UNIVERSITY OF CALIFORNIA

Translating a Trillion Points of Open Data into Diagnostics, Therapies and New Insights in Health and Disease

Atul Butte, MD, PhD

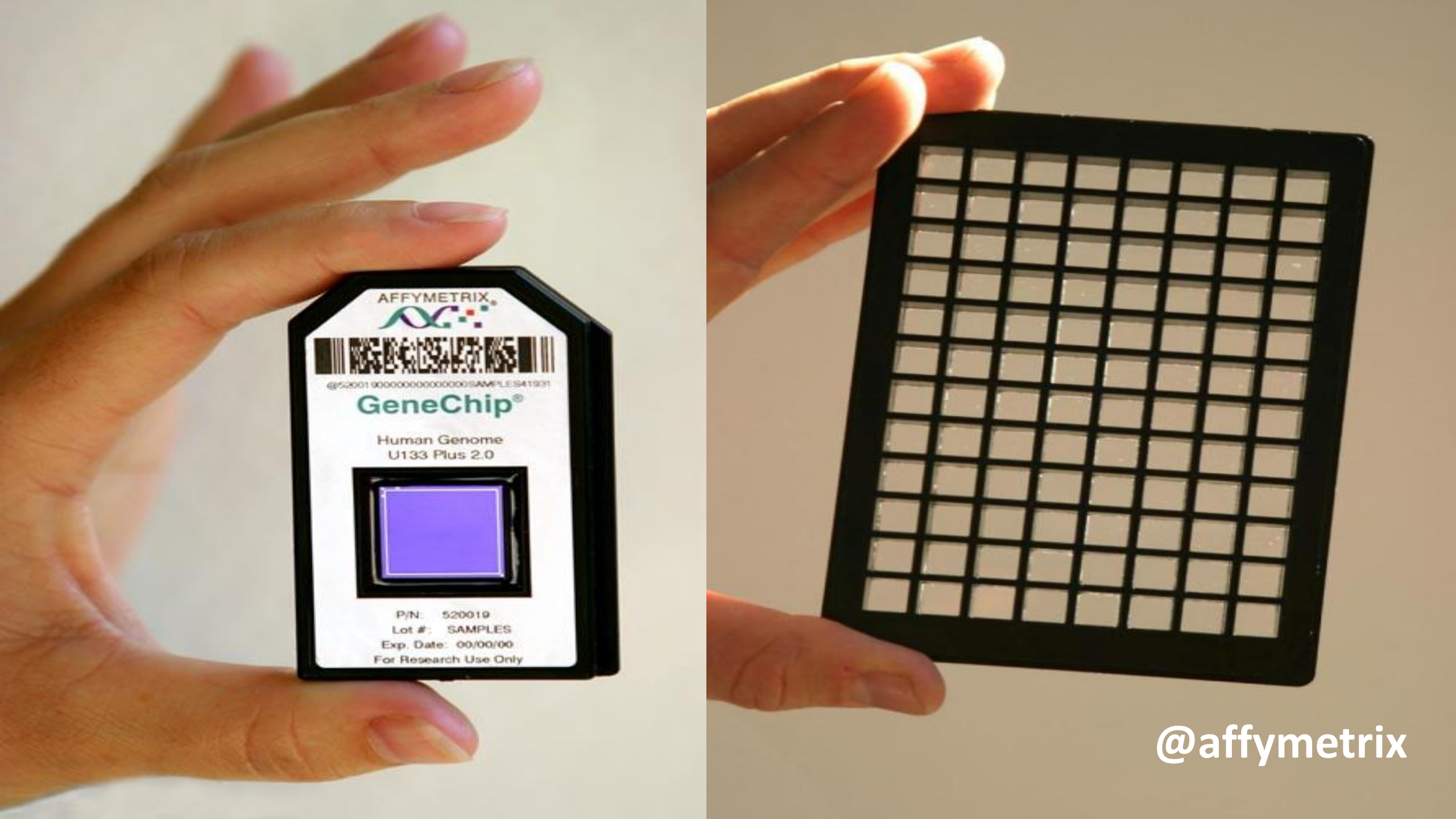
Chief Data Scientist, University of California Health (UC Health)

Director, Bakar Computational Health Sciences Institute, UCSF

Priscilla Chan and Mark Zuckerberg Distinguished Professor

Conflicts of Interest

- Scientific founder and advisory board membership
  - Genstruct
  - NuMedii
  - Personalis
  - Carmenta
- Honoraria for talks
  - Lilly
  - Pfizer
  - Siemens
  - Bristol Myers Squibb
  - AstraZeneca
  - Roche
  - Genentech
  - Warburg Pincus
  - CRG
  - AbbVie
  - Westat
- Past or present consultancy
  - Lilly
  - Johnson and Johnson
  - Roche
  - NuMedii
  - Genstruct
  - Tercica
  - Ecoeos
  - Helix
  - Ansh Labs
  - uBiome
  - Prevendia
  - Samsung
  - Assay Depot
  - Regeneron
  - Verinata
  - Pathway Diagnostics
  - Geisinger Health
  - Covance
  - Wilson Sonsini Goodrich & Rosati
  - Orrick
  - 10X Genomics
  - GNS Healthcare
  - Gerson Lehman Group
  - Coatue Management
- Corporate Relationships
  - Northrop Grumman
  - Genentech
  - Optum
  - Aptalis
  - Allergan
  - Astellas
  - Thomson Reuters
  - Intel
  - SAP
  - SV Angel
  - Progenity
  - Illumina
- Speakers' bureau
  - None
- Companies started by students
  - Carmenta
  - Serendipity
  - Stimulomics
  - NunaHealth
  - Praedicat
  - MyTime
  - Flipora
  - Tumbl.in
  - Polyglot
  - Iota Health
  - Ongevity Health


AFFYMETRIX
GeneChip®
Human Genome
U133 Plus 2.0
P/N: 520019
Lot #: SAMPLES
Exp. Date: 00/00/00
For Research Use Only
@affymetrix

DNA microarrays allow researchers to analyse the expression of a huge number of genes simultaneously.

GENOMICS

Gene data to hit milestone

With close to one million gen… researchers can identify dise… …cly accessible repositories, …ter a laboratory.

BY MONYA BAKER

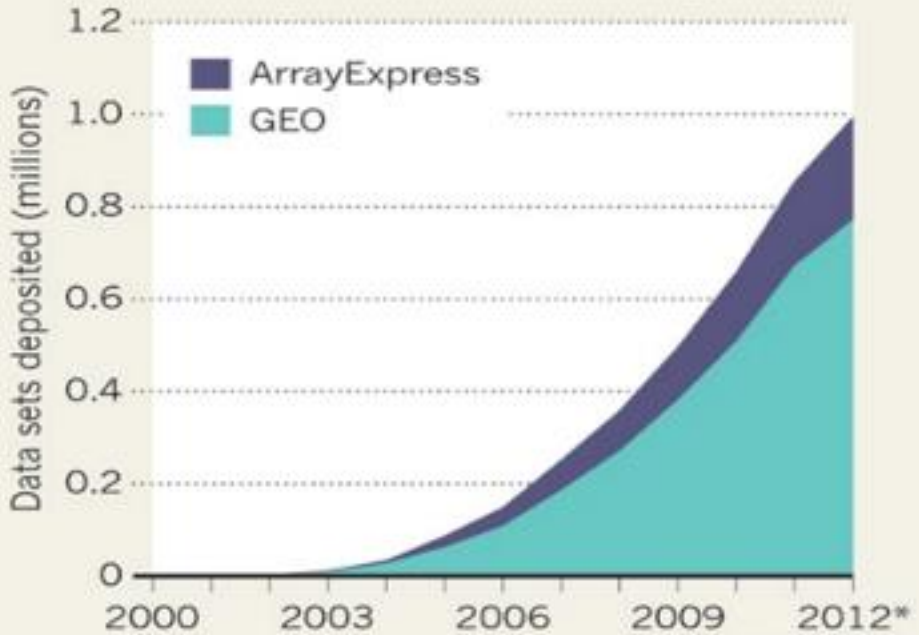
Purvesh Khatri sits in front of an oversize… computer screen, trawling for treasure i… a sea of genetic data. Entering the searc… term 'breast cancer' into a public repositor… called the Gene Expression Omnibus (GEO… the postdoctoral researcher retrieves a list … 1,170 experiments, representing nearly 33,00… samples and a hoard of gene-expression dat… that could reveal previously unseen patterns…

That is exactly the kind of search that le… Khatri's boss, Atul Butte, a bioinformatician … the Stanford School of Medicine in Californi… to identify a new drug target for diabetes. Afte… downloading data from 130 gene-expressio… studies in mice, rats and humans, Butte looke… for genes that were expressed at higher levels i…

…pository at the European Bioinformatics …titute (EBI) in Hinxton, UK. Some time in … next few weeks, the number of deposited …a sets will top one million (see 'Data dump'). …The result is an unprecedented resource that …mises to drive down costs and speed up pro- …ss in understanding disease. Gene-sequence …a are already shared extensively, but expres- …n data are more complex and can reveal …ich genes are the most active in, say, liver …sus brain cells, or in diseased versus healthy …ue. And because studies often look at many

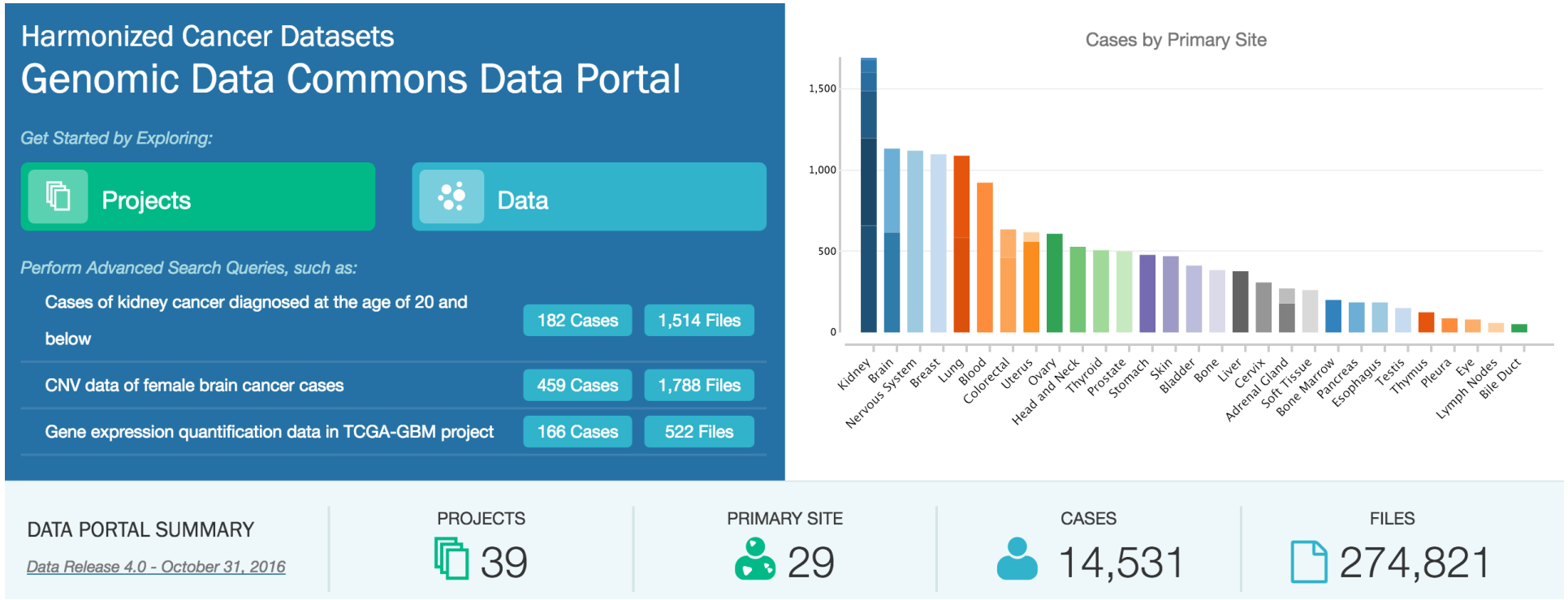
DATA DUMP

The number of gene-expression data sets in publicly available databases has climbed to nearly one million over the past decade.



bit.ly/genedata

Cancer researchers share data



The Cancer Genome Atlas

- 14 thousand cases
- 39 types of cancers
- 13 types of data: molecular, clinical, sequencing

NCBI

dbGaP
About dbGaP
Browse dbGaP
Controlled Access
Email Alerts
dbGaP Tutorial
Security Procedures
FTP Download
Publications
Contact dbGaP

