Biomedical Data Sources
The Power of Big Data

Aik Choon Tan, Ph.D.
Vice-Chair and Senior Member
Department of Biostatistics and Bioinformatics
aikchoon.tan@moffitt.org
6/13/2022

Outline

- Why Big Data is important in your biomedical research?
- Examples of Biomedical Big Data
  - GDC, TCGA
  - GEO
  - Pharmacogenomics - CMAP, LINCS, GDSC, CCLE
  - Imaging - Cancer Imaging data (TCIA)
  - Immunology - ImmPort
  - Clinical Trials
  - Mobile Data
  - Other data types - Social Media Data
- Conclusions

Simplified View on Disease

Cell, Vol. 61, 759–767, June 1, 1990, Copyright © 1990 by Cell Press

A Genetic Model for Colorectal Tumorigenesis

Eric R. Fearon and Bert Vogelstein
The Oncology Center
Program in Human Genetics
The Johns Hopkins University School of Medicine
Baltimore, Maryland 21231

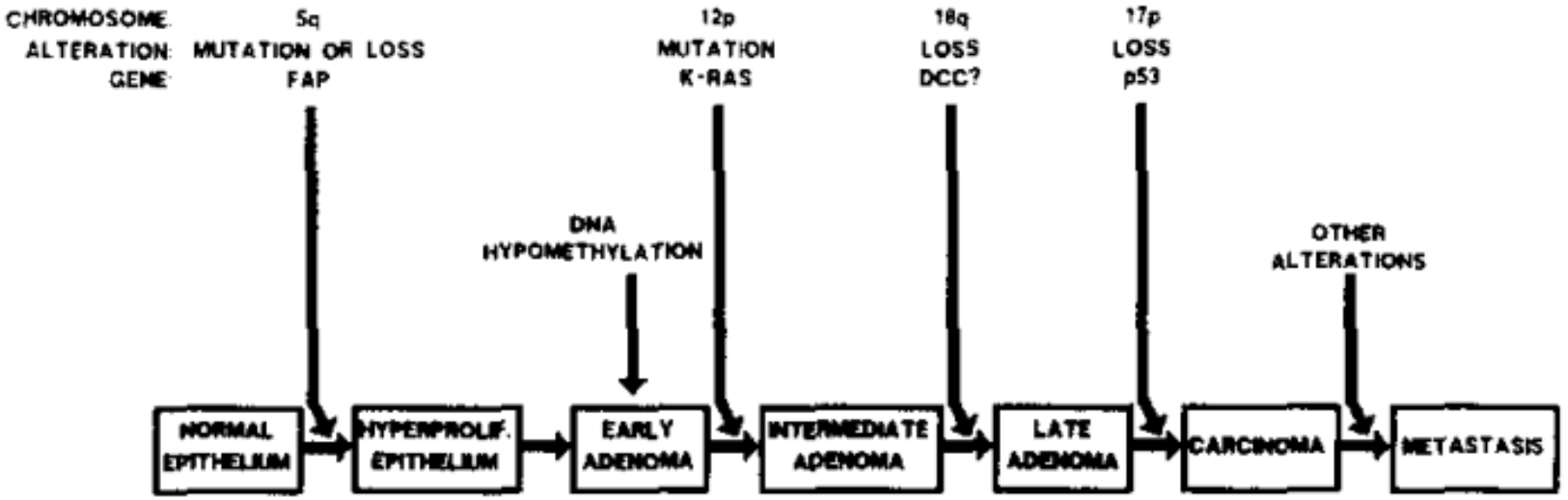


Figure 3. A Genetic Model for Colorectal Tumorigenesis

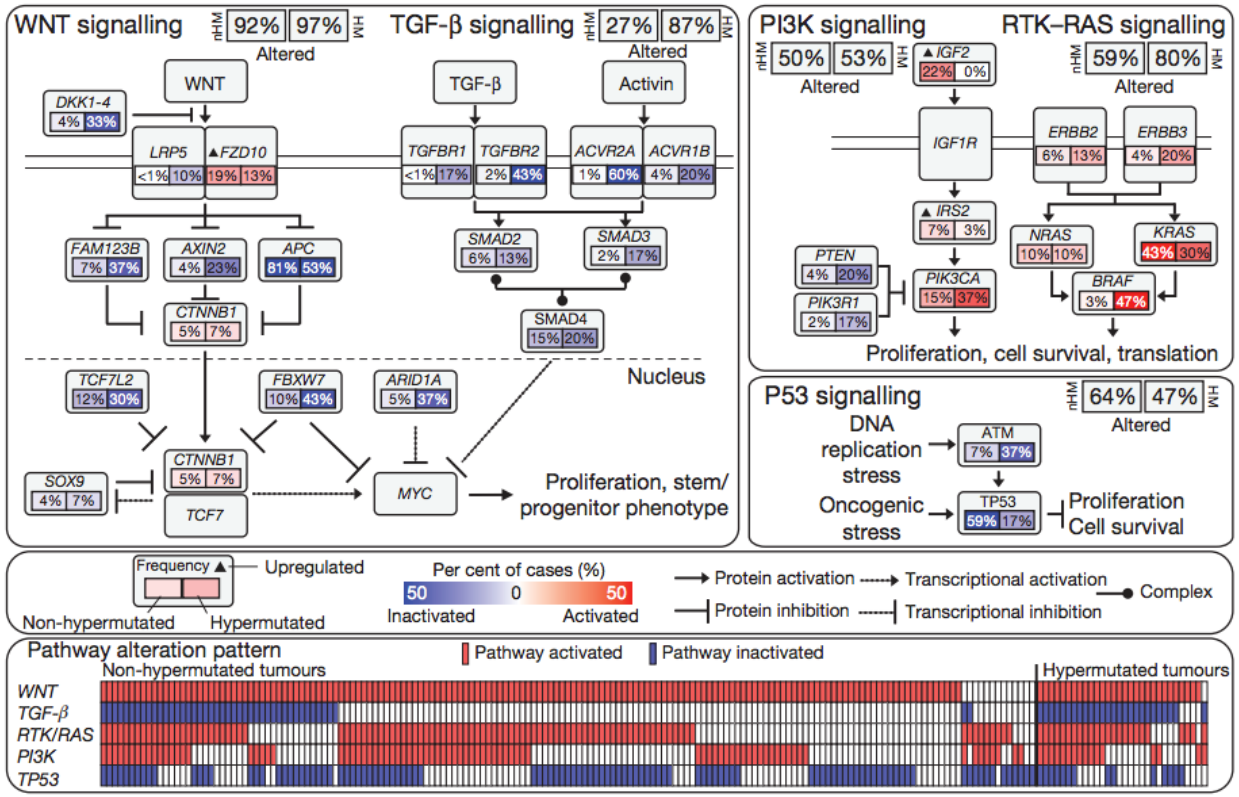
In reality … Complex Networks in Disease

ARTICLE

doi:10.1038/nature11252

Comprehensive molecular characterization of human colon and rectal cancer

The Cancer Genome Atlas Network*



Computational Systems Biology

insight overview

Computational systems biology

Hiroaki Kitano

Sony Computer Science Laboratories Inc., 3-14-13 Higashi-gotanda, Shinagwa, Tokyo 141-0022, ERATO Kitano Symbiotic Systems Project, Japan Science and Technology Corporation, and The Systems Biology Institute, Suite 6A, M31, 6-31-15 Jingu-mae, Shibuya, Tokyo 150-0001, School of Fundamental Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan, and Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125, USA (e-mail: kitano@csl.sony.co.jp)

To understand complex biological systems requires the integration of experimental and computational research — in other words a systems biology approach. Computational biology, through pragmatic modelling and theoretical exploration, provides a powerful foundation from which to address critical scientific questions head-on. The reviews in this Insight cover many different aspects of this energetic field, although all, in one way or another, illuminate the functioning of modular circuits, including their robustness, design and manipulation. Computational systems biology addresses questions fundamental to our understanding of life, yet progress here will lead to practical innovations in medicine, drug discovery and engineering.

(Kitano, Nature, 2002)

Computational Systems Biology

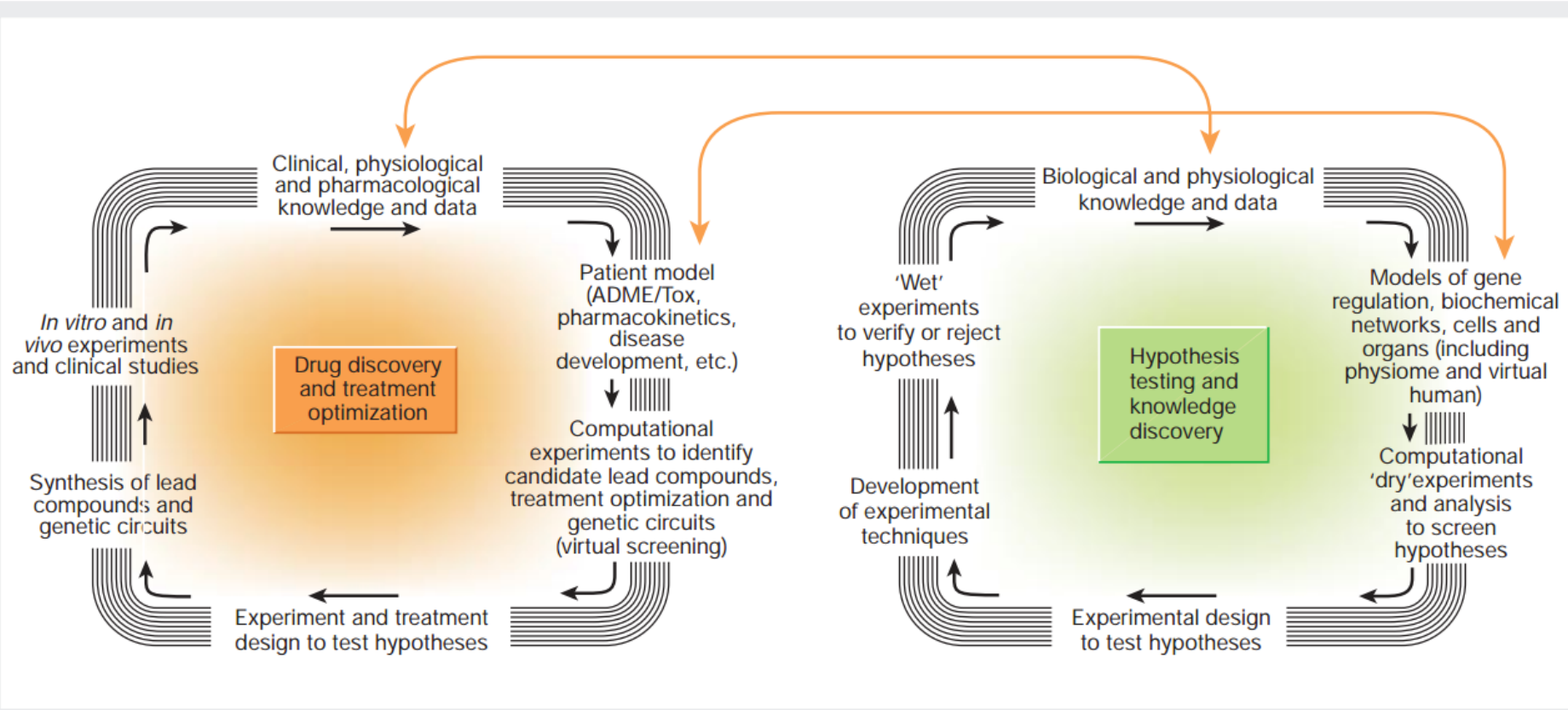


Figure 1 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments[18]. With sufficient progress in basic systems biology, this cycle can be applied to drug discovery and the development of new treatments. In the future, *in silico* experiments and screening of lead candidates and multiple drug systems, as well as introduced genetic circuits, will have a key role in the 'upstream' processes of the pharmaceutical industry, significantly reducing costs and increasing the success of product and service development.

(Kitano, Nature, 2002)

Data Driven Biology

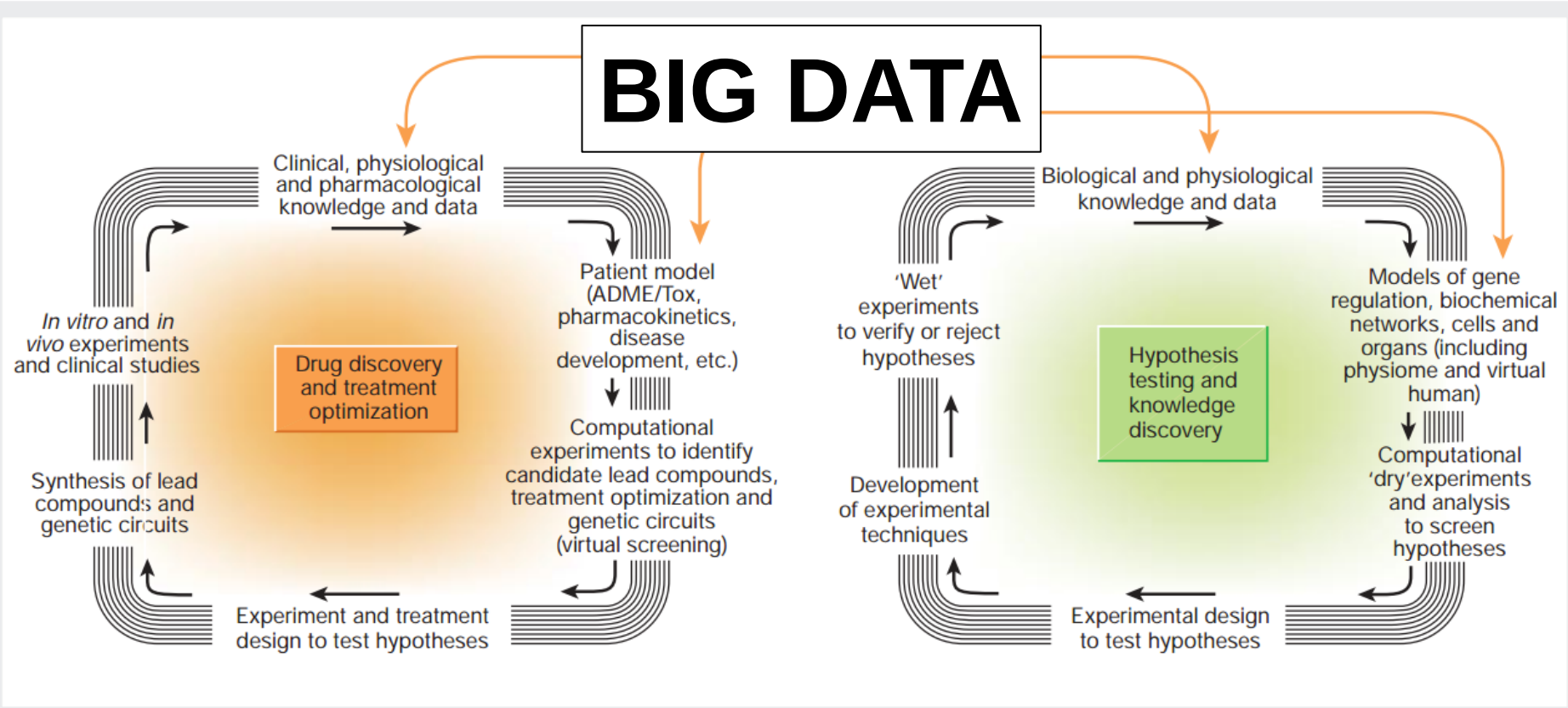


Figure 1 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments[18]. With sufficient progress in basic systems biology, this cycle can be applied to drug discovery and the development of new treatments. In the future, *in silico* experiments and screening of lead candidates and multiple drug systems, as well as introduced genetic circuits, will have a key role in the 'upstream' processes of the pharmaceutical industry, significantly reducing costs and increasing the success of product and service development.

(Kitano, Nature, 2002)

Big Data ("Omics") : Maps and Catalogs

- Maps: Structure
  - Genetic Map
  - Physical Map
  - Sequence Map
- Maps: Molecular Function
  - Gene Map
  - Evolutionary Conservation Map
  - Chromatin State Map
  - 3-D Folding Map
- Maps: Disease
  - Inherited Variation Map
  - Disease Association Map
  - Evolutionary Selection Map
  - Cancer Gene Map
- Catalogs: Signatures
  - Gene Expression
  - Protein Expression

Look up Table in Biology

(like periodic table in chemistry)

(Adapted from Eric Lander's slide)

Genomics Data Commons

portal.gdc.cancer.gov

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home | Projects | Exploration | Analysis | Repository | Quick Search | Manage Sets | Login | Cart 1 | GDC Apps

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects | Exploration | Analysis | Repository

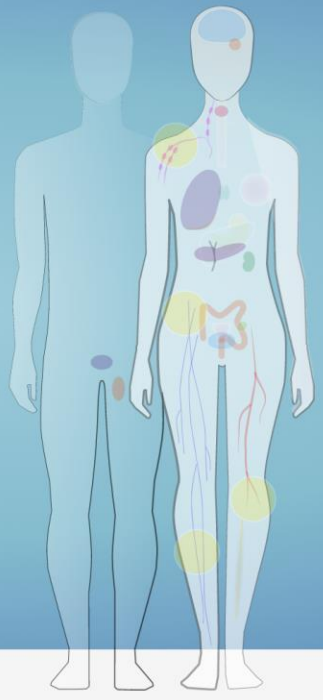
e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

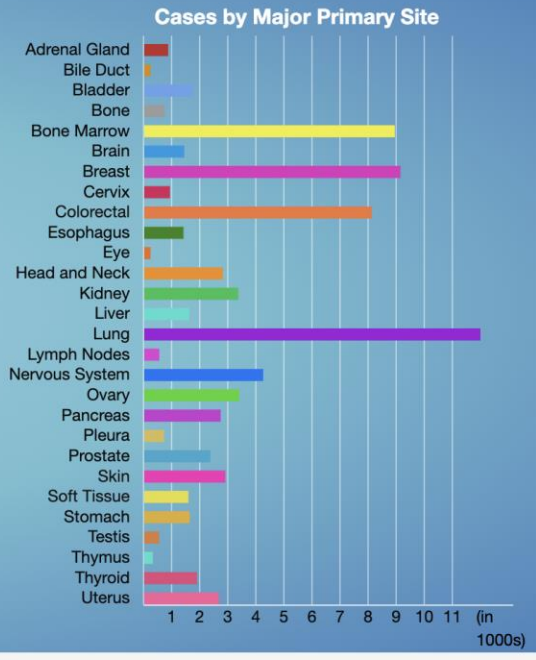
Data Portal Summary

Data Release 33.1 - May 31, 2022

PROJECTS	PRIMARY SITES	CASES
72	67	85,552

FILES	GENES	MUTATIONS
828,104	21,759	2,683,650





GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

Data Portal | Website | API | Data Transfer Tool | Documentation | Data Submission Portal | Legacy Archive | Publications

The Cancer Genome Atlas Project

- The Cancer Genome Atlas (TCGA) began as a three-year pilot in 2006 with an investment of $50 million each from the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI).
- Pilot Project: Comprehensive Characterized of three tumor types: Glioblastoma (GBM), Ovarian and lung cancers (> 200 samples)
- 2009 expand to 20 common tumor types

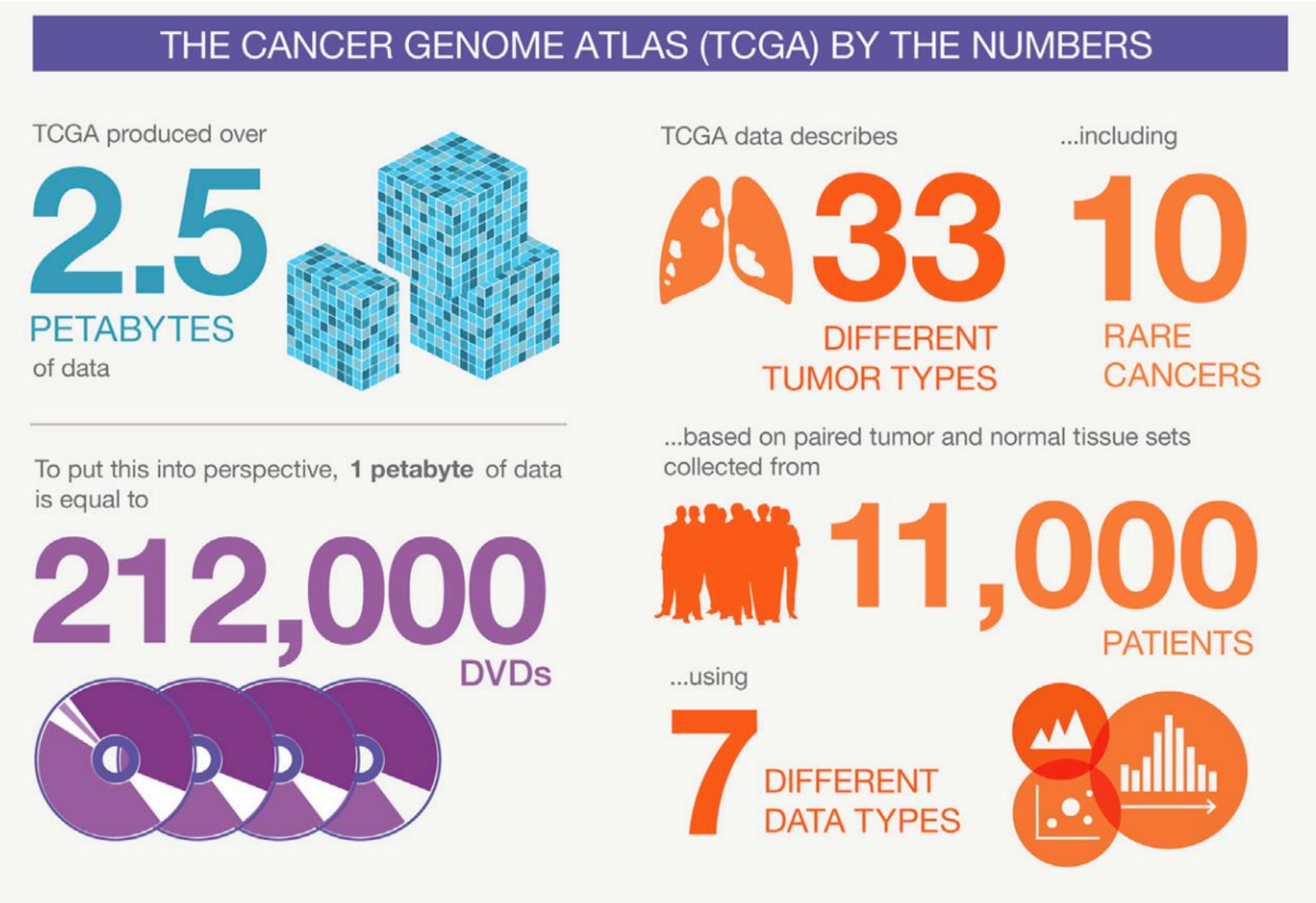
BOX 1 TCGA: MISSION AND STRATEGY

Important information about the biological relevance of the molecular changes in cancer can be obtained through combined analysis of multiple different types of data.

For that reason, TCGA's principal aims are to generate, quality control, merge, analyze and interpret molecular profiles at the DNA, RNA, protein and epigenetic levels for hundreds of clinical tumors representing various tumor types and their subtypes. Cases that meet quality assurance specifications are characterized using technologies that assess the sequence of the exome, copy number variation (measured by SNP arrays), DNA methylation, mRNA expression and sequence, microRNA expression and transcript splice variation. Additional platforms applied to a subset of the tumors, including whole-genome sequencing and RPPAs, provide additional layers of data to complement the core genomic data sets and clinical data. By the end of 2015, the TCGA Research Network plans to have achieved the ambitious goal of analyzing the genomic, epigenomic and gene expression profiles of more than 10,000 specimens from more than 25 different tumor types.

TCGA has other, complementary aims as well: to promote the development and application of new technologies, to detect cancer-specific molecular alterations, to make data and results freely available to the scientific community, to develop tools and standard operating procedures that can serve other large-scale profiling projects and to build cadres of individuals (including experimentalists, computational biologists, statistical analysts, computer scientists and administrative staff) with the expertise to carry out such large-scale, team science projects. As of 24 July 2013, TCGA had mapped molecular patterns across 7,992 total cases representing 27 tumor types. The data, along with tools for exploring them, are publicly available at http://www.cancergenome.nih.gov/. Eight 'marker papers' (comprehensive initial publications on each of the tumor types) have been published so far[8–12,14,27].

The Cancer Genome Atlas (TCGA)

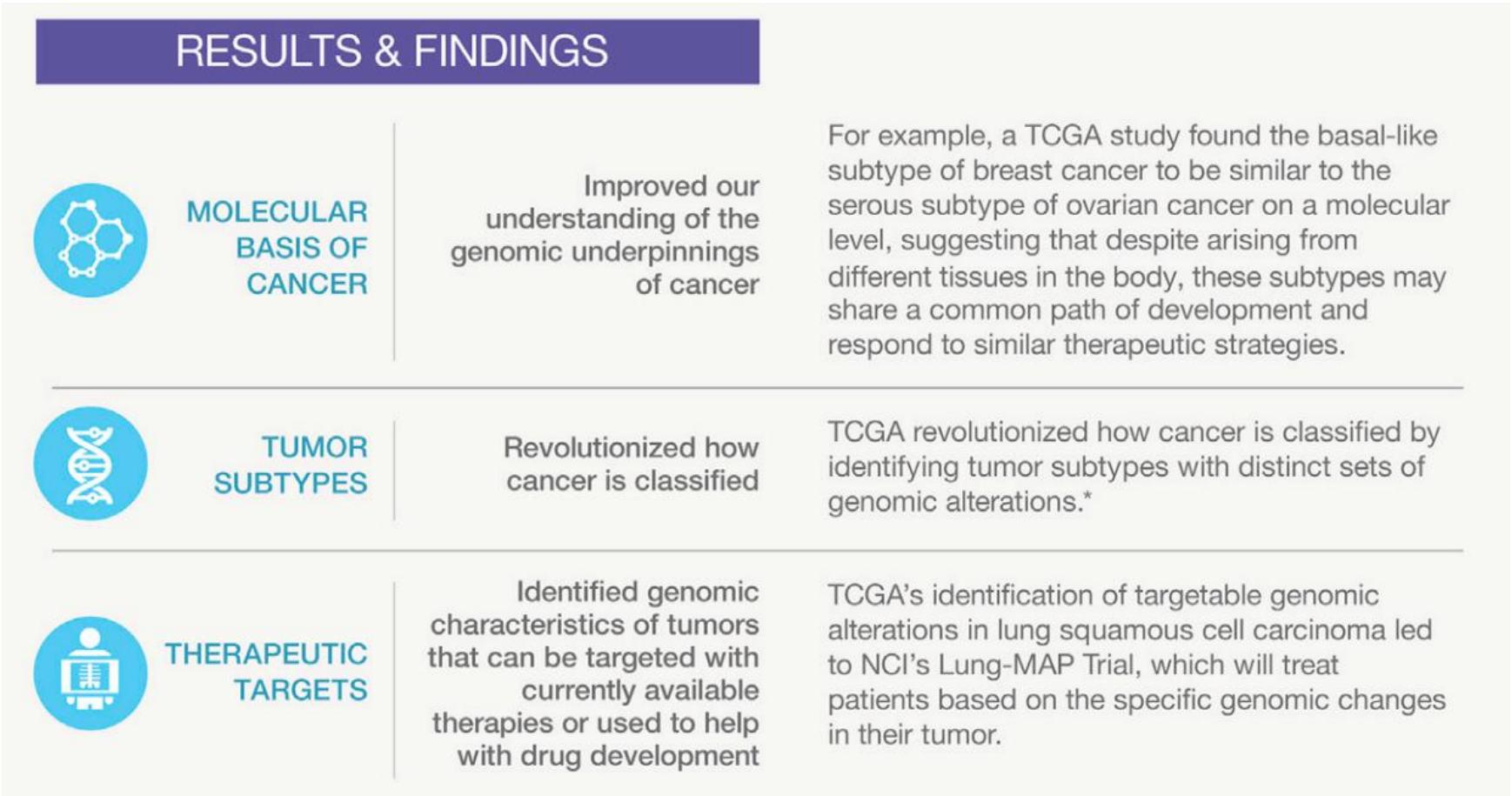


(Hutter & Zenklusen (2018) *Cell*)

Cancer type	Prevalence	TCGA cases assessed	Key findings
Breast lobular carcinoma	3,327,552	203	FOXA1 elevated in lobular carcinoma, GATA3 in ductal carcinoma; lobular enriched for PTEN loss and Akt activation
Breast ductal carcinoma		784	Four distinct subtypes: basal, Her2, luminal A, and luminal B; most common driver mutations: TP53, PIK3CA, GATA3; basal subtype similar to serous ovarian cancer
Prostate cancer	3,085,209	333	Highly heterogeneous with 26% driven by unknown alterations; ETS gene fusions or mutations in SPOP, FOXA1, or IDH1 define seven subtypes; actionable lesions in PI3K, MAPK, and DNA repair pathways
Colorectal adenocarcinoma	1,317,247	276	Colon and rectal cancers have similar genomic profiles; hypermutated subtype associated with favorable prognosis; new potential drivers: ARID1A, SOX9, FAM123B/WTX
Cutaneous melanoma	1,169,351	331	Established four subtypes: BRAF mutant, RAS mutant, NF1 mutant, and triple wild-type based on driver mutations; higher levels of immune lymphocyte infiltration correlated with better survival
Thyroid carcinoma	726,646	496	Majority driven by RAS or BRAFV600E mutations
Endometrial carcinoma	710,228	373	Classified endometrial cancers into four categories: POLE ultramutated, MSI (microsatellite instability) hypermutated, copy-number low, copy-number high
Uterine carcinosarcoma		57	Strong and varied degree of epithelial-mesenchymal transition; TP53 mutations in 91% of samples; PI3K alterations in half
Invasive urothelial bladder carcinoma	696,440	131	Increased risk associated with smoking; frequently mutated; TP53 inactivated in 76% of tumors, ERBB2 (HER2), genes in the RTK/RAS pathways altered in 44%
Lung adenocarcinoma	527,228	230	High mutation burden; 76% have activation of receptor tyrosine kinase pathways
Lung squamous cell carcinoma		178	High average number of mutations and copy-number aberrations; almost all have mutation in TP53; many have inactivating mutations in HLA-A that may aid immune evasion
Clear cell renal cell carcinoma	483,225	446	Commonly mutated genes: VHL, SED2, and the PI3K/AKT/mTOR pathway; metabolic shift similar to the "Warburg effect" correlates with a poor prognosis
Kidney papillary carcinoma		161	81% of type 1 tumors had MET alteration; type 2 tumors were heterogeneous, with alterations to CDKN2A, SETD2, TFE3, or increased expression of NRF2–ARE pathway; loss of CDKN2A expression and CpG island methylation phenotype associated with poor outcome
Chromophobe renal cell carcinoma		66	Extremely low mutation burden; metabolic shift distinct from the "Warburg effect" shift in clear cell carcinoma; TP53 and PTEN were frequently mutated; TERT gene promoter was frequently altered
Cervical cancer	256,078	228	Identification of HPV-negative, endometrial-like cancers with mutations in KRAS, ARID1A, and PTEN; amplification of CD274 and PDCD1LG2; frequent alterations in MED1, ERBB3, CASP8, HLA-A, and TGFBR2 and fusions involving lncRNA BCAR4; nearly three-quarters had alterations in either or both of the PI3K/MAPK and TGF-beta pathways
Testicular germ cell cancer	251,194	150	
Ovarian serous adenocarcinoma	222,060	489	Mutations: TP53 (in 96%), BRCA1 and BRCA2 (in 21%) and are associated with more favorable outcomes
Glioblastoma multiforme	162341	206	GBM subtypes Classical, Mesenchymal, and Proneural are defined by EGFR, NF1, and PDGFRA/IDH1 mutations, respectively; over 40% have mutations in chromatin-modifiers; frequently mutated: TP53, PIK3R1, PIK3CA, IDH1, PTEN, RB1, LZTR1
Lower-grade glioma		293	Defined three subtypes correlating with patient outcomes: IDH1 mutant with 1p/19q deletion, IDH mutant without 1p/19q deletion, and IDH wild-type
Stomach adenocarcinoma	95,764	295	Identified four subytpes characterized by EBV infection, microsatellite instability, genomic stability, and chromosomal instability
Liver hepatocellular carcinoma	66,771	363	TERT promoter mutations in 44%; TP53 commonly mutated or under-expressed; CTNNBB1 significantly mutated; many tumors had high levels of lymphocyte infiltration or overexpressed immune checkpoint genes
Cholangiocarcinoma		38	Low expression of CDKN2, BAP1, and ARID1 genes and overexpression of FGFR2 and IDH1/2 genes; four subtypes defined
Pancreatic ductal adenocarcinoma	64,668	150	KRAS mutations present in 93% of tumors; mutations in RREB1 or other members of RAS-MAPK signaling pathway
Esophageal carcinoma	45,547	164	Squamous cell carcinomas had frequent amplifications of CCND1, SOX2, and TP63; adenocarcinomas had frequent amplifications in ERBB2, VEGFA, GATA4, and GATA6
Acute myeloid leukemia		200	Low mutation burden—only 13 coding mutations on average per tumor; classified driver events into nine categories including transcription factor fusions, histone modifier mutations, and spliceosome mutations
Head and neck squamous cell carcinoma		279	HPV-positive associated with shortened or deleted TRAF3, HPV-negative characterized by co-amplification of 11q13 and 11q22, smoking-related characterized by TP53 mutations, CDKN2A inactivation, CNVs
Sarcoma		206	TP53, ATRX, and RB1 are recurrently mutated across all types; synovial sarcomas expressed fusions in SSX1 or SSX2 and TERT; JUN amplification associates with worse survival in dedifferentiated liposarcoma
Paraganglioma and pheochromocytoma		173	Four distinct subtypes: Wnt-altered, cortical admixture, pseudohypoxia, and kinase signaling; MAML3 fusion gene and CSDE1 somatic mutation define and drive the poor prognosis Wnt-altered subtype
Thymoma		124	
Adrenocortical carcinoma		91	Overexpression of IGF2, mutations in TP53, PRKAR1A and other genes, and copy-number alterations were common; hypoploidy followed by whole-genome doubling may be a driving mechanism
Mesothelioma		87	
Uveal melanoma		80	Complex mutation in BAP1 gene; identified distinct subdivisions of disomy 3 (D3) and monosomy 3 (M3) subtypes; in M3, mutually exclusive EIF1AX and SRSF2/SF3B1 mutations have distinct methylation profiles and prognoses

(Blum, Wang, & Zenklusen, Cell 2018)

The Cancer Genome Atlas (TCGA)



(Hutter & Zenklusen (2018) *Cell*)

Components of the TCGA Research Network

- **Biospecimen Core Resource (BCR)** – Tissue samples are carefully cataloged, processed, checked for quality and stored, complete with important medical information about the patient.
- **Genome Characterization Centers (GCCs)** – Several technologies will be used to analyze genomic changes involved in cancer. The genomic changes that are identified will be further studied by the Genome Sequencing Centers.
- **Genome Sequencing Centers (GSCs)** – High-throughput Genome Sequencing Centers will identify the changes in DNA sequences that are associated with specific types of cancer.
- **Proteome Characterization Centers (PCCs)** – The centers, a component of NCI's Clinical Proteomic Tumor Analysis Consortium, will ascertain and analyze the total proteomic content of a subset of TCGA samples.
- **Data Coordinating Center (DCC)** – The information that is generated by TCGA will be centrally managed at the DCC and entered into the TCGA Data Portal and Cancer Genomics Hub as it becomes available. Centralization of data facilitates data transfer between the network and the research community, and makes data analysis more efficient. The DCC manages the TCGA Data Portal.
- **Cancer Genomics Hub (CGHub)** – Lower level sequence data will be deposited into a secure repository. This database stores cancer genome sequences and alignments.
- **Genome Data Analysis Centers (GDACs)** – Immense amounts of data from array and second-generation sequencing technologies must be integrated across thousands of samples. These centers will provide novel informatics tools to the entire research community to facilitate broader use of TCGA data.

