Lucile Packard Children's Hospital AT STANFORD

Transforming a Trillion Points of Data into Diagnostics, Therapeutics, and New Insights into Disease

Atul Butte, MD, PhD abutte@stanford.edu

Chief, Division of Systems Medicine, @atulbutte
Department of Pediatrics,
Department of Medicine, and, by courtesy,
Computer Science

Center for Pediatric Bioinformatics, LPCH

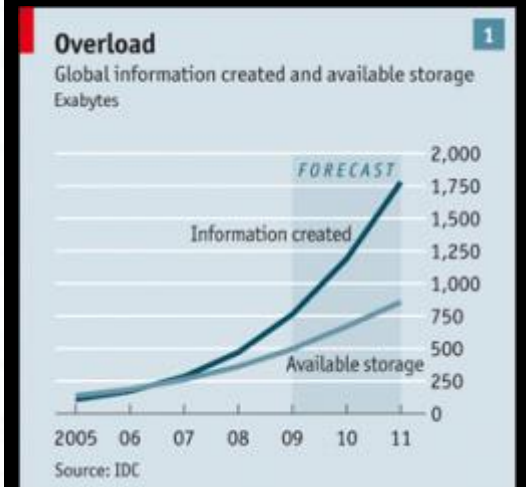
Stanford University

Butte Lab
Systems Medicine • Stanford Pediatrics • Packard Children's Hospital

Disclosures

- Scientific founder and advisory board membership
  - Genstruct
  - NuMedii
  - Personalis
  - Carmenta
- Past or present consultancy
  - Lilly
  - Johnson and Johnson
  - Roche
  - NuMedii
  - Genstruct
  - Tercica
  - Ansh Labs
  - Prevendia
  - Samsung
  - Assay Depot
- Honoraria
  - Lilly
  - Pfizer
  - Siemens
  - Bristol Myers Squibb
- Speakers' bureau
  - None
- Companies started by students
  - Carmenta
  - Serendipity
  - NuMedii
  - Stimulomics
  - NunaHealth
  - Praedicat
  - Flipora

Kilo
Mega
Giga
Tera
Peta
Exa
Zetta



WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

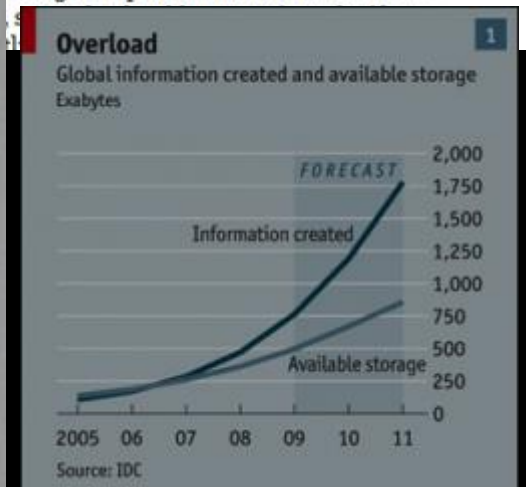
By Chris Anderson 06.23.08

Illustration: Marian Bantjes

THE PETABYTE AGE:

"All models are wrong, but some are useful."

imed statistician George Box 30 years ago, and ght. But what choice did we have? Only models, nological equations to theories of human



The Next Scientific Revolution

The Four Paradigms of Science

THEORY
Beginning in ancient Greece and China, people tried to explain their observations through natural laws instead of supernatural causes.

EXPERIMENTATION
By the 17th century, scientists like Isaac Newton tried to make predictions for new phenomena and would verify hypotheses by conducting experiments.

COMPUTATION AND SIMULATION
The advent of high-performance computers in the latter half of the 20th century allowed scientists to explore regimes inaccessible to experiment and theory, such as climate modeling or galaxy formation, by numerically solving systems of equations on a large scale and in fine detail.