Other Services

MeSH Browser
Clinical Queries

Browse dbGaP

TUTORIAL | ABO

db GaP GENOTYPE and PHENOTYPE

By Studies | By Diseases | Advanced Search

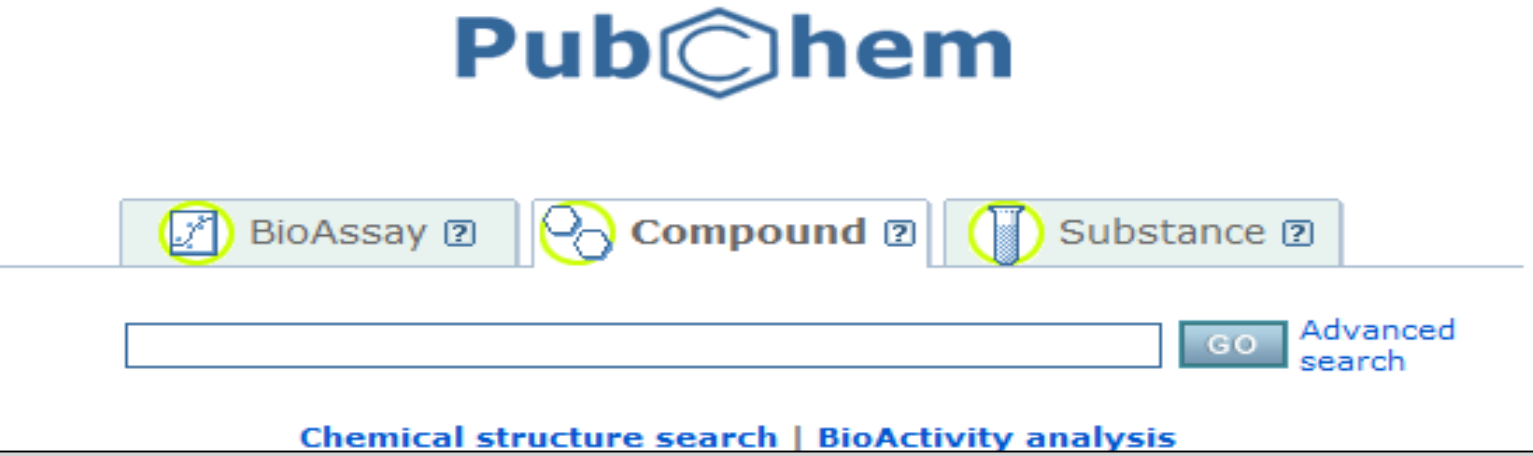
Study	Embargo Release	Details	Participants	Type of Study	Project
CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)	Feb 13, 2009	V D A	1991	Case-control	CIDR
Framingham SHARe	Version 1: Oct 19, 2008 Version 2: Feb 01, 2009 Version 3: Jul 08, 2009	V D A	14277	Longitudinal	SHARe
GAIN: Collaborative Association Study of Psoriasis	Aug 13, 2008	V D A	2875	Case-control	GAIN
GAIN: Genotyping the 270 HapMap samples for GAIN by Broad		V D A	-	Parent-offspring trios	
GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen		V D A	-	Parent-offspring trios	
GAIN: International Multi-Center ADHD Genetics Project	Mar 26, 2008	V D A	2835	Parent-offspring trios	GAIN
GAIN: Linking Genome-Wide Association Study of Schizophrenia	Version 1: Nov 07, 2008 Version 2: Dec 03, 2008	V D A	5066	Case-control	GAIN
GAIN: Major Depression: Stage 1 Genomewide Association in Population-Based Samples	Jul 09, 2008	V D A	3741	Case-control	GAIN
GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes	Jul 09, 2008	V D A	1825	Case-control	GAIN
GAIN: Whole Genome Association Study of Bipolar Disorder	Version 1: Nov 25, 2008 Version 2: Dec 01, 2008	V D A	3261	Case-control	GAIN
GAW16 Framingham and Simulated Data	Oct 19, 2008	V D A	7130	Longitudinal, population-based	SHARe
Genome-wide Association Studies in the Hutterites		V D A	632	Population-based	University of Chicago
Genome-wide Association Study of Neuroblastoma		V D A	1032	Case-control	COG
Genome-wide Study in Amyotrophic Lateral Sclerosis and Controls: First Stage Analysis	Jun 26, 2008	V D A	544	Case-control	NINDS
Ischemic Stroke Genetics Study (ISGS)	Jun 26, 2008	V D A	485	Case-control	NINDS
Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration	Mar 03, 2008	V D A	1550	Case-control	MJFF
NEI Age-Related Eye Disease Study (AREDS)	Jun 11, 2007	V D A	600	Case-control	NEI
NINDS Parkinson's Disease	Oct 12, 2007	V D A	535	Case-control	NINDS
NINDS Parkinsonism Study	Oct 12, 2007	V D A	1283	Case-set	NINDS
NINDS Repository Cerebrovascular Disease/Stroke Study	Jun 26, 2008	V D A	870	Case-set	NINDS
NIND				Case-set	NINDS
NIND				Control-set	NINDS
POP				ulation samples, Control-set	NHGRI
SEARCH GWA Study of Statin-Induced Myopathy		V D A	175	Case-control	University of Oxford
Study of Irish Amyotrophic Lateral Sclerosis (SIALS)		V D A	432	Case-control	NINDS
The Finland-United States Investigation of NIDDM Genetics (FUSION) study		V D A	2335	Case-control	University of Michigan
Whole Genome Association Study of Systemic Lupus Erythematosus		V D A	4651	Case-control	

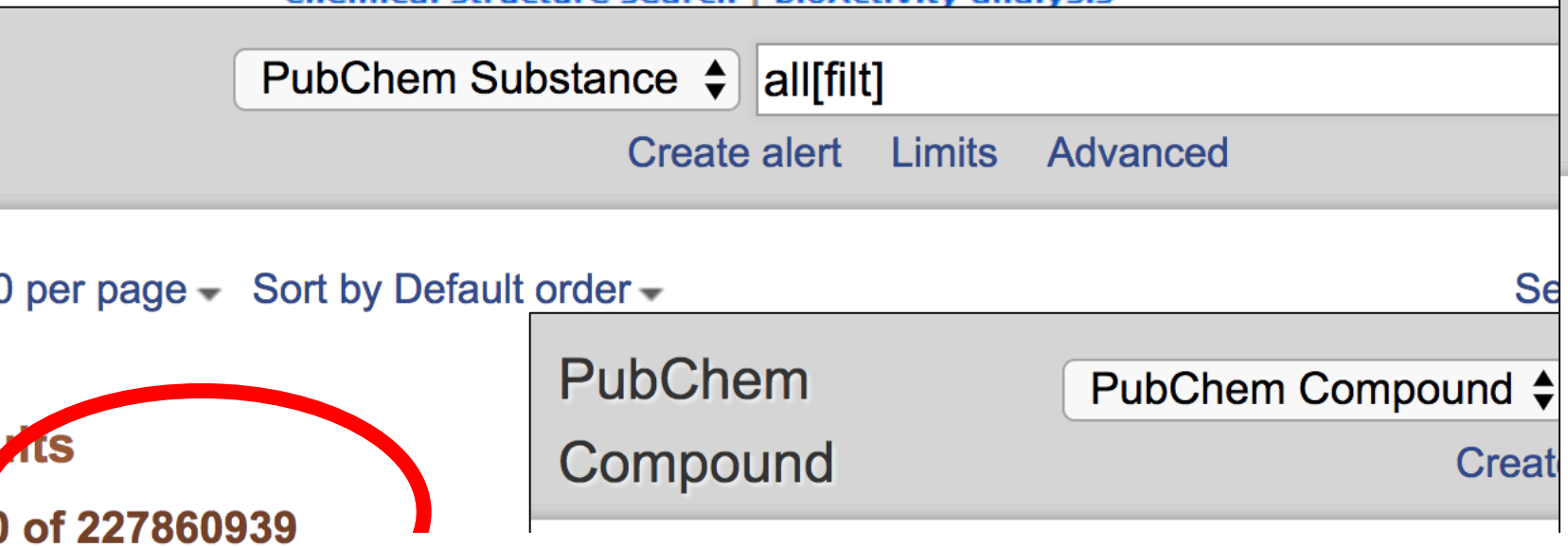
Genetics researchers share data

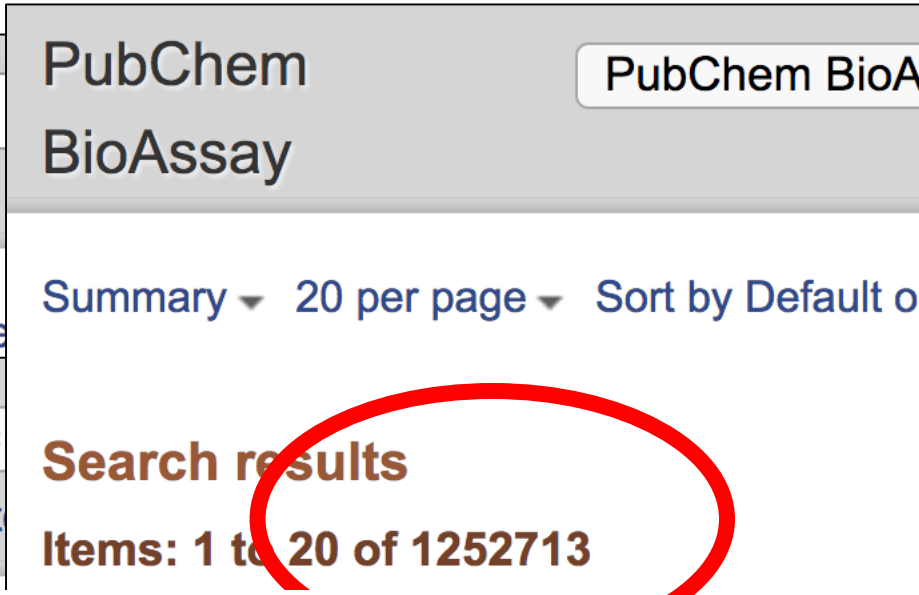
227 million substances x
1.3 million assays

More than a billion measurements within a grid of 300 trillion cells

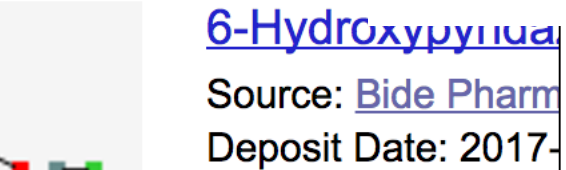
71 million meet Lipinski 5
1.2 million active substances

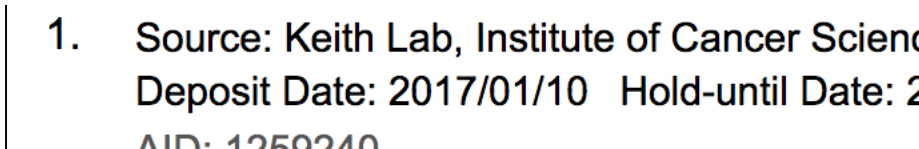






Chemical biologists share data

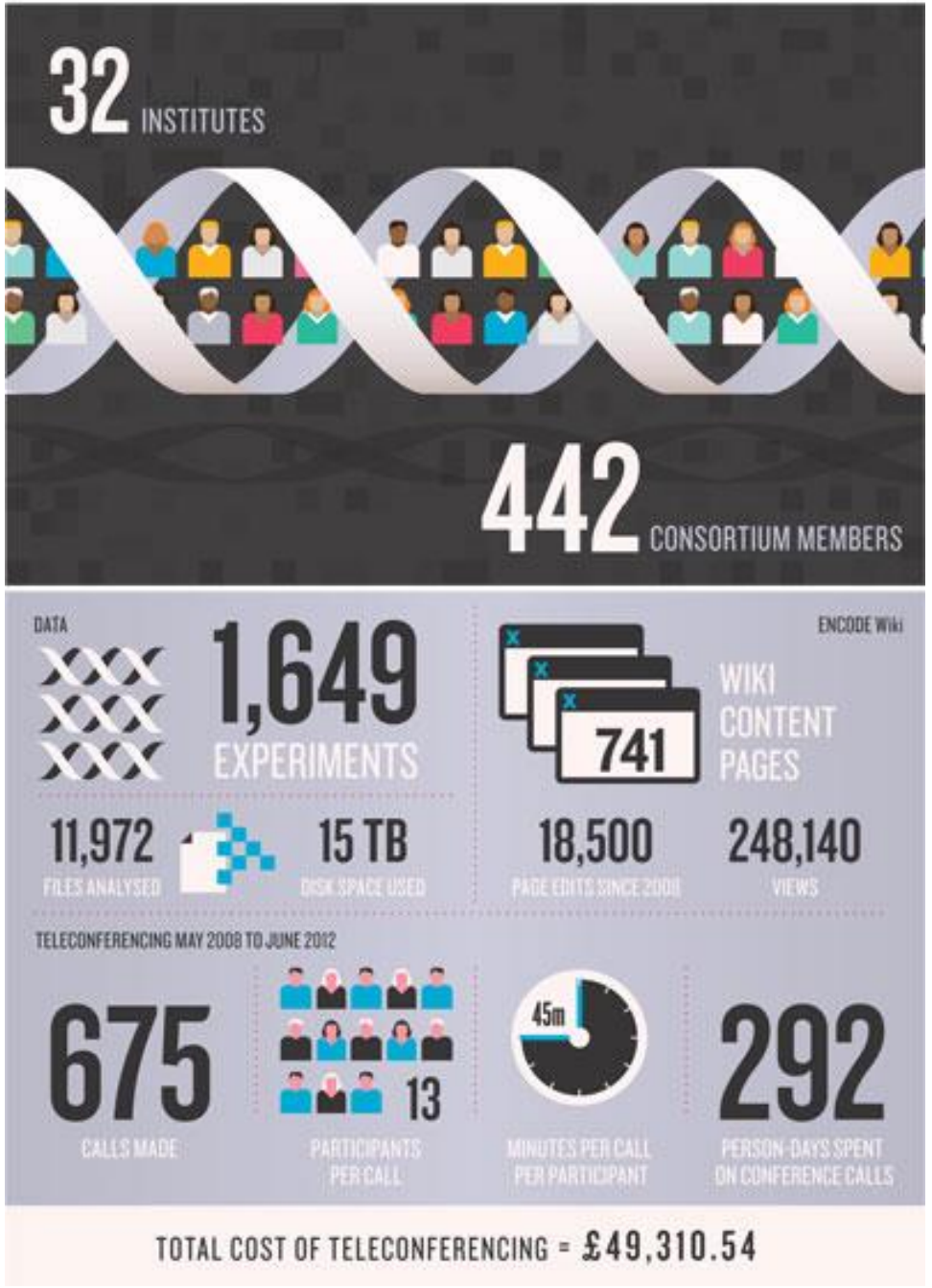




ENCODE

BY THE NUMBERS

The ENCODE project involved hundreds of people from around the world, and a lot of editing, disk space and phone calls.



Molecular biologists share data

Even immunologists and trialists can share data!

immport.org

Sanchita Bhattacharya
Zicheng Hu
Elizabeth Thomson
and many more

NORTHROP GRUMMAN

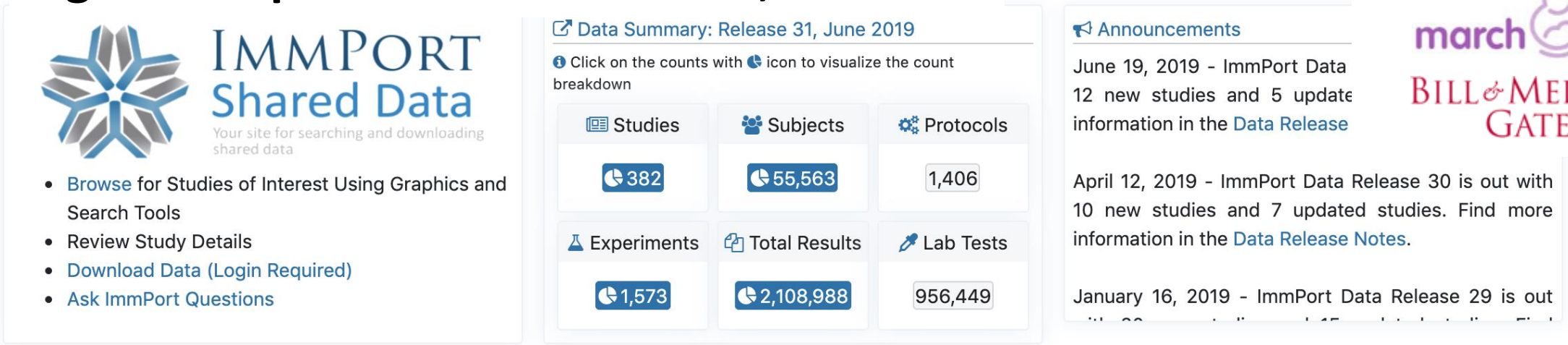
UCSF

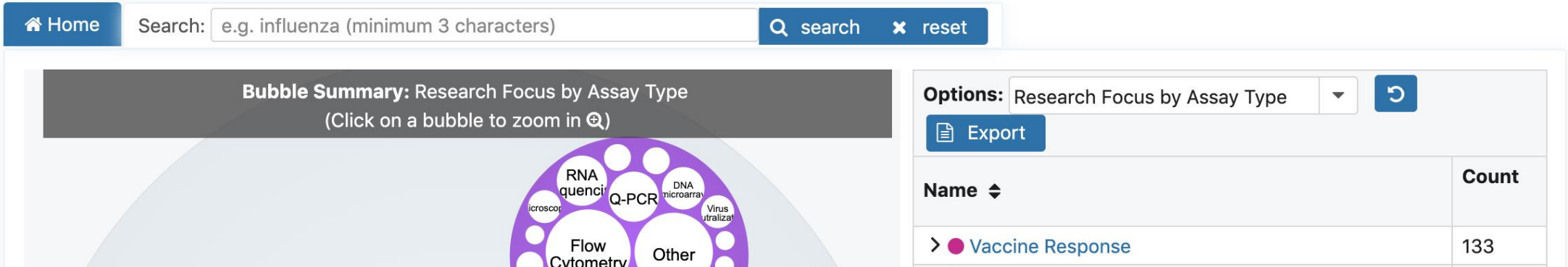
NIH National Institute of Allergy and Infectious Diseases

march of dimes

BILL & MELINDA GATES foundation

Download 380+ studies today
Drug repositioning, new patient subsets, digital comparative effectiveness, more!