CGC Centers and Data Types

Center	Data Type
BCGAC	Illumina miRNA-seq
BROAD	SNP 6.0 (copy number)
HMS Harvard	Illumina DNA-seq
JHU/USC	Methylation
MD Anderson	RPPA
UNC	Agilent Microarray (gene exp)
	Illumina RNA-seq

Pan-Cancer Analysis of TCGA Data

Focus

Focus on TCGA Pan-Cancer Analysis

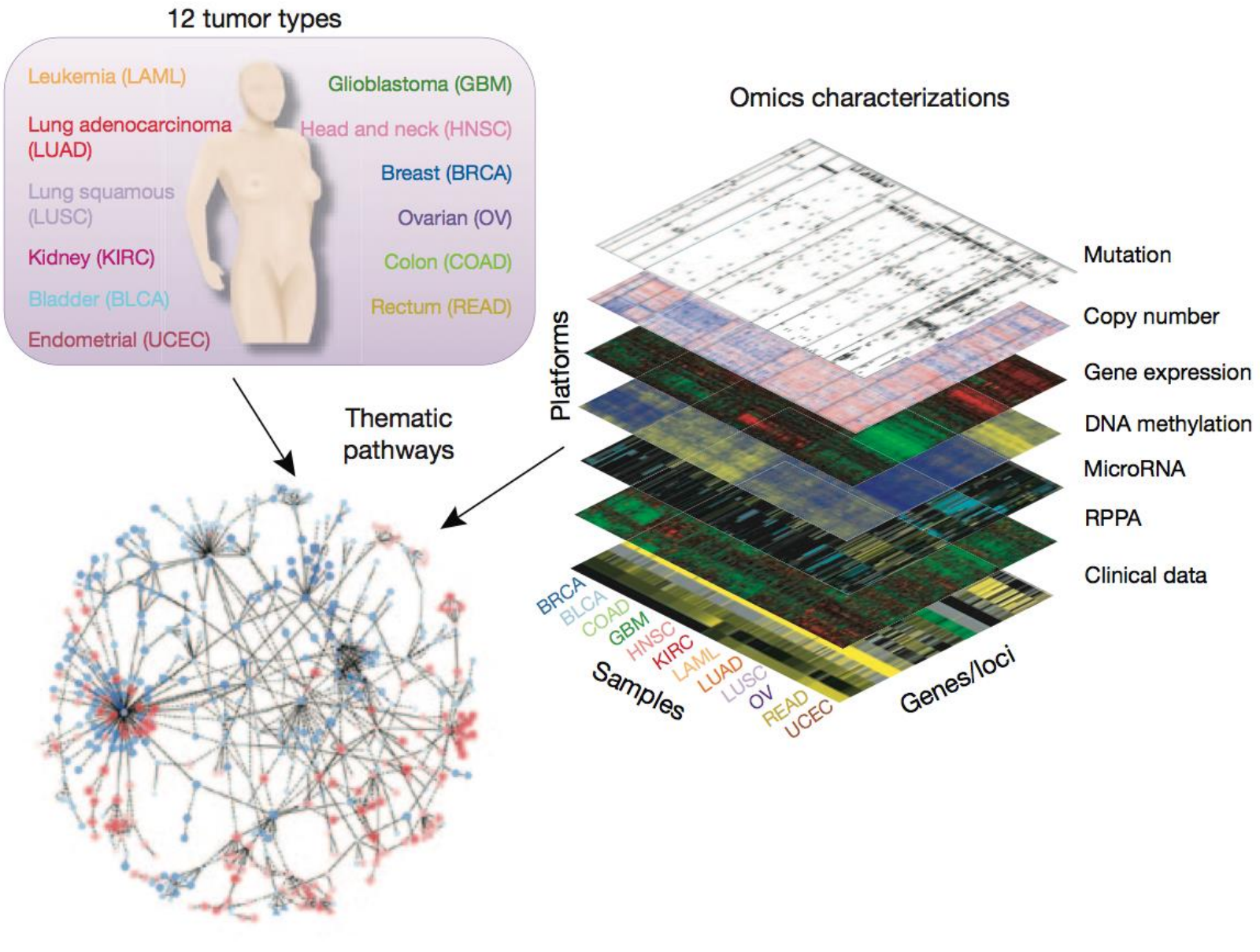
Focus issue: October 2013 Volume **45**, No 10

- **Focus contents**
- **Editorial**
- **Foreword**
- **Analysis**
- **Commentaries**
- **TCGA**
- **Podcast**

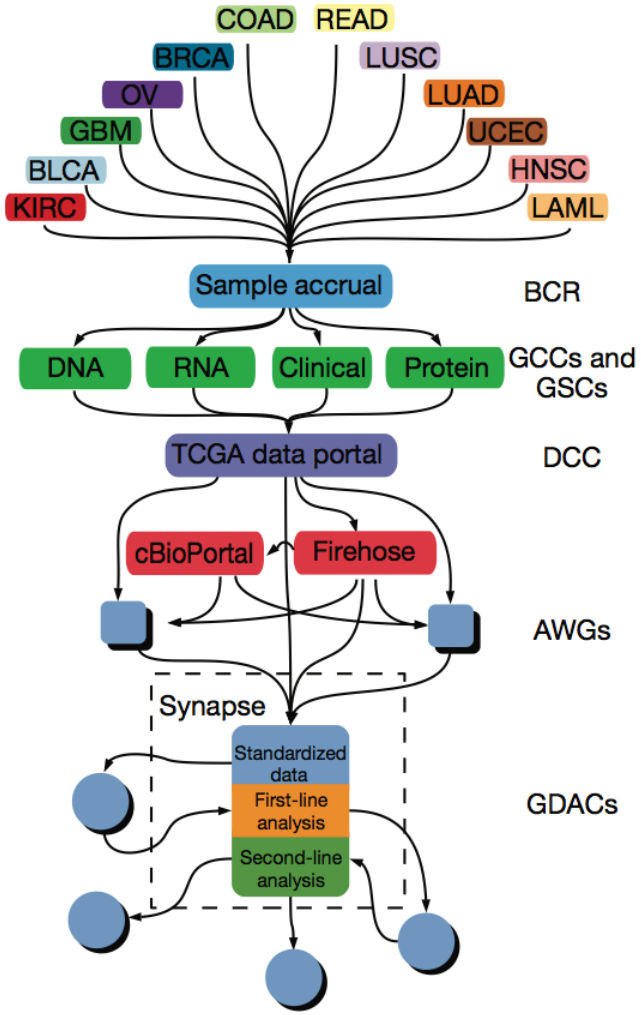
Genomic alterations in diverse cell types at different sites in the body give rise to hundreds of different forms of cancer and the ways in which these changes give rise to tumors with different biology, pathology and treatment strategies are beginning to be characterized. The Cancer Genome Atlas Research Network has catalogued the aberrations in the DNA, chromatin and RNA of the genomes of thousands of tumors relative to matched normal cellular genomes and have analyzed their epigenetic and protein consequences. Here, the Pan-cancer initiative examines the similarities and differences among the genomic and cellular alterations found in the first dozen tumor types to be profiled by TCGA. This first look across cancer offers new tools in genomics and bioinformatics and the prospect of repurposing targeted therapies directed by the molecular pathology of the tumors in addition to their clinical classification.

http://www.nature.com/ng/focus/tcga/index.html

Pan-Cancer Analysis of TCGA Data



Pan-Cancer Analysis of TCGA Data



The Cancer Genome Atlas Pan-Cancer analysis project

The Cancer Genome Atlas Research Network[1], John N Weinstein[2,3], Eric A Collisson[4], Gordon B Mills[3], Kenna R Mills Shaw[5,6], Brad A Ozenberger[7], Kyle Ellrott[8,9], Ilya Shmulevich[10], Chris Sander[11] & Joshua M Stuart[8,9]

The Cancer Genome Atlas (TCGA) Research Network has profiled and analyzed large numbers of human tumors to discover molecular aberrations at the DNA, RNA, protein and epigenetic levels. The resulting rich data provide a major opportunity to develop an integrated picture of commonalities, differences and emergent themes across tumor lineages. The Pan-Cancer initiative compares the first 12 tumor types profiled by TCGA. Analysis of the molecular aberrations and their functional roles across tumor types will teach us how to extend therapies effective in one cancer type to others with a similar genomic profile.

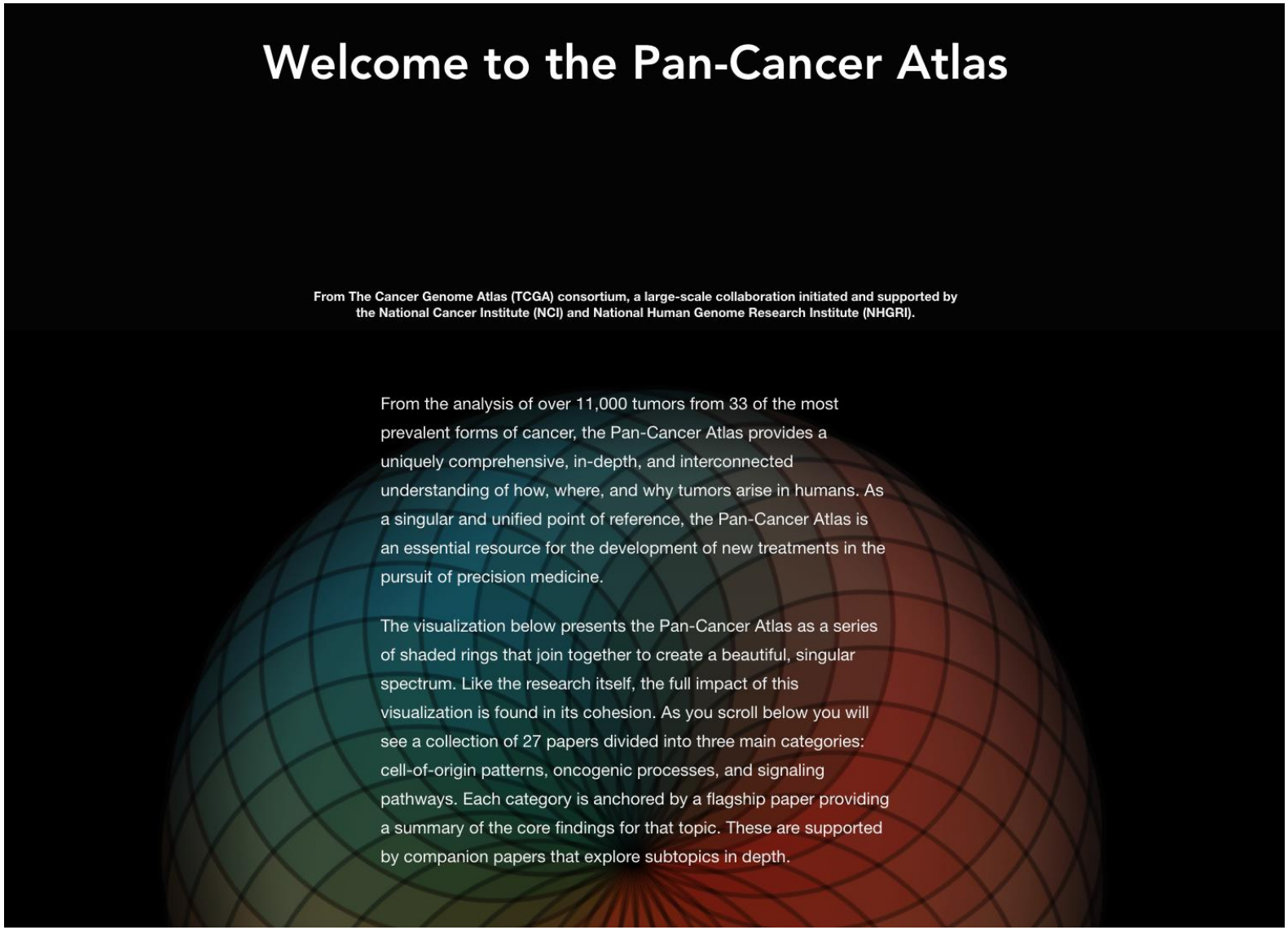
Table 1 Data freeze used by the Pan-Cancer project as defined on 21 December 2012

Cancer	RPPA[a]	DNA methylation[b]	Copy number[c]	Mutation[d]	microRNA[e]	Expression[f]
LUSC	195	358	345	178	332	227
READ	130	162	164	69	143	71
GBM	214	405	578	290	501	495
LAML	NA	194	198	197	187	179
HNSC	212	310	310	277	309	303
BLCA	54	126	126	99	121	96
KIRC	423	457	457	417	442	431
UCEC	200	512	511	248	497	333
LUAD	237	431	357	229	365	355
OV	332	592	577	316	454	581
BRCA	408	888	887	772	870	817
COAD	269	420	422	155	407	192
Total	2,674	4,855	4,932	3,247	4,628	4,080

Tabulated are the numbers of unique tumor samples available for each tumor type (rows) and each measurement platform (columns). NA, not available.
[a]Reverse-phase protein arrays measuring protein and phosphoprotein abundance. [b]DNA methylation at CpG islands. [c]Microarray-based measurement of copy number. [d]Samples subjected to whole-exome sequencing to determine single-nucleotide and structural variants. [e]Sequencing of microRNAs. [f]RNA sequencing and microarray gene expression analysis.

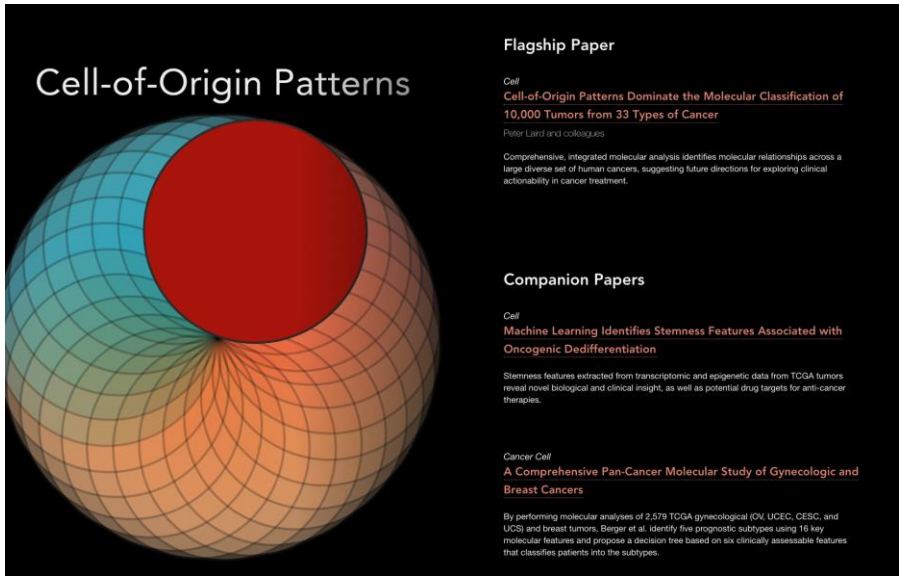
Pan-Cancer Atlas

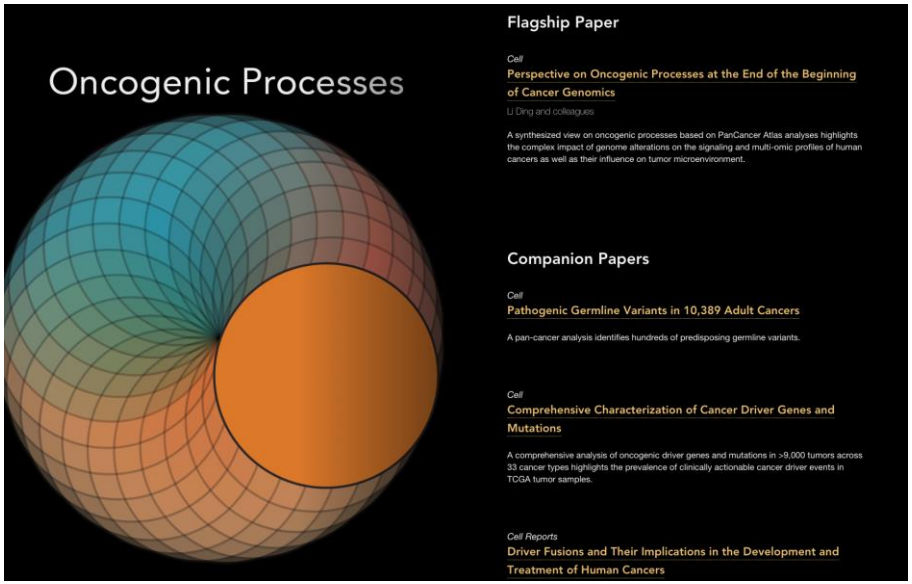
https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html

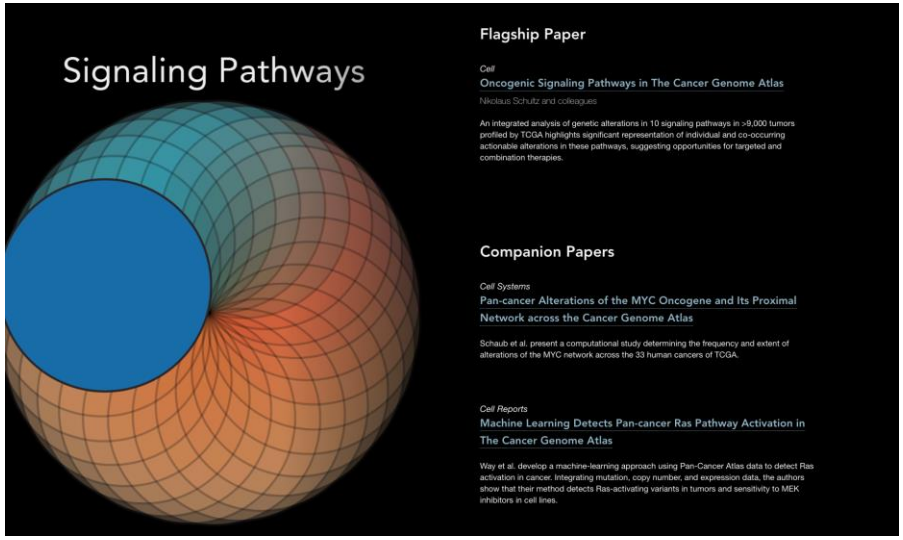


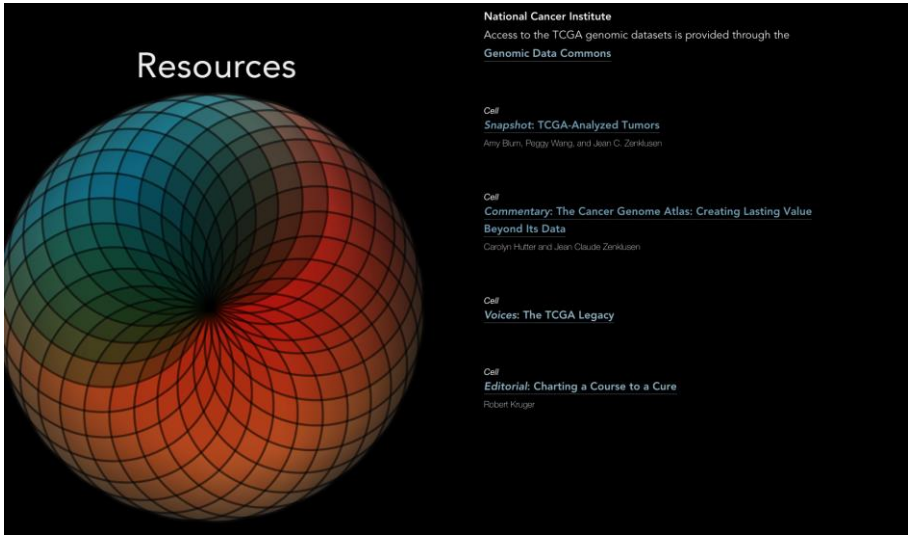
Pan-Cancer Atlas

https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html

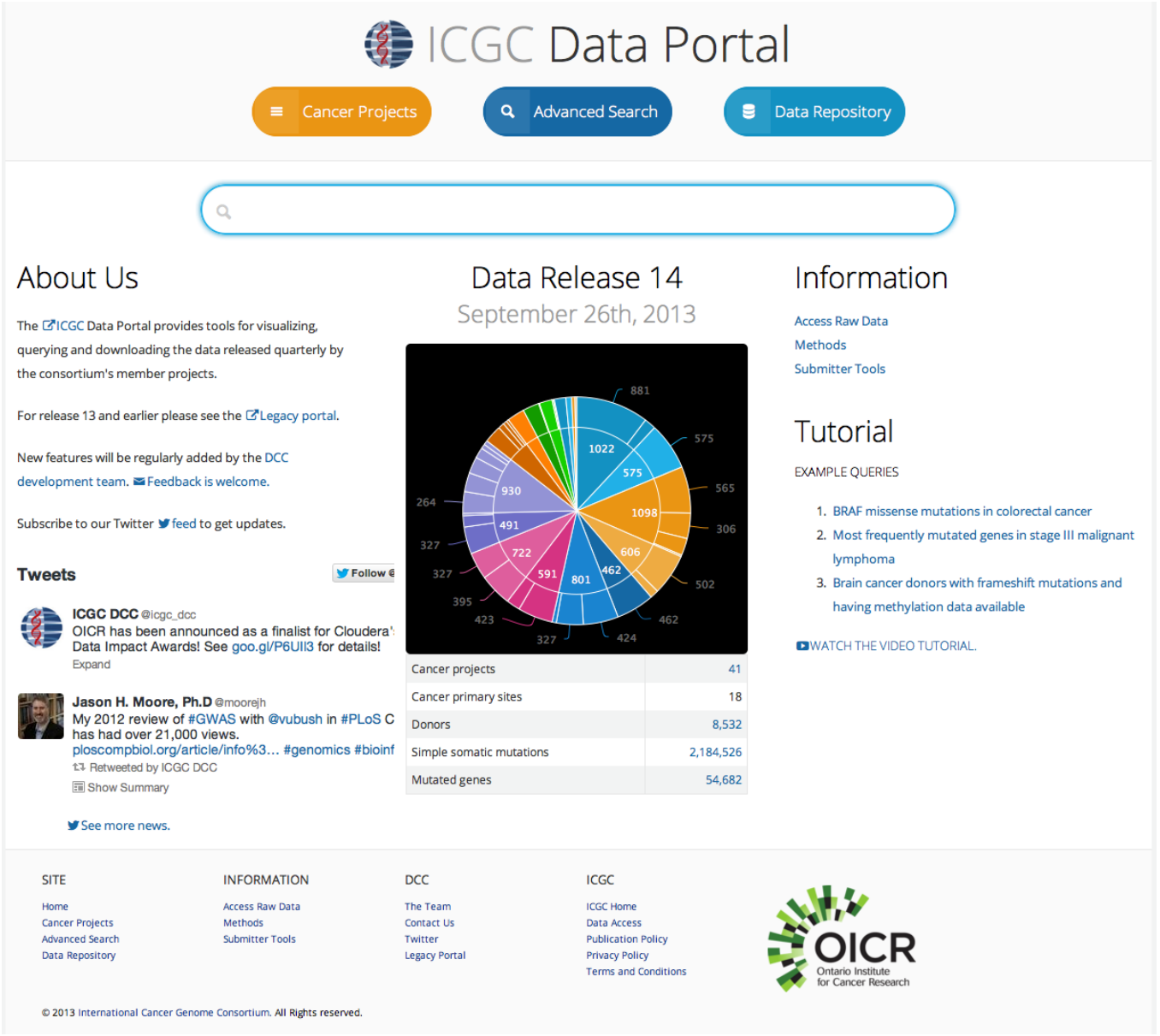






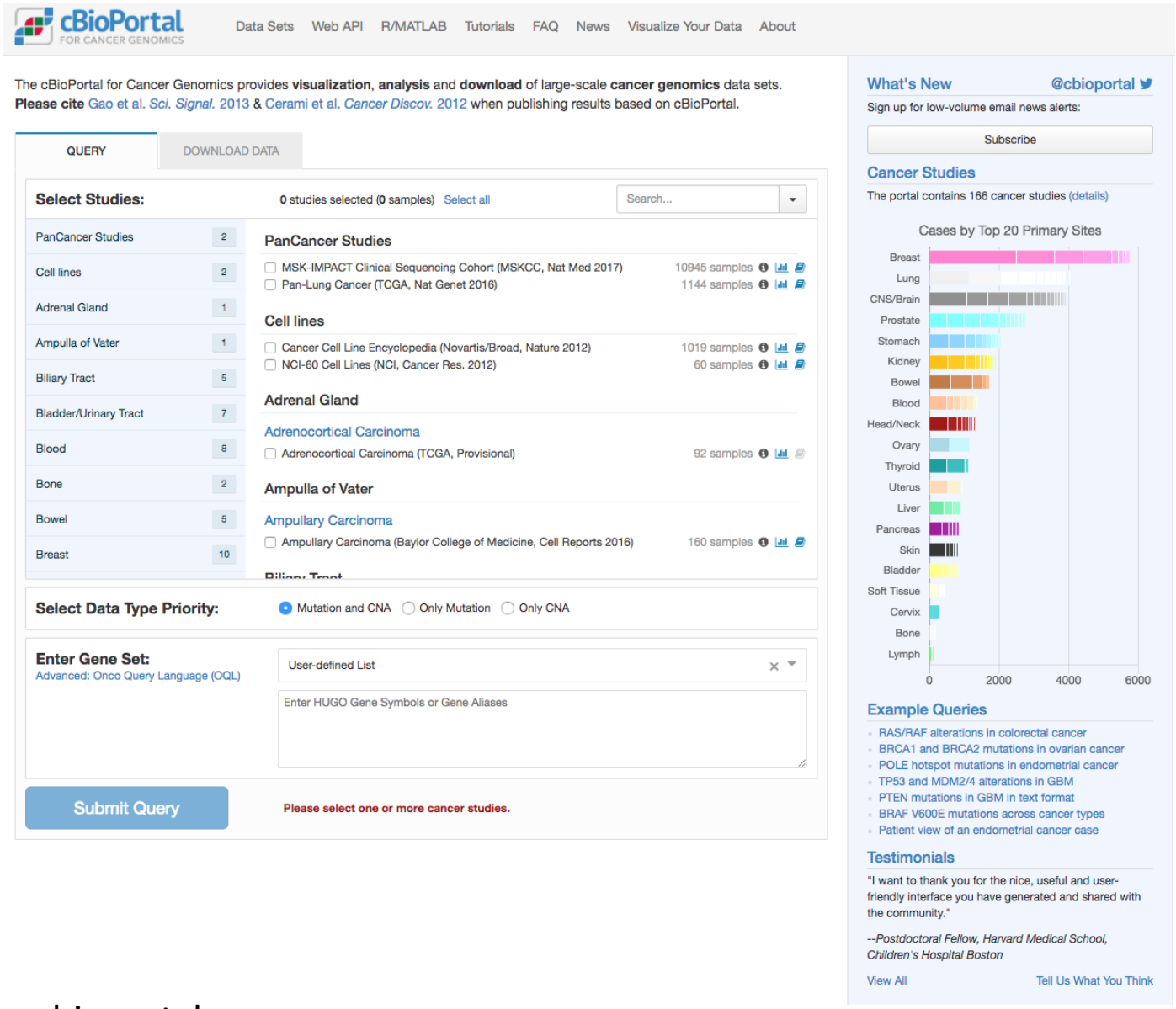


ICGC Data Portal



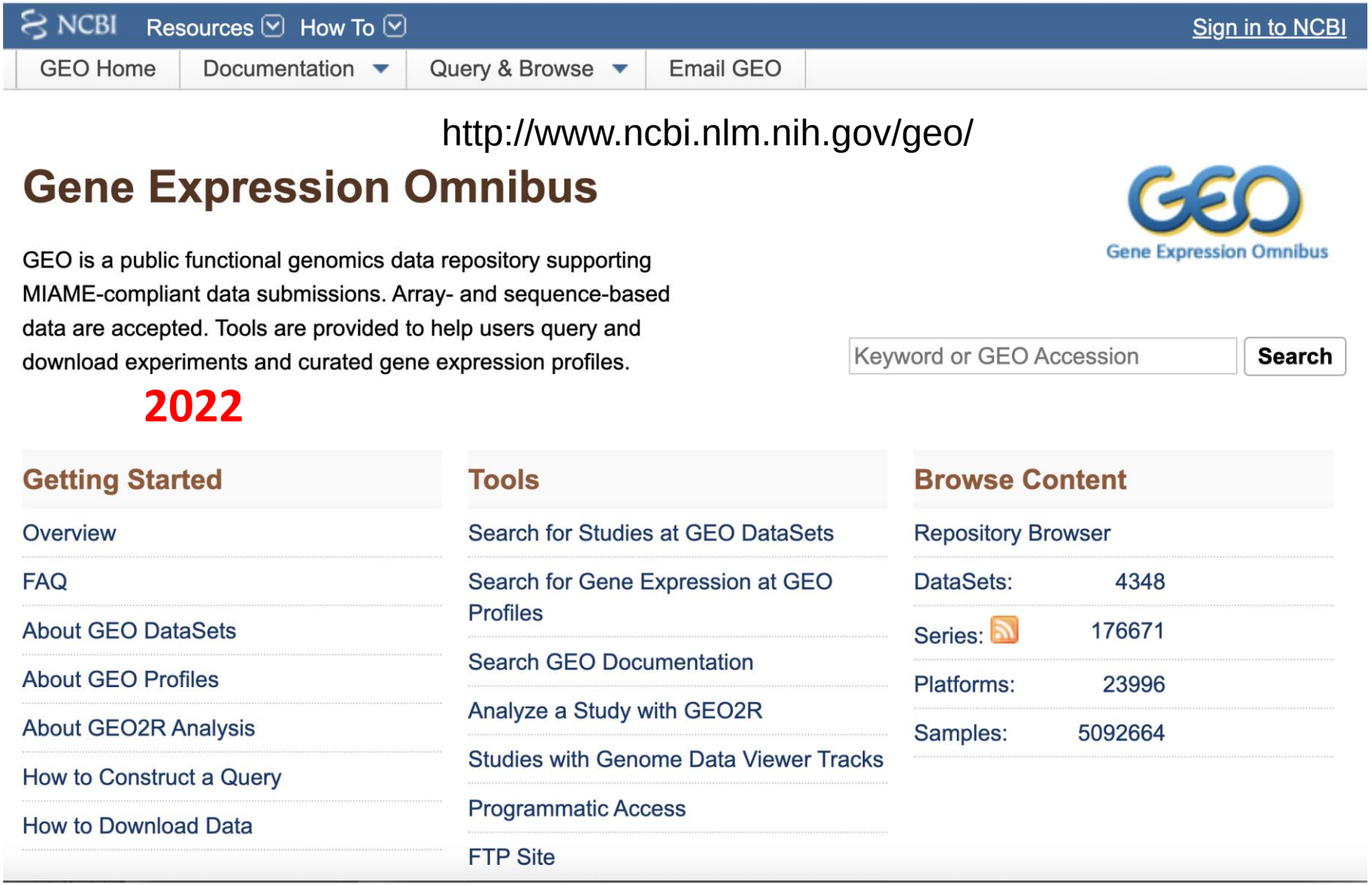
http://dcc.icgc.org/

cBio Portal for Cancer Genomics

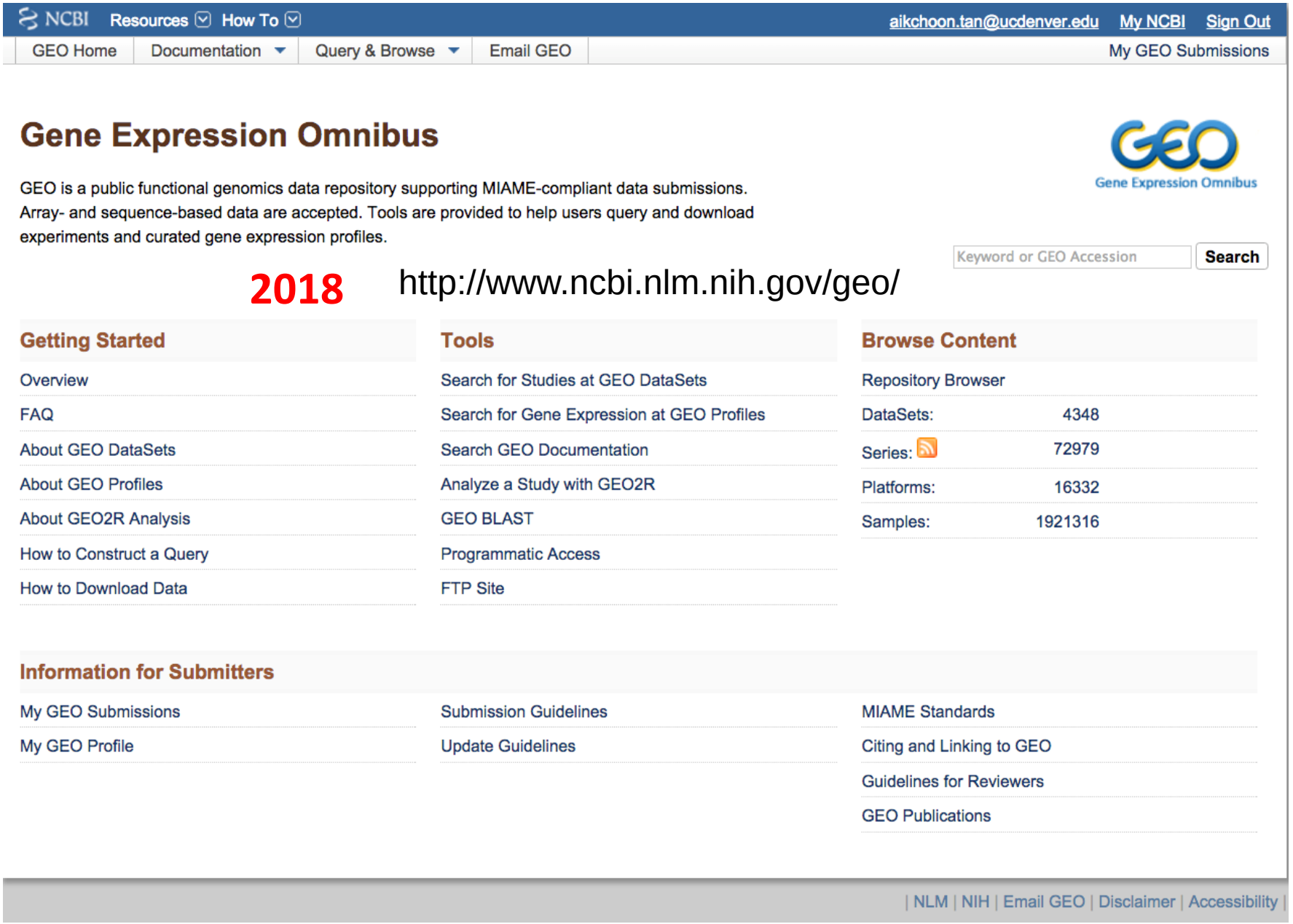


http://www.cbioportal.org

Gene Expression Omnibus (GEO)

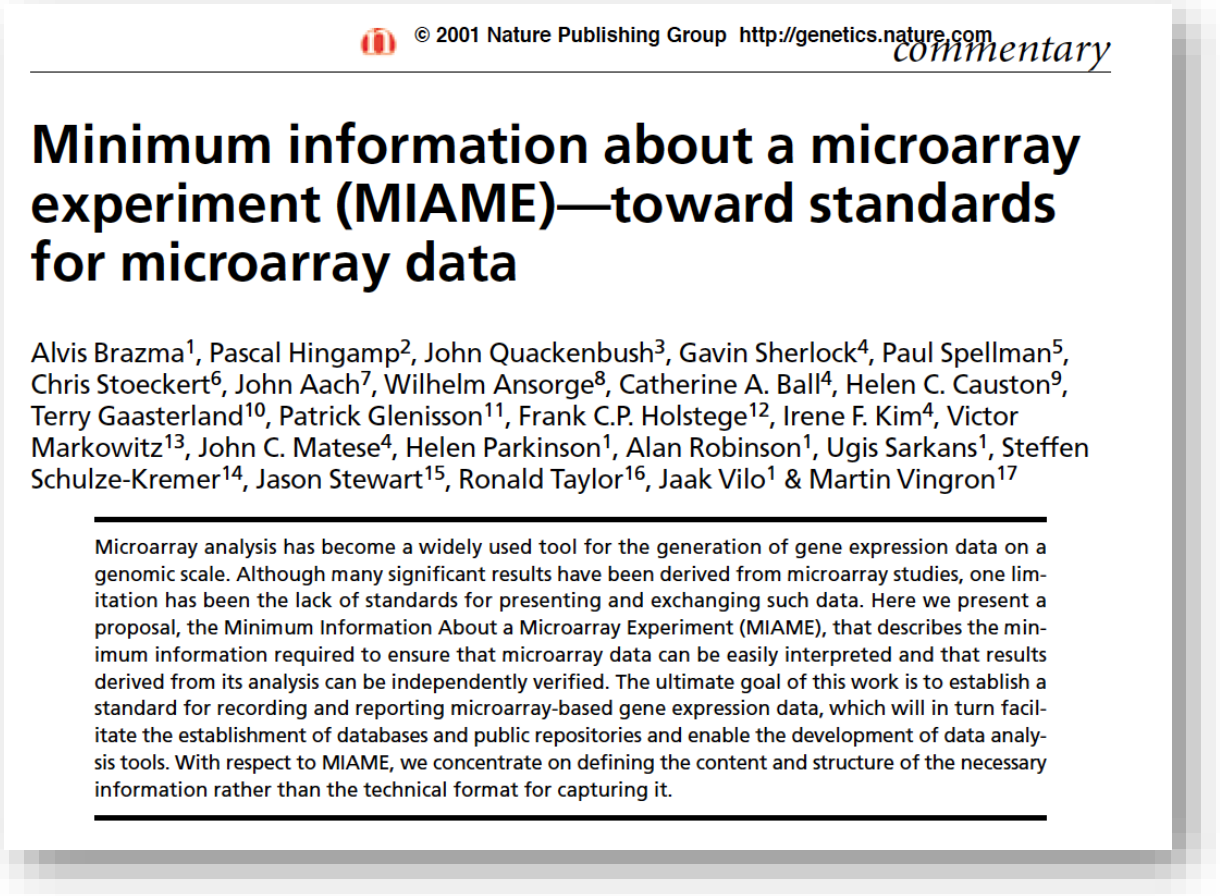


Gene Expression Omnibus (GEO)



MINiML format

MINiML/: This directory includes files in MINiML (MIAME Notation in Markup Language) format. MINiML is essentially an XML rendering of SOFT format, and the files provided here are the XML-equivalents of the Series and Platform family files provided in the SOFT/ directory.

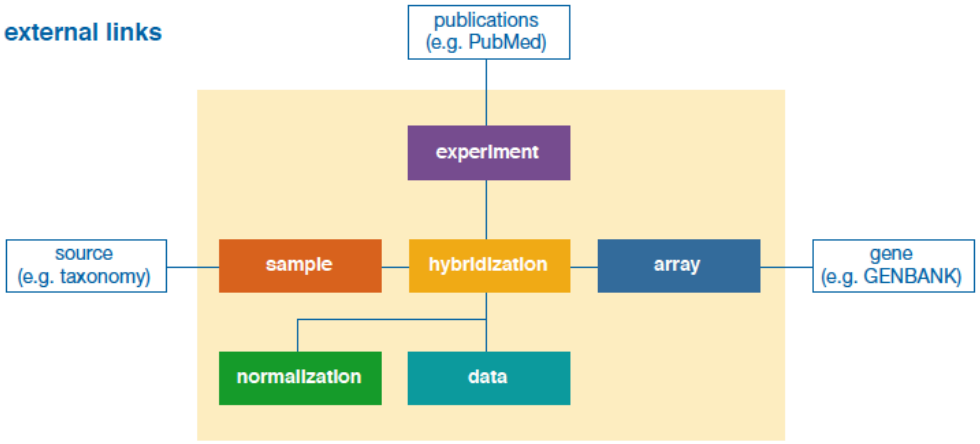


Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

Alvis Brazma[1], Pascal Hingamp[2], John Quackenbush[3], Gavin Sherlock[4], Paul Spellman[5], Chris Stoeckert[6], John Aach[7], Wilhelm Ansorge[8], Catherine A. Ball[4], Helen C. Causton[9], Terry Gaasterland[10], Patrick Glenisson[11], Frank C.P. Holstege[12], Irene F. Kim[4], Victor Markowitz[13], John C. Matese[4], Helen Parkinson[1], Alan Robinson[1], Ugis Sarkans[1], Steffen Schulze-Kremer[14], Jason Stewart[15], Ronald Taylor[16], Jaak Vilo[1] & Martin Vingron[17]

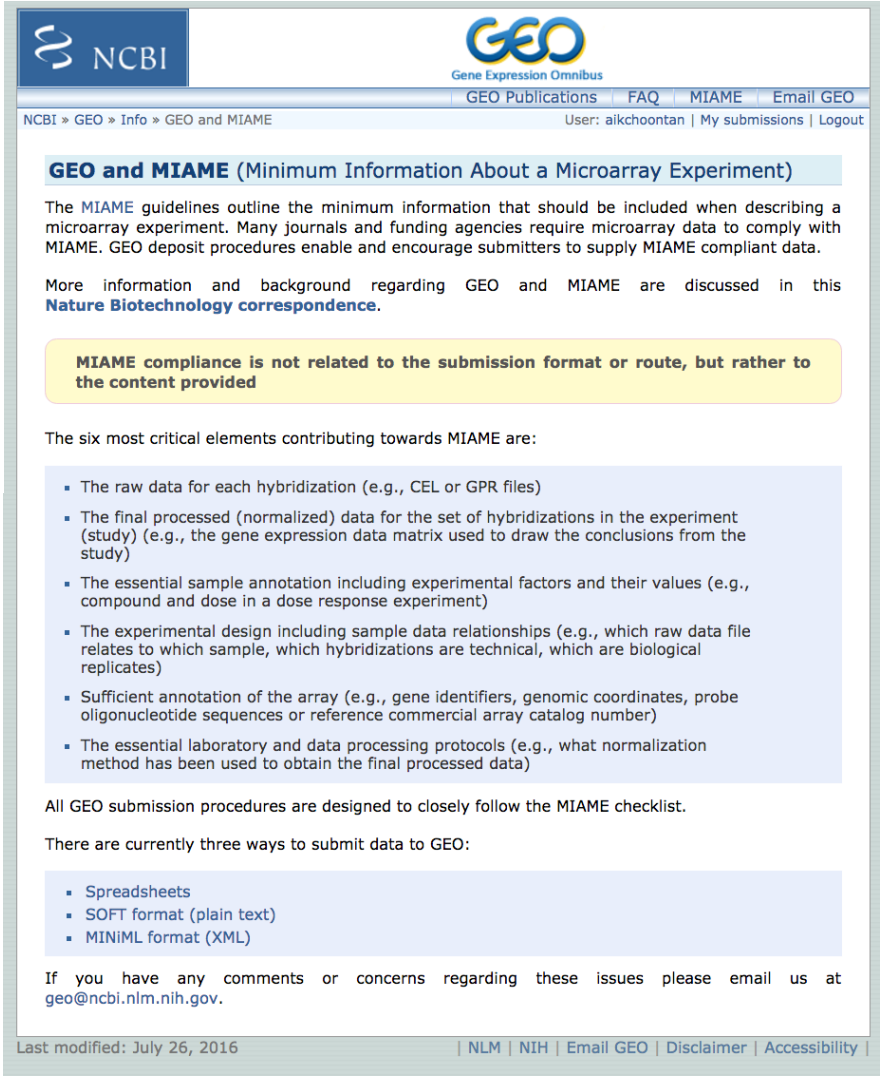
Microarray analysis has become a widely used tool for the generation of gene expression data on a genomic scale. Although many significant results have been derived from microarray studies, one limitation has been the lack of standards for presenting and exchanging such data. Here we present a proposal, the Minimum Information About a Microarray Experiment (MIAME), that describes the minimum information required to ensure that microarray data can be easily interpreted and that results derived from its analysis can be independently verified. The ultimate goal of this work is to establish a standard for recording and reporting microarray-based gene expression data, which will in turn facilitate the establishment of databases and public repositories and enable the development of data analysis tools. With respect to MIAME, we concentrate on defining the content and structure of the necessary information rather than the technical format for capturing it.