DATA MINING
Using more-powerful computers, scientists begin with the data and direct programs to mine enormous databases for relationships. In essence, they use computers to discover the rules by studying the data.

nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2

Data's shameful neglect

Research canno

Brooks Hanson is Deputy Editor for physical sciences at

Making Data Maximally Available

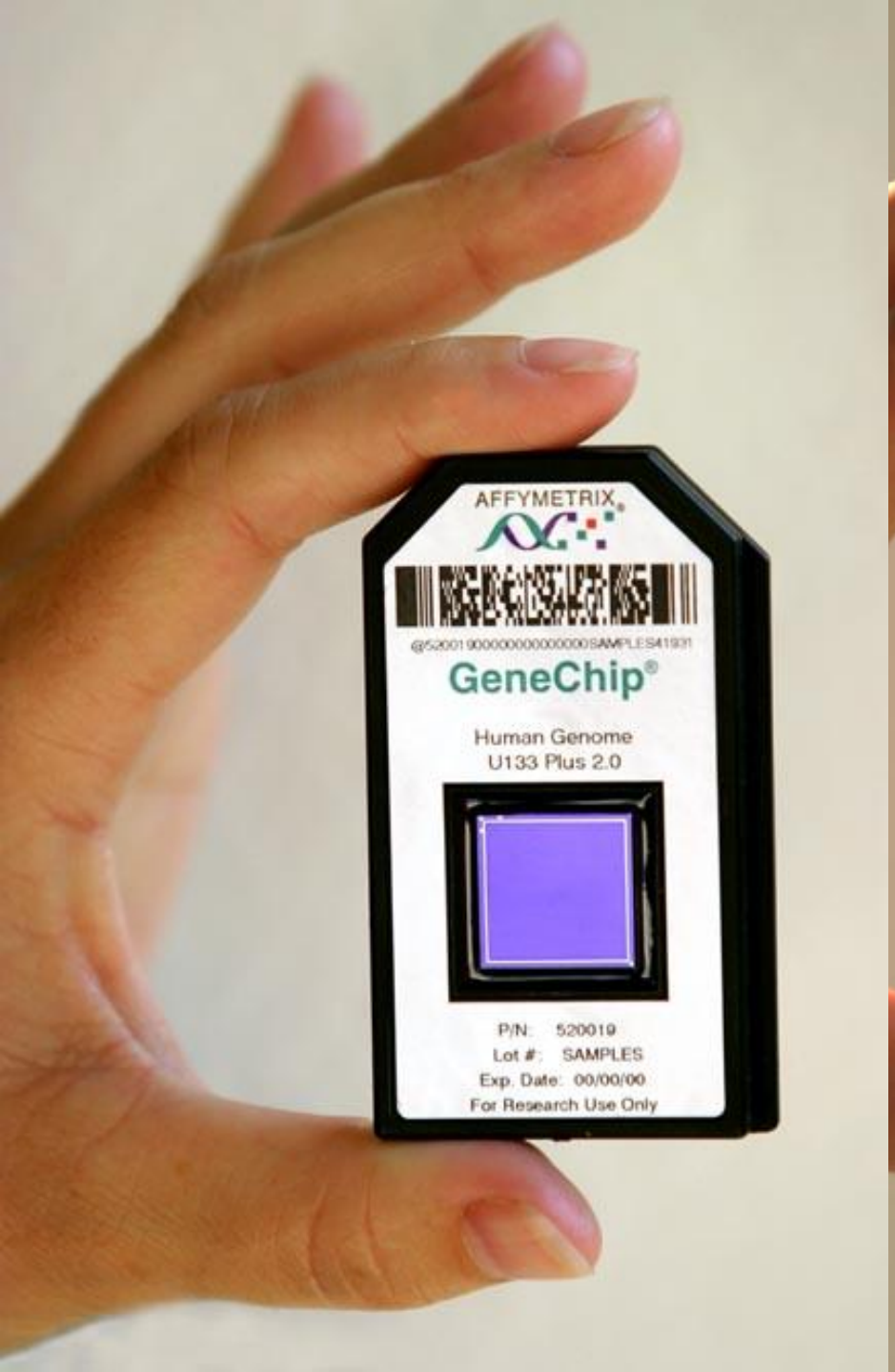
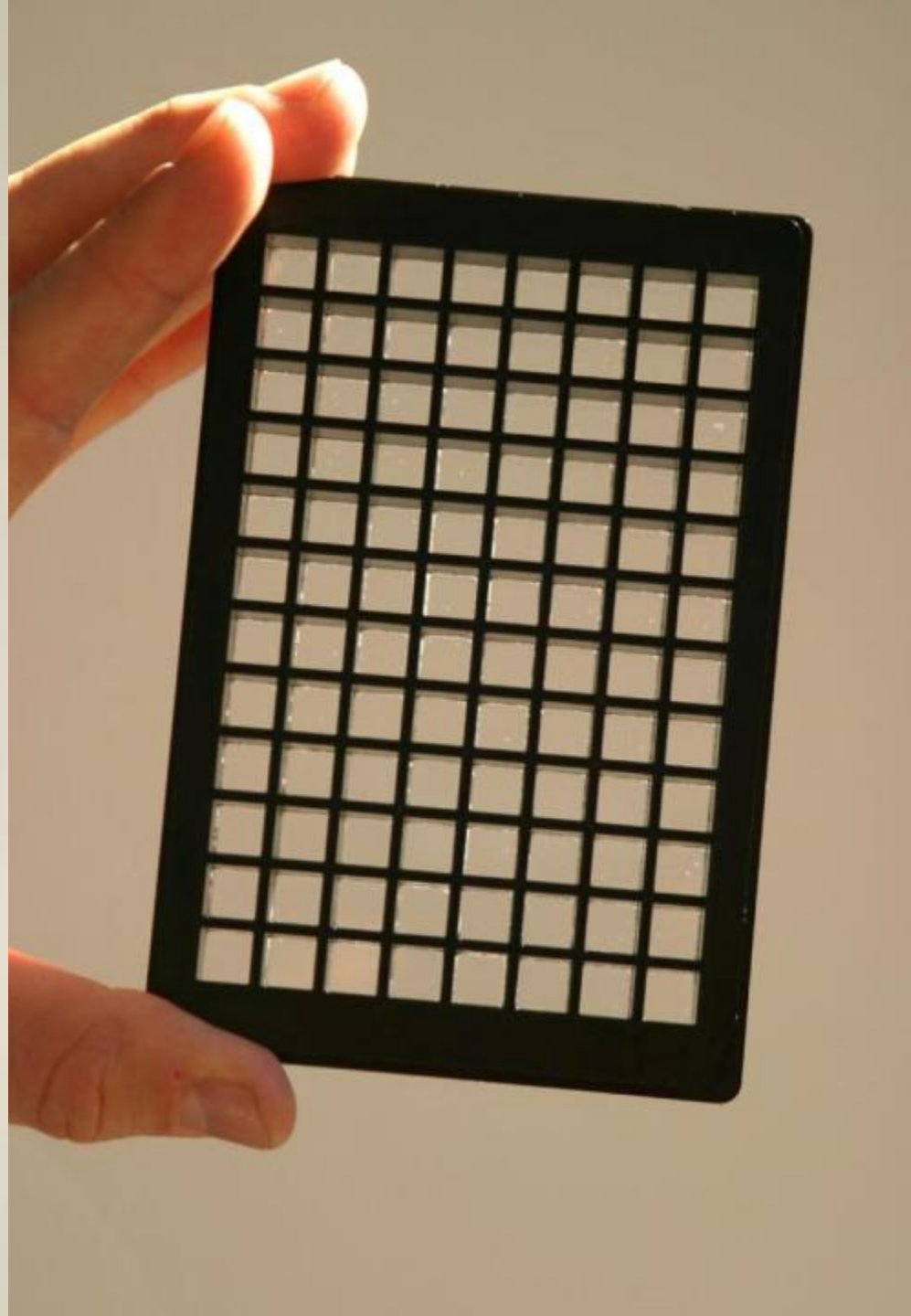
Sharing research data to improve public health

