Association Study

(1 SNP compared to ALL Phenotypes)

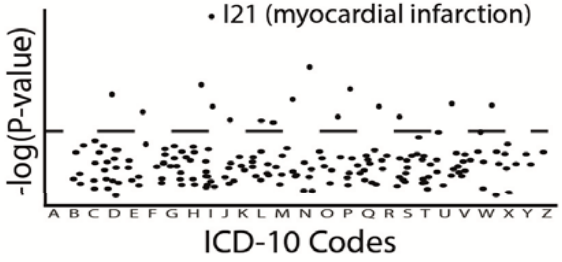
allele G patients group

allele A patients group



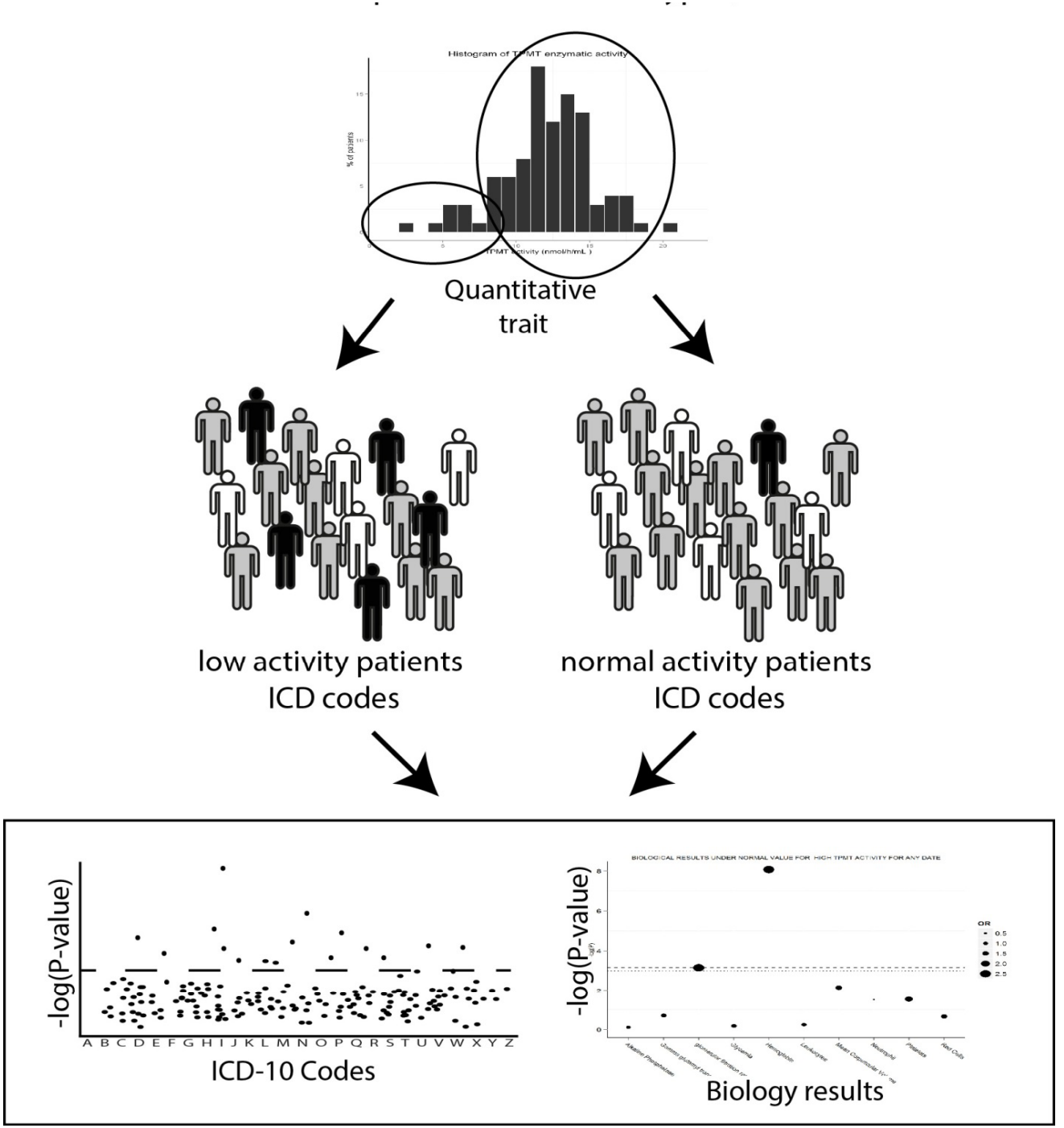
allele G patients phenotype

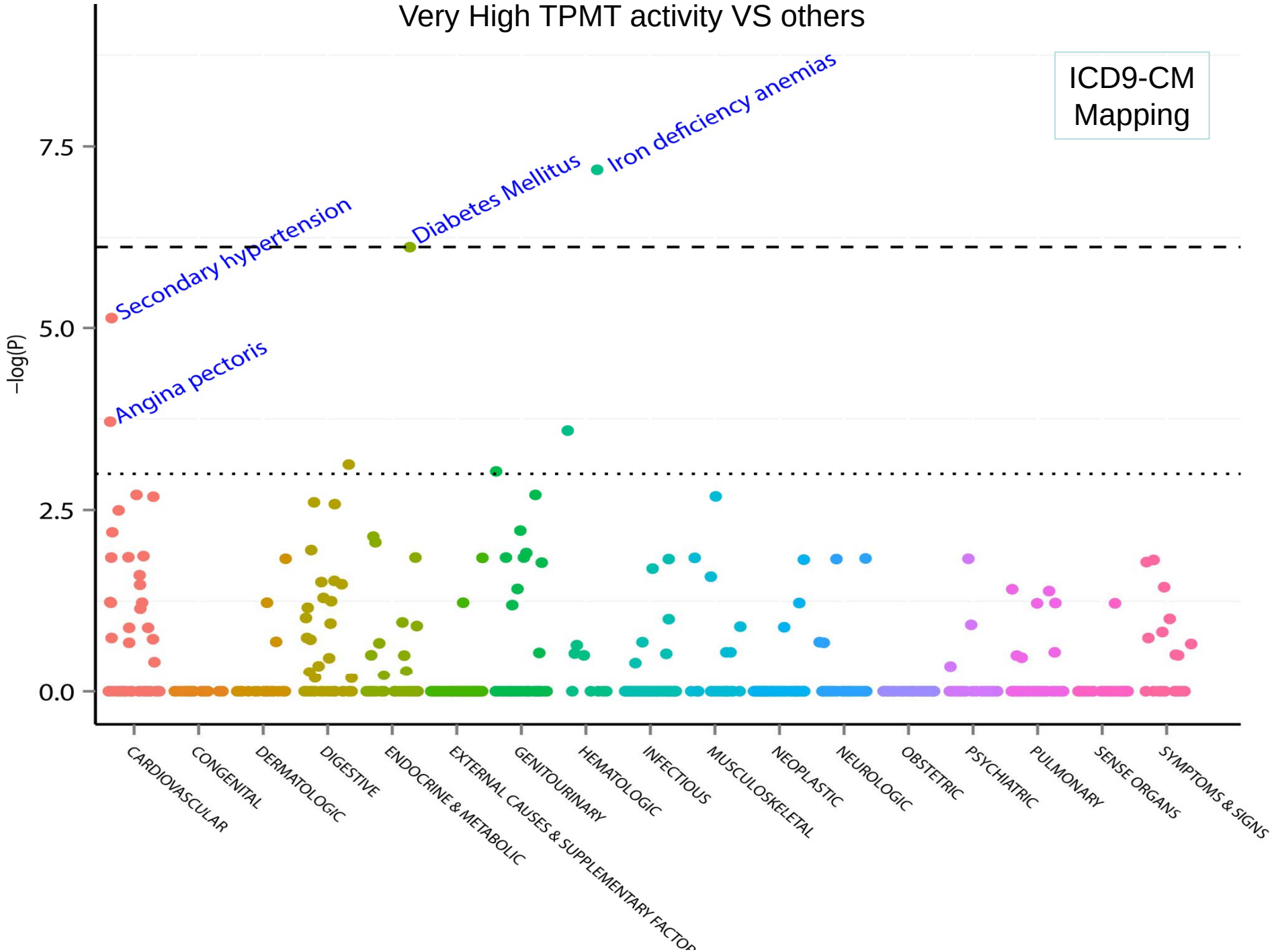