Gene Expression Omnibus (GEO)



Six Parts of MIAME

1. Experimental design: the set of hybridization experiments as a whole
2. Array design: each array used and each element (spot, feature) on the array
3. Samples: samples used, extract preparation and labeling
4. Hybridizations: procedures and parameters
5. Measurements: images, quantification and specifications
6. Normalization controls: types, values and specifications



NCBI

Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » Info » GEO and MIAME

User: aikchoontan | My submissions | Logout

GEO and MIAME (Minimum Information About a Microarray Experiment)

The MIAME guidelines outline the minimum information that should be included when describing a microarray experiment. Many journals and funding agencies require microarray data to comply with MIAME. GEO deposit procedures enable and encourage submitters to supply MIAME compliant data.

More information and background regarding GEO and MIAME are discussed in this **Nature Biotechnology correspondence**.

MIAME compliance is not related to the submission format or route, but rather to the content provided

The six most critical elements contributing towards MIAME are:

- The raw data for each hybridization (e.g., CEL or GPR files)
- The final processed (normalized) data for the set of hybridizations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
- The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
- The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridizations are technical, which are biological replicates)
- Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
- The essential laboratory and data processing protocols (e.g., what normalization method has been used to obtain the final processed data)

All GEO submission procedures are designed to closely follow the MIAME checklist.

There are currently three ways to submit data to GEO:

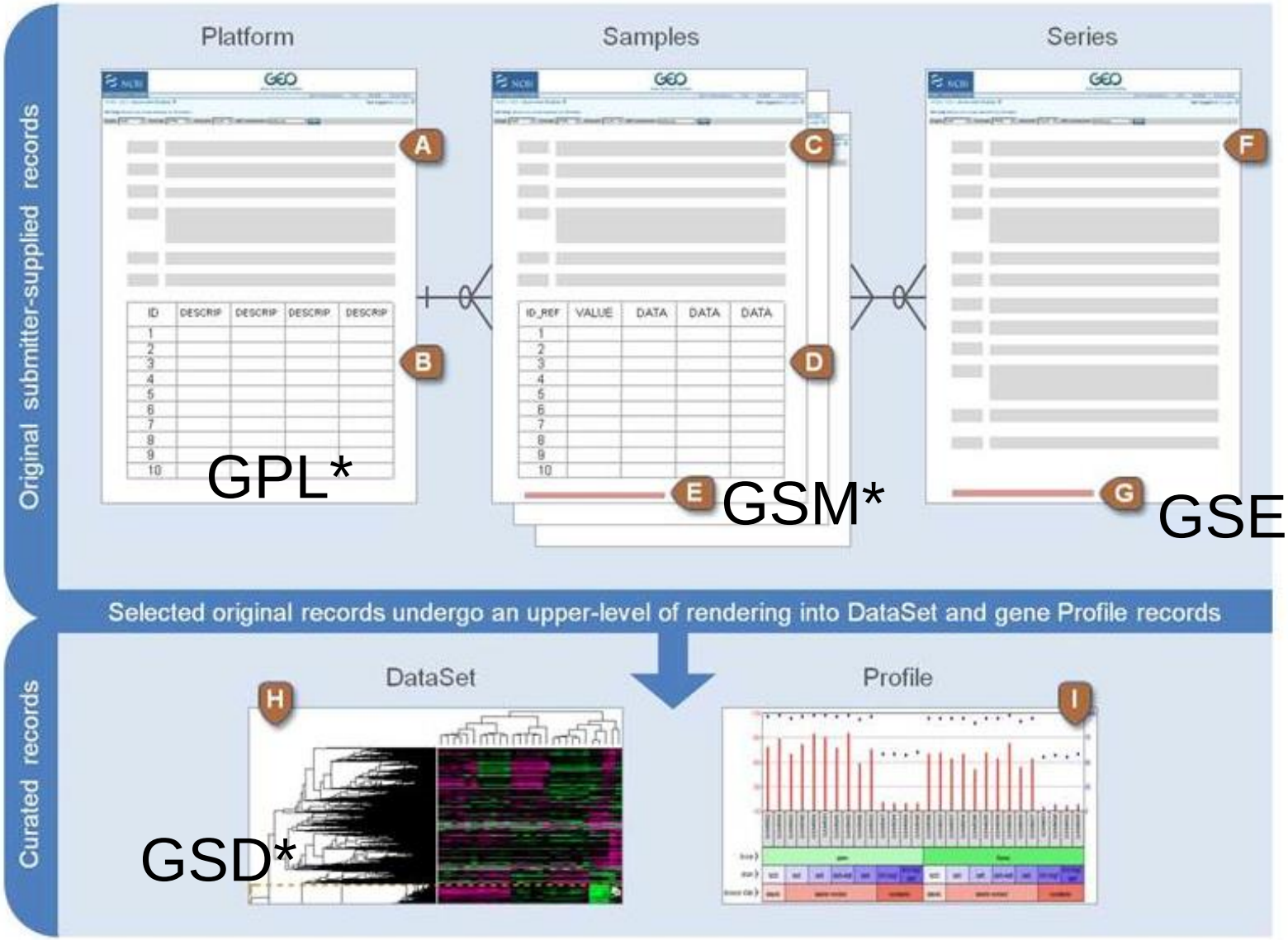
- Spreadsheets
- SOFT format (plain text)
- MINiML format (XML)

If you have any comments or concerns regarding these issues please email us at geo@ncbi.nlm.nih.gov.

Last modified: July 26, 2016

| NLM | NIH | Email GEO | Disclaimer | Accessibility |

NCBI GEO Data Formats



Drug Repurposing: The Impact of Big Data

RESEARCH ARTICLE

DRUG DISCOVERY

Discovery and Preclinical Validation of Drug Indications Using Compendia of Public Gene Expression Data

Marina Sirota,[1,2,3]* Joel T. Dudley,[1,2,3]* Jeewon Kim,[4] Annie P. Chiang,[1,2,3] Alex A. Morgan,[1,2,3] Alejandro Sweet-Cordero,[1,5] Julien Sage,[1,5,6] Atul J. Butte[1,3,5]†

Published 17 August 2011; revised 28 September 2011

The application of established drug compounds to new therapeutic indications, known as drug repositioning, offers several advantages over traditional drug development, including reduced development costs and shorter paths to approval. Recent approaches to drug repositioning use high-throughput experimental approaches to assess a compound's potential therapeutic qualities. Here, we present a systematic computational approach to predict novel therapeutic indications on the basis of comprehensive testing of molecular signatures in drug-disease pairs. We integrated gene expression measurements from 100 diseases and gene expression measurements on 164 drug compounds, yielding predicted therapeutic potentials for these drugs. We recovered many known drug and disease relationships using computationally derived therapeutic potentials and also predict many new indications for these 164 drugs. We experimentally validated a prediction for the antiulcer drug cimetidine as a candidate therapeutic in the treatment of lung adenocarcinoma, and demonstrate its efficacy both in vitro and in vivo using mouse xenograft models. This computational method provides a systematic approach for repositioning established drugs to treat a wide range of human diseases.

RESEARCH ARTICLE

DRUG DISCOVERY

Computational Repositioning of the Anticonvulsant Topiramate for Inflammatory Bowel Disease

Joel T. Dudley,[1,2,3]* Marina Sirota,[1,2,3]* Mohan Shenoy,[4] Reetesh K. Pai,[5] Silke Roedder,[1,3] Annie P. Chiang,[1,2,3] Alex A. Morgan,[1,2,3] Minnie M. Sarwal,[1,3] Pankaj Jay Pasricha,[4] Atul J. Butte[1,3]†

Inflammatory bowel disease (IBD) is a chronic inflammatory disorder of the gastrointestinal tract for which there are few safe and effective therapeutic options for long-term treatment and disease maintenance. Here, we applied a computational approach to discover new drug therapies for IBD in silico, using publicly available molecular data reporting gene expression in IBD samples and 164 small-molecule drug compounds. Among the top compounds predicted to be therapeutic for IBD by our approach were prednisolone, a corticosteroid used to treat IBD, and topiramate, an anticonvulsant drug not previously described to have efficacy for IBD or any related disorders of inflammation or the gastrointestinal tract. Using a trinitrobenzenesulfonic acid (TNBS)–induced rodent model of IBD, we experimentally validated our topiramate prediction in vivo. Oral administration of topiramate significantly reduced gross pathological signs and microscopic damage in primary affected colon tissue in the TNBS-induced rodent model of IBD. These findings suggest that topiramate might serve as a therapeutic option for IBD in humans and support the use of public molecular data and computational approaches to discover new therapeutic options for disease.

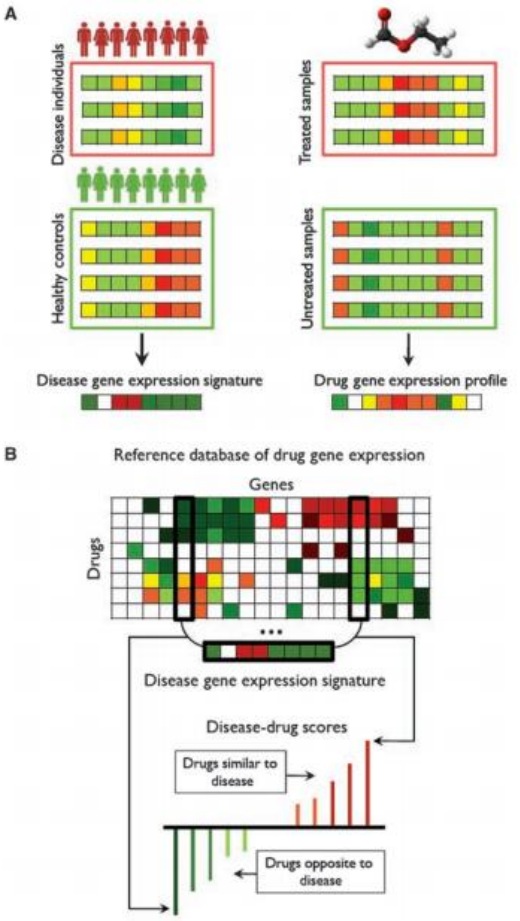


Fig. 1. Analytic workflow. (**A**) Two gene expression collections are used: a set of disease-associated gene expression data with corresponding controls and a set of gene expression data from tissue treated with drugs and small molecules with corresponding controls. SAM is used to obtain a signature of significantly up- and down-regulated genes for each disease. Rank normalization and the preprocessing procedure previously described (*25*) are used to create a reference database of drug gene expression. (**B**) A modification to the Connectivity Map method (*25*) is used to query the disease signature against the drug reference expression set to assign a drug-disease score to each drug-disease pair based on profile similarity. These scores are interpreted, resulting in a list of candidate therapeutics for each disease of interest.

The Connectivity Map Project

The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease

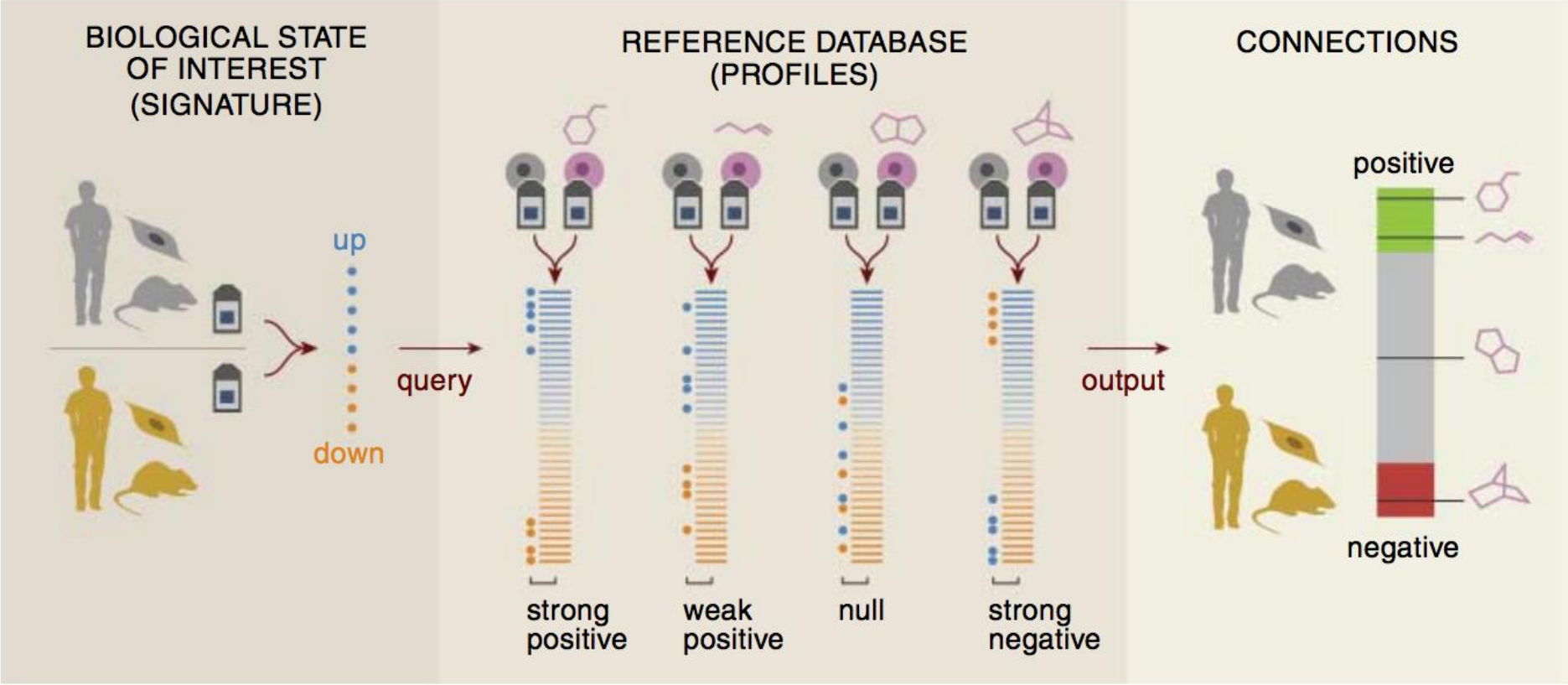
Justin Lamb,[1*] Emily D. Crawford,[1†] David Peck,[1] Joshua W. Modell,[1] Irene C. Blat,[1] Matthew J. Wrobel,[1] Jim Lerner,[1] Jean-Philippe Brunet,[1] Aravind Subramanian,[1] Kenneth N. Ross,[1] Michael Reich,[1] Haley Hieronymus,[1,2] Guo Wei,[1,2] Scott A. Armstrong,[2,3] Stephen J. Haggarty,[1,4] Paul A. Clemons,[1] Ru Wei,[1] Steven A. Carr,[1] Eric S. Lander,[1,5,6] Todd R. Golub[1,2,3,5,7*]

To pursue a systematic approach to the discovery of functional connections among diseases, genetic perturbation, and drug action, we have created the first installment of a reference collection of gene-expression profiles from cultured human cells treated with bioactive small molecules, together with pattern-matching software to mine these data. We demonstrate that this "Connectivity Map" resource can be used to find connections among small molecules sharing a mechanism of action, chemicals and physiological processes, and diseases and drugs. These results indicate the feasibility of the approach and suggest the value of a large-scale community Connectivity Map project.

Science 2006

http://science.sciencemag.org/content/sci/313/5795/1929.full.pdf

The Connectivity Map Project



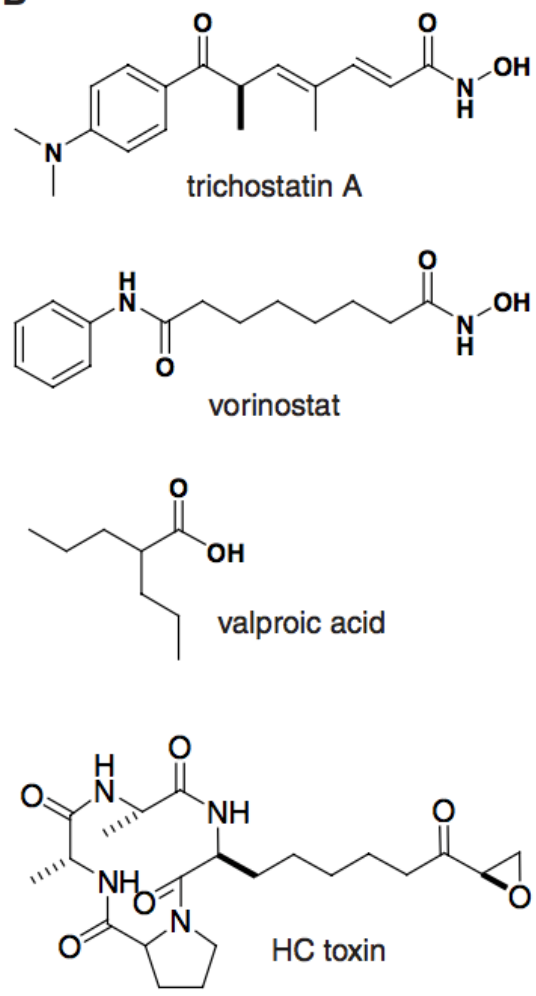
The Connectivity Map Project

Fig. 2. HDAC Inhibitors. (**A**) HDAC inhibitors are highly ranked with an external HDAC inhibitor signature. The "bar-view" is constructed from 453 horizontal lines, each representing an individual treatment instance, ordered by their corresponding connectivity scores with the Glaser *et al.* (*14*) signature (+1, top; −1, bottom). All valproic acid ($n = 18$), trichostatin A ($n = 12$), vorinostat ($n = 2$), and HC toxin ($n = 1$) instances in the data set are colored in black. Colors applied to the remaining instances reflect the sign of their scores (green, positive; gray, null; red, negative). The rank, name [instance id], concentration, cell line, and connectivity score for each of the selected HDAC inhibitor instances is shown. Unabridged results from this query are provided as Result S1. (**B**) Chemical structures.

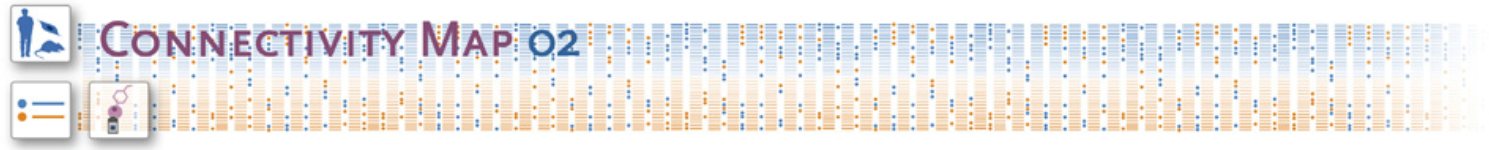
A

rank	perturbagen	dose	cell	score
1	vorinostat [1000]	10 μM	MCF7	1
2	trichostatin A [873]	1 μM	MCF7	0.969
3	trichostatin A [992]	100 nM	MCF7	0.931
4	trichostatin A [1050]	100 nM	MCF7	0.929
5	vorinostat [1058]	10 μM	MCF7	0.917
6	trichostatin A [981]	1 μM	MCF7	0.915
7	HC toxin [909]	100 nM	MCF7	0.914
8	trichostatin A [1112]	100 nM	MCF7	0.908
9	trichostatin A [1072]	1 μM	MCF7	0.906
10	trichostatin A [1014]	1 μM	MCF7	0.893
11	trichostatin A [332]	100 nM	MCF7	0.882
12	trichostatin A [331]	100 nM	MCF7	0.846
13	trichostatin A [448]	100 nM	PC3	0.788
14	valproic acid [345]	10 mM	MCF7	0.743
15	valproic acid [23]	1 mM	MCF7	0.735
16	valproic acid [1047]	1 mM	MCF7	0.733
17	trichostatin A [413]	100 nM	ssMCF7	0.725
18	valproic acid [410]	10 mM	HL60	0.725
19	valproic acid [458]	1 mM	PC3	0.680
33	valproic acid [409]	1 mM	HL60	0.634
39	valproic acid [1020]	500 μM	MCF7	0.619
52	valproic acid [346]	2 mM	MCF7	0.582
61	valproic acid [1078]	500 μM	MCF7	0.563
71	valproic acid [629]	1 mM	SKMEL5	0.539
72	valproic acid [347]	500 μM	MCF7	0.539
73	valproic acid [989]	1 mM	MCF7	0.538
76	valproic acid [433]	1 mM	PC3	0.528
89	trichostatin A [364]	100 nM	HL60	0.507
92	valproic acid [497]	1 mM	ssMCF7	0.501
297	valproic acid [348]	50 μM	MCF7	0
388	valproic acid [994]	200 μM	MCF7	0
403	valproic acid [1002]	50 μM	MCF7	0
419	valproic acid [1060]	50 μM	MCF7	-0.537

B



The Connectivity Map Project



username:

password:

sign in

email me my password | register as a new user

The Connectivity Map (also known as cmap) is a collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules and simple pattern-matching algorithms that together enable the discovery of functional connections between drugs, genes and diseases through the transitory feature of common gene-expression changes. You can learn more about cmap from our papers in *Science* and *Nature Reviews Cancer*.

This web interface provides access to the current version (**build 02**) of Connectivity Map which contains more than 7,000 expression profiles representing 1,309 compounds. It is designed to allow biologists, pharmacologists, chemists and clinical scientists to use cmap without the need for any specialist ability in the analysis of gene-expression data. The previous version (**build 01**) of Connectivity Map can be accessed here.

A brief tutorial can be found by clicking 'getting started' under the 'help' tab after log in. Detailed help and a definition of cmap terms can be found by clicking 'topics', also under the 'help' tab. For everything else, please contact us.

The Connectivity Map is based at The Broad Institute of MIT and Harvard in Cambridge, Massachusetts. The cmap team is Justin Lamb, Xiaodong Lu, Dave Peck, Matt Wrobel, Aravind Subramanian, Irene Blat, Josh Modell, Jim Lerner, Elizabeth Liu and Emily Crawford. Jean-Philippe Brunet, Ken Ross, Michael Reich, Paul Clemons, Kathy Seiler, Steve Haggarty, Bang Wong, Maria Nemchuk, Ru Wei, Steve Carr, Christopher Johnson, Stephen Johnson, the MSigDB curation team, and the Genetic Analysis Platform contribute invaluable expertise and assistance. Todd Golub and Eric Lander provide institutional leadership for the project.

privacy statement | terms and conditions

BROAD INSTITUTE *The Broad Institute is a research collaboration of MIT, Harvard and its affiliated Hospitals, and the Whitehead Institute, created to bring the power of genomics to medicine.*

The NIH LINCS Program

(www.lincsproject.org)

- LINCS (Library of Integrated Network-based Cellular Signatures) Program
- LINCS aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents

The NIH LINCS Project

https://clue.io

CLUE

Tools Projects Partnering | Log in

ConnectivityMap

BROAD INSTITUTE

Unravel biology with the world's largest perturbation-driven gene expression dataset.

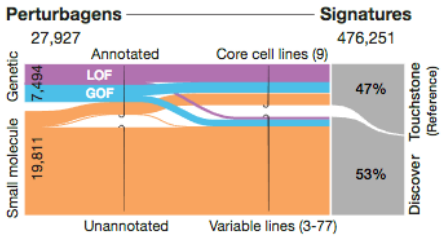
> TYPE COMPOUND, GENE, MoA, OR PERTURBAGEN CLASS TO SEE OVERVIEW
> TYPE A SLASH CHARACTER "/" TO SEE LIST OF COMMANDS

DATA VERSION: 1.0.1.1 / SOFTWARE VERSION: 1.1.1.15

Data and Tools

The CMap dataset of cellular signatures catalogs transcriptional responses of human cells to chemical and genetic perturbation. Here you can find the 1.3M L1000 profiles and the tools for their analysis.

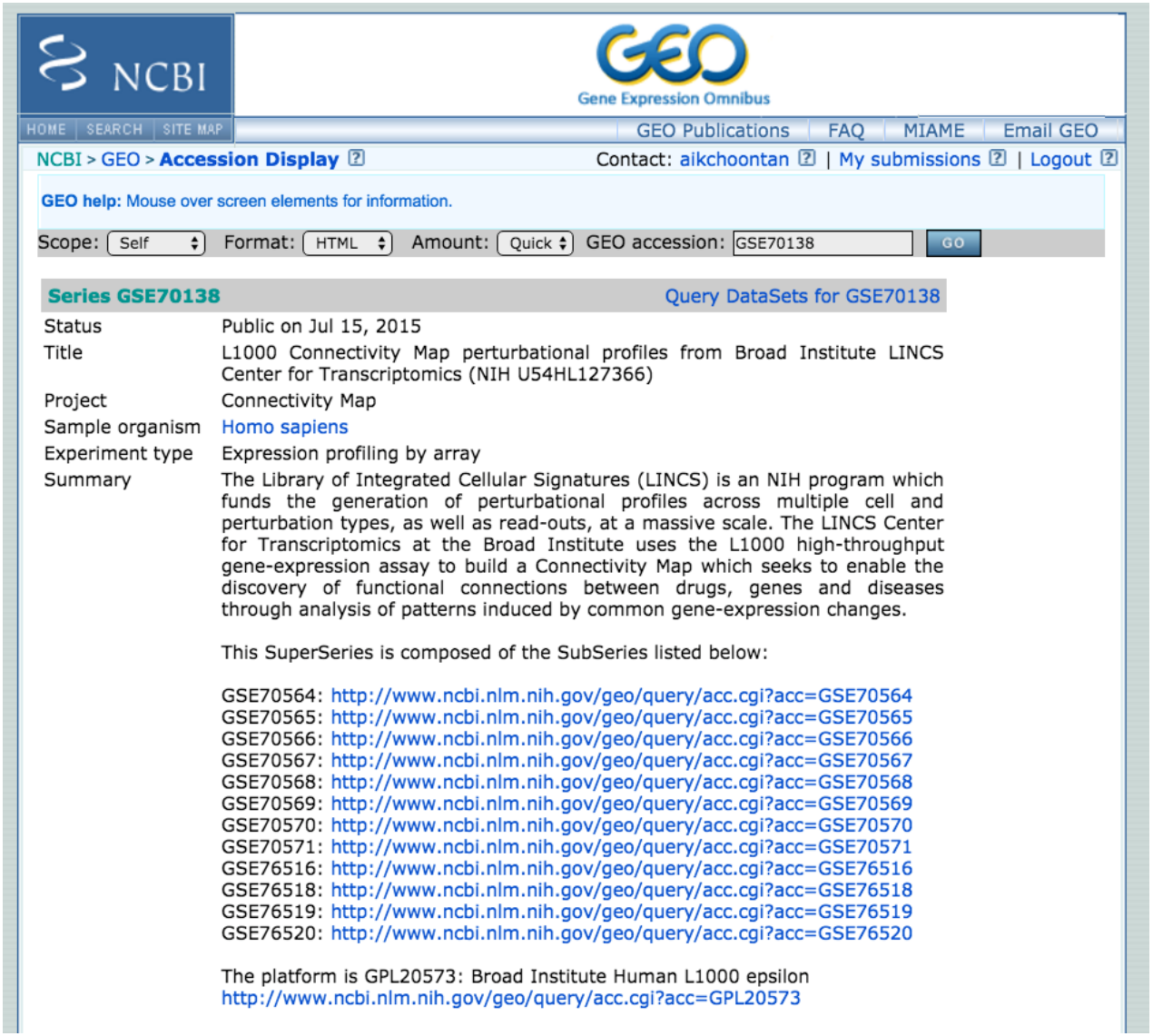
A total of 27,927 perturbagens have been profiled to produce 476,251 expression signatures. About half of those signatures make up the Touchstone (reference) dataset generated from testing well-annotated genetic and small-molecular perturbagens in a core panel of cell lines. The remainder make up the Discover dataset, generated from profiling uncharacterized small molecules in a variable number of cell lines.



Start exploring the data by using the text-box on this page to look up perturbagens of interest in Touchstone. To see the suite of tools, including apps to query your gene expression signatures and analyze resulting connections, click on Tools in the menu bar.

27,927 perturbagens
476,251 expression signatures

The NIH LINCS Project
(Data available in NCBI GEO)



NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display** Contact: aikchoontan | My submissions | Logout

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE70138 GO

Series GSE70138 Query DataSets for GSE70138

Status	Public on Jul 15, 2015
Title	L1000 Connectivity Map perturbational profiles from Broad Institute LINCS Center for Transcriptomics (NIH U54HL127366)
Project	Connectivity Map
Sample organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	The Library of Integrated Cellular Signatures (LINCS) is an NIH program which funds the generation of perturbational profiles across multiple cell and perturbation types, as well as read-outs, at a massive scale. The LINCS Center for Transcriptomics at the Broad Institute uses the L1000 high-throughput gene-expression assay to build a Connectivity Map which seeks to enable the discovery of functional connections between drugs, genes and diseases through analysis of patterns induced by common gene-expression changes.

This SuperSeries is composed of the SubSeries listed below:

GSE70564: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70564
GSE70565: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70565
GSE70566: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70566
GSE70567: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70567
GSE70568: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70568
GSE70569: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70569
GSE70570: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70570
GSE70571: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70571
GSE76516: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76516
GSE76518: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76518
GSE76519: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76519
GSE76520: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76520

The platform is GPL20573: Broad Institute Human L1000 epsilon
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL20573

The NIH LINCS Project
(Data available in NCBI GEO)

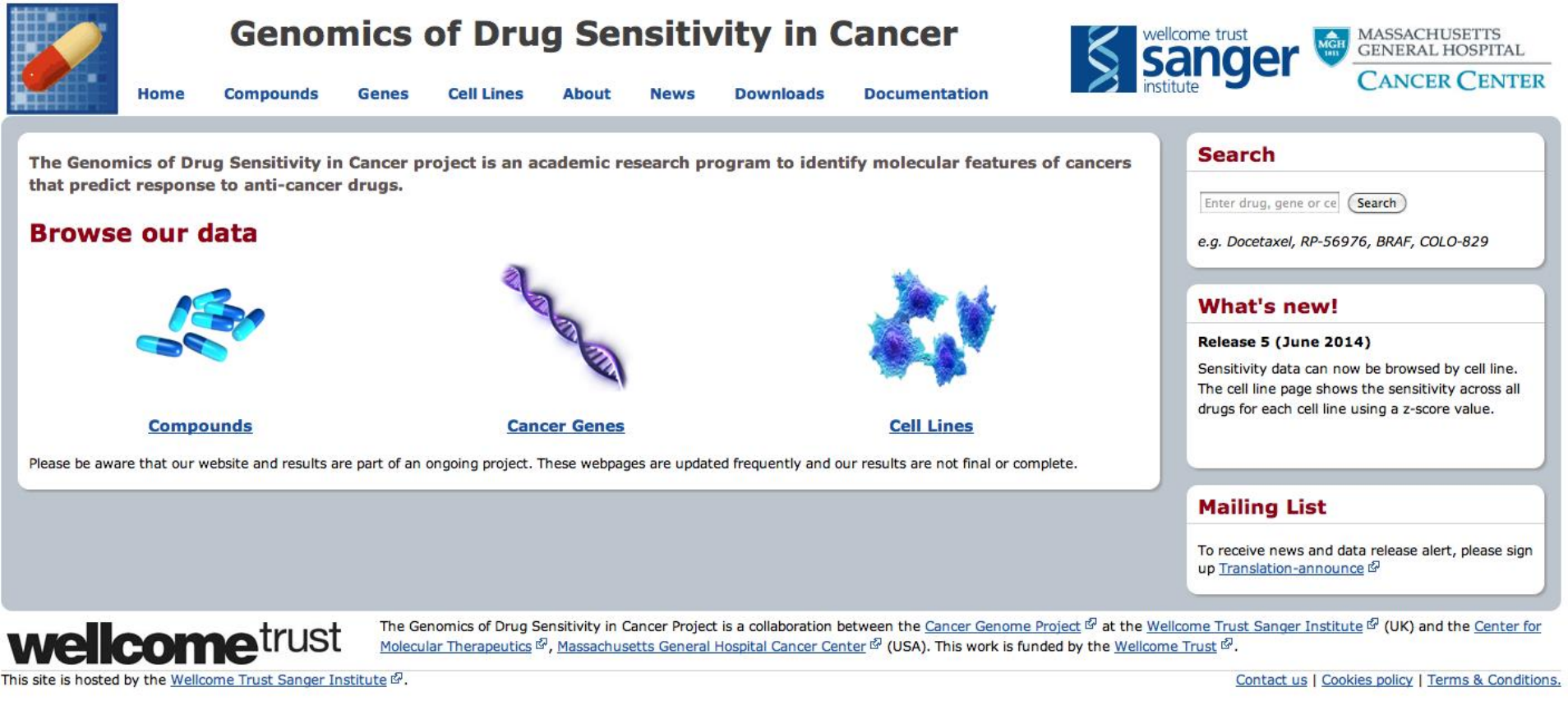
Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

Supplementary file	Size	Download	File type/resource
GSE70138_Broad_LINCS_Level2_GEX_n115209x978_2015-12-31.gct.gz	208.6 Mb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level2_GEX_n78980x978_2015-06-30.gct.gz	144.4 Mb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level3_INF_mlr12k_n115209x22268_2015-12-31.gct.gz	6.2 Gb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level3_INF_mlr12k_n78980x22268_2015-06-30.gct.gz	4.3 Gb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level4_ZSPCINF_mlr12k_n115209x22268_2015-12-31.gct.gz	6.6 Gb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level4_ZSPCINF_mlr12k_n78980x22268_2015-06-30.gct.gz	4.5 Gb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level4_ZSVCINF_mlr12k_n115209x22268_2015-12-31.gct.gz	6.7 Gb	(ftp)(http)	GCT
GSE70138_Broad_LINCS_Level4_ZSVCINF_mlr12k_n78980x22268_2015-06-30.gct.gz	4.6 Gb	(ftp)(http)	GCT
GSE70138_GEO_CMap_LINCS_User_Guide_v1_1.pdf	135.1 Kb	(ftp)(http)	PDF

Raw data provided as supplementary file

Processed data included within Sample table

Genomics of Drug Sensitivity in Cancer (GDSC COSMIC)



http://www.cancerrxgene.org/

ARTICLE

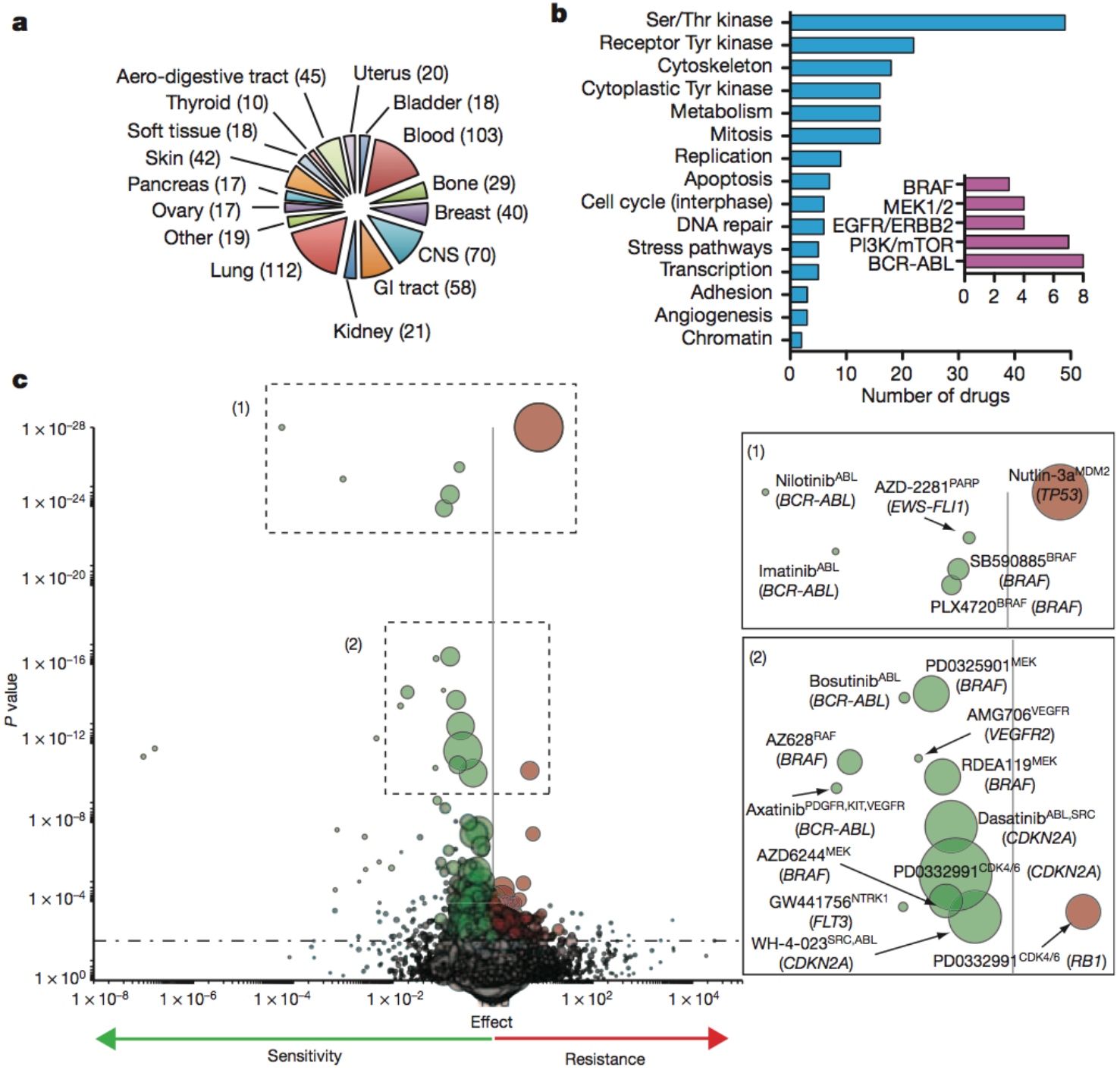
doi:10.1038/nature11005

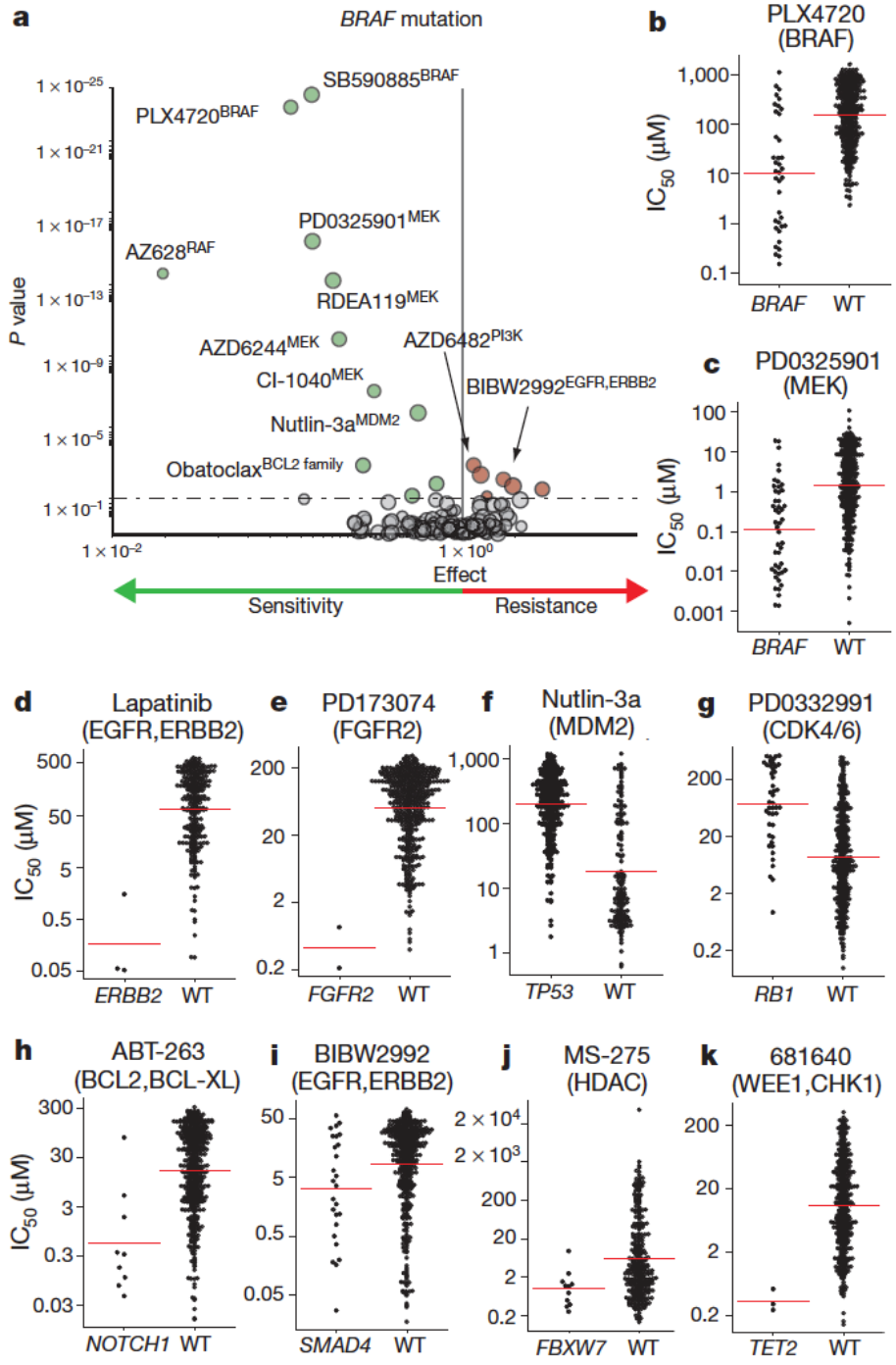
Systematic identification of genomic markers of drug sensitivity in cancer cells

