

**ΕΛΛΗΝΙΚΗ ΦΑΡΜΑΚΕΥΤΙΚΗ ΕΤΑΙΡΕΙΑ**

---

**ΗΛΕΚΤΡΟΝΙΚΑ «ΕΡΓΑΛΕΙΑ»  
ΕΓΚΡΙΤΗΣ ΑΝΑΛΥΣΗΣ ΤΗΣ  
ΒΙΒΛΙΟΓΡΑΦΙΑΣ**

**Καθηγητής Παναγιώτης Βλάμος**

**Πρόεδρος Τμήματος Πληροφορικής  
Διευθυντής Εργαστηρίου Βιοπληροφορικής  
& Ανθρώπινης Ηλεκτροφυσιολογίας  
Ιονίου Πανεπιστημίου**



## **Contemporary scene**

---

- **fundamental shift of biomedical research domain towards integrative and translational methodologies**
- **needs related to the collection, management, integration, analysis and dissemination of large-scale, heterogeneous biomedical data sets.**
- **well-established and broadly adopted theoretical and practical frameworks intended to address such needs are still largely developmental**

# Challenges

---

- the management of multi-dimensional and heterogeneous data sets
- the size and complexity of data sets that must be managed and analyzed are growing at an extremely rapid rate
- the data management practices currently used in most research settings are both labor intensive and rely upon technologies that have not been designed to handle such multi-dimensional data
- the translational science community demands the creation and delivery of information management platforms capable of adapting to and supporting heterogeneous workflows and data sources



## Future needs

---

**The rise of precision medicine and the availability of high-throughput molecular analyses in the context of clinical care have increased the need for adequate tools for translational researchers to manage and explore these data**

**T0**

**T1**

**T2**

**T3**

**T4**

**BASIC SCIENCE  
RESEARCH**

Preclinical  
and animal  
studies

Defining  
mechanisms,  
targets, and lead  
molecules

**TRANSLATION TO  
HUMANS**

Proof of concept  
Phase 1  
clinical trials

New methods  
of diagnosis,  
treatment, and  
prevention

**TRANSLATION TO  
PATIENTS**

Phase 2  
and 3  
clinical trials

Controlled studies  
leading to  
effective care

**TRANSLATION TO  
PRACTICE**

Phase 4  
clinical trials  
and clinical  
outcomes  
research

Delivery of  
recommended  
and timely care to  
the right patient

**TRANSLATION TO  
COMMUNITY**

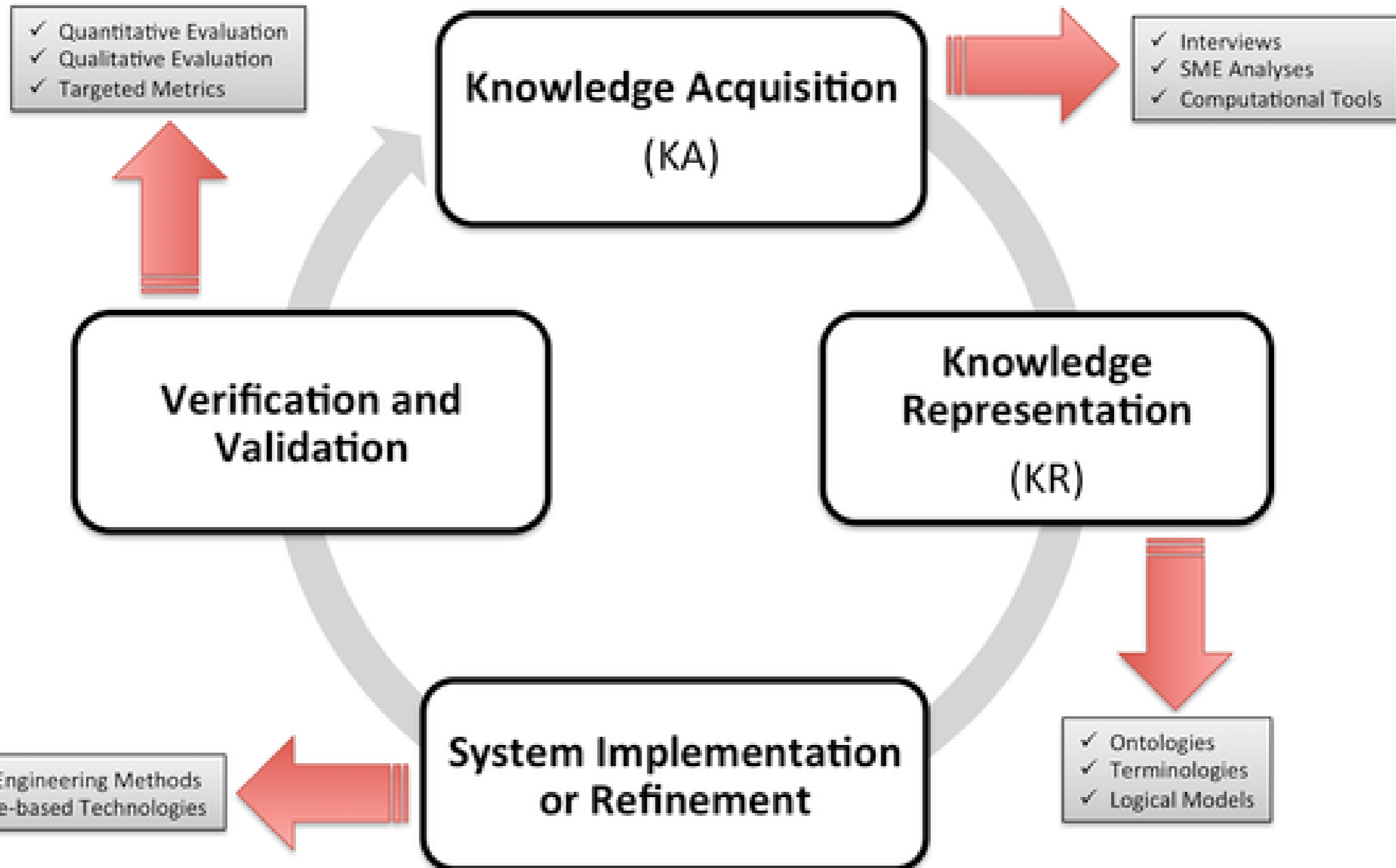
Population-  
level  
outcomes  
research