ImmPort redistributes data from major NIAID-funded programs and more

Data from 380+ trials and studies already released, involving:

- Immune Tolerance Network (ITN)
- Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus (AMP)
- Human Immunology Project Consortium (HIPC)
- Atopic Dermatitis Research Network (ADRN)
- Clinical Trials in Organ Transplantation (CTOT) and in Children (CTOT-C)
- Population Genetics Analysis Program
- Protective Immunity for Special Populations
- HLA Region Genomics in Immune-mediated Diseases
- Modeling Immunity for Biodefense
- Reagent Development for Innate Immune Receptors
- Adjuvant Development Program
- Innate Immune Receptors and Adjuvant Discovery Program
- Maintenance of Macaque Specific Pathogen-Free Breeding Colonies
- Non-human Primate Transplantation Tolerance Cooperative Study Group
- Immunity in Neonates and Infants
- Consortium for Food Allergy Research
- Development of Sample Sparing Assays for Monitoring Immune Responses (U24)
- Asthma and Allergic Diseases Cooperative Research Centers
- HLA and KIR Region Genomics in Immune-Mediated Diseases
- Systems Approach to Immunity and Inflammation
- Immunobiology of Xenotransplantation
- Cooperative Study Group for Autoimmune Disease Prevention
- Informatics Methodology and Secondary Analyses for Immunology Data in ImmPort
- Centers for Medical Countermeasures against Radiation Consortium
- Inner City Asthma Consortium

Collaborations with NIAMS, NCI, Bill and Melinda Gates Foundation, and March of Dimes

Matching Records: 38 **Sorted by:** Release Date (Desc) then Announcement Number

Announcement Number	Related Announc.	Issuing Organization	Release Date	Opening Date (SF424 Only) ?	Expiration Date	Activity Code(s)	Title
RFA-AI-18-042	See Related	NIAID	10/05/2018	01/02/2019	02/02/2019	U01	Fc-Dependent Mechanisms of Antibody-Mediated Killing (U01 Clinical Trial Not Allowed)
RFA-AI-18-026	See Related	NIAID	06/19/2018	11/13/2018	12/14/2018	R01	Modeling and Simulation to Optimize HIV Prevention Research (MS OPR) (R01 Clinical Trial not allowed)
RFA-AI-18-023	See Related	NIAID	06/14/2018	09/04/2018	10/05/2018	R01	Immune Mechanisms at the Maternal-Fetal Interface (R01 Clinical Trial Optional)
RFA-AI-18-010	See Related	NIAID	04/05/2018	06/02/2018	07/03/2018	U01	Impact of Initial Influenza Exposure on Immunity in Infants (U01 Clinical Trial Not Allowed)
PAR-18-712	See Related	NIAID	03/15/2018	05/05/2018	05/08/2021	R01	Investigations on Primary Immunodeficiency Diseases/Inborn Errors of Immunity (R01 Clinical Trial Not Allowed)
RFA-AI-18-002	See Related	NIAID	03/08/2018	05/22/2018	06/23/2018	U19	Autoimmunity Centers of Excellence, Basic Research Program (U19 Clinical Trial Not Allowed)
RFA-AI-18-003	See Related	NIAID	03/08/2018	05/22/2018	06/23/2018	UM1	Autoimmunity Centers of Excellence, Clinical Research Program (UM1 Clinical Trial Required)
RFA-AI-17-040	See Related	NIAID	12/04/2017	02/28/2018	03/29/2018	U19	Cooperative Centers on Human Immunology (U19 Clinical Trial Optional)
RFA-AI-17-037	See Related	NIAID	10/25/2017	01/22/2018	02/23/2018	R01	Immunity in the Elderly (R01) Clinical Trial Optional
RFA-AI-17-034	See Related	NIAID	10/20/2017	01/21/2018	02/22/2018	U01	Maintaining Immunity After Immunization (U01 - Clinical Trial Optional)
RFA-AI-16-078	See Related	NIAID	11/28/2016	02/15/2017	03/16/2017	U01	Limited Competition: Clinical Trials in Organ Transplantation in Children (CTOT-C): Mechanistic Ancillary Studies (U01)
RFA-AI-16-027	See Related	NIAID	09/07/2016	02/03/2017	03/04/2017	U19	T Cell Reagent Resource for the Study of Allergic Diseases (U19)

Department of Health and Human Services

Part 1. Overview Information

Participating Organization(s)	National Institutes of Health (NIH)
Components of Participating Organizations	National Institute of Allergy and Infectious Diseases (NIAID)
Funding Opportunity Title	Informatics Methodology and Secondary Analyses for Immunology Data in ImmPort (UH2 Clinical Trial Not Allowed)
Activity Code	UH2, Exploratory/Developmental Cooperative Agreement Phase I
Announcement Type	Reissue of PAR-16-253
Related Notices	• **April 08, 2019** - Notice of Change in Expiration Date of PAR-19-229. See Notice NOT-AI-19-055.
Funding Opportunity Announcement (FOA) Number	PAR-19-229

dash | UCSF Library

Search

Explore Data | Help | Login

An easy-to-use data publication service

"hypotheses come and go, but data remain"

Santiago Ramón y Cajal, 1897

Search 599 datasets

Search

Why share your data?

https://datashare.ucsf.edu/stash

Why use our service?

Straightforward compliance. Submit your data to satisfy publisher and funder requirements for preservation and availability with a minimum of effort. Optionally link data to the article and funding organization.

Any field. Any format. Submit data in any file format from any field of research. Share all of the data from a project in one place.

Get discovered. Each landing page and dataset is optimized for search engines. Include any relevant geo-location information to take advantage of our built-in geospatial search.

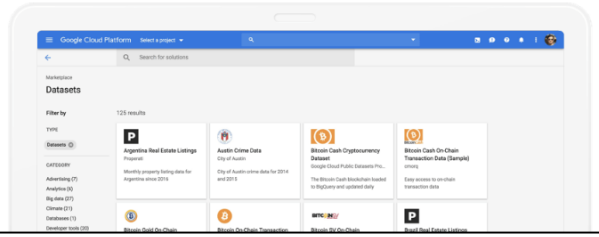
Credit for your work. Get an informative landing page to facilitate re-use of your data and a citable DOI so you can get credit. Add your ORCID identifier and we will automatically send the information to your ORCID profile.

Google Cloud Public Datasets

Uncover new insights with high-demand public datasets.

View one-pager | Explore public datasets

Explore public datasets in Google Cloud



Google Cloud Public Datasets facilitate access to high-demand public datasets, making it easy for you to access and uncover new insights in the cloud. By analyzing thes datasets hosted in BigQuery and Cloud Storage, you can seamlessly experience the full value of Google Cloud wi ease.

Google Dataset Search Beta

Search for Datasets

Try boston education data or weather site:noaa.gov

Learn more about including your datasets in Dataset Search.

Campus News • June 28, 2019

University of California Announces Collaboration with Janssen to Expand Data Science Research in Healthcare

New Fellowship Program Facilitated by Johnson & Johnson Innovation to Recruit Data Scientists for High-Impact, Data-Driven Healthcare Research

By Laura Kurtzman

The University of California, Berkeley (UC Berkeley), and the University of California, San Francisco (UCSF), today announced a collaboration with Janssen Research & Development, LLC, part of the Janssen Pharmaceutical Companies of Johnson & Johnson (Janssen), to launch a new data science fellowship program that will explore innovative data-driven approaches to improve human health and train the next generation of leaders in the healthcare data sciences. This program will be the first of its kind in the San Francisco Bay Area, which already serves as a worldwide hub for the tech and

10 reasons why to archive and share study data openly

- Reproducibility
- Transparency
- Support public policy
- Return data to the community
- Visibility into failed trials
- Speed results reporting
- Enable learning
- Enable new ventures
- **New science**
- Trust and Believability

Reproducibility

Repeatability of published microarray gene expression analyses

John P A Ioannidis[1–3], David B A
Mario Falchi[8,9], Cesare Furlanell
Michael Nitzberg[5], Grier P Page[4]

Given the complexity of microarray-l
studies, guidelines er
data availability. Sev
and several public da
publicly available, ar
whe
scie
in 1
publ
figu
tean
and
be r

nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current issue | Archive | Audio & Video | For A

Archive | Volume 500 | Issue 7460 | News | Article

NATURE | NEWS

NIH mulls rules for validating key results

US biomedical agency could enlist independent labs for verification.

Meredith Wadman

31 July 2013

PDF | Rights & Permissions

In biomedical science, at least one thing is apparently reproducible: a steady stream of studies that show the irreproducibility of many important experiments.

In a 2011 internal survey, pharmaceutical firm Bayer HealthCare of Leverkusen, Germany, was unable to validate the relevant preclinical research for almost two-thirds of 67 in-house projects. Then, in 2012, scientists at Amgen, a drug company based in Thousand Oaks, California, reported

Return data to the community

TheScientist

EXPLORING LIFE, INSPIRING INNOVATION

News | Magazine | Multimedia | Subjects | Surveys

The Scientist » Magazine » Features

Do-It-Yourself Medicine

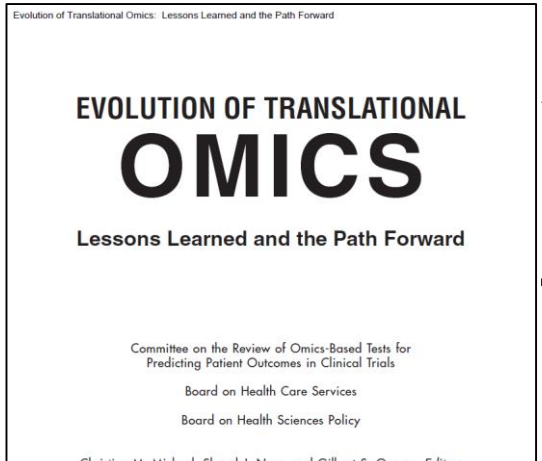
**Patients are sidestepping clinical research and using themselves as guinea pi
treatments for fatal diseases. Will they hurt themselves, or science?**

By Jef Akst | March 1, 2013

9 Comments | Like 270 | Pin it 3 | +1 17 | Link this

On August 10,
Butler dilute
percent sodium ch
slightly salted wat
injected 1 liter of t

Transparency



Evolution of Translational Omics: Lessons Learned and the Path Forward

EVOLUTION OF TRANSLATIONAL

OMICS

Lessons Learned and the Path Forward

Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials

Board on Health Care Services

Board on Health Sciences Policy

code — in protecting patient confidentiality, for example. In such cases, authors should justify the omission and assure independent reproducibility by alternative means.

The quality of scientific output will benefit from setting these standards. As a community, we owe it to patients and to the public to do what we can to ensure the validity of the research we publish.

Keith Baggerly *on behalf of 7 co-authors*, The University of Texas MD Anderson Cancer*

Support public policy

BMJ

BMJ 2013;346:f2157 doi: 10.1136/bmj.f2157 (Published 4 April 2013) Page 1 of 2

NEWS

Roche offers researchers access to all Tamiflu trials

Deborah Cohen

BMJ

More than three years after the Cochrane Collaboration first asked Roche for the full clinical study reports for its influenza drug oseltamivir (Tamiflu), the Swiss company has offered the

follows that regulators were in the same situation as we were: lacking data to come to firm conclusions."

