Biomedical Data Sources and Integration
The Power of Big Data

Aik Choon Tan

aikchoon.tan@ucdenver.edu

9/7/2018

http://tanlab.ucdenver.edu/labHomePage/teaching/BSBT6111/

Outline

- Why Big Data is important in your biomedical research?
- Examples of Biomedical Big Data
  - GEO
  - CMAP, LINCS
  - Clinical Trials
  - Genomics and Phenotype data
  - Mobile Data
  - 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

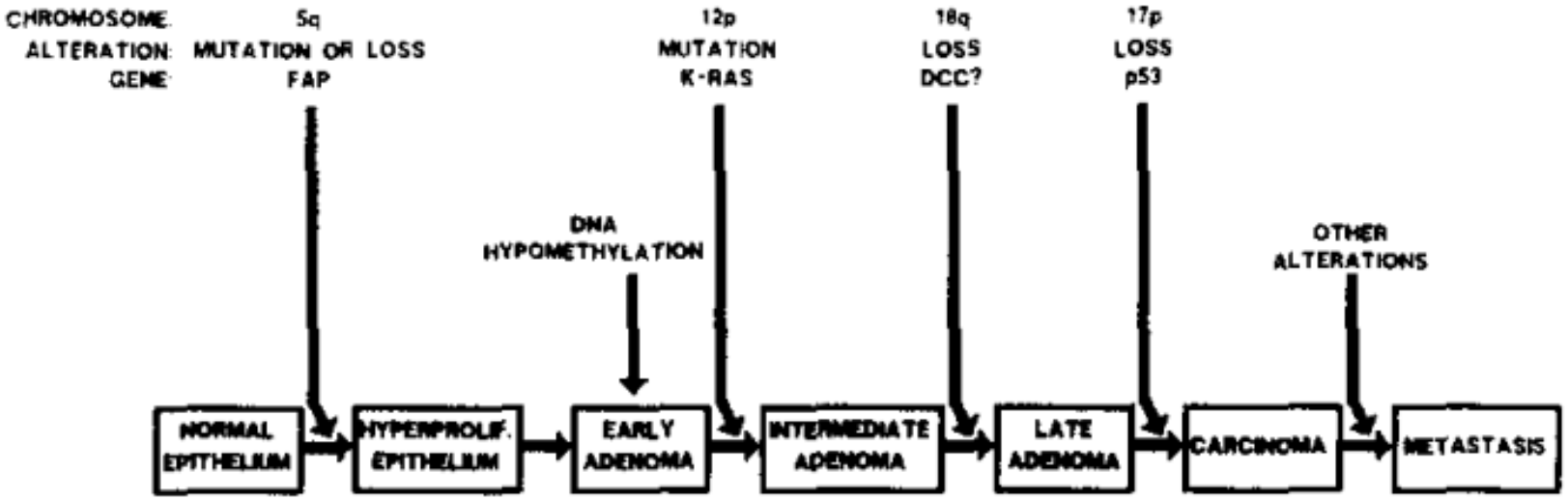


Figure 3. A Genetic Model for Colorectal Tumorigenesis

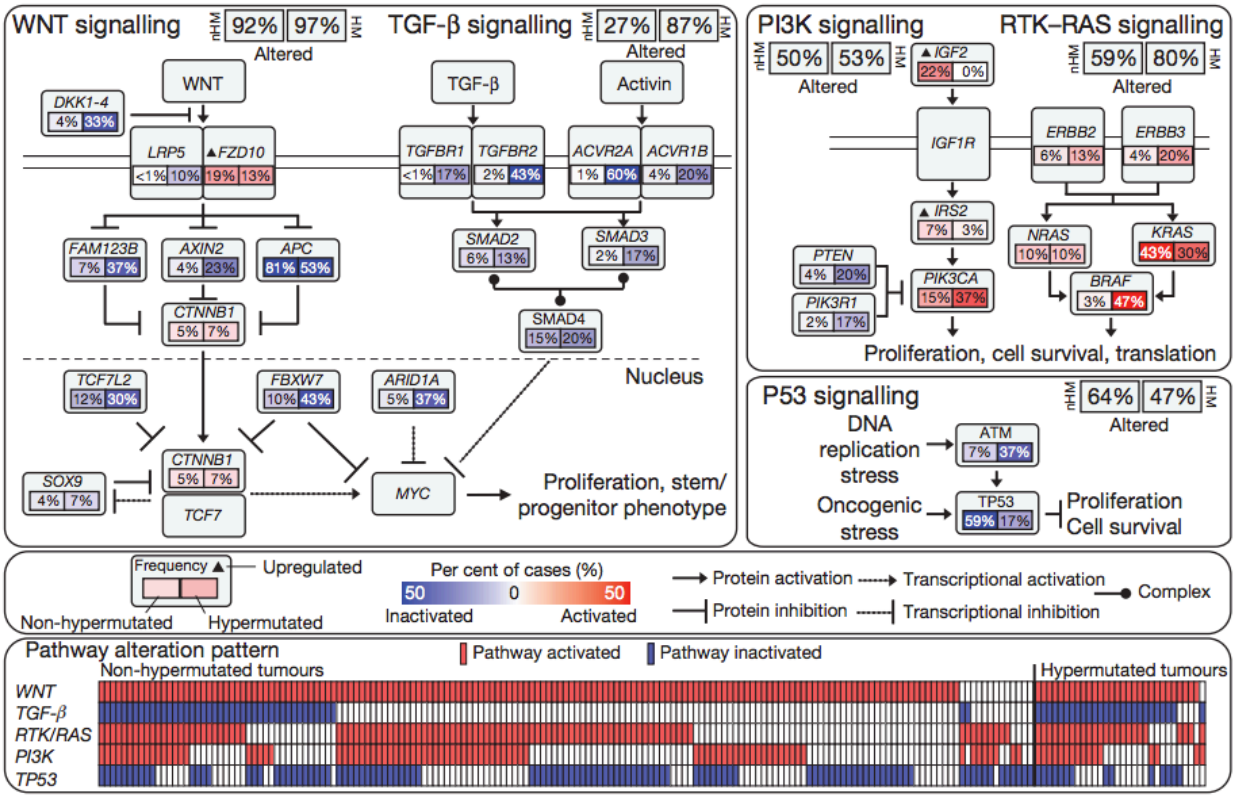
In reality … Complex Networks in Disease

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

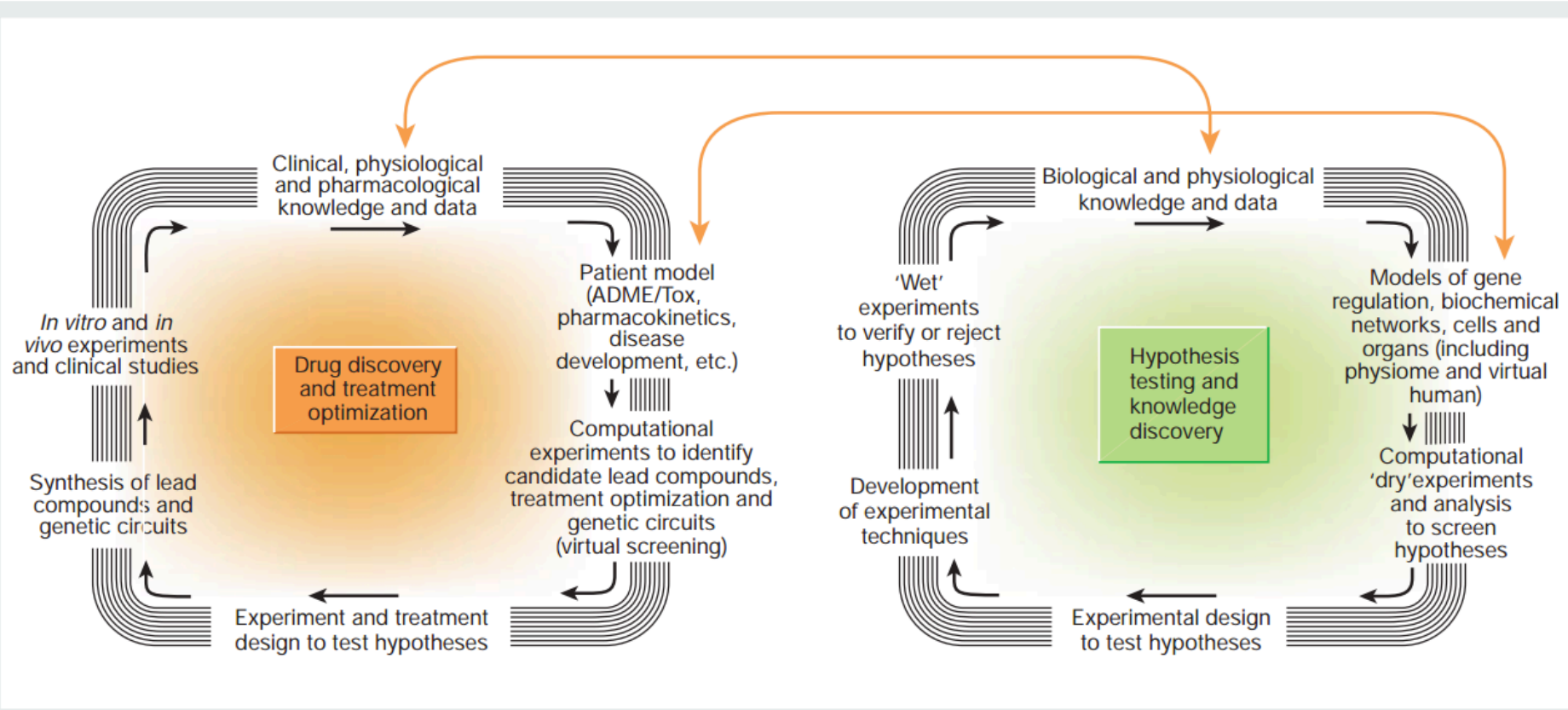


Figure 1 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments[18]. With sufficient progress in basic systems 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

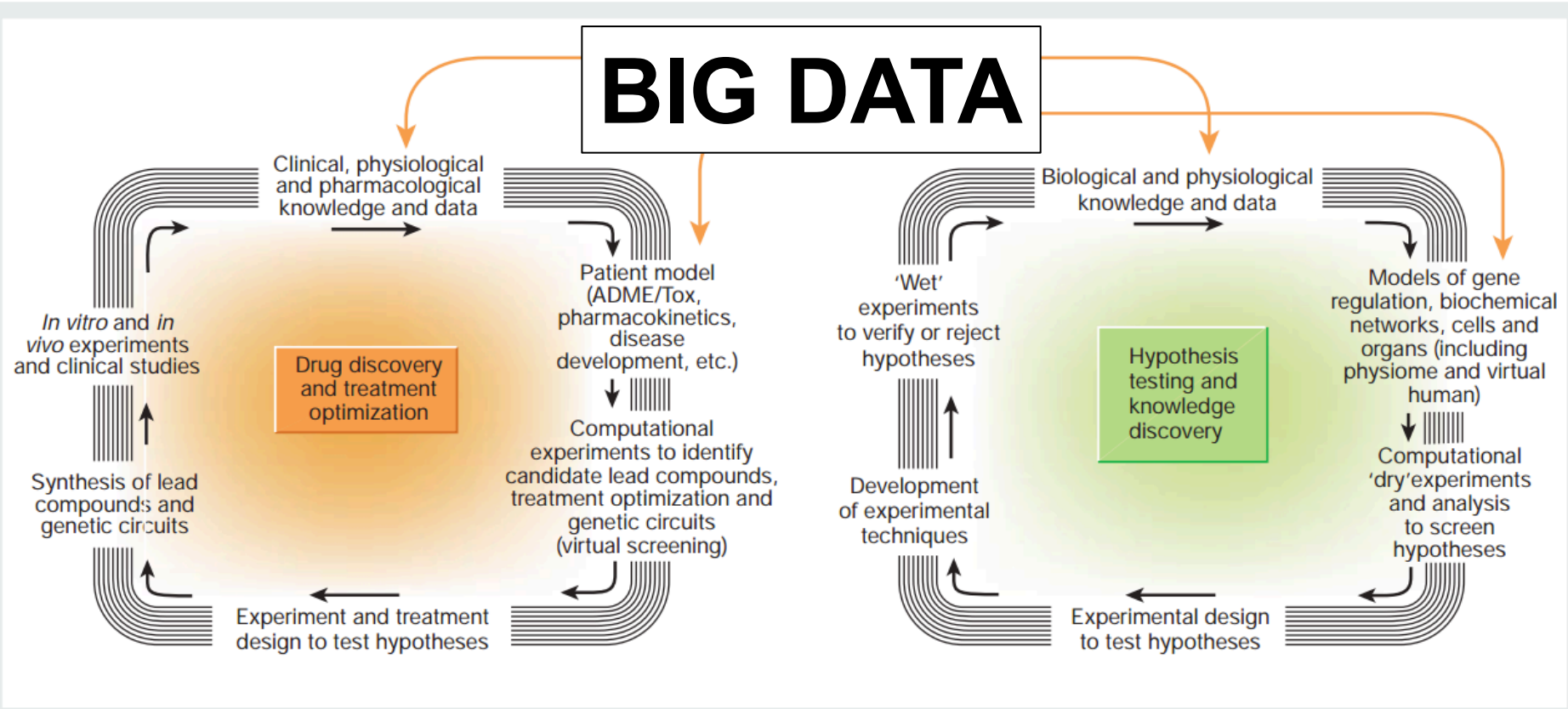


Figure 1 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments[18]. With sufficient progress in basic systems 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

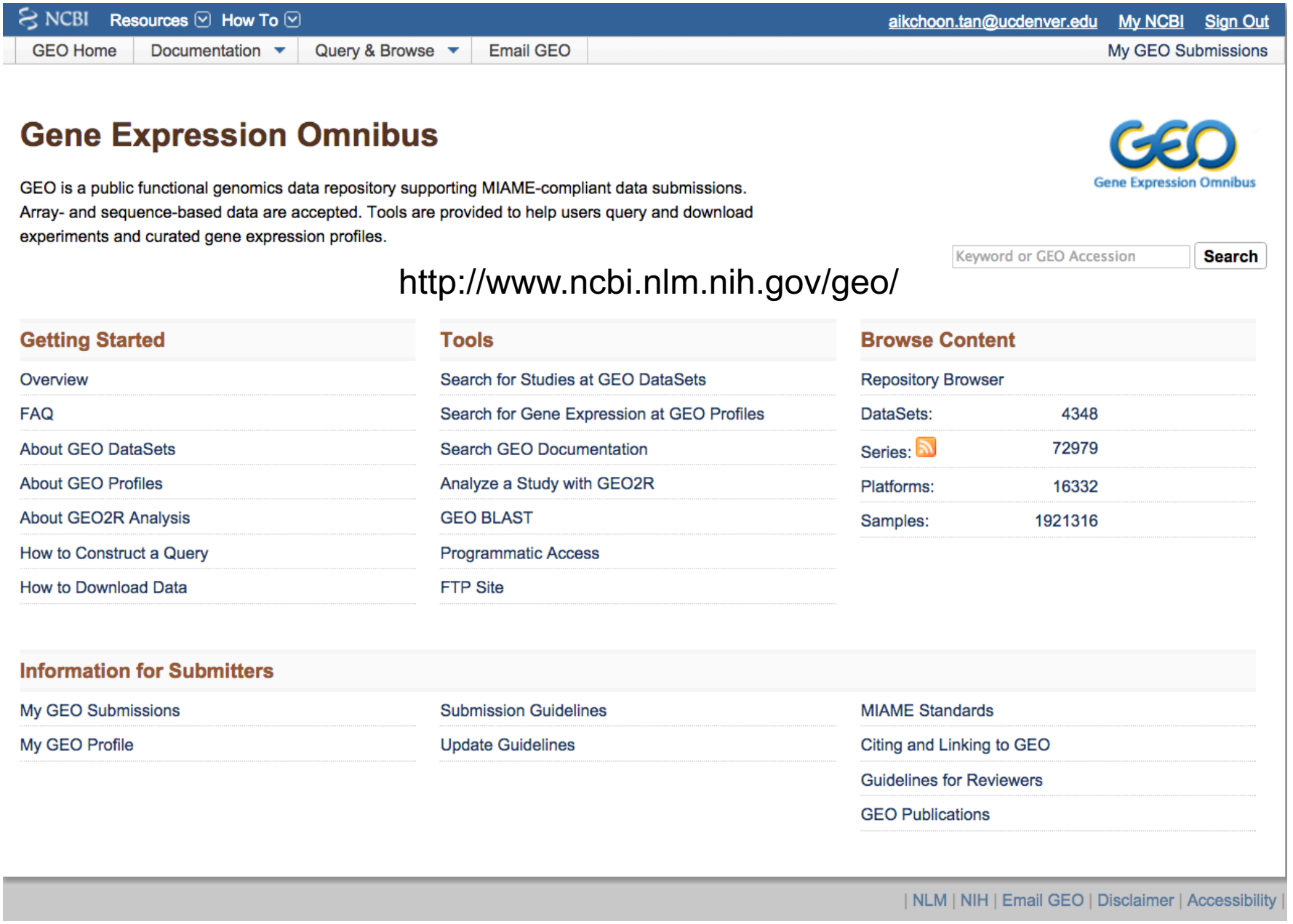
- Maps: Structure
  - Genetic Map
  - Physical Map
  - Sequence Map
- Maps: Molecular Function
  - Gene Map
  - Evolutionary Conservation Map
  - Chromatin State Map
  - 3-D Folding Map
- Maps: Disease
  - Inherited Variation Map
  - Disease Association Map
  - Evolutionary Selection Map
  - Cancer Gene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

Look up Table in Biology

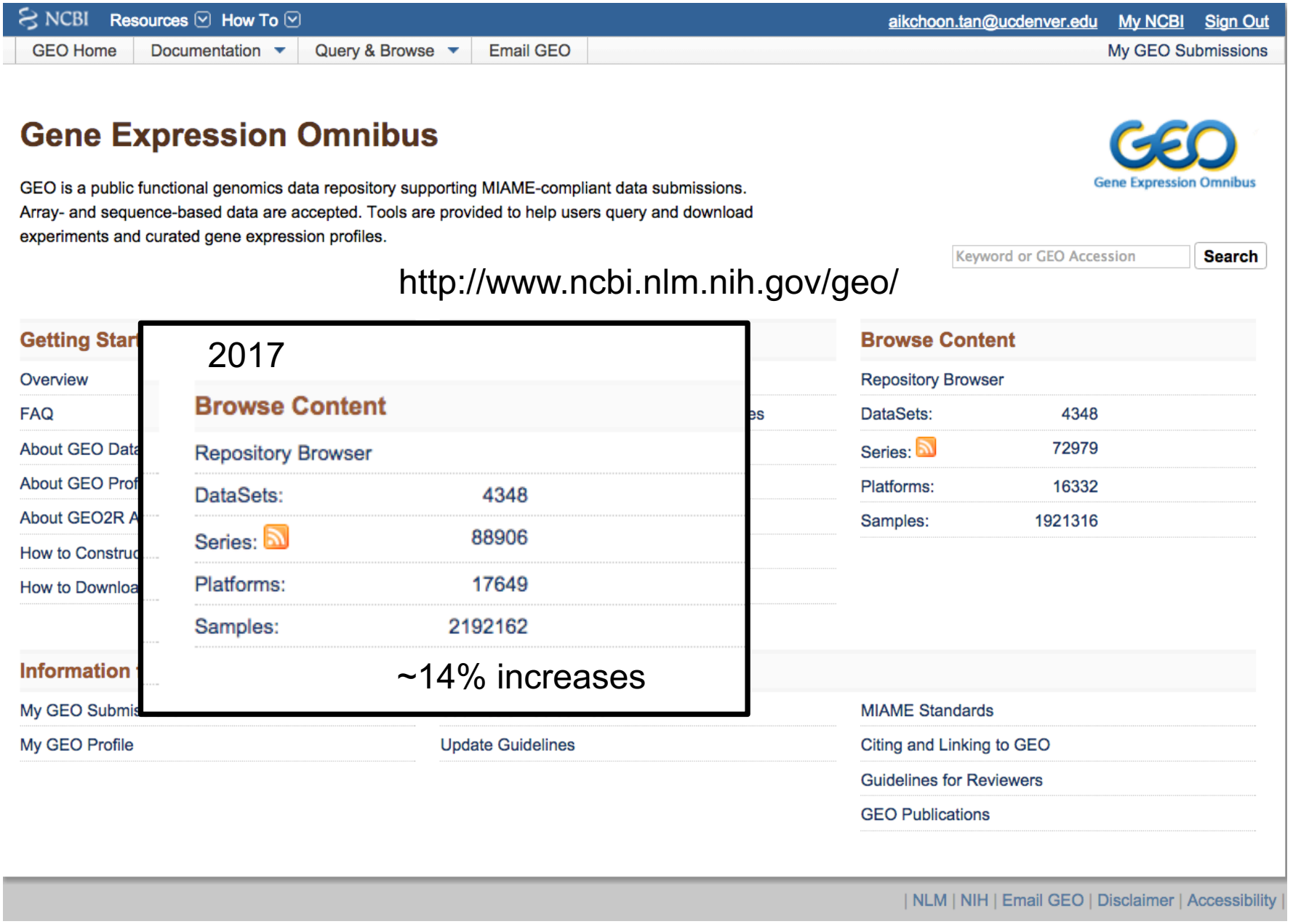
(like periodic table in chemistry)