True benefit  
to society

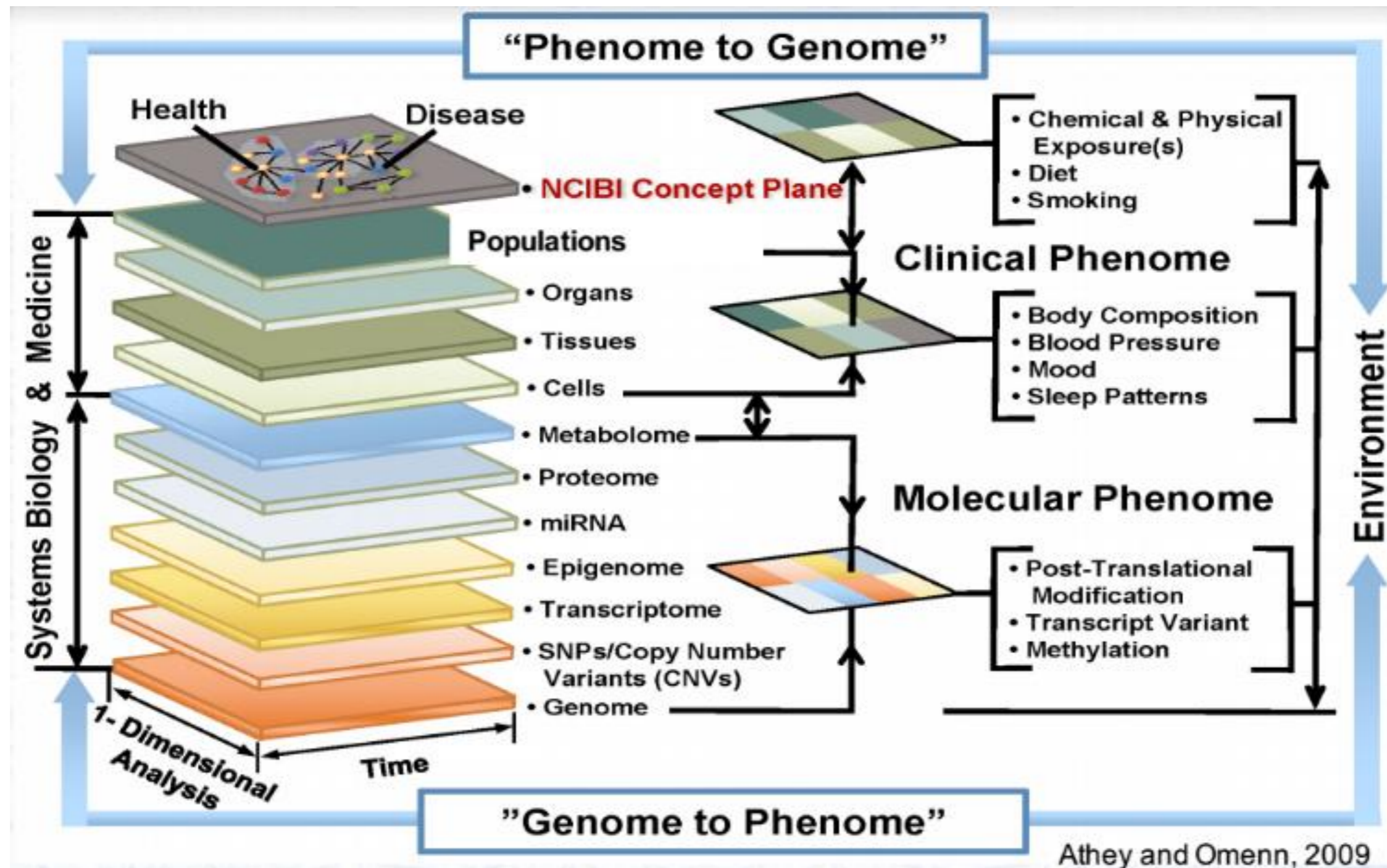
Translation from basic science to human studies

Translation of new data into the clinic  
and health decision making

# Key Definitions



# Translational Research Data is Big Data



Volume, Variety, Velocity



# **Immediate demands**

**informatics methods  
that connect molecular entities to clinical entities**

translational research platforms

**able to integrate  
large data sets of clinical information  
with omics data**

# **Translational research platforms: Overview**

**should provide**

**(i) the storage and integration of clinical and omics data;**

**(ii) an analysis framework, enabling scientists to explore their data and generate hypotheses; and**

**(iii) additional information cross-referenced from external databases**



# **Translational research platforms: Overview**

- ◆ **BRISK: Biology-Related Information Storage Kit (2011)**
- ◆ **iDASH: Integrating data for analysis, anonymization and sharing (2011)**
- ◆ **tranSMART (2010)**

## **Cancer**

- ◆ **G-DOC Georgetown Database of Cancer (2012)**
- ◆ **cBio Cancer Genomics Portal (2012)**
- ◆ **iCOD: Integrated Clinical Omics Database (2010)**
- ◆ **caTRIP (2006)**



# **Translational research platforms: Overview**

- ◆ **BRISK: Biology-Related Information Storage Kit (2011)**
- ◆ **iDASH: Integrating data for analysis, anonymization and sharing (2011)**
- ◆ **tranSMART (2010)**

## **Cancer**

- ◆ **G-DOC Georgetown Database of Cancer (2012)**
- ◆ **cBio Cancer Genomics Portal (2012)**
- ◆ **iCOD: Integrated Clinical Omics Database (2010)**
- ◆ **caTRIP (2006)**



Platform	tranSMART	cBioPortal	BRISK	iDASH	iCOD	G-DOC	caTRIP
Community							
Institution/project initiators	Johnson & Johnson, USA	Memorial Sloan-Kettering Cancer Center, NY, USA	University of British Columbia, Canada	iDASH team, USA	Tokyo Medical Dental University, Japan	Georgetown University, DC, USA	Institution/project initiators
Funding	Initially Johnson & Johnson funded—now public-private consortia	NIH and industry grants	Public and private consortium	NIH	Public	HSS	NIH
Software availability	Open source	Free for academic use, commercial licenses	Open source	Open source	Not distributed	Not distributed	Not distributed

Platform	tranSMART	cBioPortal	BRISK	iDASH	iCOD	G-DOC	caTRIP
Clinical data							
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcomes	Yes	Yes	No	Yes	Yes	Yes	Yes
Biological results	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Images	No	No	No	Yes	Yes	No	Unknown
Structured clinical research data	Yes	Yes	Yes	Yes	Yes	Yes	Yes
'omics' data							
mRNA expression	Yes	Yes	No	Yes	Yes	Yes	Unknown
miRNA expression	No	Yes	No	Yes	No	Yes	Unknown
SNPs	Yes	Yes	Yes	Yes	No	Yes	Unknown
Copy number variations	Yes	Yes	Yes	Unknown	No	Yes	Unknown
DNA methylation	No	Yes	Yes	Unknown	No	Yes	Unknown