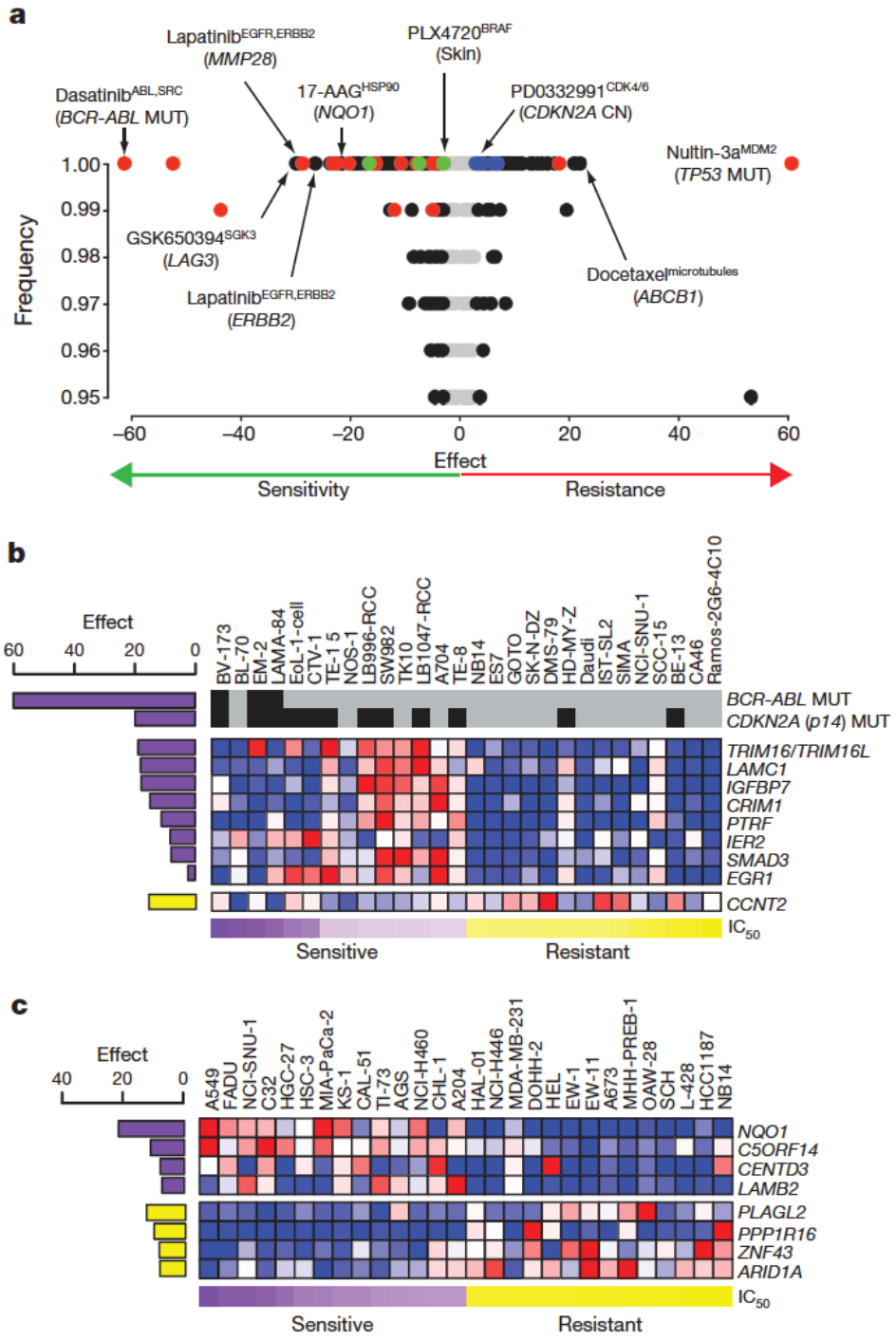
Mathew J. Garnett[1]*, Elena J. Edelman[2]*, Sonja J. Heidorn[1]*, Chris D. Greenman[1]†, Anahita Dastur[2], King Wai Lau[1], Patricia Greninger[2], I. Richard Thompson[1], Xi Luo[2], Jorge Soares[1], Qingsong Liu[3,4], Francesco Iorio[1,5], Didier Surdez[6], Li Chen[2], Randy J. Milano[2], Graham R. Bignell[1], Ah T. Tam[2], Helen Davies[1], Jesse A. Stevenson[2], Syd Barthorpe[1], Stephen R. Lutz[2], Fiona Kogera[1], Karl Lawrence[1], Anne McLaren-Douglas[1], Xeni Mitropoulos[2], Tatiana Mironenko[1], Helen Thi[2], Laura Richardson[1], Wenjun Zhou[3,4], Frances Jewitt[1], Tinghu Zhang[3,4], Patrick O'Brien[1], Jessica L. Boisvert[2], Stacey Price[1], Wooyoung Hur[3,4], Wanjuan Yang[1], Xianming Deng[3,4], Adam Butler[1], Hwan Geun Choi[3,4], Jae Won Chang[3,4], Jose Baselga[2], Ivan Stamenkovic[7], Jeffrey A. Engelman[2], Sreenath V. Sharma[2]†, Olivier Delattre[6], Julio Saez-Rodriguez[5], Nathanael S. Gray[3,4], Jeffrey Settleman[2], P. Andrew Futreal[1], Daniel A. Haber[2,8], Michael R. Stratton[1], Sridhar Ramaswamy[2], Ultan McDermott[1] & Cyril H. Benes[2]

639 cell lines treated with 130 drugs

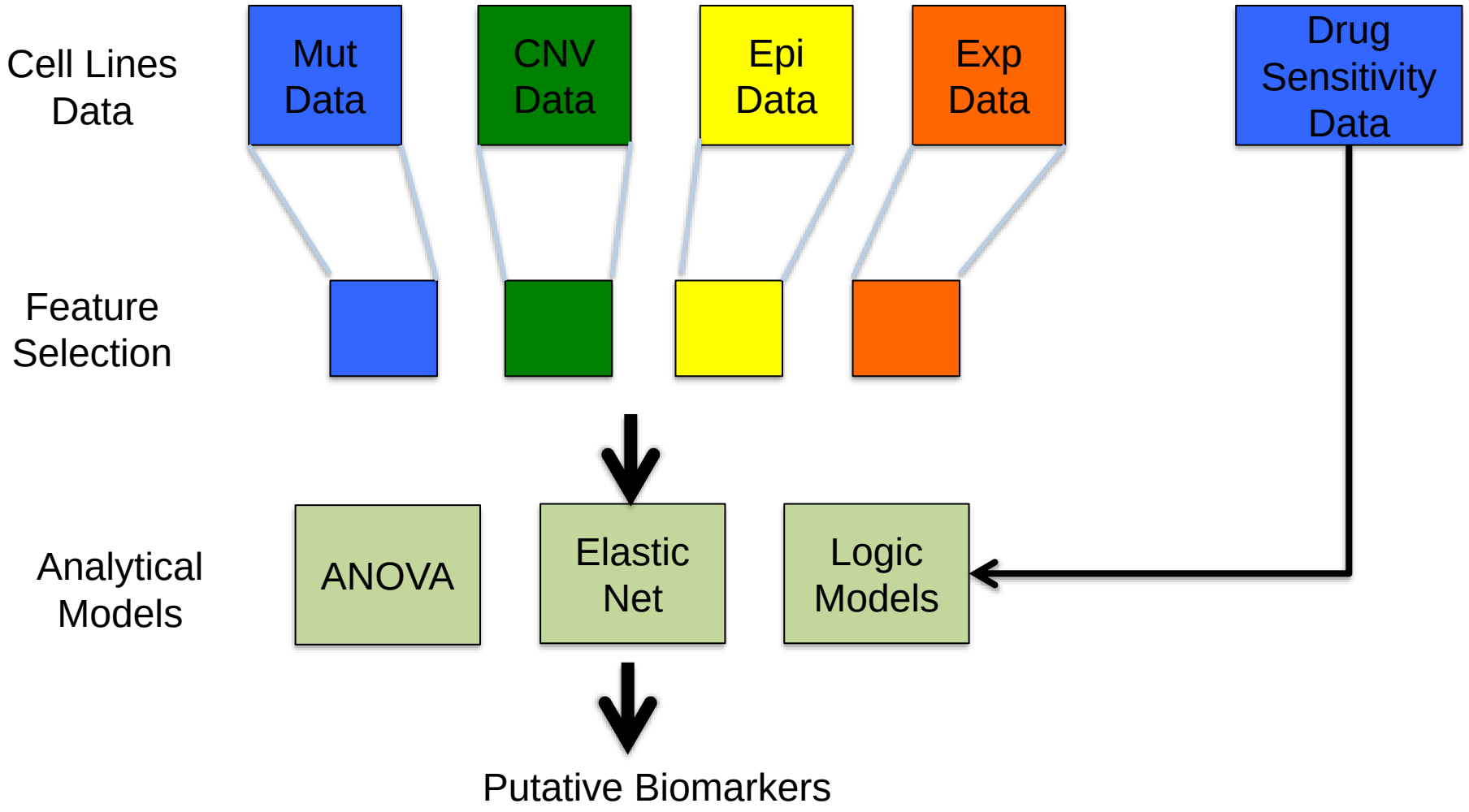
Clinical responses to anticancer therapies are often restricted to a subset of patients. In some cases, mutated cancer genes are potent biomarkers for responses to targeted agents. Here, to uncover new biomarkers of sensitivity and resistance to cancer therapeutics, we screened a panel of several hundred cancer cell lines—which represent much of the tissue-type and genetic diversity of human cancers—with 130 drugs under clinical and preclinical investigation. In aggregate, we found that mutated cancer genes were associated with cellular response to most currently available cancer drugs. Classic oncogene addiction paradigms were modified by additional tissue-specific or expression biomarkers, and some frequently mutated genes were associated with sensitivity to a broad range of therapeutic agents. Unexpected relationships were revealed, including the marked sensitivity of Ewing's sarcoma cells harbouring the *EWS* (also known as *EWSR1*)-*FLI1* gene translocation to poly(ADP-ribose) polymerase (PARP) inhibitors. By linking drug activity to the functional complexity of cancer genomes, systematic pharmacogenomic profiling in cancer cell lines provides a powerful biomarker discovery platform to guide rational cancer therapeutic strategies.


a
Aero-digestive tract (45)
Uterus (20)
Thyroid (10)
Bladder (18)
Soft tissue (18)
Blood (103)
Skin (42)
Pancreas (17)
Bone (29)
Ovary (17)
Breast (40)
Other (19)
CNS (70)
Lung (112)
GI tract (58)
Kidney (21)
b
Ser/Thr kinase
Receptor Tyr kinase
Cytoskeleton
Cytoplastic Tyr kinase
Metabolism
Mitosis
Replication
Apoptosis
Cell cycle (interphase)
DNA repair
Stress pathways
Transcription
Adhesion
Angiogenesis
Chromatin
BRAF
MEK1/2
EGFR/ERBB2
PI3K/mTOR
BCR-ABL
0 2 4 6 8
0 10 20 30 40 50
Number of drugs
c
(1)
(2)
P value
1 × 10−28
1 × 10−24
1 × 10−20
1 × 10−16
1 × 10−12
1 × 10−8
1 × 10−4
1 × 10^0
1 × 10−8 1 × 10−6 1 × 10−4 1 × 10−2 1 × 10^2 1 × 10^4
Effect
Sensitivity
Resistance
(1)
Nilotinib ABL (BCR-ABL)
AZD-2281 PARP (EWS-FLI1)
Nutlin-3a MDM2 (TP53)
Imatinib ABL (BCR-ABL)
SB590885 BRAF (BRAF)
PLX4720 BRAF (BRAF)
(2)
Bosutinib ABL (BCR-ABL)
PD0325901 MEK (BRAF)
AMG706 VEGFR (VEGFR2)
AZ628 RAF (BRAF)
RDEA119 MEK (BRAF)
Axatinib PDGFR,KIT,VEGFR (BCR-ABL)
Dasatinib ABL,SRC (CDKN2A)
AZD6244 MEK (BRAF)
PD0332991 CDK4/6 (CDKN2A)
GW441756 NTRK1 (FLT3)
WH-4-023 SRC,ABL (CDKN2A)
PD0332991 CDK4/6 (RB1)


a
BRAF mutation
SB590885BRAF
PLX4720BRAF
PD0325901MEK
AZ628RAF
RDEA119MEK
AZD6244MEK
AZD6482PI3K
CI-1040MEK
BIBW2992EGFR,ERBB2
Nutlin-3aMDM2
ObatoclaxBCL2 family
P value
1 × 10−25
1 × 10−21
1 × 10−17
1 × 10−13
1 × 10−9
1 × 10−5
1 × 10−1
1 × 10−2
1 × 100
Effect
Sensitivity
Resistance
b
PLX4720 (BRAF)
IC50 (μM)
1,000
100
10
1
0.1
BRAF
WT
c
PD0325901 (MEK)
100
10
1
0.1
0.01
0.001
BRAF
WT
d
Lapatinib (EGFR,ERBB2)
500
50
5
0.5
0.05
ERBB2
WT
e
PD173074 (FGFR2)
200
20
2
0.2
FGFR2
WT
f
Nutlin-3a (MDM2)
1,000
100
10
1
TP53
WT
g
PD0332991 (CDK4/6)
200
20
2
0.2
RB1
WT
h
ABT-263 (BCL2,BCL-XL)
300
30
3
0.3
0.03
NOTCH1
WT
i
BIBW2992 (EGFR,ERBB2)
50
5
0.5
0.05
SMAD4
WT
j
MS-275 (HDAC)
2 × 104
2 × 103
200
20
2
0.2
FBXW7
WT
k
681640 (WEE1,CHK1)
200
20
2
0.2
TET2
WT


a
Lapatinib EGFR,ERBB2 (MMP28)
PLX4720 BRAF (Skin)
Dasatinib ABL,SRC (BCR-ABL MUT)
17-AAG HSP90 (NQO1)
PD0332991 CDK4/6 (CDKN2A CN)
Nultin-3a MDM2 (TP53 MUT)
GSK650394 SGK3 (LAG3)
Lapatinib EGFR,ERBB2 (ERBB2)
Docetaxel microtubules (ABCB1)
Frequency
1.00
0.99
0.98
0.97
0.96
0.95
−60
−40
−20
0
20
40
60
Effect
Sensitivity
Resistance
b
Effect
60
40
20
0
BV-173
BL-70
EM-2
LAMA-84
EoL-1-cell
CTV-1
TE-1 5
NOS-1
LB996-RCC
SW982
TK10
LB1047-RCC
A704
TE-8
NB14
ES7
GOTO
SK-N-DZ
DMS-79
HD-MY-Z
Daudi
IST-SL2
SIMA
NCI-SNU-1
SCC-15
BE-13
CA46
Ramos-2G6-4C10
BCR-ABL MUT
CDKN2A (p14) MUT
TRIM16/TRIM16L
LAMC1
IGFBP7
CRIM1
PTRF
IER2
SMAD3
EGR1
CCNT2
IC50
Sensitive
Resistant
c
Effect
40
20
0
A549
FADU
NCI-SNU-1
C32
HGC-27
HSC-3
MIA-PaCa-2
KS-1
CAL-51
TI-73
AGS
NCI-H460
CHL-1
A204
HAL-01
NCI-H446
MDA-MB-231
DOHH-2
HEL
EW-1
EW-11
A673
MHH-PREB-1
OAW-28
SCH
L-428
HCC1187
NB14
NQO1
C5ORF14
CENTD3
LAMB2
PLAGL2
PPP1R16
ZNF43
ARID1A
IC50
Sensitive
Resistant

Analytical Framework for Biomarkers Discovery in GDSC



Querying Drug in GDSC

Drug : Erlotinib

Targets	EGFR
PubCHEM	176870
Legacy data	Legacy data for Erlotinib

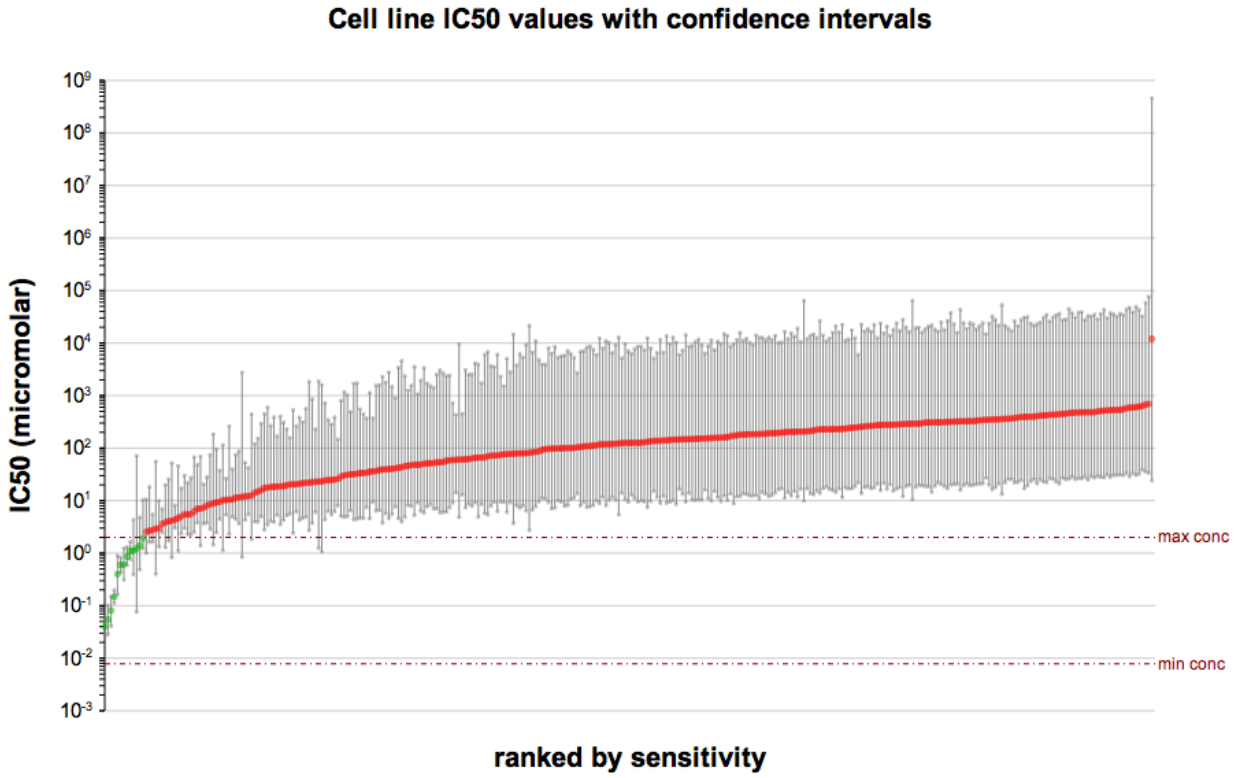
Help

- Description of statistical approaches
- Interpreting Volcano plots
- Interpreting EN heatmaps
- Detail of download files

Analysis of drug sensitivity data

Overview | Volcano plot | Volcano Data | Elastic net | Scatter plots | Download data

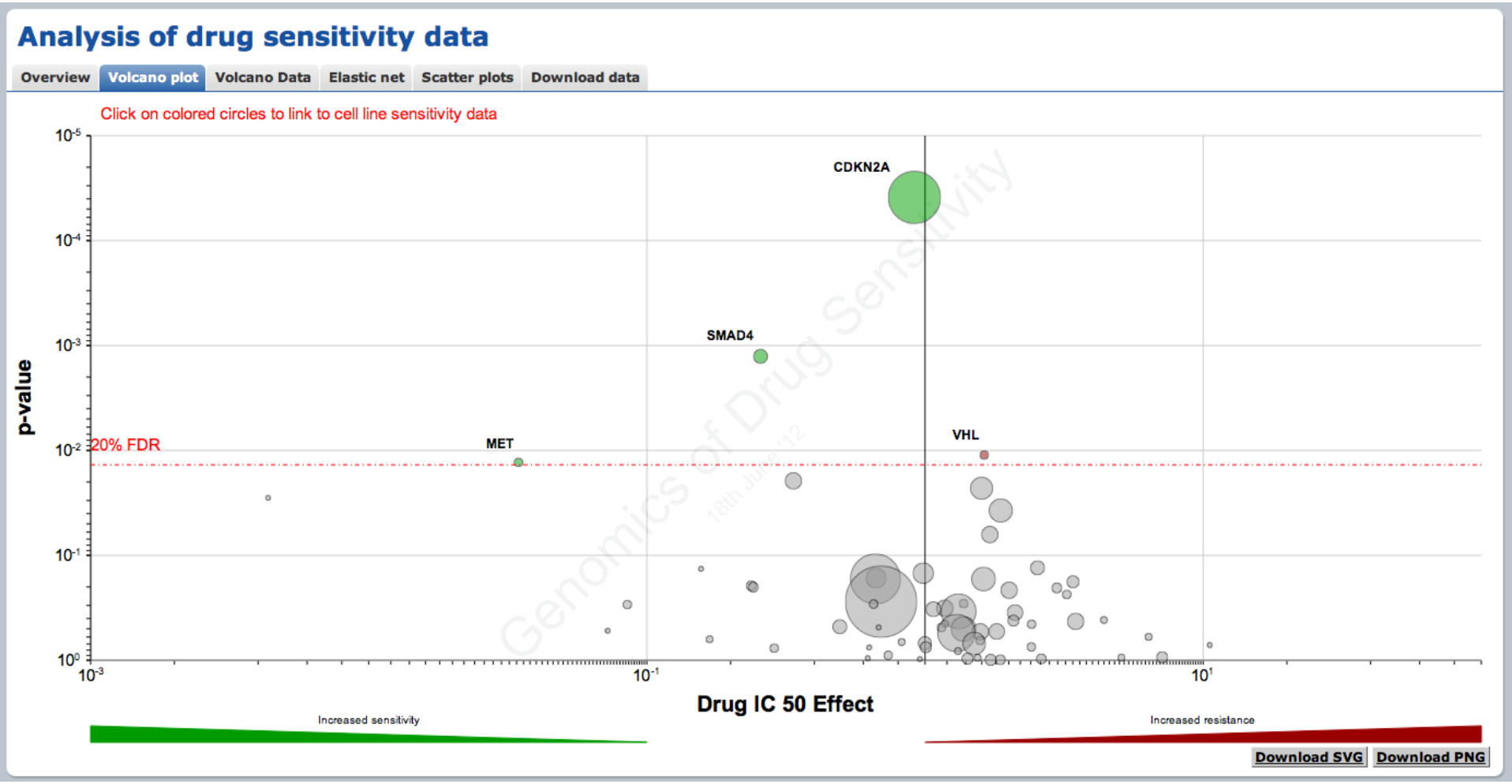
IC50 | IC90 | IC75 | IC25 | AUC



Number of cell lines screened	329
Maximum IC50	1.20e+4
Geometric mean	78.9
Minimum IC50	0.0400
Min screening concentration	0.00781
Max screening concentration	2.00

Download SVG | Download PNG

Volcano Plot



Querying Drug in GDSC

Analysis of drug sensitivity data

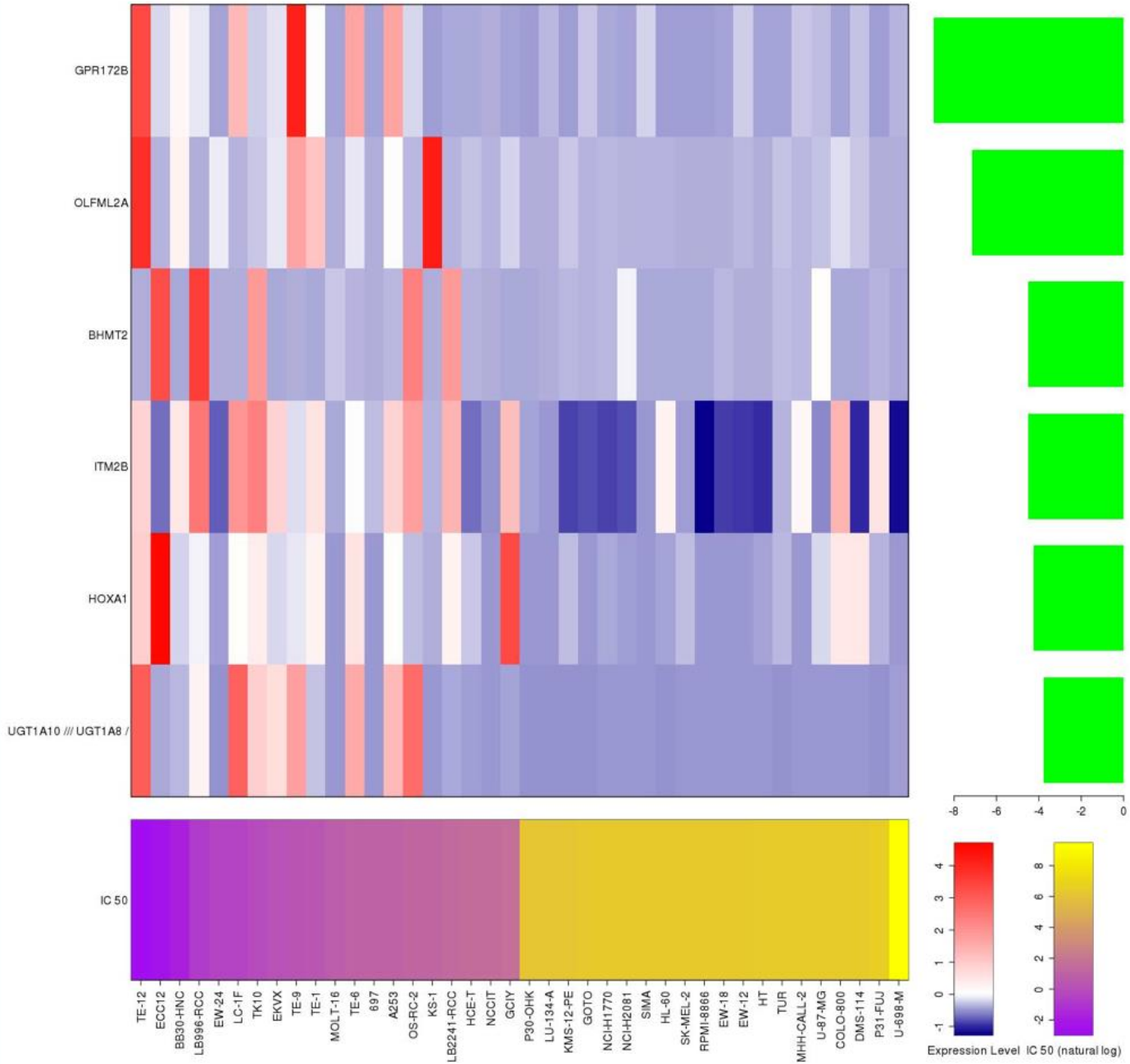
Overview | Volcano plot | Volcano Data | Elastic net | Scatter plots | Download data

Filter: 64 entries 1 2 3 25 per page CSV TAB XLS

Gene	Effect	P-value	No. of mutations
CDKN2A	0.916	0.0000387	112
SMAD4	0.256	0.00127	8
VHL	1.63	0.0110	3
MET	0.0345	0.0130	3
NF2	0.336	0.0195	11
NRAS	1.60	0.0229	20
WT1	0.00434	0.0283	1
KRAS	1.87	0.0373	22
CCND1	1.71	0.0633	11
MLH1	2.54	0.131	8
SETD2	0.157	0.134	1
KDM6A	0.985	0.148	17
PIK3CA	0.667	0.165	16
CDKN2a(p14)	0.663	0.168	102
BRAF	1.62	0.168	23
CDKN2C	3.41	0.178	6
BCR_ABL	0.237	0.195	4
MDM2	0.241	0.201	4
MSH2	2.98	0.204	4
MLLT3	2.01	0.215	11
FGFR2	3.23	0.236	3
TP53	0.695	0.276	209
SMARCA4	1.38	0.288	3
JAK2	0.653	0.291	3
BRCA2	0.0851	0.294	3

Analysis of drug sensitivity data

Overview | Volcano plot | Volcano Data | Elastic net | Scatter plots | Download data



Correlating Mutation with Drug

Analysis of drug sensitivity data

Overview | Volcano plot | Volcano Data | Elastic net | **Scatter plots** | Download data

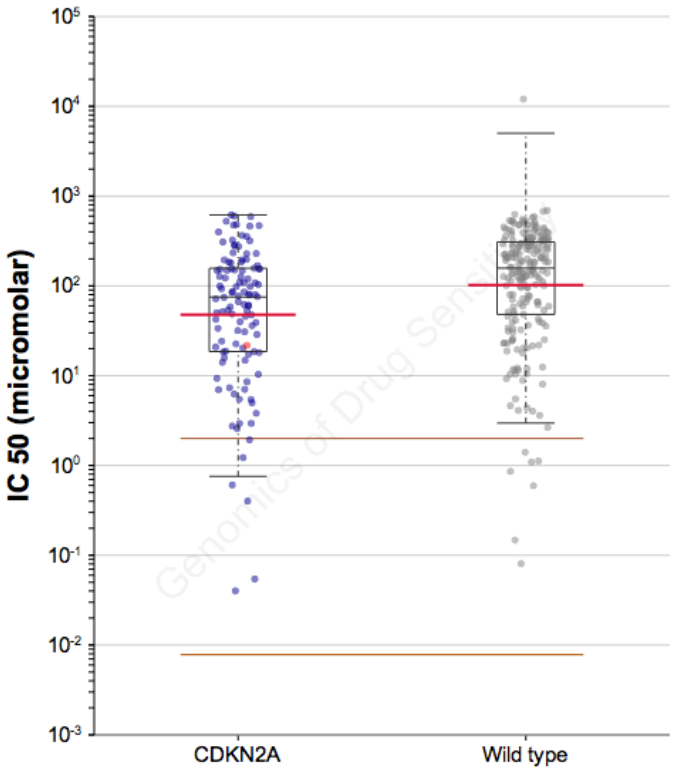
Select gene to plot

- CDKN2A
- SMAD4
- VHL
- MET

CDKN2A

By mutation type | By tissue type

Click on circles to link to cell line information



Screening concentration: 0.0078125 (lower brown line) - 2.0000 (upper brown line)

	CDKN2A	Wild type	Selected
Number of cell lines	112 of 112	217 of 217	
Upper quartile	156.99	308.74	
Median	75.207	157.81	
Geometric mean (red line)	47.741	102.18	
Lower quartile	18.582	48.156	
Mann-Whitney test p value	0.000019161		
Coding mutation	111	0	☑
Deletion	1	0	☑
Amplification	0	0	☑

Check all | Clear all

Download SVG | Download PNG

Correlating Mutation in Specific Tumor Type

Analysis of drug sensitivity data

Overview | Volcano plot | Volcano Data | Elastic net | Scatter plots | Download data

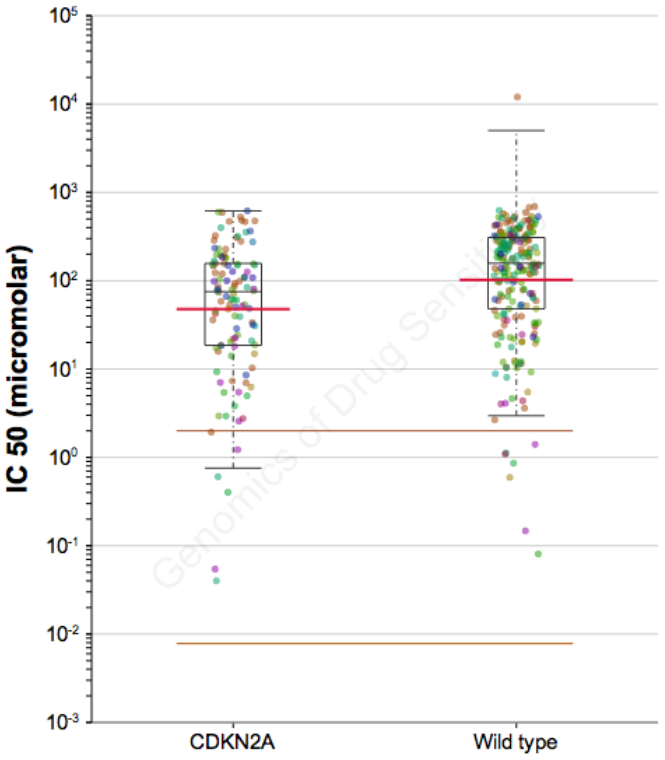
Select gene to plot

- CDKN2A
- SMAD4
- VHL
- MET

CDKN2A

By mutation type | By tissue type

Click on circles to link to cell line information



Download SVG | Download PNG

Screening concentration: 0.0078125 (lower brown line) - 2.0000 (upper brown line)

	CDKN2A	Wild type	Selected
Number of cell lines	112 of 112	217 of 217	
Upper quartile	156.99	308.74	
Median	75.207	157.81	
Geometric mean (red line)	47.741	102.18	
Lower quartile	18.582	48.156	
Mann-Whitney test p value	0.000019161		
Bladder	1	1	☑
Blood	33	66	☑
Bone	5	13	☑
Breast	0	12	☑
Central Nervous System	18	29	☑
Gastro-intestinal tract	4	21	☑
Kidney	7	2	☑
Lung	13	46	☑
Ovary	1	2	☑
Pancreas	2	0	☑
Skin	12	6	☑
Soft tissue	3	5	☑
Thyroid	2	0	☑
Upper aerodigestive	10	4	☑
Uterus	0	5	☑
Other tissue type	1	5	☑

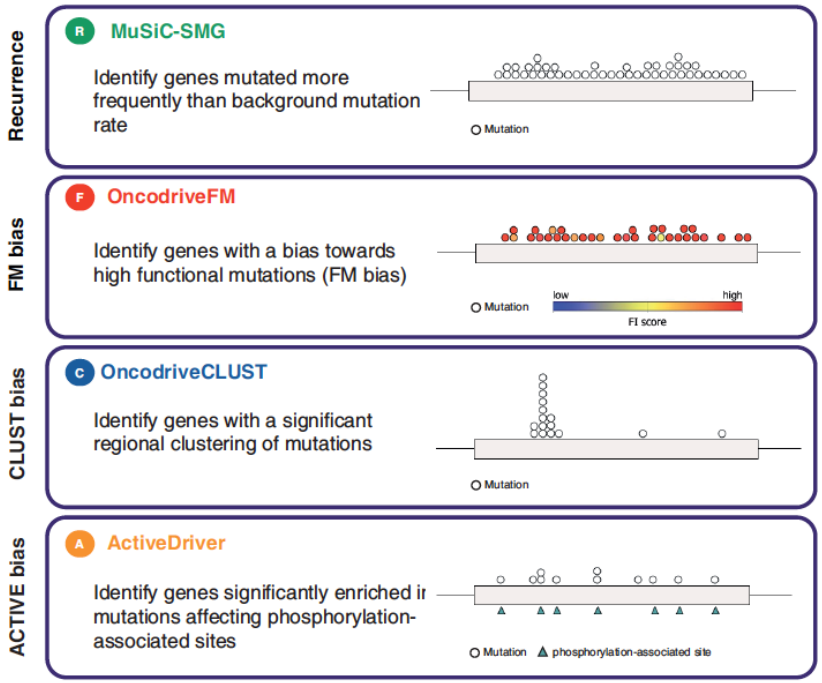
Check all | Clear all

High Confidence Drivers

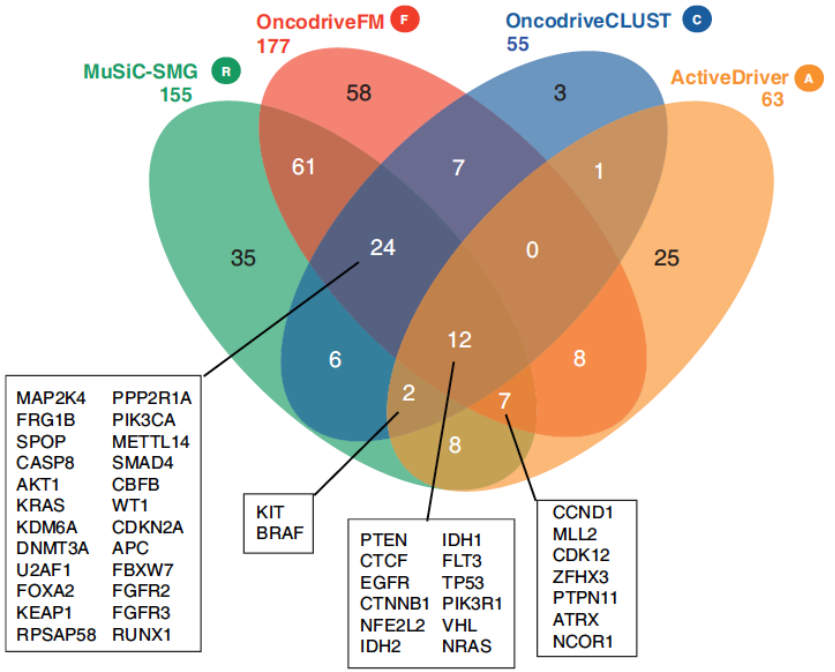
Comprehensive identification of mutational cancer driver genes across 12 tumor types