(Adapted from Eric Lander's slide)

Gene Expression Omnibus (GEO)

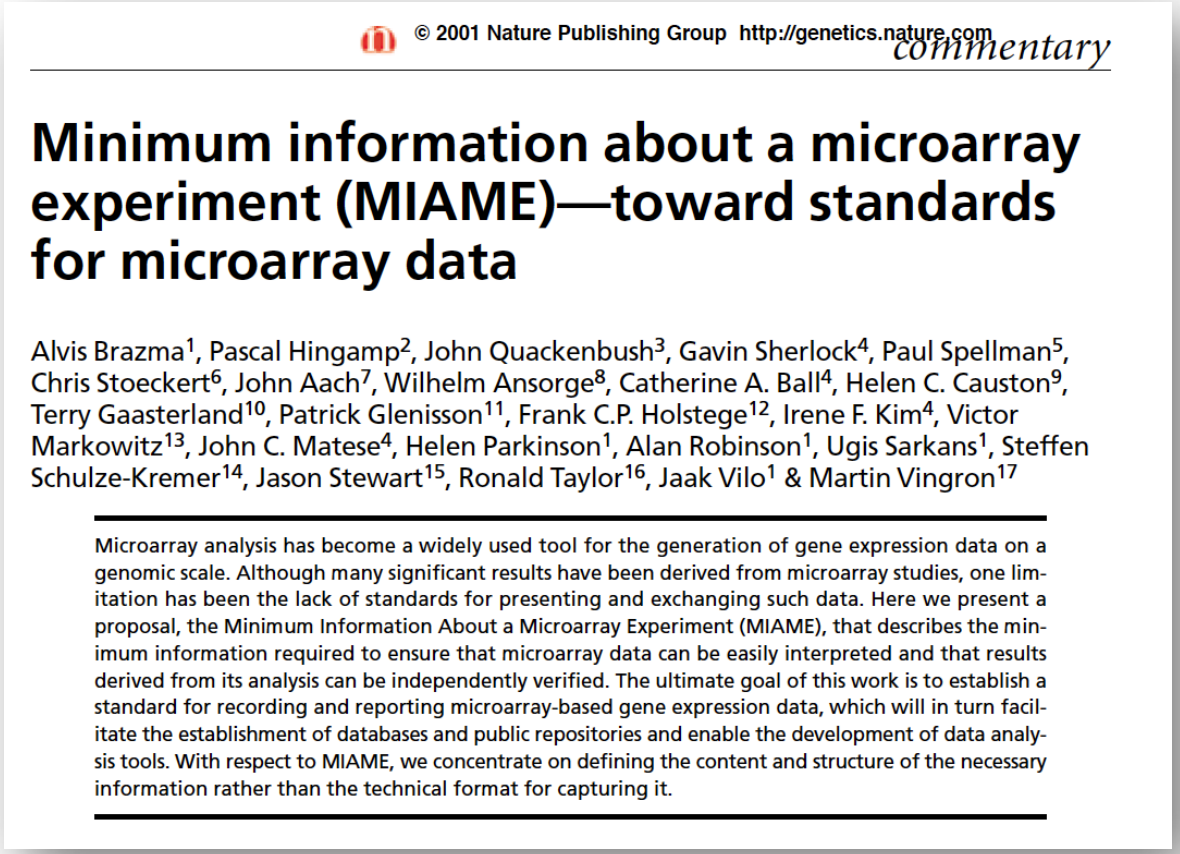


Gene Expression Omnibus (GEO)



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.

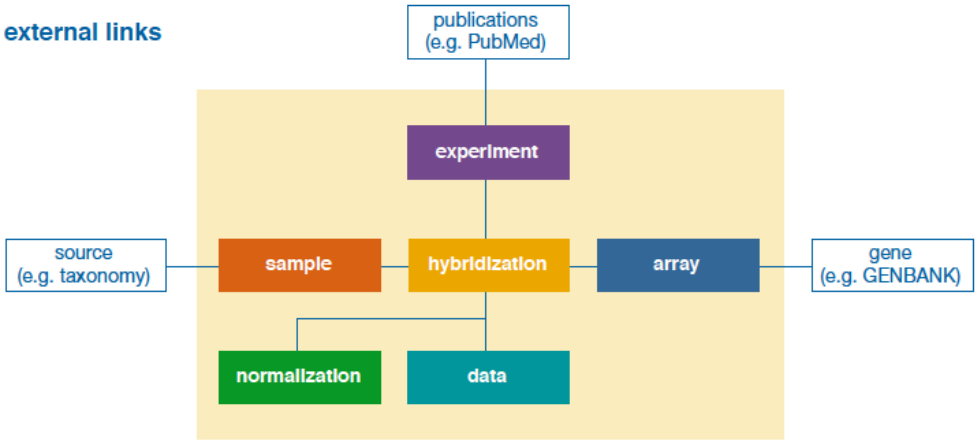


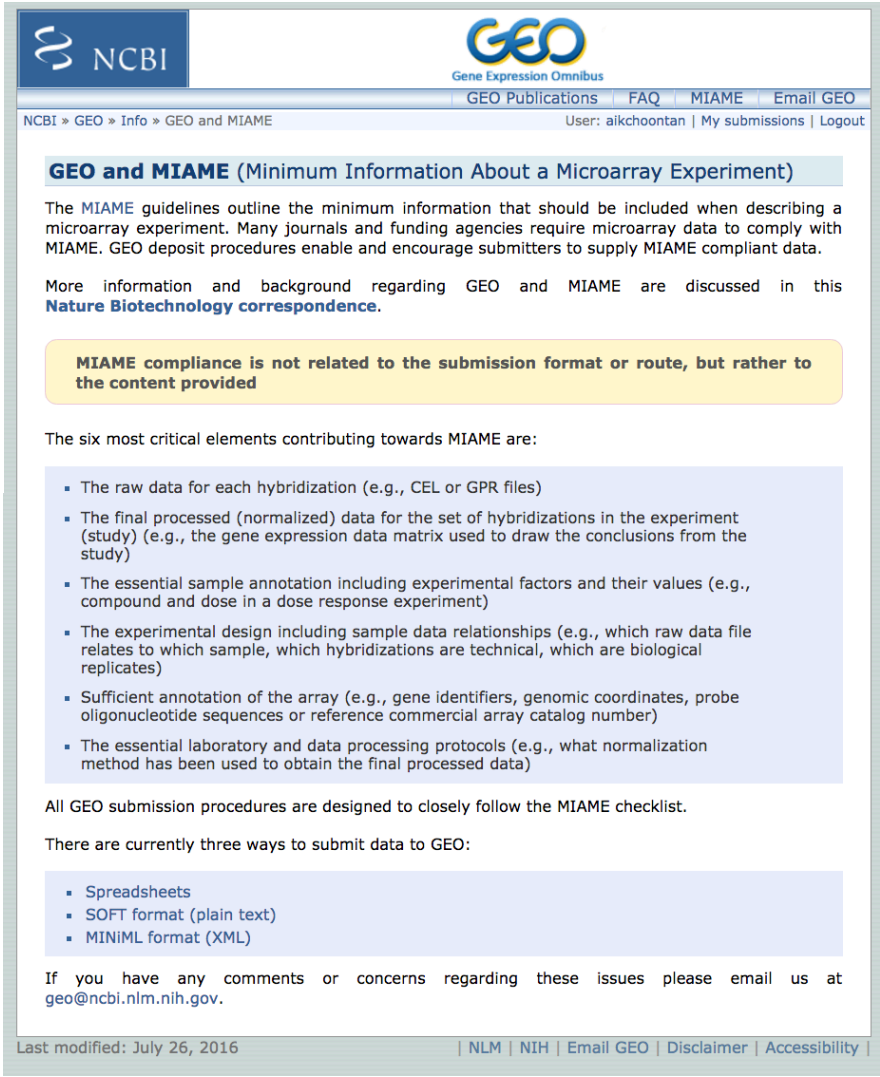
Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

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)





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.

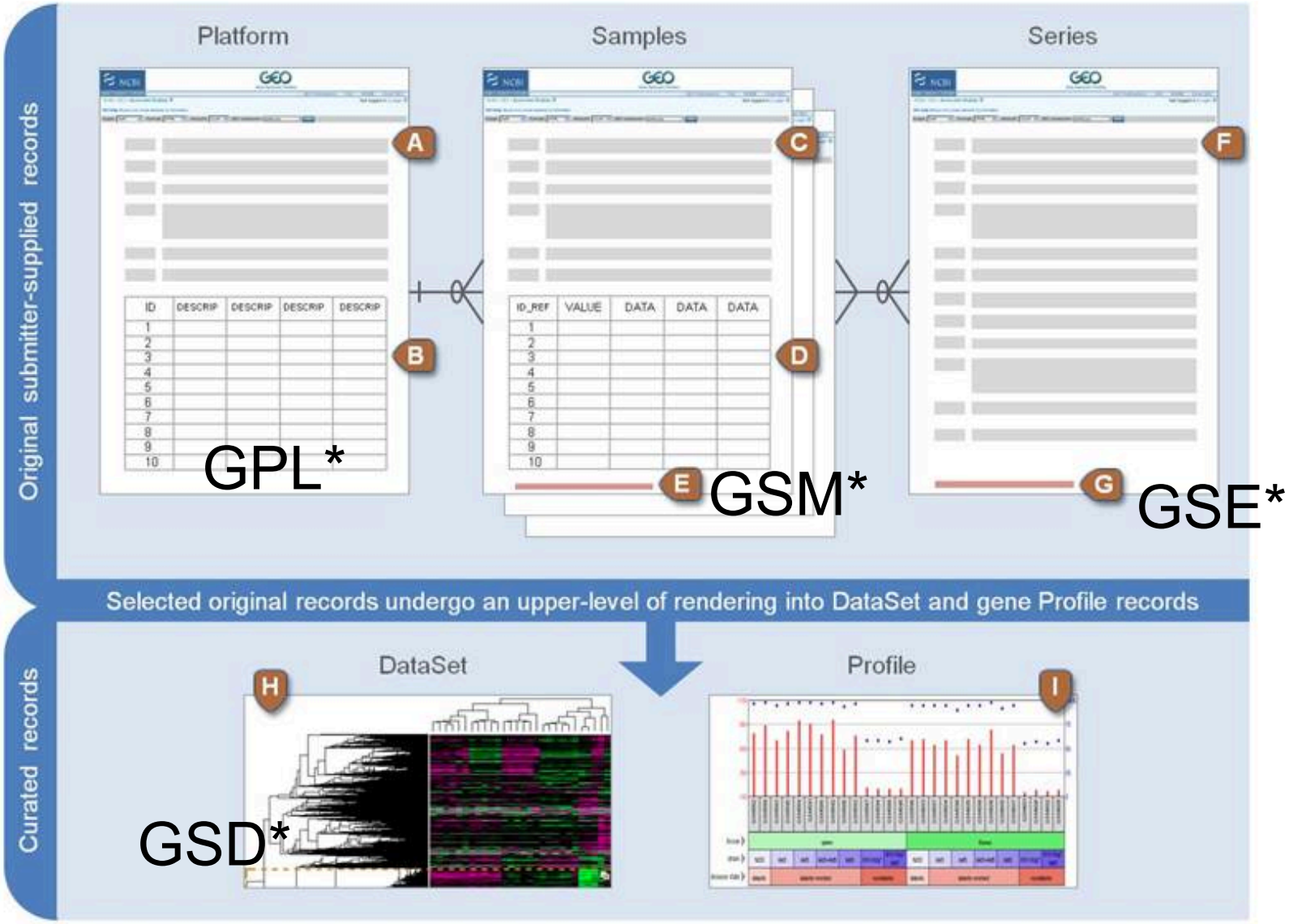
Last modified: July 26, 2016

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

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 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.

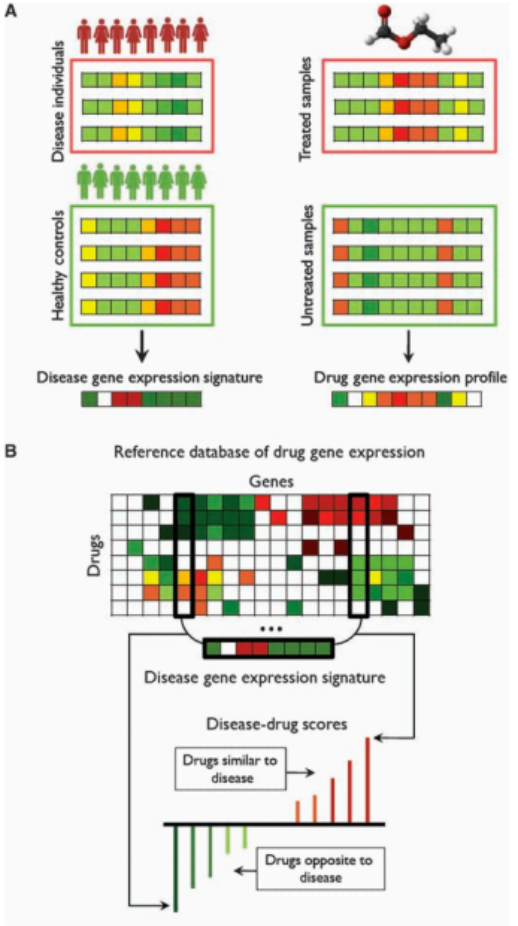


Fig. 1. Analytic workflow. (**A**) Two gene expression collections are used: a set of disease-associated gene expression data with corresponding controls and a set of gene expression data from tissue treated with drugs and small molecules with corresponding controls. SAM is used to obtain a signature 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

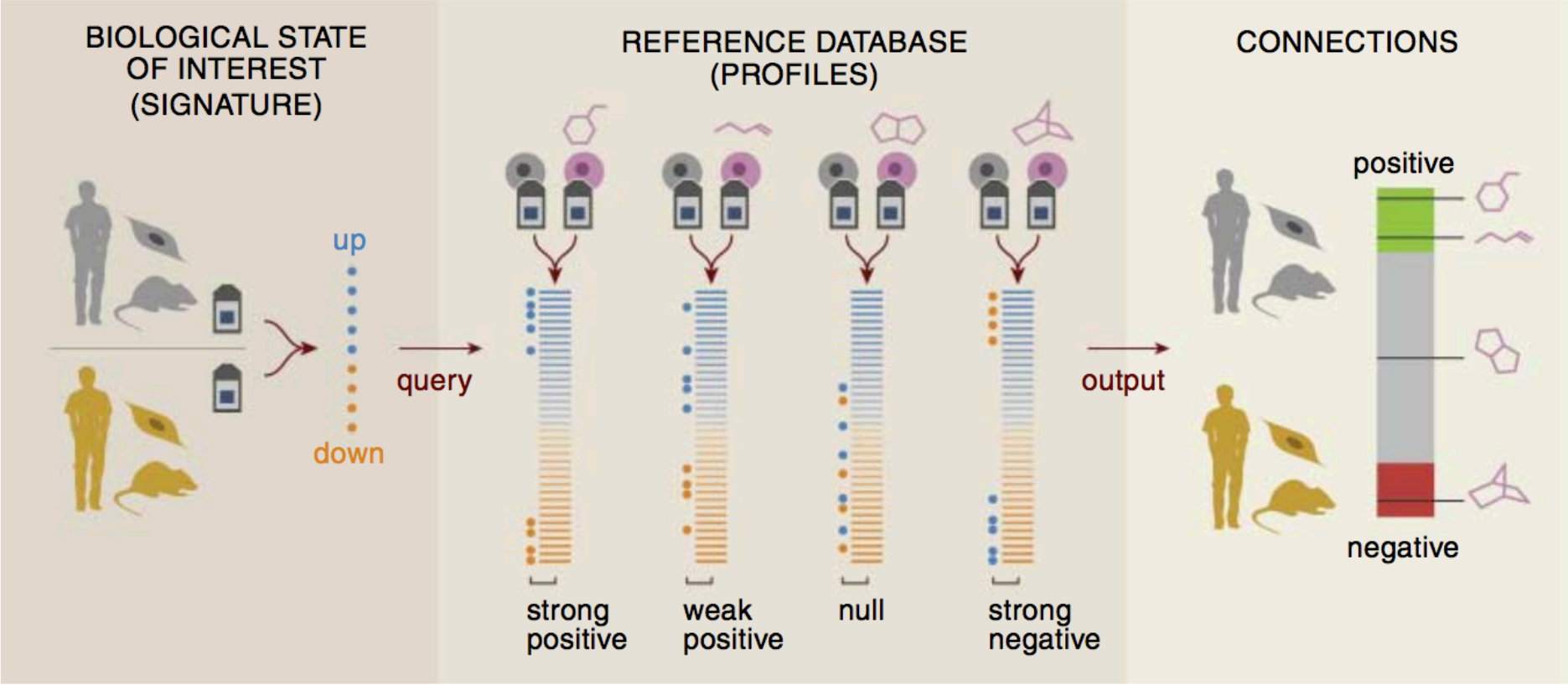
Justin Lamb,[1*] Emily D. Crawford,[1†] David Peck,[1] Joshua W. Modell,[1] Irene C. Blat,[1] Matthew J. Wrobel,[1] Jim Lerner,[1] Jean-Philippe Brunet,[1] Aravind Subramanian,[1] Kenneth N. Ross,[1] Michael Reich,[1] Haley Hieronymus,[1,2] Guo Wei,[1,2] Scott A. Armstrong,[2,3] Stephen J. Haggarty,[1,4] 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



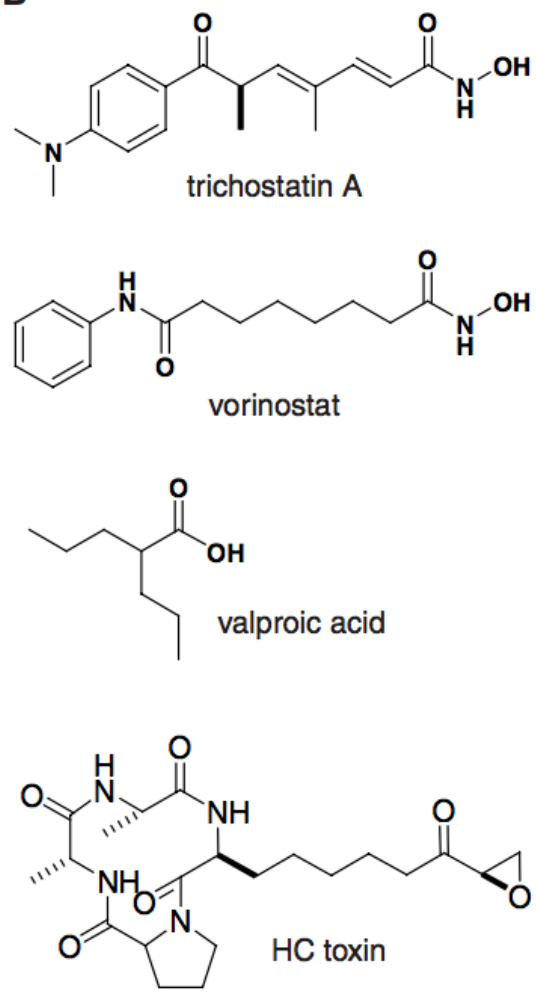
The Connectivity Map Project

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.