allele A patients phenotype

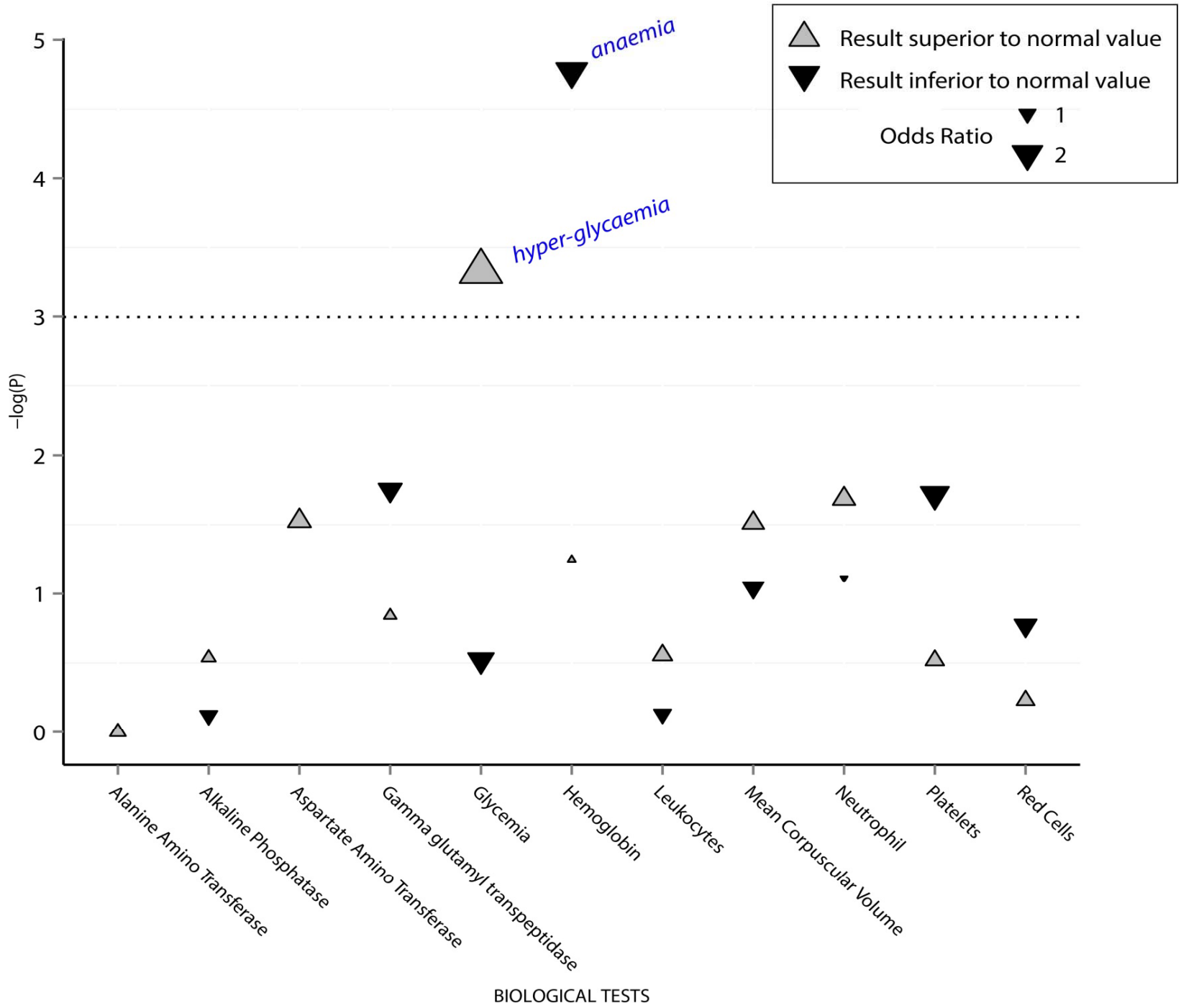


compare ALL DIAGNOSIS to find differences between cases and controls

TPMT


Histogram of TPMT enzymatic activity
% of patients
TPMT activity (nmol/h/mL)
Quantitative trait
low activity patients ICD codes
normal activity patients ICD codes
-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
BIOLOGICAL RESULTS UNDER NORMAL VALUE FOR HIGH TPMT ACTIVITY FOR ANY DATE
-log(P-value)
OR
0.5
1.0
1.5
2.0
2.5
Biology results