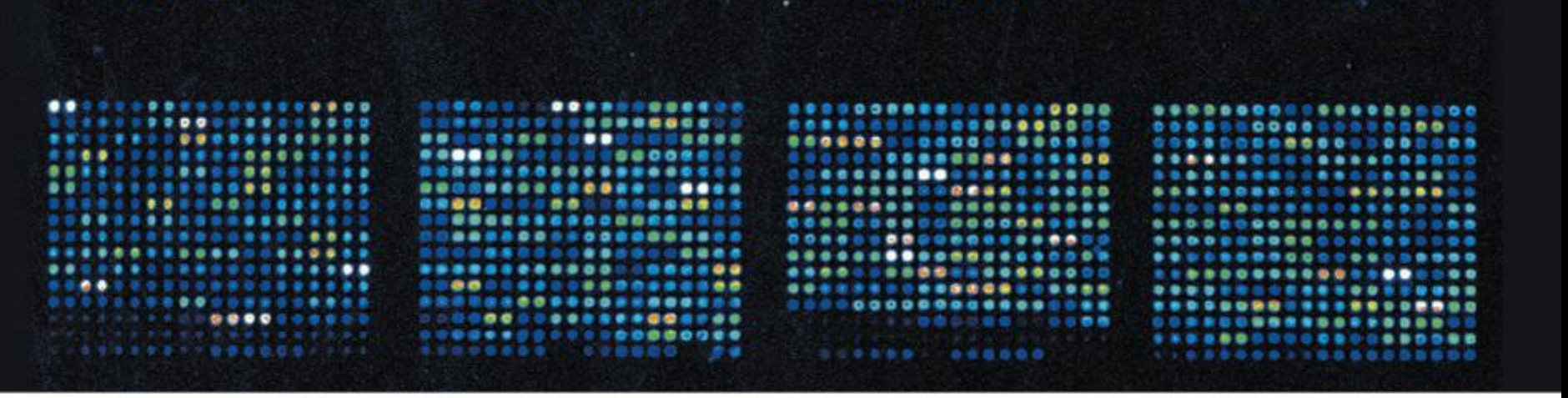
The purpose of medical research is to analyse and understand health and disease. A key and expensive element is the study of populations to explore how interactions between behaviour and environment, in the context of genetic diversity, determine causation and variation in

that every last ounce of knowledge will be wrung from the research.

Ensuring data are made widely available to the research community accelerates the pace of discovery and enhances the efficiency of the research enterprise.

Published Online
January 10, 2011
DOI:10.1016/S0140-6736(10)62234-9


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




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 gene-expression data sets now in publi researchers can identify disease trends without ever having to e

BY MONYA BAKER

Purvesh Khatri sits in front of an oversized computer screen, trawling for treasure in a sea of genetic data. Entering the search term 'breast cancer' into a public repository called the Gene Expression Omnibus (GEO), the postdoctoral researcher retrieves a list of 1,170 experiments, representing nearly 33,000 samples and a hoard of gene-expression data that could reveal previously unseen patterns.

That is exactly the kind of search that led Khatri's boss, Atul Butte, a bioinformatician at the Stanford School of Medicine in California, to identify a new drug target for diabetes. After downloading data from 130 gene-expression

for discovery," he says. Those are for validating hypotheses. The beauty of analysing data from multiple experiments is that biases and artefacts should cancel out between data sets, helping true relationships to stand out, Butte says. "There is safety in numbers."

