Creating clinical and 'omics information commons using i2b2 and tranSMART

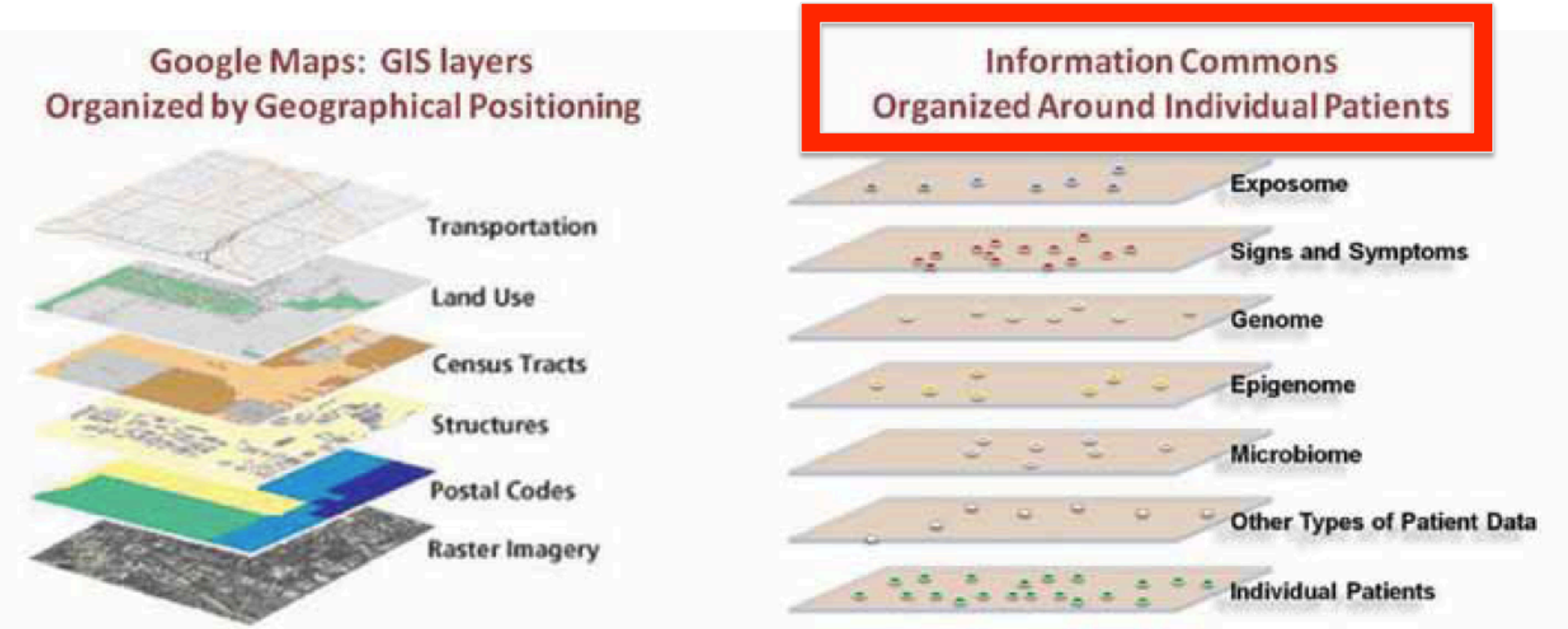
Paul Avillach, MD, PhD

Harvard Medical School

HEGP Hospital, Paris – Paris Descartes University

INSERM UMRS 872 eq 22

Erasmus MC University

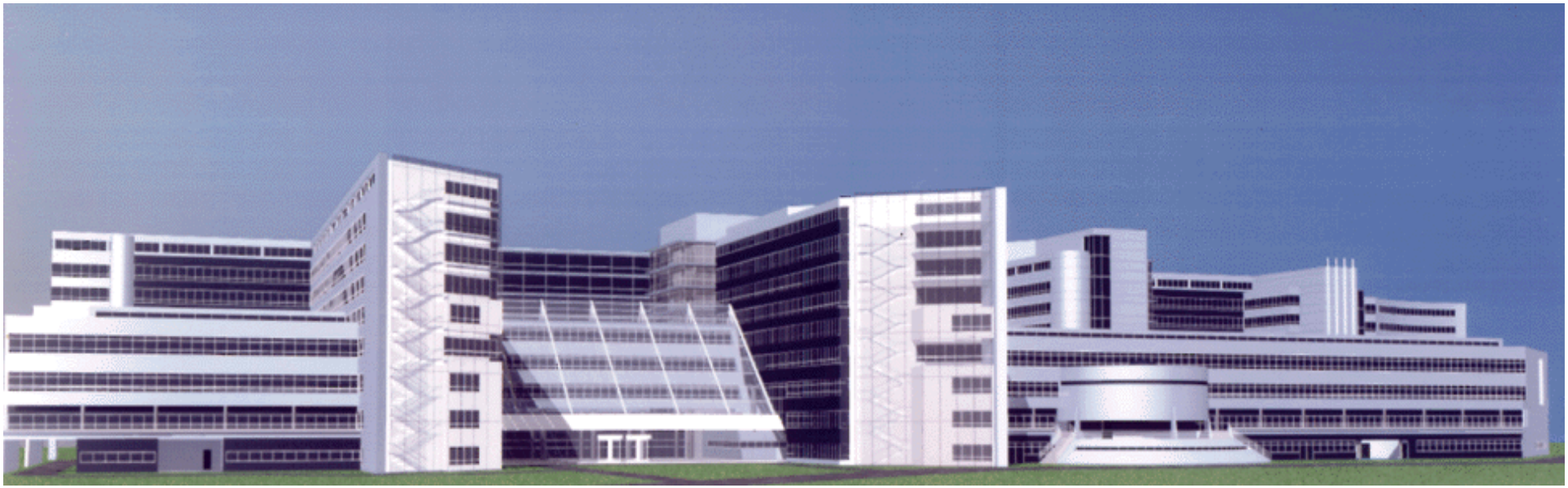


Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease

Report from National academy of science, USA, 2011

HEGP background

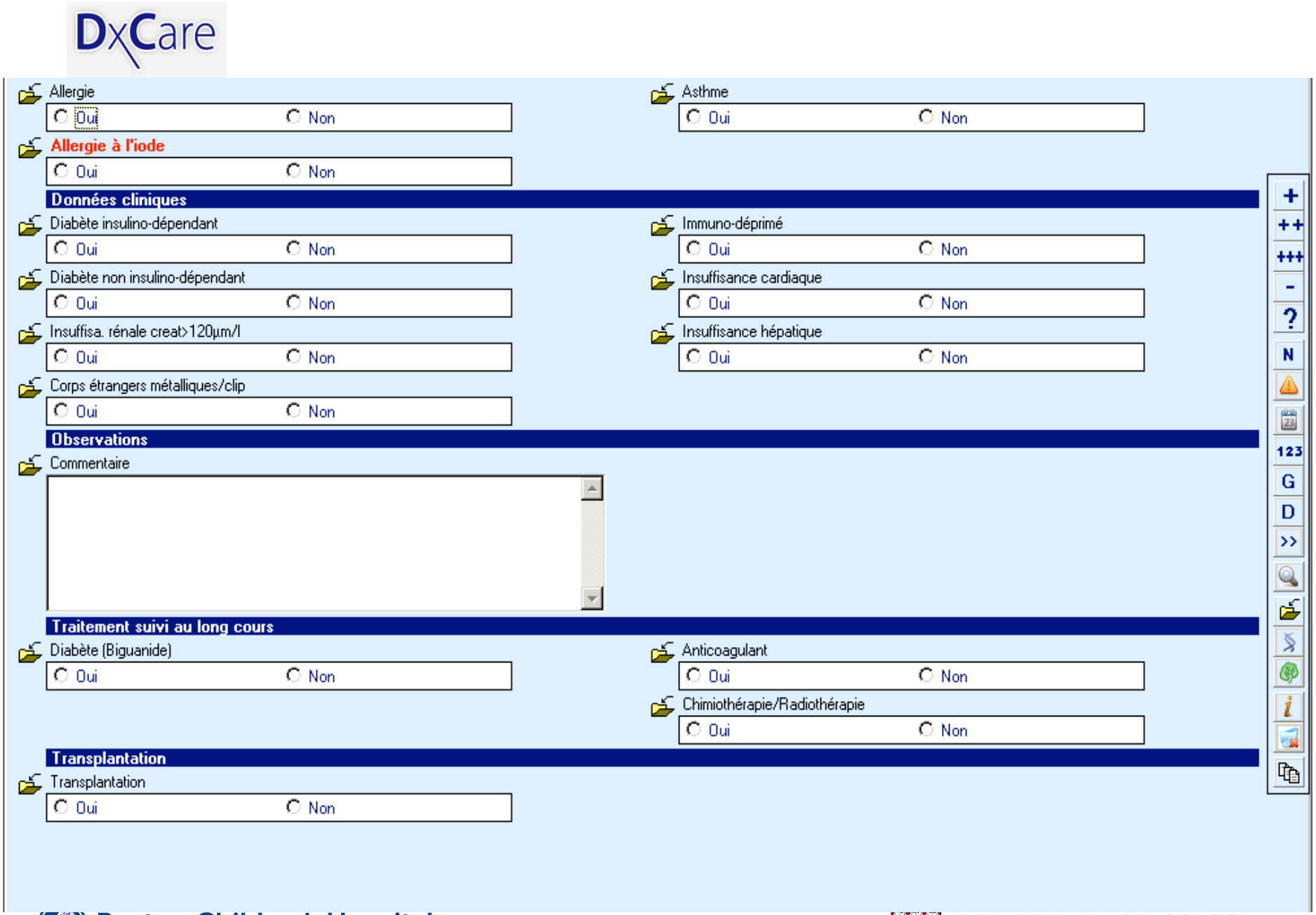
Opening : July 2000-



Hôpital Laennec
(1634)

Hôpital Boucicaut

Hôpital Broussais

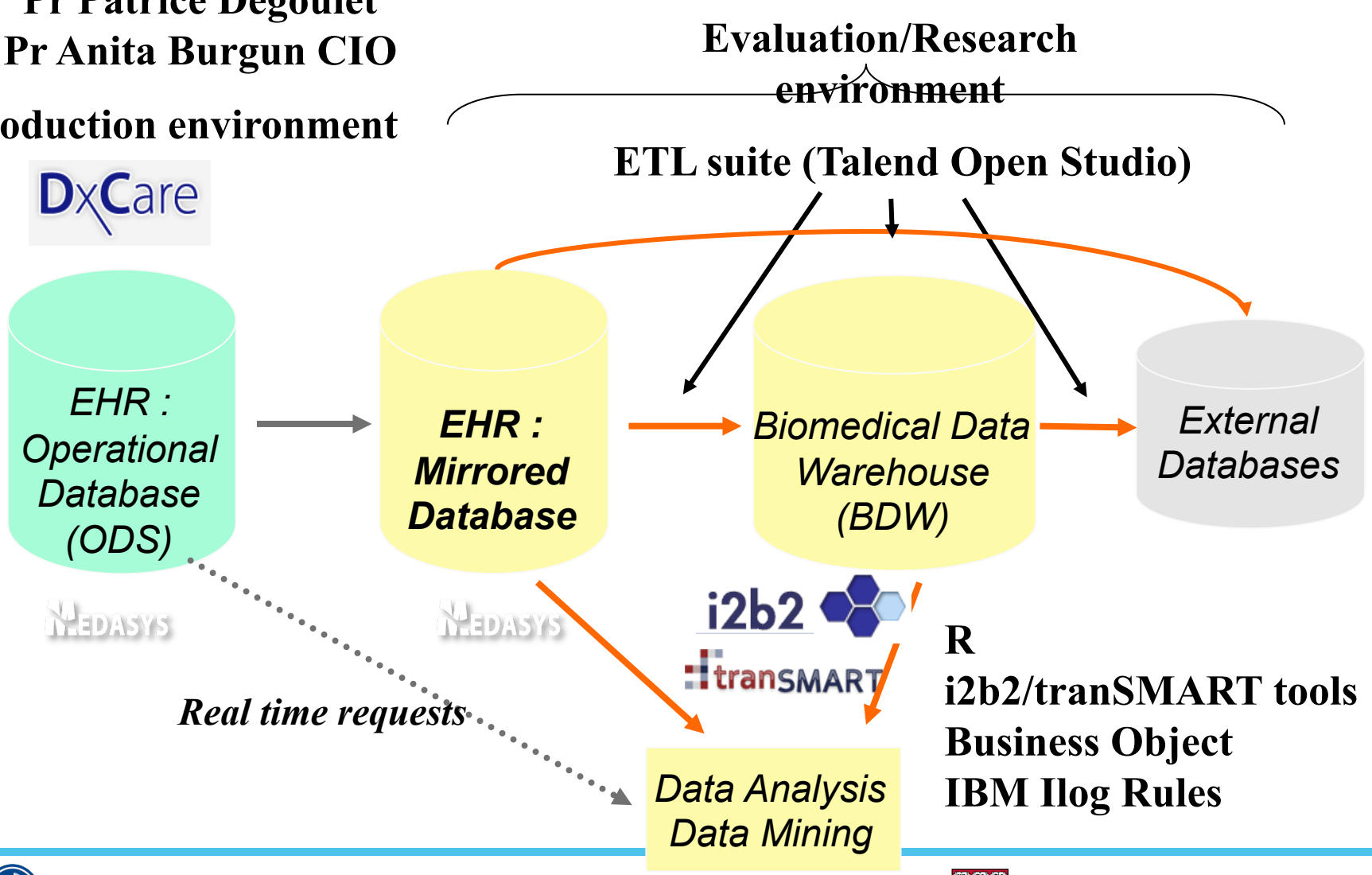

DxCare
Allergie
Oui
Non
Asthme
Oui
Non
Allergie à l'iode
Oui
Non
Données cliniques
Diabète insulino-dépendant
Oui
Non
Immuno-déprimé
Oui
Non
Diabète non insulino-dépendant
Oui
Non
Insuffisance cardiaque
Oui
Non
Insuffisa. rénale creat>120µm/l
Oui
Non
Insuffisance hépatique
Oui
Non
Corps étrangers métalliques/clip
Oui
Non
Observations
Commentaire
Traitement suivi au long cours
Diabète (Biguanide)
Oui
Non
Anticoagulant
Oui
Non
Chimiothérapie/Radiothérapie
Oui
Non
Transplantation
Transplantation
Oui
Non
+
++
+++
-
?
N
123
G
D
>>