Platform	tranSMA RT	cBioPortal	BRISK	iDASH	iCOD	G-DOC	caTRIP
Clinical data							
Demographics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Outcomes	Yes	Yes	No	Yes	Yes	Yes	Yes
Biological results	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Images	No	No	No	Yes	Yes	No	Unknown
Structured clinical research data	Yes	Yes	Yes	Yes	Yes	Yes	Yes
'omics' data							
mRNA expression	Yes	Yes	No	Yes	Yes	Yes	Unknown
miRNA expression	No	Yes	No	Yes	No	Yes	Unknown
SNPs	Yes	Yes	Yes	Yes	No	Yes	Unknown
Copy number variations	Yes	Yes	Yes	Unknown	No	Yes	Unknown

# Desiderata for translational research platforms

---

**Privacy**

**Interoperability and standards**

**Heterogeneity of granularity of the data models**

**Deployment and maintenance**

## **Closing the loop.....**

---

**The ultimate goal of translational medicine is enabling personalized care. Nowadays efforts are made to populate translational research platforms with patient data to fuel discovery. Allowing real-time data-driven decision algorithms to leverage translation research results in the context of clinical care should be a short-term objective.**



# **caTRIP – 2006**

## **Cancer Translational Research Informatics Platform**

---

### **Description**

Translational research system for the caBIG project (cancer Biomedical Informatics Grid™), initiated by the U.S. National Cancer Institute.

### **Goal**

data-integration project with a goal to develop an open-source information network across the United States for secure data exchange on cancer research.

### **Aims**

to create a platform that allows physicians to find patients with similar characteristics, analyze their clinical outcomes and find information about treatments that were administered with success across the caBIG data network.



# caTRIP – 2006

## Cancer Translational Research Informatics Platform

---

**Oncologists could access** a rich **live** data network that could provide strong statistically significant facts in mere minutes.

This could benefit clinicians to query data from a cohort of preexisting patients to help guide treatment of another patient, **rather than relying on single cancer centers** or limited facts published in the literature.

caBIG project was **cancelled** in 2011, few cancer centers were using caBIG's clinical data management tools or its cloud computing infrastructure.

Also caBIG applications **cost** millions of dollars to build.

The majority of the life sciences research tools developed for caBIG have had **limited** use or impact, according to the report.



# tranSMART (2010) - Foundation members

## Gold Members (7)

ORACLE®



ConvergeHEALTH  
by **Deloitte**.