Very High TPMT activity VS others
ICD9-CM Mapping
−log(P)
7.5
5.0
2.5
0.0
Secondary hypertension
Diabetes Mellitus
Iron deficiency anemias
Angina pectoris
CARDIOVASCULAR
CONGENITAL
DERMATOLOGIC
DIGESTIVE
ENDOCRINE & METABOLIC
EXTERNAL CAUSES & SUPPLEMENTARY FACTOR
GENITOURINARY
HEMATOLOGIC
INFECTIOUS
MUSCULOSKELETAL
NEOPLASTIC
NEUROLOGIC
OBSTETRIC
PSYCHIATRIC
PULMONARY
SENSE ORGANS
SYMPTOMS & SIGNS


anaemia
hyper-glycaemia
Result superior to normal value
Result inferior to normal value
Odds Ratio
1
2
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

European Medical Information Framework

EMIF project: European Medical Information Framework

- 58 partners
- 58 Million euros
- Started 1st jan 2013
- 5 year project
- PI: Prof. Simon Lovestone (KCL) & Bart Vannieuwenhuyse (Janssen)
- PI Platform: Prof. Johan van der Lei
- Detect new biomakers:
  - prediposition Alzheimer's disease
  - Metabolic complication diabetes

Innovative Medicines Initiative

EMIF project: European Medical Information Framework

3 projects, 3 datasources:

- EHR derived patient data: **48 millions patient records** : observational studies (**Jerboa tool)**
- Cohort clinical + ‘omics data from **: tranSMART option**
  - AD
  - Metabolic :

  100 000 patients

- **WP11** : Semantic harmonization
- Same concepts across all database : ETL process Jerboa & tranSMART

Acknowledgments HEGP

Informatics & Public Health Dept :

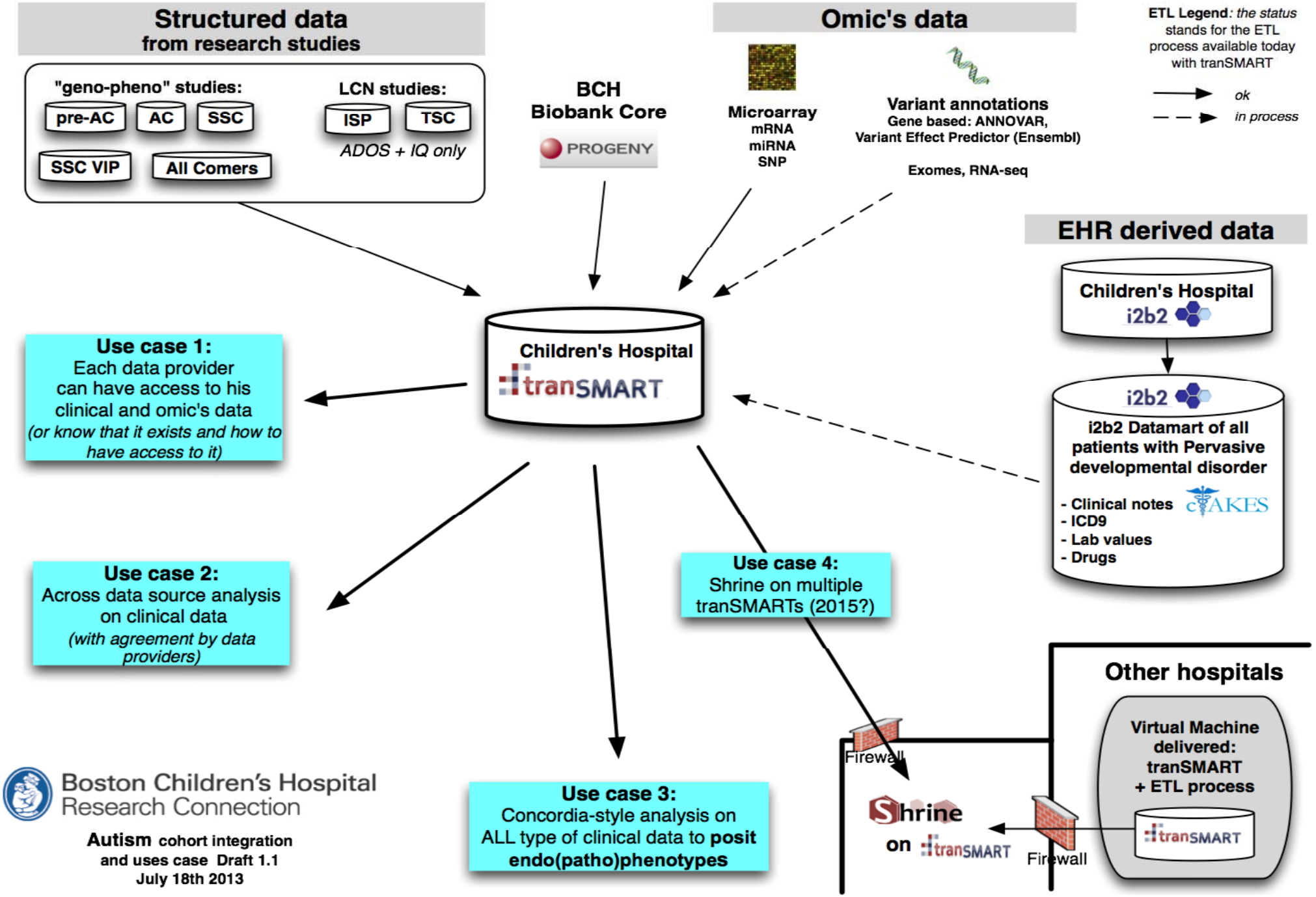
- **Eric Zapletal,**
- **Vincent Canuel,**
- **Antoine Neuraz,**
- **Fabien Joubert,**
- **Anita Burgun,**
- **Patrice Degoulet**