HEGP BDW

EHR/BDW integration

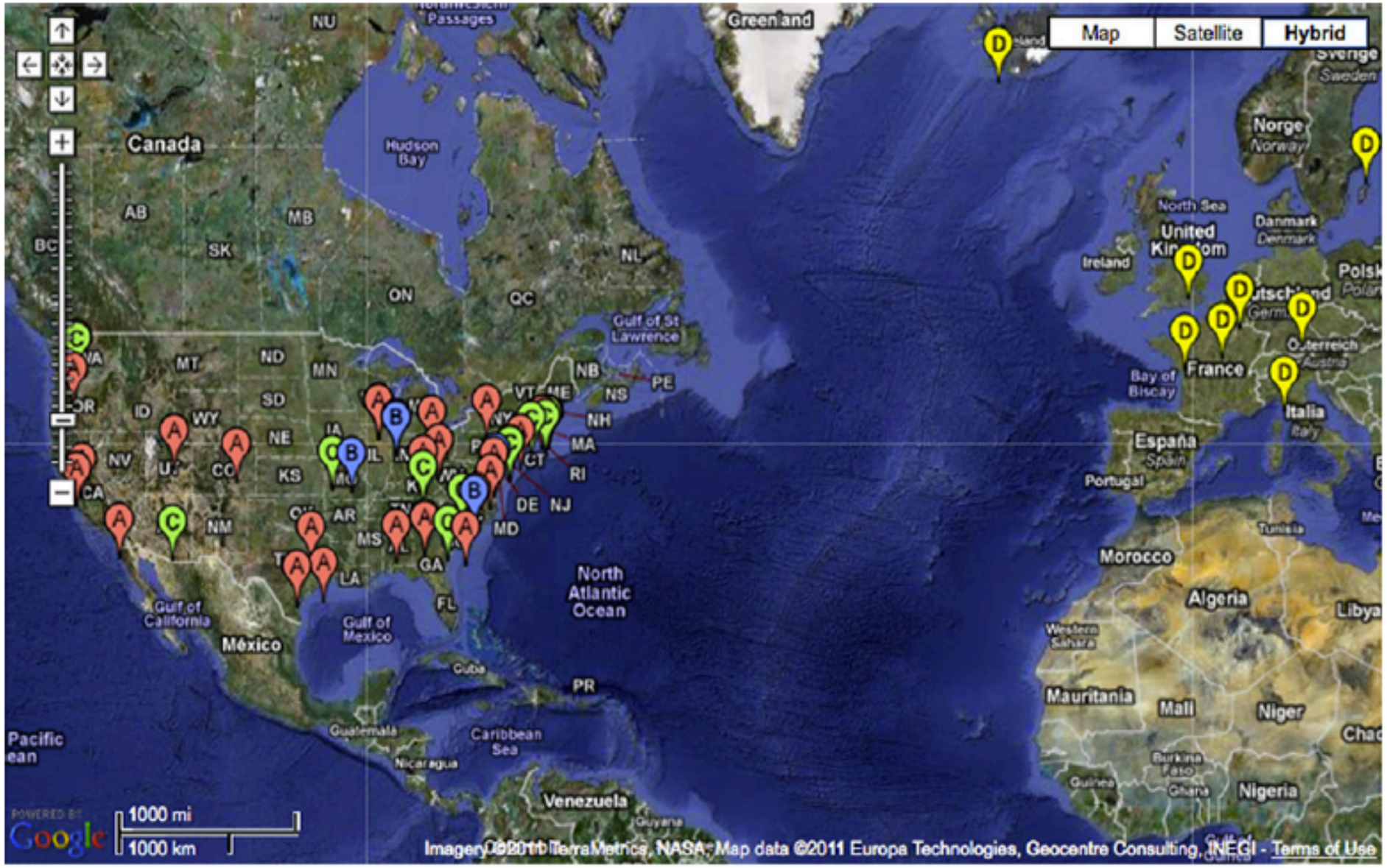
Pr Patrice Degoulet
Pr Anita Burgun CIO

Production environment



i2b2

PI: Isaac Kohane – Harvard - CBMI



Boston Children's Hospital
Until every child is well™

HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

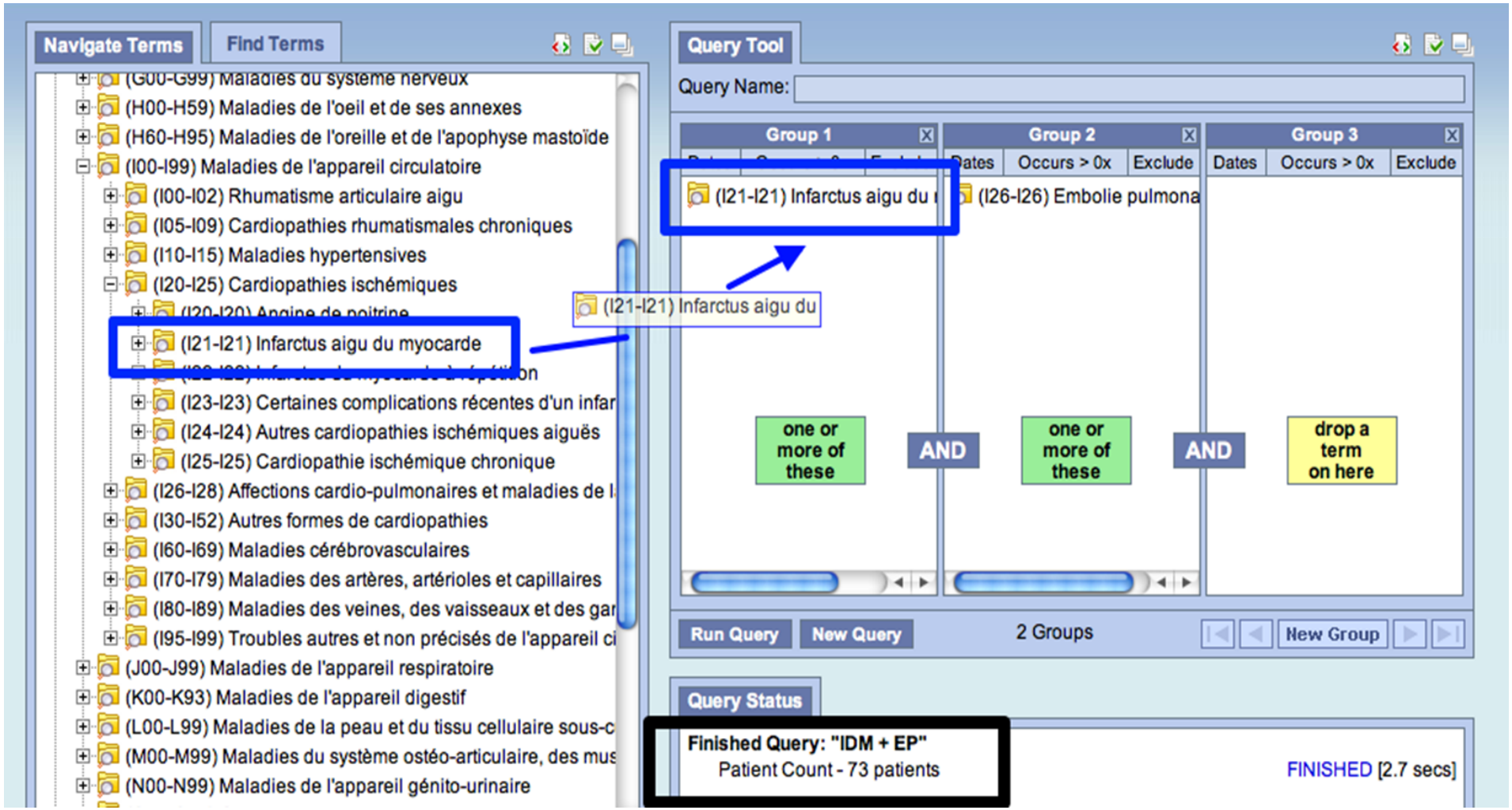
HEGP i2b2

Type of data

		H	OV	Start date	*# unique patients*	*# values*
	Demographic (age, sex, Hospital vital status)	X	X	1971	**606 524**	
CLINICAL	Vital signs (temperature, blood pressure, weight, …)	X	X	2000	141 164	14 213 951
CLINICAL	Diagnostic codes (DRG ICD10)	X		1995	305 369	2 626 792
CLINICAL	Procedures (French CCAM codes)	X		2004	241 482	3 200 482
CLINICAL	Clinical data (DxCare questionnaires)	X	X	1971	391 218	46 506 217
CLINICAL	Free text reports*: Hospitalization, Surgery, consultations, …	X	X	2004	289 614	1 961 985
CLINICAL	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

HEGP CDW 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)

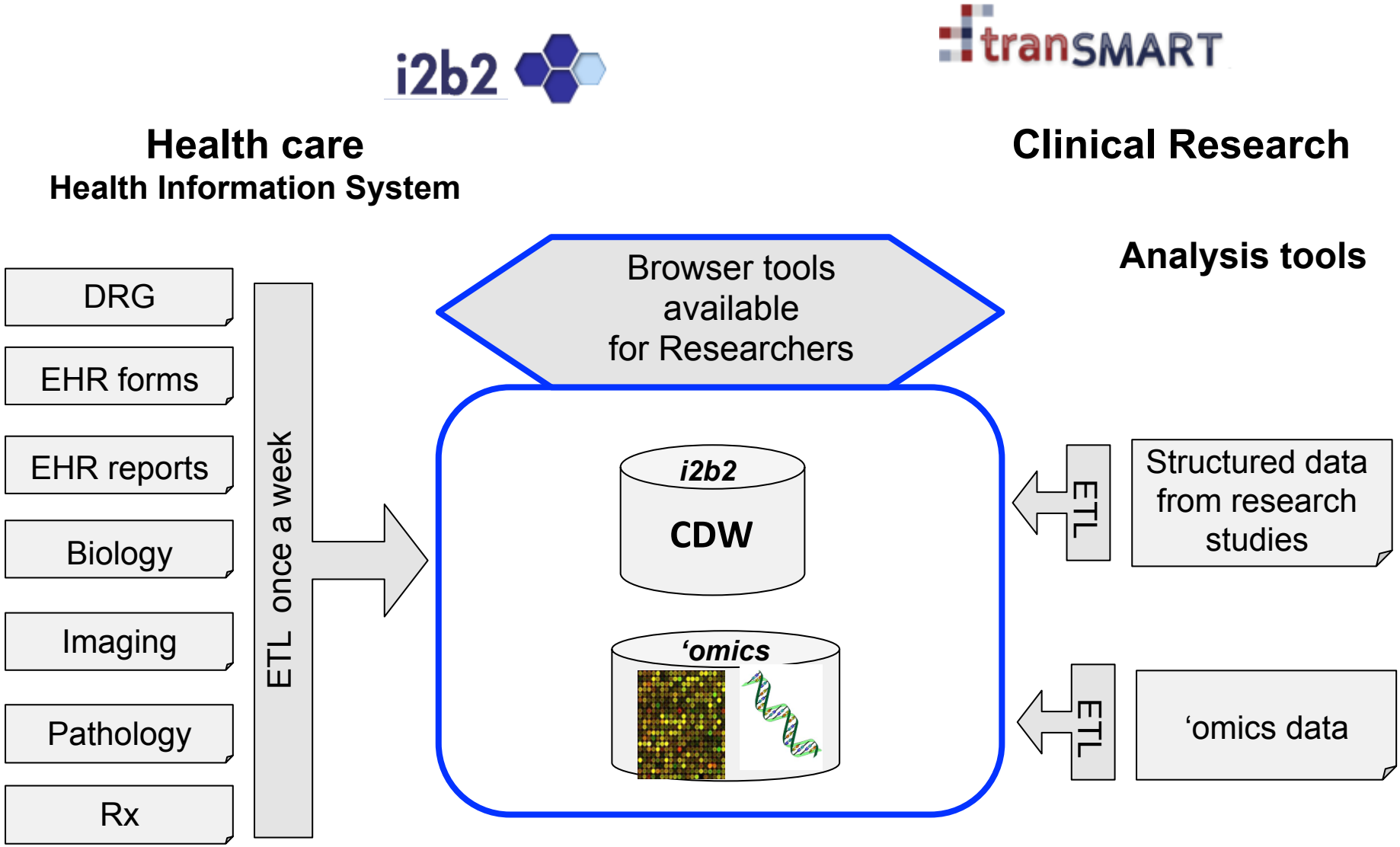
Access rights

- **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


i2b2
tranSMART
Health care
Health Information System
Clinical Research
Analysis tools
DRG
EHR forms
EHR reports
Biology
Imaging
Pathology
Rx
ETL once a week
Browser tools available for Researchers
i2b2
CDW
'omics
ETL
Structured data from research studies
ETL
'omics data

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 May, 2012

- Today, driven and maintained by the tranSMART Foundation & communitie

http://transmartfoundation.org

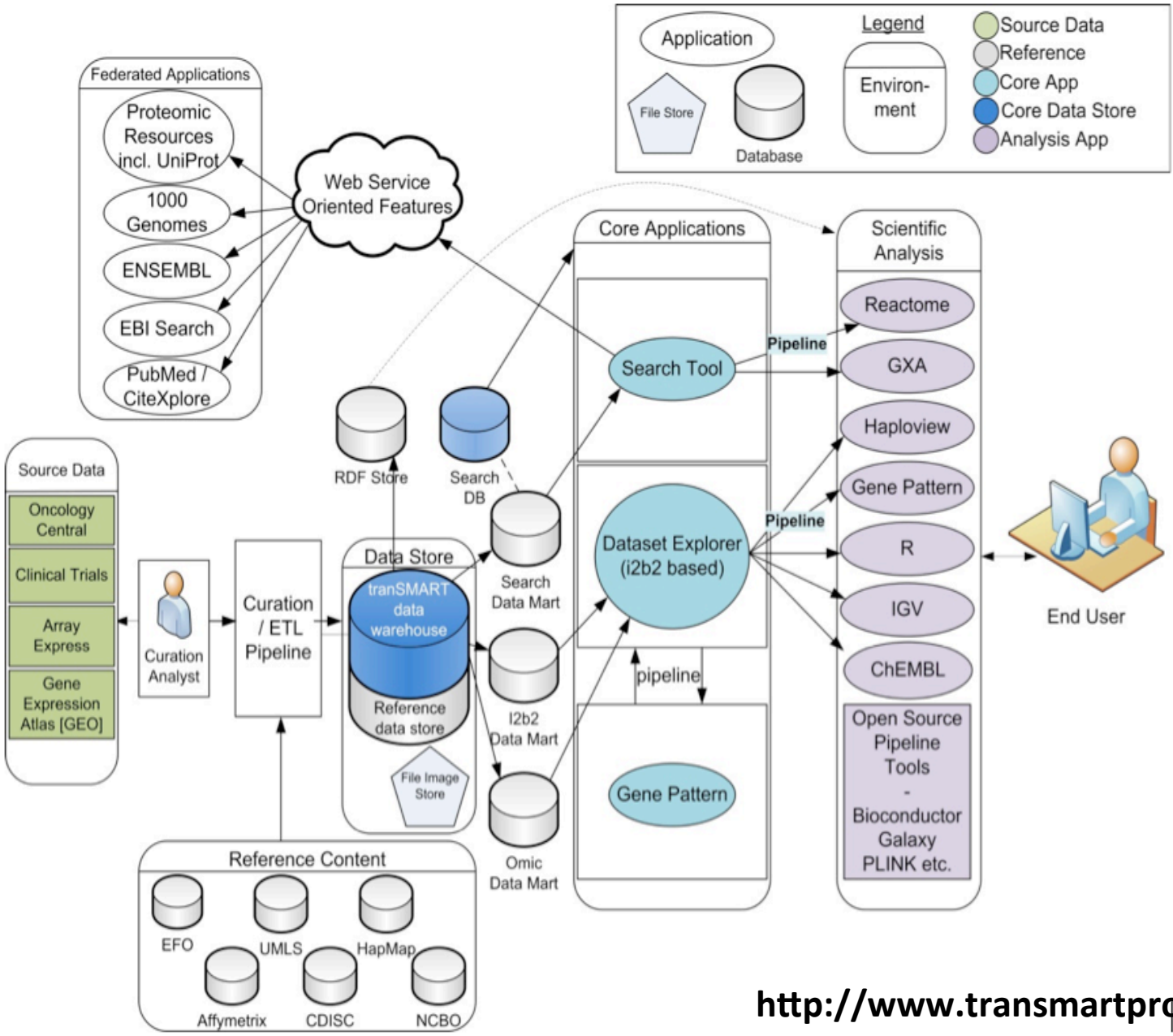
tranSMART

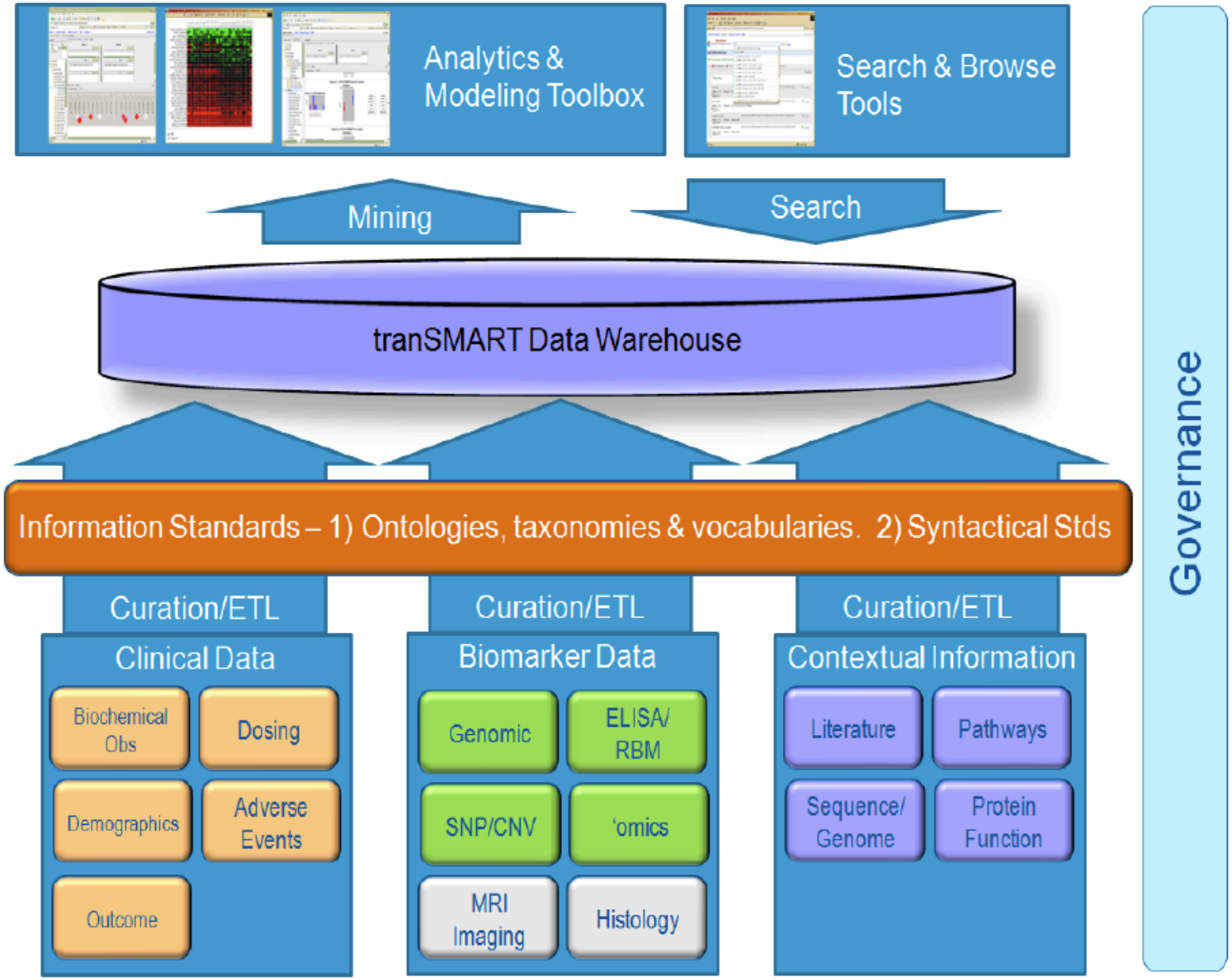
Objectives :

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

"Omics" data integration

tranSMART





- Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med. 2010; 8:68.