## Silver Members (15)

PHILIPS



assureX  
health



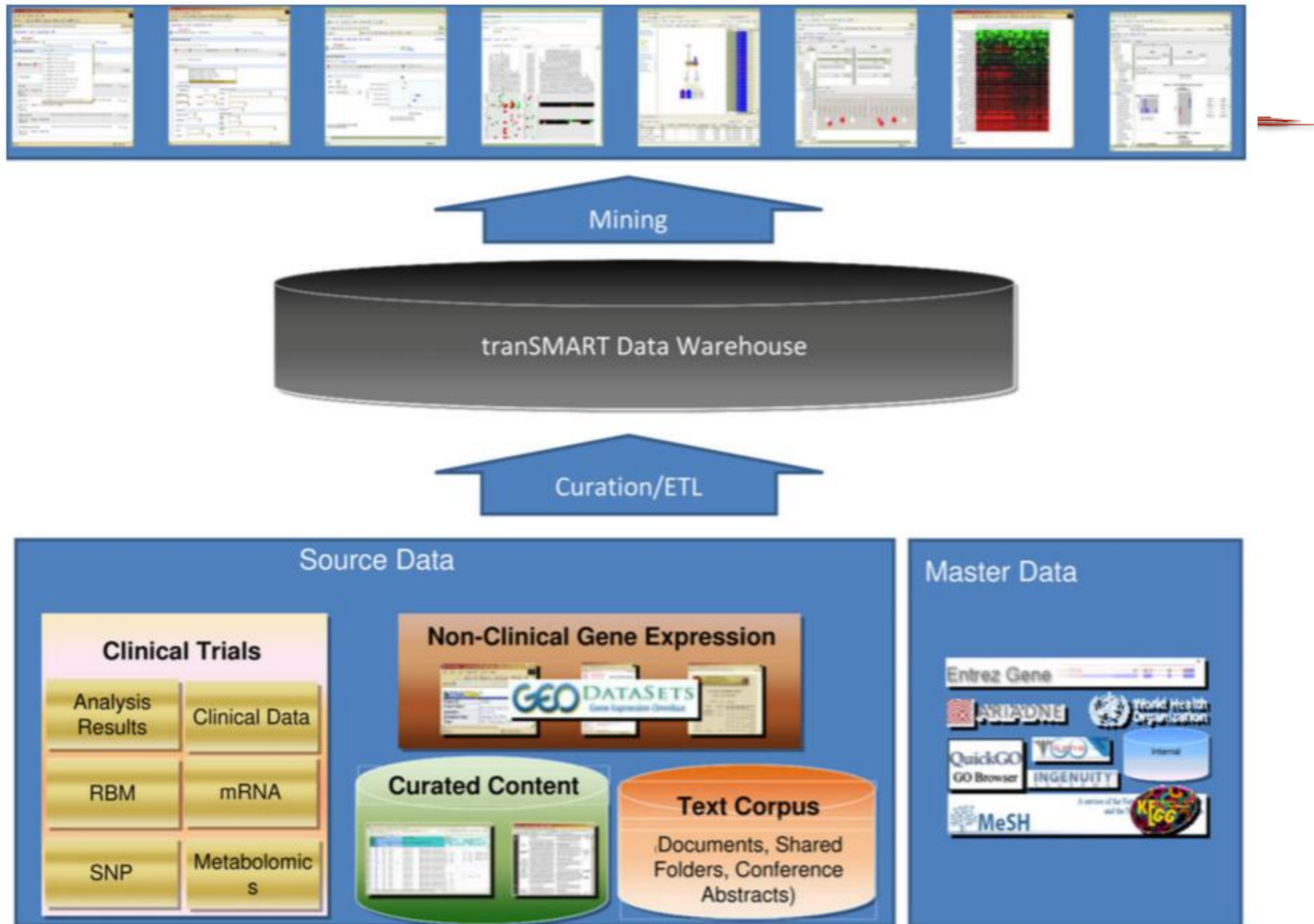
abbvie



Imperial College  
London  
Data Science Institute

## Affiliate Members (3)





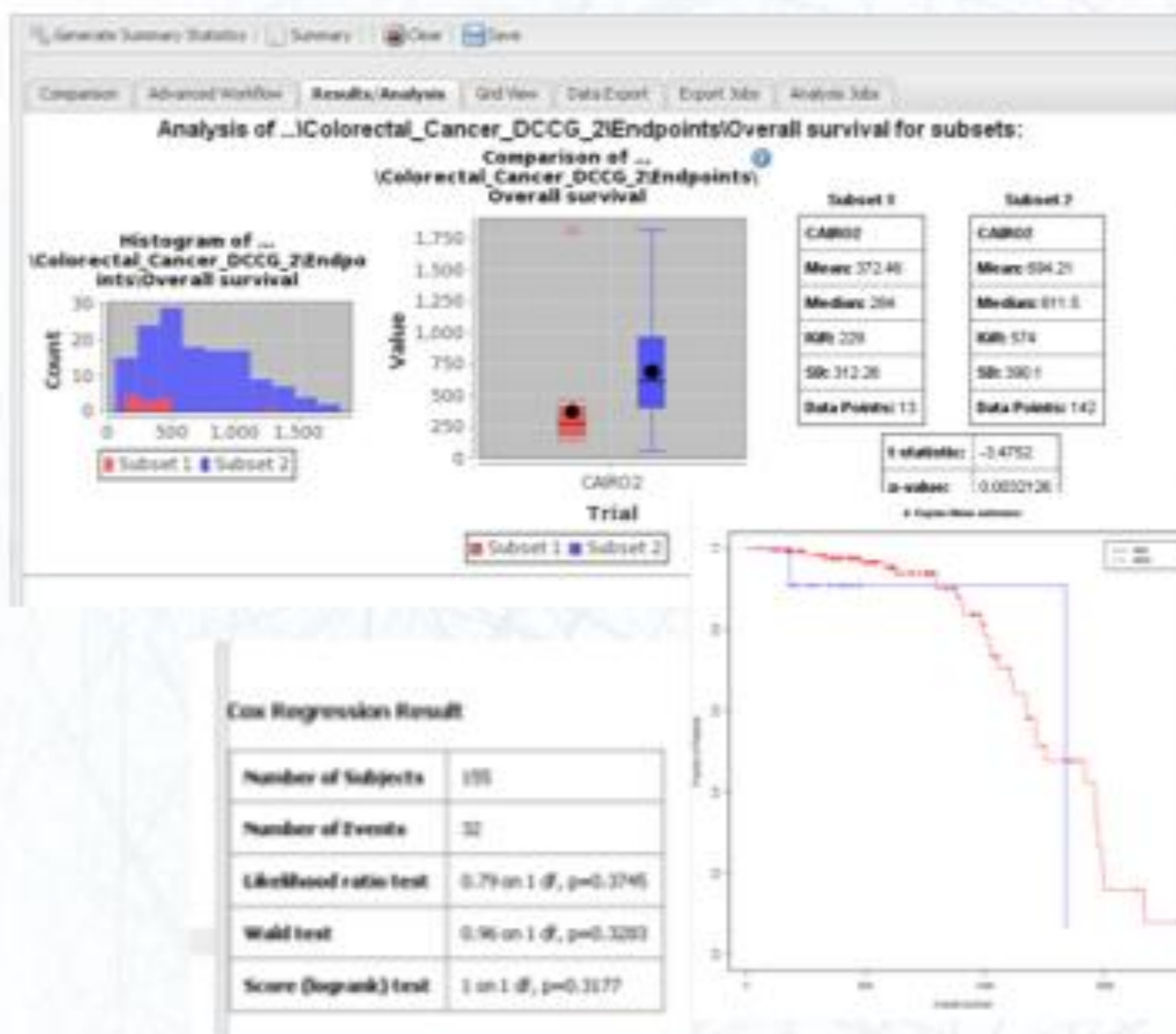
# Cohort comparison

---

- Analysis goals
  - Compare survival and chromosomal abnormalities across disease subtypes
- Available data
  - Clinical data: E.g. staging, age, gender, treatment arm
  - Non-omics data: MSI/MSS (microsatellite instability), mutation data
  - Genomics: Comparative genomic hybridization (arrayCGH)
- Methods
  - Histograms, boxplots, t-test, Kaplan-Meier plot, Cox regression, arrayCGH group test

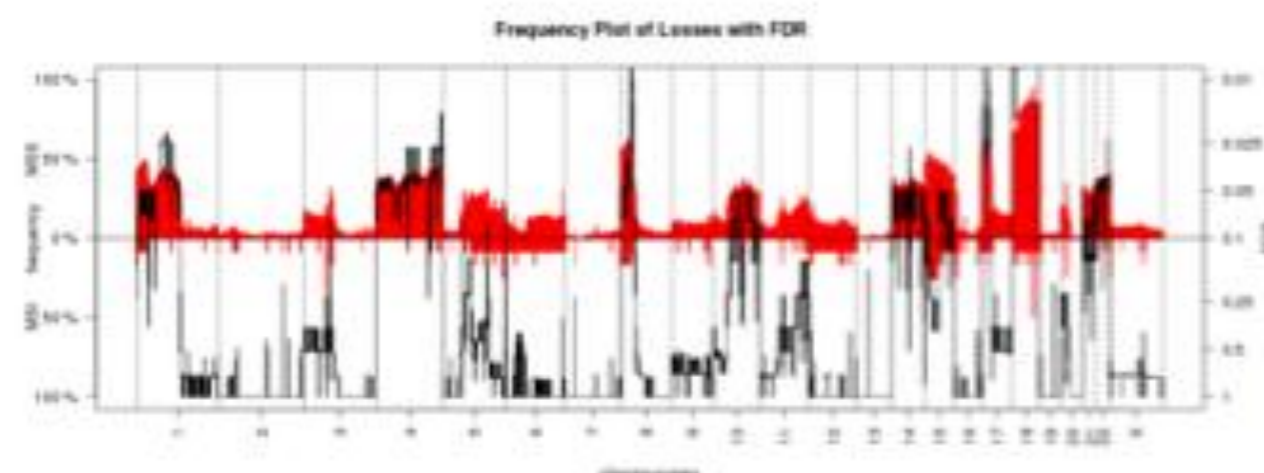


# Results: Comparison of cohorts



of 21 genes/chr: 162/923

chr	start	end	prob	id
1p36.32	3618012	3718012	0.0286	0.07400380206
1p36.32	378110	408110	0.0286	0.0707607427
1p36.32	410310	421807	0.0286	0.0707607427
1p36.32	423440	430000	0.0286	0.0707607427
1p36.32	441701	454157	0.0286	0.1148072044
1p36.32	458202	510004	0.0286	0.1148072044
1p36.32/1p36.31	5114802	5211801	0.0284	0.1182020447
1p36.31	5225219	541428	0.0286	0.1182020447