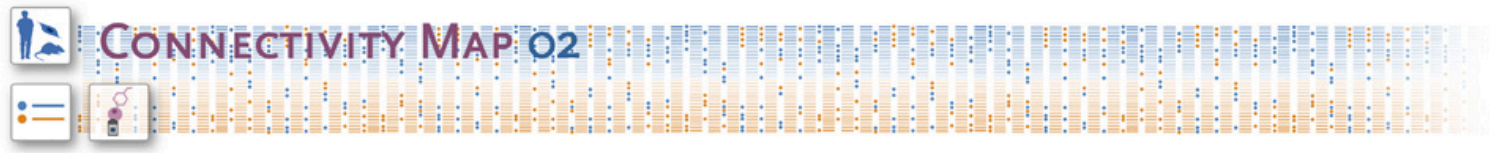
A

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

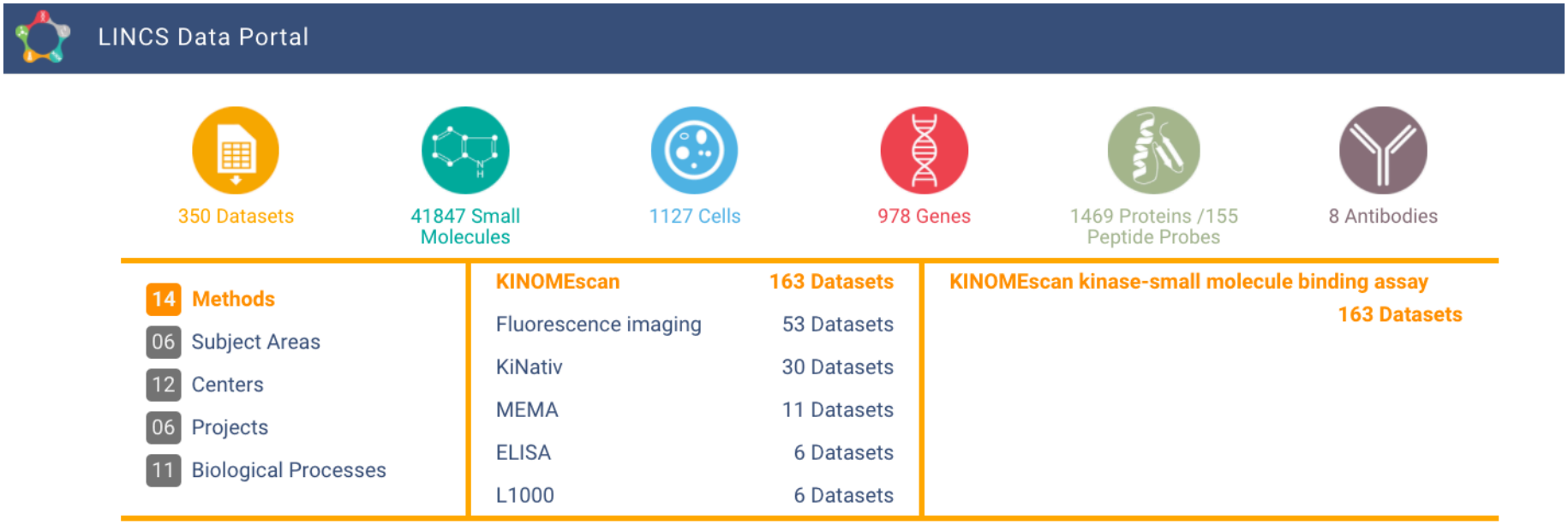
BROAD INSTITUTE *The Broad Institute is a research collaboration of MIT, Harvard and its affiliated Hospitals, and the Whitehead Institute, created to bring the power of genomics to medicine.*

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

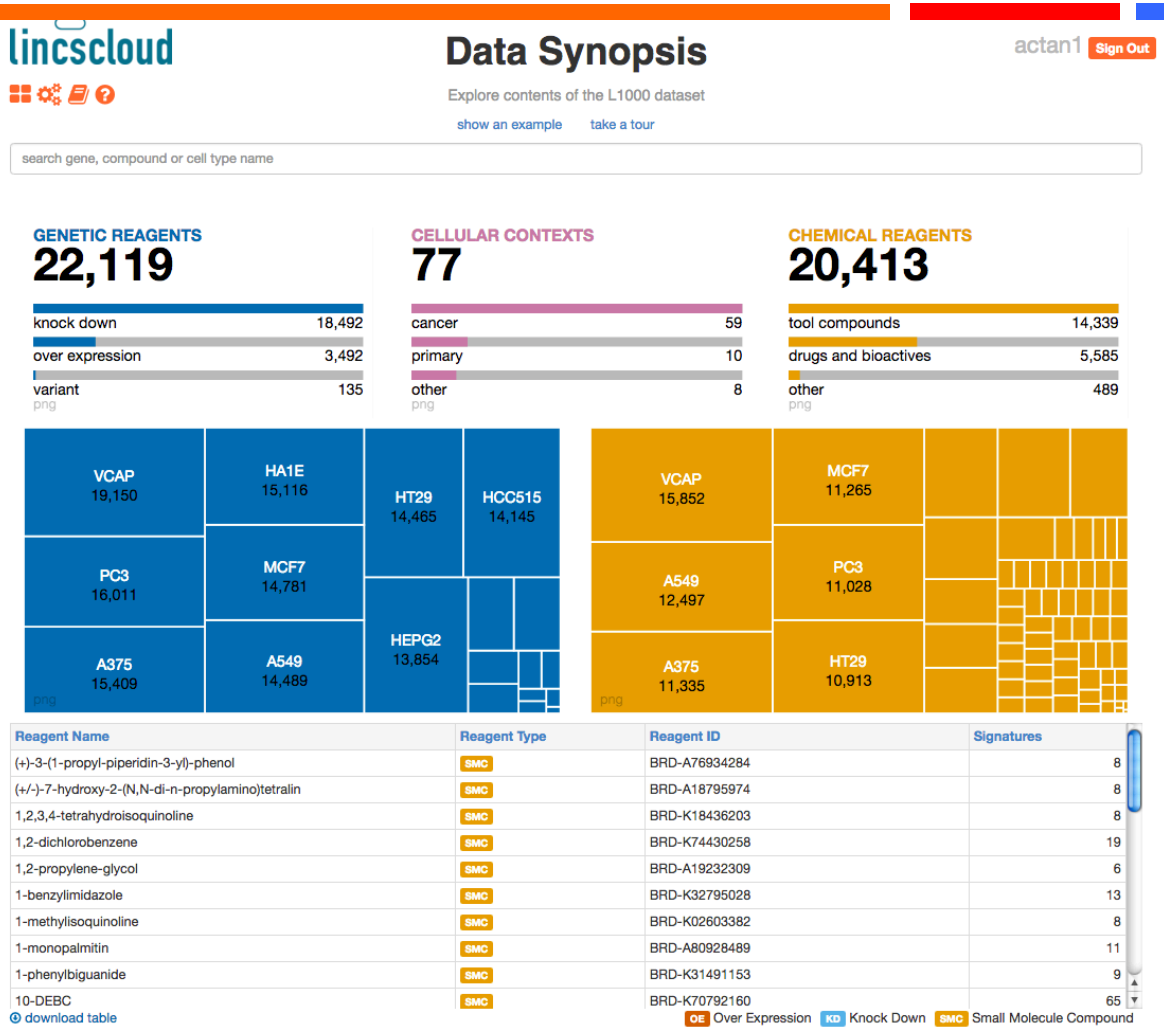
The NIH LINCS Project



http://lincsportal.ccs.miami.edu/dcic-portal/

The NIH LINCS Project

(The Broad Institute)



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

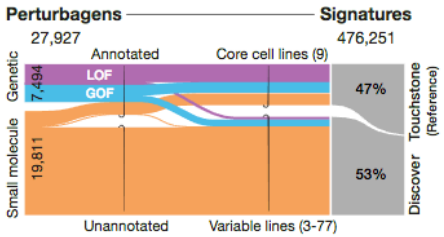
DATA VERSION: 1.0.1.1 / SOFTWARE VERSION: 1.1.1.15

27,927 perturbagens
476,251 expression signatures

Data and Tools

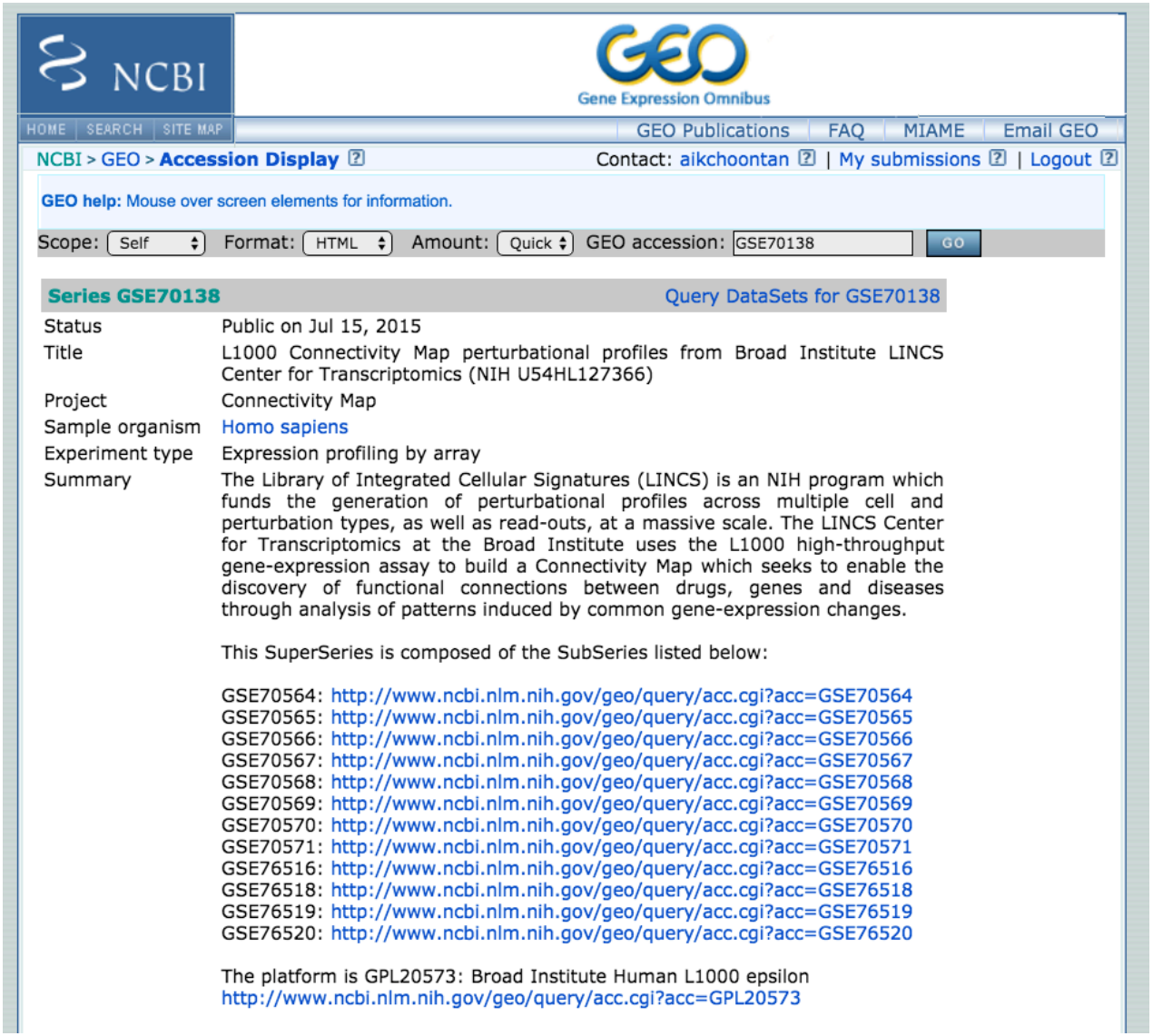
The CMap dataset of cellular signatures catalogs transcriptional responses of human cells to chemical and genetic perturbation. Here you can find the 1.3M L1000 profiles and the tools for their analysis.

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.

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

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS

Data Sharing Viewpoints
(Clinical Trials Data)

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

The aerial view of the concept of data sharing is beautiful. What could be better than having high-quality information carefully reexamined for the possibility that new nuggets of useful data are lying there, previously unseen? The potential for leveraging existing results for even more benefit pays appropriate increased tribute to the patients who put themselves at risk to generate the data. The moral imperative to honor their collective sacrifice is the trump card that takes this trick.

However, many of us who have actually conducted clinical research, managed clinical studies and data collection and analysis, and curated data sets have concerns about the details. The first concern is that someone not involved in the generation and collection of the data may not understand the choices made in defining the parameters. Special problems arise if data are to be combined from independent studies and considered comparable. How heterogeneous were the study populations? Were the eligibility criteria the same? Can it be assumed that the differences in study populations, data collection and analysis, and treatments, both protocol-specified and unspecified, can be ignored?

A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as "research parasites."

This issue of the *Journal* offers a product of data sharing that is exactly the opposite. The new investigators arrived on the scene with their own ideas and worked symbiotically, rather than parasitically, with the investigators holding the data, moving the field forward in a way that neither group could have done on its own. In this case, Dalerba and colleagues[1] had a hypothesis that colon cancers arising from more primitive colon epithelial precursors might be more aggressive tumors at greater risk of relapse and might be more likely to benefit from adjuvant treatment. They found a gene whose expression appeared to correlate with the expression of genes that characterize more mature colon cancers on gene-expression arrays and whose product was reliably measurable in resected colon cancer specimens by immunohistochemistry. To assess the clinical value of this potential biomarker, they needed a sufficiently large group of patients whose archived tissues could be used to assess biomarker expression and who had been treated in relatively homogeneous way.

They proposed a collaboration with the National Surgical Adjuvant Breast and Bowel Project (NSABP) cooperative group, a research consortium funded by the National Cancer Institute that has conducted seminal research in the treatment of breast and bowel cancer for the past 50 years. The NSABP provided access to tissue and to clinical trial results on an individual patient basis. This symbiotic collaboration found that a small proportion (4%) of colon cancers did not express the biomarker and that the survival of patients with those tumors was poorer than that of patients whose tumors expressed the biomarker. Furthermore, when the effect of adjuvant chemotherapy was assessed, nearly all

There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as "research parasites."

Data Sharing Viewpoints

More on Data Sharing

TO THE EDITOR: For all the understandable uproar over the term "research parasites" — an inflammatory term that gives short shrift to how open data changed our understanding of Tamiflu, Paxil, and other treatments — those of us who support increased data sharing should realize that Drazen and Longo[1,2] were giving voice to an opinion that many researchers privately hold. After all, it is only human nature that some feel wary of a policy that seems to require them to do extra work that other people will then use for their own academic advancement.

The best way to create a world with more data sharing is to hear out these concerns fairly and figure out how to address them. For example, tenure committees and National Institutes of Health funding reviews should give abundant credit to anyone who originates a data set that other scientists find useful. If data sharing is in the self-interest of whoever collected the data, data sharing as a policy will be on better footing.

Stuart Buck, J.D., Ph.D.

Laura and John Arnold Foundation
Houston, TX
stuartbuck@gmail.com

No potential conflict of interest relevant to this letter was reported.

Research Parasite
@dataparasite

Reanalyzing your data. Disproving what you posited. Stealing ideas you haven't yet had.

In your data
for.researchparasites.com
Joined January 2016

Data Sharing Viewpoints

Bioinformatics, 2016, 1–1
doi: 10.1093/bioinformatics/btw090
OXFORD

Editorial

ISCB's initial reaction to *New England Journal of Medicine* editorial on data sharing

The recent editorial by Dr Longo and Dr Drazen in the *New England Journal of Medicine* (Longo and Drazen, 2016) has stirred up quite a bit of controversy. As Executive Officers of the International Society of Computational Biology, Inc. (ISCB), we express our deep concern about the restrictive and potentially damaging opinions voiced in this editorial, and while ISCB works to write a detailed response, we felt it necessary to promptly address the editorial with this reaction. Although some of the concerns voiced by the authors of the editorial are worth considering, large parts of the statement purport an obsolete view of hegemony over data that is neither in line with today's spirit of open access nor furthering an atmosphere where the potential of data can be fully realized.

ISCB acknowledges that the additional comment on the editorial (Drazen, 2016) eases some of the polemics unfortunately without addressing some of the core issues. We still feel, however, that we need to contrast the opinion voiced in the editorial with what we consider the axioms of our scientific society, statements that lead into a fruitful future of data-driven science:

i. Data produced with public money should be public in benefit of the science and society
ii. Restrictions on the use of public data hamper science and slow progress
iii. Open data is the best way to combat fraud and misinterpretations

Current large data collections proceed from many sources, are continually accumulated, and require a variety of analytical approaches. Data generation and data analysis overlap in time and are continually updated with new data sets produced by new techniques and new analysis methodologies. Furthermore, in many cases current science functions in consortia in which scientists collaborate toward common goals while preserving their own scientific objectives. Dividing scientists into data providers and data analysts is simplistic and gives a misleading impression of the actual state of biological and biomedical science.

ISCB very much supports collaboration between disciplines, including experimental and clinical as well as bioinformatics, as the best way forward to address complex biological problems. But this collaboration cannot be based on imposed restrictions to data access and cannot be contained in professional silos. (The use of expressions such as 'research parasites' clearly does not help.)

Many bio-communities have made significant progress by endorsing open data policies and, gratefully, public funding agencies have connected to the spirit that they are distributing taxpayers' money to science and that, therefore, the data that are generated in the course belong to the public. It is, perhaps, natural that some areas of biomedical research are slow in adopting these policies. History and the confidential nature of the relevant data are surely among the reasons. However, in our opinion data hegemony is another, a reason that has to be overcome. The sooner these barriers to progress are removed the sooner the patients will benefit from the current flourishing of biomedical research.

Conflict of Interest: none declared.