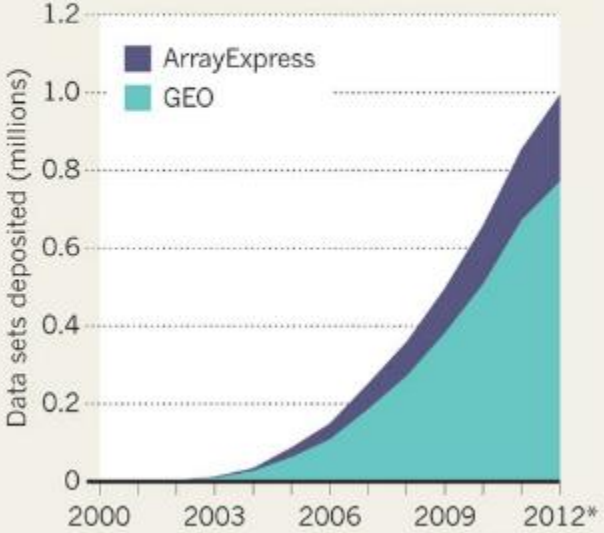
And those numbers are rising rapidly. Since 2002, many scientific journals have required that data from gene-expression studies be deposited in public databases such as GEO, which is maintained by the National Center for Biotechnology Information in Bethesda, Maryland, and ArrayExpress, a large gene-expression

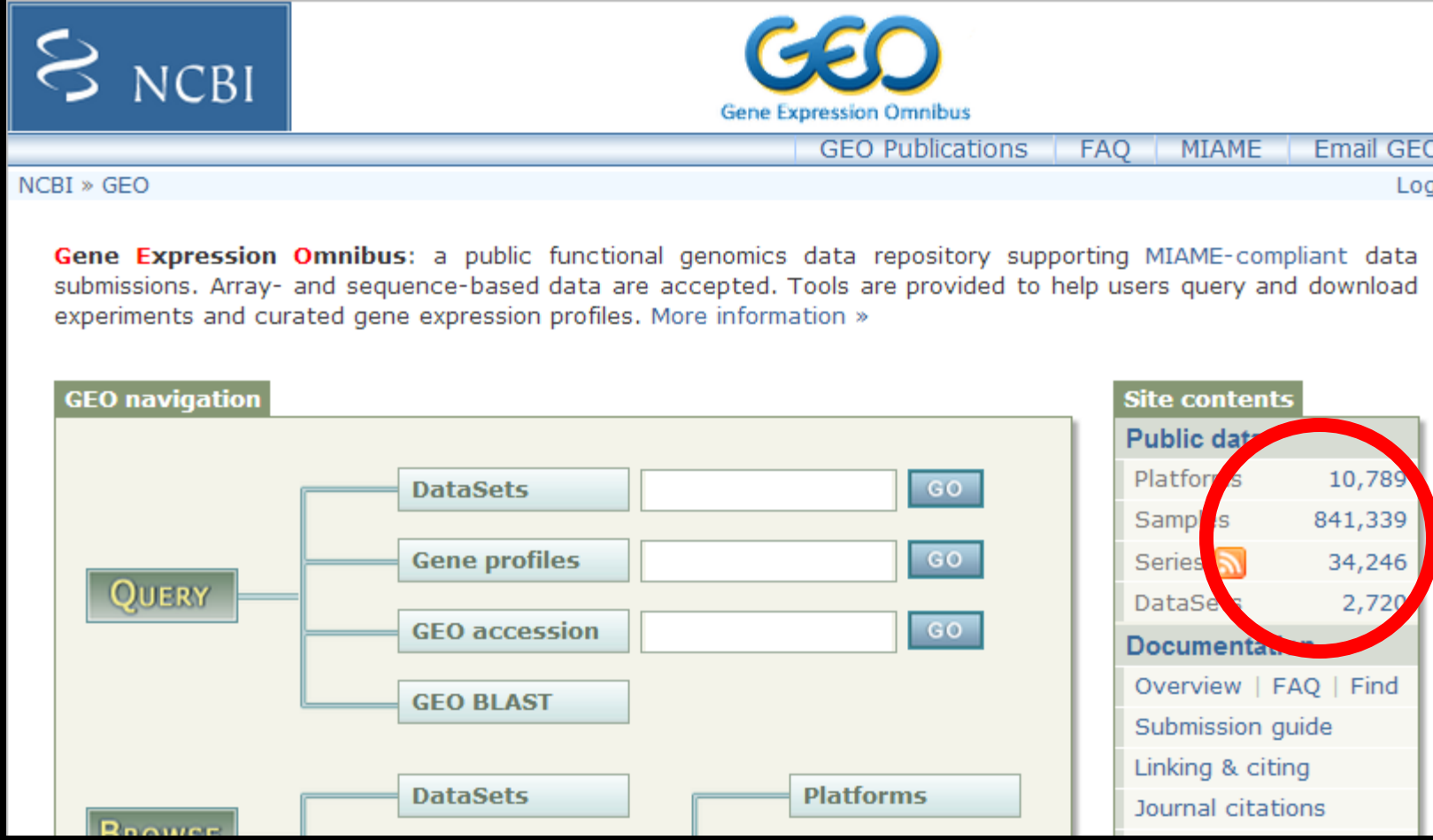
re
In
th
da
pr
gr
da
si
w
ve
tis
ge
as
or