Visibility into failed trials

BMJ

RESEARCH

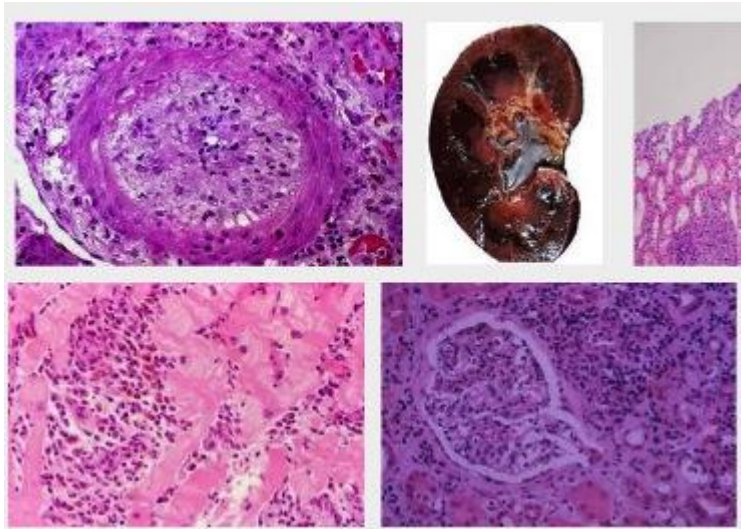
BMJ 2013;347:f6104 doi: 10.1136/bmj.f6104 (Published 29 October 2013)

Non-publication of large randomized clinical trials: cross sectional analysis

OPEN ACCESS

Christopher W Jones *attending physician*[1], Lara Handler *school of medicine liaison librarian*[2], Karen E Crowell *clinical information specialist*[2], Lukas G Keil *research assistant*[3], Mark A Weaver *assistant professor*[4], Timothy F Platts-Mills *assistant professor*[3]

Enable learning



Speed results reporting

Scientists voice fears over ethics of drug trials remaining unpublished

Almost a third of large clinical trials in the US still not published five years after being finished, scientists write in BMJ

Sarah Boseley
The Guardian, Tuesday 29 October 2013 19.30 EDT

Enable new ventures

NEXTBIO
now part of Illumina

Pathwork Diagnostics

All the world's genomic data

NHS National Institute for Health Research

Welcome to the UK Clinical Trials Gateway

New Science

nature genetics

ANALYSIS

An APOBEC cytidine deaminase mutagenesis pattern widespread in human cancers

Steven A Roberts[1], Michael S Lawrence[2], Leszek J Klimczak[3], Sara A Grimm[3], David Fargo[3], Petar Stojano
Adam Kiezun[2], Gregory V Kryukov[2,4], Scott L Carter[2], Gordon Saksena[2], Shawn Harris[5],
Ruchir R Shah[5], Michael A Resnick[1], Gad Getz[2,6–8] & Dmitry A Gordenin[1,8]

Recent studies indicate that a subclass of APOBEC cytidine deaminases, which convert cytosine to uracil during RNA editing and retrovirus or retrotransposon restriction, may induce mutation clusters in human tumors. We show here that throughout cancer genomes APOBEC-mediated mutagenesis is pervasive and correlates with APOBEC mRNA levels. Mutation clusters in whole-genome and exome c to the stringent criteria indicative of a pattern. Applying these criteria to 954 exomes from 14 cancer types, mostly f Atlas (TCGA), showed a significant pre mutation pattern in bladder, cervical, and lung cancers, reaching 68% of all samples. Within breast cancer, the HE was clearly enriched for tumors with t pattern, suggesting that this type of mu linked with cancer development. The A pattern also extended to cancer-associ ubiquitous APOBEC-mediated mutage

Genome instability triggers the develo cancers[1,2]. Radiation and chemical dama as culprits in theories of carcinogeni normal enzymatic activities can also be and mutation. Cytidine deaminases, wh to uracil, likely contribute to DNA da cytidine deaminase (AID), a key enzyme only initiates the hypermutation and cla immunoglobulin genes but also can mu a limited number of secondary targets,

implicated in carcinogenesis[5]. In addition to AID, the human encodes several homologous APOBEC (apolipoprotein B editing enzyme, catalytic polypeptide-like) cytidine deamin function in innate immunity as well as in RNA editing[6]. human cell culture studies showed that a subclass of APOB mutational specificity for TC motifs (with the mutated bas

[1]Laboratory of Molecular Genetics, National Insti Sciences, Durham, North Carolina, USA. [2]The Br Harvard, Cambridge, Massachusetts, USA. [3]Integ Institute of Environmental Health Sciences, Durh [4]Harvard Medical School, Boston, Massachusetts Durham, North Carolina, USA. [6]Massachusetts G Boston, Massachusetts, USA. [7]Department of Pa Hospital, Boston, Massachusetts, USA. [8]These au work. Correspondence should be addressed to D.A G.G. (gadgetz@broadinstitute.org).

Received 28 January; accepted 20 June; publish doi:10.1038/ng.2702

Methods

Abstract • Introduction • Results • Discussion • Methods • References • Acknowledgments • Author information • Supplementary information

Genome and exome data sets.

Genome and exome data sets were obtained from publications[20, 21] or from the TCGA data portal (see URLs; Controlled Data Access HTTP Directory). The catalog of base substitutions identified by whole-genome sequencing in 21 breast cancers was downloaded from the website provided in ref. 12 (seeURLs).

Hyperlinks to TCGA data sets and references to published mutation lists are provided in Supplementary Table 3.

Cluster analysis.

Clusters and colocalization between clusters and rearrangement breakpoints in whole-genome data sets were identified as described in ref. 13. Analysis of mutation clustering in exomes was conducted similarly to that in whole-genome data sets. Briefly, we first filtered out mutations identical to variants in dbSNP. These SNPs generally constituted a small percentage (0.9–12.1%) of all exome mutations for a given cancer type. However, LUSC, KIRC, PRAD and STAD samples contained somewhat higher numbers of mutations identical to variants in dbSNP (19.5–25.1%). Notably, each prefiltered mutation was included in the total number of mutations in the genome, which would thereby only increase the *P* values of clusters. We next identified groups of closely

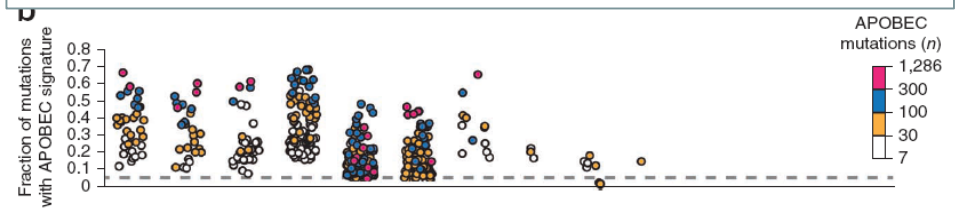
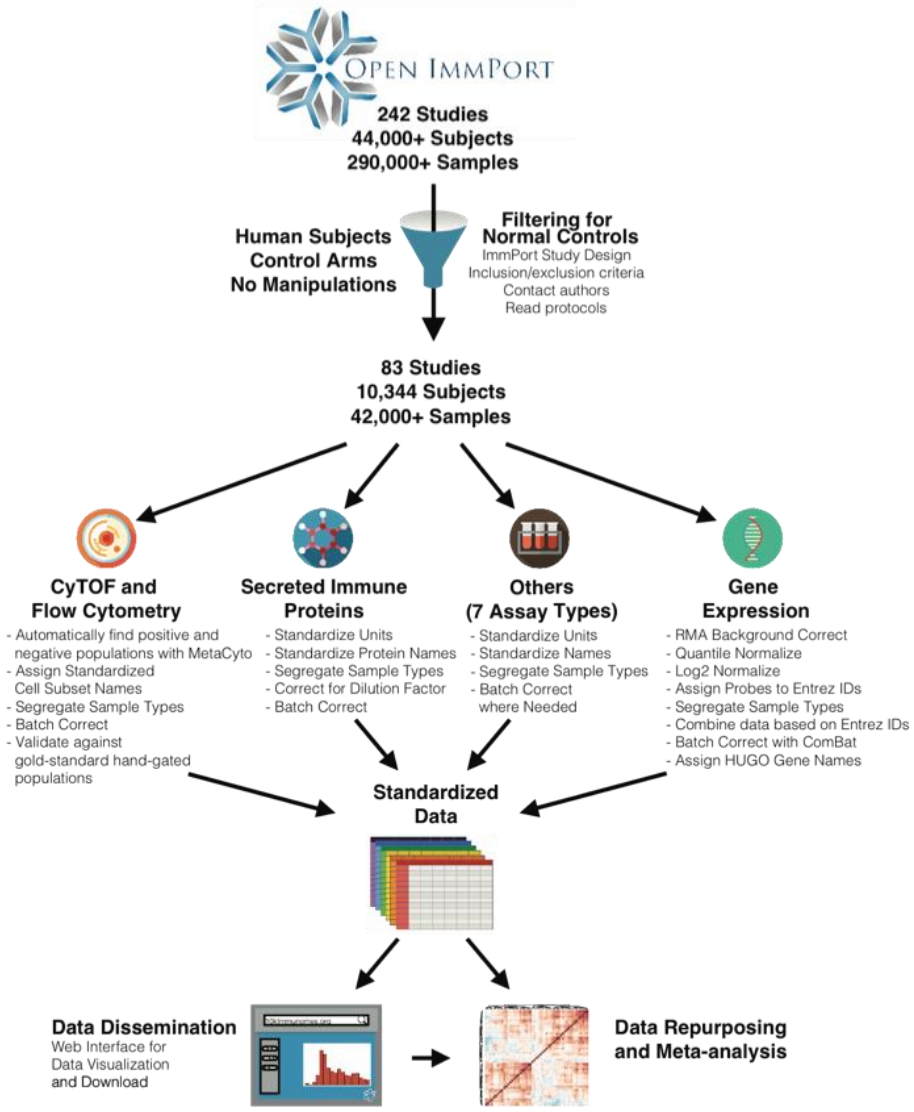
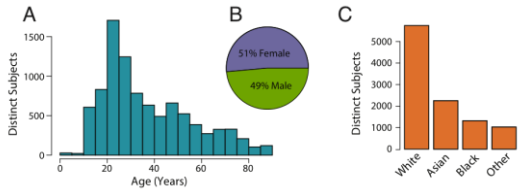


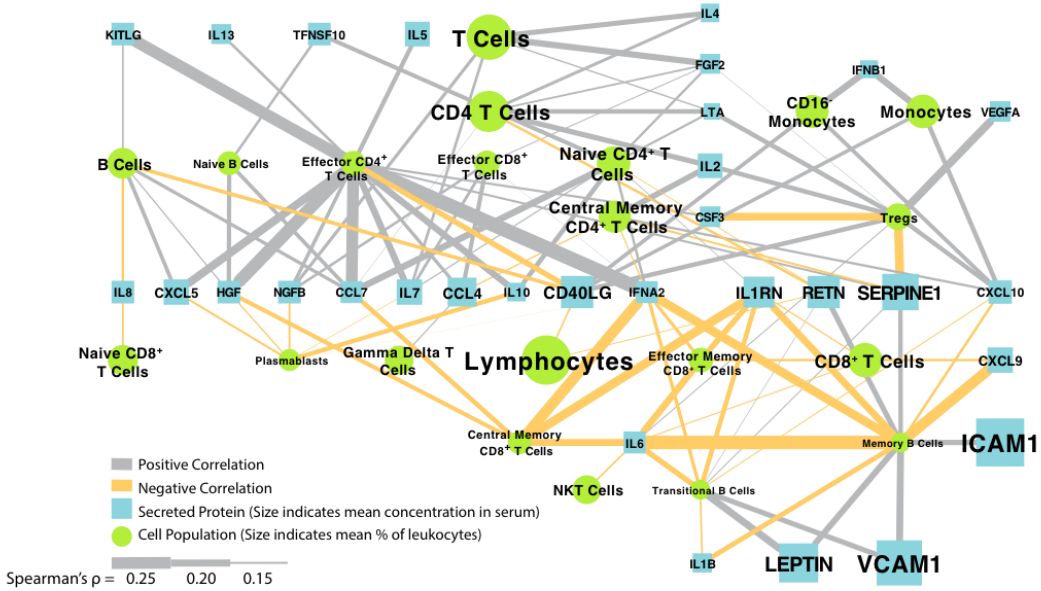
Figure 2 Presence of an APOBEC mutation pattern in exome data sets from different cancer types. (**a**,**b**) Fold enrichment (**a**) and mutation load (**b**) of the APOBEC mutation pattern were determined in each of 2,680 whole exome–sequenced tumors representing 14 cancer types. Samples were categorized by the statistical significance of the APOBEC mutation pattern and the magnitude of enrichment. The significance of the APOBEC mutation pattern was calculated by one-sided Fisher's exact test comparing the ratio of the number of C-to-T or C-to-G substitutions and complementary G-to-A or G-to-C substitutions that occur in and out of the APOBEC target motif (TCW or WGA) to an analogous ratio for all cytosines or guanines that reside inside and outside of the TCW or WGA motif within a sample fraction of the genome (Benjamini-Hochberg–corrected q value < 0.05). The number of tumor samples in each category is presented in each pie chart in **a**. Samples with q value > 0.05 are represented in black. These samples are excluded from the scatter graphs in **a**,**b**. Color scales indicate the magnitude of enrichment in **a** and the number of APOBEC signature mutations in **b** for samples with $q < 0.05$. Dashed lines indicate effects expected with random mutagenesis. Cancer types are abbreviated as in TCGA: cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), bladder urothelial carcinoma (BLCA), head and neck squamous cell carcinoma (HNSC), breast invasive carcinoma (BRCA), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), uterine corpus endometrioid carcinoma (UCEC), ovarian serous cystadenocarcinoma (OV), stomach adenocarcinoma (STAD), rectum adenocarcinoma (READ), colon adenocarcinoma (COAD), prostate adenocarcinoma (PRAD), kidney renal clear-cell carcinoma (KIRC) and acute myeloid leukemia (LAML).

The 10,000 Immunome Project: From the control groups of 242 manually curated experiments





Data available in the 10,000 Immunomes Project	
Total Samples	**42117**
Total Distinct Subjects	**10344**
MEASUREMENT	**SUBJECTS**
Secreted Proteins	4835
ELISA	4035
Multiplex ELISA	1286
Virus Titer	3609
Virus Neutralization Titer	2265
HAI Titer	1344
Clinical Lab Tests	2639
Complete Blood Count	1684
Comprehensive Metabolic Panel	664
Fasting Lipid Profile	664
Questionnaire	1422
Cytometry	1415
Flow Cytometry (PBMC)	907
CyTOF (PBMC)	583
Flow Cytometry (Whole Blood)	164
HLA Type	1093
Gene Expression Array	476
Whole Blood	311
PBMC	165



Kelly Zalocusky
Sanchita Bhattacharya
@ImmPortDB

Cell Reports bit.ly/10kimmunome
http://10kimmunomes.org/

Share successful, failed, and so-so data

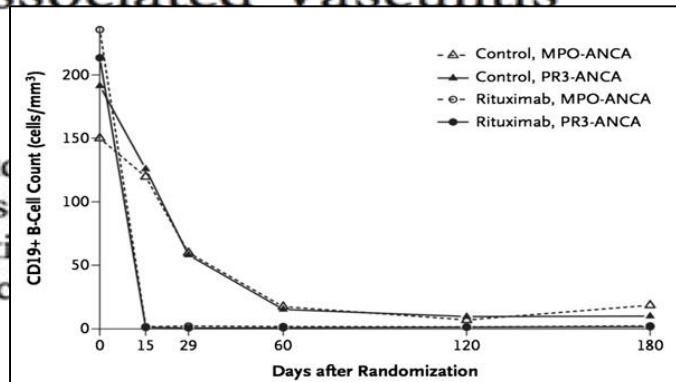
- Rituximab in ANCA-Associated Vasculitis (RAVE) trial of new approach to the induction of remission
- But even though rituximab was found to be non-inferior than cyclophosphamide, which drug is the right one to use?