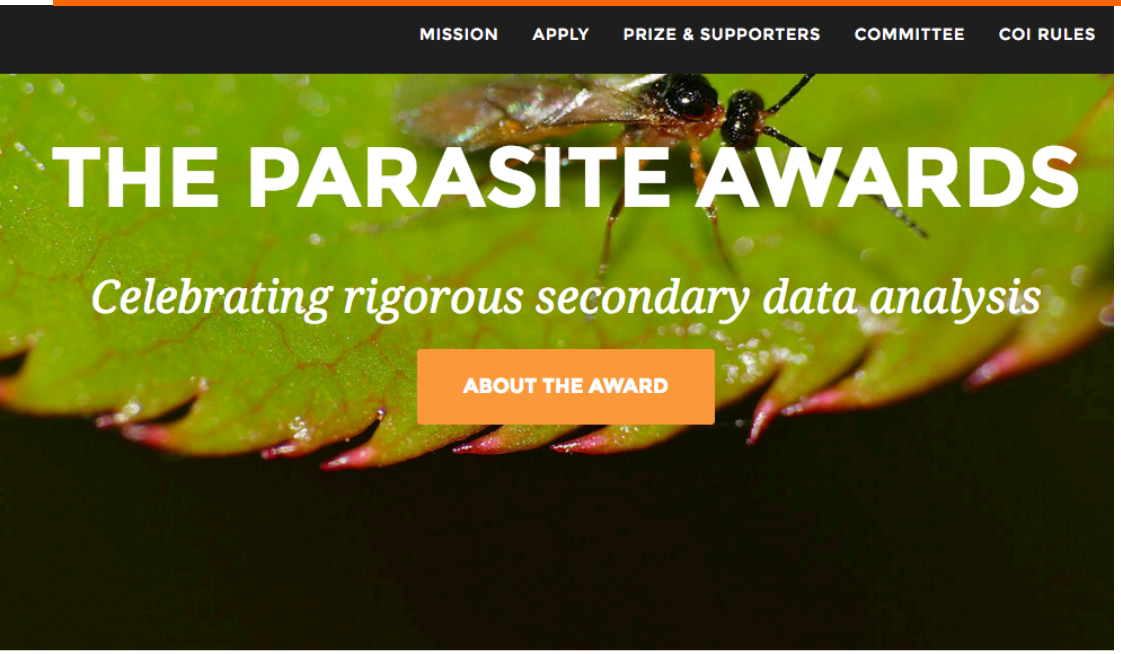
Bonnie Berger, Theresa Gaasterland, Thomas Lengauer, Christine A. Orengo, Bruno Gaeta, Scott Markel and Alfonso Valencia*
International Society for Computational Biology, Inc. (ISCB), 9650 Rockville Pike Bethesda, Maryland 20814, USA.

*To whom correspondence should be addressed.

References

Drazen,J.M. Data sharing and the journal. *N. Engl. J. Med.* doi: 10.1056/NEJMe1601087.
Longo,D.L. and Drazen,J.M. (2016) Data sharing. *N. Engl. J. Med.*, 374, 276–277.

Data Sharing Viewpoints



THE "PARASITES"

PSB Awards for rigorous secondary data analysis.

The act of generating new hypotheses from existing data is a major component in the process of science. Dr. Albert Szent-Györgyi has been quoted as saying "discovery consists of seeing what everybody has seen, and thinking what nobody has thought." Recent advances in data sharing, combined with the expectation that publicly funded research will be shared, have led to projects that consist largely of secondary analysis of data. The practitioners of this craft may analyze or combine these data in ways that answer scientific questions that the initial investigators did not consider. In a 2016 editorial, the New England Journal of Medicine termed these people "research parasites."

The Parasite awards, given annually, recognize outstanding contributions to the rigorous secondary analysis of data. This practice of secondary analysis plays a key role in scientific ecosystem: conclusions that persist through substantial reanalysis are expected to be more credible; and analyses that extract more knowledge from underutilized data make the practice of science more efficient.

The Parasites currently consist of two awards: the first recognizes an outstanding contribution from a junior parasite (postdoctoral, graduate, or undergraduate trainee), and the second recognizes an individual for a sustained period of exemplary research parasitism.

http://researchparasite.com/

Data Sharing Viewpoints

ELIGIBILITY & APPLICATION

How to apply for an award.

APPLICATION PROCESS

For either award, submit an application by **October 14, 2016 at 5PM HST** (Hawaii Standard Time) to parasite.award@gmail.com. An application requires:

- A nomination letter describing how each selected paper meets the criteria for the award. Self nominations are encouraged, and all nominees must be aware that they have been nominated.
- Junior Parasite (aka the sporozoite): a PDF of one paper published after peer review on which the application will be judged.
- Sustained Parasitism (aka the merozoite): PDFs of three papers published after peer review on which the application will be judged.

The award winners will be recognized at the Pacific Symposium on Biocomputing each year, and listed on the PSB website, along with links to the winning papers.

ELIGIBILITY

Selection criteria (both awards) for the work in question:

- The awardee must not have been involved the design of the experiments that generated the data.
- The awardee published independently of the original investigators, and the original investigators are not authors of the secondary analyses but are appropriately credited in the manuscripts.
- The awardee may have extended, replicated or disproved what the original investigators had posited.
- The awardee has provided source code and intermediate or final results in a manner that enhances reproducibility.

Additional selection criteria for the Junior Parasite award:

- The awardee must have published the work at the training stage of their career (postdoctoral, graduate, or undergraduate). If the awardee has assumed a position as an independent investigator she or he should not have been in that position for more than 2 years.
- The award will be based on work described in a single manuscript (submitted alongside the nomination letter).

Additional selection criteria for the Sustained Parasitism award:

- The awardee must be in an independent investigator position in academia, industry or public sector.
- The awardee must be a last or corresponding author on the three manuscripts submitted alongside the nomination letter.
- At least a five-year period must have elapsed between the publication of the first manuscript and the final manuscript.

PRIZE & SUPPORTERS

Those who make this possible.

PRIZES

The winners of each award will receive:

- a $500 prize.
- a free one-year electronic subscription to Nature Genetics.
- an article-processing charge waiver for an article in *Scientific Data*.
- a Gordon and Betty Moore Foundation Klean Kanteen and notebook.

Financial support for the award has been provided by: **Nature Genetics**, **The Arnold Foundation**, **The Gordon and Betty Moore Foundation** (via GBMF 4552 to CSG), and **Casey Greene**.

TRAVEL SUPPORT

Travel support is available to the recipient of the Junior Parasite award. Generous sponsorship from **GigaScience** and **Scientific Data** will allow us to cover the costs of economy airfare and hotel for the duration of the meeting. Support from **GigaScience**, **Scientific Data**, and **Nature Genetics** will allow us to cover the cost of registration for the Pacific Symposium on Biocomputing, where the award is announced.

GigaScience

GigaScience aims to revolutionize reproducibility of analyses, data dissemination, organization, understanding, and use through open access and open data publication of 'big data' studies across the life and biomedical sciences.

Nature Genetics

Nature Genetics publishes research that encompasses genetic and functional genomic studies. Current emphasis is on common and complex diseases and on the functional mechanism, architecture and evolution of gene networks.

Scientific Data

Scientific Data is an open-access journal for descriptions of scientifically valuable datasets from a broad range of research disciplines – helping make research data more available, citable, discoverable, interpretable, reusable and reproducible.

http://researchparasite.com/

Data Sharing Viewpoints

AWARD RECIPIENTS

Exemplars of research parasitism.

Kun-Hsing Yu
2017 Junior Parasite

Erick Turner
2017 Sustained Parasitism

http://researchparasite.com/

Data Sharing Viewpoints

KUN-HSING YU

Winner of the 2017 Junior Parasite Award.

The 2017 junior parasite award recipient was Dr. Kun-Hsing Yu. In the research for which he was nominated, Dr. Yu and colleagues employed existing datasets and software in an innovative new analysis. They reanalyzed TCGA histopathology images and Stanford Tissue Microarray data and extracted features using methods built into CellProfiler. They employed a number of different machine learning approaches using packages for the R programming language, which were mentioned and cited in the manuscript. They also provided source code and data for the analyses under an open license.

Dr. Yu says:

> I would like to take this opportunity to thank the PSB Parasite Award Committee for organizing this award and my Ph.D. co-advisors Professors Michael Snyder and Russ Altman for supervising my work and nominating me for the award.
>
> Research parasitism, or secondary data analysis, plays a key role in the scientific ecosystem. With data reanalysis, we can ensure the reproducibility of scientific investigations, make the most of the underutilized data, and integrate data from different sources to generate novel biomedical insights. Currently, most biomedical data is trapped in silos, which hinders scientific progress and improvement of healthcare. Biomedical informaticians routinely integrate diverse data types and are in a great position to revolutionize biomedical investigations by breaking the silos, accelerating the scientific process, and translating big data into deep knowledge in biomedicine.
>
> It takes a whole ecosystem to advance science, and secondary data analysis is indispensable for biomedical investigations in the 21st century. As a biomedical informatician, I am very proud of being able to contribute to the ecosystem in a revolutionary way.

ERICK TURNER

Winner of the 2017 Sustained Parasitism Award.

The 2017 sustained parasite award recipient was Dr. Erick Turner. In the research for which he was nominated [1, 2, 3], Dr. Turner and colleagues identified pervasive publication bias. According to published literature, nearly all clinical trials of antidepressants that they evaluated were positive. However at the FDA level, only half showed a significant positive effect. Dr. Turner and his collaborators have continued to identify reporting biases for a sustained period.

Dr. Turner says:

> In each of these studies, I and my colleagues have "parasitically" compared published peer-reviewed journal articles to FDA drug approval packages. These are freely available to the public but unfortunately user-unfriendly. To correct this, we have been developing OpenTrialsFDA, a project aimed at making the FDA's trove of drug data easier to access and use. The goal is to grow the community of "FDA parasites" so that researchers and others, including journalists, can get a much more complete and "unspun" picture of how safe and effective our drugs really are.
>
> OpenTrialsFDA is currently one of 6 finalists for the Open Science Prize.

http://researchparasite.com/

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

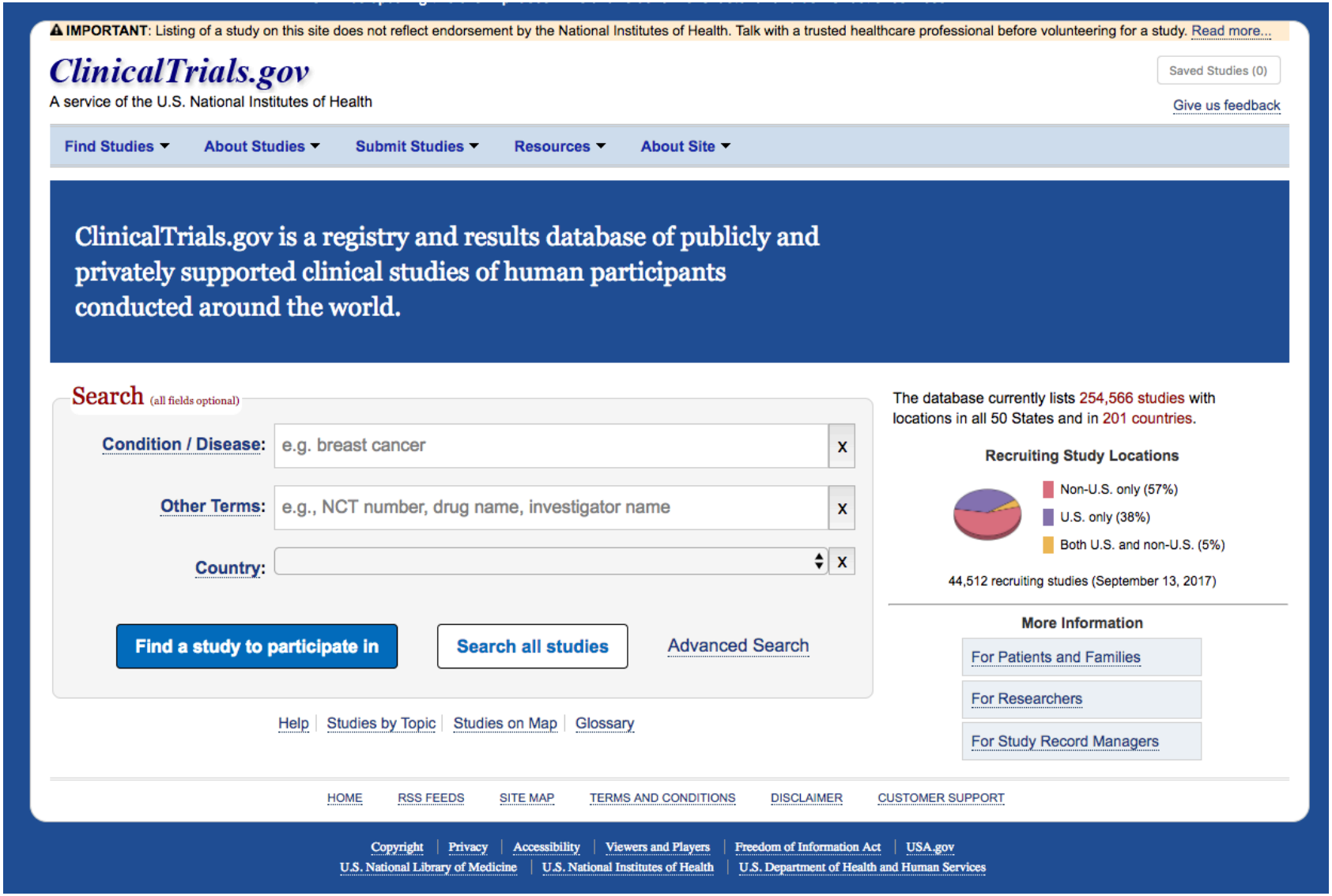
Learn More
- Tutorials for using ClinicalTrials.gov
- Glossary of common site terms
- For the press
- Using our RSS feeds

HOME | RSS FEEDS | SITE MAP | TERMS AND CON

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

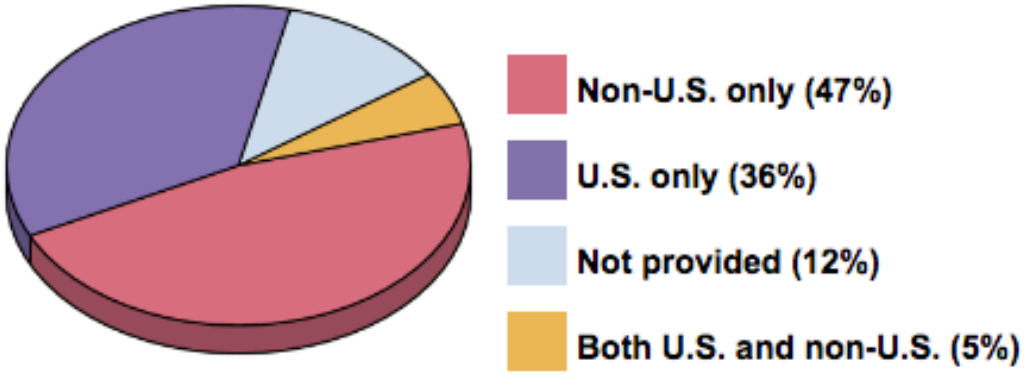
Copyright | Privacy | Accessibility | Viewers and Players | Freedom of Information Act | USA.gov
U.S. National Library of Medicine | U.S. National Institutes of Health | U.S. Department of Health and Human Services

Clinical Trials Data



Clinical Trials Data

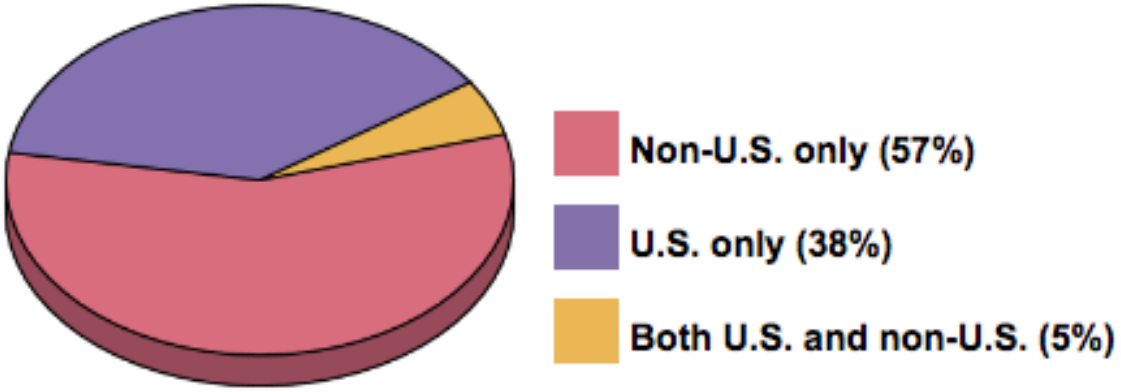
Percentage of Registered Studies by Location (as of September 13, 2017)
Total N = 254,566 studies



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

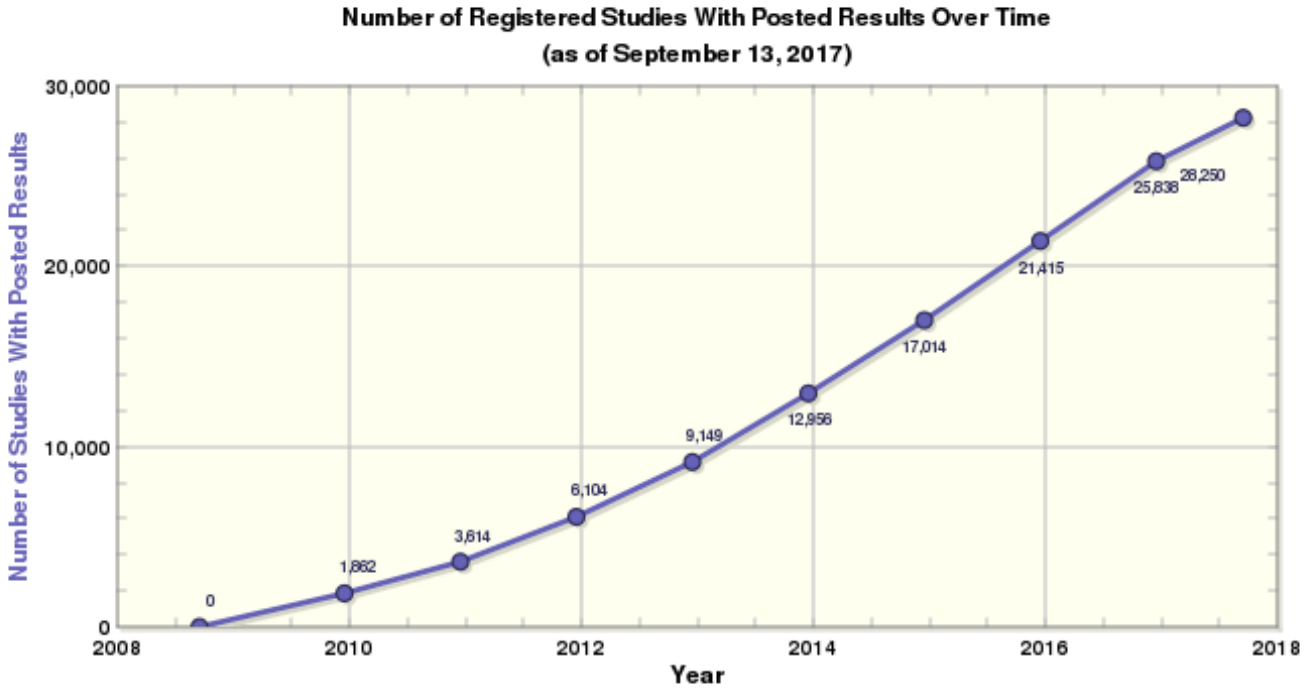
The graph and table below show the total number of studies registered on ClinicalTrials.gov since 2000, based on the First Received date. The first version of ClinicalTrials.gov was made available to the public on February 29, 2000.