Pilot Study
HEGP - Paris

tranSMART

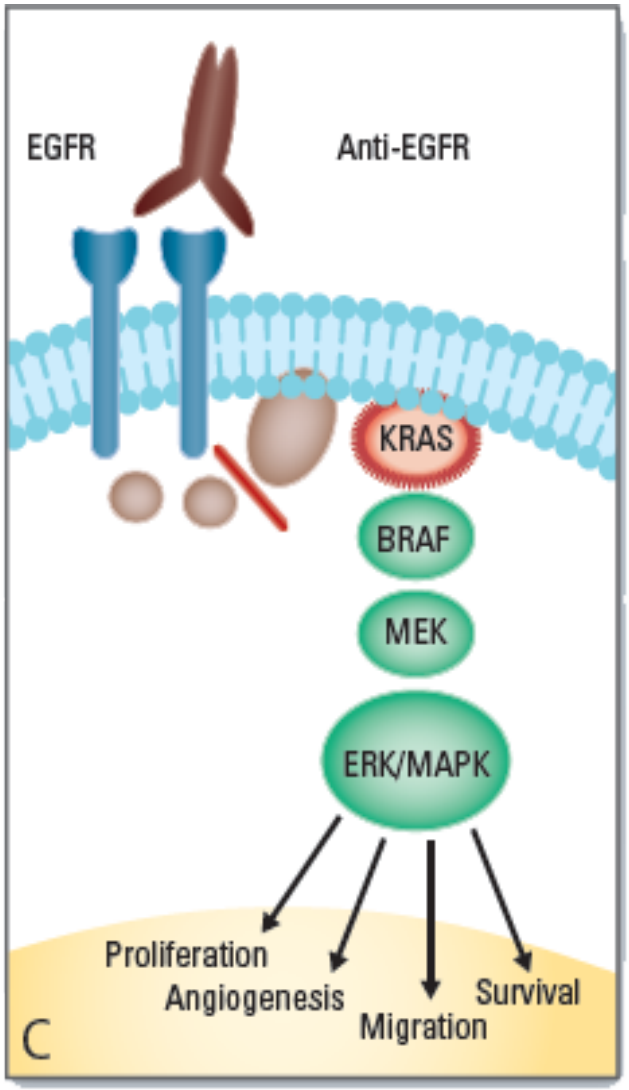
VOLUME 27 · NUMBER 35 · DECEMBER 10 2009

JOURNAL OF CLINICAL ONCOLOGY

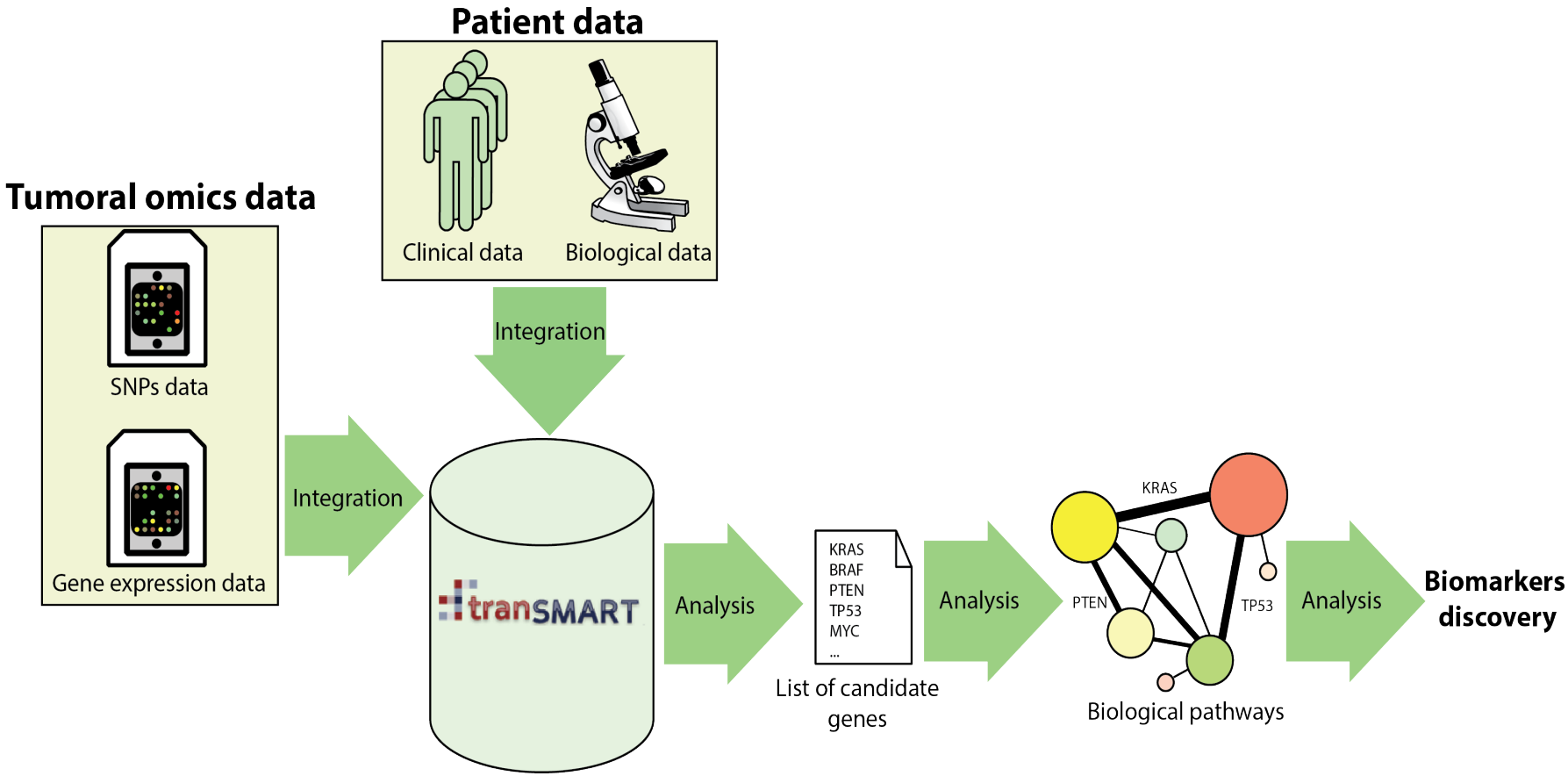
ORIGINAL REPORT

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

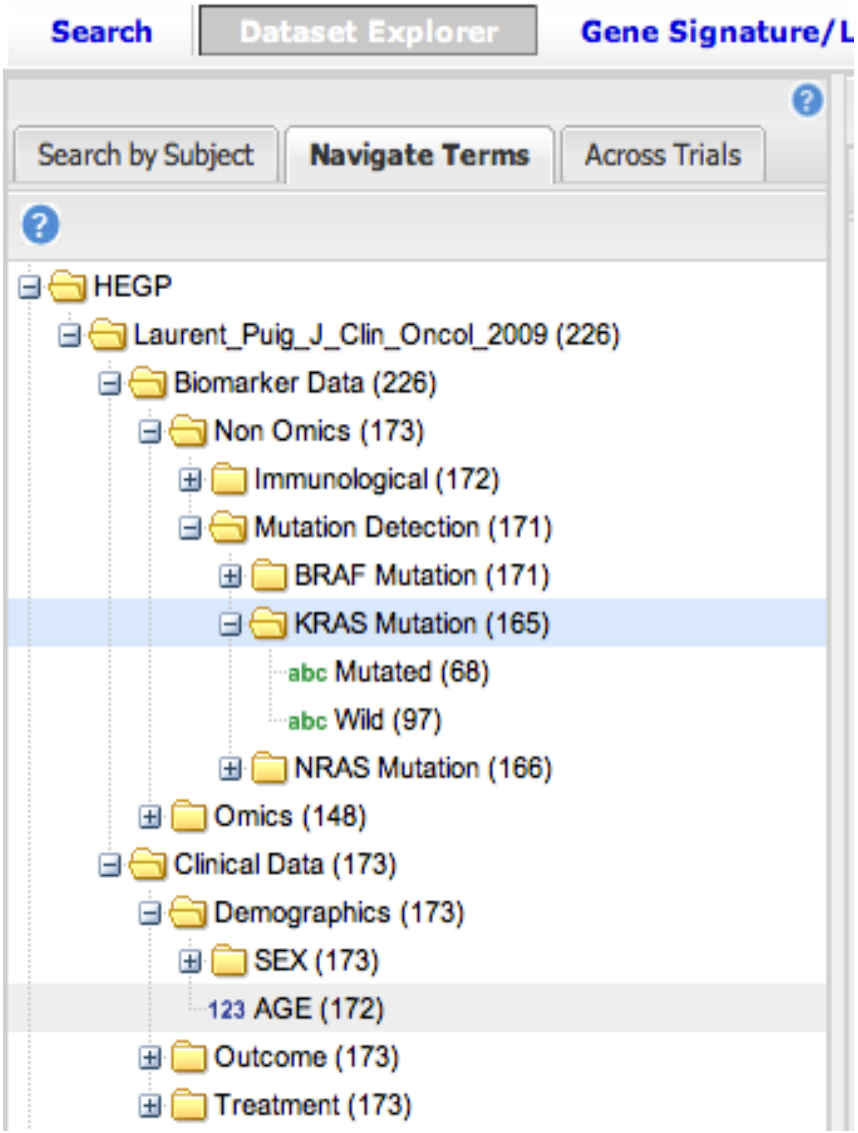
Pierre Laurent-Puig, Anne Cayre, Gilles Manceau, Emmanuel Buc, Jean-Baptiste Bachet, Thierry Lecomte, Philippe Rougier, Astrid Lievre, Bruno Landi, Valérie Boige, Michel Ducreux, Marc Ychou, Fréderic Bibeau, Olivier Bouché, Julia Reid, Steven Stone, and Frédérique Penault-Llorca


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

tranSMART



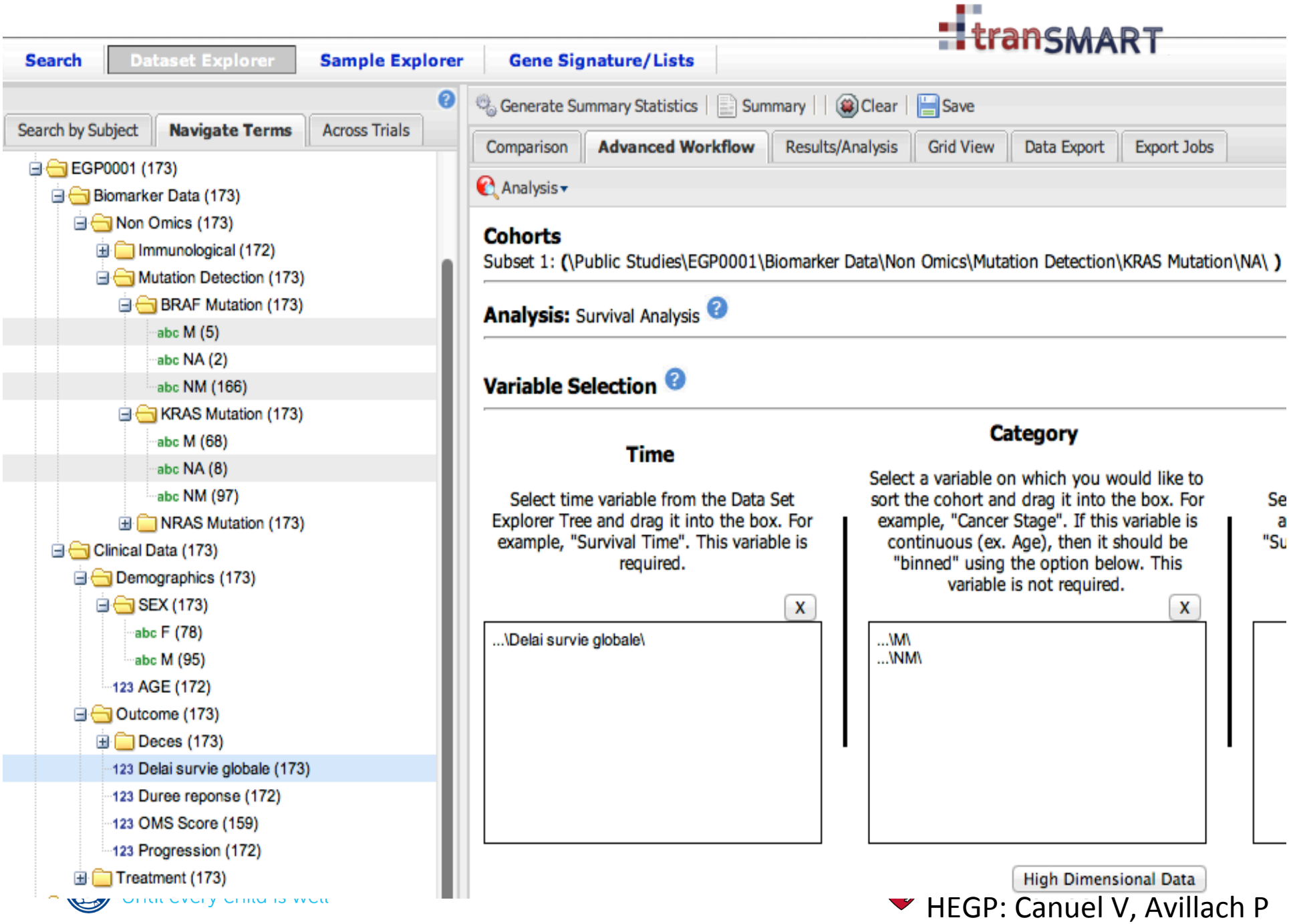
HEGP: Canuel V, Avillach P



HEGP: Canuel V, Avillach P

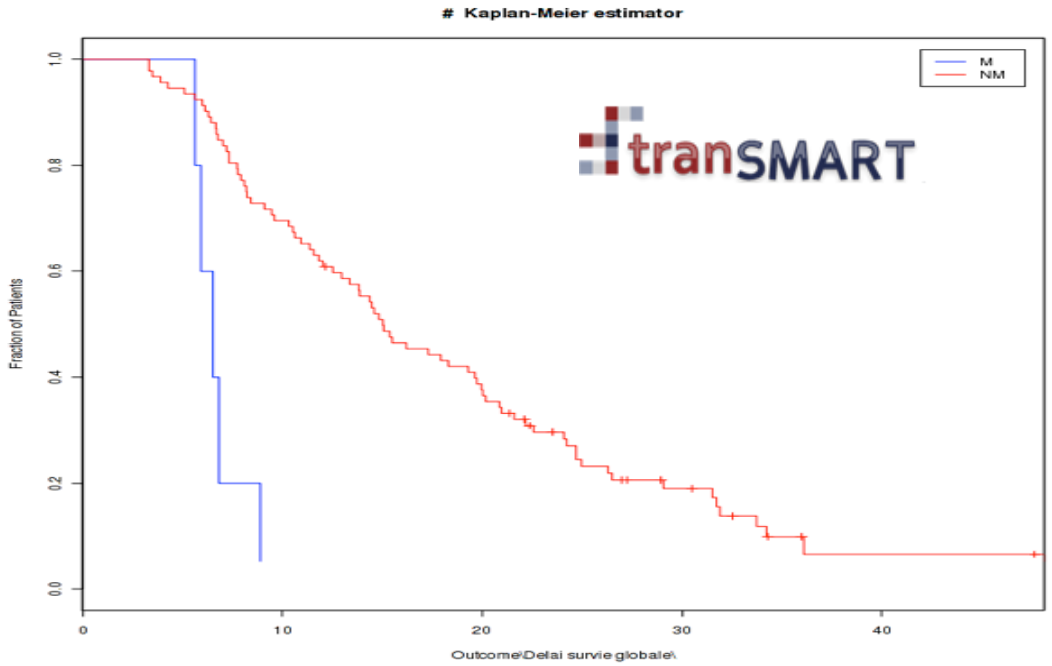
NÂ°interne ADN	Age	Sexe	BILANKRAS	Mutation BRAF	BILAN MUT NRAS	nbligne	Toxicit	Meilleure reponse	Progression	Duree rep	DÃ©cÃ¨s	dÃ©lai survie globale	OMS	EGFR copy number	Score de HIRSCH	Mutation PIK3CA	PTEN_CYTO	PTEN_MB	PTEN_NX
1	[illegible]	M		NM	NM	1	2	1	1	58.14	oui	26.3	1	10	positif	NM	80	20	20
5	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			