DATA DUMP

DATA DUMP

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




NCBI
Gene Expression Omnibus
GEO Publications
FAQ
MIAME
Email GEO
NCBI » GEO
Log
Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. More information »
GEO navigation
QUERY
DataSets
GO
Gene profiles
GO
GEO accession
GO
GEO BLAST
DataSets
Platforms
Site contents
Public data
Platforms 10,789
Samples 841,339
Series 34,246
DataSets 2,720
Documentation
Overview | FAQ | Find
Submission guide
Linking & citing
Journal citations

Total 1.1 million microarrays available

Doubles every 2-3 years

Butte AJ. Translational Bioinformatics: coming of age. *JAMIA*, 2008.

NCBI Resources How To

GEO DataSets

GEO DataSets | breast cancer | Sear

Save search Limits Advanced

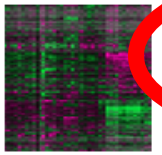
Display Settings: Summary, 20 per page, Sorted by Default order

Send to:

Results: 1 to 20 of 35583

<< First < Prev Page 1 of 1780 Next > Last >>

1. **Breast cancer**: histologically normal breast epithelium

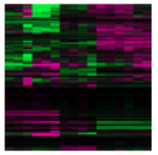


Analysis of histological normal breast epithelia from both ER- and ER+ **breast cancer** patients and prophylactic mastectomy patients, and normal breast epithelia from reduction mammoplasty patients. Results provide insight into the mechanisms underlying **breast cancer** initiation and progression.

Organism: Homo sapiens
Type: Expression profiling by array, count, 2 disease state, 4 specimen sets
Platform: GPL96 Series: GSE20437 42 Samples
Download data: GEO (CEL)

DataSet Accession: GDS3716 ID: 3716
PubMed Full text in PMC Similar studies GEO Profiles Analyze DataSet

2. Actein effect on **breast cancer** cell line: dose response and time course



Analysis of MDB-MB-453 **breast cancer** cells treated with 20 or 40 ug/ml actein for 6 or 24 hours. Actein is a triterpene glycoside from the herb black cohosh and inhibits the growth of cancer cells in vitro. Results provide insight into the molecular basis of this inhibitory effect.

Organism: Homo sapiens
Type: Expression profiling by array, transformed count, 2 agent, 3 dose, 2 time sets
Platform: GPL571 Series: GSE7848 16 Samples
Download data: GEO (CEL)

DataSet Accession: GDS3638 ID: 3638
PubMed Similar studies GEO Profiles Analyze DataSet

Filter your results:

All (35583)
DataSets (90)
Platforms (27)
Samples (34162)
Series (1304)

Top Organisms [
Homo sapiens (330
Mus musculus (242
Rattus norvegicus (
Canis lupus familiar
Human herpesvirus
More...