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

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

Big Data Mining of Adverse Events in ClinicalTrials.gov

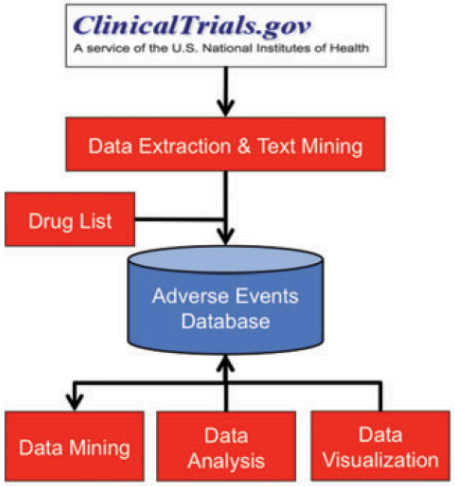
ORIGINAL ARTICLE

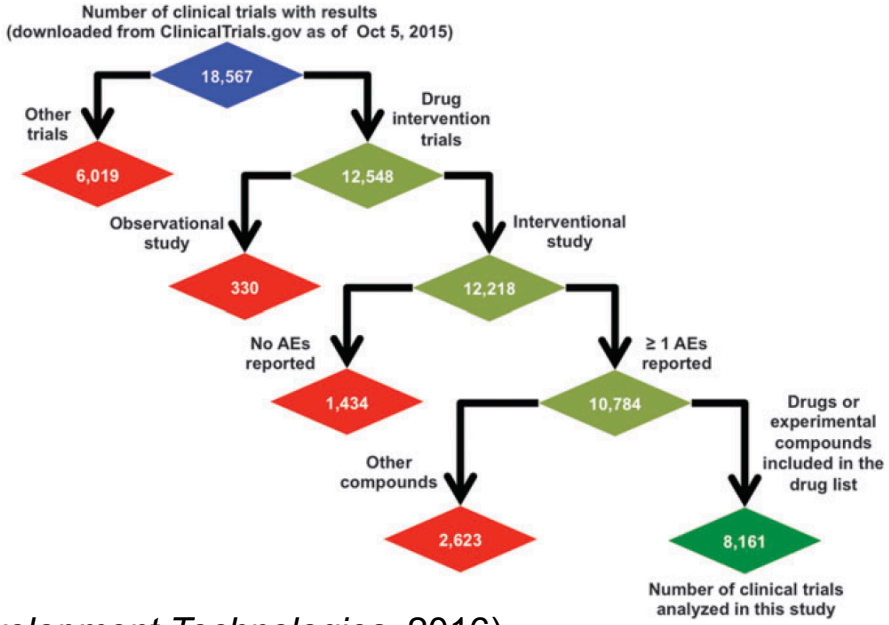
Big Data Mining and Adverse Event Pattern Analysis in Clinical Drug Trials

Callie Federer,[1,2,] Minjae Yoo,[1,*] and Aik Choon Tan[1,3,4]*

ABSTRACT

Drug adverse events (AEs) are a major health threat to patients seeking medical treatment and a significant barrier in drug discovery and development. AEs are now required to be submitted during clinical trials and can be extracted from ClinicalTrials.gov (https://clinicaltrials.gov/), a database of clinical studies around the world. By extracting drug and AE information from ClinicalTrials.gov and structuring it into a database, drug-AEs could be established for future drug development and repositioning. To our knowledge, current AE databases contain mainly U.S. Food and Drug Administration (FDA)-approved drugs. However, our database contains both FDA-approved and experimental compounds extracted from ClinicalTrials.gov. Our database contains 8,161 clinical trials of 3,102,675 patients and 713,103 reported AEs. We extracted the information from ClinicalTrials.gov using a set of python scripts, and then used regular expressions and a drug dictionary to process and structure relevant information into a relational database. We performed data mining and pattern analysis of drug-AEs in our database. Our database can serve as a tool to assist researchers to discover drug-AE relationships for developing, repositioning, and repurposing drugs.





(*ASSAY and Drug Development Technologies*, 2016)

Big Data Mining of Adverse Events in ClinicalTrials.gov

Table 2. Summary Statistics of the Database

Description	Counts
Number of clinical trials	8,161
Number of patients	3,102,675
Number of drugs	1,248
Number of FDA-approved drugs	634
Number of non-FDA-approved drugs	614
Number of cohorts	20,739
Number of adverse event names	31,267
Number of adverse event categories	26
Number of reported adverse events	713,103
Number of conditions	3,279

FDA, U.S. Food and Drug Administration.

Big Data Mining of Adverse Events in ClinicalTrials.gov

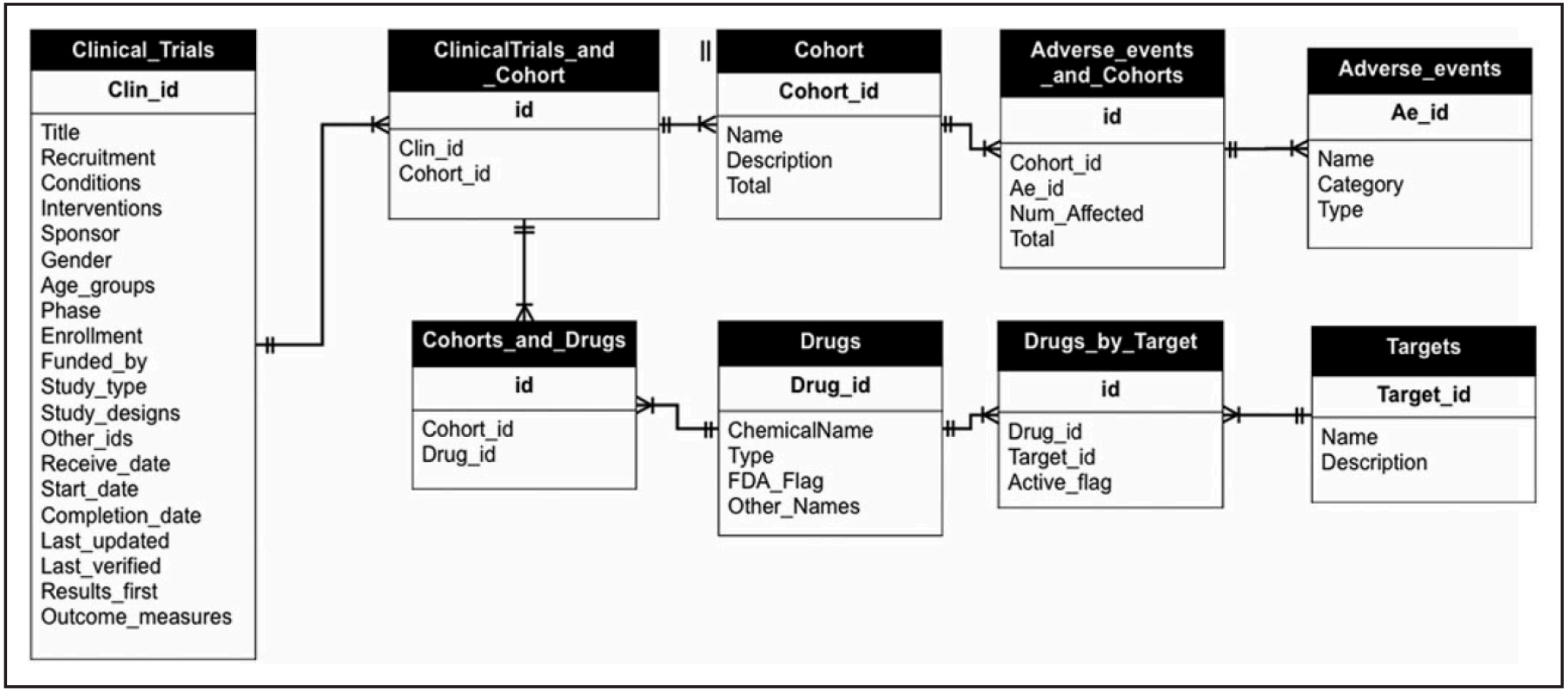


Fig. 2. Entity-relationship model of the AEDB. AEDB, adverse event database.

Big Data Mining of Adverse Events in ClinicalTrials.gov

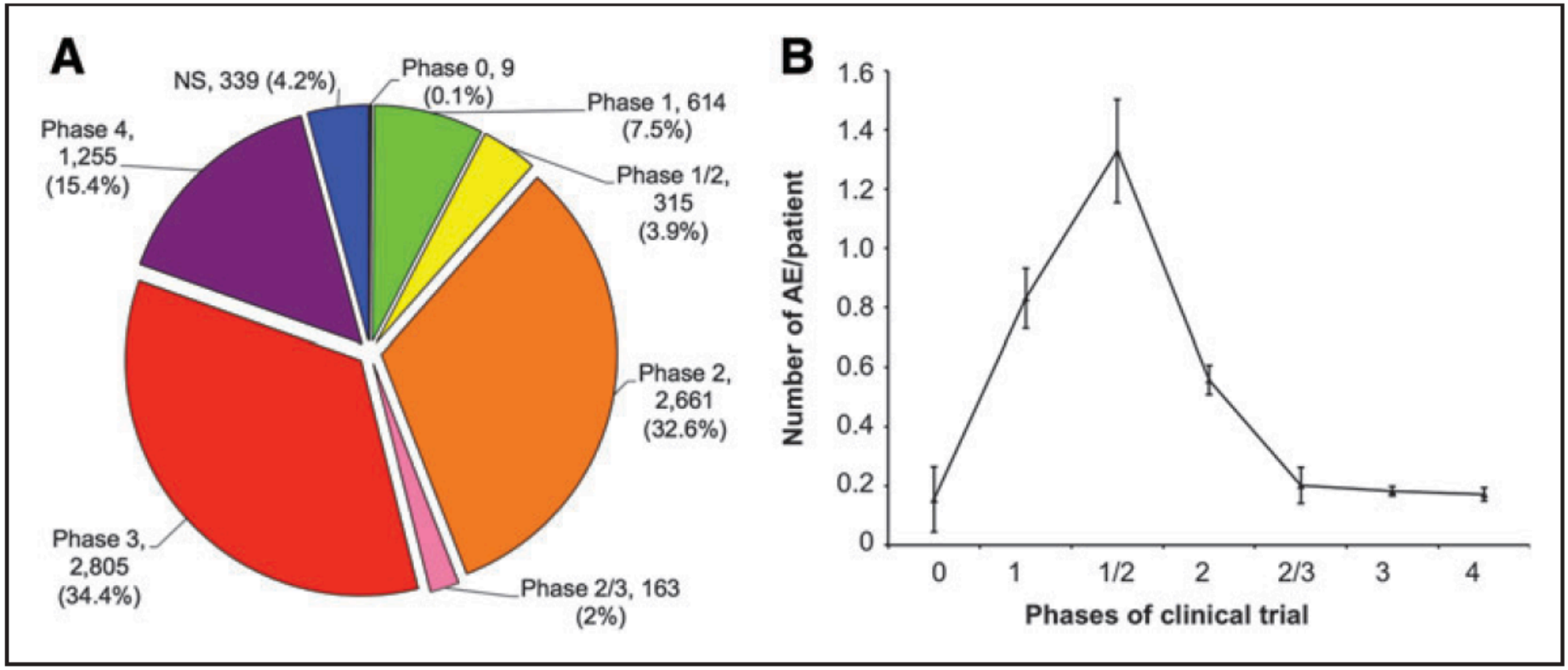


Fig. 4. AEs in different phases of clinical trials. **(A)** Distribution of the different phases of clinical trials. **(B)** Average number of AEs per patient in different phases of clinical trials. N.S., not specified. Error bar represents the standard error of the mean. Color images available online at www.liebertpub.com/adt

Big Data Mining of Adverse Events in ClinicalTrials.gov

Table 3. Vascular Event Proportional Reporting Ratios for the Five Kinase Inhibitors Commonly Used to Treat Chronic Myelogenous Leukemia Patients

Kinase Inhibitors	Vascular Adverse Events						
	Peripheral Arterial Occlusive Disease	Embolism	Hypertension	Platelet Dysfunction	Hyperglycemia	Hair Loss Alopecia	Vascular Disorders
Imatinib	7.416	4.874	4.550	10.929	4.944	4.398	5.481
Dasatinib	NA	2.959	8.161	17.624	10.720	4.427	4.263
Nilotinib	31.497	2.070	10.541	11.604	14.998	4.239	4.810
Bosutinib	NA	5.457	7.719	10.197	5.272	4.443	3.719
Ponatinib	374.810	NA	41.811	69.044	NA	7.486	9.158
Placebo	2.065	1.861	1.957	0.326	1.404	0.000	1.836

NA, not applicable due to no data.

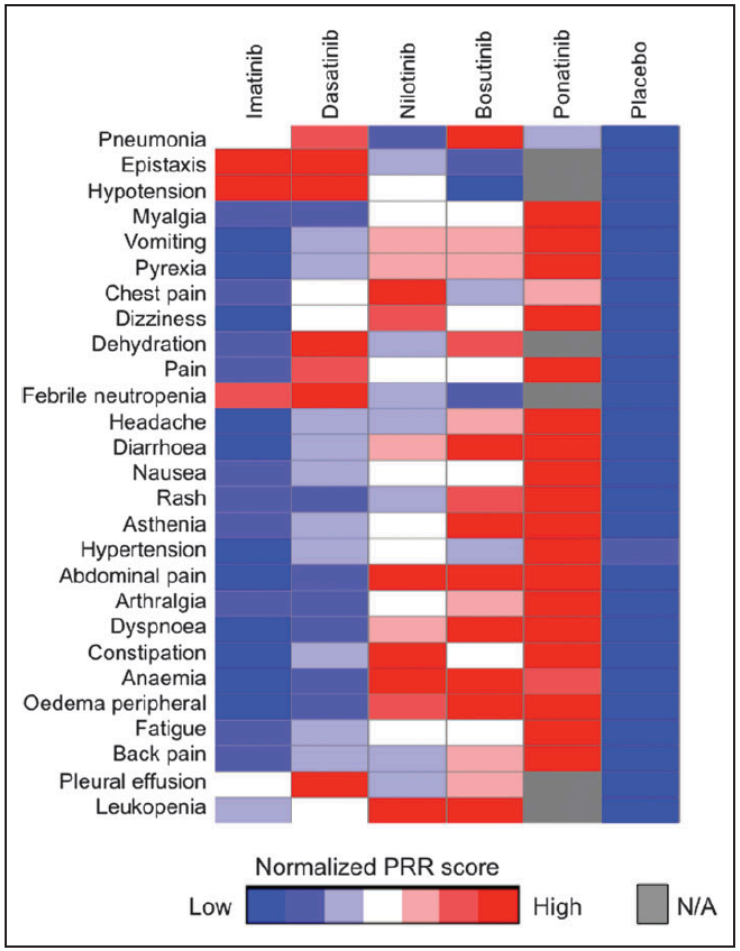


Fig. 6. Kinase inhibitor-AE relationships. Heatmap of the PRR of the top 10 AEs of imatinib, dasatinib, nilotinib, bosutinib, ponatinib, and placebo. The PRR is normalized per AE, where *red* and *blue* colors indicate high and low frequencies, respectively. PRR, proportional reporting ratio. Color images available online at www.liebertpub.com/adt

Other Data Mining Studies in ClinicalTrials.gov

Extracting genetic alteration information for personalized cancer therapy from ClinicalTrials.gov

RECEIVED 15 September 2015
REVISED 7 December 2015
ACCEPTED 13 January 2016

Jun Xu,[1] Hee-Jin Lee,[1] Jia Zeng,[2] Yonghui Wu,[1] Yaoyun Zhang,[1] Liang-Chin Huang,[1] Amber Johnson,[2] Vijaykumar Holla,[2] Ann M. Bailey,[2] Trevor Cohen,[1] Funda Meric-Bernstam,[2,3] Elmer V. Bernstam,[1,4] Hua Xu[1]

ABSTRACT

Objective: Clinical trials investigating drugs that target specific genetic alterations in tumors are important for promoting personalized cancer therapy. The goal of this project is to create a knowledge base of cancer treatment trials with annotations about genetic alterations from ClinicalTrials.gov.

Methods: We developed a semi-automatic framework that combines advanced text-processing techniques with manual review to curate genetic alteration information in cancer trials. The framework consists of a document classification system to identify cancer treatment trials from ClinicalTrials.gov and an information extraction system to extract gene and alteration pairs from the Title and Eligibility Criteria sections of clinical trials. By applying the framework to trials at ClinicalTrials.gov, we created a knowledge base of cancer treatment trials with genetic alteration annotations. We then evaluated each component of the framework against manually reviewed sets of clinical trials and generated descriptive statistics of the knowledge base.

Results and Discussion: The automated cancer treatment trial identification system achieved a high precision of 0.9944. Together with the manual review process, it identified 20 193 cancer treatment trials from ClinicalTrials.gov. The automated gene-alteration extraction system achieved a precision of 0.8300 and a recall of 0.6803. After validation by manual review, we generated a knowledge base of 2024 cancer trials that are labeled with specific genetic alteration information. Analysis of the knowledge base revealed the trend of increased use of targeted therapy for cancer, as well as top frequent gene-alteration pairs of interest. We expect this knowledge base to be a valuable resource for physicians and patients who are seeking information about personalized cancer therapy.