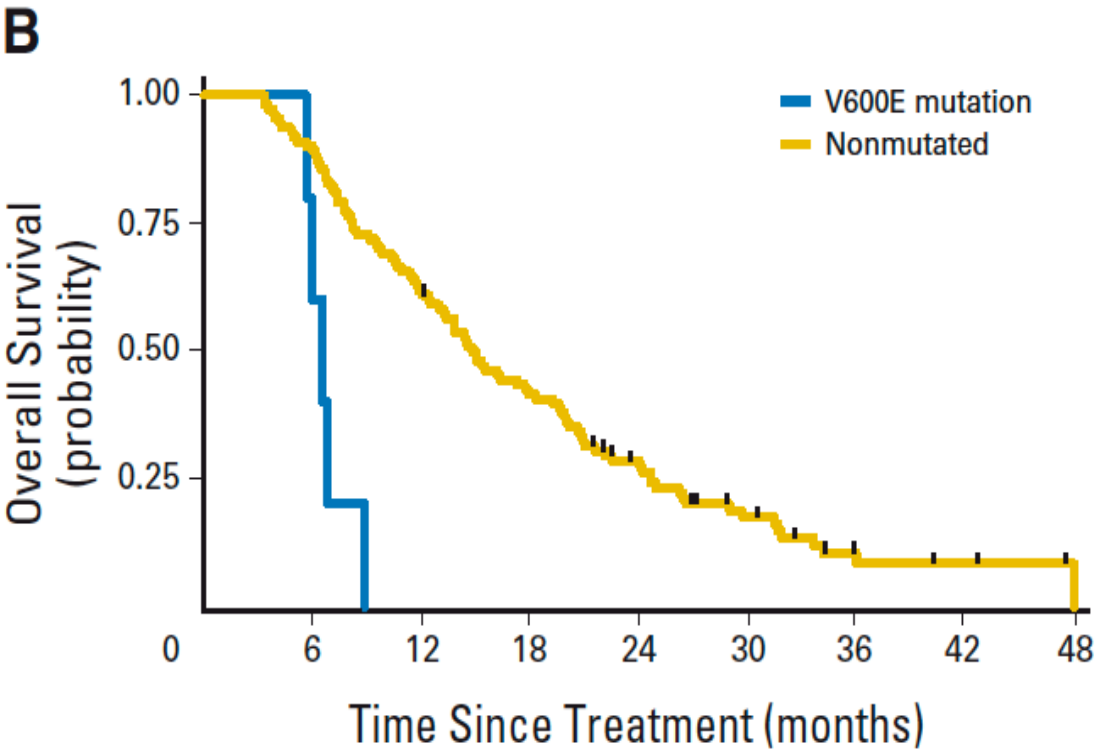


HEGP: Canuel V, Avillach P

- R module in tranSMART

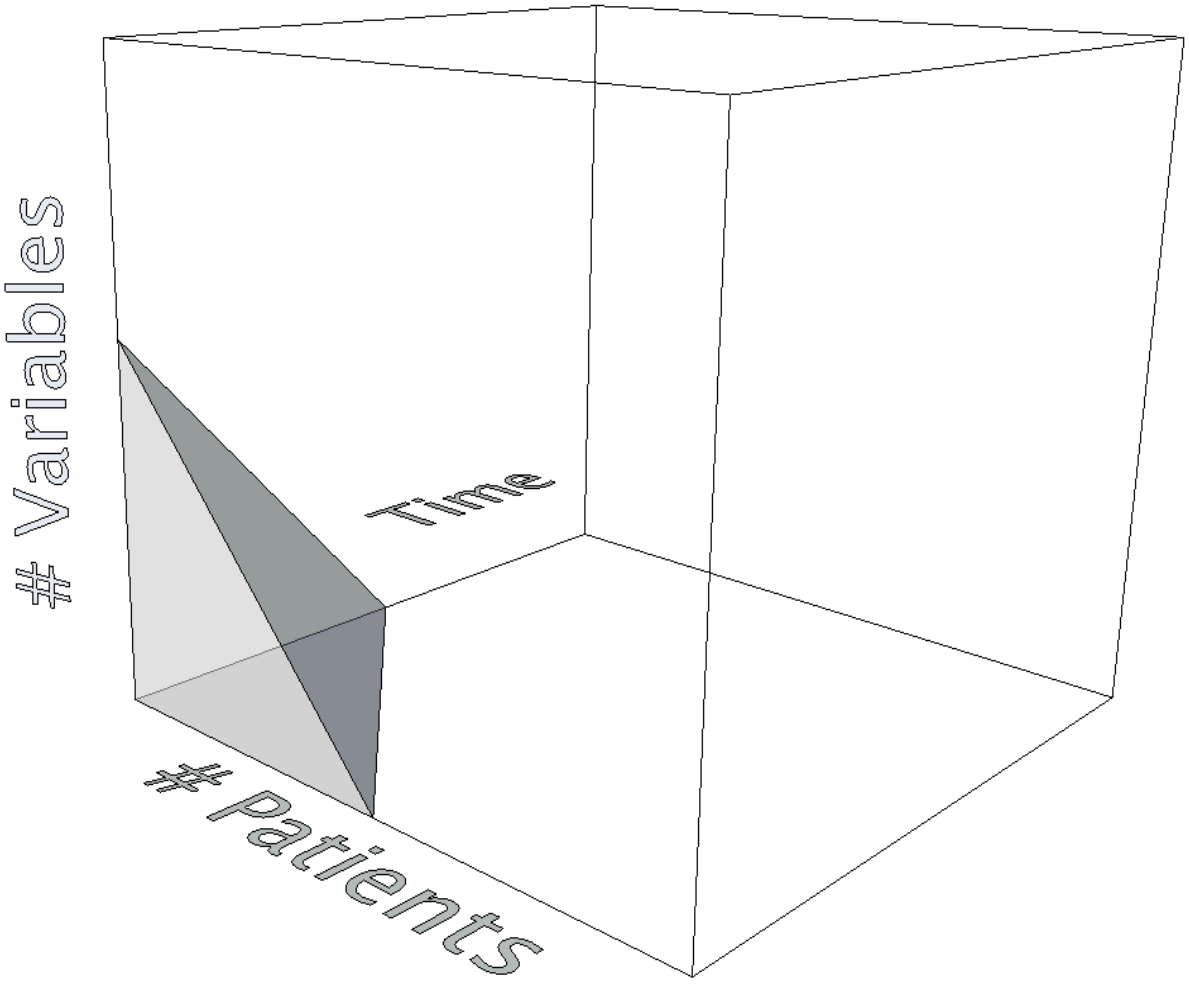


- Published figure in JCO

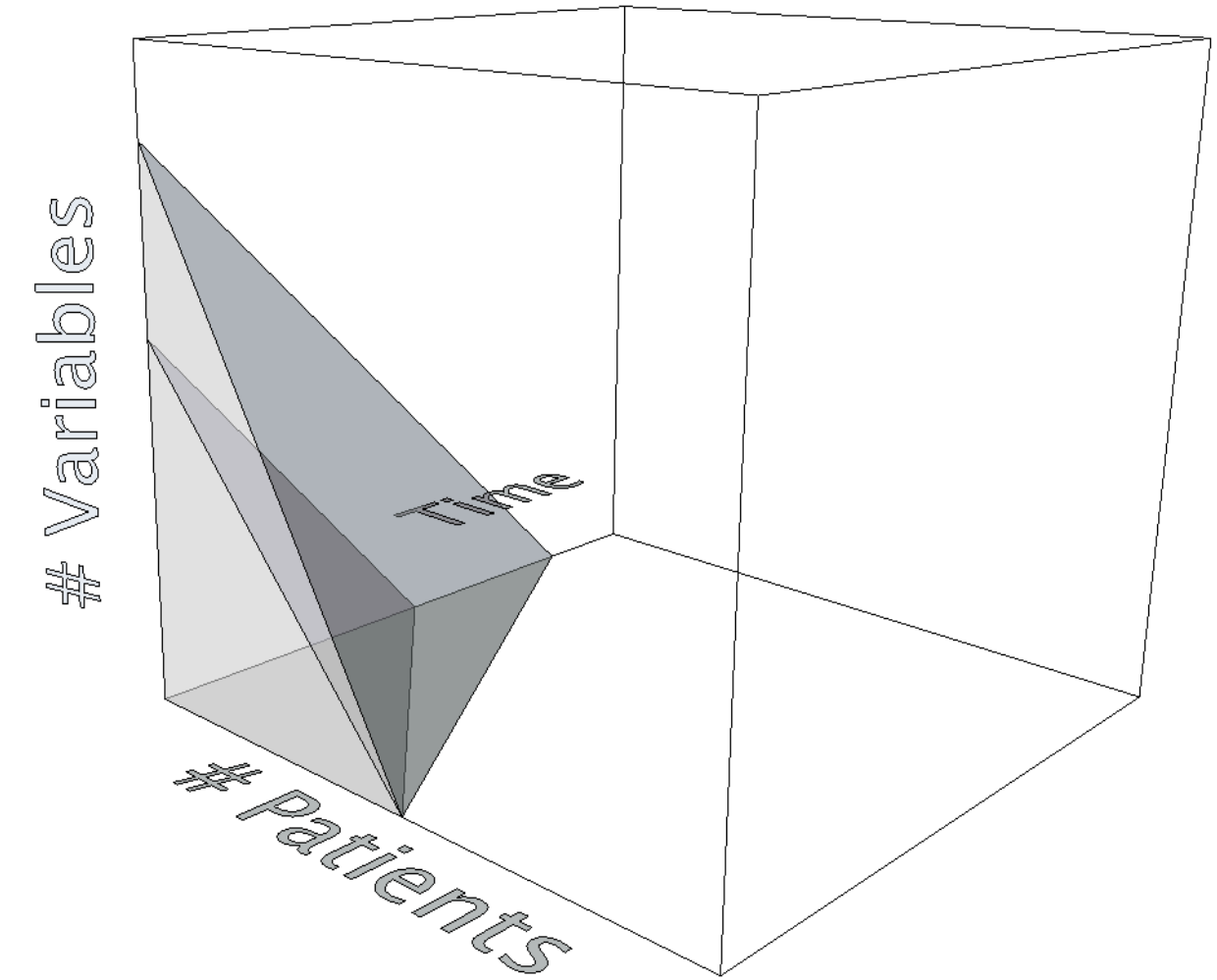


HEGP: Canuel V, Avillach P

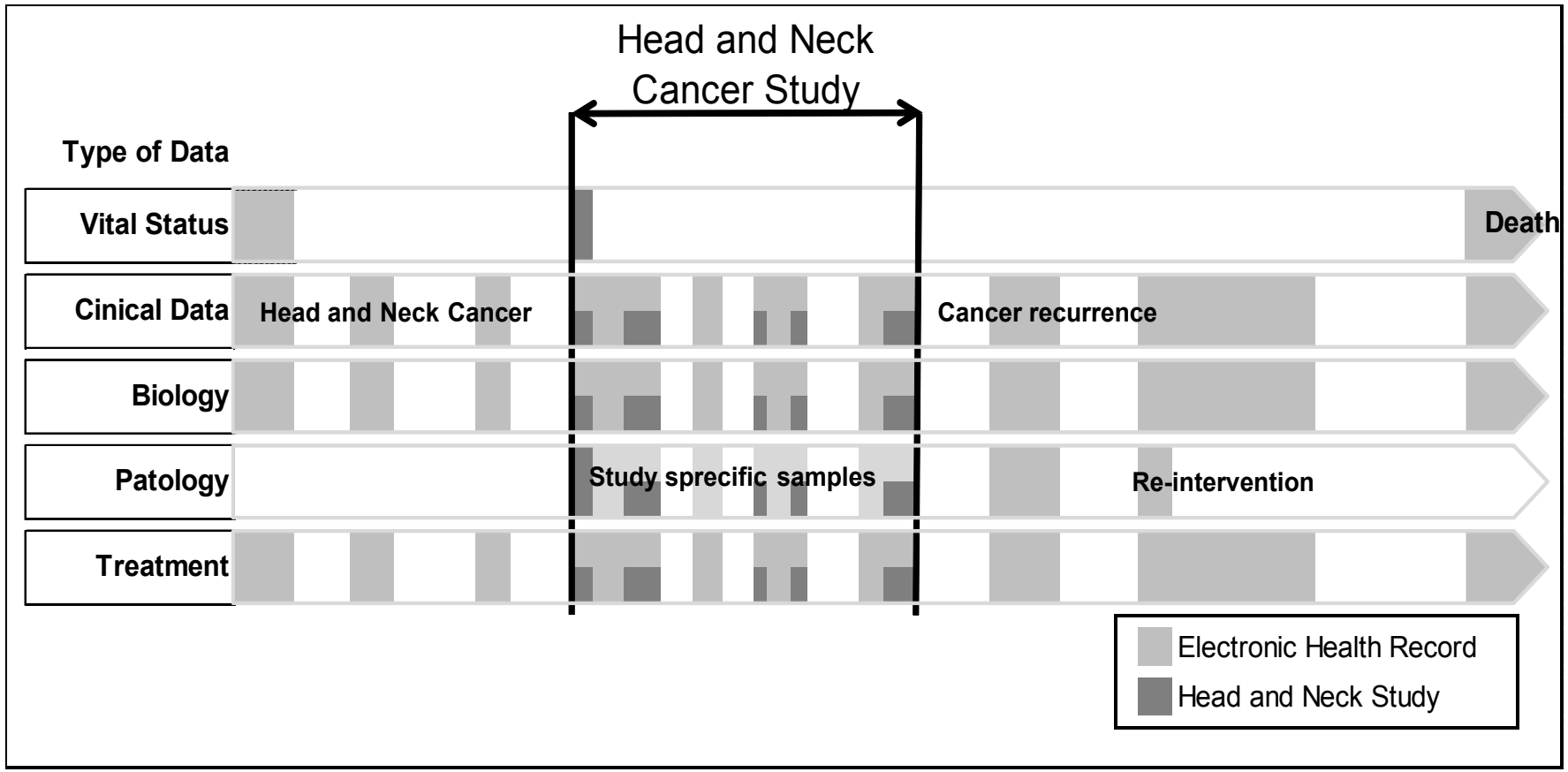
Phenotypic augmentation



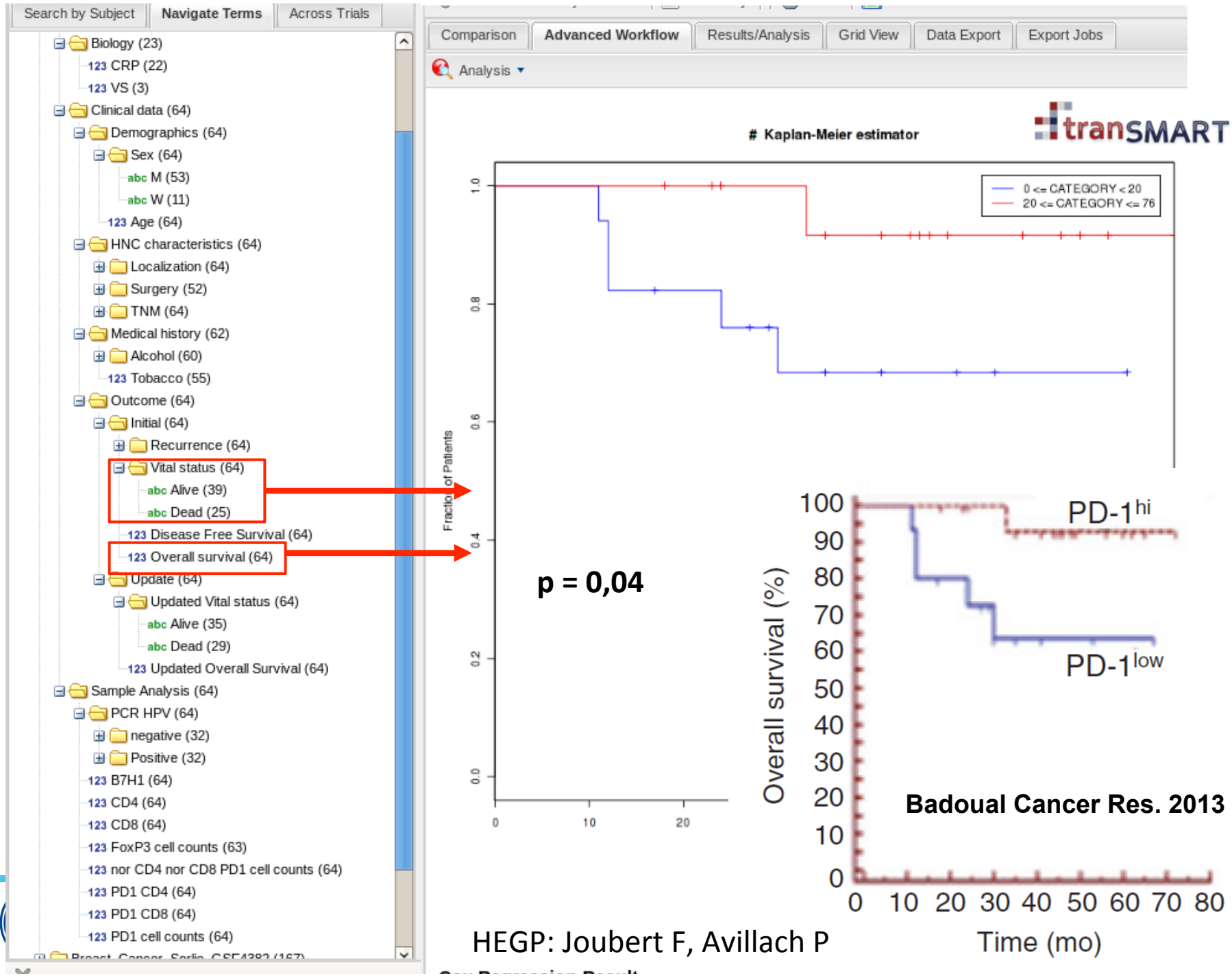
Phenotypic augmentation

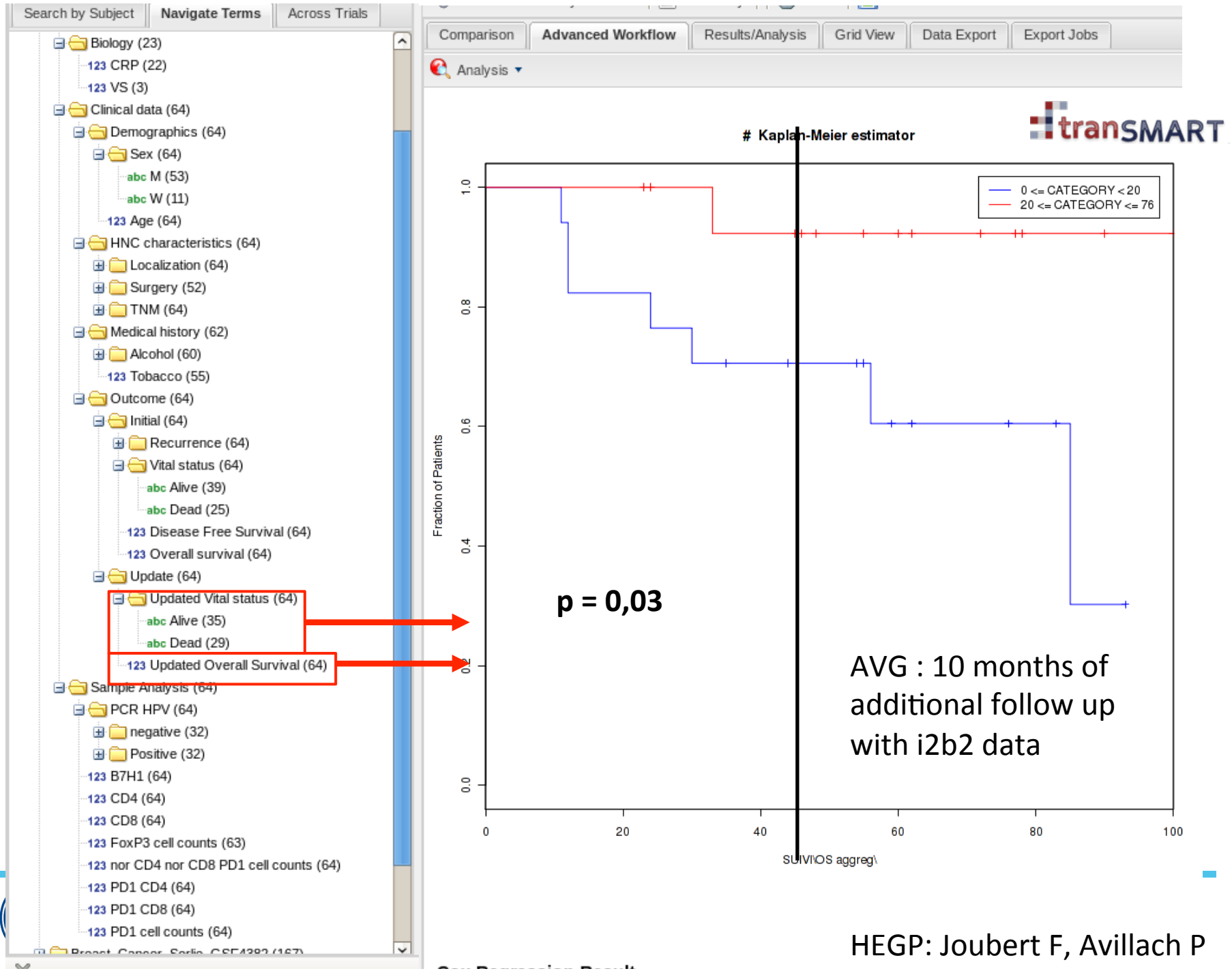


tranSMART + i2b2 = Phenotypic augmentation



HEGP: Joubert F, Avillach P


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
Kaplan-Meier estimator
tranSMART
0 <= CATEGORY < 20
20 <= CATEGORY <= 76
Fraction of Patients
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
p = 0,04
Overall survival (%)
100
90
80
70
60
50
40
30
20
10
0
PD-1hi
PD-1low
0 10 20 30 40 50 60 70 80
Time (mo)
Badoual Cancer Res. 2013
HEGP: Joubert F, Avillach P

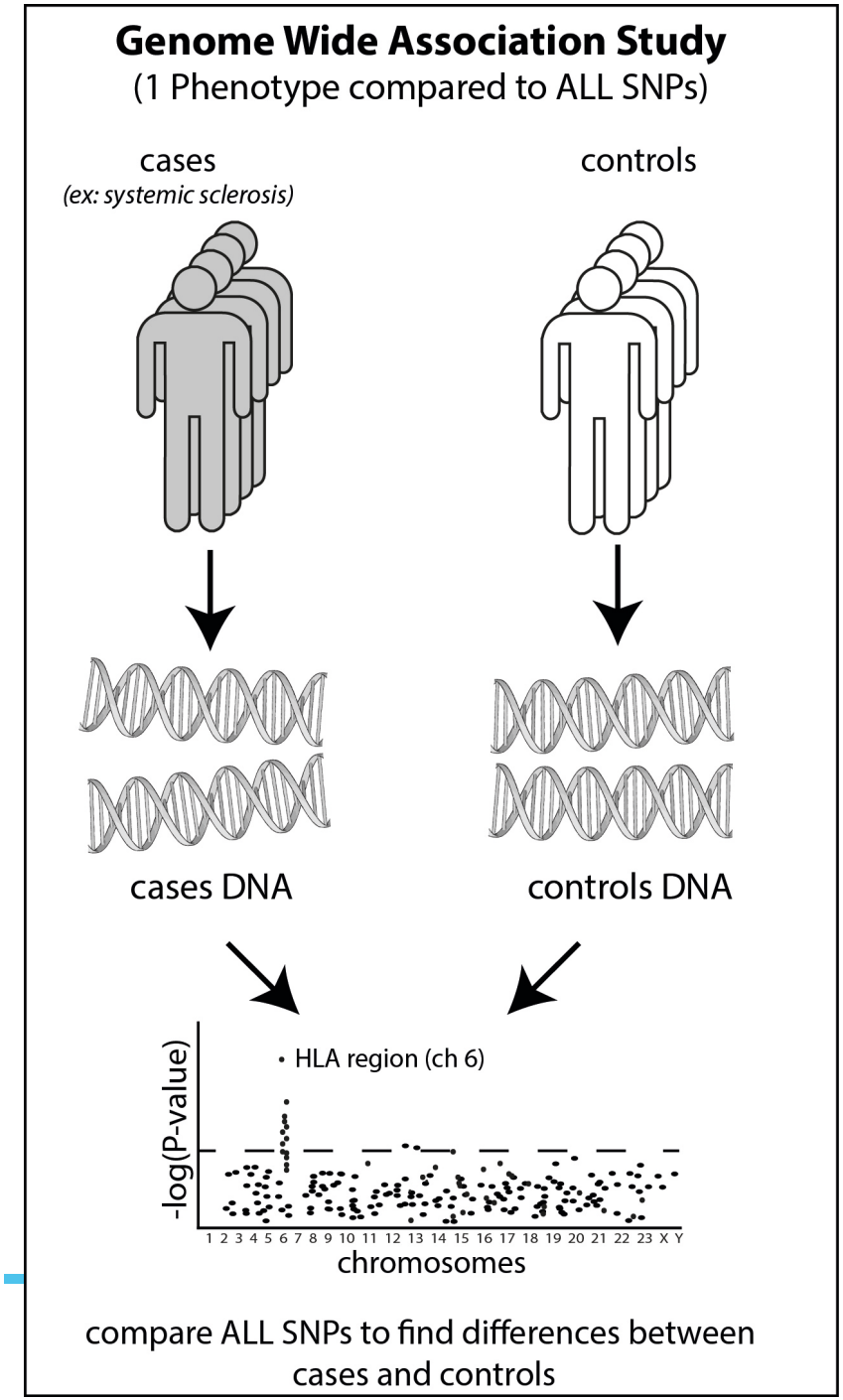

Search by Subject
Navigate Terms
Across Trials
Biology (23)
123 CRP (22)
123 VS (3)
Clinical data (64)
Demographics (64)
Sex (64)
abc M (53)
abc W (11)