Find related data

Database: Select

Find items

Browse dbGaP

db GaP GENOTYPE and PHENOTYPE

By Studies | By Diseases | Advanced Search

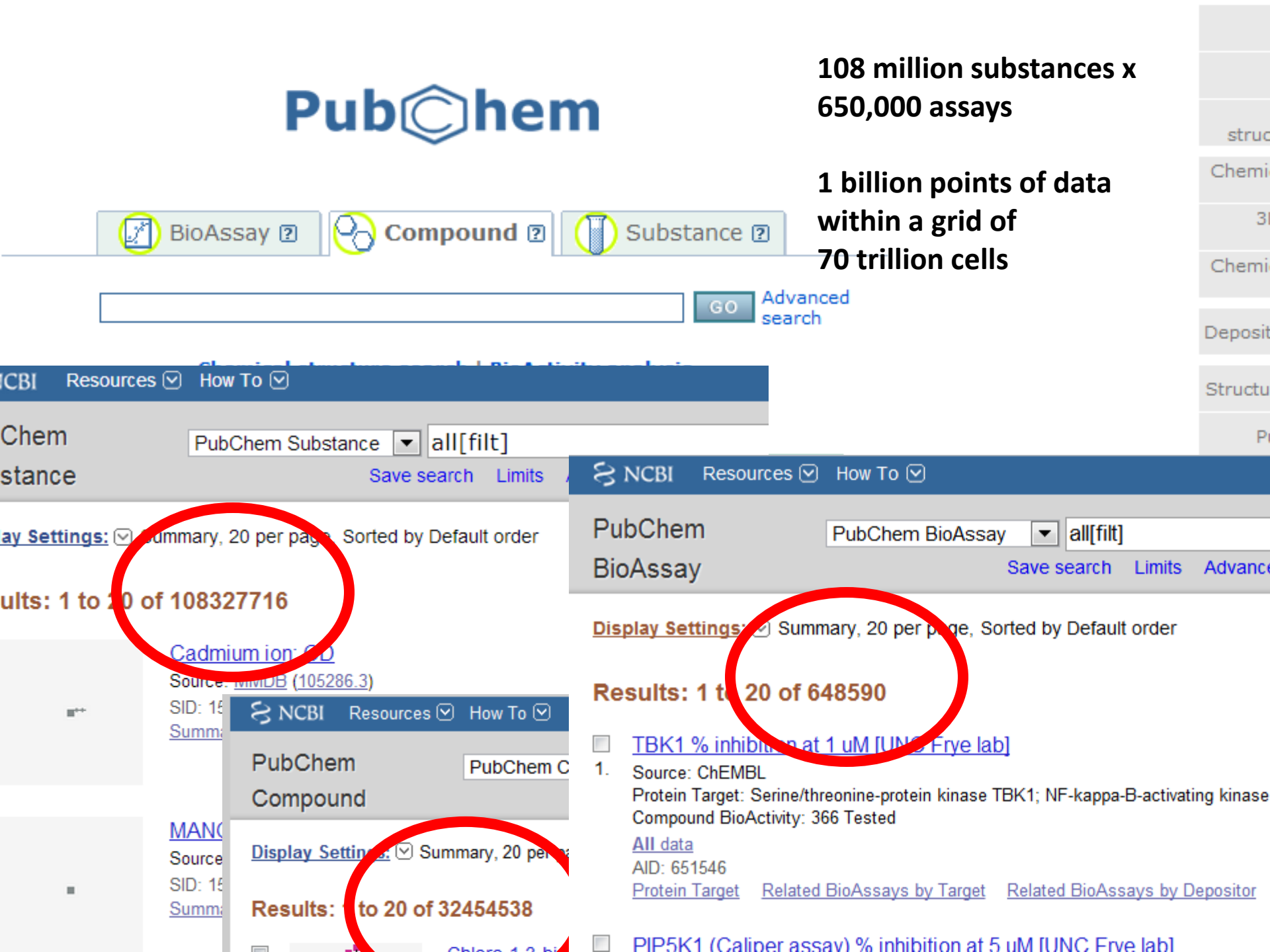
Study	Embargo Release	Details	Participants	Type of St
CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)	Feb 13, 2009	V D A	1991	Case-cont
Framingham SHARe	Version 1: Oct 19, 2008 Version 2: Feb 01, 2009 Version 3: Jul 08, 2009	V D A	14277	Longitudir
GAIN: Collaborative Association Study of Psoriasis	Aug 13, 2008	V D A	2875	Case-cont
GAIN: Genotyping the 270 HapMap samples for GAIN by Broad		V D A	-	Parent-offsprir
GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen		V D A	-	Parent-offsprir
GAIN: International Multi-Center ADHD Genetics Project	Mar 26, 2008	V D A	2835	Parent-offsprir
GAIN: Linking Genome-Wide Association Study of Schizophrenia	Version 1: Nov 07, 2008 Version 2: Dec 03, 2008	V D A	5066	Case-cont
GAIN: Major Depression: Stage 1 Genomewide Association in Population-Based Samples	Jul 09, 2008	V D A	3741	Case-cont
GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes	Jul 09, 2008	V D A	1825	Case-cont
GAIN: Whole Genome Association Study of Bipolar Disorder	Version 1: Nov 25, 2008 Version 2: Dec 01, 2008	V D A	3261	Case-cont
GAW16 Framingham and Simulated Data	Oct 19, 2008	V D A	7130	Longitudir population-b
Genome-wide Association Studies in the Hutterites		V D A	632	Population-b
Genome-wide Association Study of Neuroblastoma		V D A	1032	Case-cont
Genome-wide Study in Amyotrophic Lateral Sclerosis and Controls: First Stage Analysis	Jun 26, 2008	V D A	544	Case-cont
Ischemic Stroke Genetics Study (ISGS)	Jun 26, 2008	V D A	485	Case-cont
Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration	Mar 03, 2008	V D A	1550	Case-cont