Other Data Mining Studies in ClinicalTrials.gov

Figure 1: Overview of the 2-step framework

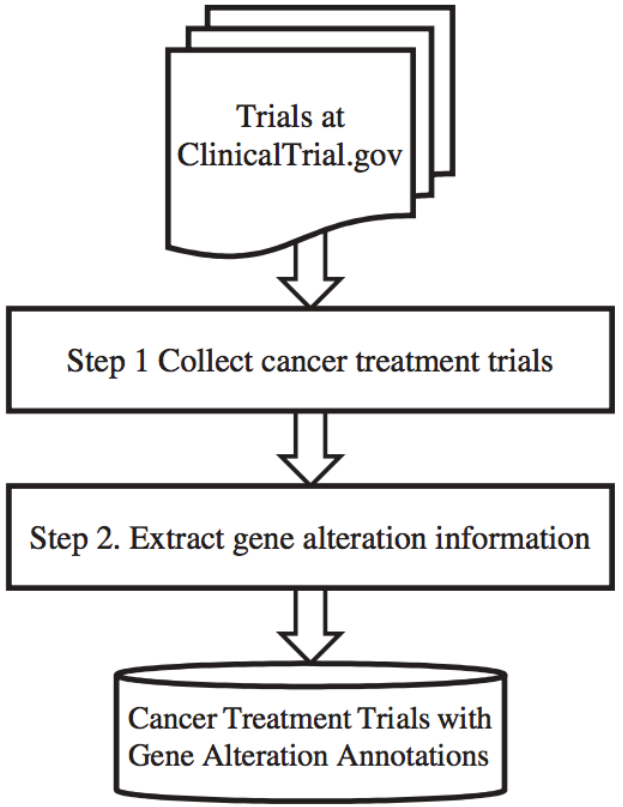
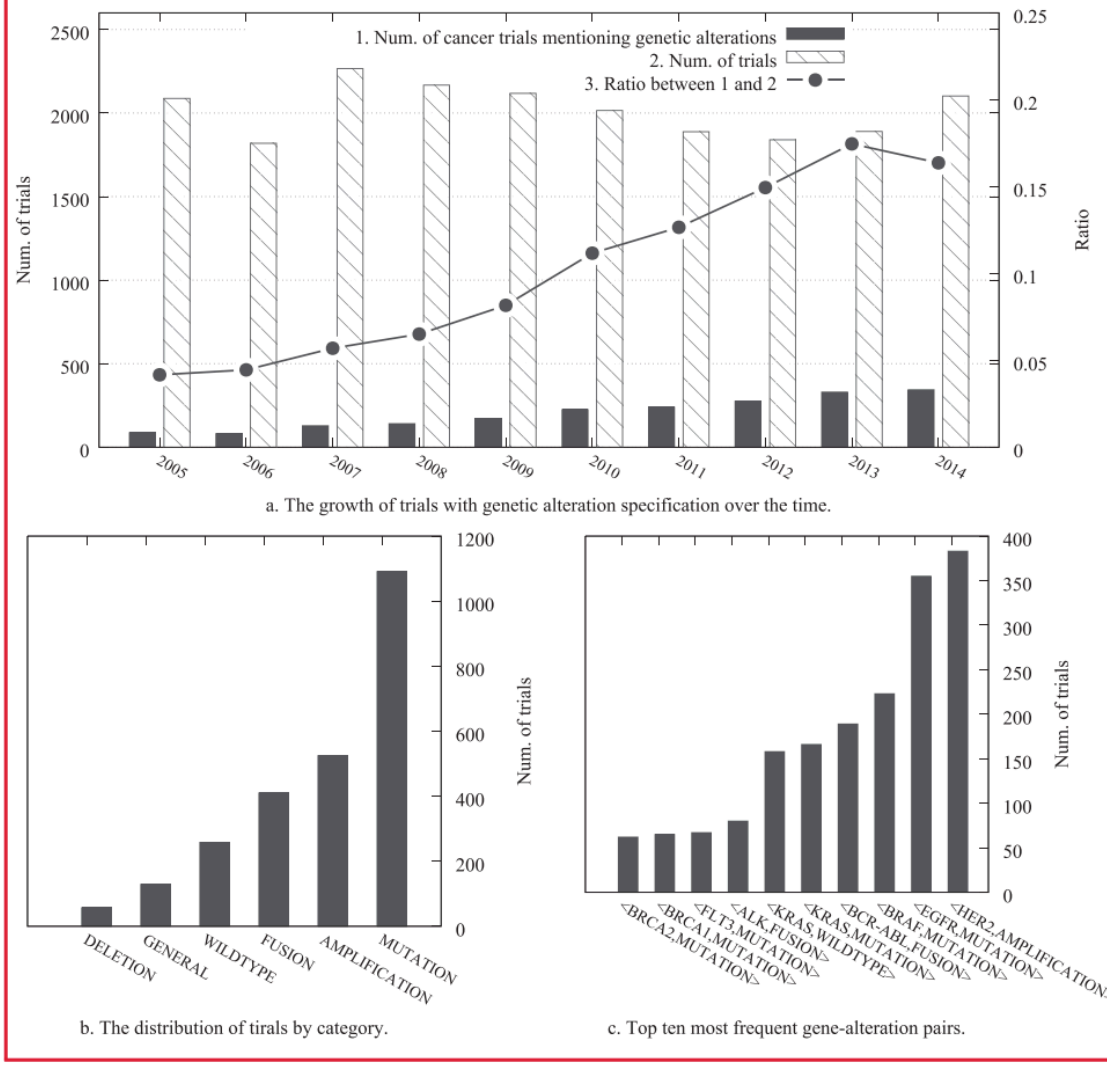


Figure 6: Descriptive statistics of the genetic alteration knowledge base of cancer trials



Other Data Mining Studies in ClinicalTrials.gov

Learning disease relationships from clinical drug trials

RECEIVED 11 August 2015
REVISED 23 December 2015
ACCEPTED 3 January 2016

Bryan Haslam[1] and Luis Perez-Breva[2]

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.
OXFORD UNIVERSITY PRESS

ABSTRACT

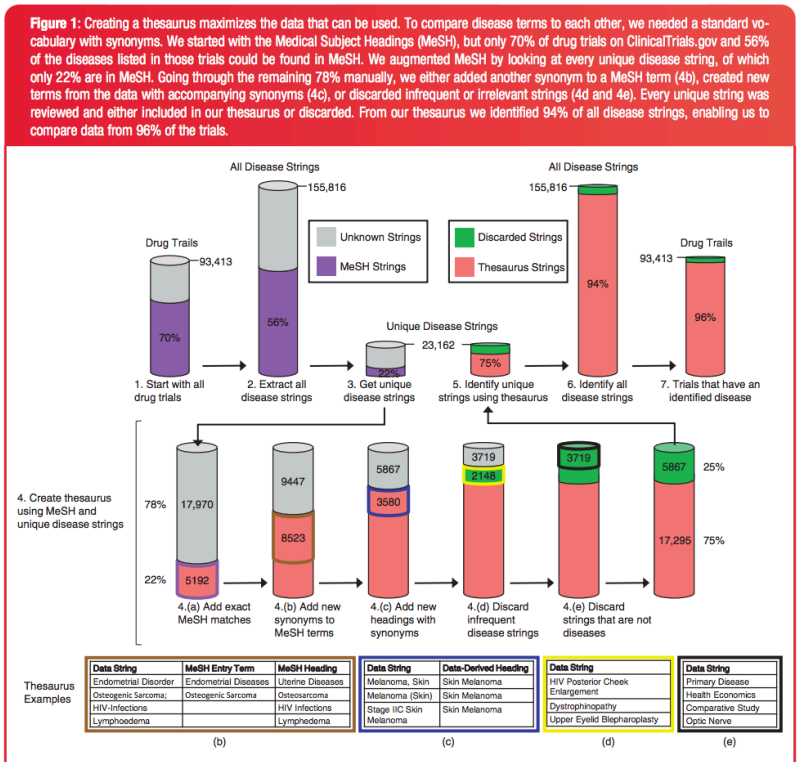
Objective Our objective is to test the limits of the assumption that better learning from data in medicine requires more granular data. We hypothesize that clinical trial metadata contains latent scientific, clinical, and regulatory expert knowledge that can be accessed to draw conclusions about the underlying biology of diseases. We seek to demonstrate that this latent information can be uncovered from the whole body of clinical trials.

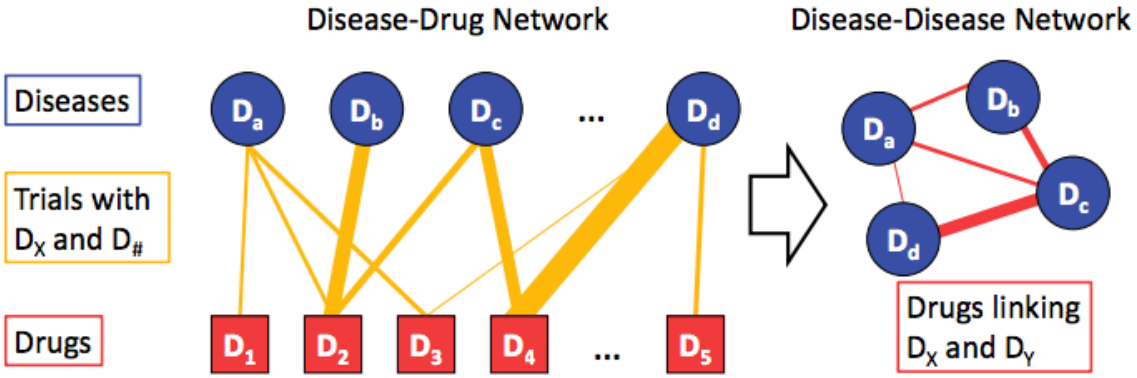
Materials and Methods We extract free-text metadata from 93 654 clinical drug trials and introduce a representation that allows us to compare different trials. We then construct a network of diseases using only the trial metadata. We view each trial as the summation of expert knowledge of biological mechanisms and medical evidence linking a disease to a drug believed to modulate the pathways of that disease. Our network representation allows us to visualize disease relationships based on this underlying information.

Results Our disease network shows surprising agreement with another disease network based on genetic data and on the Medical Subject Headings (MeSH) taxonomy, yet also contains unique disease similarities.

Discussion and Conclusion The agreement of our results with other sources indicates that our premise regarding latent expert knowledge holds. The disease relationships unique to our network may be used to generate hypotheses for future biological and clinical research as well as drug repurposing and design. Our results provide an example of using experimental data on humans to generate biologically useful information and point to a set of new and promising strategies to link clinical outcomes data back to biological research.

Other Data Mining Studies in ClinicalTrials.gov





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

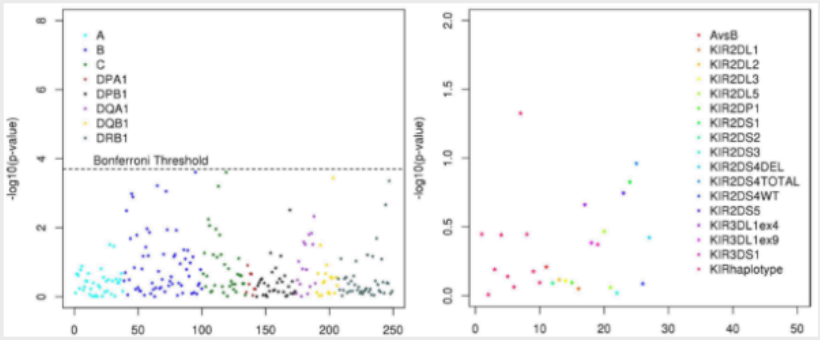
more

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



Sponsored by:

- National Institute of Allergy and Infectious Diseases (NIAID)
- National Institutes of Health (NIH)
- Health and Human Services (HHS)

Contact Us | Privacy Policy | Disclaimer | Accessibility

Recommended Browsers: Chrome, Firefox, Safari v7+, Internet Explorer v9.0+

Connect with Us

Like 79
Twitter
RSS Feed
Google Group

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

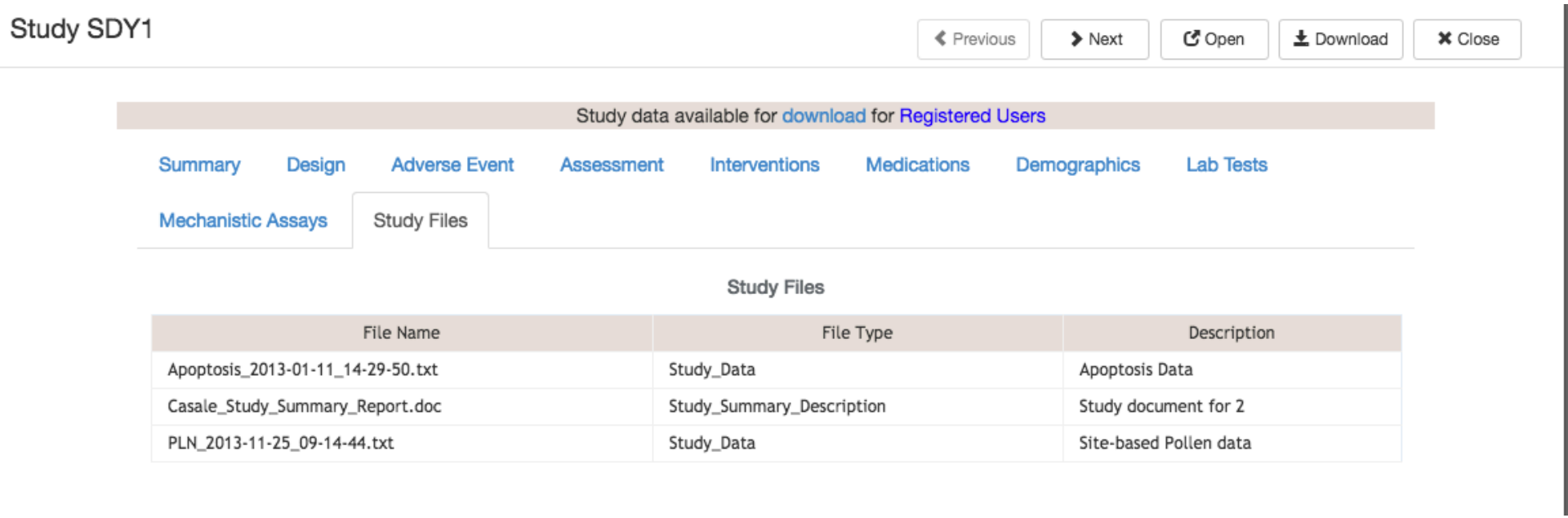
ARM4 = Immunotherapy with anti-IgE

ARM3 = Placebo Immunotherapy with anti-IgE

ARM2 = Immunotherapy with placebo anti-IgE

ARM1 = Placebo Immunotherapy with placebo anti-IgE

Clinical Trials Data



The Resilience Project

The Resilience Project

Join the Search. Be a Hero

The Resilience Project aims to discover hidden factors that protect people from disease.

Led by the Icahn Institute for Genomics at Mount Sinai, in collaboration with Sage Bionetworks and others worldwide, we are searching for people who, according to medical textbooks, should be sick but have somehow escaped typical signs and symptoms of disease.

These people are "resilient," protected by undiscovered genetic or environmental factors. Finding and studying these resilient individuals could pave the way to disease prevention and new treatments.

https://www.youtube.com/watch?v=Yagdvqn2YMU

Recently, we reported the first systematic search for resilience to hundreds of childhood diseases. The largest study of its kind, this retrospective study of more than 589,000 genomes was a key first step for the Resilience Project and was performed in collaboration with researchers from 23andMe, BGI, the Ontario Institute for Cancer Research, and other institutions. Click here to view the full study published in Nature Biotechnology in April 2016.

nature biotechnology ARTICLES

The Resilience Project

Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases

Rong Chen[1,2,12], Lisong Shi[1,2,12], Jörg Hakenberg[1,2], Brian Naughton[3,11], Pamela Sklar[1,2,4], Jianguo Zhang[5], Hanlin Zhou[5], Lifeng Tian[6], Om Prakash[7], Mathieu Lemire[8], Patrick Sleiman[6], Wei-yi Cheng[1,2], Wanting Chen[5], Hardik Shah[1,2], Yulan Shen[5], Menachem Fromer[1,2,4], Larsson Omberg[9], Matthew A Deardorff[6], Elaine Zackai[6], Jason R Bobe[1,2], Elissa Levin[1,2], Thomas J Hudson[8], Leif Groop[7], Jun Wang[10], Hakon Hakonarson[6], Anne Wojcicki[3], George A Diaz[1,2], Lisa Edelmann[1,2], Eric E Schadt[1,2] & Stephen H Friend[1,2,9]

Genetic studies of human disease have traditionally focused on the detection of disease-causing mutations in afflicted individuals. Here we describe a complementary approach that seeks to identify healthy individuals resilient to highly penetrant forms of genetic childhood disorders. A comprehensive screen of 874 genes in 589,306 genomes led to the identification of 13 adults harboring mutations for 8 severe Mendelian conditions, with no reported clinical manifestation of the indicated disease. Our findings demonstrate the promise of broadening genetic studies to systematically search for well individuals who are buffering the effects of rare, highly penetrant, deleterious mutations. They also indicate that incomplete penetrance for Mendelian diseases is likely more common than previously believed. The identification of resilient individuals may provide a first step toward uncovering protective genetic variants that could help elucidate the mechanisms of Mendelian diseases and new therapeutic strategies.

Nature Biotech 2016

The Resilience Project

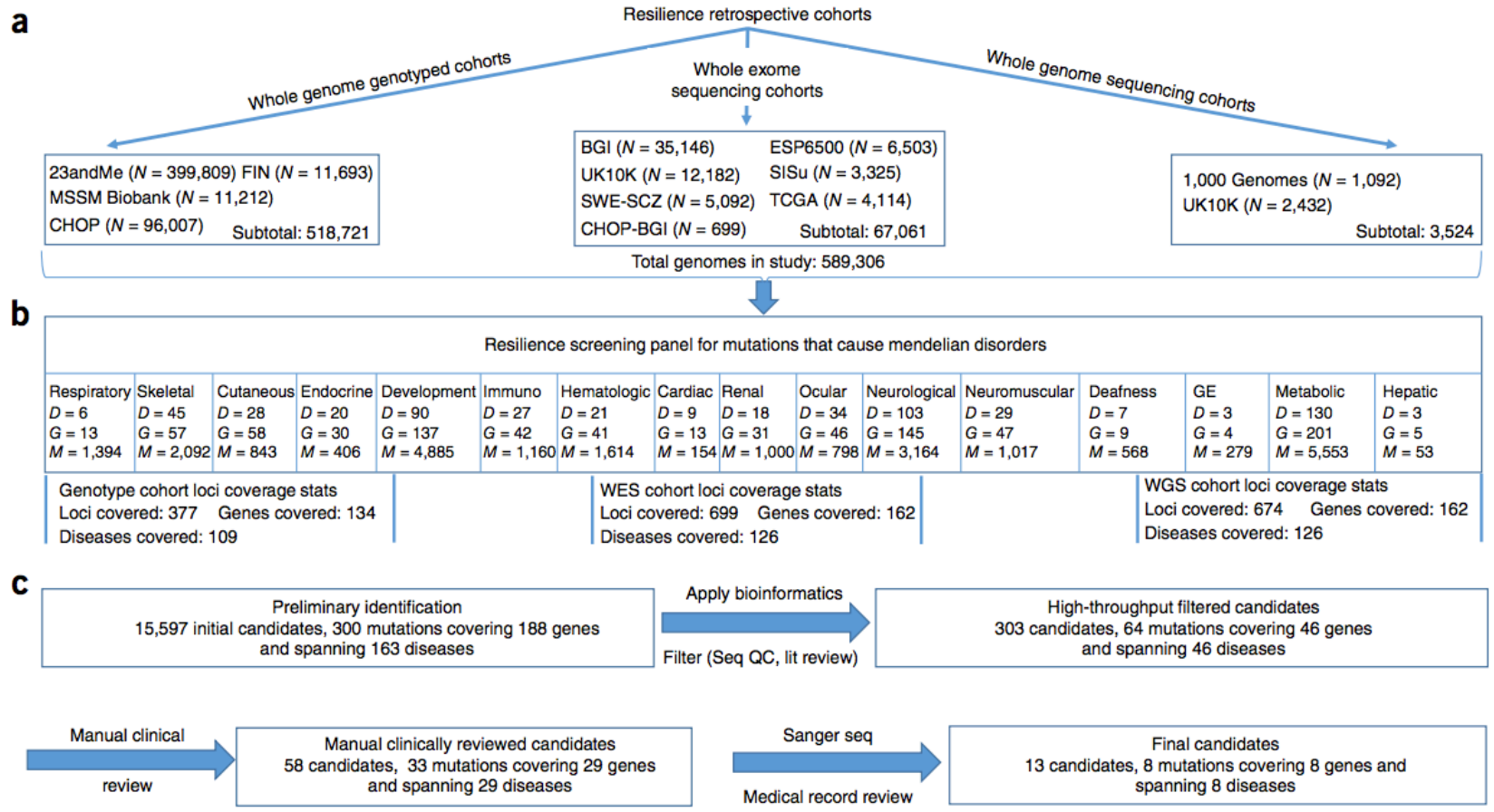


Figure 1 Study design and results for the retrospective search for resilient individuals. (**a**) A summary of the different cohorts and the genomic data available on those cohorts (see **Table 2** for more details). (**b**) The disease-causing genes and mutations that were assembled to construct our screening panel (more details in **Table 1** and **Supplementary Tables 1** and **2**). The D, G and M variables denote the number of diseases, genes and mutations, respectively, represented on our screening panel in the respective disease categories. The coverage statistics indicate the coverage achieved for the core allele panel in the genotype, WES and WGS cohorts. (**c**) Summaries for the different stages of the filtering process to identify candidate resilient individuals (see **Supplementary Fig. 1** and **Tables 3** and **4** for more details).