David Tamborero[1]*, Abel Gonzalez-Perez[1]*, Christian Perez-Llamas[1], Jordi Deu-Pons[1], Cyriac Kandoth[2], Jüri Reimand[3], Michael S. Lawrence[4], Gad Getz[4], Gary D. Bader[3], Li Ding[2,5,6,7] & Nuria Lopez-Bigas[1,8]

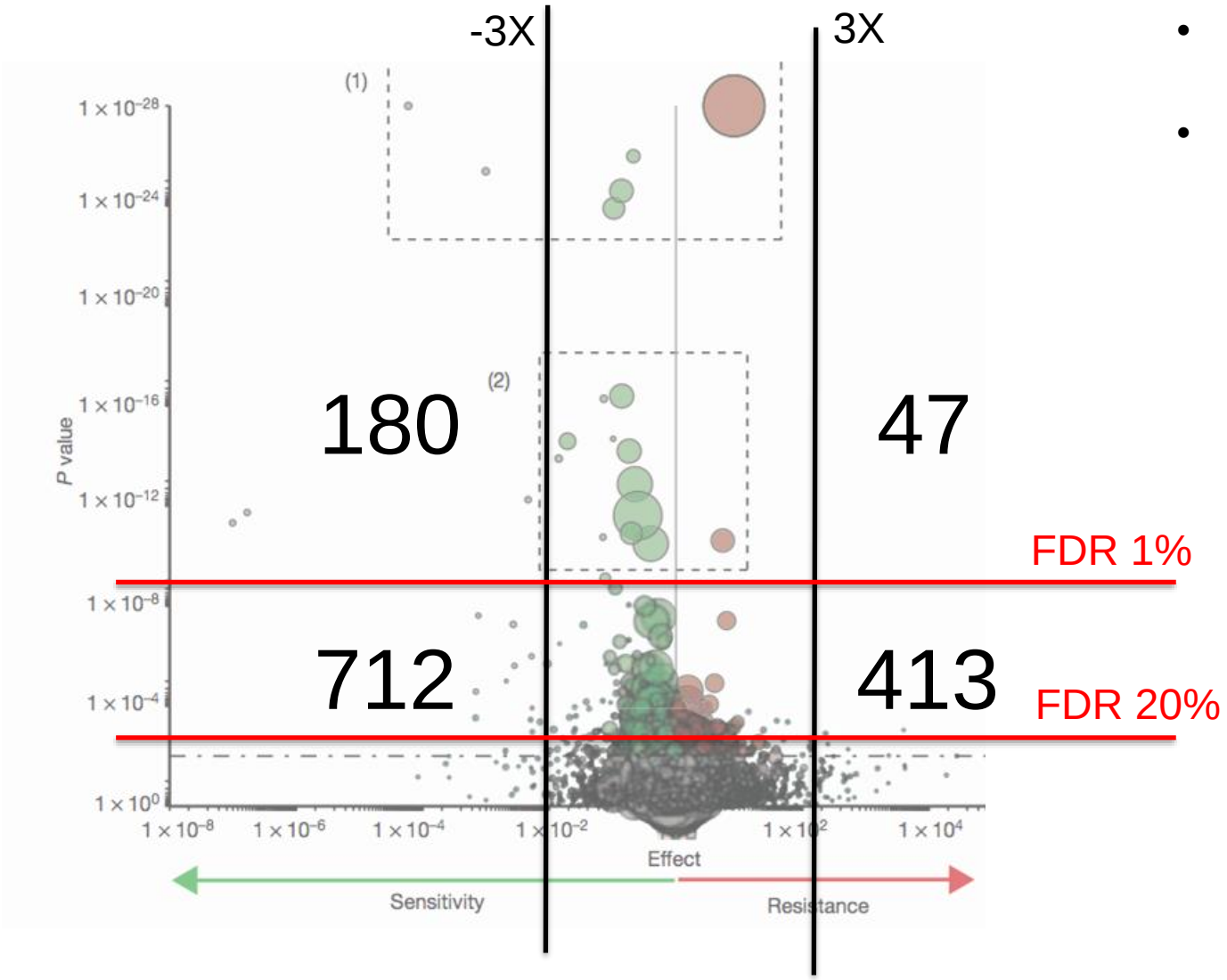
A Signals of positive selection used to identify driver genes



B High Confidence Drivers (HCDs) detected by each method



Stringent Filters using HCD



- 143,000 drug-gene interactions
- 2,286 significance associations ($p < 0.05$, FDR < 20%)

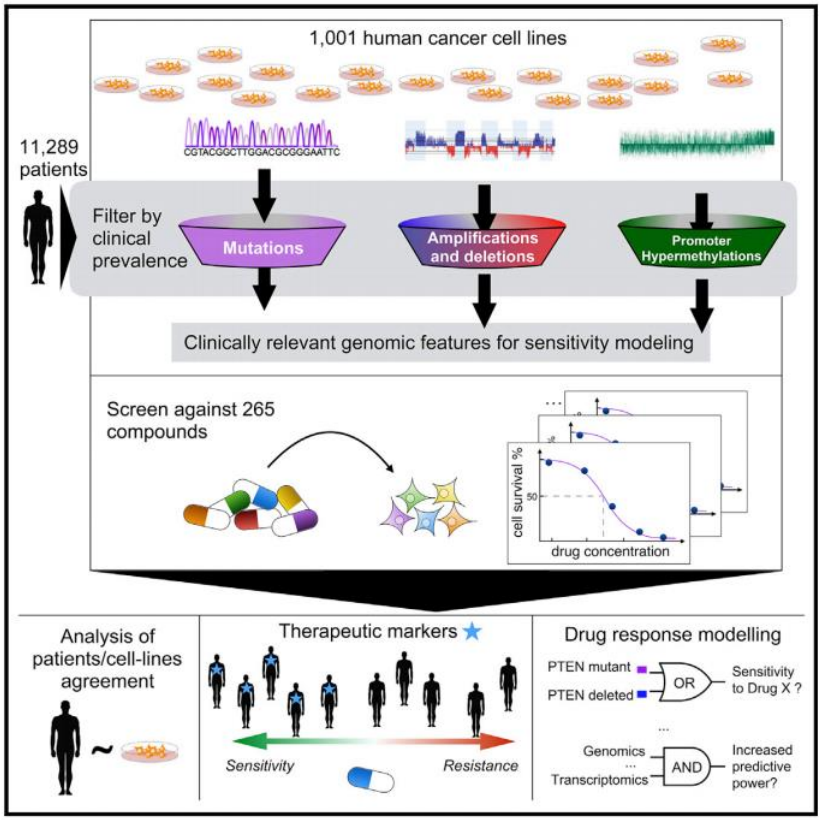
GDSC 2.0

A Landscape of Pharmacogenomic Interactions in Cancer

Francesco Iorio,[1,2,20] Theo A. Knijnenburg,[3,4,20] Daniel J. Vis,[4,20] Graham R. Bignell,[2,20] Michael P. Menden,[1,5,20] Michael Schubert,[1] Nanne Aben,[4,6] Emanuel Gonçalves,[1] Syd Barthorpe,[2] Howard Lightfoot,[2] Thomas Cokelaer,[1,2,17] Patricia Greninger,[7] Ewald van Dyk,[4] Han Chang,[8] Heshani de Silva,[8] Holger Heyn,[9] Xianming Deng,[10,11,18] Regina K. Egan,[7] Qingsong Liu,[10,11] Tatiana Mironenko,[2] Xeni Mitropoulos,[7] Laura Richardson,[2] Jinhua Wang,[10,11] Tinghu Zhang,[10,11] Sebastian Moran,[9] Sergi Sayols,[9,19] Maryam Soleimani,[2] David Tamborero,[12] Nuria Lopez-Bigas,[12,13] Petra Ross-Macdonald,[8] Manel Esteller,[9,13,14] Nathanael S. Gray,[10,11] Daniel A. Haber,[7,15] Michael R. Stratton,[2] Cyril H. Benes,[7] Lodewyk F.A. Wessels,[4,6,16,21] Julio Saez-Rodriguez,[1,5,21] Ultan McDermott,[2,21,*] and Mathew J. Garnett[2,21,*]

Highlights

- We integrate heterogeneous molecular data of 11,289 tumors and 1,001 cell lines
- We measure the response of 1,001 cancer cell lines to 265 anti-cancer drugs
- We uncover numerous oncogenic aberrations that sensitize to an anti-cancer drug
- Our study forms a resource to identify therapeutic options for cancer sub-populations



GDSC Download

Resources

This page allows access to all of our drug sensitivity data and the genomic datasets used in our analyses.

Data File Download Tool

> **Download dataset of your choice using** Data Download Tool .

Data Files

Data Class	Data Type	Objects	Brief description (with links)	Details	Last updated
Annotated	Reference	Cell lines	Annotated list of Cell lines	List of Cell lines included in the study with molecular and drug-response data availability, microsatellite instability status, growth properties and media, TCGA and COSMIC tissue classification	July 4th 2016
Annotated	Reference	Drugs	Screened compounds	List of screened drugs including targets, targeted process/pathways, clinical stage	July 4th 2016
Annotated	Reference	Cell lines	Shared cell lines and drugs (GDSC, CCLE & CTRP)	To aid comparison we list the cell lines and drugs within the GDSC study that are also in the CCLE (Barretina et al, Nature 2012) and/or CTRP (Seashore-Ludlow et al, Cancer Discovery 2015) studies	July 4th 2016
Copy number	Raw	Cell lines	Copy number data for Cell lines	Affymetrix SNP6 cel files at EGA (EGAS00001000978)	July 4th 2016
Copy number	Preprocessed	Cell lines	Gene level copy number data	Copy number data for all genes across all samples derived from PICNIC analysis of Affymetrix SNP6 segmentation data (available via COSMIC Cell Lines Project)	July 4th 2016
Copy number	Preprocessed	Cell lines	RACS in cell lines	Recurrently altered chromosomal segments in cell lines identified from the analysis of patient tumours. As described in Iorio F, et al. Cell. 2016	July 4th 2016
Copy number	Preprocessed	Cell lines	RACSs CNV BEMs for cell lines	Binary event matrix with status in cell lines of recurrently altered chromosomal segments (RACS) identified from the analysis of patient tumours. As described in Iorio F, et al. Cell. 2016.	July 4th 2016
Drug	Raw	Cell lines Drugs	compound sensitivity data for Cell lines	GDSC drug screening data	July 4th 2016
Drug	Preprocessed	Cell lines/Drugs	log(IC50) and AUC values	Natural log half maximal inhibitory concentration (IC50) and Area under the dose-response curve (AUCs) values for all screened cell line/drug combinations	July 4th 2016
Drug	Preprocessed	Cell lines/Drugs	ANOVA results	Results from Pan-Cancer and cancer specific ANOVA	July 4th 2016
Expression	Raw	Cell lines	Expression array data for Cell lines	Affymetrix Human Genome U219 array data at ArrayExpress (E-MTAB-3610)	July 4th 2016
Expression	Preprocessed	Cell lines	RMA normalised expression data for Cell lines	RMA normalised basal expression profiles for all the Cell lines	July 4th 2016
Methylation	Raw	Cell lines	DNA methylation data for Cell lines	IlluminaHumanMethylation450 BeadChip data at GEO (GSE68379)	July 4th 2016
Sequencing	Raw	Cell lines	WES data for Cell lines	Illumina HiSeq 2000 Whole exome sequence BAM files at EGA (EGAS00001000978)	July 4th 2016
Sequencing	Preprocessed	Cell lines	Cell-line sequence variants	List of genomic variants found in Cell lines by whole exome sequencing	July 4th 2016
Sequencing	Preprocessed	Cell lines	Sequencing BEMs for Cell lines	Binary event matrix with status in cell lines of selected cancer genes identified from the analysis of patient tumours. As described in Iorio F, et al. Cell. 2016.	July 4th 2016

Results of data analysis (from 'A landscape of pharmacogenomic interactions in cancer', Iorio F et al. Cell. 2016)

Click here to navigate linked webpages to access data and results associated with our publication "A landscape of Pharmacogenomic Interations in Cancer". We provide these as a resource but for our most current please access data through the GDSC website.

Archive

Click here to download all releases.

LETTER

doi:10.1038/nature11003

The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

Jordi Barretina[1,2,3]†*, Giordano Caponigro[4]*, Nicolas Stransky[1]*, Kavitha Venkatesan[4]*, Adam A. Margolin[1]†*, Sungjoon Kim[5], Christopher J. Wilson[4], Joseph Lehár[4], Gregory V. Kryukov[1], Dmitriy Sonkin[4], Anupama Reddy[4], Manway Liu[4], Lauren Murray[1], Michael F. Berger[1]†, John E. Monahan[4], Paula Morais[1], Jodi Meltzer[4], Adam Korejwa[1], Judit Jané-Valbuena[1,2], Felipa A. Mapa[4], Joseph Thibault[5], Eva Bric-Furlong[4], Pichai Raman[4], Aaron Shipway[5], Ingo H. Engels[5], Jill Cheng[6], Guoying K. Yu[6], Jianjun Yu[6], Peter Aspesi Jr[4], Melanie de Silva[4], Kalpana Jagtap[4], Michael D. Jones[4], Li Wang[4], Charles Hatton[3], Emanuele Palescandolo[3], Supriya Gupta[1], Scott Mahan[1], Carrie Sougnez[1], Robert C. Onofrio[1], Ted Liefeld[1], Laura MacConaill[3], Wendy Winckler[1], Michael Reich[1], Nanxin Li[5], Jill P. Mesirov[1], Stacey B. Gabriel[1], Gad Getz[1], Kristin Ardlie[1], Vivien Chan[6], Vic E. Myer[4], Barbara L. Weber[4], Jeff Porter[4], Markus Warmuth[4], Peter Finan[4], Jennifer L. Harris[5], Matthew Meyerson[1,2,3], Todd R. Golub[1,3,7,8], Michael P. Morrissey[4]*, William R. Sellers[4]*, Robert Schlegel[4]* & Levi A. Garraway[1,2,3]*

978 cell lines treated with 24 drugs

compound name	other name
PHA-665752	
AZD6244	Selumetinib
Tozasertib	VX680
Nutlin-3	
Sorafenib	Nexavar
PF-2341066	Crizotinib
Staurosporine	
Docetaxel	Taxotere
Paclitaxel	Taxol
L-685458	
Irinotecan	Camptosar
Topotecan	Hycamtin
Cisplastin	
AEW541	
TAE684	
Panobinostat	Faridak
Erlotinib	Tarceva
PD-0332991	
Lapatinib	Tykerb
LBW242	
RAF265	
TKI258	Dovitinib
Vandetanib	Zactima
17-AAG	Tanespimycin
PD-0325901	
AZD0530	Saracatinib

Cancer Cell Lines Encyclopedia (CCLE) (Broad and Novartis)

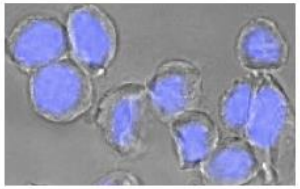
CCLE Cancer Cell Line Encyclopedia

BROAD INSTITUTE

Username sign in
Password Forgot Password?
Register Why Register?

HOME | BROWSE▲ | ANALYSIS TOOLS | HELP▲ | ABOUT

Broad-Novartis Cancer Cell Line Encyclopedia (CCLE)



The Cancer Cell Line Encyclopedia (CCLE) project is a collaboration between the Broad Institute, and the Novartis Institutes for Biomedical Research and its Genomics Institute of the Novartis Research Foundation to conduct a detailed genetic and pharmacologic characterization of a large panel of human cancer models, to develop integrated computational analyses that link distinct pharmacologic vulnerabilities to genomic patterns and to translate cell line integrative genomics into cancer patient stratification. The CCLE provides public access to genomic data, analysis and visualization for about 1000 cell lines.

The CCLE is an ongoing project and some data are not complete yet. The CCLE website is subject to periodic changes and improvements. Please visit regularly!

This project is funded by Novartis.

What you can do on this portal

Search *Cell line, Annotation, Gene*

Search CCLE

Search for information

Enter a keyword to search for genes, news items and publications. Search results for a gene include links to annotations and analyses.

Data Sets

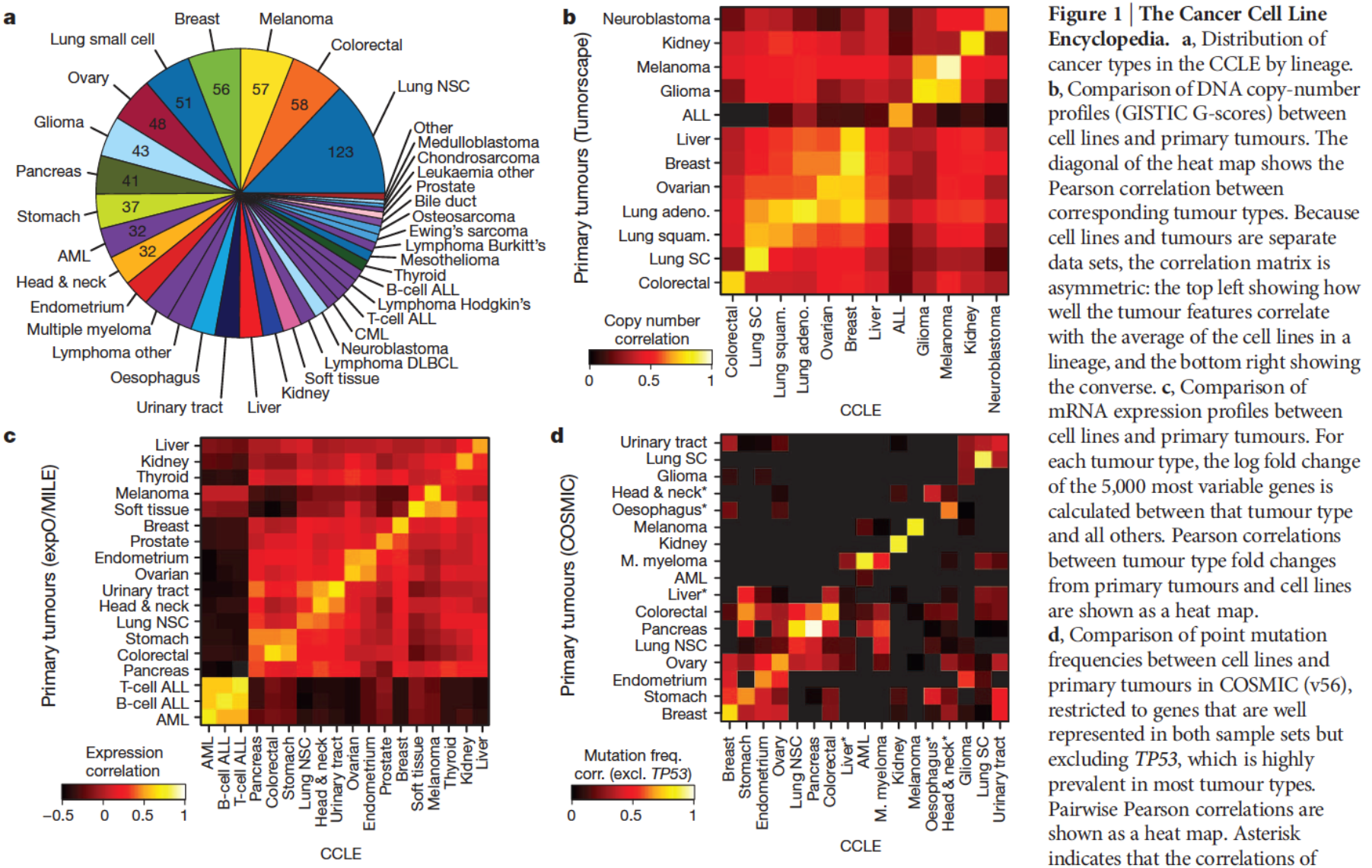
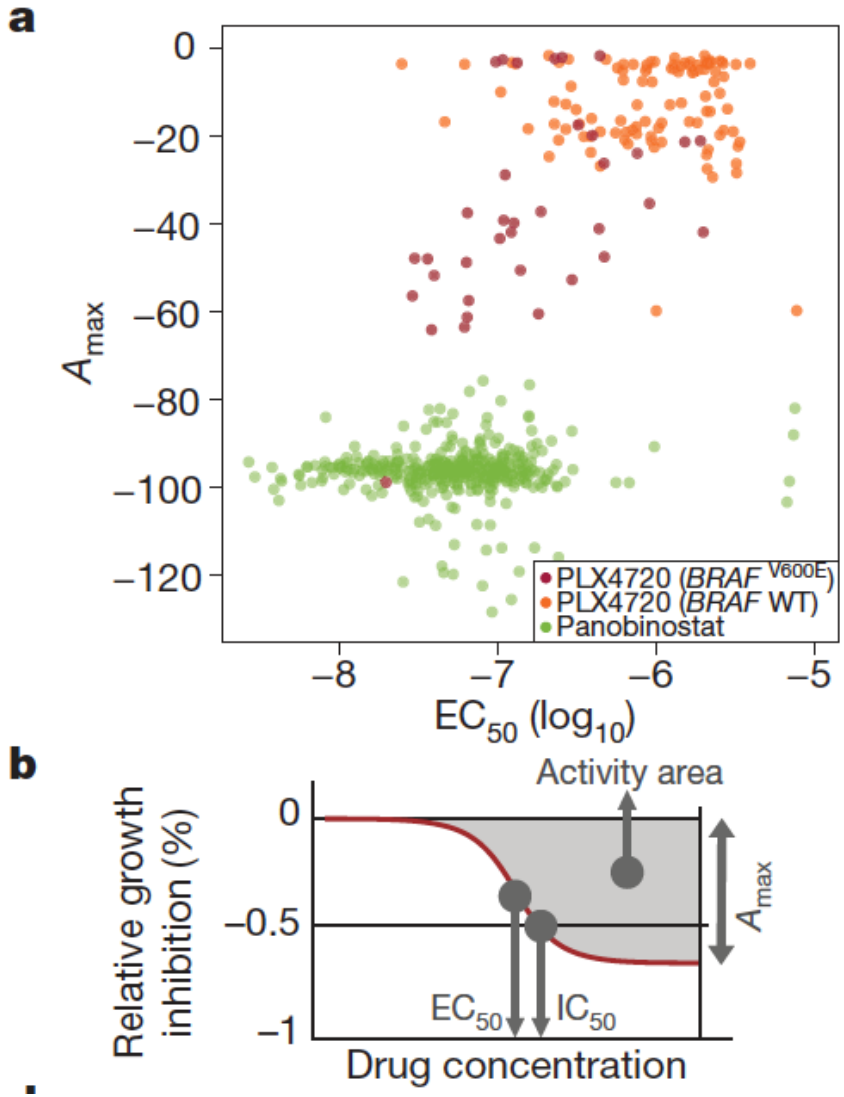


Figure 1 | The Cancer Cell Line Encyclopedia. **a**, Distribution of cancer types in the CCLE by lineage. **b**, Comparison of DNA copy-number profiles (GISTIC G-scores) between cell lines and primary tumours. The diagonal of the heat map shows the Pearson correlation between corresponding tumour types. Because cell lines and tumours are separate data sets, the correlation matrix is asymmetric: the top left showing how well the tumour features correlate with the average of the cell lines in a lineage, and the bottom right showing the converse. **c**, Comparison of mRNA expression profiles between cell lines and primary tumours. For each tumour type, the log fold change of the 5,000 most variable genes is calculated between that tumour type and all others. Pearson correlations between tumour type fold changes from primary tumours and cell lines are shown as a heat map. **d**, Comparison of point mutation frequencies between cell lines and primary tumours in COSMIC (v56), restricted to genes that are well represented in both sample sets but excluding *TP53*, which is highly prevalent in most tumour types. Pairwise Pearson correlations are shown as a heat map. Asterisk indicates that the correlations of oesophageal, liver, and head and neck cancer mutation frequencies are restored when including *TP53*.


a
0
−20
−40
−60
−80
−100
−120
A_{max}
−8
−7
−6
−5
EC_{50} (log_{10})
PLX4720 (*BRAF* ^{V600E})
PLX4720 (*BRAF* WT)
Panobinostat
b
Activity area
Relative growth inhibition (%)
0
−0.5
−1
EC_{50}
IC_{50}
A_{max}
Drug concentration

Browse Data

PAGE INFO▲

Data Type

☑ Affy SNP ☑ Drug data ☑ Gene expression ☑ Mutation ☑ Sample annotation

check all uncheck all

Cancer Cell Line Encyclopedia (55.7GB)

Include previously released (deprecated) files

DNA Copy Number (41.3GB) Affy SNP (Published)

Affymetrix SNP6.0 arrays.
Raw Affymetrix CEL files were converted to a single value for each probe set representing a SNP allele or a copy number probe. Copy numbers were then inferred based upon estimating probe set specific linear calibration curves, followed by normalization by the most similar HapMap normal samples. Segmentation of normalized log2 ratios (specifically, log2(CN/2)) was performed using the circular binary segmentation (CBS) algorithm.

▶ Show Available Data

mRNA expression (8.0GB) Gene expression (Published)

Affymetrix U133+2 arrays.
Raw Affymetrix CEL files were converted to a single value for each probe set using Robust Multi-array Average (RMA) and normalized using quantile normalization. Either the original Affymetrix U133+2 CDF file or a redefined custom CDF file (ENTREZG - v15) was used for the summarization.

▶ Show Available Data

Cell Line Annotations (196.1KB) Sample annotation (Published)

▶ Show Available Data

Oncomap mutations (464.8KB) Mutation (Published)

Oncomap mutation data.
The mutations were assessed in 33 genes (381 specific mutations) using Oncomap 3.0 core. See the associated publication.

▶ Show Available Data

Hybrid capture sequencing (6.4GB) Mutation (In process)

List of mutations and indels in 1651 genes, determined by targeted massively parallel sequencing. Note: the hybrid capture process might yield sequences in genes outside of the target list; those were kept in the analysis and mutations in these genes are present in the files below.

▶ Show Available Data

Pharmacological profiling (8.0MB) Drug data (Published)

Pharmacologic profiles for 24 anticancer drugs across 504 cell lines.

▶ Show Available Data

mRNA expression (8.0GB) Gene expression (Published)

Affymetrix U133+2 arrays.
Raw Affymetrix CEL files were converted to a single value for each probe set using Robust Multi-array Average (RMA) and normalized using quantile normalization. Either the original Affymetrix U133+2 CDF file or a redefined custom CDF file (ENTREZG - v15) was used for the summarization.

▼ Show Available Data

CCLE_Expression_2012-09-29.res	589.6MB	17-Oct-2012	RMA-normalized mRNA expression data. (Affymetrix annotations)
CCLE_Expression_Entrez_2012-09-29.gct	167.2MB	17-Oct-2012	Gene-centric RMA-normalized mRNA expression data. (ENTREZG v15 CDF file, see the Brainarray website for more details about the probe set annotations; gct file format.
CCLE_Expression.Arrays.sif_2012-10-18.txt	124.6KB	18-Oct-2012	Expression arrays samples info file.
CCLE_Expression_Entrez_2012-10-18.res	204.6MB	18-Oct-2012	Gene-centric RMA-normalized mRNA expression data. (ENTREZG v15 CDF file, see the Brainarray website for more details about the probe set annotations; res file format)
CCLE_Expression_Entrez_IGV_2012-09-29.tdf	2.0GB	19-Oct-2012	Gene-centric RMA-normalized mRNA expression data (ENTREZG v15 CDF). This format is only meant to be used with the Integrative Genomics Viewer (IGV, available in the "Analysis tools" section) and the value for each probe set is median-centered and divided by the MAD (robust z-score).
CCLE_Expression.Arrays_2013-03-18.tar.gz	5.0GB	19-Mar-2013	Raw CEL files in a compressed archive.
CCLE_expression_CN_muts_GENEE_2010-04-16.gctx	63.4MB	18-Sep-2013	The gctx file contains the Affymetrix mRNA expression levels, DNA copy number, mutations and indels in 1651 genes by targeted massively parallel sequencing, pharmacologic profiles for 24 anticancer drugs, and cell line annotations.

Analysis

Pharmacological profiling (8.0MB) Drug data (Published)

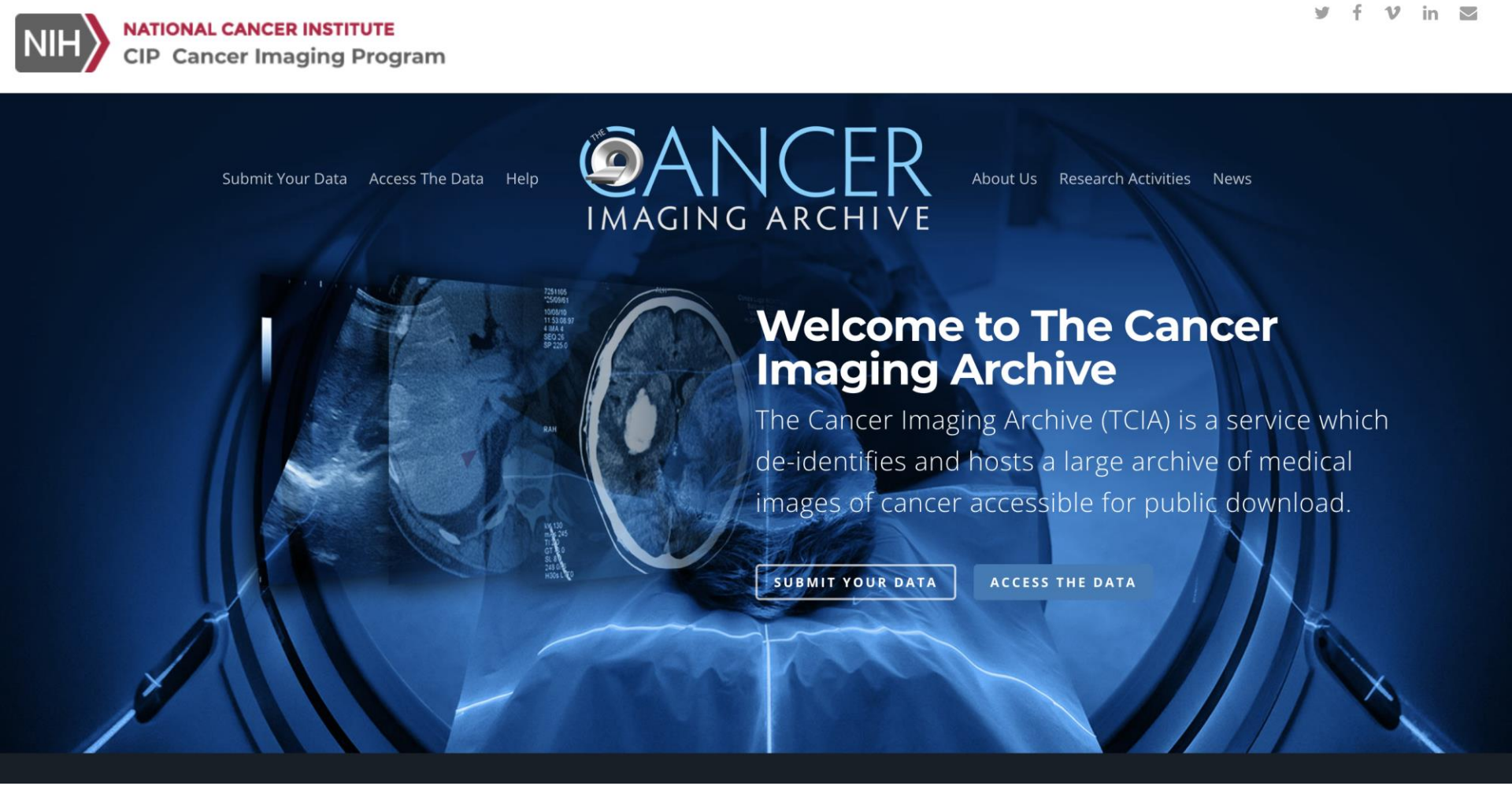
Pharmacologic profiles for 24 anticancer drugs across 504 cell lines.

▼ Show Available Data

CCLE_NP24.2009_Drug_data_2012.02.20.csv	2.3MB	20-Jun-2012	Pharmacologic profiles for 24 anticancer drugs across 504 CCLE lines.
CCLE_NP24.2009_profiling_2012.02.20.csv	2.4KB	17-Apr-2012	List of the 24 drugs profiled across 504 CCLE lines.
CCLE_GNF_data_090613.xls	5.7MB	06-Sep-2013	

Analysis

Cancer Imaging Data



Cancer Imaging Data

Submit Your Data | Access The Data | Help | CANCER IMAGING ARCHIVE | About Us | Research Activities | News

TCIA data are organized as "collections"; typically these are patient cohorts related by a common disease (e.g. lung cancer), image modality or type (MRI, CT, digital histopathology, etc) or research focus. Supporting data related to the images such as patient outcomes, treatment details, genomics and image analyses are also provided when available. Try using the filter box above the table to quickly find collections of interest using keywords. Column headers can also be clicked to change the sorting method.