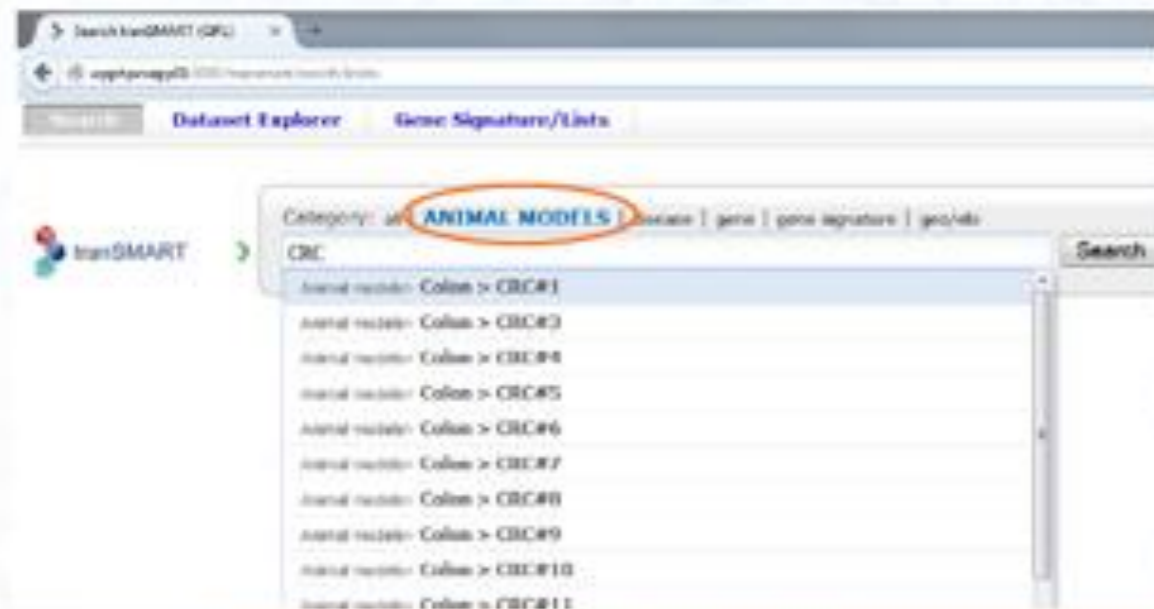
Comparison of chromosomal alterations between cohorts

Comparison of overall survival in subjects with MSI vs MSS

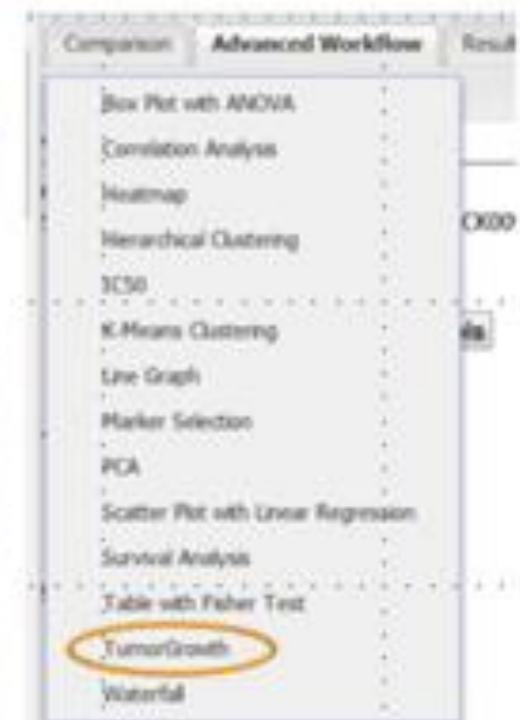
# Collaboration, Analysis, Customization

- **Goals**
  - Create central data storage for a preclinical oncology group at a large pharmaceutical company
  - Incorporate a custom R script that processes tumor volume and weight data
- **Available data**
  - In vivo studies using animal model
  - animal model metadata

# Results: Collaboration, Analysis, Customization



Custom R script incorporated



Search tranSMART (CPU)

Dataset Explorer Gene Signatures/Lists

Category: **ANIMAL MODELS** | Disease | Gene | Gene Signature | Geo/ids

Search

transSMART

Colon > CRC#1

Colon > CRC#2

Colon > CRC#3

Colon > CRC#4

Colon > CRC#5

Colon > CRC#6

Colon > CRC#7

Colon > CRC#8

Colon > CRC#9

Colon > CRC#10

Colon > CRC#11

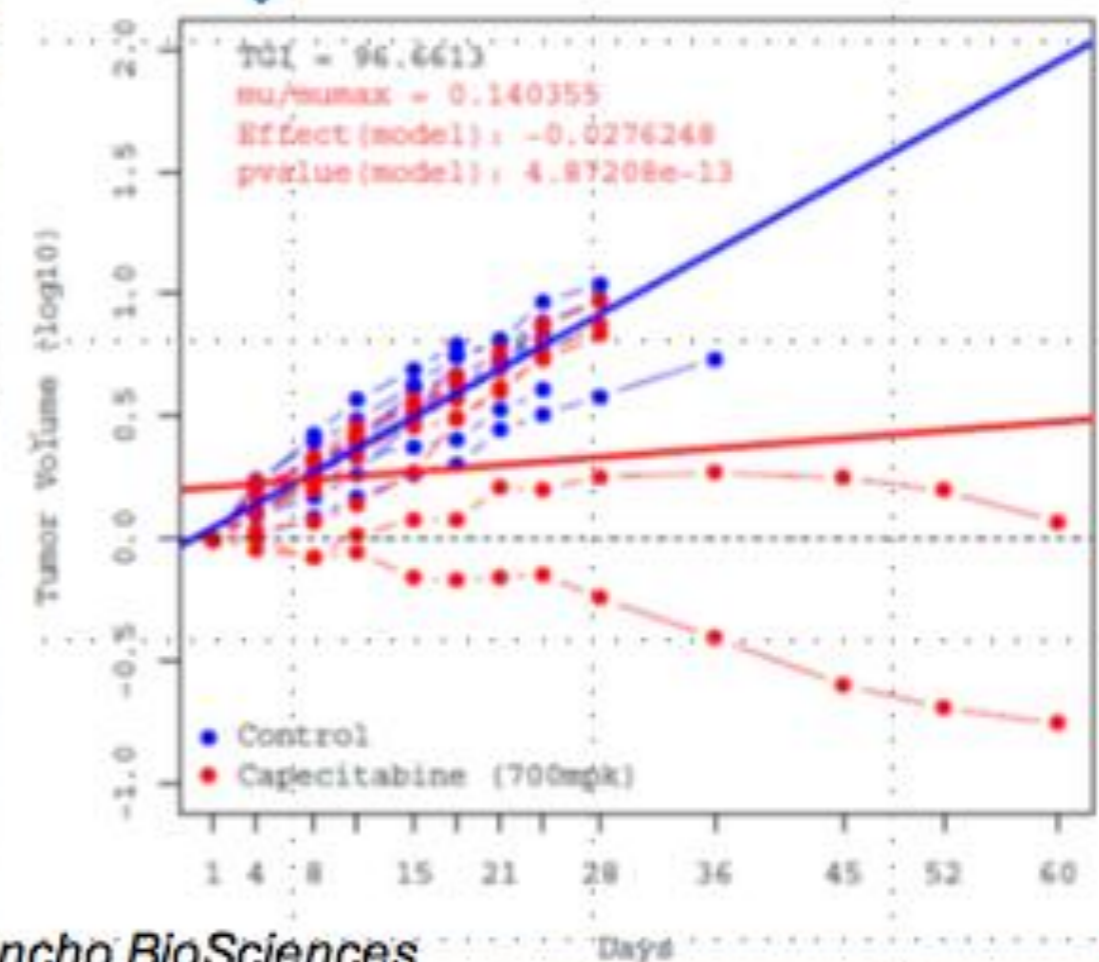
About 1 results found

Filters: advanced view clear all

Model info (1)