The Resilience Project

Table 2 Data sources used in current retrospective study

Sample source	Sample type	Sample size	Technology	Population
TCGA	Matched normal tissues for 17 tumor types	4,114	WES and WGS	No population-specific data acquired
Mount Sinai BioBank	Various diseases	11,212	Genotyping array	Self-reported ethnicities
23andMe	Mixed	399,809	Genotyping array	No population-specific data acquired
1000 Genomes Projects	Healthy	1,092	Low pass WGS	African, American, Asian and European; subcategories available
ESP6500	Various diseases	6,503	WES	African-American and European-American (both USA)
UK10K[a]	Cohorts; neurodevelopmental disorders; obesity samples; rare diseases	14,614	Partly WGS, partly WES	Mostly UK and Finland; no population-specific data acquired
SISu[a,b]	Case-control mixed	3,325	WES	Finnish
FINN[a,c]	Case-control mixed	11,693	Genotyping array	Finnish
CHOP-BGI	Case-control mixed	699	WES	Mixed
CHOP	Case-control mixed	96,007	Genotyping array	Mixed
BGI	Case-control mixed	35,146	Partly WGS, partly WES	Mixed
SWE-SCZ	Schizophrenia cases and controls	5,092	WES	Swedish (some samples with partial Finnish ancestry)
Total WES/WGS		70,585		
Total genotyping		518,721		
Grand total		589,306		

[a]For detailed data, see **Supplementary Table 4**. [b]SISu, Sequencing Initiative Suomi (http://www.sisuproject.fi/): consortia including FINRISK, GoT2D (only the Fusion and Botnia studies), H2000, METSIM, NFBC66 and Finnish samples from the 1000 Genomes projects. [c]FINN, a subset of cohorts from SISu: FINRISK, EUFAM, Finnish Twin study and Migraine Study, with genome-wide genotype data.

The Resilience Project

Table 4 13 Candidates identified in the Resilience Project

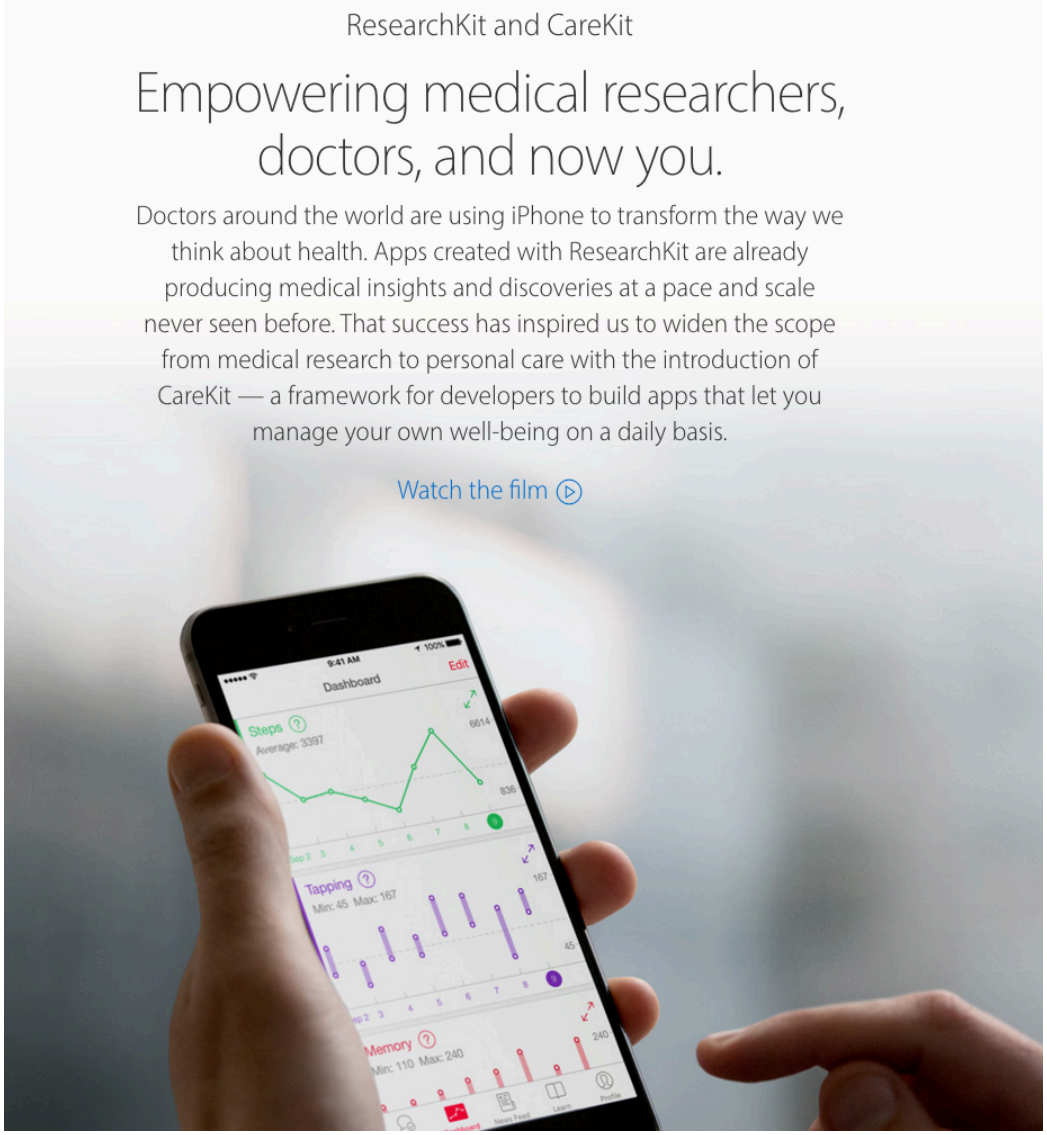
Phenotype	Gene	Mutation (cDNA; protein (reference))	Genomic coordinate (hg19)	Mutation severity	Candidate confidence	Panel source	No. of candidates	Zygosity	Data source	Level of support for candidacy[a]	Sample status	Population carrier frequency[b] 1KG	Population carrier frequency[b] ESP
Cystic fibrosis	*CFTR*	c.1558G>T; p.V520F (NM_000492.3)	Chr7 117199683	Severe pulmonary disease, childhood-onset	Strong	Core allele panel	3	hom	23andMe	C1,C2,C3, G1,G2,G3	2 adults, one declared no manifestation	0.00	0.00
Smith-Lemli-Opitz syndrome	*DHCR7*	c.964-1G>C (NM_001360.2)	Chr11: 71146886	Severe developmental disorder, probably embryonic lethal	Strong	Core allele panel	2	hom	UK10K	C1,C2, G1,G2	Not obtained	0.0052	0.011
Familial dysautonomia	*IKBKAP*	c.2204+6T>C (NM_003640.3)	Chr9: 111662096	Severe neurological disease, high mortality in early childhood	Strong	Core allele panel	1	hom	23andMe	C1,C2, G1,G2,G3	No disease reported by individual	0.00	0.0012 (only in EA)
Epidermolysis Bullosa simplex	*KRT14*	c.373C>T; p.R125C (NM_000526.4)	Chr17: 39742714	Severe dermatologic condition, infantile onset	Strong	Core allele panel	1	het	BGI	C1,C2,C3, G1,G2	No disease reported by individual	0.00	0.00
Pfeiffer syndrome	*FGFR1*	c.755C>G; p.P252R (NM_023110.2)	Chr8: 38282208	Severe congenital skeletal dysplasia with variable expressivity	Strong[c]	Core allele panel	1	het	SWE-SCZ	C1,C2,C3, G1,G2,G3	No abnormal morphology reported in discharged health information	0.00	0.00
APECED	*AIRE*	c.769C>T; p.R257* (NM_000383.2)	Chr21: 45709656	Severe childhood-onset autoimmune disease	Strong	Core allele panel	1	hom	23andMe	C1,C2,C3, G1,G2	No disease reported by individual	0.00	0.00015
Acampomelic campomelic dysplasia	*SOX9*	c.1320C>G; p.Y440* (NM_000346.3)	Chr17: 70120318	Severe skeletal dysplasia with early childhood death	Strong	Expanded panel	1	het	FINN	C1,C2, G1,G2	Not obtained	0.00	0.00
Atelosteogenesis	*SLC26A2*	c.835C>T; p.R279W (NM_000112.3)	Chr5: 149359991	Severe early-onset skeletal dysplasia with variable expressivity	Moderate[d]	Expanded panel	3	hom	23andMe	C1,C2, G1,G2	Not obtained	0.0028	0.0023

[a]See **Table 5** for code definitions. [b]Carrier frequencies from combined ethnicities. [c]Individual was categorized as strong candidate due to lack of dysmorphic features. [d]Individual with variable phenotypes have been reported with the mutation[37].
EA, European American.

Table 5 Status codes for different levels of support identified during follow up of candidate resilient individuals

Support type	Status code	Status description for different levels of support for candidacy
Clinical validation	C1	Pass criteria for severity and penetrance for specific mutation set and reviewed by clinical specialist
	C2	Reference in literature found that can be cited for that mutation
	C3	Individual's clinical record examined - lacking classical presentation by "chart review" and family history
	C4	Individual is able to be recontacted to confirm atypical clinical presentation
Genetic validation	G1	Genotype call made
	G2	Review of primary sequencing/genotyping data
	G3	Resequencing of the sample
	G4	Work-up to rule out mosaic
Biomedical validation	B	Clinical test performed to determine if the individual harbor expected biomedical characteristics (enzyme activity, blood count, organ function etc.)

Apple ResearchKit (mobile data)



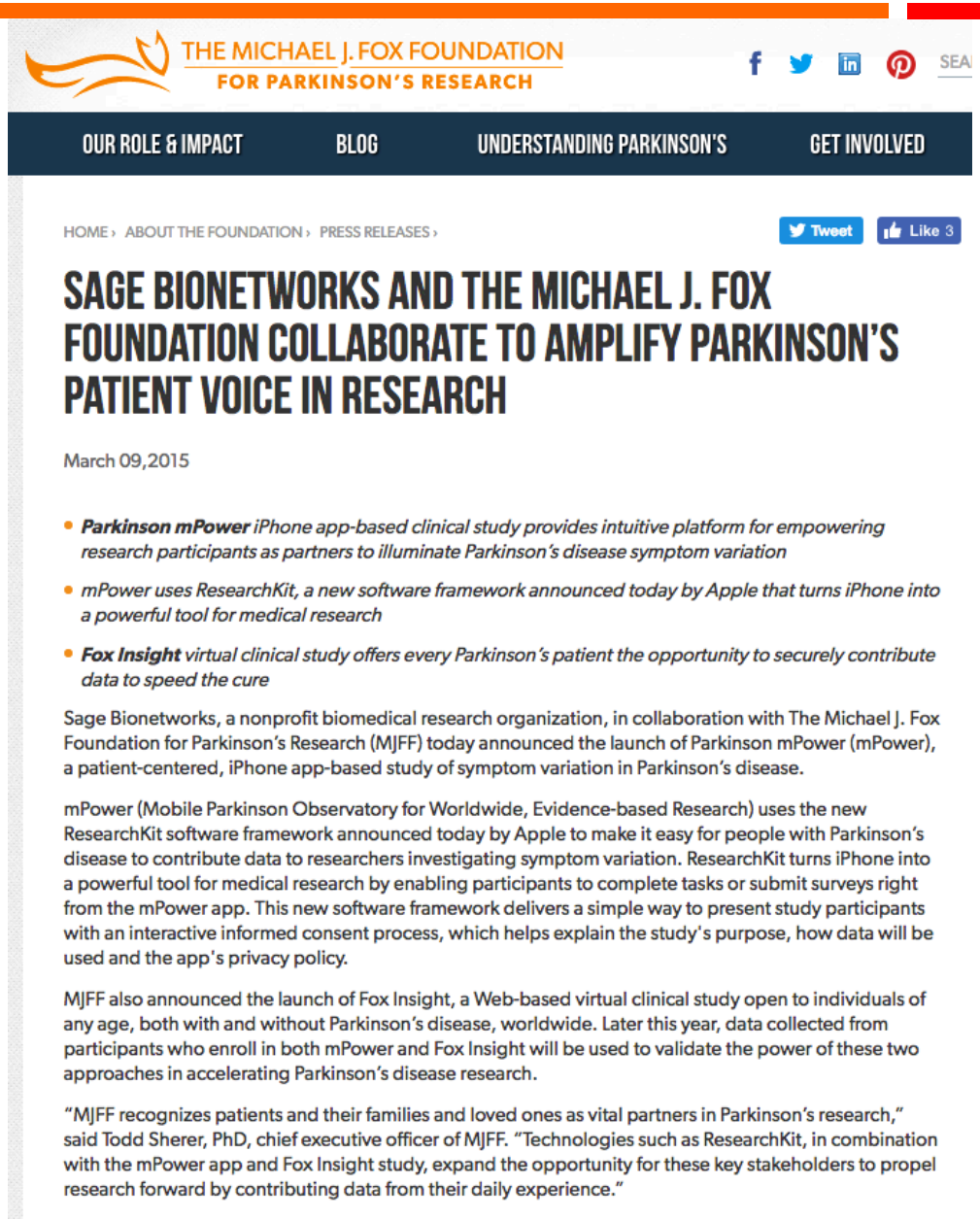
http://www.apple.com/researchkit/

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*

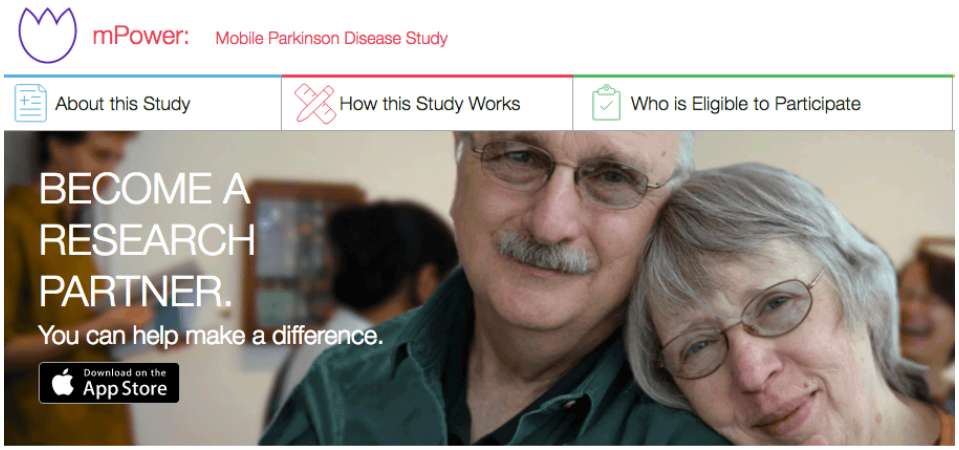
Sage Bionetworks, a nonprofit biomedical research organization, in collaboration with The Michael J. Fox Foundation for Parkinson's Research (MJFF) today announced the launch of Parkinson mPower (mPower), a patient-centered, iPhone app-based study of symptom variation in Parkinson's disease.

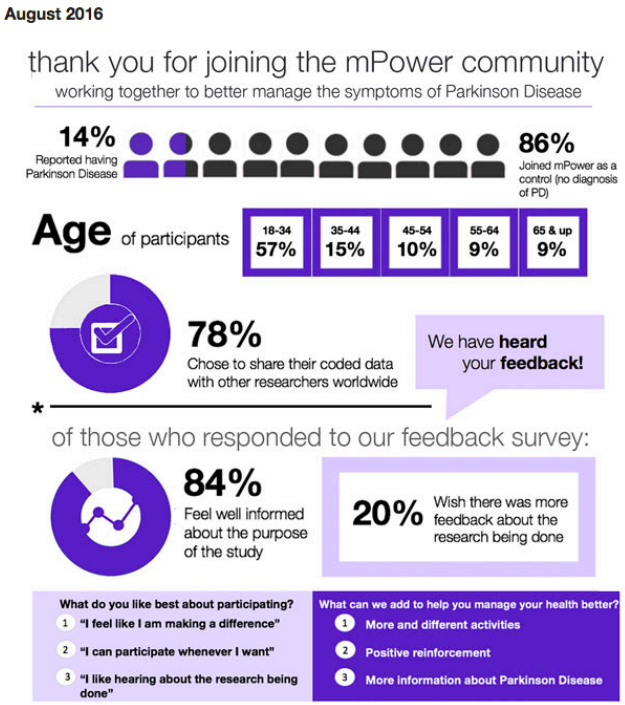
mPower (Mobile Parkinson Observatory for Worldwide, Evidence-based Research) uses the new ResearchKit software framework announced today by Apple to make it easy for people with Parkinson's disease to contribute data to researchers investigating symptom variation. ResearchKit turns iPhone into a powerful 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

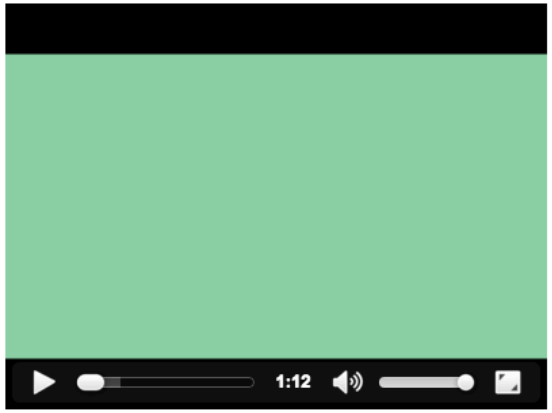
How can we better manage the symptoms of Parkinson's disease (PD) together? Whether you have PD, are touched by someone who has or has had PD or you want to help, we invite you to participate in this study. Become a research partner!

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

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]

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

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

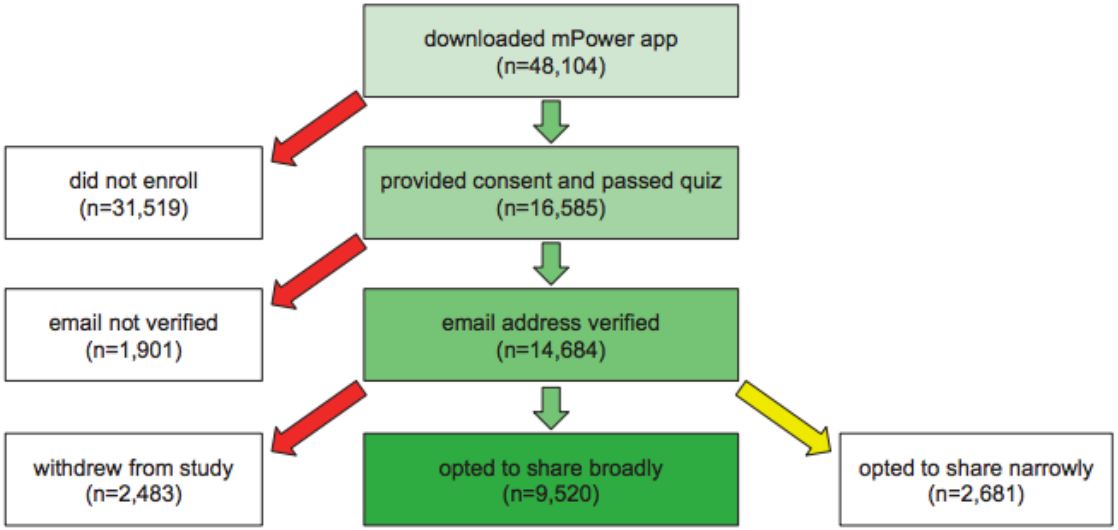


Figure 1. mPower study cohort description.

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.