Show 125 entries

Previous Next

Filter table:

Collection	Cancer Type	Location	Species	Subjects	Image Types	Supporting Data	Access	Status	Updated
AHOD0831	Hodgkin Lymphoma	Various	Human	165	CR,CT,DX,MR,NM,OT,PT,SC,XA	Clinical	Limited	Complete	2022-06-08
Hungarian-Colorectal-Screening	Colorectal Cancer	Colon	Human	200	Pathology	Clinical, Image Analyses	Public	Complete	2022-06-03
FDG-PET-CT-Lesions	Lymphoma, Melanoma, Non-small Cell Lung Cancer	Lung, Lymph, Skin	Human	900	CT, PT, SEG	Clinical, Image Analyses, Software/Source Code	Limited	Complete	2022-06-02
ACRIN 6698/I-SPY2 Breast DWI	Breast Cancer	Breast	Human	385	MR, SEG	Clinical, Image Analyses	Public	Complete	2022-05-02
I-SPY2 Trial	Breast Cancer	Breast	Human	719	MR, SEG	Clinical, Image Analyses	Public	Complete	2022-05-02
DLBCL-Morphology	Diffuse Large B-Cell Lymphoma	Lymph Node	Human	209	Pathology	Clinical, Image Analyses	Public	Complete	2022-04-28
HER2 tumor ROIs	HER2+ Breast Cancer	Breast	Human	273	Pathology	Image Analyses	Public	Complete	2022-04-28
LDCT-and-Projection-data	Various	Head, Chest, Abdomen	Human	299	CT	Clinical	Limited	Complete	2022-03-31

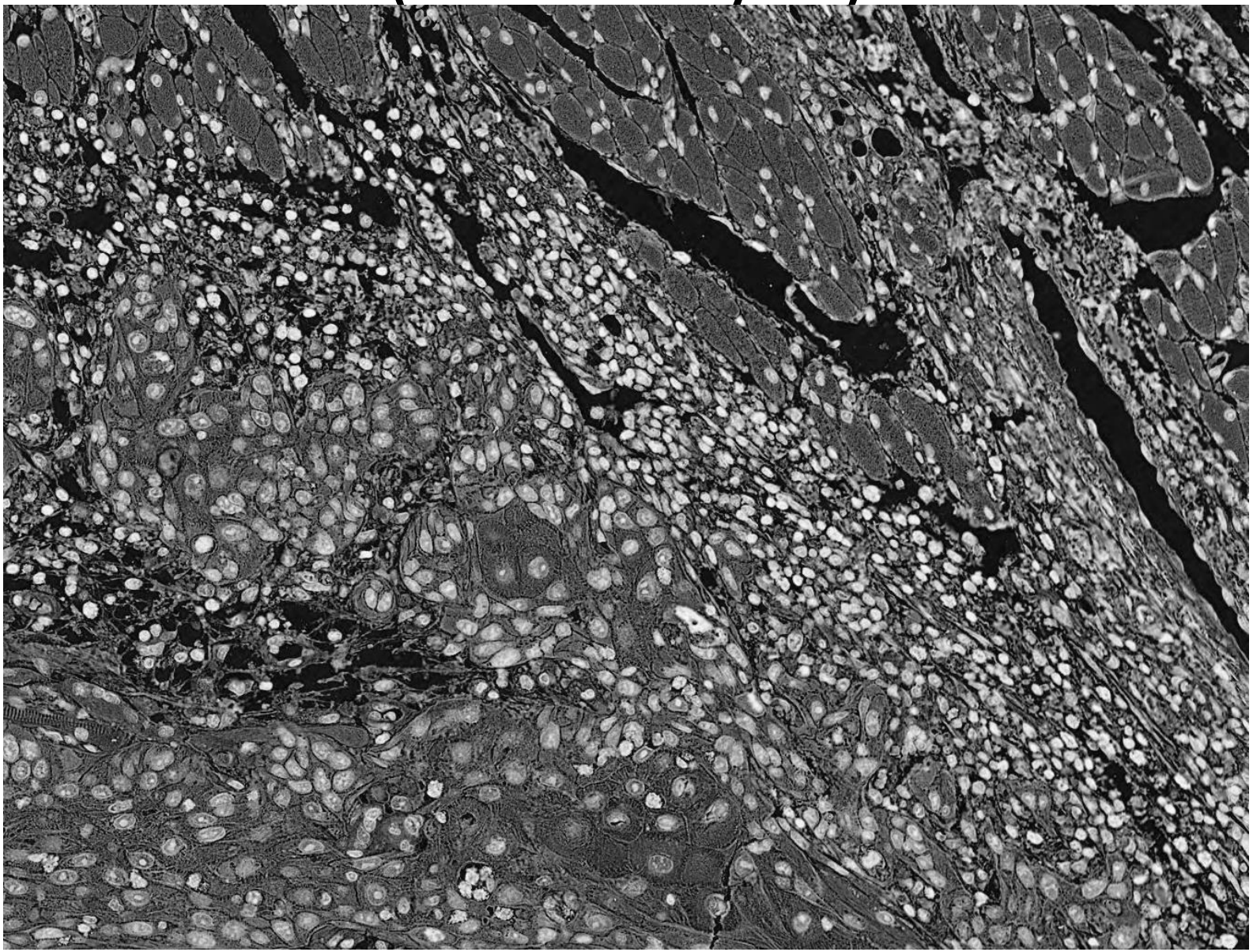
Cancer Imaging Data

Submit Your Data | Access The Data | Help | THE CANCER IMAGING ARCHIVE | About Us | Research Activities | News

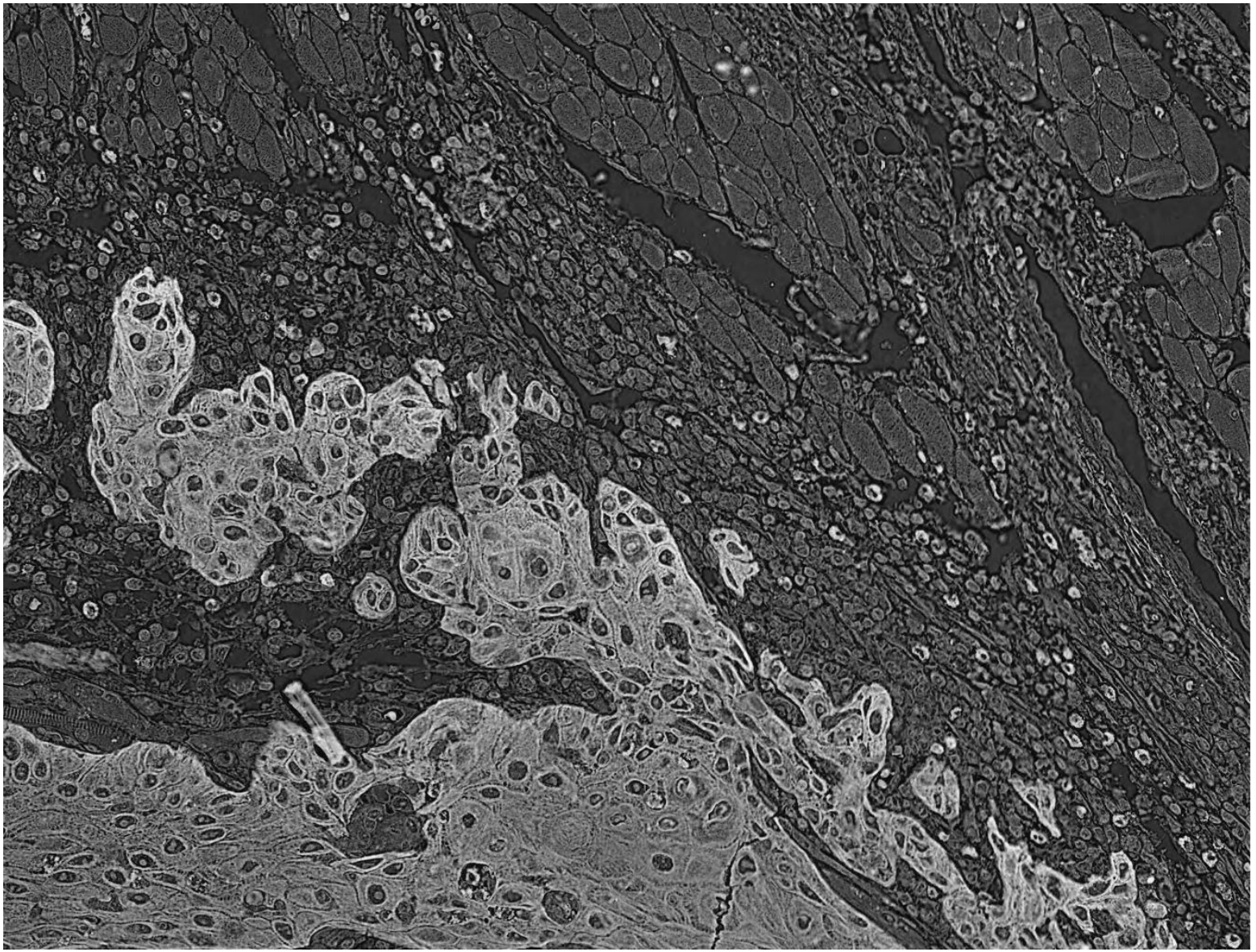
ACRIN-FMISO-Brain (ACRIN 6684)	Glioblastoma Multiforme	Brain	Human	45	CT, MR, PT	Clinical	Limited	Complete	2019-09-16
Mouse-Astrocytoma	Glioblastoma Multiforme	Head	Mouse	48	MR		Public	Complete	2017-03-30
TCGA-GBM	Glioblastoma Multiforme	Brain	Human	262	MR, CT, DX, Pathology	Clinical, Genomics, Image Analyses	Limited	Complete	2014-05-08
ICDC-Glioma (GLIOMA01)	Glioma	Head	Canine	78	MR, Pathology	Genomics	Public	Complete	2021-01-12
HNSCC-mIF-mIHC-comparison	Head and Neck Cancer	Head-Neck	Human	8	Pathology	Image Analyses	Public	Complete	2020-07-31
Head-Neck-Radiomics-HN1	Head and Neck Cancer	Head-Neck	Human	137	CT, PT, RTSTRUCT, SEG	Clinical	Limited	Complete	2020-07-29
AAPM RT-MAC Grand Challenge 2019	Head and Neck Cancer	Head-Neck	Human	55	MR, RTSTRUCT		Limited	Complete	2020-07-21
CPTAC-HNSCC	Head and Neck Cancer	Head-Neck	Human	112	CT, MR, SC, Pathology	Clinical, Genomics, Proteomics	Limited	Ongoing	2020-06-30

Example – Histopathology Data

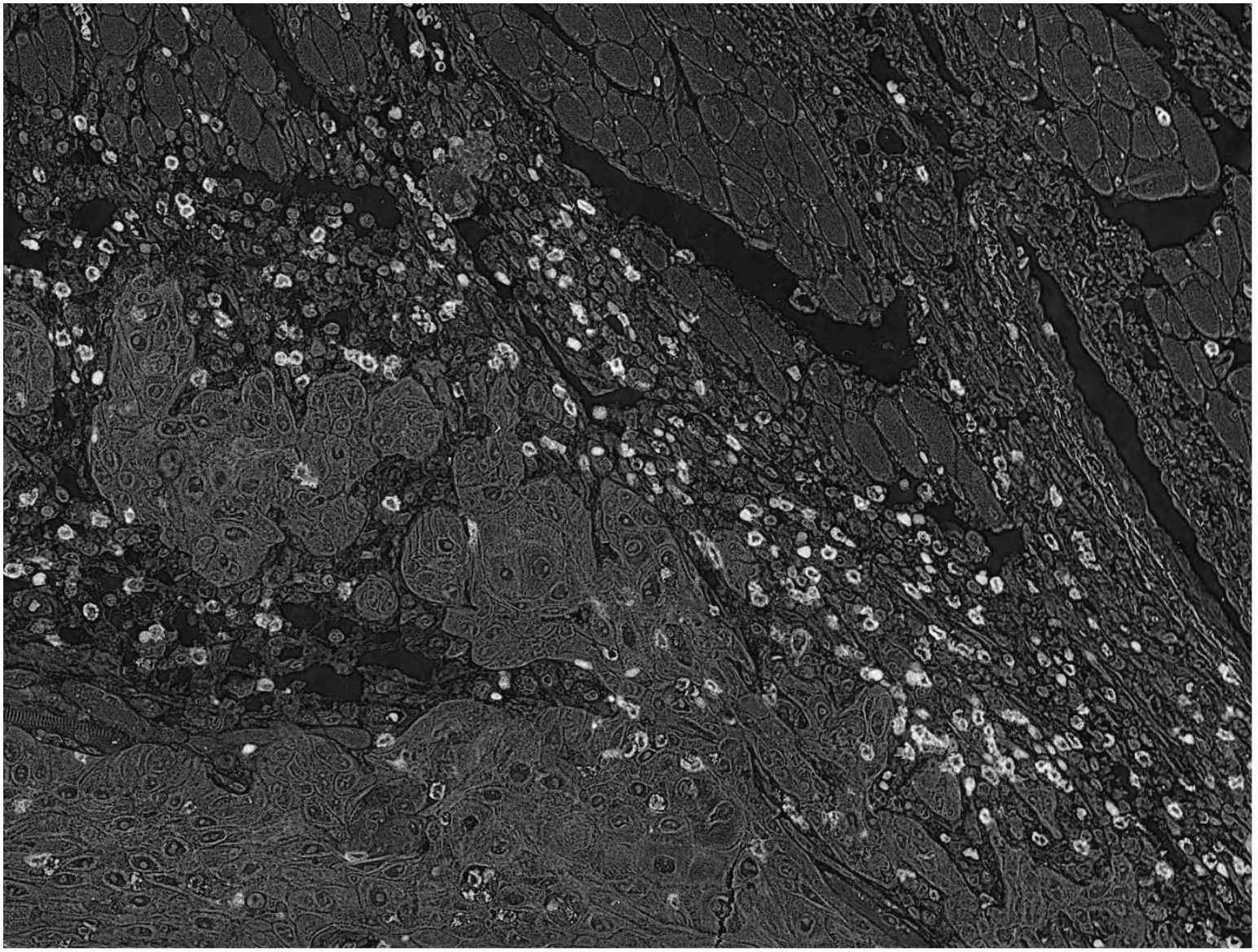
Example – Histopathology Data (Hematoxylin)



Example – Histopathology Data (mIHC – PCK)



Example – Histopathology Data (mIHC – CD8+)



Clinical Trials Data

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

ClinicalTrials.gov is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world. Learn more about clinical studies and about this site, including relevant history, policies, and laws.

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

ClinicalTrials.gov currently lists 224,696 studies with locations in all 50 States and in 192 countries.

Text Size

Search for Studies
Example: "Heart attack" AND "Los Angeles"
Search
Advanced Search | See Studies by Topic
See Studies on Map

Search Help
- How to search
- How to find results of studies
- How to read a study record

Locations of Recruiting Studies

Total N = 39,469 studies
(Data as of September 07, 2016)

- See more trends, charts, and maps

For Patients and Families
- How to find studies
- See studies by topic
- Learn about clinical studies
- Learn more

For Researchers
- How to submit studies
- Download content for analysis
- About the results database
- Learn more

For Study Record Managers
- Why register?
- How to register your study
- FDAAA 801 requirements
- Learn more

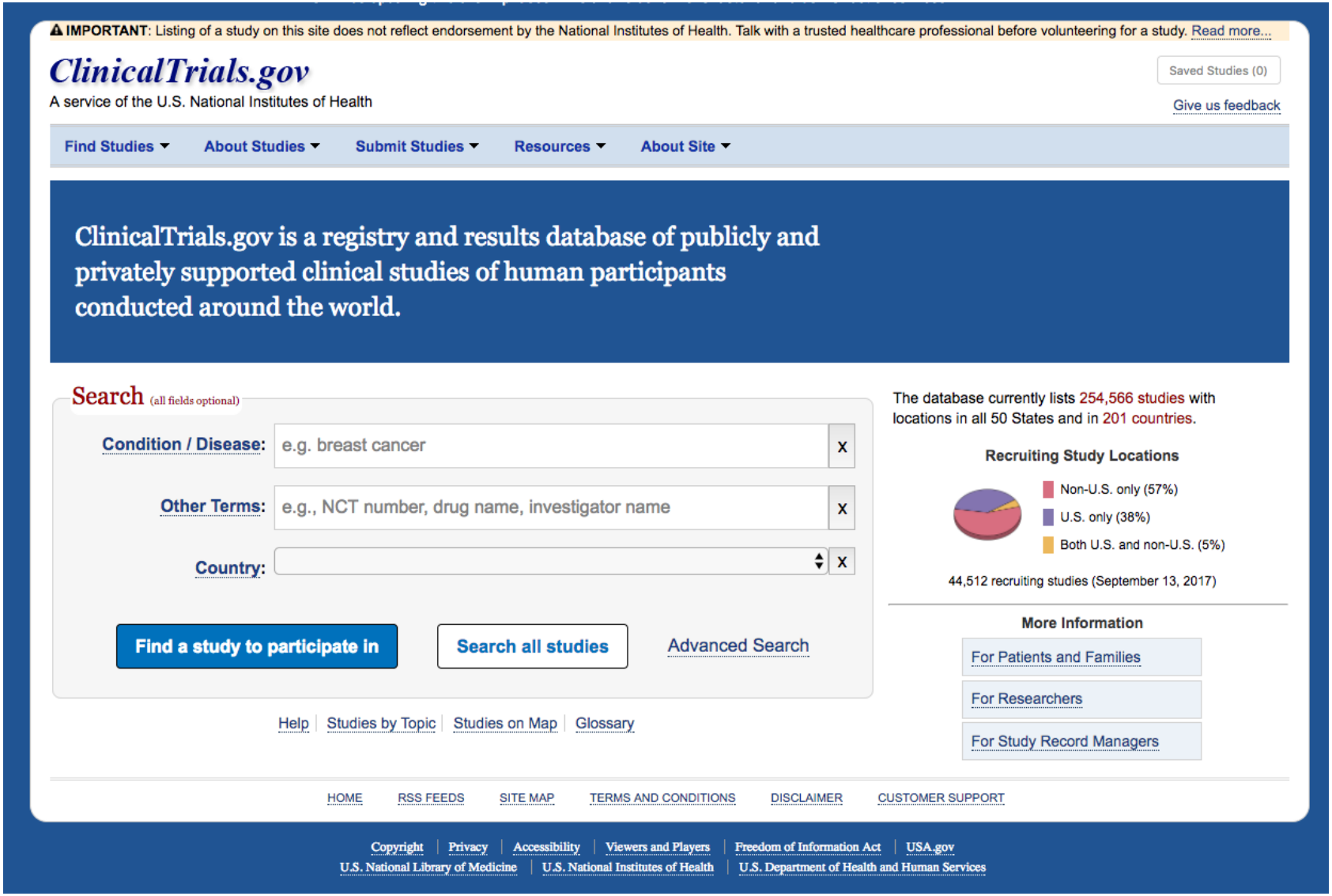
Learn More
- Tutorials for using ClinicalTrials.gov
- Glossary of common site terms
- For the press
- Using our RSS feeds

HOME | RSS FEEDS | SITE MAP | TERMS AND CON

FDAAA 801 (Sept 2007):
Expands registry and adds results reporting requirements

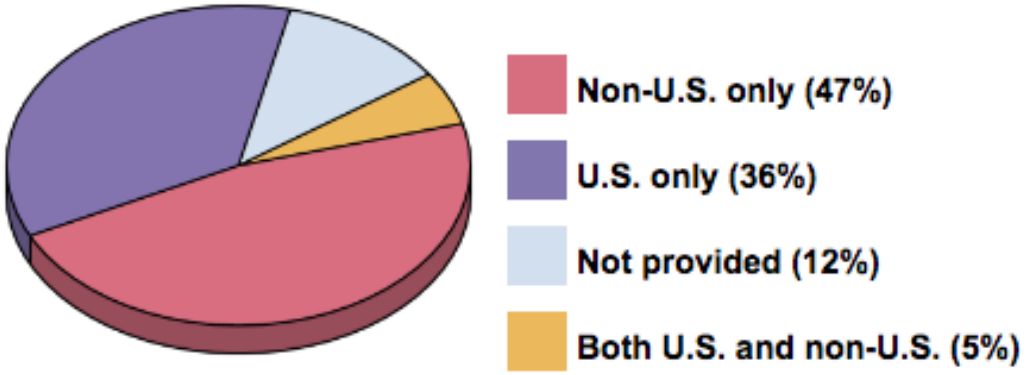
Copyright | Privacy | Accessibility | Viewers and Players | Freedom of Information Act | USA.gov
U.S. National Library of Medicine | U.S. National Institutes of Health | U.S. Department of Health and Human Services

Clinical Trials Data



Clinical Trials Data

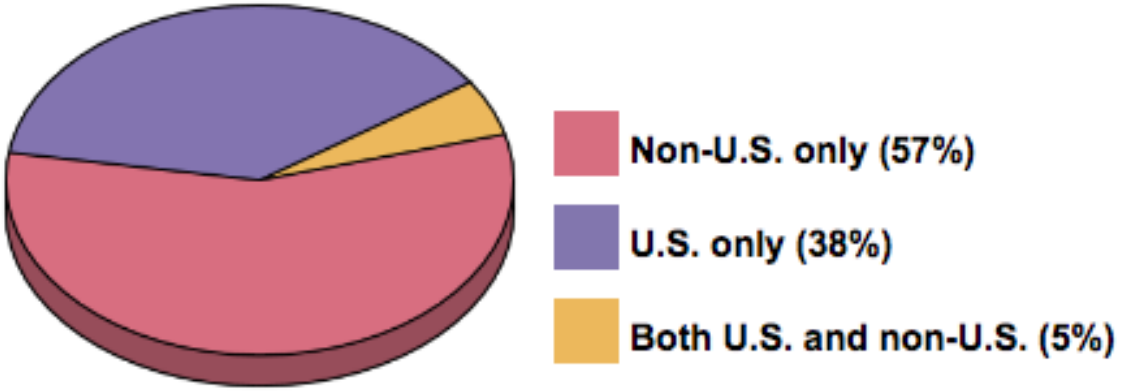
Percentage of Registered Studies by Location (as of September 13, 2017)
Total N = 254,566 studies



Location	Number of Registered Studies and Percentage of Total (as of September 13, 2017)
Non-U.S. only	119,471 (47%)
U.S. only	91,048 (36%)
Not provided	30,092 (12%)
Both U.S. and non-U.S.	13,955 (5%)
Total	254,566

Clinical Trials Data

Percentage of Recruiting Studies by Location (as of September 13, 2017)
Total N = 44,512 studies



Location	Number of Recruiting Studies and Percentage of Total (as of September 13, 2017)
Non-U.S. only	25,249 (57%)
U.S. only	16,965 (38%)
Both U.S. and non-U.S.	2,298 (5%)
Total	44,512

Clinical Trials Data

Study and Intervention Type (as of September 13, 2017)		Number of Registered Studies and Percentage of Total	Number of Studies With Posted Results and Percentage of Total***
Total		254,566	28,250
Interventional		203,299 (79%)	26,497 (93%)
Type of Intervention*	**Drug or biologic**	121,523	21,219
	Behavioral, other	61,185	4,610
	Surgical procedure	21,834	1,454
	Device**	24,440	3,185
Observational		50,095 (19%)	1,753 (6%)
Expanded Access		444	N/A

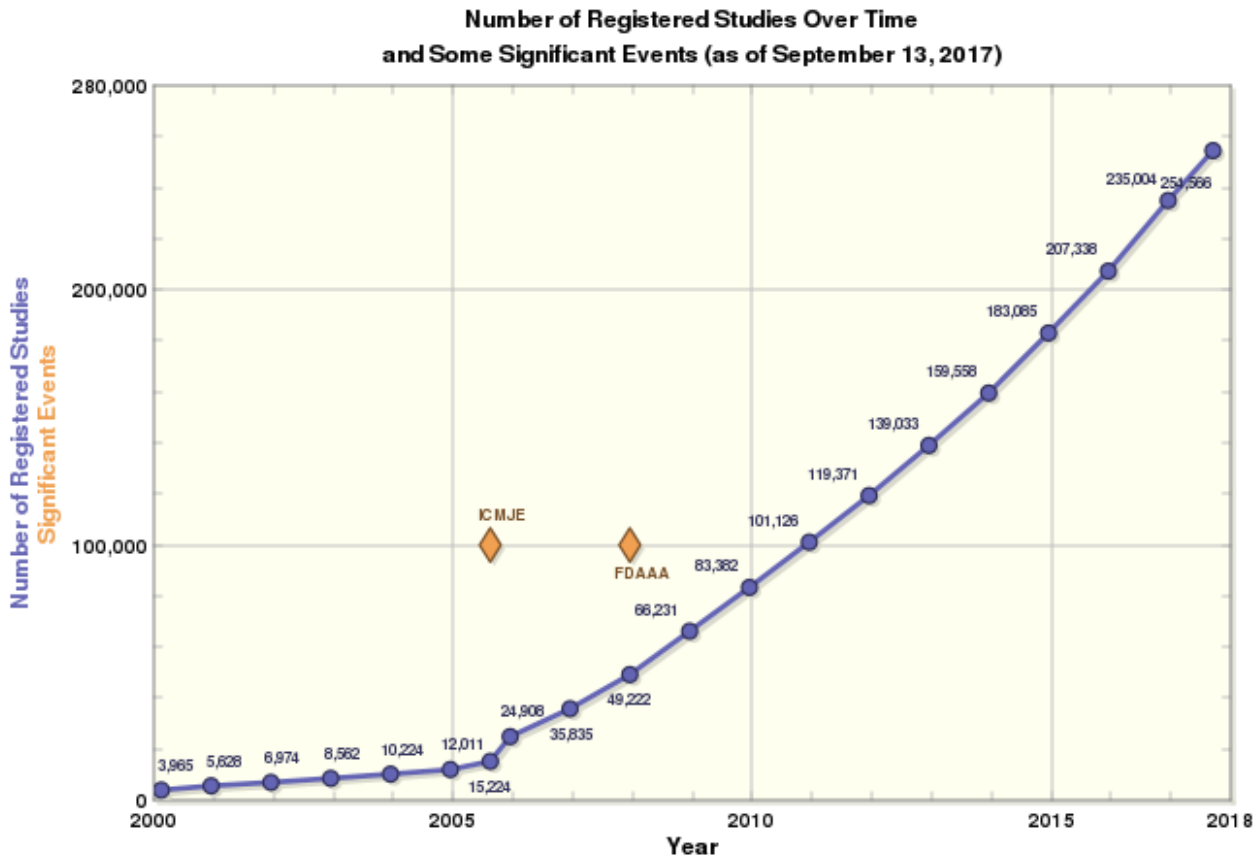
* A study may include more than one type of intervention, meaning that a single study may be counted more than once. Because of this, the sum of counts by type of intervention do not equal the total number of interventional studies.
** A total of 728 applicable device clinical trials were submitted as "delayed posting" under the Food and Drug Administration Amendments Act of 2007 (FDAAA). That is, the Responsible Party indicated that the trial includes a device not previously approved or cleared by the Food and Drug Administration (U.S. FDA) for any use. These trials are not included in the counts of trials with at least one device.
*** Results are required to be submitted only for certain studies. For example, results submission is generally not required for observational studies; trials completed before 2008; and trials that include drugs, biologics, or devices not previously approved by the U.S. FDA for any use. See FDAAA 801 Requirements for further information.

N/A = not applicable

Clinical Trials Data

Number of Registered Studies Over Time

The graph and table below show the total number of studies registered on ClinicalTrials.gov since 2000, based on the First Received date. The first version of ClinicalTrials.gov was made available to the public on February 29, 2000.

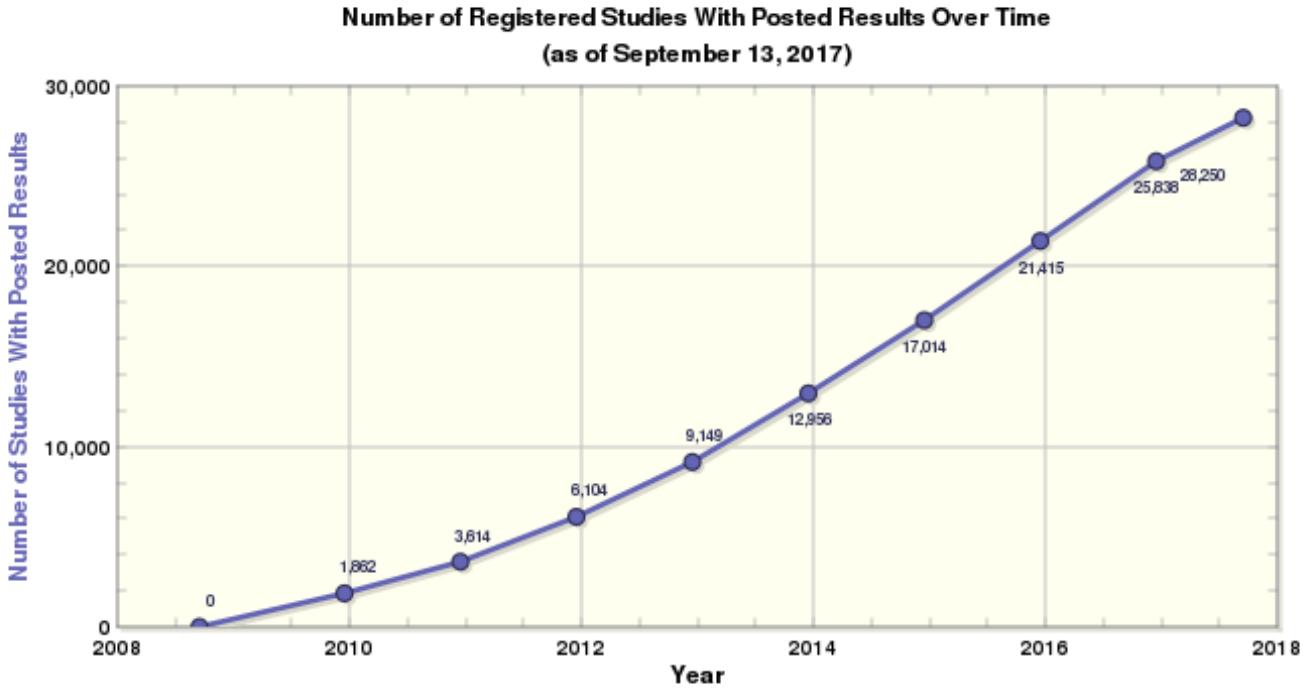


Source: https://ClinicalTrials.gov

Clinical Trials Data

Number of Registered Studies With Posted Results Over Time

The graph and table below show the number of registered studies with results posted on ClinicalTrials.gov, based on the Results First Received date. ClinicalTrials.gov launched its results database in September 2008, at which time sponsors or investigators were allowed to begin submitting results for their registered studies. The results database was developed to accommodate the results submission requirements outlined in FDAAA. See About the Results Database for more information.



Source: https://ClinicalTrials.gov

Reporting Clinical Trials Data

BMJ

BMJ 2011;344:d7292 doi: 10.1136/bmj.d7292 (Published 3 January 2012)

RESEARCH

Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis

OPEN ACCESS

Joseph S Ross *assistant professor of medicine*[1 2], Tony Tse *program analyst at ClinicalTrials.gov*[3], Deborah A Zarin *director of ClinicalTrials.gov*[3], Hui Xu *postgraduate house staff trainee*[4], Lei Zhou *postgraduate house staff trainee*[4], Harlan M Krumholz *Harold H Hines Jr professor of medicine and professor of investigative medicine and of public health*[2 5 6]

[1]Section of General Internal Medicine, Department of Medicine, Yale University School of Medicine, New Haven, CT, USA; [2]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT; [3]Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; [4]Fuwai Hospital and Cardiovascular Institute, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [5]Robert Wood Johnson Clinical Scholars Program and Section of Cardiovascular Medicine, Department of Medicine, Yale University School of Medicine, New Haven, CT; [6]Section of Health Policy and Administration, Yale University School of Epidemiology and Public Health, New Haven, CT

Abstract

Objective To review patterns of publication of clinical trials funded by US National Institutes of Health (NIH) in peer reviewed biomedical journals indexed by Medline.

Design Cross sectional analysis.

Setting Clinical trials funded by NIH and registered within ClinicalTrials.gov (clinicaltrials.gov), a trial registry and results database maintained by the US National Library of Medicine, after 30 September 2005 and updated as having been completed by 31 December 2008, allowing at least 30 months for publication after completion of the trial.

Main outcome measures Publication and time to publication in the biomedical literature, as determined through Medline searches, the last of which was performed in June 2011.

Results Among 635 clinical trials completed by 31 December 2008, 294 (46%) were published in a peer reviewed biomedical journal, indexed by Medline, within 30 months of trial completion. The median period of follow-up after trial completion was 51 months (25th-75th centiles 40-68 months), and 432 (68%) were published overall. Among published trials, the median time to publication was 23 months (14-36 months). Trials completed in either 2007 or 2008 were more likely to be published within 30 months of study completion compared with trials completed before 2007 (54% (196/366) *v* 36% (98/269); P<0.001).

Conclusions Despite recent improvement in timely publication, fewer than half of trials funded by NIH are published in a peer reviewed biomedical journal indexed by Medline within 30 months of trial completion. Moreover, after a median of 51 months after trial completion, a third of trials remained unpublished.

Introduction

Today, there is an increasing emphasis on the successful translation of results from research into practice. This requires the timely dissemination of findings. While research results might be submitted directly to regulatory agencies, such as the Food and Drug Administration (FDA), physicians and policy makers generally depend on peer reviewed publications to learn about findings from clinical trials. Extensive research has shown, however, that the results of studies are often not shared publicly in a timely way and that between 25% and 50% of clinical trials remain unpublished even several years after completion,[1-16] although this work was largely focused on industry funded studies. There are many possible reasons behind the delayed or non-publication of results from clinical trials, including lack of incentive to disseminate negative or unsupportive findings, time constraints, limited resources, changing interests, or even failure to have an article accepted by a journal.

Understanding the patterns of publication of research findings among publicly funded research, as opposed to industry funded research, is important because of the funding and the expectation for public benefit. Within the United States, the National Institutes of Health (NIH) is the leading and largest government agency responsible for biomedical and health related research and invests more than $12bn (about £7600m or €8900m) of public resources in funding research in people or in clinical research, $3.5bn explicitly on clinical trials.[17] These costs do not include the considerable contributions and costs incurred by the participants in the research. Previous work suggests that

Correspondence to: J S Ross, Section of General Internal Medicine, Yale University School of Medicine, PO Box 208093, New Haven, CT 0520, USA joseph.ross@yale.edu

Abstract

Objective To review patterns of publication of clinical trials funded by US National Institutes of Health (NIH) in peer reviewed biomedical journals indexed by Medline.

Design Cross sectional analysis.

Setting Clinical trials funded by NIH and registered within ClinicalTrials.gov (clinicaltrials.gov), a trial registry and results database maintained by the US National Library of Medicine, after 30 September 2005 and updated as having been completed by 31 December 2008, allowing at least 30 months for publication after completion of the trial.

Main outcome measures Publication and time to publication in the biomedical literature, as determined through Medline searches, the last of which was performed in June 2011.

Results Among 635 clinical trials completed by 31 December 2008, 294 (46%) were published in a peer reviewed biomedical journal, indexed by Medline, within 30 months of trial completion. The median period of follow-up after trial completion was 51 months (25th-75th centiles 40-68 months), and 432 (68%) were published overall. Among published trials, the median time to publication was 23 months (14-36 months). Trials completed in either 2007 or 2008 were more likely to be published within 30 months of study completion compared with trials completed before 2007 (54% (196/366) *v* 36% (98/269); P<0.001).

Conclusions Despite recent improvement in timely publication, fewer than half of trials funded by NIH are published in a peer reviewed biomedical journal indexed by Medline within 30 months of trial completion. Moreover, after a median of 51 months after trial completion, a third of trials remained unpublished.

Clinical Trials Data

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

Example: "Heart attack" AND "Los Angeles"
Search for studies: Search
Advanced Search | Help | Studies by Topic | Glossary

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

Home > Find Studies > Search Results > Study Record Detail Text Size

Trial record **36 of 114** for: vemurafenib
◂ Previous Study | Return to List | Next Study ▸

A Study of Vemurafenib (RO5185426) in Comparison With Dacarbazine in Previously Untreated Patients With Metastatic Melanoma (BRIM 3)

This study has been completed.

Sponsor:
Hoffmann-La Roche

Information provided by (Responsible Party):
Hoffmann-La Roche

ClinicalTrials.gov Identifier:
NCT01006980

First received: October 30, 2009
Last updated: July 5, 2016
Last verified: December 2015
History of Changes

Full Text View | Tabular View | Study Results | Disclaimer | How to Read a Study Record

Purpose

This randomized, open-label study will evaluate the efficacy, safety and tolerability of **vemurafenib** (**RO5185426**) as compared to dacarbazine in previously untreated patients with metastatic melanoma. Patients will be randomized to receive either **vemurafenib** 960 mg orally twice daily or dacarbazine 1000 mg/m2 intravenously every 3 weeks. Anticipated time on study treatment is until disease progression or unacceptable toxicity occurs. Patients in the dacarbazine arm may cross over to **vemurafenib** treatment.

Condition	Intervention	Phase
Malignant Melanoma	Drug: **Vemurafenib** Drug: Dacarbazine	Phase 3

Study Type: Interventional
Study Design: Allocation: Randomized
Endpoint Classification: Safety/Efficacy Study
Intervention Model: Parallel Assignment
Masking: Open Label
Primary Purpose: Treatment

Official Title: BRIM 3: A Randomized, Open-Label, Controlled, Multicenter, Phase III Study in Previously Untreated Patients With Unresectable Stage IIIC or Stage IV Melanoma With V600E BRAF Mutation Receiving **Vemurafenib** (**RO5185426**) or Dacarbazine

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation

Paul B. Chapman, M.D., Axel Hauschild, M.D., Caroline Robert, M.D., Ph.D., John B. Haanen, M.D., Paolo Ascierto, M.D., James Larkin, M.D., Reinhard Dummer, M.D., Claus Garbe, M.D., Alessandro Testori, M.D., Michele Maio, M.D., David Hogg, M.D., Paul Lorigan, M.D., Celeste Lebbe, M.D., Thomas Jouary, M.D., Dirk Schadendorf, M.D., Antoni Ribas, M.D., Steven J. O'Day, M.D., Jeffrey A. Sosman, M.D., John M. Kirkwood, M.D., Alexander M.M. Eggermont, M.D., Ph.D., Brigitte Dreno, M.D., Ph.D., Keith Nolop, M.D., Jiang Li, Ph.D., Betty Nelson, M.A., Jeannie Hou, M.D., Richard J. Lee, M.D., Keith T. Flaherty, M.D., and Grant A. McArthur, M.B., B.S., Ph.D., for the BRIM-3 Study Group*

ABSTRACT

BACKGROUND
Phase 1 and 2 clinical trials of the BRAF kinase inhibitor vemurafenib (PLX4032) have shown response rates of more than 50% in patients with metastatic melanoma with the BRAF V600E mutation.

METHODS
We conducted a phase 3 randomized clinical trial comparing vemurafenib with dacarbazine in 675 patients with previously untreated, metastatic melanoma with the BRAF V600E mutation. Patients were randomly assigned to receive either vemurafenib (960 mg orally twice daily) or dacarbazine (1000 mg per square meter of body-surface area intravenously every 3 weeks). Coprimary end points were rates of overall and progression-free survival. Secondary end points included the response rate, response duration, and safety. A final analysis was planned after 196 deaths and an interim analysis after 98 deaths.

RESULTS
At 6 months, overall survival was 84% (95% confidence interval [CI], 78 to 89) in the vemurafenib group and 64% (95% CI, 56 to 73) in the dacarbazine group. In the interim analysis for overall survival and final analysis for progression-free survival, vemurafenib was associated with a relative reduction of 63% in the risk of death and of 74% in the risk of either death or disease progression, as compared with dacarbazine (P<0.001 for both comparisons). After review of the interim analysis by an independent data and safety monitoring board, crossover from dacarbazine to vemurafenib was recommended. Response rates were 48% for vemurafenib and 5% for dacarbazine. Common adverse events associated with vemurafenib were arthralgia, rash, fatigue, alopecia, keratoacanthoma or squamous-cell carcinoma, photosensitivity, nausea, and diarrhea; 38% of patients required dose modification because of toxic effects.

CONCLUSIONS
Vemurafenib produced improved rates of overall and progression-free survival in patients with previously untreated melanoma with the BRAF V600E mutation. (Funded by Hoffmann–La Roche; BRIM-3 ClinicalTrials.gov number, NCT01006980.)

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Chapman at the Department of Medicine, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10065, or at chapmanp@mskcc.org.

Drs. Flaherty and McArthur contributed equally to this article.

*Members of the BRAF Inhibitor in Melanoma 3 (BRIM-3) study group are listed in the Supplementary Appendix at NEJM.org.

This article (10.1056/NEJMoa1103782) was published on June 5, 2011, and updated on March 13, 2012, at NEJM.org.

N Engl J Med 2011;364:2507-16.

Clinical Trials Data

Results First Received: July 29, 2011

Study Type:	Interventional
Study Design:	Allocation: Randomized; Endpoint Classification: Safety/Efficacy Study; Intervention Model: Parallel Assignment; Masking: Open Label; Primary Purpose: Treatment
Condition:	Malignant Melanoma
Interventions:	Drug: **Vemurafenib** Drug: Dacarbazine

Participant Flow

Hide Participant Flow

Recruitment Details

Key information relevant to the recruitment process for the overall study, such as dates of the recruitment period and locations
No text entered.

Pre-Assignment Details

Significant events and approaches for the overall study following participant enrollment, but prior to group assignment
675 participants were randomized, 337 to vemurafenib and 338 to dacarbazine. One participant randomized to dacarbazine was treated in error with vemurafenib throughout the study and is included in the Vemurafenib arm in the table below and for exposure and safety analyses and is included in the dacarbazine arm for efficacy analyses.

Reporting Groups

	Description
Vemurafenib	Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).

Participant Flow: Overall Study

	Vemurafenib	**Dacarbazine**
STARTED	**337**	**338**
Treated	**336**	**293**
COMPLETED	**0**	**0**
NOT COMPLETED	**337**	**338**
Randomized but Not Treated	**1**	**45**
Adverse Event	**25**	**5**
Death	**13**	**12**
Progression	**257**	**218**
Withdrawal of Consent	**4**	**6**
Refuse Treatment	**9**	**6**
Protocol Violation	**2**	**3**
Reason Not Specified	**26**	**43**

Clinical Trials Data

Baseline Characteristics

Hide Baseline Characteristics

Population Description

Explanation of how the number of participants for analysis was determined. Includes whether analysis was per protocol, intention to treat, or another method. Also provides relevant details such as imputation technique, as appropriate.
No text entered.

Reporting Groups

	Description
Vemurafenib	Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).
Total	Total of all reporting groups

Baseline Measures

	Vemurafenib	**Dacarbazine**	**Total**
Number of Participants [units: participants]	**337**	**338**	**675**
Age, Customized [units: participants]			
< 65 years	**244**	**270**	**514**
>=65 years	**93**	**68**	**161**
Gender [units: participants]			
Female	**137**	**157**	**294**
Male	**200**	**181**	**381**

Clinical Trials Data

1. Primary: Overall Survival [Time Frame: From randomization (initiated January 2010) to December 30 2010. Median follow-up time in the vemurafenib group was 3.75 months (range 0.3 to 10.8) and in the dacarbazine group was 2.33 months (range <0.1 to 10.3).]

Measure Type	Primary
Measure Title	Overall Survival
Measure Description	An Overall survival event was defined as death due to any cause. The number of participants with overall survival events is reported.
Time Frame	From randomization (initiated January 2010) to December 30 2010. Median follow-up time in the vemurafenib group was 3.75 months (range 0.3 to 10.8) and in the dacarbazine group was 2.33 months (range <0.1 to 10.3).
Safety Issue	No

Population Description

Explanation of how the number of participants for analysis was determined. Includes whether analysis was per protocol, intention to treat, or another method. Also provides relevant details such as imputation technique, as appropriate.
The intent-to-treat (ITT) population was defined as all randomized participants, whether or not study treatment was received. The ITT population was analyzed according to the treatment assigned at randomization. Overall survival was assessed on participants randomized at least 15 days prior to the clinical cutoff date of December 30, 2010.

Reporting Groups

	Description
Vemurafenib	Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).