123 Age (64)
HNC characteristics (64)
Localization (64)
Surgery (52)
TNM (64)
Medical history (62)
Alcohol (60)
123 Tobacco (55)
Outcome (64)
Initial (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
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
SUIVI\OS aggreg\
p = 0,03
AVG : 10 months of additional follow up with i2b2 data
HEGP: Joubert F, Avillach P

PheWAS study
on i2b2/tranSMART

Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, Degoulet P, Burgun A, Loriot MA, **Avillach P**.

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

PLoS Comput Biol. 2013 *in press*


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

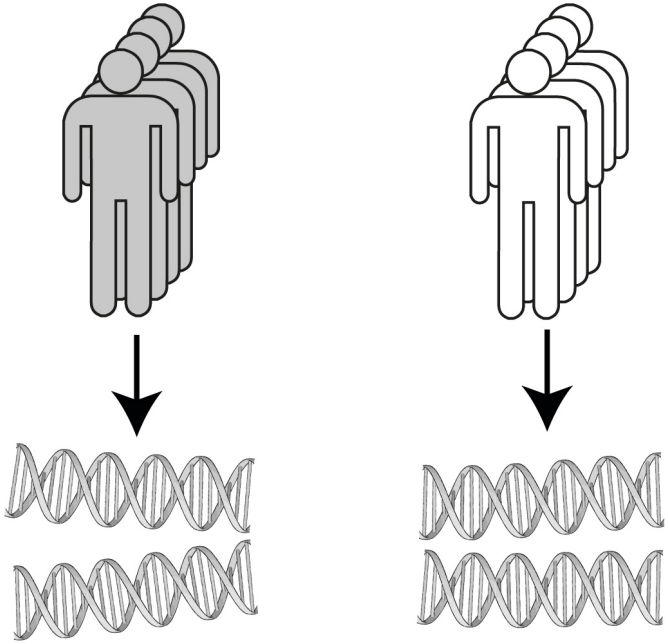
Genome Wide Association Study

(1 Phenotype compared to ALL SNPs)

cases

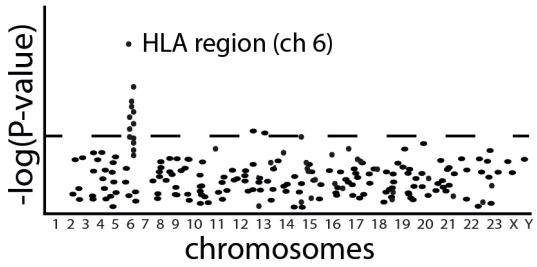
(ex: systemic sclerosis)

controls



cases DNA

controls DNA



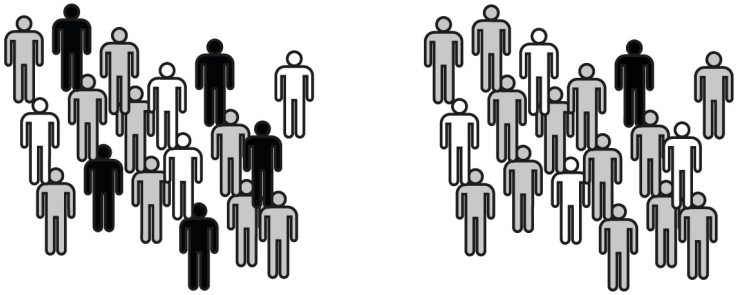
compare ALL SNPs to find differences between cases and controls

Phenome Wide Association Study

(1 SNP compared to ALL Phenotypes)