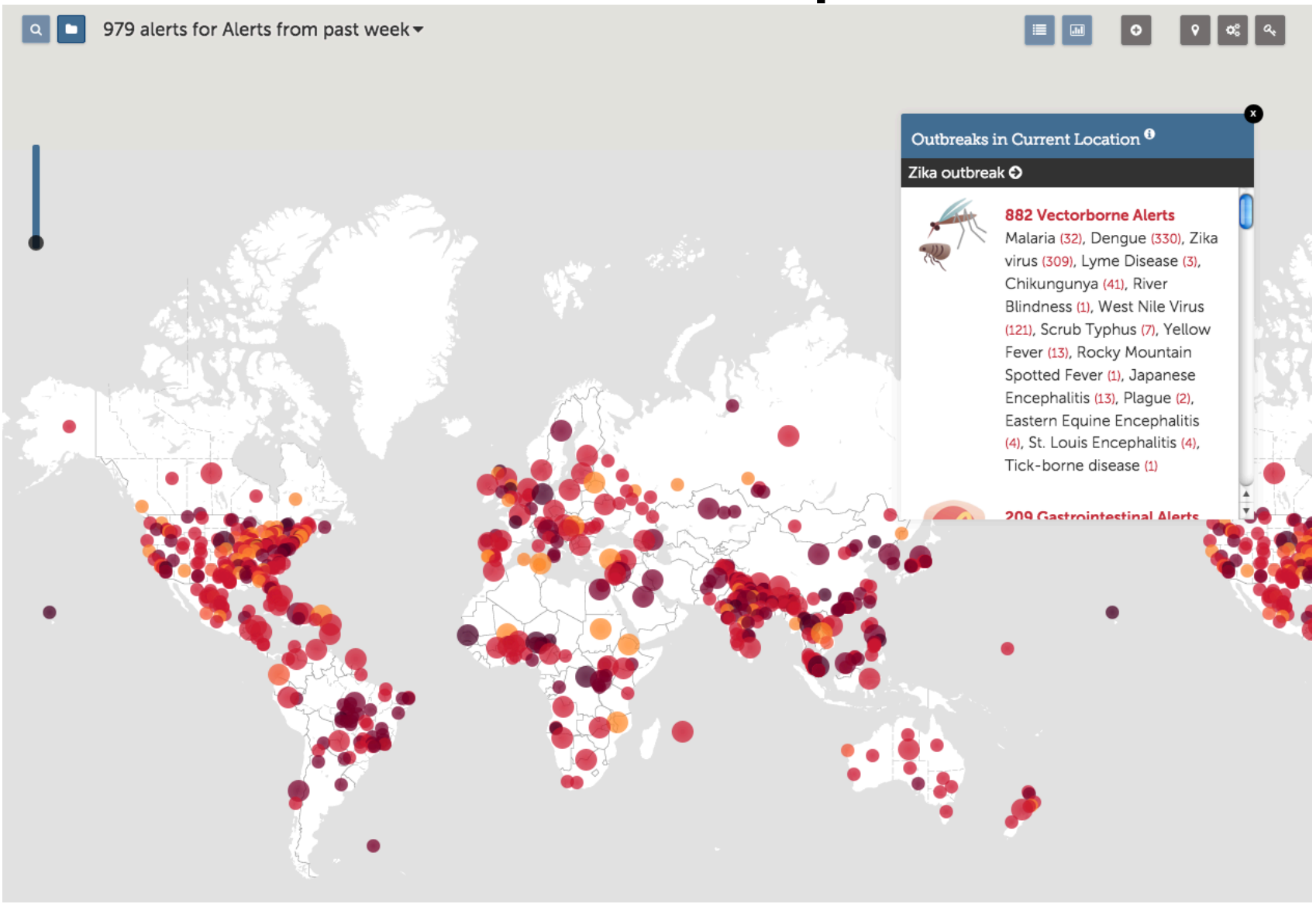
Software Tools

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

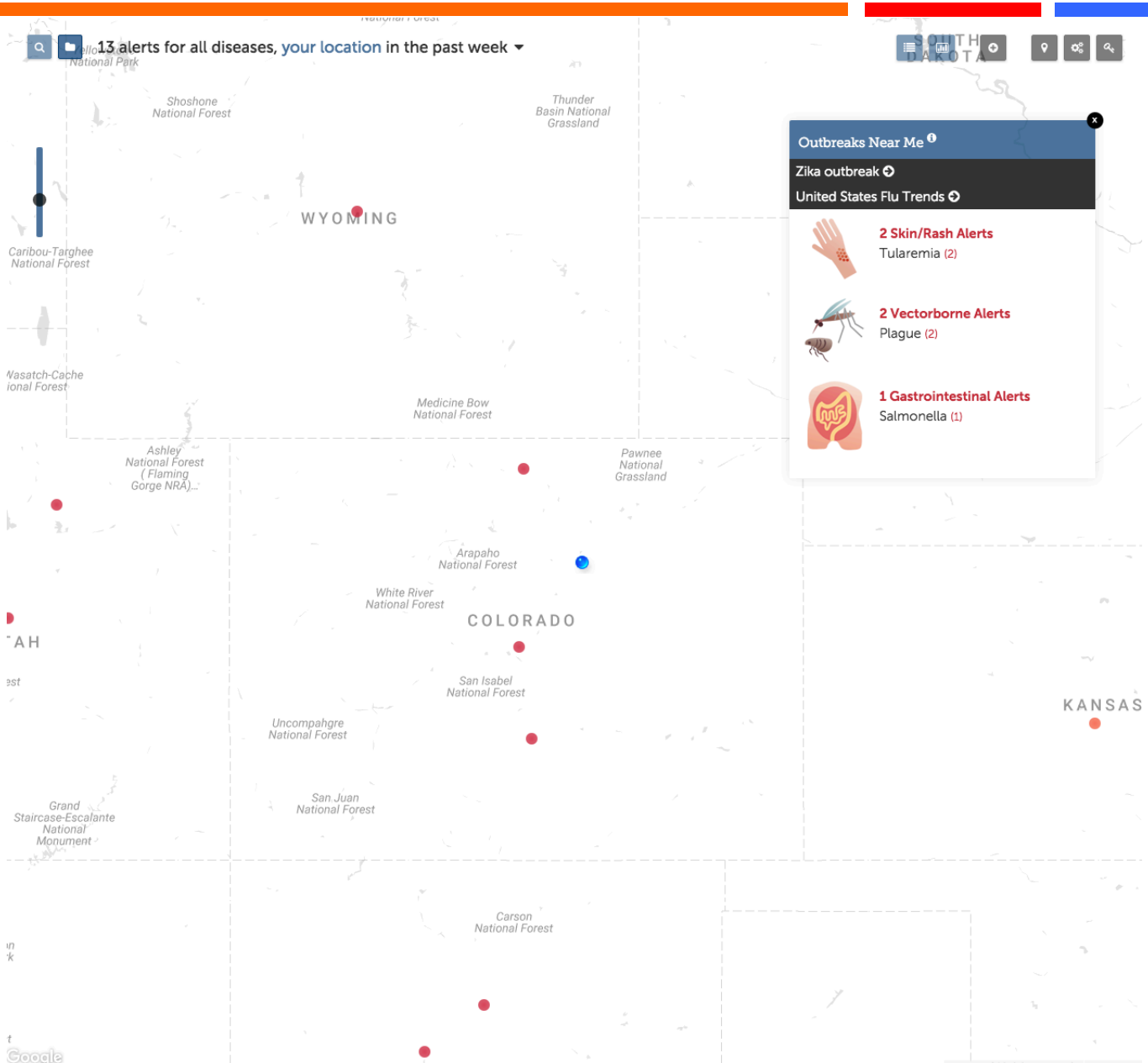
- Google Maps
- GoogleMapAPI for PHP
- Google Translate API
- xajax PHP AJAX library

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

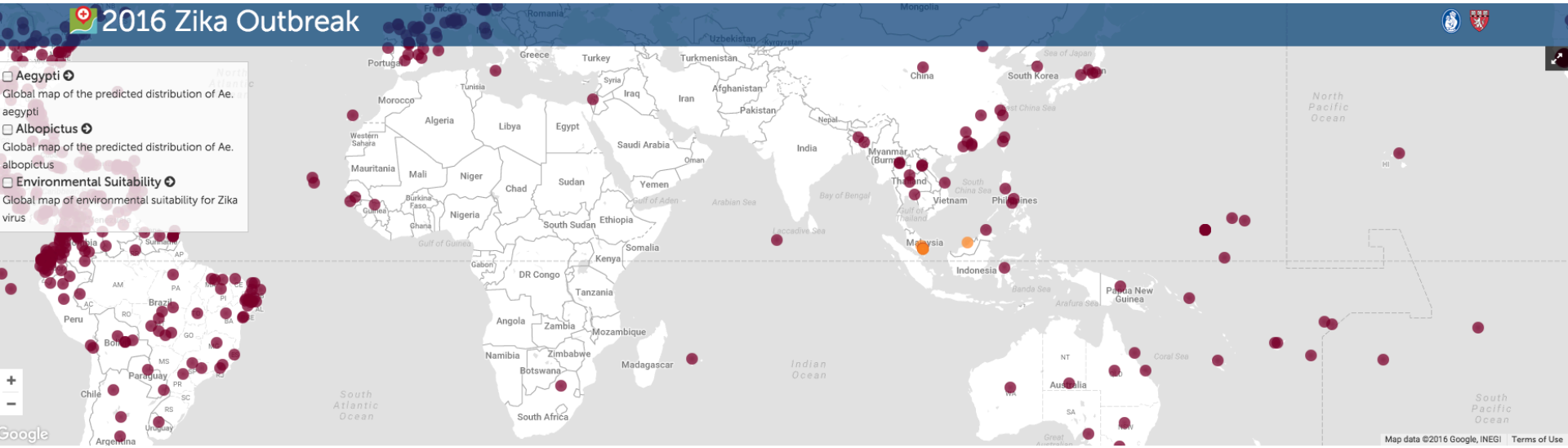
HealthMap

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

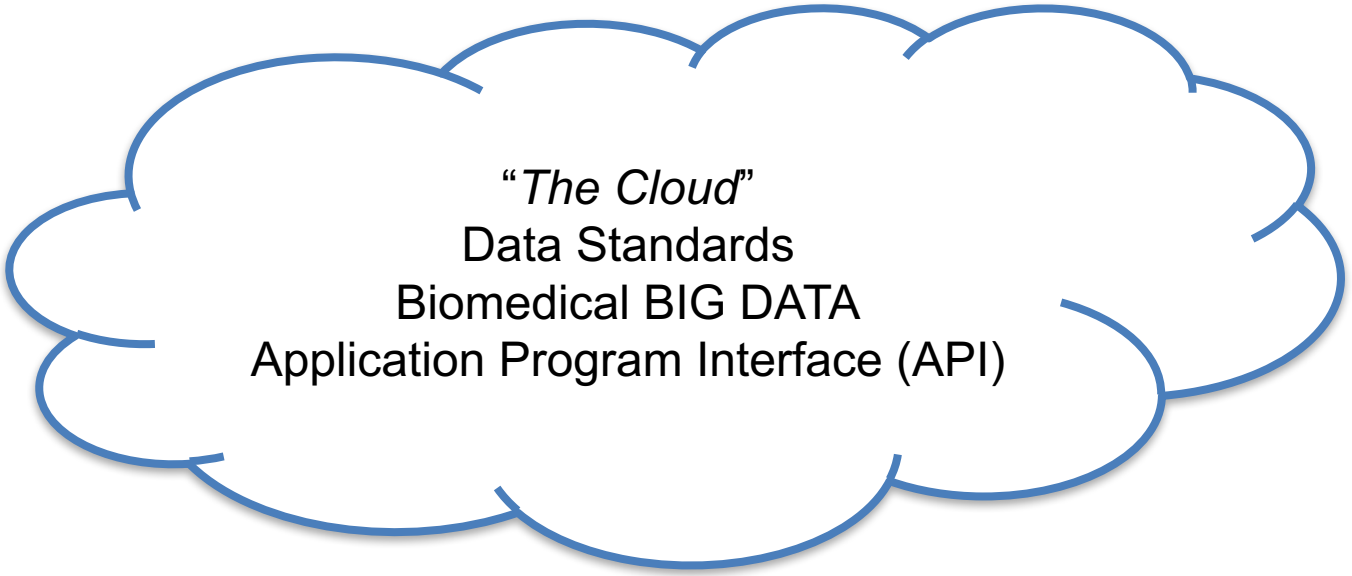
Malaysia reports new case of infection in women zika ... - Radio Havana Cuba

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

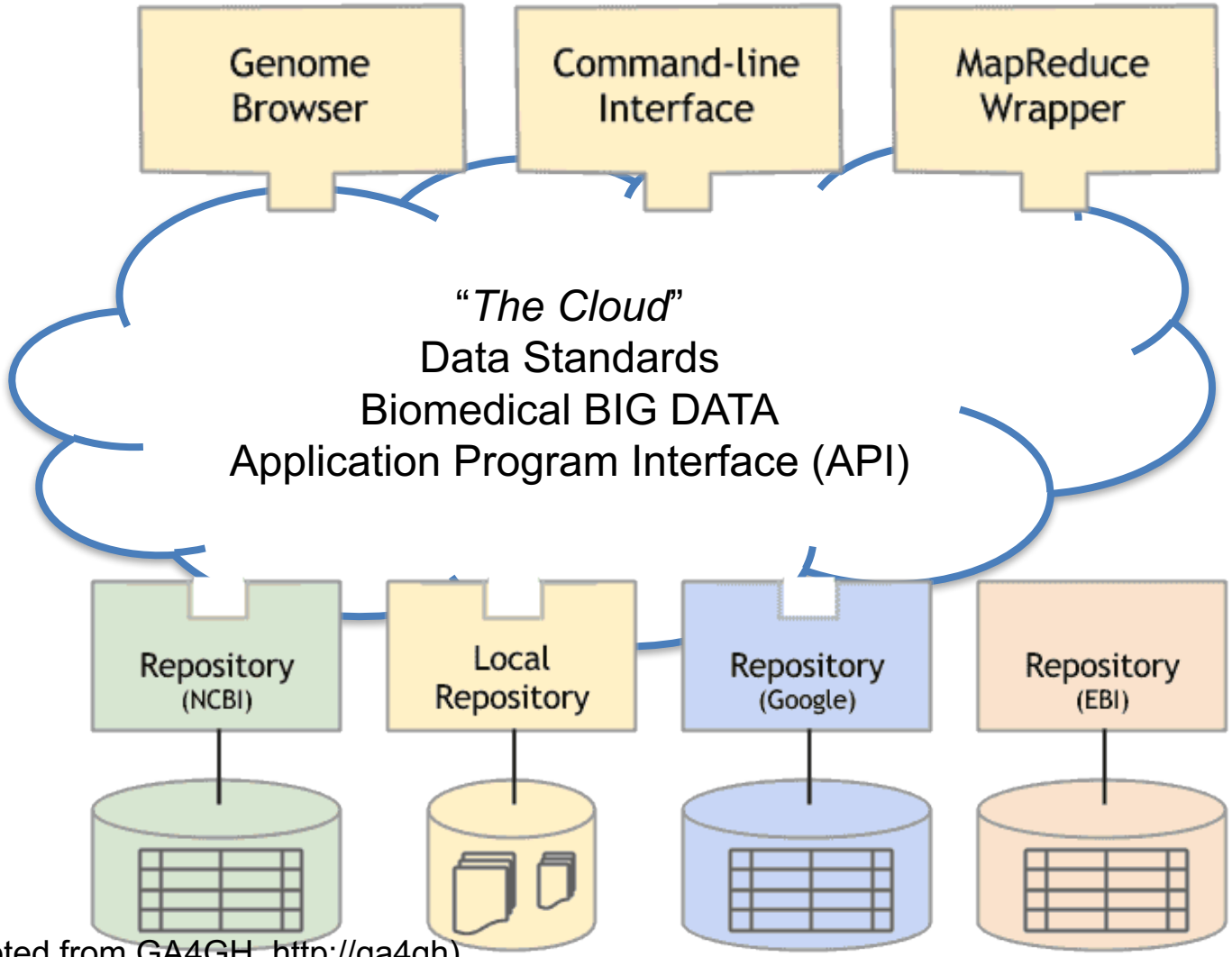
ZIKA VIRUS UPDATE: As of 12pm, 7 September, MOH has confirmed eight new cases of locally transmitted Zika virus... https://t.co/FSvCbil0Cl

- 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

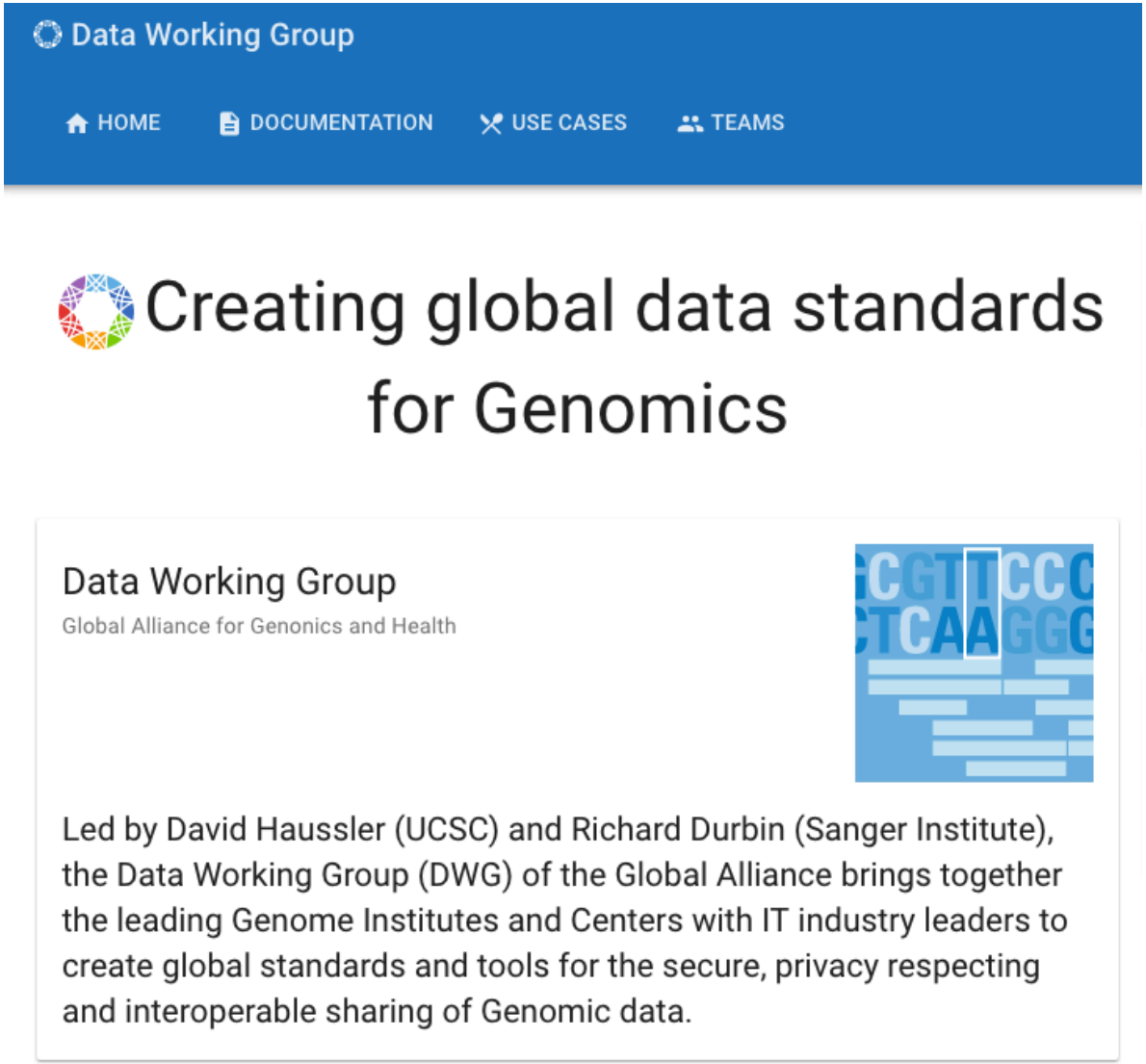


Ideal World of Biomedical Big data
Interoperability



(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 to find unexpected and interesting *knowledge*