Measured Values

	Vemurafenib	**Dacarbazine**
Number of Participants Analyzed [units: participants]	**336**	**336**
Overall Survival [units: participants]		
Participants with events	**43**	**75**
Participants without events	**293**	**261**

Statistical Analysis 1 for Overall Survival

Groups [1]	All groups
Method [2]	Log Rank
P Value [3]	<0.0001
Hazard Ratio (HR) [4]	0.37
95% Confidence Interval	0.26 to 0.55

Clinical Trials Data

2. Primary: Progression-free Survival [Time Frame: From randomization (initiated January 2010) to December 30 2010.]

Measure Type	Primary
Measure Title	Progression-free Survival
Measure Description	A progression-free survival (PFS) event was defined as disease progression or death due to any cause. Tumor response (progression) was assessed according to the Response Evaluation Criteria In Solid Tumors (RECIST) version 1.1 criteria using computed tomography (CT) scans or magnetic resonance imaging (MRI).
Time Frame	From randomization (initiated January 2010) to December 30 2010.
Safety Issue	No

Population Description

Explanation of how the number of participants for analysis was determined. Includes whether analysis was per protocol, intention to treat, or another method. Also provides relevant details such as imputation technique, as appropriate.
The analysis population for PFS consisted of all ITT participants randomized by October 27, 2010 (at least 9 weeks prior to the clinical cutoff date of December 30, 2010). The 9-week interval was chosen to allow time for participants to have had their first scheduled post baseline tumor assessment CT scan.

Reporting Groups

	Description
Vemurafenib	Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).

Measured Values

	Vemurafenib	**Dacarbazine**
Number of Participants Analyzed [units: participants]	**275**	**274**
Progression-free Survival [units: participants]		
Participants with events	**104**	**182**
Participants without events	**171**	**92**

Statistical Analysis 1 for Progression-free Survival

Groups [1]	All groups
Method [2]	Log Rank
P Value [3]	<.0001
Hazard Ratio (HR) [4]	0.26
95% Confidence Interval	0.20 to 0.33

Clinical Trials Data

3. Secondary: Participants With a Best Overall Response (BOR) of Complete Response or Partial Response [Time Frame: From randomization (initiated January 2010) until December 30, 2010]

Measure Type	Secondary
Measure Title	Participants With a Best Overall Response (BOR) of Complete Response or Partial Response
Measure Description	BOR was defined as a complete response (CR) or partial response (PR) confirmed per Response Evaluation Criteria In Solid Tumors (RECIST) version 1.1. Participants who never received study treatment and treated participants without any post-baseline tumor assessments were considered as non-responders. CR: Disappearance of all target lesions, all non-target lesions and no new lesion. Any pathological lymph nodes must have had reduction in the short axis to <10 mm. PR: At least a 30% decrease in the sum of diameters of target lesions, no progression in non-target lesion and no new lesion.
Time Frame	From randomization (initiated January 2010) until December 30, 2010
Safety Issue	No

Population Description

Explanation of how the number of participants for analysis was determined. Includes whether analysis was per protocol, intention to treat, or another method. Also provides relevant details such as imputation technique, as appropriate.
The analysis population consisted of all ITT participants randomized by September 22, 2010 (at least 14 weeks prior to the clinical cutoff date of December 30, 2010). The 14-week interval was chosen as it was the minimum time needed to observe a confirmed overall response according to protocol-specified schedule for the first two tumor assessments.

Reporting Groups

	Description
Vemurafenib	Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).

Measured Values

	Vemurafenib	**Dacarbazine**
Number of Participants Analyzed [units: participants]	**219**	**220**
Participants With a Best Overall Response (BOR) of Complete Response or Partial Response [units: participants]		
Responders	**106**	**12**
Non-responders	**113**	**208**

No statistical analysis provided for Participants With a Best Overall Response (BOR) of Complete Response or Partial Response

Clinical Trials Data

Time Frame	Baseline through the end of study (maximum exposure: 57.07 months)
Additional Description	No text entered.

Reporting Groups

	Description
Vemurafenib	Adverse events reported for this group include those occurring in participants receiving vemurafenib starting at their baseline visit. Participants received continuous oral doses of vemurafenib (RO5185426) 960 mg twice a day. Participants took four 240 mg tablets in the morning and four 240 mg tablets in the evening (960 mg twice a day for a total daily dose of 1920 mg).
Dacarbazine	Adverse events reported for this group include those occurring in participants receiving dacarbazine starting at their baseline visit until study discontinuation or treatment switch. Dacarbazine was administered intravenously 1000 mg/m^2 up to 60 minutes on Day 1 of every 3 weeks (3 weeks was one cycle length).
Vemurafenib After Crossover	Adverse events reported for this group include those occurring following switch to vemurafenib in those participants who switched from dacarbazine to vemurafenib during the study.

Serious Adverse Events

	Vemurafenib	Dacarbazine	Vemurafenib After Crossover
Total, serious adverse events			
# participants affected / at risk	**165/336 (49.11%)**	**52/293 (17.75%)**	**44/84 (52.38%)**
Blood and lymphatic system disorders			
Anaemia † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**0/293 (0.00%)**	**2/84 (2.38%)**
Bone marrow failure † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Lymphadenitis † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Neutropenia † [1]			
# participants affected / at risk	**1/336 (0.30%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Thrombocytopenia † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Cardiac disorders			
Acute myocardial infarction † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**0/293 (0.00%)**	**1/84 (1.19%)**
Atrial fibrillation † [1]			
# participants affected / at risk	**3/336 (0.89%)**	**0/293 (0.00%)**	**0/84 (0.00%)**
Atrial tachycardia † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Cardiac arrest † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**1/293 (0.34%)**	**0/84 (0.00%)**
Cardiac failure † [1]			
# participants affected / at risk	**0/336 (0.00%)**	**0/293 (0.00%)**	**1/84 (1.19%)**

Clinical Trials Data

Results Point of Contact:

Name/Title: Medical Communications
Organization: Hoffman-LaRoche
phone: 800-821-8590

Publications automatically indexed to this study by ClinicalTrials.gov Identifier (NCT Number):

Yamazaki N, Kiyohara Y, Sugaya N, Uhara H. Phase I/II study of vemurafenib in patients with unresectable or recurrent melanoma with BRAF(V) (600) mutations. J Dermatol. 2015 Jul;42(7):661-6. doi: 10.1111/1346-8138.12873. Epub 2015 Apr 17.

Frederick DT, Salas Fragomeni RA, Schalck A, Ferreiro-Neira I, Hoff T, Cooper ZA, Haq R, Panka DJ, Kwong LN, Davies MA, Cusack JC, Flaherty KT, Fisher DE, Mier JW, Wargo JA, Sullivan RJ. Clinical profiling of BCL-2 family members in the setting of BRAF inhibition offers a rationale for targeting de novo resistance using BH3 mimetics. PLoS One. 2014 Jul 1;9(7):e101286. doi: 10.1371/journal.pone.0101286. eCollection 2014.

McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, Ribas A, Hogg D, Hamid O, Ascierto PA, Garbe C, Testori A, Maio M, Lorigan P, Lebbé C, Jouary T, Schadendorf D, O'Day SJ, Kirkwood JM, Eggermont AM, Dréno B, Sosman JA, Flaherty KT, Yin M, Caro I, Cheng S, Trunzer K, Hauschild A. Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. Lancet Oncol. 2014 Mar;15(3):323-32. doi: 10.1016/S1470-2045(14)70012-9. Epub 2014 Feb 7.

Lacouture ME, Duvic M, Hauschild A, Prieto VG, Robert C, Schadendorf D, Kim CC, McCormack CJ, Myskowski PL, Spleiss O, Trunzer K, Su F, Nelson B, Nolop KB, Grippo JF, Lee RJ, Klimek MJ, Troy JL, Joe AK. Analysis of dermatologic events in vemurafenib-treated patients with melanoma. Oncologist. 2013;18(3):314-22. doi: 10.1634/theoncologist.2012-0333. Epub 2013 Mar 1.

Su F, Viros A, Milagre C, Trunzer K, Bollag G, Spleiss O, Reis-Filho JS, Kong X, Koya RC, Flaherty KT, Chapman PB, Kim MJ, Hayward R, Martin M, Yang H, Wang Q, Hilton H, Hang JS, Noe J, Lambros M, Geyer F, Dhomen N, Niculescu-Duvaz I, Zambon A, Niculescu-Duvaz D, Preece N, Robert L, Otte NJ, Mok S, Kee D, Ma Y, Zhang C, Habets G, Burton EA, Wong B, Nguyen H, Kockx M, Andries L, Lestini B, Nolop KB, Lee RJ, Joe AK, Troy JL, Gonzalez R, Hutson TE, Puzanov I, Chmielowski B, Springer CJ, McArthur GA, Sosman JA, Lo RS, Ribas A, Marais R. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. N Engl J Med. 2012 Jan 19;366(3):207-15. doi: 10.1056/NEJMoa1105358.

Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA; BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med. 2011 Jun 30;364(26):2507-16. doi: 10.1056/NEJMoa1103782. Epub 2011 Jun 5.

Responsible Party:	Hoffmann-La Roche
ClinicalTrials.gov Identifier:	NCT01006980 History of Changes
Other Study ID Numbers:	NO25026
	2009-012293-12
Study First Received:	October 30, 2009
Results First Received:	July 29, 2011
Last Updated:	July 5, 2016
Health Authority:	United States: Food and Drug Administration

ImmPort

Bioinformatics For The Future of Immunology

ImmPort is funded by the NIH, NIAID and DAIT in support of the NIH mission to share data with the public. Data shared through ImmPort has been provided by NIH-funded programs, other research organizations and individual scientists ensuring these discoveries will be the foundation of future research.

Private Data

- Upload
- Validator
- Search/Download

Shared Data

- Tutorials
- Gene Lists
- Search/Download

Data Analysis

- Analysis Workflow
- Automated Clustering
- Tutorials

Announcements

Shared Data: 255 Studies; 49319 Subjects; 1141 Experiments; 257 Assessments; 191 Lab Test Panels

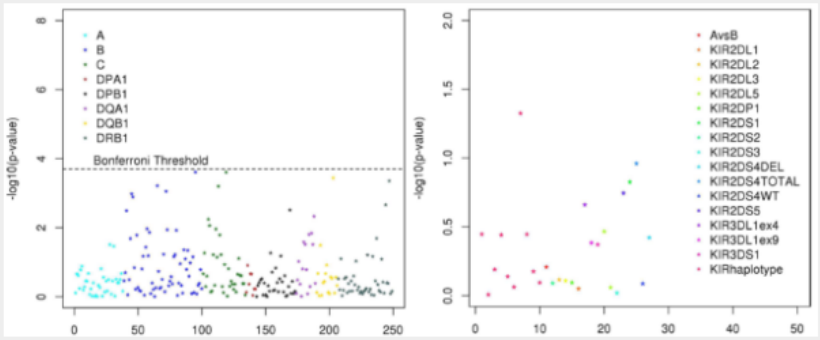
more

June 16, 2017 - ImmPort Data Release 22 is out with 13 new studies shared. New studies include clinical trial data provided by the Auto Immunity Centers of Excellence - see studies SDY625, SDY655, SDY824 and SDY961 for details. HIPC II data shared from Donna Farber Lab at Columbia University looks at CMV-specific T cell

Study: The Immunogenetics of Measles Immunity

Infection by measles virus can lead to rash, fever, encephalitis, and death. The measles vaccine remains the best prevention however immune responses to the vaccine vary greatly. Host genotype is an important determinant in this response with several immunoregulatory genes known to play a role. In this study, measles vaccine response was analyzed with human leukocyte antigen (HLA) and killer cell immunoglobulin-like receptor (KIR) genotypes. While several HLA alleles showed possible associations, KIR alleles were not implicated in measle vaccine response.

PubMed ID: 28158231 Study: SDY839



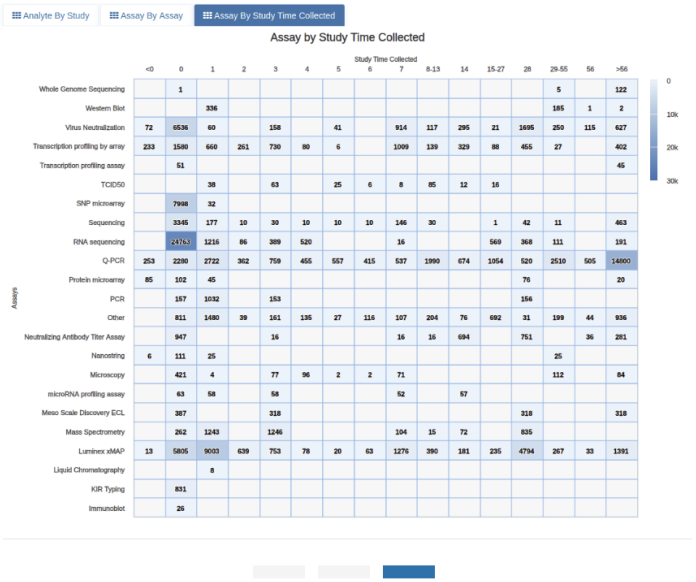
https://immport.niaid.nih.gov/

Data Summary: Release 43, March 2022

Click on the counts with icon to visualize the count breakdown

Studies	Subjects	Diseases
548	82506	130
Experiments	**Total Results**	**Lab Tests**
2023	6298929	1268118

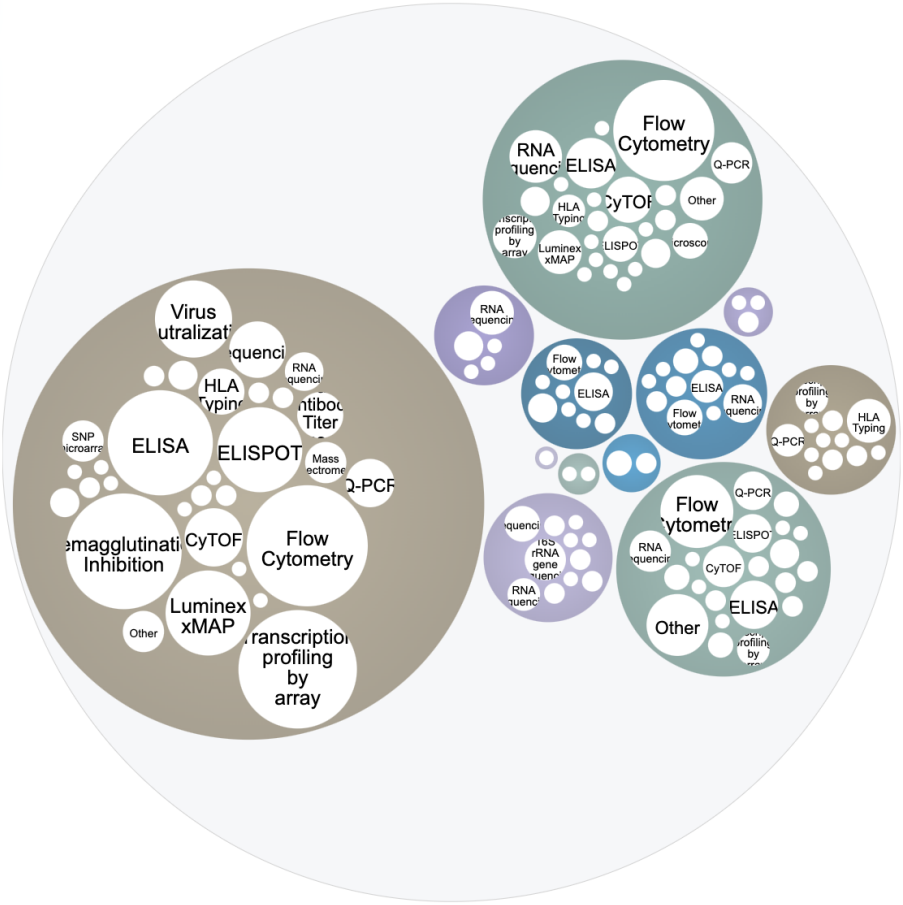
Matrix Summary - Assay By Study Time Collected



Bubble Summary: Research Focus by Assay Type

Click on a bubble to zoom in

Research Focus by Assay Type



https://immport.niaid.nih.gov/

Filter Option

Clinical Trial

- [ ] N (0)
- [x] Y (60)

Study Type

- [ ] Intervention Longitudinal (1)
- [ ] Interventional (40)
- [ ] Longitudinal (0)
- [ ] Observational (19)

Research Focus

- [ ] Atopy/Allergy (16)
- [ ] Autoimmune (14)
- [ ] Immune Response (0)
- [ ] Infection Response (0)
- [ ] Transplantation (27)
- [ ] Vaccine Response (4)

Species

- [ ] Anas platyrhynchos (0)
- [ ] Gallus gallus (0)
- [ ] Homo sapiens (60)
- [ ] Macaca fascicularis (0)
- [ ] Macaca mulatta (0)
- [ ] Mus musculus (0)
- [ ] Mustela putorius furo (0)
- [ ] Sus scrofa domesticus (0)

Biosample Type

- [ ] Bodily fluid (15)
- [ ] Cell (2)
- [ ] Colon (0)
- [ ] DNA (1)
- [ ] Ileum (0)
- [ ] Inguinal lymph node (0)
- [ ] Jejunum (0)
- [ ] Kidney (1)
- [ ] Lung (0)
- [ ] Lung lymph node (0)
- [ ] Mesenteric lymph node (0)
- [ ] Not_Specified (4)
- [ ] Organ (0)
- [ ] Other (6)
- [ ] PBMC (2)
- [ ] Plasma (1)
- [ ] Protein (0)
- [ ] RNA (0)
- [ ] Serum (12)
- [ ] Spleen (0)

Found 60 Studies in 2048 ms

SDY1 — Efficacy and Safety Evaluation of Allergen Immunotherapy Co-Administered with Omalizumab (an anti-IgE Monoclonal Antibody)

Download

A series of allergy shots may reduce symptoms of seasonal ragweed allergies. This study will determine whether taking a drug called omalizumab (also known as Xolair) before getting the allergy shots is more effective than allergy shots alone or other treatments, such as prescription antihistamines.

Thomas Casale, Creighton University School of Medicine

SDY10 — Role of Antimicrobial Peptides in Host Defense Against Vaccinia Virus

Download

Atopic dermatitis (AD) is a chronic inflammatory skin disorder characterized by recurrent viral skin infections. Recent studies have demonstrated that the skin of people with AD my have decreased antimicrobial peptide (AMP) expression. The purpose of this study is to compare small pox virus replication and the number of AMPs and other antiviral molecules in people with AD, as compared to those ...

Donald Leung, National Jewish Health

SDY13 — Analysis of the Response of Subjects with Atopic Dermatitis to Oral Vitamin D3 by Measurement of Antimicrobial Peptide Expression in Skin and Saliva

Download

The goal of the Atopic Dermatitis Vaccinia Network (ADVN) is to research methods for preventing atopic dermatitis (AD) patients from contracting eczema vaccinatum (EV), a potentially fatal complication of smallpox vaccinations. A critical host defense defect uncovered in patients with AD is their apparent relative lack of expression of antimicrobial peptides (AMPs), specifically cathelicidins,...

Donald Leung, National Jewish Health
Jon Hanifin, Oregon Health & Science University
Richard Gallo, University of California at San Diego

SDY131 — Pediatric Kidney Transplant Without Calcineurin Inhibitors (CN01)

Download

The purpose of this study is to see the effect of using drugs other than calcineurin inhibitors to improve the rate of kidney transplant failure.

Kidney transplantation can help children with end-stage kidney disease. However, it has been difficult to find treatment for donor graft rejection that does not have a lot of side effects. Researchers hope to find treatments (immunosuppressan...

Study data available for download for Registered Users

Summary | Design | Adverse Event | Assessment | Interventions | Medications | Demographics | Lab Tests | Mechanistic Assays | Study Files

Accession:	SDY1
Title:	Efficacy and Safety Evaluation of Allergen Immunotherapy Co-Administered with Omalizumab (an anti-IgE Monoclonal Antibody)
PI:	Thomas Casale - Creighton University School of Medicine
Type:	Interventional
Condition Studied:	Seasonal allergy to ragweed
Brief Description:	A series of allergy shots may reduce symptoms of seasonal ragweed allergies. This study will determine whether taking a drug called omalizumab (also known as Xolair) before getting the allergy shots is more effective than allergy shots alone or other treatments, such as prescription antihistamines.
Start Date:	2003-04-01
Schematic: Show	
Detailed Description:	Allergic rhinitis affects 20 to 40 million Americans annually. Allergy symptoms, which can range from mild to seriously debilitating, may affect quality of life. Left untreated, allergic rhinitis can exacerbate or trigger more serious conditions, such as asthma and sinus inflammation. Individuals with allergies react to harmless particles such as dust or pollen. Proteins in the blood called IgE antibodies treat the harmless particles as invaders and trigger an immune system response. The immune response results in harmful inflammation of healthy tissues. In ragweed allergy, inflammation occurs in the airways and causes familiar allergy symptoms like sneezing, coughing, and general discomfort. Omalizumab is an investigational drug that has been shown to block the effects of IgE antibodies. The blocking effect of omalizumab is temporary, but giving the drug to people before their regular allergy shots may make the shots more effective. Participants in this study will be randomly assigned to receive injections of omalizumab or a placebo before an accelerated course of allergy shots (given over 12 weeks). The participants will return for follow-up for up to one year, and they may have as many as 27 study visits.
Objectives:	**Primary Objective:** To examine whether omalizumab given prior to RIT followed by 12 weeks of dual omalizumab and IT is more effective than RIT followed by IT alone in preventing the symptoms of ragweed-induced SAR. **Secondary Objective:** To examine whether omalizumab given prior to RIT followed by 12 weeks of dual omalizumab and IT is safe and more effective than omalizumab alone or placebo in preventing the symptoms of ragweed-induced SAR; to assess the immunologic mechanisms

Study data available for download for Registered Users

Summary | Design | Adverse Event | Assessment | Interventions | Medications | Demographics | Lab Tests | Mechanistic Assays | Study Files

Arms or Cohorts

Accession	Name	Description	Population Selection Rule
ARM4	Immunotherapy with anti-IgE	Omalizumab pre-treatment, ragweed RIT, omalizumab + ragweed IT	Randomized 1:1:1:1 to 4 treatment groups
ARM3	Placebo Immunotherapy with anti-IgE	Omalizumab pre-treatment, placebo RIT, omalizumab + placebo IT	Randomized 1:1:1:1 to 4 treatment groups
ARM2	Immunotherapy with placebo anti-IgE	Placebo omalizumab pre-treatment, ragweed RIT, placebo omalizumab + ragweed IT	Randomized 1:1:1:1 to 4 treatment groups
ARM1	Placebo Immunotherapy with placebo anti-IgE	Placebo omalizumab pre-treatment, placebo RIT, placebo omalizumab + placebo IT	Randomized 1:1:1:1 to 4 treatment groups

Inclusion Exclusion Criteria

Criteria Category	Criteria
Inclusion	A positive skin test by prick method to ragweed pollen at Visit -01. A positive skin prick test will be defined as a ragweed pollen-induced wheal >3 mm larger in diameter than diluent control (measurements will be made 15-20 minutes after application).
Inclusion	Able to comprehend and grant a witnessed, written informed consent prior to any study procedures.
Inclusion	Female participants of child bearing age must have a negative urine pregnancy test at Visit -01 and a negative urine pregnancy test at subsequent visits. In addition, female participants must be using a medically acceptable form of birth control.
Inclusion	History of seasonal allergic rhinitis for at least 2 years with symptoms during the ragweed pollen season requiring pharmacotherapy.
Inclusion	Male or female 18 to 50 years of age.
Inclusion	Must be capable of faithfully completing the diary and of attending regularly scheduled study visits.
Inclusion	Must intend to remain in the ragweed pollen area during the entire ragweed season.
Inclusion	Participants must have a baseline serum IgE level > 10 and < 700 IU/mL.
Inclusion	Participants must meet pretrial eligibility requirements for trial enrollment (acceptable medical history, physical examination results, normal electrocardiogram and acceptable laboratory test results).
Inclusion	Willing to avoid prohibited medications for the periods indicated in the protocol.
Exclusion	Asthma (either history of, abnormal spirometry, [FEV1 <80% predicted] or use of asthma medications).
Exclusion	Chronic or intermittent use of inhaled, oral, intra-muscular, or intra-venous corticosteroids; or chronic or intermittent use of topical

Study data available for download for Registered Users

Adverse Event Summary

Show 10 entries Search:

Totals By	ARM4	ARM3	ARM2	ARM1
Grade 1 Mild Adverse Events	315	311	296	268
Grade 2 Moderate Adverse Events	122	127	150	119
Grade 3 Severe Adverse Events	29	40	46	28
Grade 4 Life Threatening or Disabling Adverse Events	0	1	1	0
Grade 5 Death Related to Adverse Events	0	0	0	0
Subjects	39	40	40	40
Subjects with Adverse Events	39	40	40	39
Total Adverse Events	466	479	493	415

Showing 1 to 8 of 8 entries Previous 1 Next

Adverse Event Detail

Show 10 entries Search:

Name Reported	Severity	Total Count	ARM4	ARM3	ARM2	ARM1
(ASSOCIATED WITH SINUSITIS DIAGNOSIS) HEADACHES INCREASED IN FREQUENCY	Grade 1 Mild Adverse Event	1		1		
(L) EXTERNAL AUDITORY CANAL IRRITATION WITH ERYTHEMA AND EXCORIATION	Grade 1 Mild Adverse Event	1				1
(L) EYELID TWITCHING, INTERMITTENT	Grade 1 Mild Adverse Event	1		1		
(L) HAND PAIN	Grade 1 Mild Adverse Event	1	1			
(L) NARE EDEMA OF TURBINATES	Grade 1 Mild Adverse Event	1		1		
(L) NASAL POLYP 70% OCCLUSION	Grade 2 Moderate Adverse Event	1				1
(L) TURBINATE EDEMA	Grade 2 Moderate Adverse Event	1				1

Study data available for download for Registered Users

Assessment Summary

Show 10 entries Search:

Assessment Name Reported	Totals By	ARM4	ARM3	ARM2	ARM1
15 mins post injection allergy skin reaction measurement	Subjects	38	40	40	40
15 mins post injection allergy skin reaction measurement	Assessment Components	2,190	2,372	2,278	2,228
24 hrs post injection allergy skin reaction measurement	Subjects	38	40	40	40
24 hrs post injection allergy skin reaction measurement	Assessment Components	2,190	2,372	2,278	2,228
Allergen History	Subjects	39	40	40	40
Allergen History	Assessment Components	390	400	400	400
Allergy Symptom History	Subjects	39	40	40	40
Allergy Symptom History	Assessment Components	273	280	280	280
Animal Exposure History	Subjects	39	40	40	40
Animal Exposure History	Assessment Components	150	129	131	150

Showing 1 to 10 of 24 entries

Previous 1 2 3 Next

Assessment Component List

Show 10 entries Search:

Assessment	Assessment Component
15 mins post injection allergy skin reaction measurement	Injection 1-A(1-10000000) ERYTH measurement
15 mins post injection allergy skin reaction measurement	Injection 1-A(1-10000000) Wheal measurement
15 mins post injection allergy skin reaction measurement	Injection 2-A(1-1000000) ERYTH measurement
15 mins post injection allergy skin reaction measurement	Injection 2-A(1-1000000) Wheal measurement
15 mins post injection allergy skin reaction measurement	Injection 3-A(1-100000) ERYTH measurement
15 mins post injection allergy skin reaction measurement	Injection 3-A(1-100000) Wheal measurement

Clinical Trials Data

Study SDY1

Previous | Next | Open | Download | Close

Study data available for download for Registered Users

Summary | Design | Adverse Event | Assessment | Interventions | Medications | Demographics | Lab Tests | Mechanistic Assays | Study Files

Interventions

Show 10 entries Search:

Intervention Name	Compound Name	Total Count	ARM4	ARM3	ARM2	ARM1
Immunotherapy	Ragweed Amb a 1	64	33		31	
Omalizumab injection	Omalizumab	79	39	40		
Omalizumab/Placebo injection	Excipients and diluents of omalizumab	80			40	40
Placebo for Immunotherapy	Histamine	71		35		36
Placebo for Rush Immunotherapy	Histamine	74		37		37
Rush Immunotherapy	Ragweed Amb a 1	75	36		39	

Showing 1 to 6 of 6 entries

Previous 1 Next

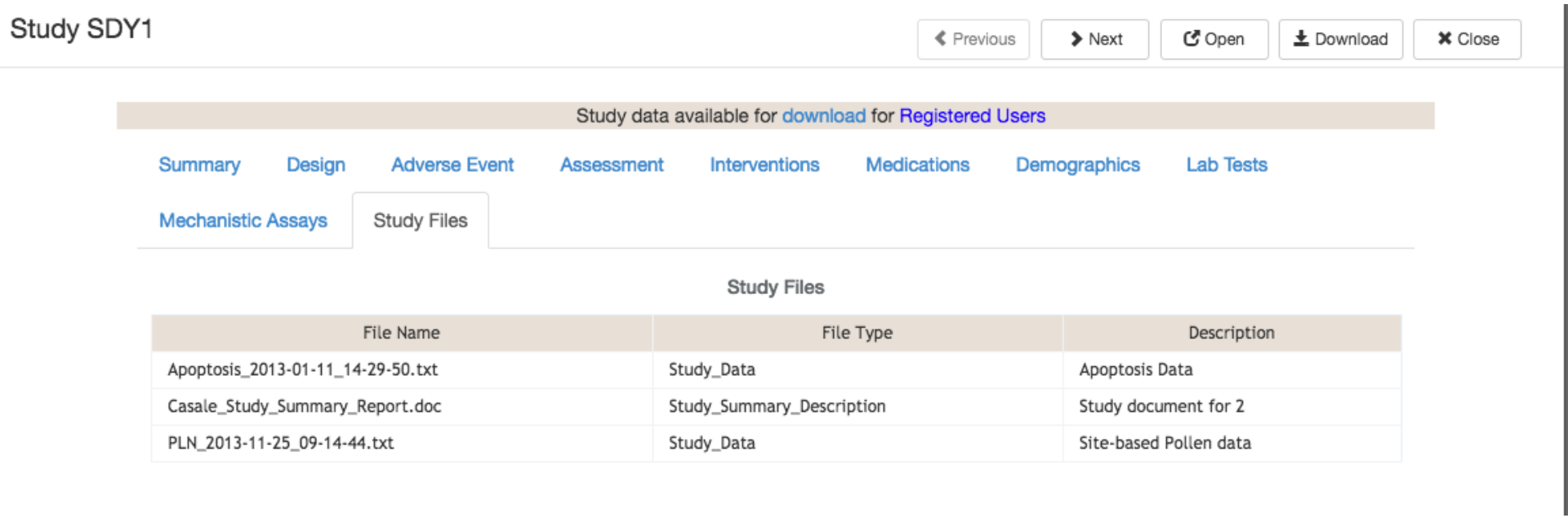
ARM4 = Immunotherapy with anti-IgE

ARM3 = Placebo Immunotherapy with anti-IgE

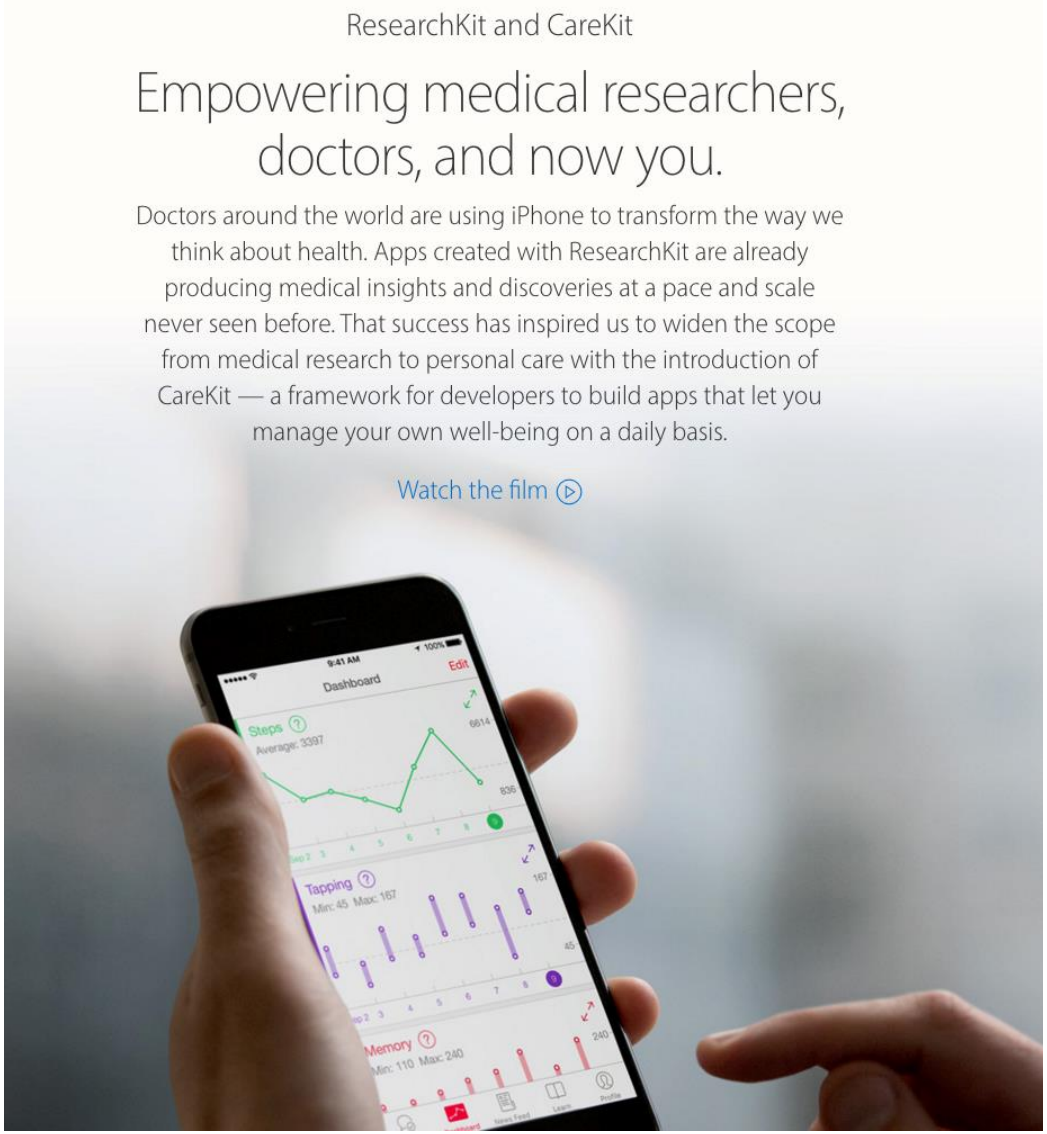
ARM2 = Immunotherapy with placebo anti-IgE

ARM1 = Placebo Immunotherapy with placebo anti-IgE

Clinical Trials Data



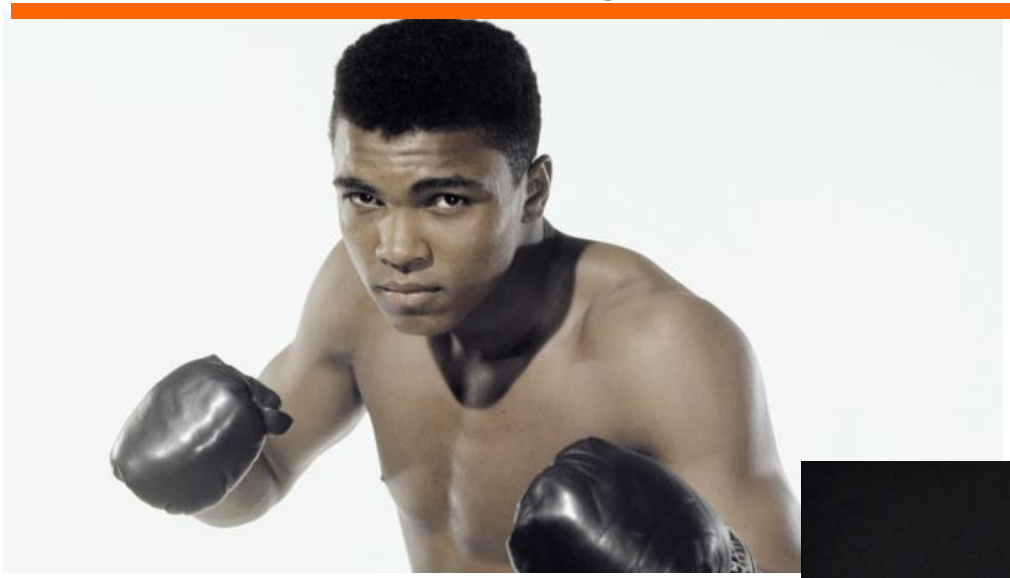
Apple ResearchKit (mobile data)



http://www.apple.com/researchkit/

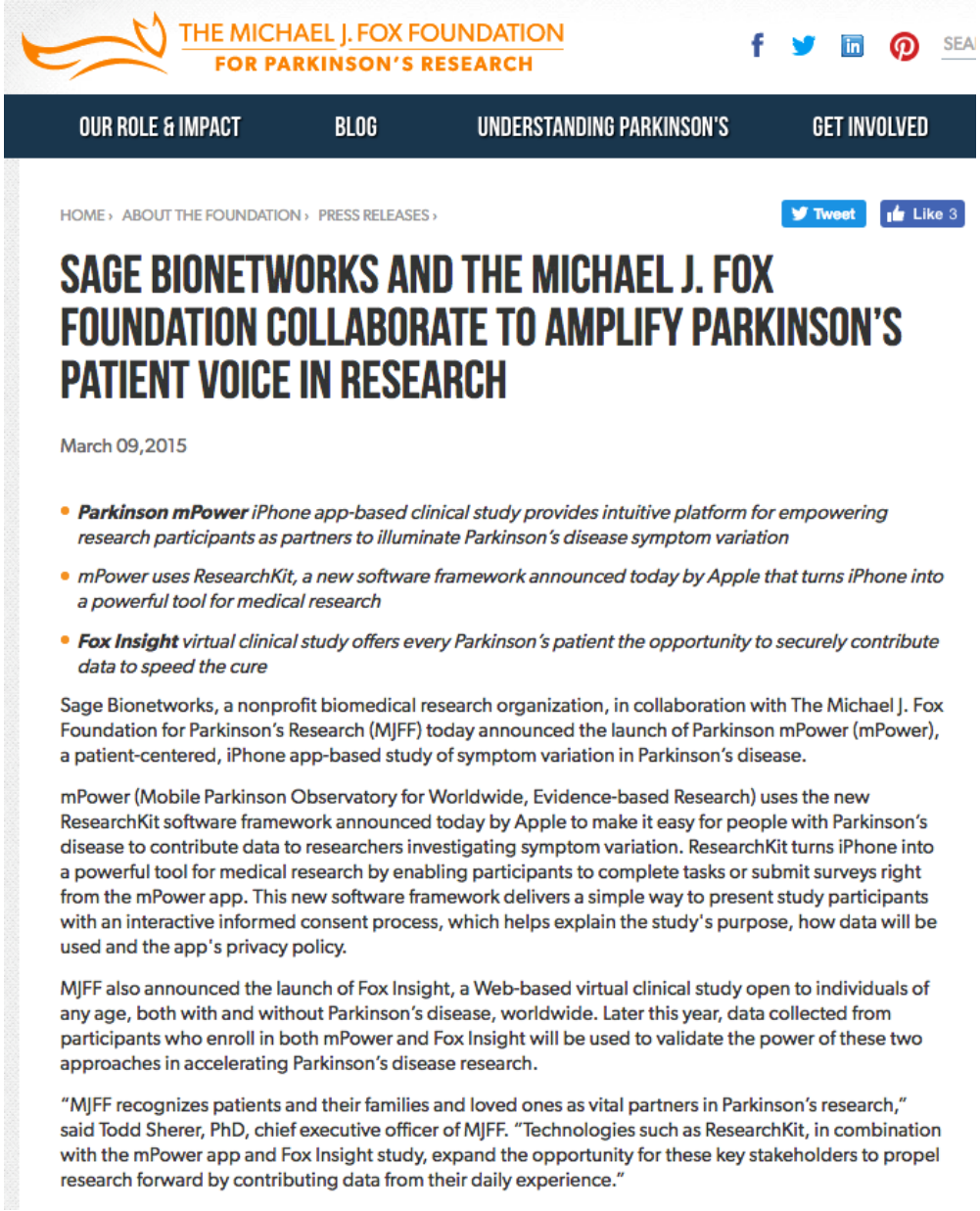
http://images.apple.com/media/us/researchkit/2016/a63aa7d4_e6fd_483f_a59d_d962016c8093/films/carekit/researchkit-carekit-cc-us-20160321_960x540.mp4

Case Study – Parkinson's Disease



http://www.history.com/this-day-in-history/muhammad-ali-refuses-army-induction
https://www.michaeljfox.org/foundation/news.html?tagid=12

mPower – ResearchKit Apps



THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

SEA

OUR ROLE & IMPACT | BLOG | UNDERSTANDING PARKINSON'S | GET INVOLVED

HOME › ABOUT THE FOUNDATION › PRESS RELEASES ›

Tweet Like 3

SAGE BIONETWORKS AND THE MICHAEL J. FOX FOUNDATION COLLABORATE TO AMPLIFY PARKINSON'S PATIENT VOICE IN RESEARCH

March 09,2015

- ***Parkinson mPower*** *iPhone app-based clinical study provides intuitive platform for empowering research participants as partners to illuminate Parkinson's disease symptom variation*
- *mPower uses ResearchKit, a new software framework announced today by Apple that turns iPhone into a powerful tool for medical research*
- ***Fox Insight*** *virtual clinical study offers every Parkinson's patient the opportunity to securely contribute data to speed the cure*

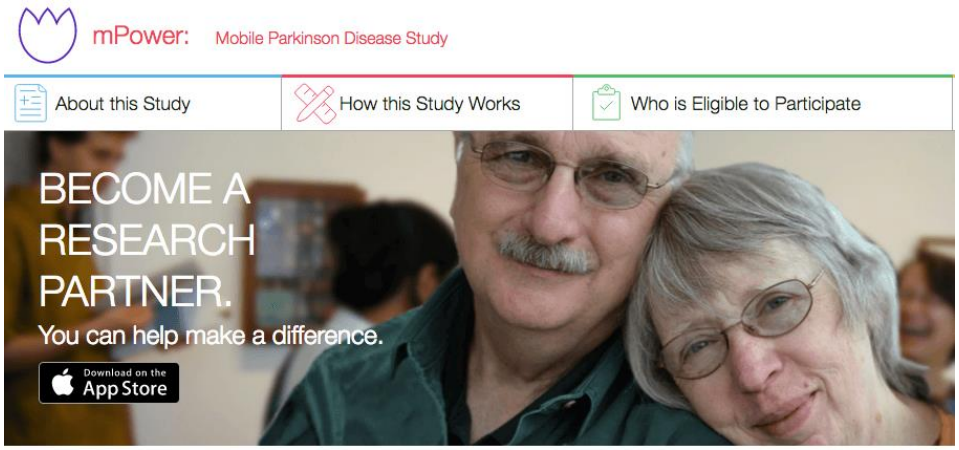
Sage Bionetworks, a nonprofit biomedical research organization, in collaboration with The Michael J. Fox Foundation for Parkinson's Research (MJFF) today announced the launch of Parkinson mPower (mPower), a patient-centered, iPhone app-based study of symptom variation in Parkinson's disease.