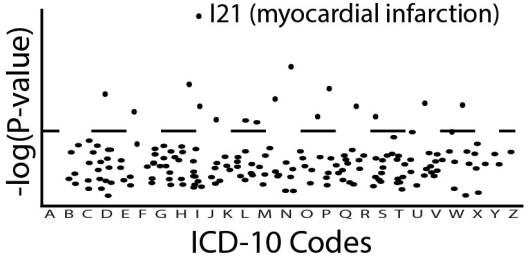
allele G patients group

allele A patients group



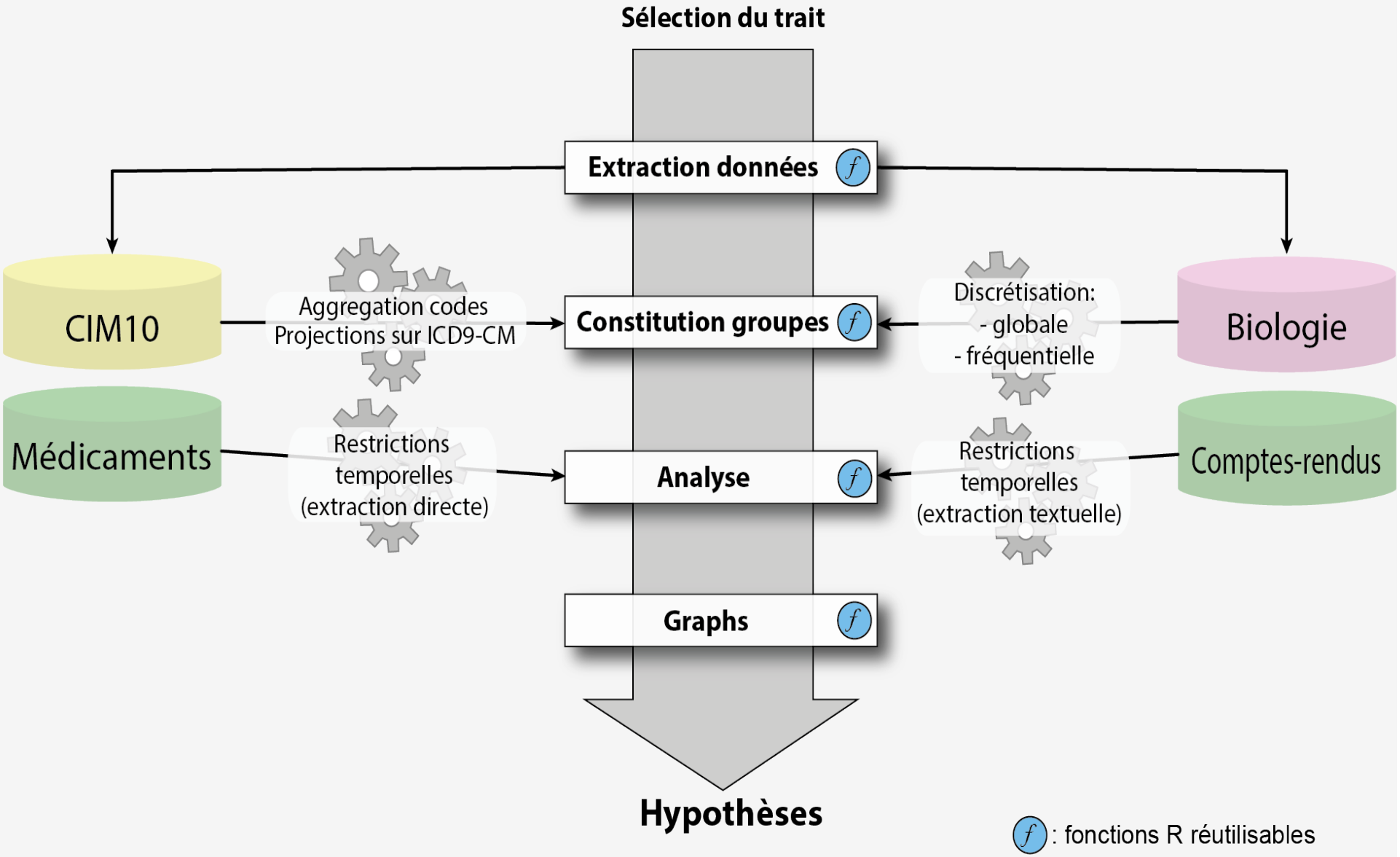
allele G patients phenotype

allele A patients phenotype

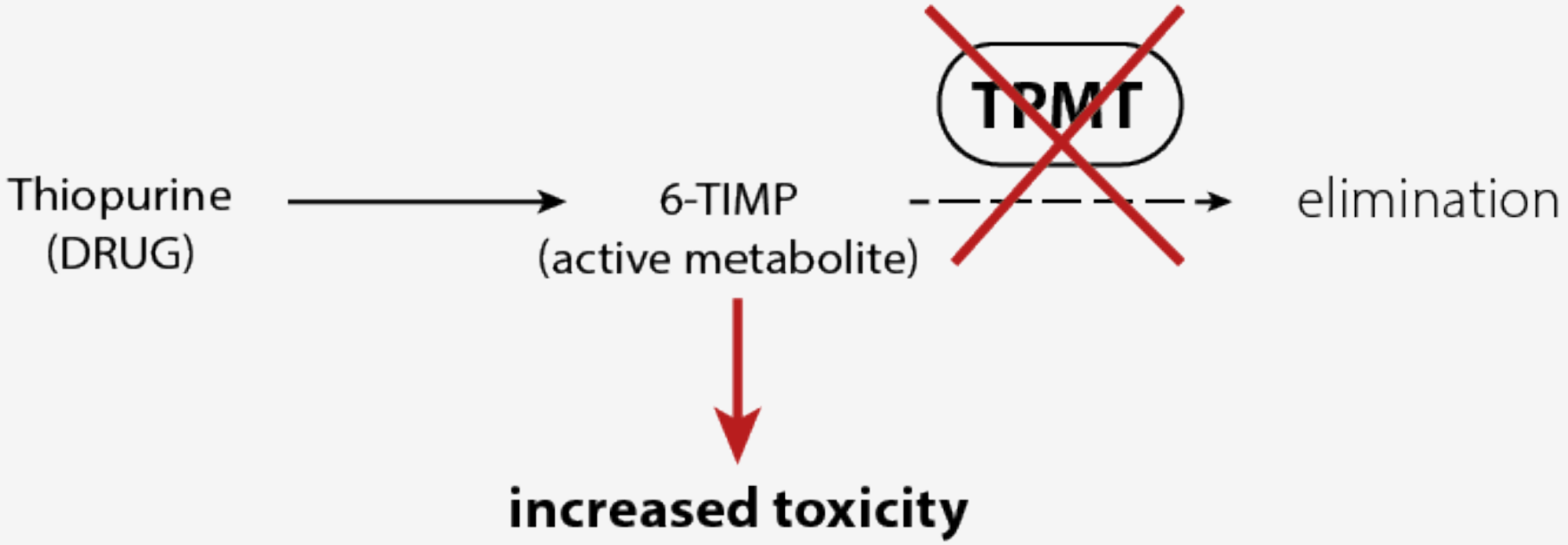


compare ALL DIAGNOSIS to find differences between cases and controls

Workflow


Sélection du trait
Extraction données
CIM10
Médicaments
Biologie
Comptes-rendus
Aggregation codes
Projections sur ICD9-CM
Constitution groupes
Discrétisation:
- globale
- fréquentielle
Restrictions
temporelles
(extraction directe)
Analyse
Restrictions
temporelles
(extraction textuelle)
Graphs
Hypothèses
f : fonctions R réutilisables

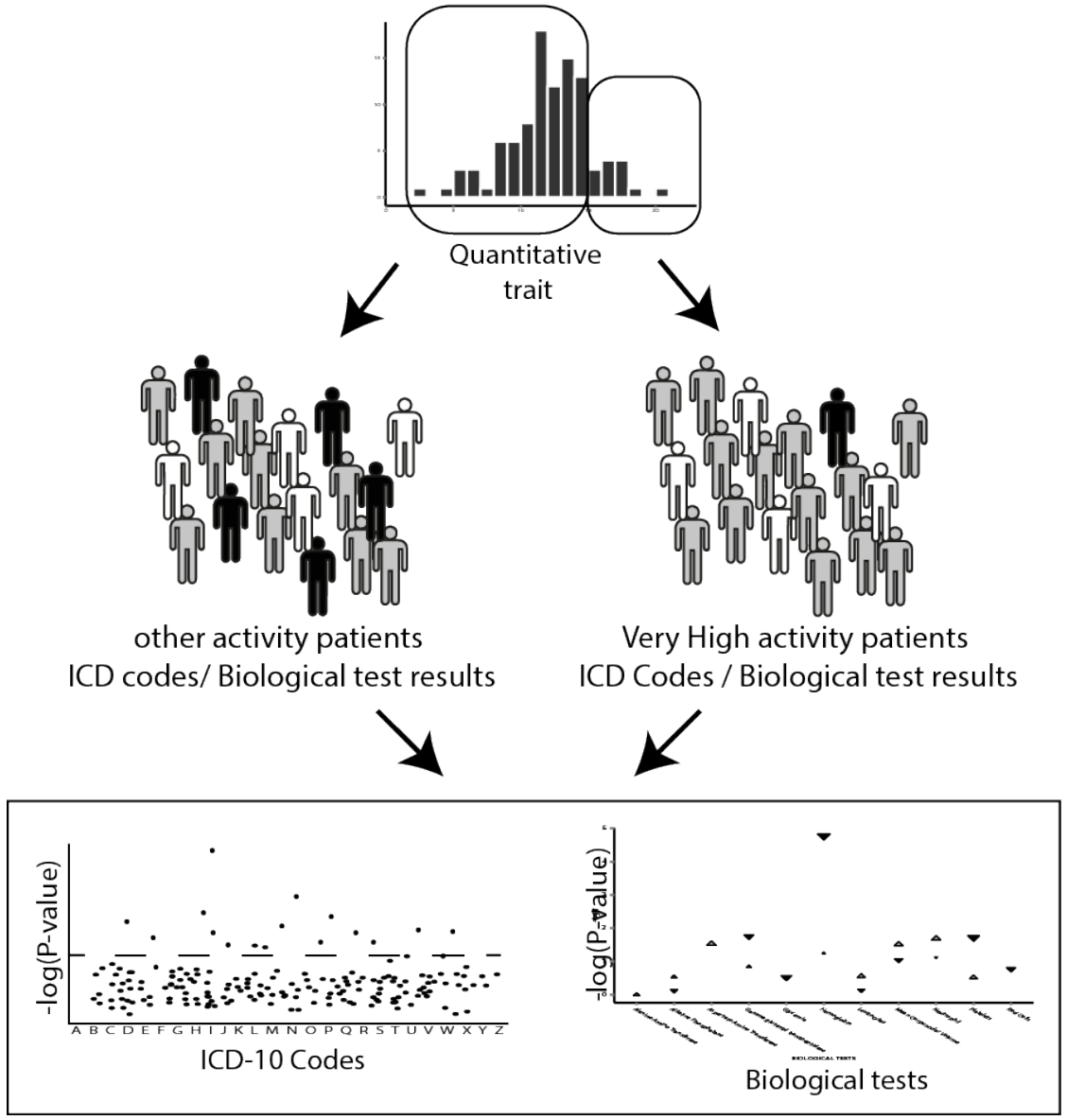
Methodes: Selection of trait: enzymatic Activity TPMT



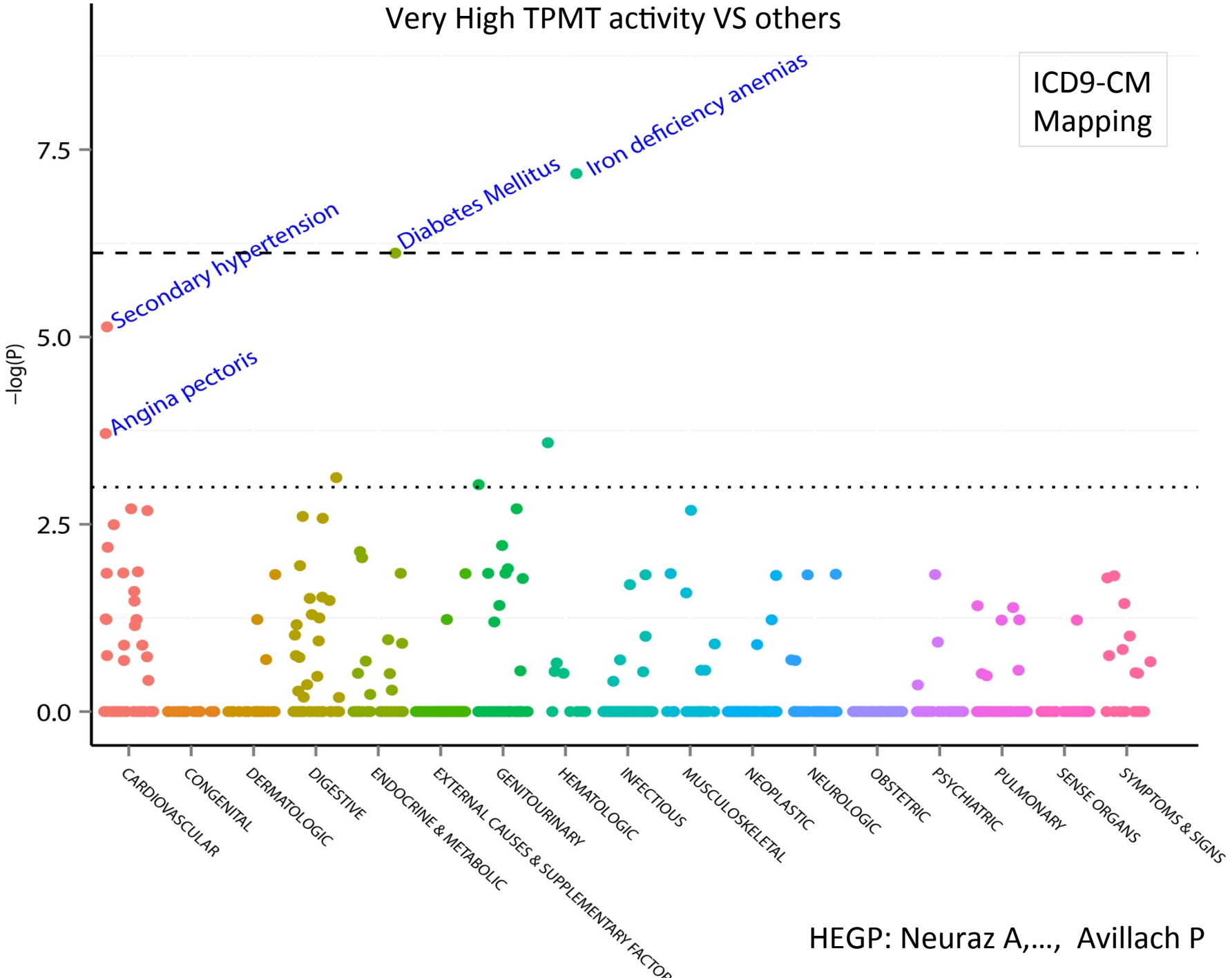
FDA & EMA recommendations

Phenotype	Low activity	Intermediate activity	Normal Activity	?
Thiopurine dose	10 % dose	30 – 70 % dose	100 % dose	