Cancer Type	Clinical and PDX model information	Patient tissue collection		Patient information		Diagnosis		Pathology		Treatment history		Additional clinical material		Materials and characterization passage for p	
		Date of collection	City of collection	Sex	Age	Staging	Primary or metastatic tissue collected	Clinical pathology data available	Verified by board certified pathologist (Rouse, China or U.S.)	Treatments administered (chemo, radio, surgery)	Response to treatment	PDXs	Matched adjacent normal	Long term frozen tissues	DBA
Colon cancer	CRC#1	2008-07-29	Shanghai	Male	79	T3	N0	M0	Primary	Yes	China	No	No	Yes	Yes

Animal model metadata stored in tranSMART



# Patient Stratification

---

- **Analysis goals**

- Association between demographic characteristics and Bortezomib response?
- Association between disease subtype and drug response?
- Any gene expression data predictive of drug response?

- **Available data**

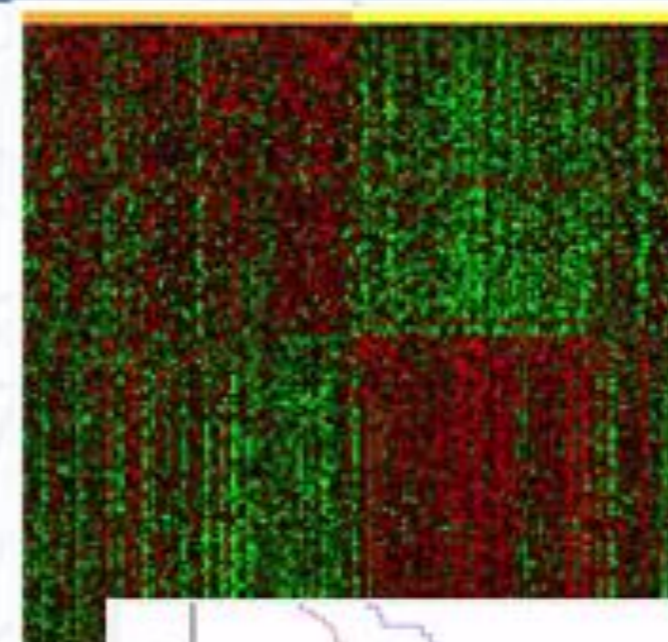
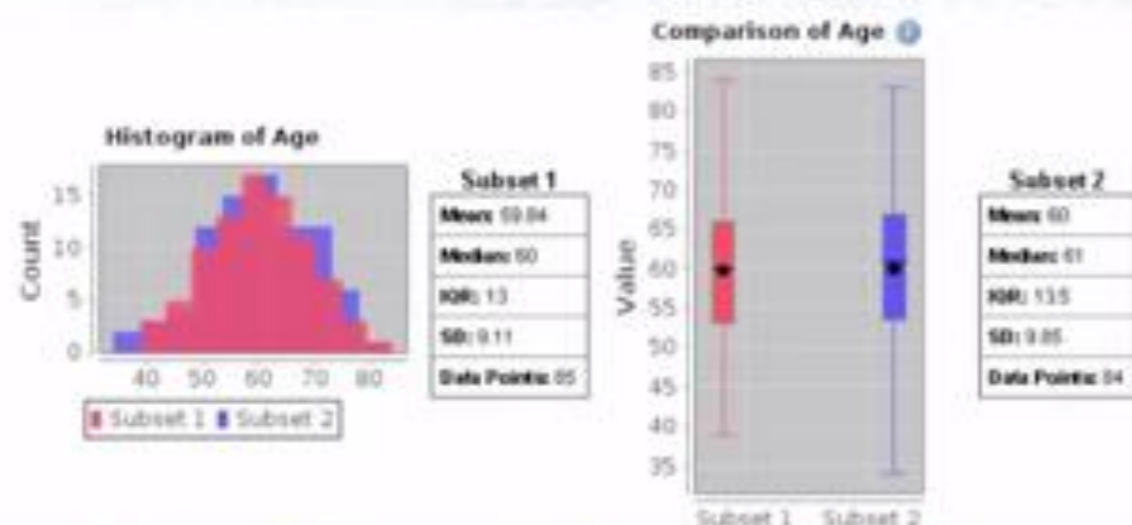
- Clinical study results from Bortezomib multiple myeloma study
  - Treated, untreated, responder, nonresponder patients
  - Demographic information
  - Disease subtype
  - Gene expression data