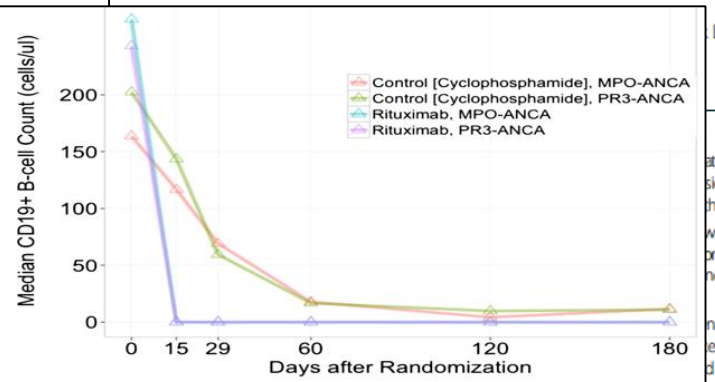
ORIGINAL ARTICLE

Rituximab versus Cyclophosphamide for ANCA-Associated Vasculitis

CONCLUSIONS

Rituximab therapy was not inferior t
tion of remission in severe ANCA-as
lapsing disease. (Funded by the Nati
eases, Genentech, and Biogen; Clinic





Nasrallah et al. Arthritis Research & Therapy (2015) 17:262
DOI 10.1186/s13075-015-0778-z

arthritis research&therapy

RESEARCH ARTICLE Open Access

Reanalysis of the Rituximab in ANCA-Associated Vasculitis trial identifies granulocyte subsets as a novel early marker of successful treatment

Dunn[5], Elizabeth Thomson[5], Jeffrey Wiser[5]

atients with anti-neutrophil cytoplasmic antibody
sion at 6 months following rituximab or
he Rituximab in ANCA-Associated Vasculitis (RAVE) trial.
wnloaded from the Immunology Database and
ories. Flow cytometry data were analyzed using
nd granulocyte populations were measured in patients

nt outcome with either treatment. We defined a
entage of hypergranular and hypogranular
d remission had a significantly higher GI at baseline

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Once-Daily Plazomicin for Complicated Urinary Tract Infections

Florian M.E. Wagenlehner, M.D., Daniel J. Cloutier, Pharm.D., Allison S. Komirenko, Pharm.D., Deborah S. Cebrik, M.S., M.P.H., Kevin M. Krause, M.B.A., Tiffany R. Keepers, Ph.D., Lynn E. Connolly, M.D., Ph.D., Loren G. Miller, M.D., M.P.H., Ian Friedland, M.D., and Jamie P. Dwyer, M.D., for the EPIC Study Group*

ABSTRACT

BACKGROUND

The increasing multidrug resistance among gram-negative uropathogens necessitates new treatments for serious infections. Plazomicin is an aminoglycoside with bactericidal activity against multidrug-resistant (including carbapenem-resistant) Enterobacteriaceae.

METHODS

We randomly assigned 609 patients with complicated urinary tract infections (UTIs), including acute pyelonephritis, in a 1:1 ratio to receive intravenous plazomicin (15 mg per kilogram of body weight once daily) or meropenem (1 g every 8 hours), with optional oral step-down therapy after at least 4 days of intravenous therapy, for a total of 7 to 10 days of therapy. The primary objective was to show the noninferiority of plazomicin to meropenem in the treatment of complicated UTIs, including acute pyelonephritis, with a noninferiority margin of 15 percentage points. The primary end points were composite cure (clinical cure and microbiologic eradication) at day 5 and at the test-of-cure visit (15 to 19 days after initiation of therapy) in the microbiologic modified intention-to-treat population.

RESULTS

Plazomicin was noninferior to meropenem with respect to the primary efficacy end points. At day 5, composite cure was observed in 88.0% of the patients (168 of 191 patients) in the plazomicin group and in 91.4% (180 of 197 patients) in the meropenem group (difference, –3.4 percentage points; 95% confidence interval [CI], –10.0 to 3.1). At the test-of-cure visit, composite cure was observed in 81.7% (156 of 191 patients) and 70.1% (138 of 197 patients), respectively (difference, 11.6 percentage points; 95% CI, 2.7 to 20.3). At the test-of-cure visit, a higher percentage of patients in the plazomicin group than in the meropenem group were found to have microbiologic eradication, including eradication of Enterobacteriaceae that were not susceptible to aminoglycosides (78.8% vs. 68.6%) and Enterobacteriaceae that produce extended-spectrum β-lactamases (82.4% vs. 75.0%). At late follow-up (24 to 32 days after initiation of therapy), fewer patients in the plazomicin group than in the meropenem group had microbiologic recurrence (3.7% vs. 8.1%) or clinical relapse (1.6% vs. 7.1%). Increases in serum creatinine levels of 0.5 mg or more per deciliter (≥40 μmol per liter) above baseline occurred in 7.0% of patients in the plazomicin group and in 4.0% in the meropenem group.

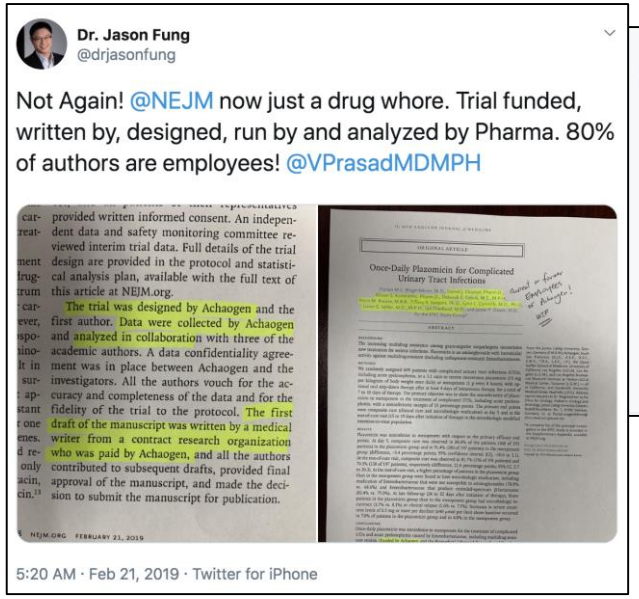
CONCLUSIONS

Once-daily plazomicin was noninferior to meropenem for the treatment of complicated UTIs and acute pyelonephritis caused by Enterobacteriaceae, including multidrug-resistant strains. (Funded by Achaogen and the Biomedical Advanced Research and Development Authority; EPIC ClinicalTrials.gov number, NCT02486627.)

From the Justus Liebig University, Giessen, Germany (F.M.E.W.); Achaogen, South San Francisco (D.J.C., A.S.K., D.S.C., K.M.K., T.R.K., L.E.C., I.F.), the David Geffen School of Medicine, University of California Los Angeles (UCLA), Los Angeles (L.G.M.), and Los Angeles Biomedical Research Institute at Harbor–UCLA Medical Center, Torrance (L.G.M.) — all in California; and Vanderbilt University Medical Center, Nashville (J.P.D.). Address reprint requests to Dr. Wagenlehner at the Clinic for Urology, Pediatric Urology and Andrology, Justus Liebig University Giessen, Rudolf-Buchheim Str. 7, 35392 Giessen, Germany, or at florian.wagenlehner@chiru.med.uni-giessen.de.

*A complete list of the principal investigators in the EPIC Study is provided in the Supplementary Appendix, available at NEJM.org.

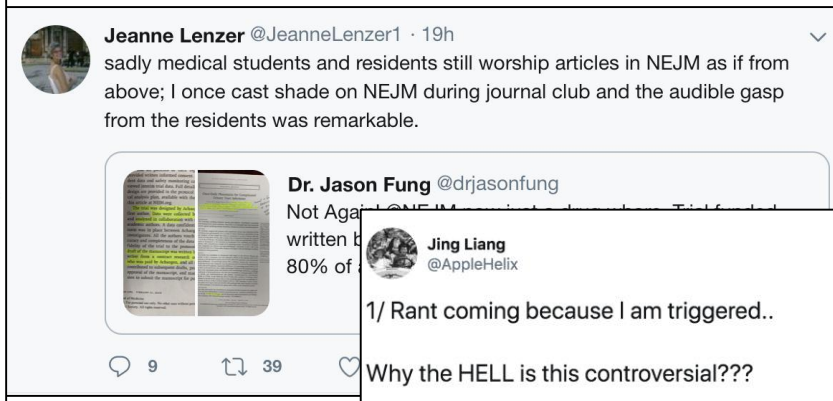
N Engl J Med 2019;380:729-40.
DOI: 10.1056/NEJMoa1801467



Dr. Jason Fung
@drjasonfung

Not Again! @NEJM now just a drug whore. Trial funded, written by, designed, run by and analyzed by Pharma. 80% of authors are employees! @VPrasadMDMPH

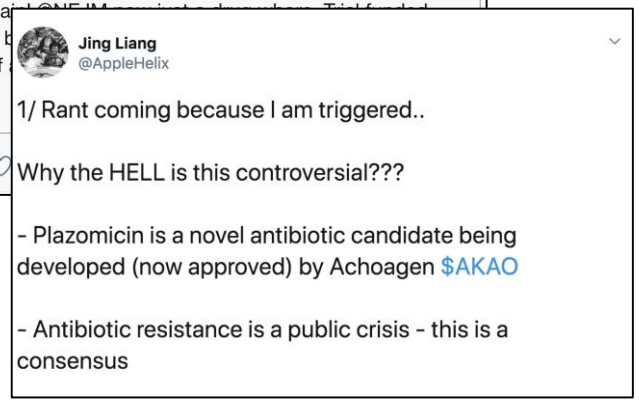
5:20 AM · Feb 21, 2019 · Twitter for iPhone



Jeanne Lenzer @JeanneLenzer1 · 19h
sadly medical students and residents still worship articles in NEJM as if from above; I once cast shade on NEJM during journal club and the audible gasp from the residents was remarkable.

Dr. Jason Fung @drjasonfung
Not Agai
written b
80% of

9 39



Jing Liang
@AppleHelix

1/ Rant coming because I am triggered..

Why the HELL is this controversial???

- Plazomicin is a novel antibiotic candidate being developed (now approved) by Achoagen $AKAO

- Antibiotic resistance is a public crisis - this is a consensus

Journal reputation is at a critical moment

- How are journals going to respond to professional skeptics?
- How does a journal respond to another journal's editor criticizing approval?
- How will journals respond to health systems (like University of California) who will now look at drug efficacy in real world clinical data? Will the data match?
- How will journals respond to payers incentivized to challenge expensive drug approvals, who will now want to see the raw data? $Billions riding on these papers.
- How will journals address the family ready to sell their house to pay for a drug for their family member?
- How will journals counter when government officials label them "fake news"?

"Trust us, it works, we've looked at the data"?! Really?

Preeclampsia: large cause of maternal and fetal death

- Incidence
  - 5-8% of all pregnancies in the U.S. and worldwide
  - 4.1 million births in the U.S. in 2009
  - Up to 300K cases of preeclampsia annually in the U.S.
- Mortality
  - Responsible for 18% of all maternal deaths in the U.S.
  - Maternal death in 56 out of every 100,000 live births in US
  - Neonatal death in 71 out of every 100,000 live births in US
- Cost
  - $20 billion in direct costs in the U.S annually
  - Average hospital stay of 3.5 days

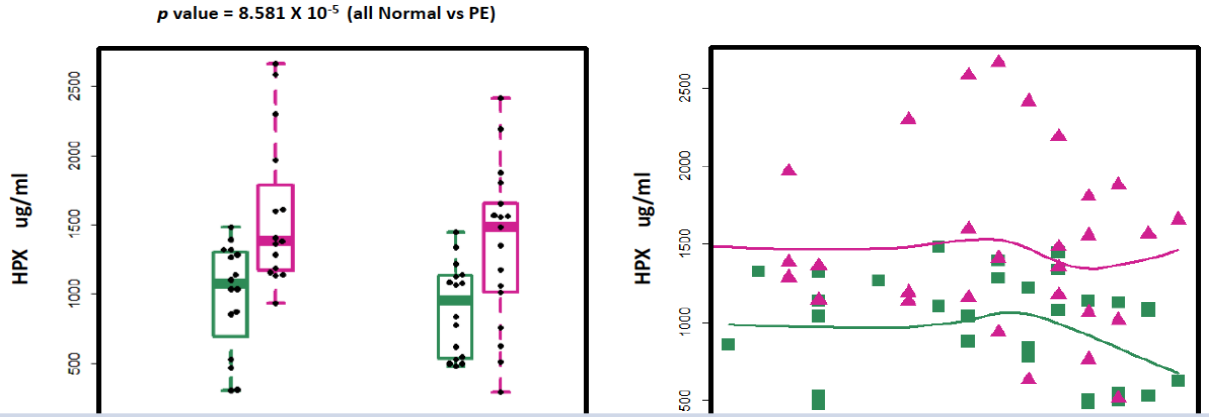
Linda Liu
Bruce Ling
Matt Cooper

Page 1 2 3 4 5 6 .. 11 Showing 1 - 25 of 266 experiments

Accession	Title	Type	Organism	Assays	Released
E-GEOD-32472	Oxygen induced complication of prematurity: from experimental data to prevention strategy	transcription profiling by array	Homo sapiens	299	01/11/2011
E-GEOD-27976	Calvarial osteoblast transcriptome analysis identifies genetic targets and extracellular matrix-mediated focal adhesion as potential biomarkers for single-suture craniosynostosis	transcription profiling by array	Homo sapiens	249	04/03/2012
E-GEOD-46510	New whole blood gene expression profile predictve of preterm birth	transcription profiling by array	Homo sapiens	154	15/05/2014
E-GEOD-37210	The application of nonsense-mediated mRNA decay inhibition to the identification of breast cancer susceptibility genes	transcription profiling by array	Homo sapiens	143	11/04/2012
E-TABM-682	Transcription profiling of human decidua basalis to identify pre-eclampsia susceptibility genes	transcription profiling by array	Homo sapiens	104	07/04/2009
E-GEOD-35574	Differentially expressed microRNAs revealed by molecular signatures of Preeclampsia and IUGR in human placenta	transcription profiling by array	Homo sapiens	94	07/02/2012
E-GEOD-41336	Cultured Cyto and Syncytio-trophoblast samples exposed to varying degrees of hypoxia (methylation)	methylation profiling by array	Homo sapiens	90	18/01/2013
E-GEOD-5999	Transcription profiling of human 27 non-	transcription	Homo sapiens	72	07/11/2008

march of dimes
prematurity research center
at STANFORD University School of Medicine

SPARK
AT STANFORD



Need a diagnostic for preeclampsia

Public big data available

March of Dimes Center for Prematurity Research

Data analyzed, diagnostic designed

SPARK grant ($50k)

Life Science Angels, other seed investors ($2 million)

Acquired by Progenity (La Jolla)

@CarmentaBio
progenity.com
bit.ly/carm_prog

STOCK WATCH Express, Wet Seal, Avago Jump

Carmenta Bioscience Secures Over $2 Million in Oversubscribed Seed Financing

Camille Samuels Accepts Seat on Carmenta Board of Di

BusinessWire

Press Release: Carmenta Bioscience, Inc. – Wed, A

Email Recommend Tweet Share +1

RELATED CONTENT

PALO ALTO, Calif.--(BUSINESS

BusinessWire
A Berkshire Hathaway Company

Progenity Acquires Carmenta Bioscience for Proprietary Preeclampsia Technology; Appoints Matthew Cooper Chief Scientific Officer