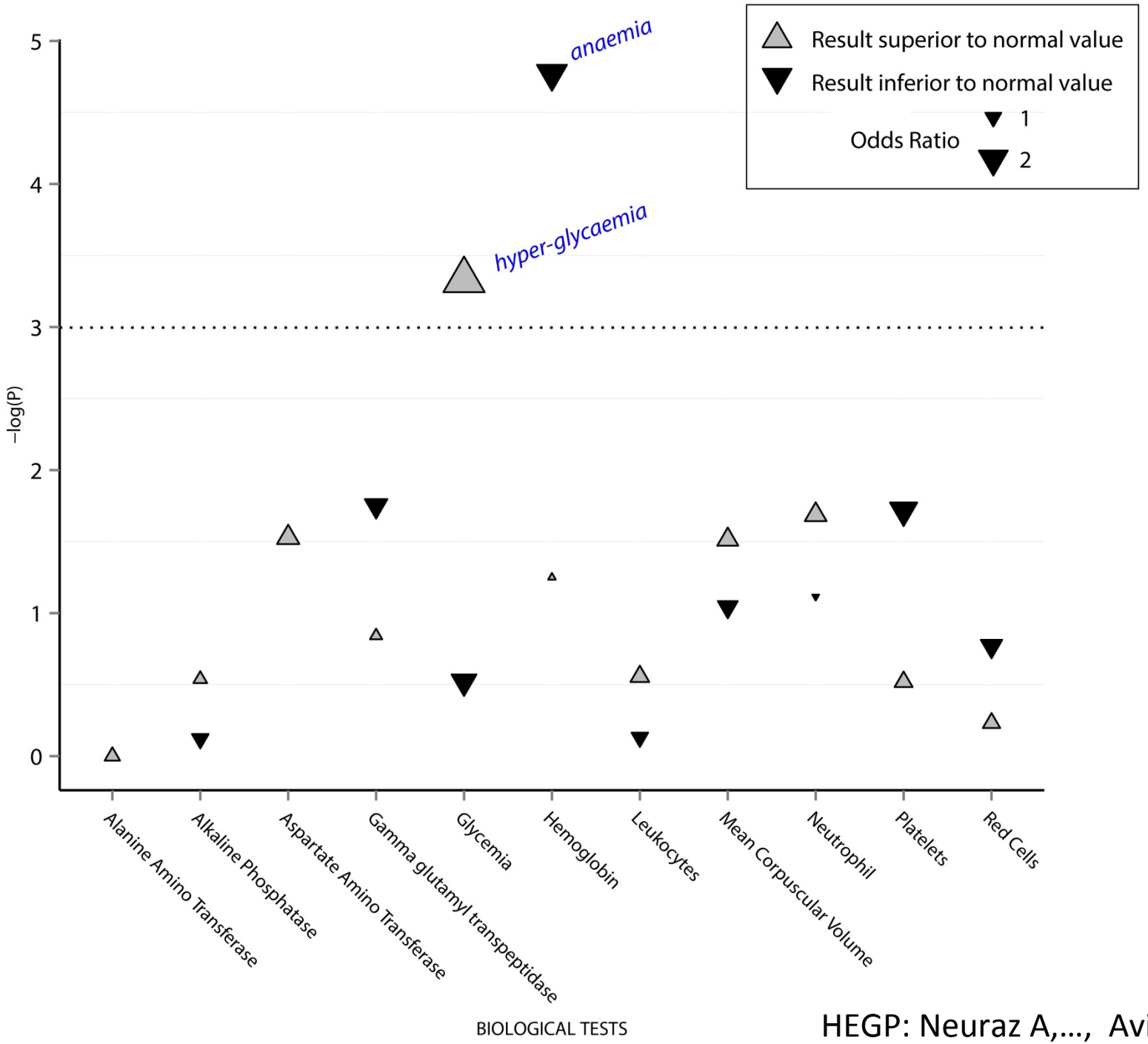
TPMT



PEGP: Neuraz A,…, Avillach P



HEGP: Neuraz A,…, Avillach P



OOL

HEGP: Neuraz A,…, Avillach P

Acknowledgments HEGP

Prof Patrice Degoulet, past CIO
Prof Anita Burgun, CIO

HÔPITAUX UNIVERSITAIRES PARIS OUEST
Hôpital européen Georges-Pompidou

Informatics & Public Health Dept :

- Eric Zapletal,
- Vincent Canuel,
- Antoine Neuraz,
- Fabien Joubert

i2b2 www.i2b2.org

tranSMART www.transmartproject.org

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

European Medical Information Framework

EMIF project: European Medical Information Framework

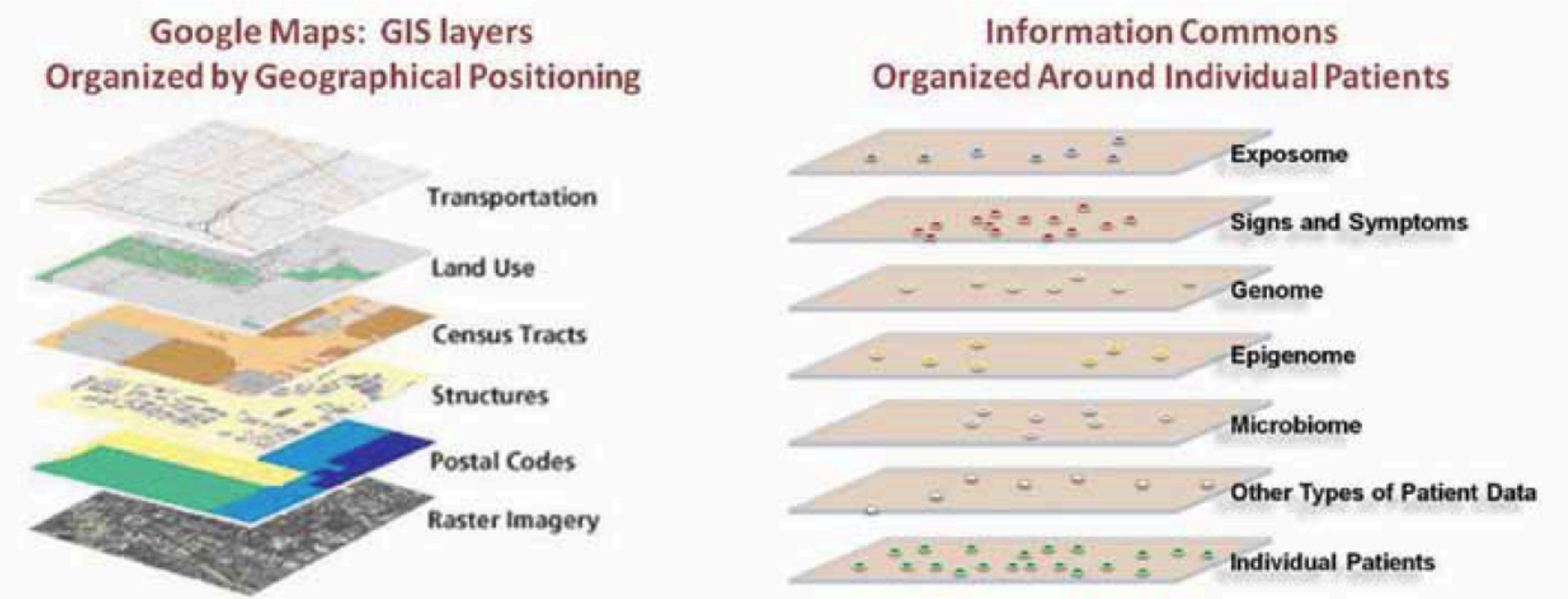
- 58 partners
- 58 Million euros
- Started 1st jan 2013
- 5 year project
- PI: Bart Vannieuwenhuyse (Janssen) & Prof. Simon Lovestone (KCL)
- PI Platform: Prof. Johan van der Lei
- Detect new biomakers:
  - prediposition Alzheimer's disease
  - Metabolic complication diabetes

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

CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL

Boston Children's Hospital Research Connection

EHR

i2b2

- Clinical notes
- ICD9
- Lab values
- Drugs

cTAKES

Selected patients

Research cohorts

Research cohorts

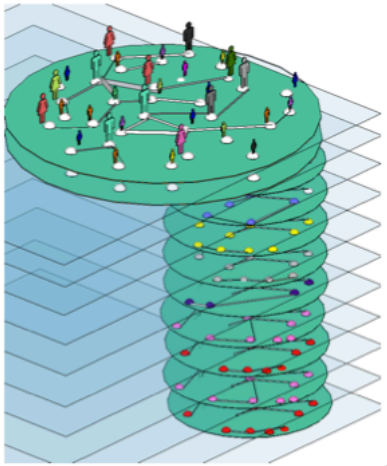
Research cohorts

Biobank Core

Patients' consent(s)

tranSMART

Patient centered Information Commons



'omics
Gene exp
SNPs
Whole Exome Seq
RNA seq

Generation of Hypotheses by PIs

"Drag & Drop"