- **Methods**

- Chi-squared
- K-means clustering
- Kaplan-Meier analysis

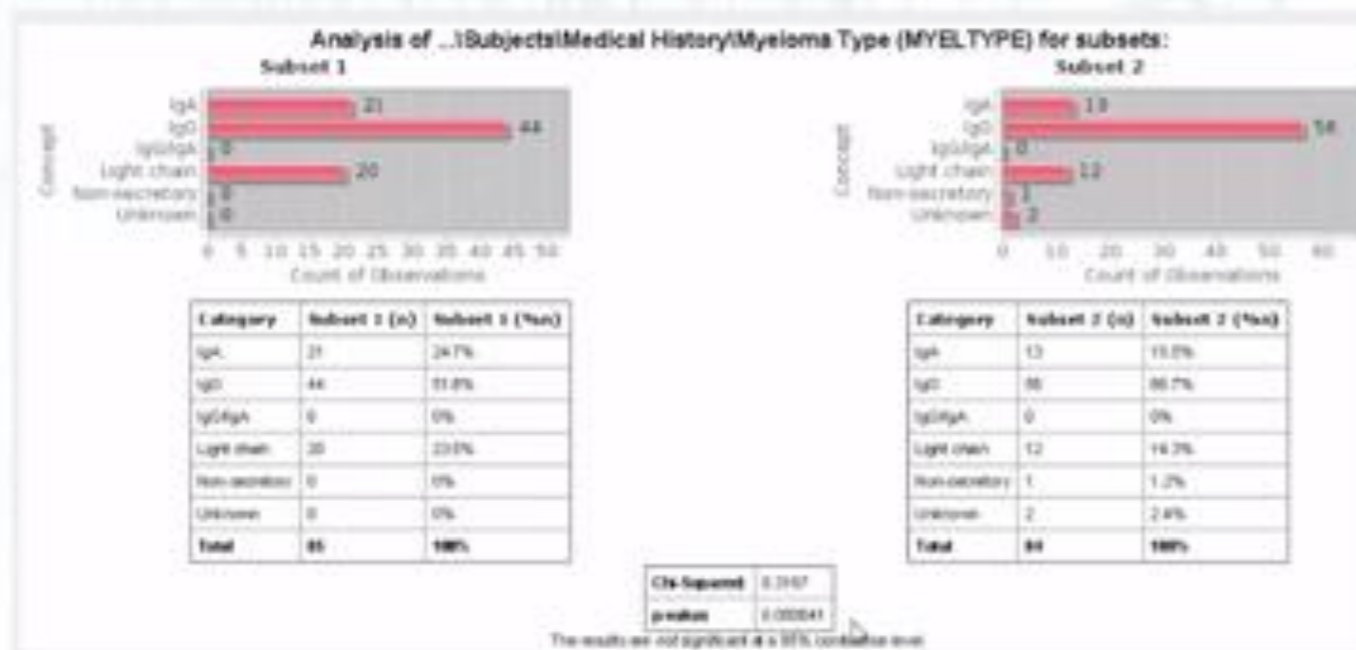


# Results: Patient Stratification



Differing gene expression pattern between responders and non-responders

Age distribution of responders and non-responders is similar



Expression of NDUFB1 correlates with OS in intermediate time frames, but not through end of study

No significant difference in response rate for different types of disease tested

# **tranSMART Foundation Datathon 1.0 (2015) :**

## **The cross neurodegenerative diseases challenge**

---

### **Specific objectives were to identify:**

- **Common biomarker changes across Parkinson and Alzheimer disease**
- **Common pathway changes across Parkinson and Alzheimer disease**
- **The normal distribution of imaging and fluid biomarkers across controls**
- **Novel hypotheses, research findings or conclusions about these neurodegenerative diseases.**



# What is a “Datathon”?

A datathon is an intense 3-day workshop that challenges researchers **to turn data and information into knowledge.**

- The datathon format is modeled after hackathons, which are focused on software development. Datathons use research questions and datasets to advance knowledge, not to develop applications.
- At a datathon, participants work in teams to frame research questions, create and implement a research design, mobilize data resources and present their findings in front of a panel of judges.



# Open Big Data: Preparing for the Datathon

---

**Alzheimer's Disease Neuroimaging Initiative (ADNI)**

**Parkinson's Progression Markers Initiative (PPMI)**

**The Cancer Genome Atlas (TCGA)**



# Open Big Data: ADNI

## The Alzheimer's Disease Neuroimaging Initiative

(ADNI) unites researchers with study data as they work to define the progression of Alzheimer's disease. ADNI researchers collect, validate and utilize data . . . Data. . . including Alzheimer's disease patients, mild cognitive impairment

subjects and elderly controls, *are available from this site.*



# Open Big Data: PPMI

---

Since 2002, the **Michael J. Fox Foundation** has been an

essential driver of PD biomarker development efforts, funding nearly \$90 million in biomarker research... MJFF has enabled researchers to identify several promising biomarker candidates. . . MJFF has invested \$60 million to develop a solution that would accelerate verification of PD

biomarkers. PPMI is the resulting “**collaboration**” between researchers, industry, government and study participants.



# Open Big Data: TCGA

---

The Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), aims to generate comprehensive, multi-dimensional maps of the key genomic changes . . . initiated in 2006 . . . showed that a national network of research and technology teams . . . could pool the results of their efforts . . . that making the data freely available would enable researchers make important discoveries.



# Open Big Data: The Reality

## Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Use Agreement

I ***will not further disclose these data*** beyond the uses outlined in this agreement and my data use application and understand that redistribution of data in any manner is Prohibited.

## Parkinson's Progression Markers Initiative (PPMI) Data Use Agreement

I ***will not further disclose these data*** beyond the uses outlined in this agreement and my data use application.

I will do my best to ensure that Investigators who utilize PPMI data use appropriate administrative, physical and technical safeguards ***to prevent use or disclosure of the data*** other than as provided for by this Agreement and will promptly report any use or disclosure of data that does not comply with the guidelines established by this Agreement.

## The Cancer Genome Atlas (TCGA) Data Use Agreement

The Requester and Approved Users agree to retain control over the data and further ***agree not to distribute data*** obtained through this Data Access Request to any entity or individual not covered in the submitted Data Access.



# Key Outcomes:

---

## Very interesting preliminary findings

- **4 SNP's that predict Parkinson's progression BETTER than Age**
- **3 Biomarkers that cross Alzheimers and Parkinsons**
- **Sets of biomarkers that predict disease stage**

Barash CI, et al 2015;6:42–4. • More “Open” Neurodegenerative Disease Data

- tranSMART Foundation, MJFF and LONI are working together to enable the distribution and access of ADNI, PPMI, LRRK2 and BioFIND datasets in tranSMART.



## Key Outcomes:

---

- Bringing data scientists, neuroscientists and biostatisticians together to leverage an integrated cross-neurodegenerative disease dataset in tranSMART validated the utility for the platform, and the value of these data when integrated.
- The Datathon produced an innovative approach using machine learning, new biomarker findings, and scientifically thorough pathway analyses.
- It successfully demonstrated that data sharing and the Datathon approach in particular can expedite discovery, as well as offer new fruitful ways to explore datasets. These new approaches are available for the research community to adopt.



## Key Outcomes:

---

- The tranSMART Platform can support many different large and distributed datasets as well as meet varied end-user needs simultaneously.
- The platform's ability to organize these datasets in useful fashions enabled participants to generate new findings and approaches for future research.
- The Platform proved to be a powerful tool in expediting research in a cost efficient way.
- The value of bringing together experts from different disciplines and organizations to work in newly formed teams on integrated datasets deployed in the tranSMART platform was evident.
- The Datathon also provided excellent technical and scientific feedback on how to improve the tranSMART platform for use by the research community in real scientific projects.