Matthew Herper
Forbes Staff

FOLLOW

PHARMA & HEALTHCARE 8/11/2013 @ 11:10AM | 74,127 views

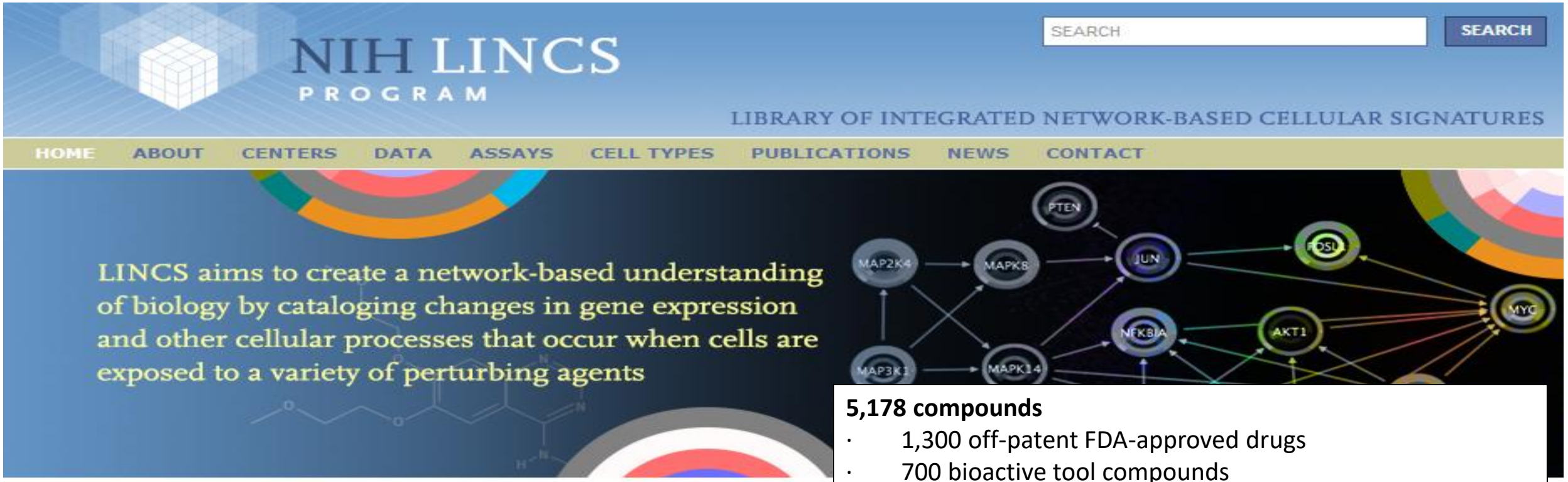
How Much Does Pharmaceutical Innovation Cost? A Look At 100 Companies

+ Comment Now + Follow Comments

Company	Ticker	Number of drugs approved	R&D Spending Per Drug ($Mil)	Total R&D Spending 1997-2011 ($Mil)
AstraZeneca	AZN	5	11,790.93	58,955
GlaxoSmithKline	GSK	10	8,170.81	81,708
Sanofi	SNY	8	7,909.26	63,274
Roche Holding AG	RHHBY	11	7,803.77	85,841
Pfizer Inc.	PFE	14	7,727.03	108,178
Johnson & Johnson	JNJ	15	5,885.65	88,285
Eli Lilly & Co.	LLY	11	4,577.04	50,347
Abbott Laboratories	ABT	8	4,496.21	35,970
Merck & Co Inc	MRK	16	4,209.99	67,360
Bristol-Myers Squibb Co.	BMY	11	4,152.26	
Novartis AG	NVS	21	3,983.13	
Amgen Inc.	AMGN	9	3,692.14	

Sources: InnoThink Center For Research In Biomedical Innovation; Tho Fundamentals via FactSet Research Systems

@MatthewHerper
bit.ly/newdrug1



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, and by using computational tools to integrate this diverse information into a comprehensive view of normal and disease states that can be applied for the development of new biomarkers and therapeutics. By generating and making public data that indicates how cells respond to various genetic and environmental stressors, the LINCS project will help us gain a more detailed understanding of cell pathways and aid efforts to develop therapies that might restore perturbed pathways and networks to their normal states.

5,178 compounds

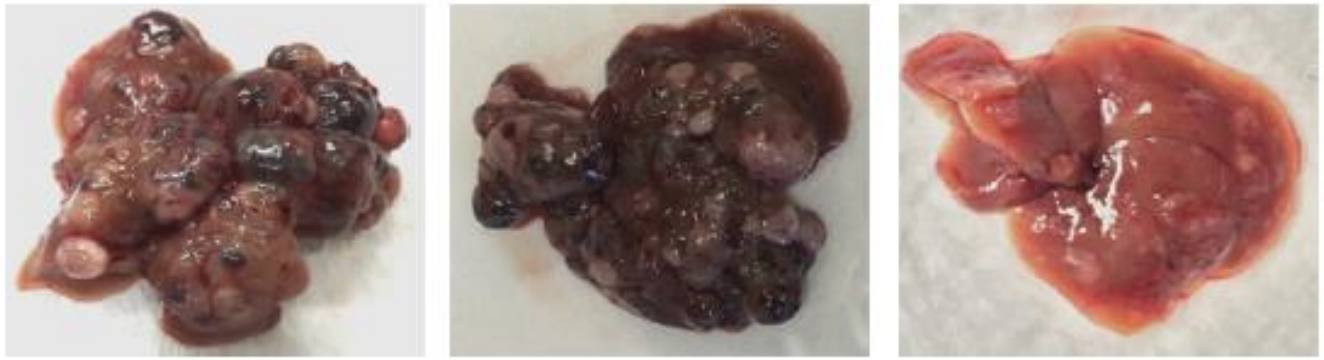
- 1,300 off-patent FDA-approved drugs
- 700 bioactive tool compounds
- 2,000+ screening hits (MLPCN and others)

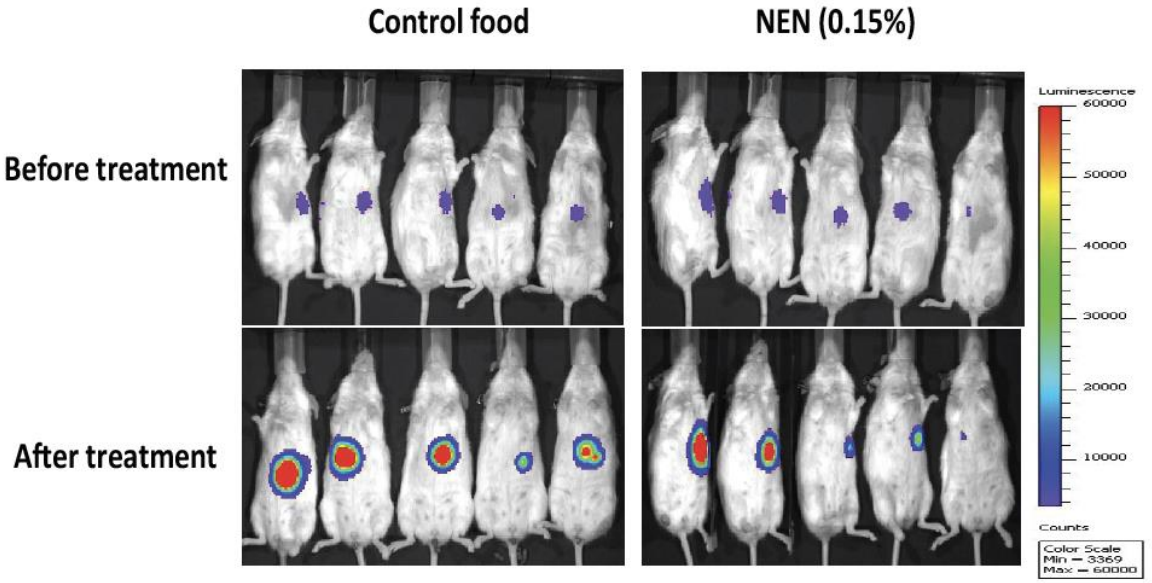
3,712 genes (shRNA + cDNA)

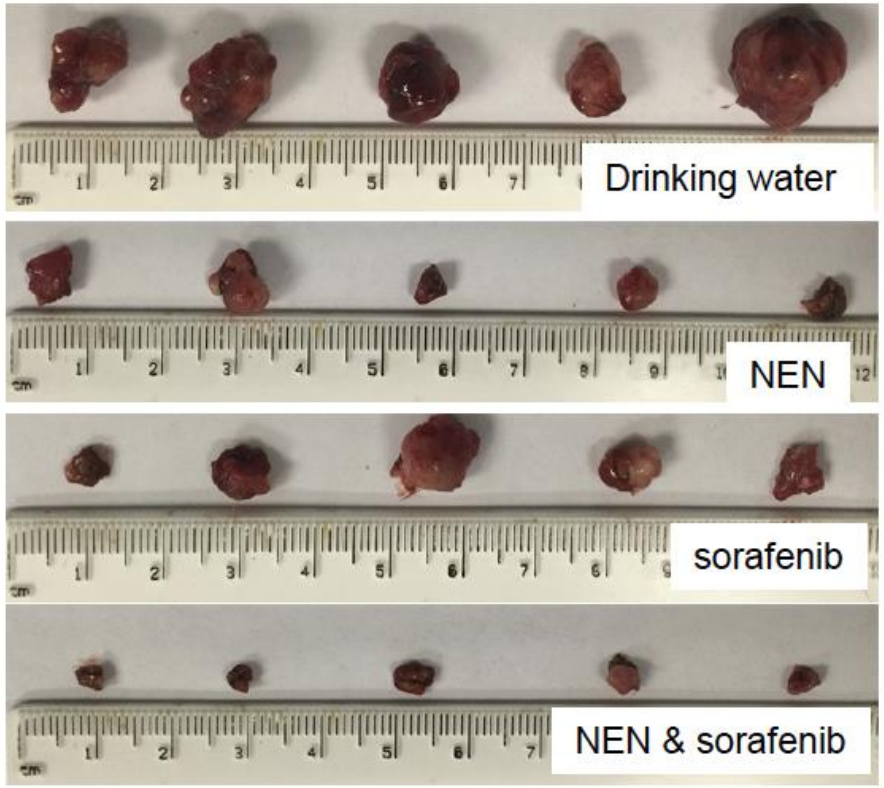
- targets/pathways of FDA-approved drugs (n=900)
- candidate disease genes (n=600)
- community nominations (n=500+)

15 cell types

- Banked primary cell types
- Cancer cell lines
- Primary hTERT immortalized
- Patient derived iPS cells
- 5 community nominated







Bin Chen
Wei Wei
Li Ma
Bin Yang
Mei-Sze Chua
Samuel So

***Gastroenterology*, 2017**

Need more drugs for more diseases

Public big data available

NIH funding

Data analyzed, method designed

Company launched, ARRA, StartX, Stanford license, first deal

Claremont Creek, Lightspeed ($3.5 million)

@NuMedii

NuMedii
Translating Big Data into new medicines

Venture capital

'Digital drug development' company NuMedii snags $3.5 million

Ron Leuty
Reporter-
San Francisco Business Times
Email | Twitter | Google+ | Twitter

NuMedii Inc., the Palo Alto startup looking to convert pages of drug safety data into faster drug-development times, lined up $3.5 million in a Series A round.

Enlarge

NuMedii CEO Gini Deshpande: Tapping old data drugs.

The oversubscribed round was led by Claremont Creek Ventures and Lightspeed Ve Partners and included Life Science Angels and others.

NuMedii's data-into-gold approach rolls a wide range of data — from public scientific data bases and other sources — into an algorithm to predict if a compound will trans

THE WALL STREET JOURNAL. | U.S.

TOP STORIES IN U.S. 1 of 12

VA Scandal Is Shinseki's Latest Battle

2 of 12

U.S. Economy Contracted Last Quarter

U.S. NEWS

Researchers Show Gains in Finding Reus

Email Print

By AMY DOCKSER

August 18, 2011

In a bit of high-tec already-approved combat.

The scientists ha have benefits in t

FierceBiotechIT

Topics: R&D

Allergan taps NuMedii's digit platform for psoriasis R&D

October 5, 2015 | By Nick Paul Taylor

SHARE

NuMedii has landed a deal that could val discovery. Allergan ($AGN) is the compa tments fo

Astellas hooks up with NuMedii to continue drug repurposing deal drive

January 15, 2016 | By Nick Paul Taylor

SHARE

Email

NuMedii, Inc. Announces New Partnership To Discover An Advance New Treatments For Idiopathic Pulmonary Fibros