High throughput analysis by Biostatistician & Bioinformatician

BCH Private network

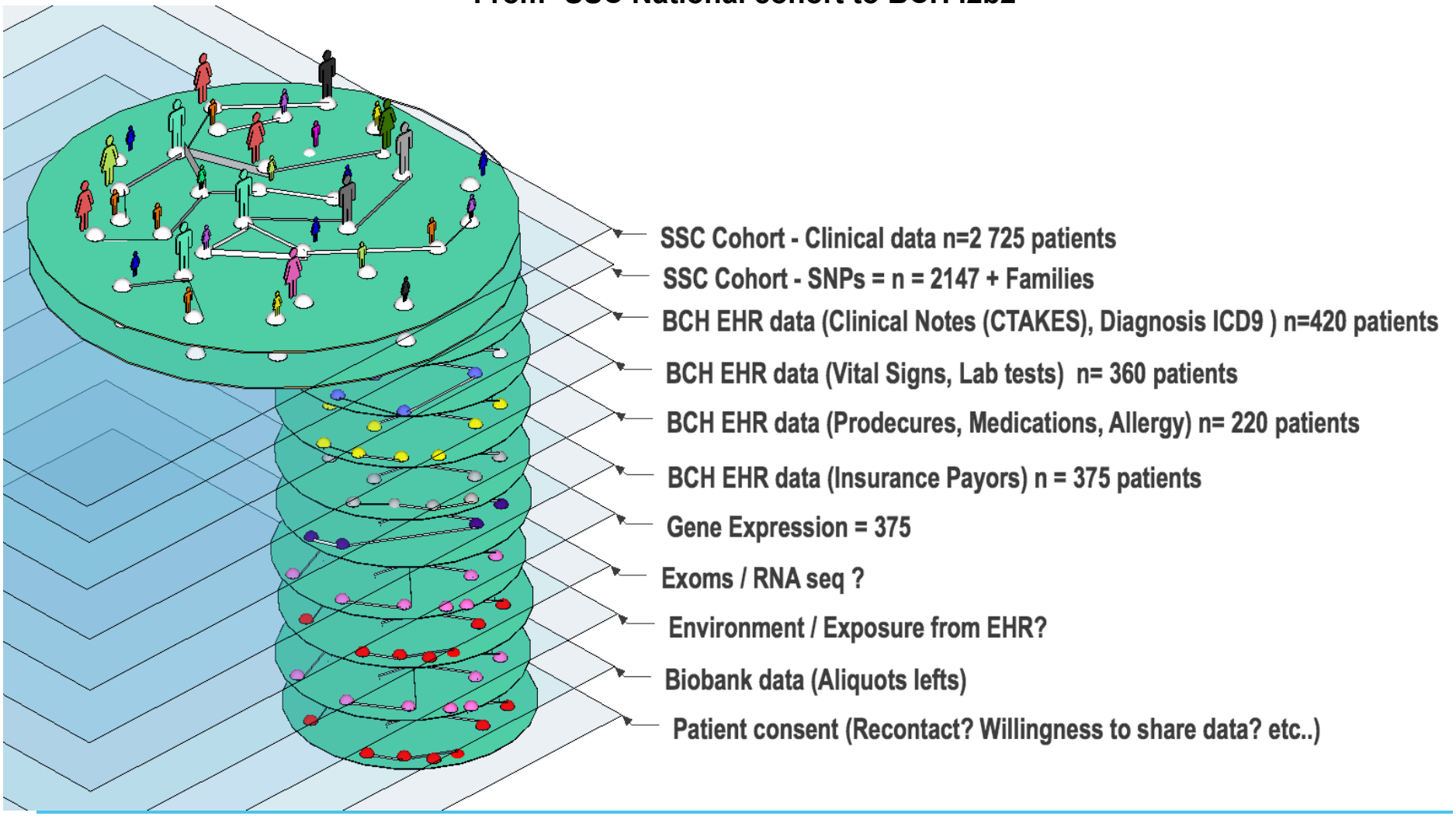
integration
analysis

Autism cohorts

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

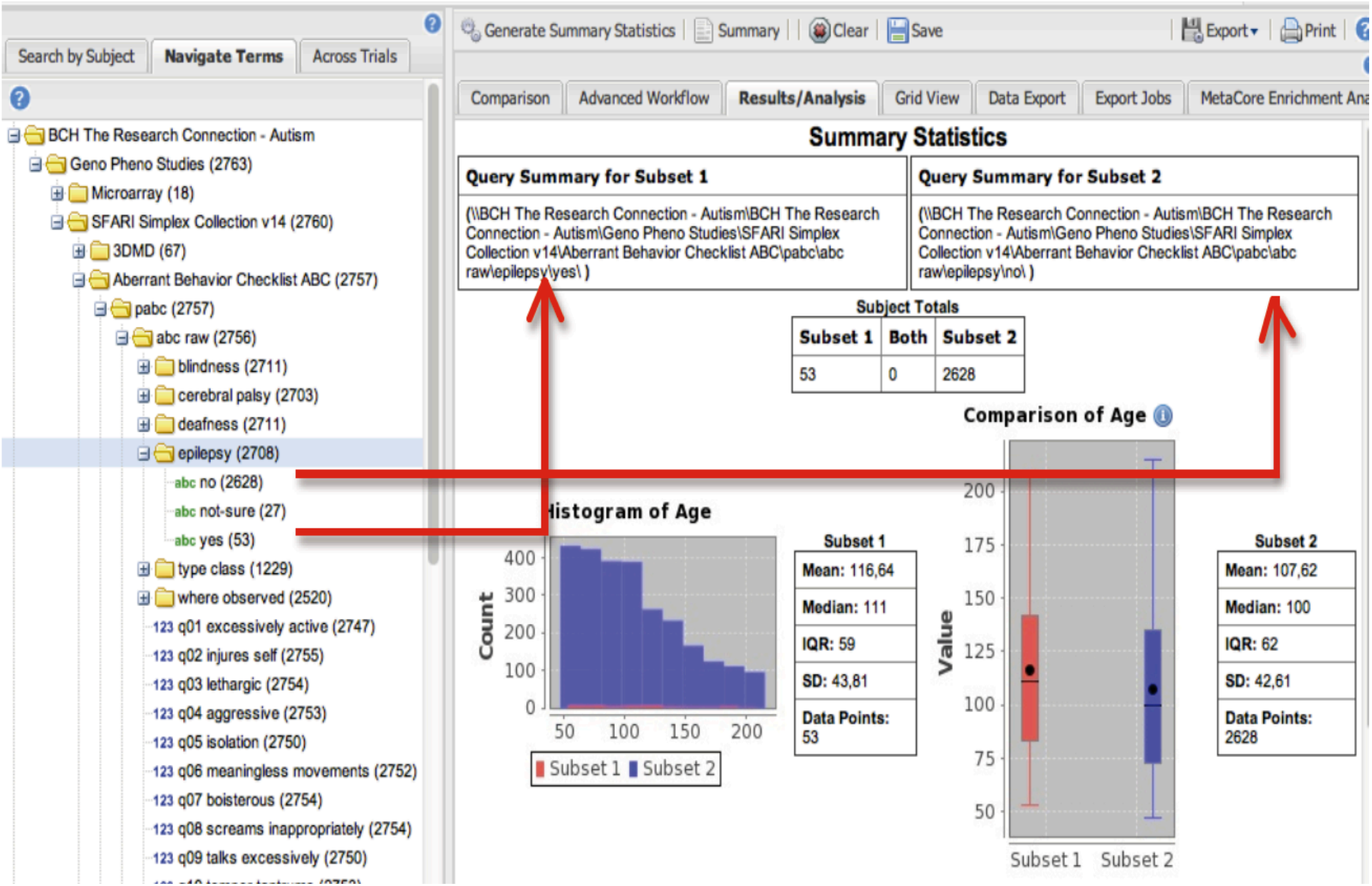
Autism information commons

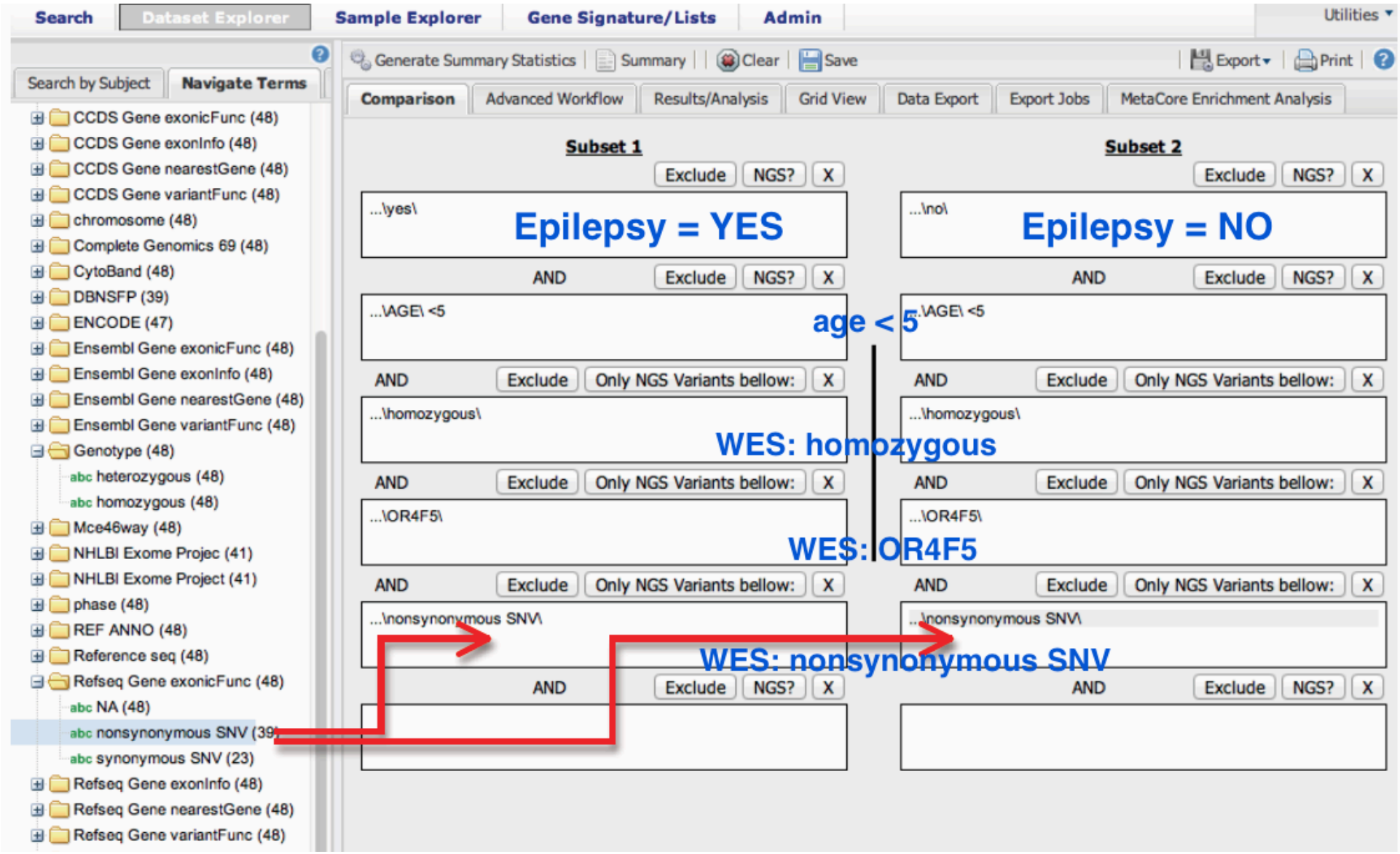
From SSC National cohort to BCH i2b2



ASDs SSC Cohort to i2b2 BCH

	# patients	# observations
Vital:Weight	367	4 727
Vital:Height	359	7 220
SER	421	17 915
Radiology Procedures	174	1 142
Procedures (CPT)	209	1 470
Medications (Prescription)	212	3 171
Pathology Repports	220	2 255
Clinical Notes	419	33 978
Medications	189	19 321
Lab tests	360	72 433
Insurance Payors	375	8 869
Medications (History)	217	835
Diagnosis (ICD9)	421	26 323
Clinic	421	14 169
AMED	164	7 142
Allergy	246	583
	AVG = 298	**TOTAL = 221 553**


Search by Subject
Navigate Terms
Across Trials
BCH The Research Connection - Autism
Geno Pheno Studies (2763)
Microarray (18)
SFARI Simplex Collection v14 (2760)
3DMD (67)
Aberrant Behavior Checklist ABC (2757)
pabc (2757)
abc raw (2756)
blindness (2711)
cerebral palsy (2703)
deafness (2711)
epilepsy (2708)
abc no (2628)
abc not-sure (27)
abc yes (53)
type class (1229)
where observed (2520)
123 q01 excessively active (2747)
123 q02 injures self (2755)
123 q03 lethargic (2754)
123 q04 aggressive (2753)
123 q05 isolation (2750)
123 q06 meaningless movements (2752)
123 q07 boisterous (2754)
123 q08 screams inappropriately (2754)
123 q09 talks excessively (2750)
Generate Summary Statistics
Summary
Clear
Save
Export
Print
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
MetaCore Enrichment Ana
Summary Statistics
Query Summary for Subset 1
(\\BCH The Research Connection - Autism\BCH The Research Connection - Autism\Geno Pheno Studies\SFARI Simplex Collection v14\Aberrant Behavior Checklist ABC\pabc\abc raw\epilepsy\yes\ )
Query Summary for Subset 2
(\\BCH The Research Connection - Autism\BCH The Research Connection - Autism\Geno Pheno Studies\SFARI Simplex Collection v14\Aberrant Behavior Checklist ABC\pabc\abc raw\epilepsy\no\ )
Subject Totals
Subset 1 | Both | Subset 2
53 | 0 | 2628
Comparison of Age
Histogram of Age
Count
Subset 1
Mean: 116,64
Median: 111
IQR: 59
SD: 43,81
Data Points: 53
Value
Subset 2
Mean: 107,62
Median: 100
IQR: 62
SD: 42,61
Data Points: 2628
Subset 1
Subset 2


Search
Dataset Explorer
Sample Explorer
Gene Signature/Lists
Admin
Utilities
Search by Subject
Navigate Terms
CCDS Gene exonicFunc (48)
CCDS Gene exonInfo (48)
CCDS Gene nearestGene (48)
CCDS Gene variantFunc (48)
chromosome (48)
Complete Genomics 69 (48)
CytoBand (48)
DBNSFP (39)
ENCODE (47)
Ensembl Gene exonicFunc (48)
Ensembl Gene exonInfo (48)
Ensembl Gene nearestGene (48)
Ensembl Gene variantFunc (48)
Genotype (48)
abc heterozygous (48)
abc homozygous (48)
Mce46way (48)
NHLBI Exome Projec (41)
NHLBI Exome Project (41)
phase (48)
REF ANNO (48)
Reference seq (48)
Refseq Gene exonicFunc (48)
abc NA (48)
abc nonsynonymous SNV (39)
abc synonymous SNV (23)
Refseq Gene exonInfo (48)
Refseq Gene nearestGene (48)
Refseq Gene variantFunc (48)
Generate Summary Statistics
Summary
Clear
Save
Export
Print
Comparison
Advanced Workflow
Results/Analysis
Grid View
Data Export
Export Jobs
MetaCore Enrichment Analysis
Subset 1
Subset 2
Exclude
NGS?
X
...\yes\
Epilepsy = YES
...\no\
Epilepsy = NO
AND
...\AGE\ <5
age < 5
Only NGS Variants bellow:
...\homozygous\
WES: homozygous
...\OR4F5\
WES: OR4F5
...\nonsynonymous SNV\
WES: nonsynonymous SNV

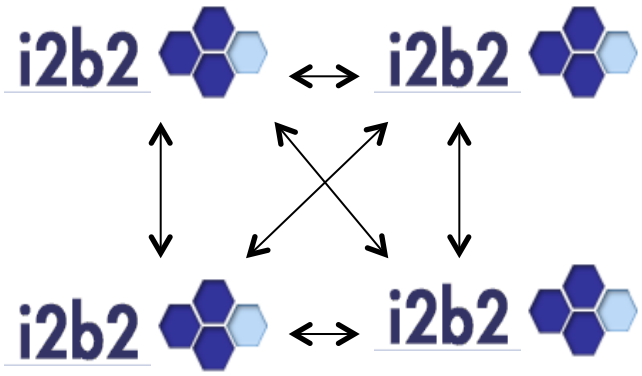
OPEN ACCESS Freely available online

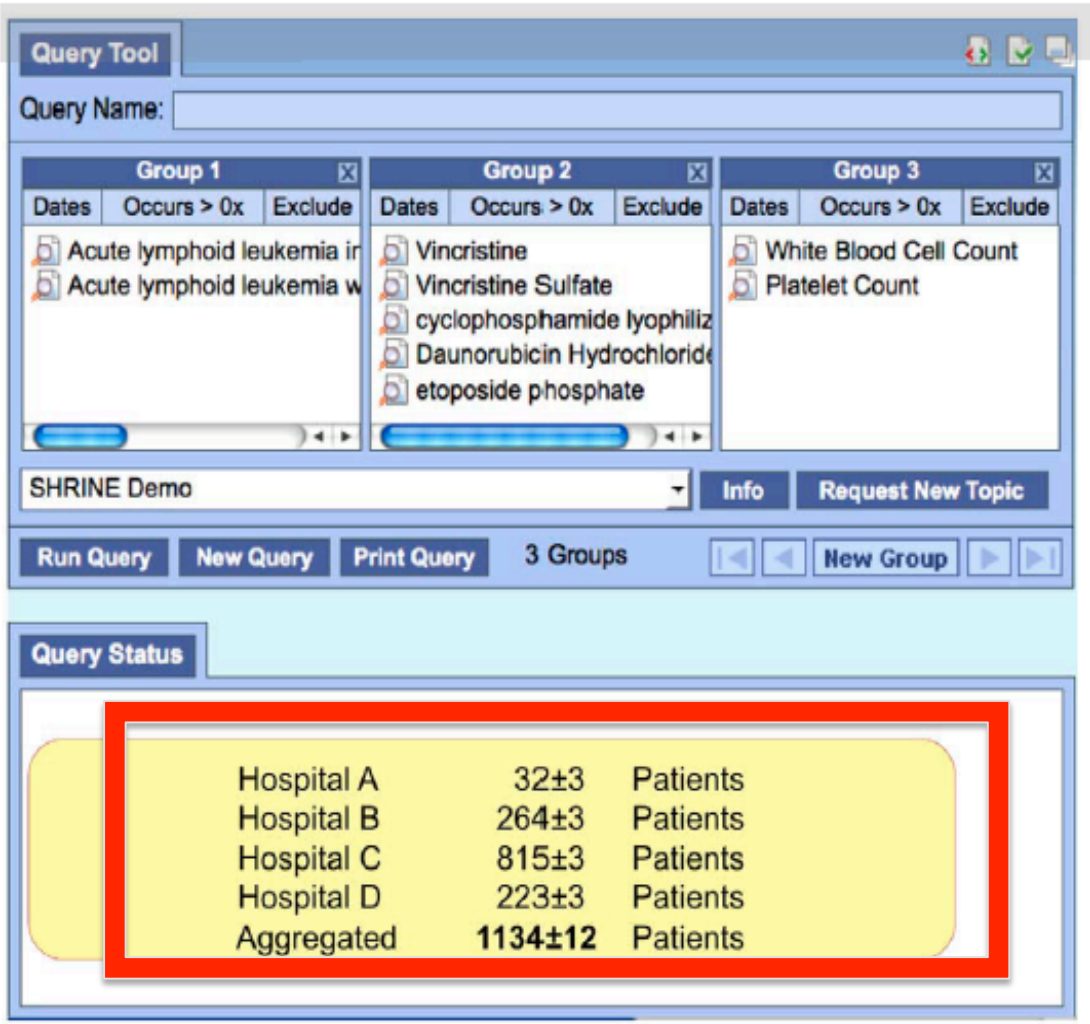
PLOS ONE

SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies

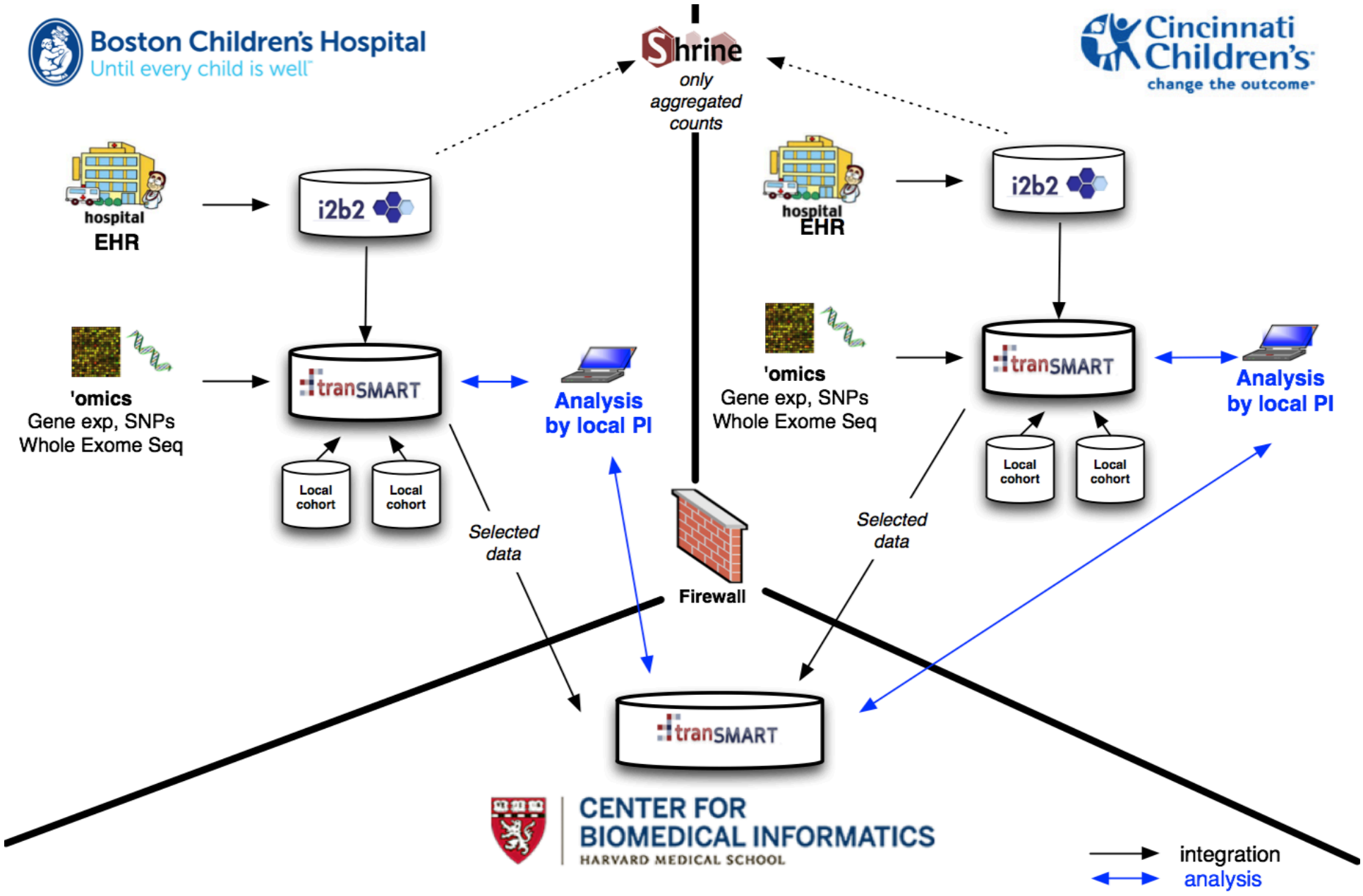
Andrew J. McMurry[1,2,3,4]*, Shawn N. Murphy[3,5,6], Douglas MacFadden[1], Griffin Weber[3,7], William W. Simons[1], John Orechia[8], Jonathan Bickel[2,9], Nich Wattanasin[5], Clint Gilbert[1], Philip Trevvett[1], Susanne Churchill[3,5], Isaac S. Kohane[1,2,3]

Shrine





Boston Children's Hospital
Until every child is well™


Boston Children's Hospital
Until every child is well
Shrine
only aggregated counts
Cincinnati Children's
change the outcome
hospital
EHR
i2b2
'omics
Gene exp, SNPs
Whole Exome Seq
tranSMART
Local cohort
Local cohort
Analysis by local PI
Selected data
Firewall
hospital
EHR
i2b2
'omics
Gene exp, SNPs
Whole Exome Seq
tranSMART
Local cohort
Local cohort
Analysis by local PI
Selected data
tranSMART
CENTER FOR BIOMEDICAL INFORMATICS
HARVARD MEDICAL SCHOOL
integration
analysis

Acknowledgments

Informatics & Public Health Dept :

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

www.i2b2.org

www.transmartproject.org

www.recomdata.com

Contact :
paul.avillach@egp.aphp.fr

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
Paul Avillach, MD, PhD

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

MassGeneral Hospital for Children
Lurie Center for Autism

Timothy Yu, MD, PhD

Laboratory of cognitive neuroscience

Charles Nelson, PhD
Vanessa Vogel-Farley
Nicole Coman

The Research Connection

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