# **G-DOC Georgetown Database of Cancer (2012)**

---

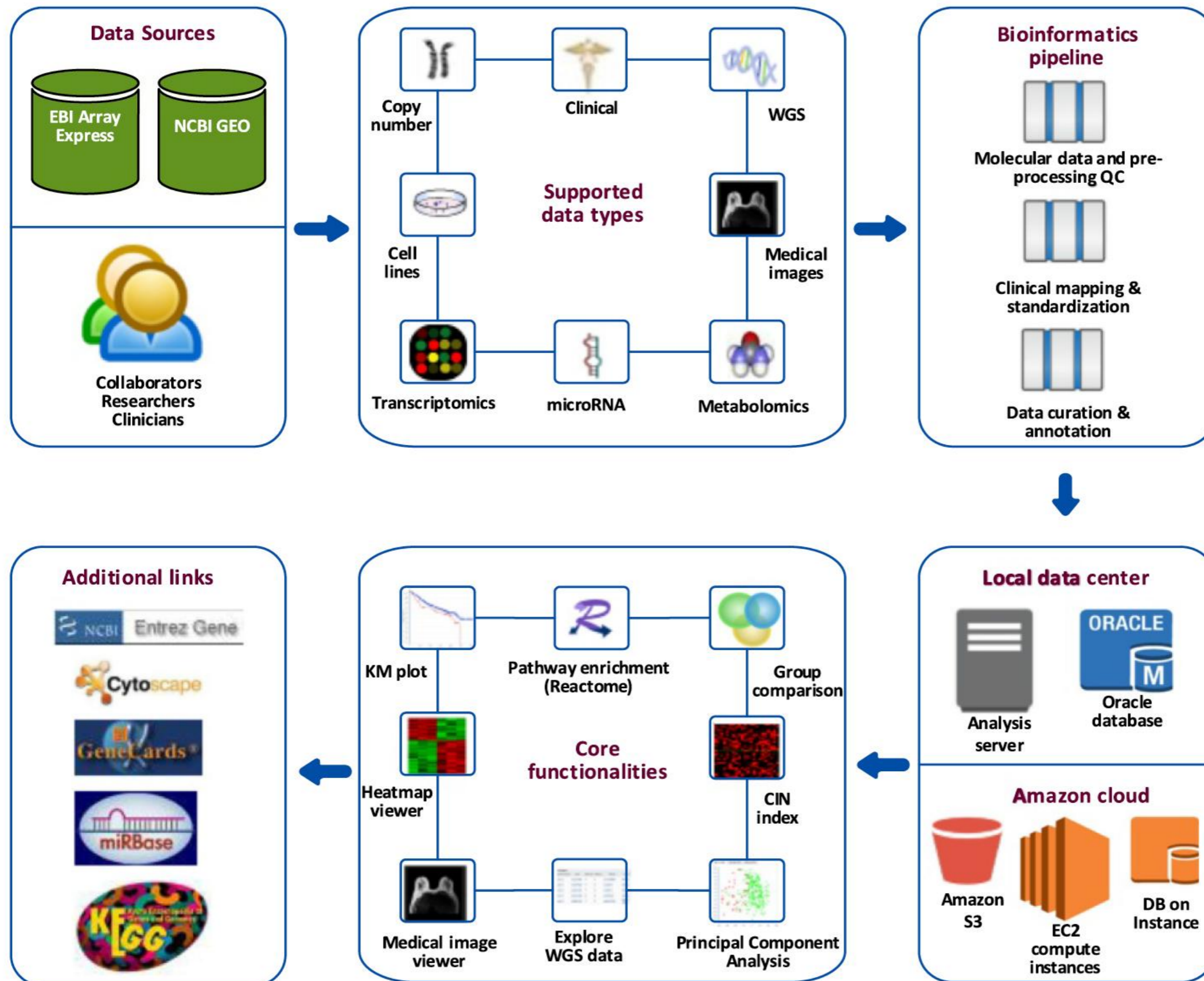
**G- Doc Plus uses advanced computational tools to handle a variety of biomedical BIG DATA.**

**It has 3 entry points for the user based on their interest.**

- 1. Precision Medicine workflow**
- 2. Translation Research workflow**
- 3. Population Genetics workflow**



# Summary of G-DOC Plus data management and analysis features



# Publications using G-DOC

Madhavan S, Gusev Y, Natarajan T, Song L, Bhuvaneshwar K, Gauba R, Pandey A, Haddad B, Goerlitz D, Cheema A, Juhl H, Kallakury B, Marshall J, Byers S, Weiner L. **Genome-wide multi-omics profiling of colorectal cancer identifies immune determinants strongly associated with relapse.** *Frontiers in Genetics*. Nov, 2013. PubMed

Madhavan S, Gauba R, Song L, Bhuvaneshwar K, Gusev Y, Byers S, Juhl H, Weiner L. **Platform for Personalized Oncology: Integrative analyses reveal novel molecular signatures associated with colorectal cancer relapse.** *AMIA Summits Transl Sci Proc*. 2013 Mar 18; 2013:118. PubMed

Madhavan S. The Development of Informatics Platforms to Help Drive Systems Medicine. **Genome-wide multi-omics profiling of colorectal cancer identifies immune determinants strongly associated with relapse.** *Handbook of Personalized Medicine: Advances in Nanotechnology, Drug Delivery and Therapy*. Pan Stanford Publishing (2013).

Gusev Y, Riggins RB, Bhuvaneshwar K, Gauba R, Sheahan L, Clarke R, and Madhavan S. **In silico discovery of mitosis regulation networks associated with early distant metastases in estrogen receptor positive breast cancers.** *Cancer Informatics* 2013:12 31-51 PubMed

Madhavan S, Gusev Y, Harris MA, Tanenbaum DM, Gauba R, Bhuvaneshwar K, Shinohara A, Rosso K, Carabet LA, Song L, Riggins RB, Dakshanamurthy S, Wang Y, Byers SW, Clarke R, and Weiner LM. **G-CODE: enabling systems medicine through innovative informatics.** *Genome Biology*, 12 (Suppl 1):P38, September 2011. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3439034/>

Elgamal, OA, Park, J, Gusev Y, Azevedo-Pouly AC, Jiang J, Roopra A and Schmittgen TD. **Tumor Suppressive Function of mir-205 in Breast Cancer Is Linked to HMGB3 Regulation,** *PLoS One*, 2013 Oct 2;8(10):e76402 PubMed



# The long-term vision

---

- **is to extend translational bioinformatics platforms to stay current with emerging omics technologies and analysis methods to continue supporting novel hypothesis generation, analysis and validation for integrative biomedical research.**
- **integrating several aspects of the disease and exposing various data elements, such as outpatient lab workup, pathology, radiology, current treatments, molecular signatures and expected outcomes over a web interface, these platforms will continue to strengthen precision medicine research.**



# The long-term vision

---

- **Using former or existing platforms' best practices**
- **Designing architectures that can survive in time**
- **Evaluating low cost solutions**
- **Learning from huge existing databases in other fields of science**
- **Thinking about metadata**
- **Believing in future high performance computing**

# BiHELab – Ελληνική Φαρμακευτική Εταιρεία

---

**HelpDesk for existing translational platforms :**

- **PhD Candidate : Αντιγόνη Αβραμούλη**
- **PhD Candidate : Σωτήρης Φωτίου**

Focusing on one platform risks  
missing an obvious signal!!!