@nanopore

@wired

Whitehead Institute

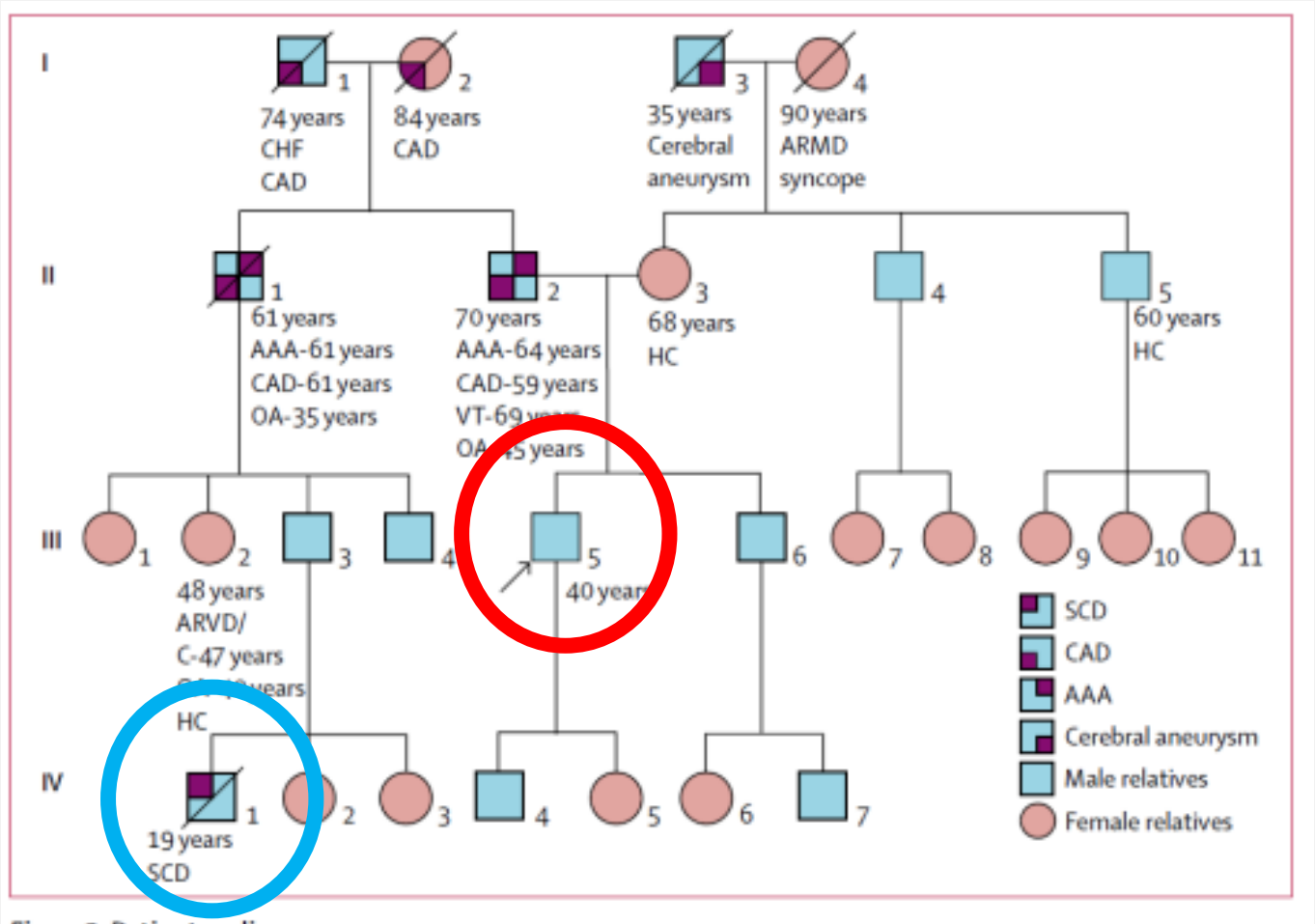
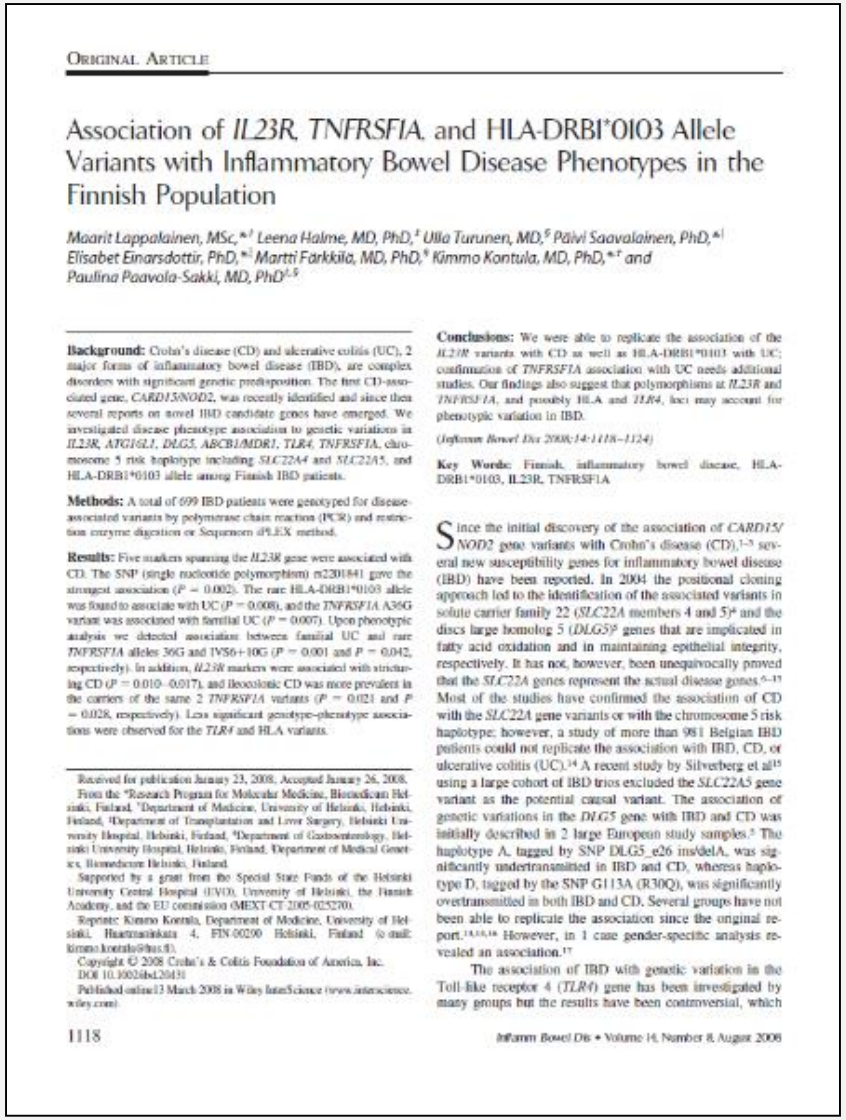


***Figure 2:* Patient pedigree**

The arrow shows the patient. Diagonal lines show relatives who are deceased. Years are age at death or diagnosis. AAA=abdominal aortic aneurysm. ARMD=age-related macular degeneration. ARVD/C=arrhythmogenic right-ventricular dysplasia or cardiomyopathy. CAD=coronary artery disease. CHF=congestive heart failure. HC=hypercholesterolaemia. HTN=hypertension. OA=osteoarthritis. SCD=sudden cardiac death (presumed). VT=paroxysmal ventricular tachycardia.

Credit: Euan Ashley, Russ Altman, Steve Quake, Lancet

Important genome differences “locked up” in publications



ORIGINAL ARTICLE

Association of *IL23R*, *TNFRSF1A*, and HLA-DRB1*0103 Allele Variants with Inflammatory Bowel Disease Phenotypes in the Finnish Population

Maarit Lappalainen, MSc,[*,†] *Leena Halme, MD, PhD,*[‡] *Ulla Turunen, MD,*[§] *Päivi Saavalainen, PhD,*[*,¶] *Elisabet Einarsdottir, PhD,*[*,¶] *Martti Farkkila, MD, PhD,*[‖] *Kimmo Kontula, MD, PhD,*[*,**] *and Paulina Paavola-Sakki, MD, PhD*[*,‖]

Background: Crohn's disease (CD) and ulcerative colitis (UC), 2 major forms of inflammatory bowel disease (IBD), are complex disorders with significant genetic predisposition. The first CD-associated gene, *CARD15/NOD2*, was recently identified and since then several reports on novel IBD candidate genes have emerged. We investigated disease phenotype association to genetic variations in *IL23R*, *ATG16L1*, *DLG5*, *ABCB1/MDR1*, *TLR4*, *TNFRSF1A*, chromosome 5 risk haplotype including *SLC22A4* and *SLC22A5*, and HLA-DRB1*0103 allele among Finnish IBD patients.

Methods: A total of 699 IBD patients were genotyped for disease-associated variants by polymerase chain reaction (PCR) and restriction enzyme digestion or Sequenom iPLEX method.

Results: Five markers spanning the *IL23R* gene were associated with CD. The SNP (single nucleotide polymorphism) rs2201841 gave the strongest association ($P = 0.002$). The rare HLA-DRB1*0103 allele was found to associate with UC ($P = 0.008$), and the *TNFRSF1A* A36G variant was associated with familial UC ($P = 0.007$). Upon phenotypic analysis we detected association between familial UC and rare *TNFRSF1A* alleles 36G and IVS6+10G ($P = 0.001$ and $P = 0.042$, respectively). In addition, *IL23R* markers were associated with stricturing CD ($P = 0.010–0.017$), and ileocolonic CD was more prevalent in the carriers of the same 2 *TNFRSF1A* variants ($P = 0.021$ and $P = 0.028$, respectively). Less significant genotype–phenotype associations were observed for the *TLR4* and HLA variants.

Conclusions: We were able to replicate the association of the *IL23R* variants with CD as well as HLA-DRB1*0103 with UC; confirmation of *TNFRSF1A* association with UC needs additional studies. Our findings also suggest that polymorphisms at *IL23R* and *TNFRSF1A*, and possibly HLA and *TLR4*, loci may account for phenotypic variation in IBD.

(*Inflamm Bowel Dis 2008;14:1118–1124*)

Key Words: Finnish, inflammatory bowel disease, HLA-DRB1*0103, IL23R, TNFRSF1A

Since the initial discovery of the association of *CARD15/NOD2* gene variants with Crohn's disease (CD),[1–5] several new susceptibility genes for inflammatory bowel disease (IBD) have been reported. In 2004 the positional cloning approach led to the identification of the associated variants in solute carrier family 22 (*SLC22A* members 4 and 5)[6] and the discs large homolog 5 (*DLG5*)[5] genes that are implicated in fatty acid oxidation and in maintaining epithelial integrity, respectively. It has not, however, been unequivocally proved that the *SLC22A* genes represent the actual disease genes.[5–13] Most of the studies have confirmed the association of CD with the *SLC22A* gene variants or with the chromosome 5 risk haplotype; however, a study of more than 981 Belgian IBD patients could not replicate the association with IBD, CD, or ulcerative colitis (UC).[14] A recent study by Silverberg et al[15] using a large cohort of IBD trios excluded the *SLC22A5* gene variant as the potential causal variant. The association of genetic variations in the *DLG5* gene with IBD and CD was initially described in 2 large European study samples.[5] The haplotype A, tagged by SNP DLG5_e26 ins/delA, was significantly undertransmitted in IBD and CD, whereas haplotype D, tagged by the SNP G113A (R30Q), was significantly overtransmitted in both IBD and CD. Several groups have not been able to replicate the association since the original report.[13,15,16] However, in 1 case gender-specific analysis revealed an association.[17]

The association of IBD with genetic variation in the Toll-like receptor 4 (*TLR4*) gene has been investigated by many groups but the results have been controversial, which

Received for publication January 23, 2008; Accepted January 26, 2008.

From the *Research Program for Molecular Medicine, Biomedicum Helsinki, Finland, †Department of Medicine, University of Helsinki, Helsinki, Finland, ‡Department of Transplantation and Liver Surgery, Helsinki University Hospital, Helsinki, Finland, §Department of Gastroenterology, Helsinki University Hospital, Helsinki, Finland, ‖Department of Medical Genetics, Biomedicum Helsinki, Finland.

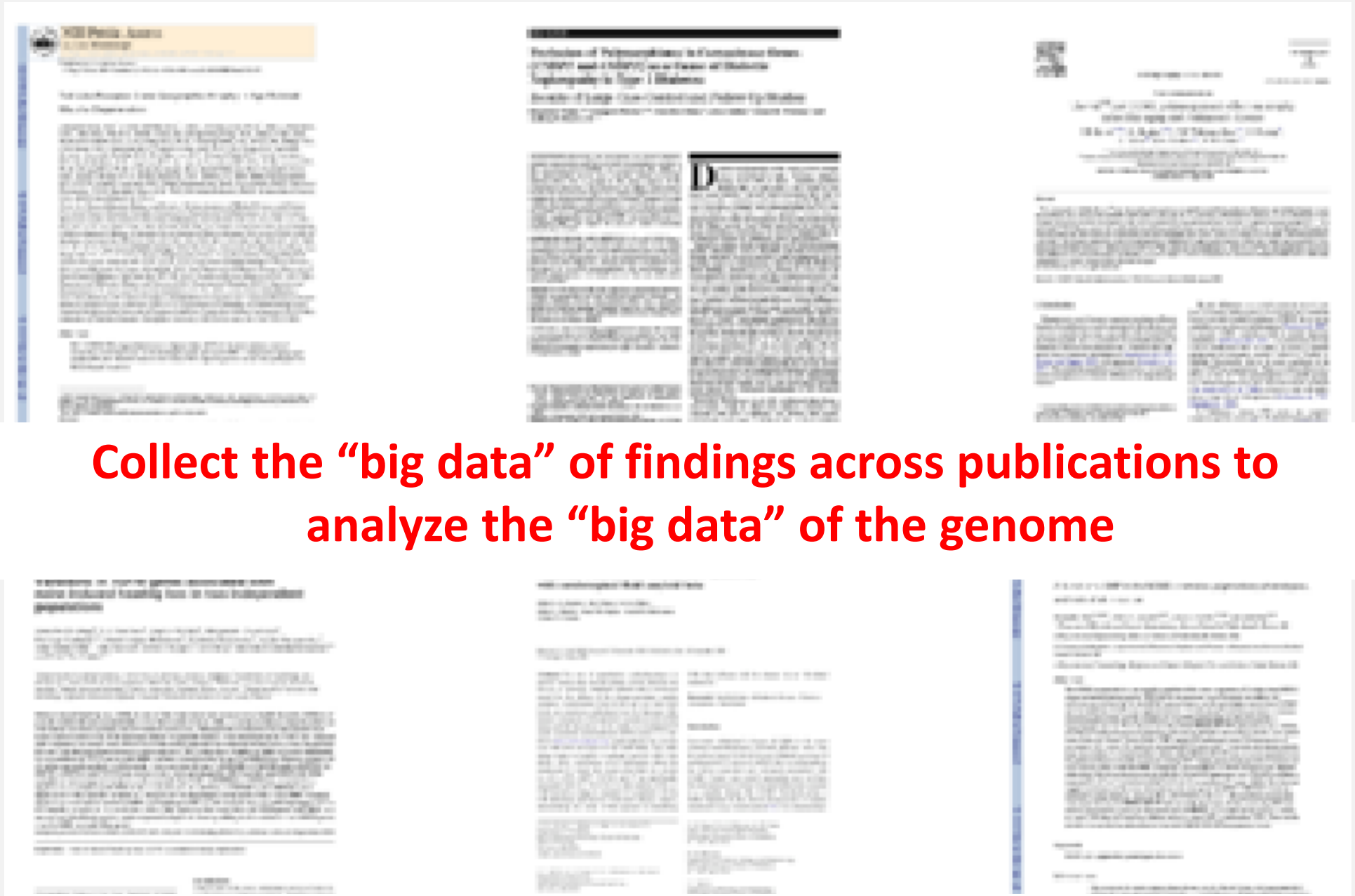
Supported by a grant from the Special State Funds of the Helsinki University Central Hospital (EVO), University of Helsinki, the Finnish Academy, and the EU commission (MEXT-CT-2005-025270).

Reprints: Kimmo Kontula, Department of Medicine, University of Helsinki, Haartmaninkatu 4, FIN-00290 Helsinki, Finland (e-mail: kimmo.kontula@hus.fi).