i2b2 www.i2b2.org

tranSMART www.transmartproject.org

Recombinant HEALTHCARE DATA WAREHOUSING www.recomdata.com

Contact :
paul.avillach@egp.aphp.fr


Structured data
from research studies
"geno-pheno" studies:
pre-AC
AC
SSC
SSC VIP
All Comers
LCN studies:
ISP
TSC
ADOS + IQ only
BCH
Biobank Core
PROGENY
Omic's data
Microarray
mRNA
miRNA
SNP
Variant annotations
Gene based: ANNOVAR,
Variant Effect Predictor (Ensembl)
Exomes, RNA-seq
ETL Legend: the status stands for the ETL process available today with tranSMART
ok
in process
EHR derived data
Children's Hospital
i2b2
Children's Hospital
tranSMART
i2b2
i2b2 Datamart of all patients with Pervasive developmental disorder
- Clinical notes
cTAKES
- ICD9
- Lab values
- Drugs
Use case 1:
Each data provider can have access to his clinical and omic's data
(or know that it exists and how to have access to it)
Use case 2:
Across data source analysis on clinical data
(with agreement by data providers)
Use case 3:
Concordia-style analysis on ALL type of clinical data to posit endo(patho)phenotypes
Use case 4:
Shrine on multiple tranSMARTs (2015?)
Other hospitals
Firewall
Shrine
on
tranSMART
Firewall
Virtual Machine delivered:
tranSMART
+ ETL process
tranSMART
Boston Children's Hospital
Research Connection
Autism cohort integration and uses case Draft 1.1
July 18th 2013

Boston Children's Hospital
Research Connection

Harvard Medical School

cbmi
Center for Biomedical Informatics

Autism Cohort

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

Division of Developmental Medicine

Leonard Rappaport, MD, MS

CH Genomic program

Ellen Hanson, PhD
Stephanie Brewster, MS, CGC
Joanna Reinwald, MS, GC
Franck Jackson

MassGeneral Hospital for Children Lurie Center for Autism

Timothy Yu, MD, PhD

Laboratory of cognitive neuroscience

Charles Nelson, MD, PhD
Vanessa Vogel-Farley
Nicole Coman

The Research Connection

Ingrid Holm, MD, MPH
Sarah Savage, MS, CGC
Catherine Clinton, MS, CGC
Wendy Wolf, PhD
Tram Tran

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

i2b2 / Partners

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

Business Intelligence and Clinical Research Informatics

Jonathan Bickel, MD, MS
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