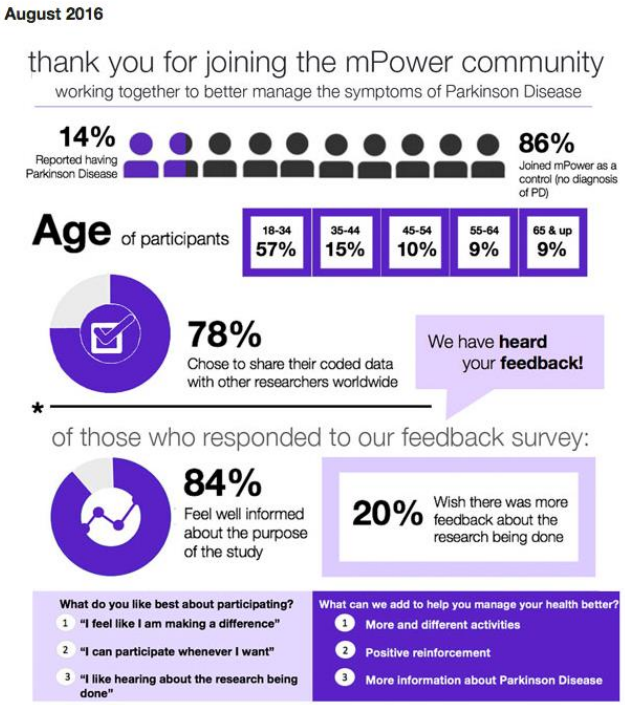
mPower (Mobile Parkinson Observatory for Worldwide, Evidence-based Research) uses the new ResearchKit software framework announced today by Apple to make it easy for people with Parkinson's disease to contribute data to researchers investigating symptom variation. ResearchKit turns iPhone into a powerful tool for medical research by enabling participants to complete tasks or submit surveys right from the mPower app. This new software framework delivers a simple way to present study participants with an interactive informed consent process, which helps explain the study's purpose, how data will be used and the app's privacy policy.

MJFF also announced the launch of Fox Insight, a Web-based virtual clinical study open to individuals of any age, both with and without Parkinson's disease, worldwide. Later this year, data collected from participants who enroll in both mPower and Fox Insight will be used to validate the power of these two approaches in accelerating Parkinson's disease research.

"MJFF recognizes patients and their families and loved ones as vital partners in Parkinson's research," said Todd Sherer, PhD, chief executive officer of MJFF. "Technologies such as ResearchKit, in combination with the mPower app and Fox Insight study, expand the opportunity for these key stakeholders to propel research forward by contributing data from their daily experience."

mPower – ResearchKit Apps





mPower – ResearchKit Apps

About this Study

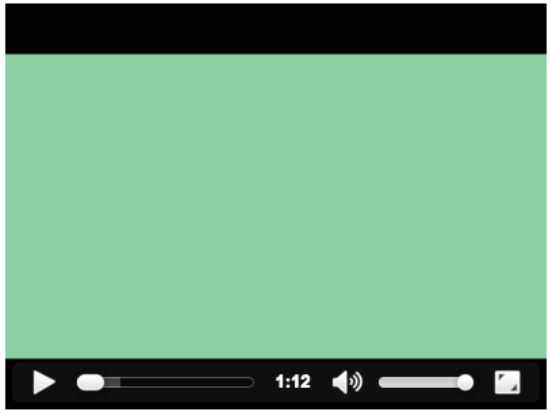
How can we better manage the symptoms of Parkinson's disease (PD) together? Whether you have PD, are touched by someone who has or has had PD or you want to help, we invite you to participate in this study. Become a research partner!

Sage Bionetworks (nonprofit) is proposing a new approach to monitor health in PD using a mobile app. We want to understand why some people with PD have different symptoms than other people with PD, why a person's symptoms and side effects can vary over time, and what can be done to help manage these differences in symptoms day to day.

Learn More

Frequently Asked Questions

How this Study Works



The mPower application uses a mix of surveys and tasks that activate phone sensors to collect and track health and symptoms of PD progression - like dexterity, balance or gait. Our goals are to learn about the variations of PD, to improve the way we describe and manage these variations, and to learn whether mobile devices and sensors can help measure PD and its progression to ultimately improve the quality of life for people with PD.

Learn More

Download the mobile app

The mobile app will help you log your symptoms.

Give consent to enroll

Understand the risks and benefits of participating.

Read the consent form.

Perform simple tasks

We'll ask you to do a few tasks and answer some questions about your health.

Track your health

You can use the health dashboard to track your health data.

Scientists make discoveries

Scientists will use your data to make breakthroughs in medical research and treatments.

mPower – ResearchKit Apps

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Research data
» Neurology
» Parkinson's disease
» Medical research

Received: 07 December 2015
Accepted: 02 February 2016
Published: 3 March 2016

The mPower study, Parkinson disease mobile data collected using ResearchKit

Brian M. Bot[1], Christine Suver[1], Elias Chaibub Neto[1], Michael Kellen[1], Arno Klein[1], Christopher Bare[1], Megan Doerr[1], Abhishek Pratap[1], John Wilbanks[1], E. Ray Dorsey[2], Stephen H. Friend[1] & Andrew D. Trister[1]

Current measures of health and disease are often insensitive, episodic, and subjective. Further, these measures generally are not designed to provide meaningful feedback to individuals. The impact of high-resolution activity data collected from mobile phones is only beginning to be explored. Here we present data from mPower, a clinical observational study about Parkinson disease conducted purely through an iPhone app interface. The study interrogated aspects of this movement disorder through surveys and frequent sensor-based recordings from participants with and without Parkinson disease. Benefitting from large enrollment and repeated measurements on many individuals, these data may help establish baseline variability of real-world activity measurement collected via mobile phones, and ultimately may lead to quantification of the ebbs-and-flows of Parkinson symptoms. App source code for these data collection modules are available through an open source license for use in studies of other conditions. We hope that releasing data contributed by engaged research participants will seed a new community of analysts working collaboratively on understanding mobile health data to advance human health.

Design Type(s)	observation design • time series design • repeated measure design
Measurement Type(s)	disease severity measurement
Technology Type(s)	Patient Self-Report
Factor Type(s)	
Sample Characteristic(s)	Homo sapiens

mPower – ResearchKit Apps

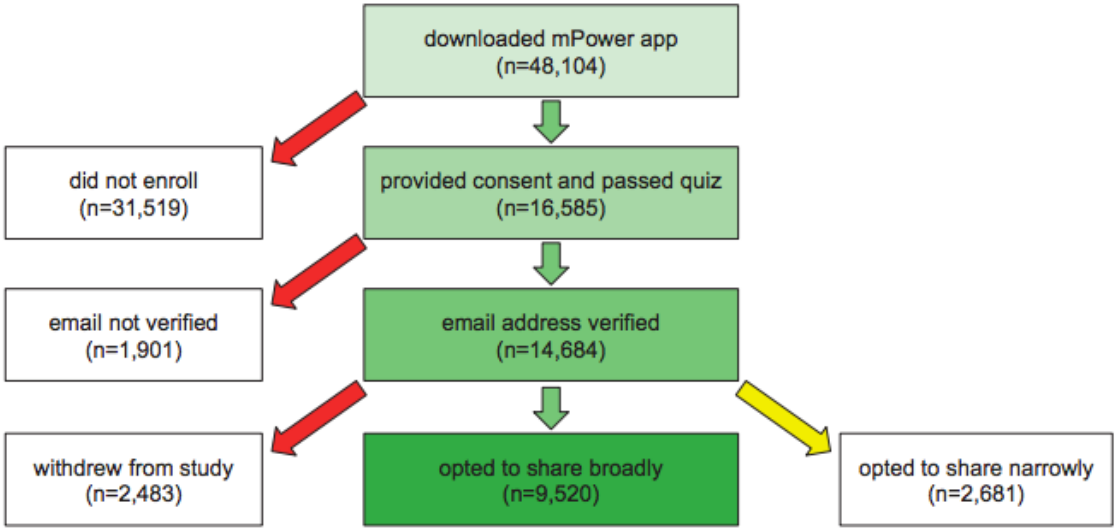


Figure 1. mPower study cohort description.

Task name	Type of task and schedule	Citation	unique participants	unique tasks
Demographics	Survey—once	Data Citation 1	6,805	6,805
PDQ8	Survey—monthly	Data Citation 2	1,334	1,641
UPDRS	Survey—monthly	Data Citation 3	2,024	2,305
Memory	Activity—t.i.d.	Data Citation 4	968	8,569
Tapping	Activity—t.i.d.	Data Citation 5	8,003	78,887
Voice	Activity—t.i.d.	Data Citation 6	5,826	65,022
Walking	Activity—t.i.d.	Data Citation 7	3,101	35,410

Table 1. Data available for each survey and activity completed by study participants.

HealthMap

About

http://www.healthmap.org/site/about

HealthMap, a team of researchers, epidemiologists and software developers at Boston Children's Hospital founded in 2006, is an established global leader in utilizing online informal sources for disease outbreak monitoring and real-time surveillance of emerging public health threats. The freely available Web site 'healthmap.org' and mobile app 'Outbreaks Near Me' deliver real-time intelligence on a broad range of emerging infectious diseases for a diverse audience including libraries, local health departments, governments, and international travelers. HealthMap brings together disparate data sources, including online news aggregators, eyewitness reports, expert-curated discussions and validated official reports, to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health. Through an automated process, updating 24/7/365, the system monitors, organizes, integrates, filters, visualizes and disseminates online information about emerging diseases in nine languages, facilitating early detection of global public health threats. Download our brochure to learn more.

HealthMap

Alert Sources

HealthMap's content is aggregated from freely available information from the following sources. Use of their logos or trademarks by HealthMap is intended only to refer specifically to the respective service; it does not imply any endorsement or affiliation.

ProMED Mail
Program for Monitoring Emerging Diseases, a program of the International Society for Infectious Diseases.

World Health Organization
The United Nations specialized agency for health.

GeoSentinel
Clinician-based sentinel surveillance of individual travelers from the International Society of Travel Medicine and CDC.

OIE - World Organisation for Animal Health
The intergovernmental organisation responsible for improving animal health worldwide.

FAO - Food and Agriculture Organization of the United Nations
An intergovernmental organization for ensuring worldwide food quality and agricultural productivity.

EuroSurveillance
Peer-reviewed European information on communicable disease surveillance and control. Published by the European Centre for Disease Prevention and Control.

Google News
A commercial news aggregation service provided by Google.

Moreover
A commercial news feed aggregation service provided by VeriSign.

Wildlife Data Integration Network
A news feed from WDIN's Global Wildlife Disease News Map. WDIN is a project at the Unversity of Wisconsin - Madison, School of Veterinary Medicine.

Baidu News 新闻
A Chinese language commercial news aggregation service provided by Baidu, the number 1 search engine in China.

SOSO Info 资讯
A Chinese language commercial news aggregation service provided by the Chinese search engine Soso.

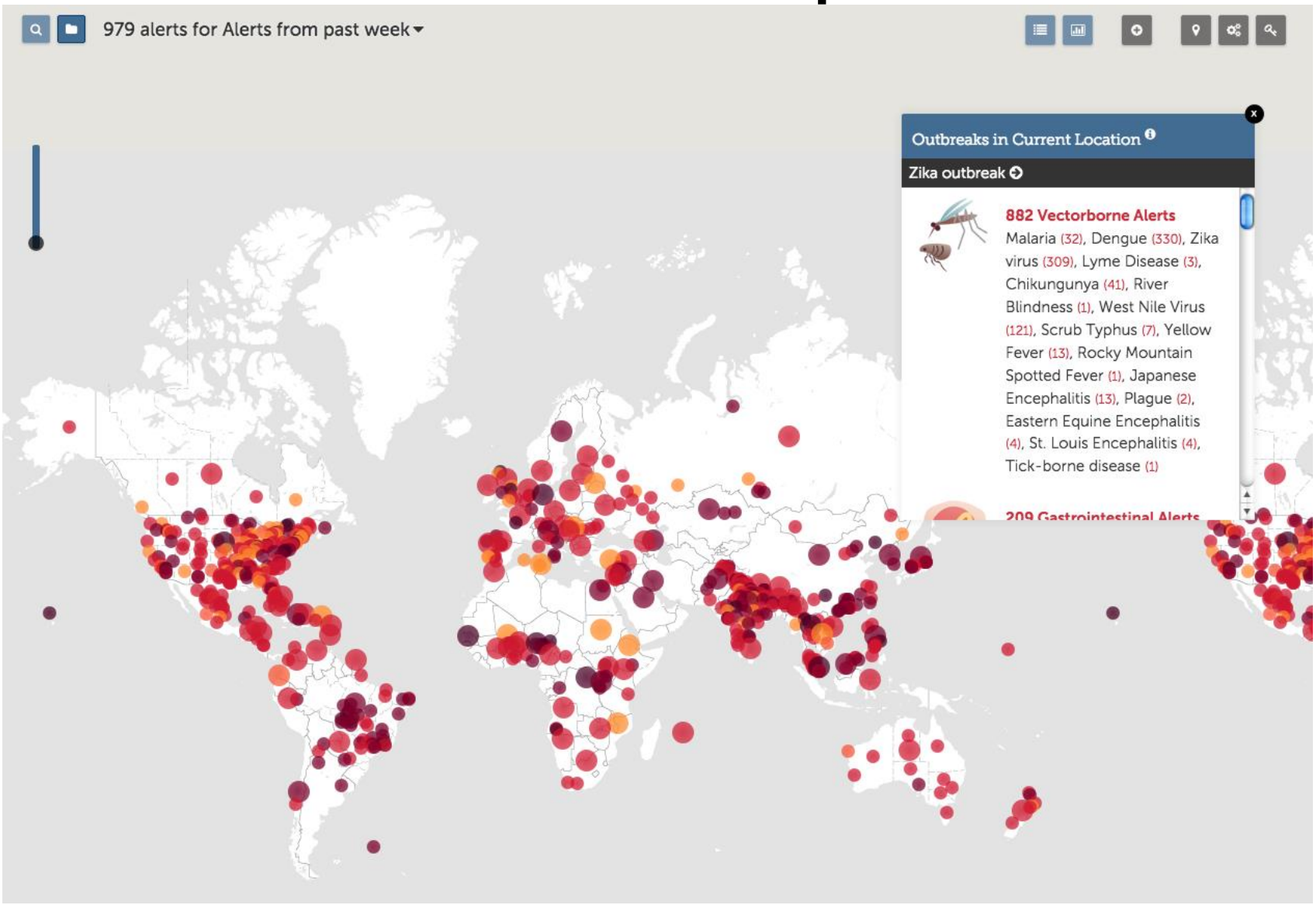
Software Tools

HealthMap is a Linux/Apache/MySQL/PHP application and relies on the following open products. Special thanks to their authors.

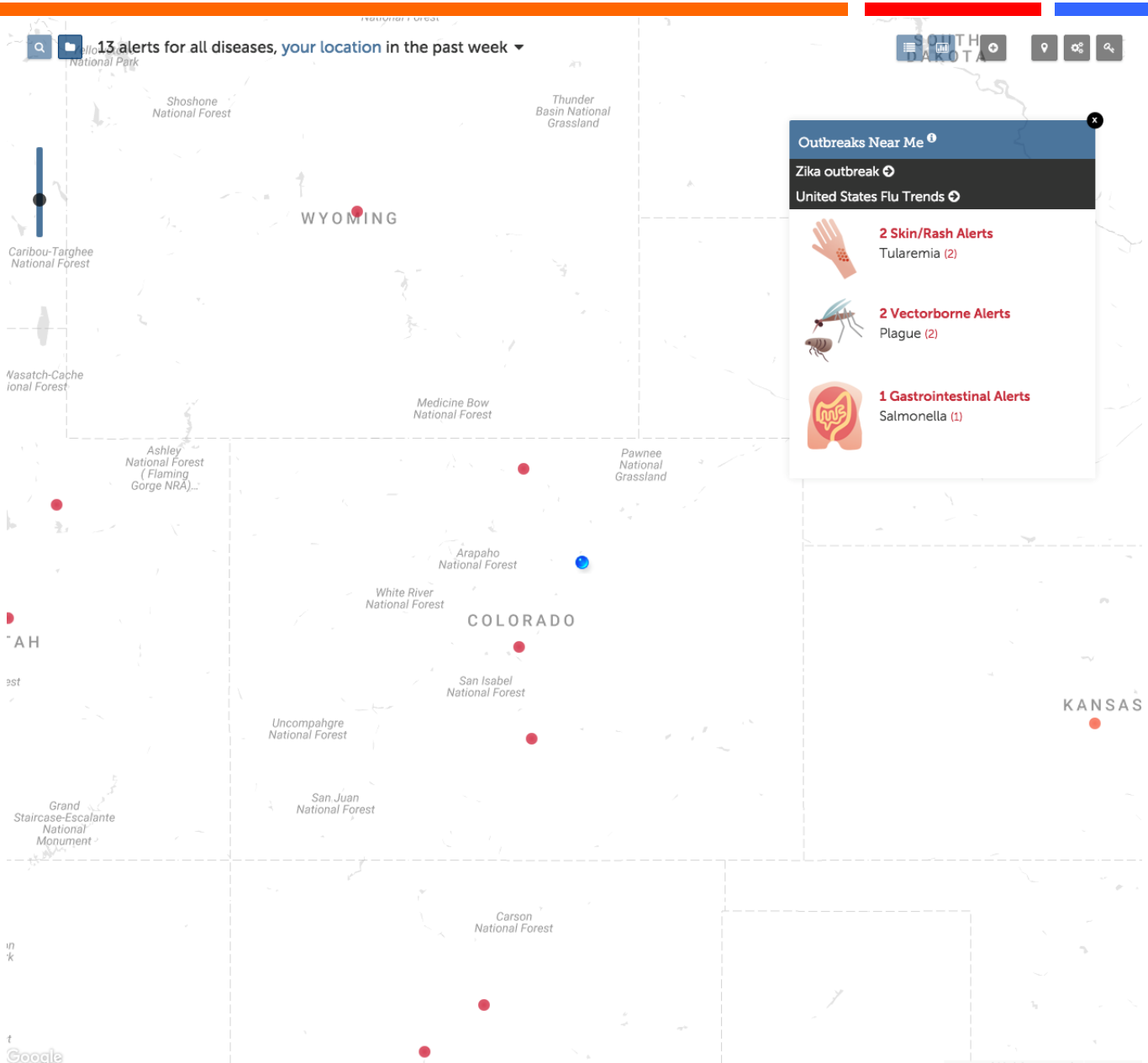
- Google Maps
- GoogleMapAPI for PHP
- Google Translate API
- xajax PHP AJAX library

HealthMap also uses Fisher-Robinson Bayesian filtering, as described by Gary Robinson in A Statistica Approach to the Spam Problem.

HealthMap


979 alerts for Alerts from past week
Outbreaks in Current Location
Zika outbreak
882 Vectorborne Alerts
Malaria (32), Dengue (330), Zika virus (309), Lyme Disease (3), Chikungunya (41), River Blindness (1), West Nile Virus (121), Scrub Typhus (7), Yellow Fever (13), Rocky Mountain Spotted Fever (1), Japanese Encephalitis (13), Plague (2), Eastern Equine Encephalitis (4), St. Louis Encephalitis (4), Tick-borne disease (1)
209 Gastrointestinal Alerts

HealthMap


13 alerts for all diseases, your location in the past week
Outbreaks Near Me
Zika outbreak
United States Flu Trends
2 Skin/Rash Alerts
Tularemia (2)
2 Vectorborne Alerts
Plague (2)
1 Gastrointestinal Alerts
Salmonella (1)
WYOMING
COLORADO
KANSAS
Shoshone National Forest
Thunder Basin National Grassland
Caribou-Targhee National Forest
Medicine Bow National Forest
Ashley National Forest (Flaming Gorge NRA)...
Pawnee National Grassland
Arapaho National Forest
White River National Forest
San Isabel National Forest
Uncompahgre National Forest
San Juan National Forest
Grand Staircase-Escalante National Monument
Carson National Forest
Google

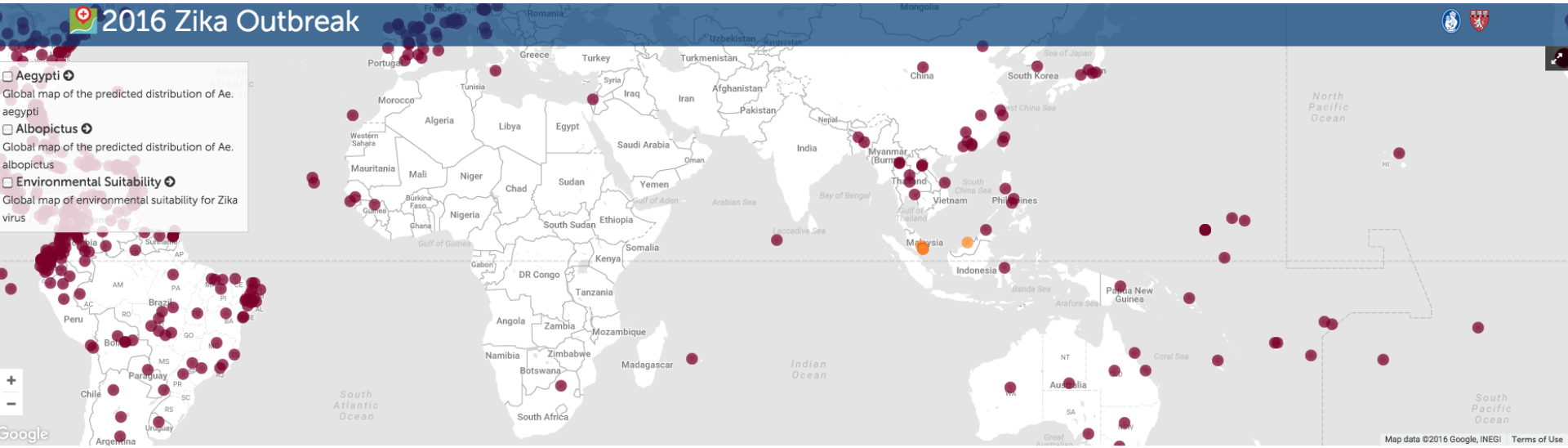
HealthMap

Display 5 results

Filter Results

Source	Date	Summary	Disease	Location	Species	Cases	Deaths	Significance
G	7 Sep 2016	Larimer's skyrocketing West Nile tally an unsolved mystery - The ...	West Nile Virus	Larimer County, Colorado, United States	Humans	28		★★★☆☆
	6 Sep 2016	PRO/PL> Bacterial leaf streak, maize - USA: 1st rep ...	Other Plant Disease	Colorado, United States	Crops			★★☆☆☆
G	6 Sep 2016	Recent Salmonella Outbreak in Utah Linked with Raw Milk - ...	Salmonella	Wasatch County, Utah, United States	Humans			★★★☆☆
G	6 Sep 2016	Recent Salmonella Outbreak in Utah Linked with Raw Milk - ...	Salmonella	Utah, United States	Humans	9		★★★☆☆
	6 Sep 2016	#USA, #Utah: Avoid Possible #Exposure to #Rabies by Avoiding #Bats ...	Rabies	Utah, United States	Bats	7		★☆☆☆☆

HealthMap


2016 Zika Outbreak
Aegypti
Global map of the predicted distribution of Ae. aegypti
Albopictus
Global map of the predicted distribution of Ae. albopictus
Environmental Suitability
Global map of environmental suitability for Zika virus
Map data ©2016 Google, INEGI
Terms of Use

HealthMap

September 07, 2016

Countries with confirmed local transmission: Chile, Brazil, Colombia, Suriname, El Salvador, Mexico, Panama, Venezuela, Honduras, French Guiana, Martinique, Puerto Rico, Bolivia, Saint Martin, Haiti, Barbados, U.S. Virgin Islands, Dominican Republic, Nicaragua, Jamaica, Curacao, Costa Rica, Republic of Trinidad and Tobago, Aruba, Bonaire, Saint Vincent and the Grenadines, France*, Canada*, New Caledonia, Sint Maarten, Laos, Philippines, Italy*, Cuba, Dominica, Bangladesh, Vietnam, Saint Lucia, Belize, Papua New Guinea, Portugal*, Republic of Nauru, Grenada, Peru, Saint Barthélemy, Germany*, Argentina, Anguilla, Spain*, Guinea-Bissau, Sint Eustatius, Saba, Turks and Caicos, Antigua and Barbuda, United States, Cayman Islands, The Bahamas, Singapore, British Virgin Islands, Malaysia

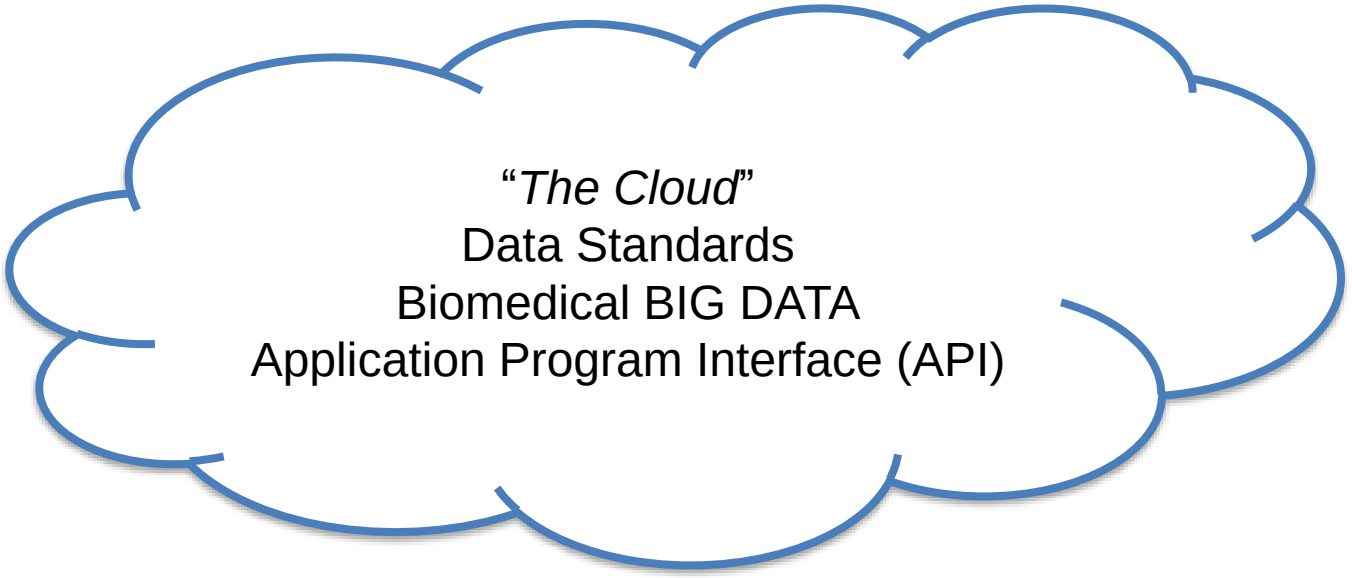
Malaysia reports new case of infection in women zika ... - Radio Havana Cuba

- Malaysia is reporting its third Zika virus case - the newest patient a pregnant woman from Johor.

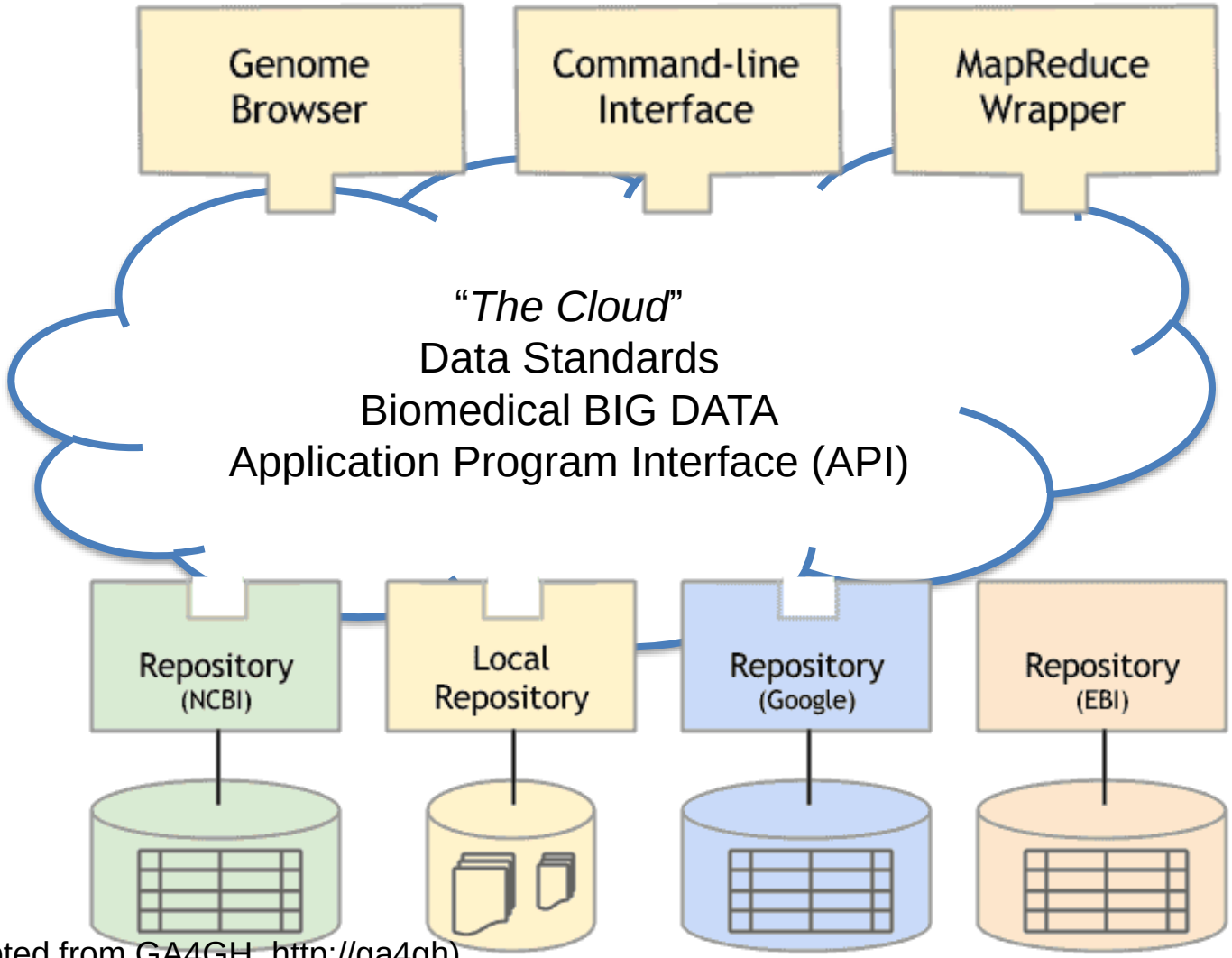
ZIKA VIRUS UPDATE: As of 12pm, 7 September, MOH has confirmed eight new cases of locally transmitted Zika virus... https://t.co/FSvCbil0Cl

- Singapore reports eight new cases:
- "As of 12pm, 7 September, MOH has confirmed eight new cases of locally transmitted Zika virus infection in Singapore. Of these, two cases are linked to the Aljunied Crescent/ Sims Drive/ Kallang Way/ Paya Lebar Way cluster, and one case is linked to the Bishan Street 12 cluster. There is a potential new cluster involving one previously reported case and a new case today. They both live in the Elite Terrace area."
- Total estimated to be 283.

Ideal World of Biomedical Big data

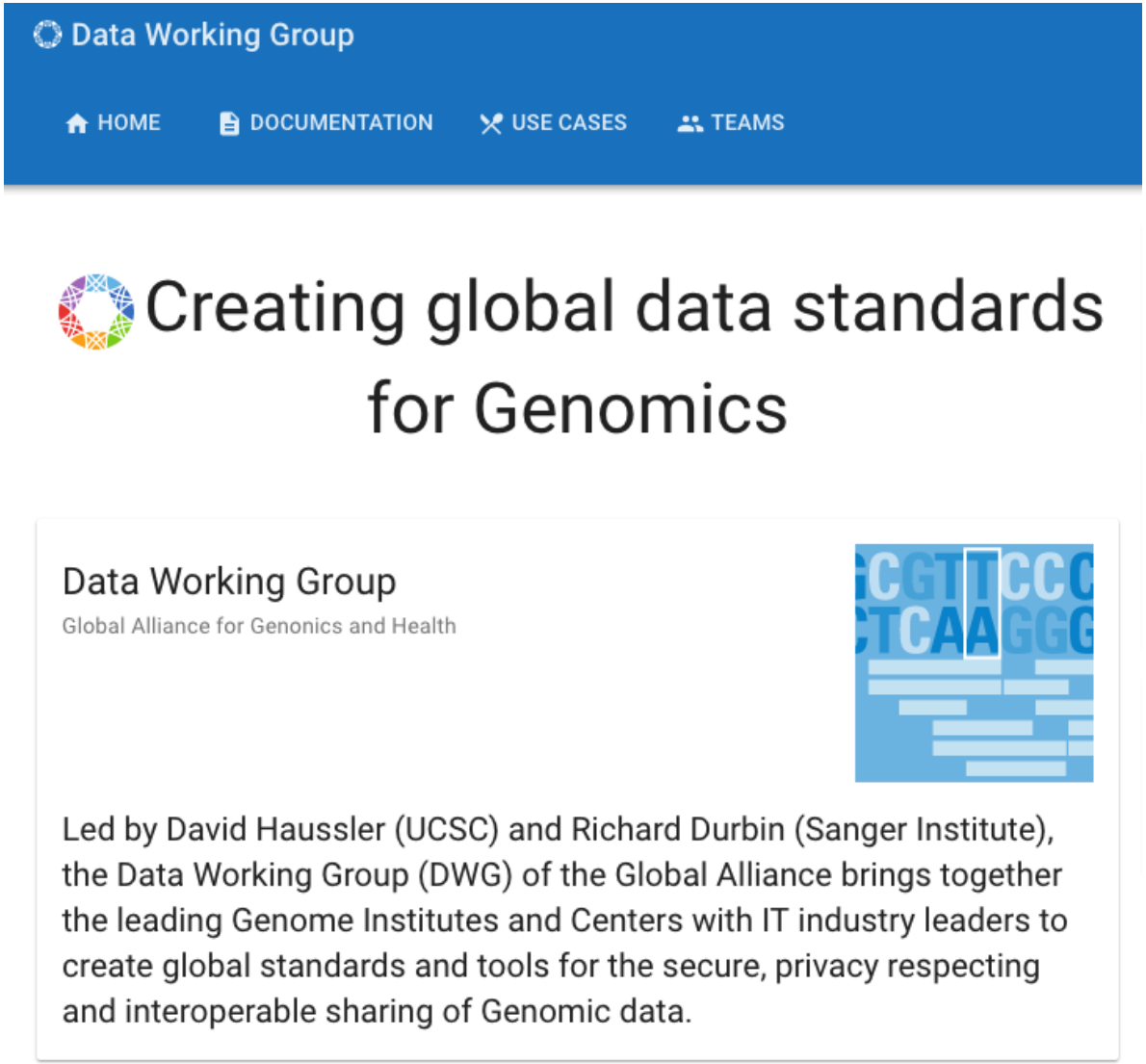


Ideal World of Biomedical Big data
Interoperability



(Image adapted from GA4GH, http://ga4gh)

Global Efforts in Creating Data Standards for Genomics



Conclusions

- Use big data to generate *hypothesis*
- Use standards for interoperability
- Share your research data and program
- Empower data driven research
- Think outside the box – use big data and AI/ML to find unexpected and interesting *knowledge*