DOI 10.1002/ibd.20431

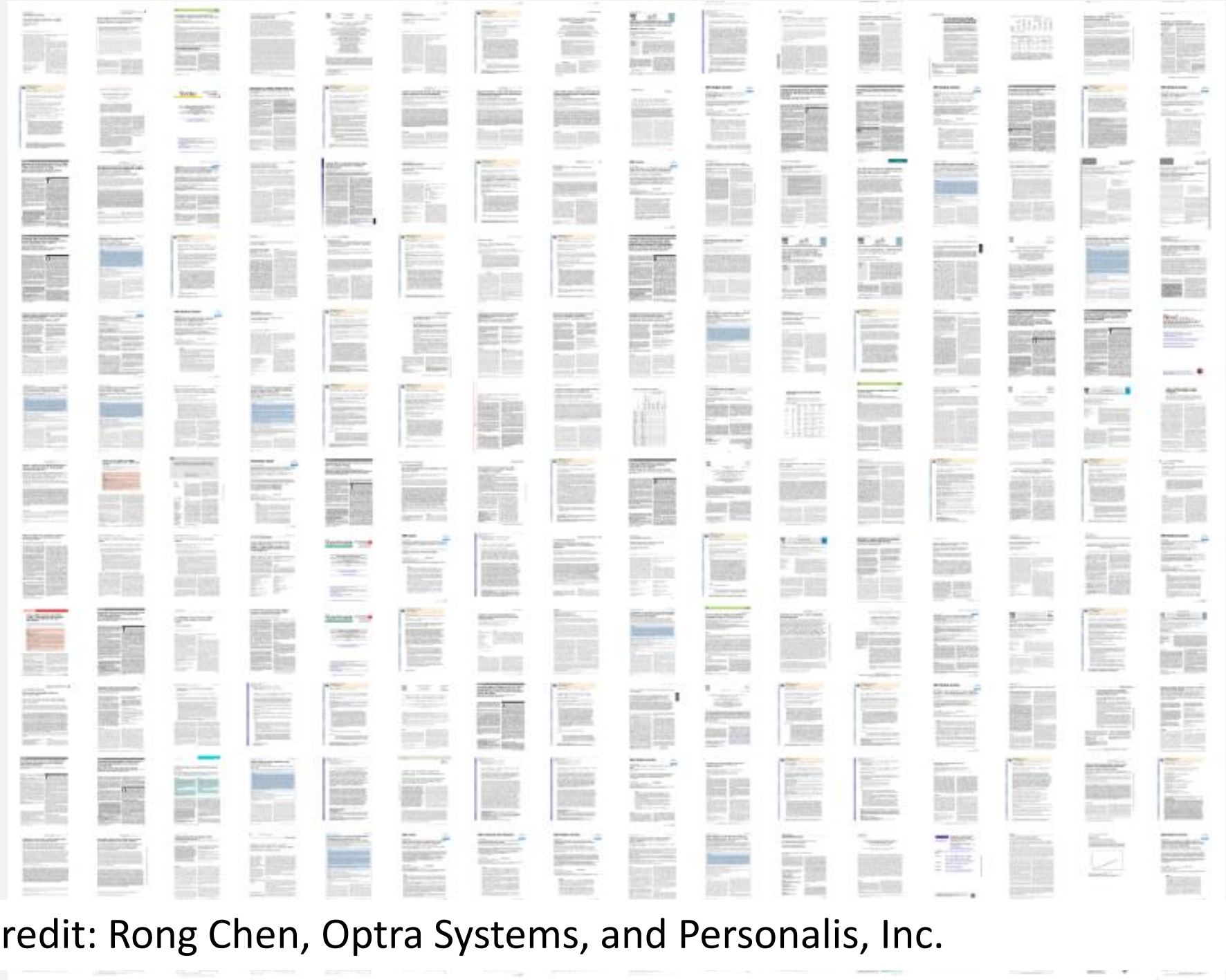
Published online 13 March 2008 in Wiley InterScience (www.interscience.wiley.com).

Credit: Rong Chen, Optra Systems, and Personalis, Inc.

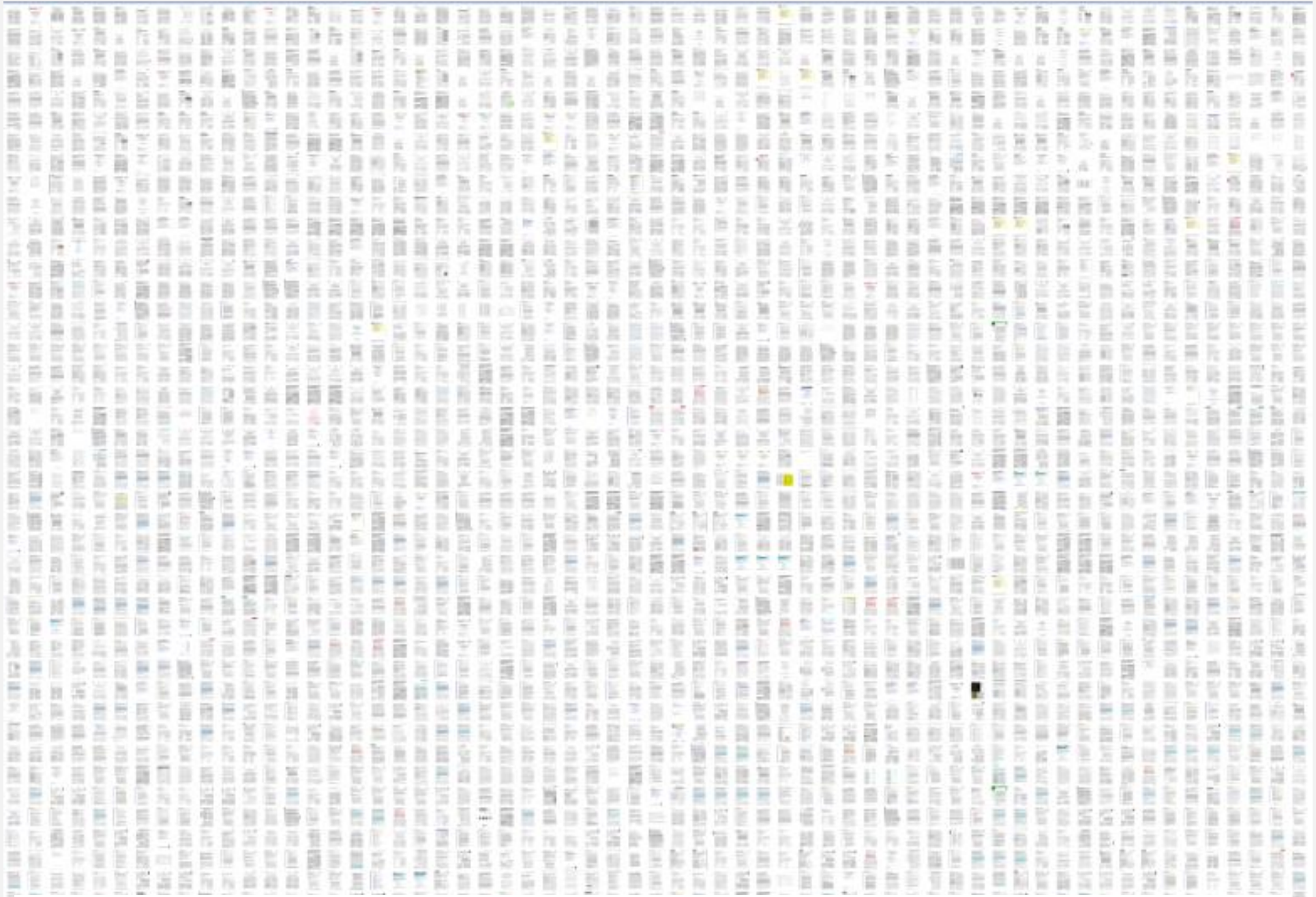


Collect the “big data” of findings across publications to analyze the “big data” of the genome

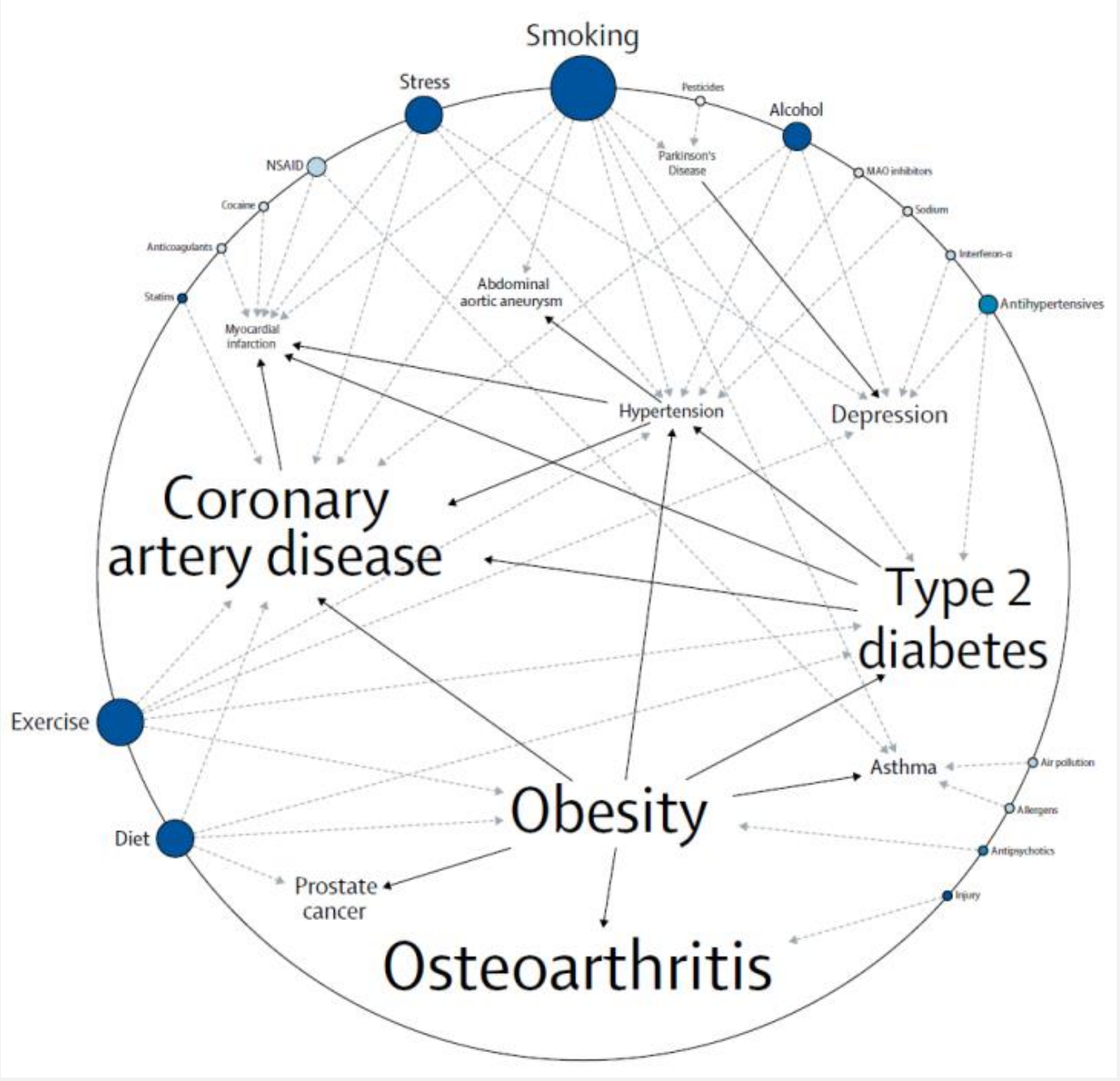
Credit: Rong Chen, Optra Systems, and Personalis, Inc.



Credit: Rong Chen, Optra Systems, and Personalis, Inc.



Credit: Rong Chen, Optra Systems, and Personalis, Inc.



Maybe the genome can be used to suggest (promote?) preventative health strategies?

Credit: Rong Chen, Alex Morgan, Joel Dudley, Lancet

Need to use genomes to predict disease (2008)

Publications available for curation

Stanford donor funding

Company launched, Stanford license

MDV, Lightspeed, Abingworth ($20 million)

Same 3 plus Wellington Shields ($22 million)

Series C ($33 million)

IPO (2019, $141 million)

STOCK WATCH Express, Wet Seal, Avago Jump

Personalis Awarded Contract From VA Million Veteran Program – Whole Genome Sequencing and Data Analysis for Over 1,000 Individuals

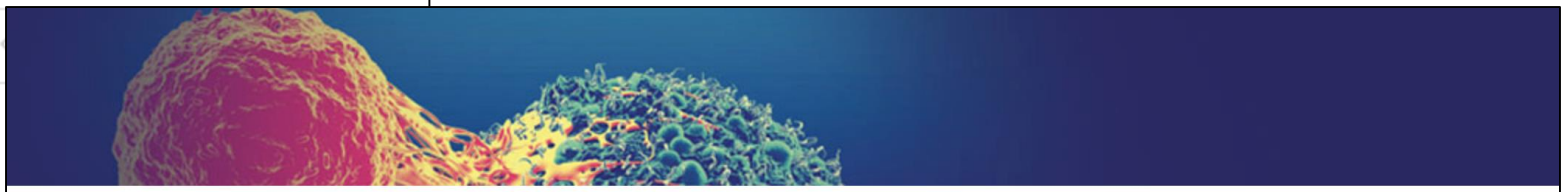
BusinessWire

Press Release: Personalis, Inc. – Tue, Mar 12, 2013 12:02 AM EDT

Recommend Tweet Share +1 Pocket

MENLO PARK, Calif.--(BUSINESS WIRE)--

The US Department of Veterans Affairs (VA) has awarded its first co
sequencing and data analysis to Personalis, Inc., of Menlo Park, CA.
samples from several VA sources, including from the Million Veteran
secure computing facility and proprietary algorithms, Personalis will
against an advanced human reference sequence, annotate both SNV/
genetic analyses to help confirm sample / data chain of custody. Pers
laboratory genetic analysis, including both DNA sequencing and gen
Illumina, Inc., of San Diego, CA.



Personalis Announces Closing of Initial Public Offering and Exercise in Full of Over Allotment Option

June 24, 2019

PDF Version

MENLO PARK, Calif.--(BUSINESS WIRE)--Jun. 24, 2019-- Personalis, Inc. (Nasdaq: PSNL), a leader in advanced genomics for cancer, today announced the closing of its initial public offering of 9,109,725 shares of common stock at a public offering price of $17.00 per share, which includes the exercise in full by the underwriters of their option to purchase up to 1188 225 additional shares of common stock. All of the shares

Four Big Lessons Learned in Building Big Data Ecosystems

- Sufficient data already exists to impact medicine
  - Diagnostics and drugs from public big data
  - More data is better, but never a reason to wait for more
  - “Retroactive crowdsourcing”: robust findings come from integrated data sets
  - Real success use-cases get data storers, data users, and data contributors excited
- Public and open data is already extremely high quality
  - Should never wait for perfect data, experiment, conditions
- Sticks seem to work better than carrots for sharing
  - Continue exponential growth, ask grantees to share more
- Need more question askers, train students to initiate science with data
  - High school → higher education → career changers