NEI Age-Related Eye Disease Study (AREDS)	Jun 11, 2007	V D A	600	Case-cont
NINDS Parkinson's Disease	Oct 12, 2007	V D A	535	Case-cont
NINDS Parkinsonism Study	Oct 12, 2007	V D A	1283	Case-se
NINDS Repository Cerebrovascular Disease/Stroke Study	Jun 26, 2008	V D A	870	Case-se
NINDS Repository Motor Neuron Disease/ALS Study	Jun 26, 2008	V D A	1790	Case-se
NINDS Repository Neurologically Normal Control Collection	Oct 12, 2007	V D A	2723	Control-s
POPRES: Population Reference Sample		V D A	5919	Population sa Control-s
SEARCH GWA Study of Statin-Induced Myopathy		V D A	175	Case-cont
Study of Irish Amyotrophic Lateral Sclerosis (SIALS)		V D A	432	Case-cont
The Finland-United States Investigation of NIDDM Genetics (FUSION) study		V D A	2335	Case-cont
Whole Genome Association Study of Systemic Lupus Erythematosus		V D A	4651	Case-cont

Browse dbGaP

db GaP GENOTYPE and PHENOTYPE

By Studies | By Diseases | Advanced Search

Study	Embargo Release	Details	Participants	Type of St
CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)	Feb 13, 2009	V D A	1991	Case-cont
Framingham SHARe	Version 1: Oct 19, 2008 Version 2: Feb 01, 2009 Version 3: Jul 08, 2009	V D A	14277	Longitudir
GAIN: Collaborative Association Study of Psoriasis	Aug 13, 2008	V D A	2875	Case-cont
GAIN: Genotyping the 270 HapMap samples for GAIN by Broad		V D A	-	Parent-offsprir
GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen		V D A	-	Parent-offsprir
GAIN: International Multi-Center ADHD Genetics Project	Mar 26, 2008	V D A	2835	Parent-offsprir
GAIN: Linking Genome-Wide Association Study of Schizophrenia	Version 1: Nov 07, 2008 Version 2: Dec 03, 2008	V D A	5066	Case-cont
GAIN: Major Depression: Stage 1 Genomewide Association in Population-Based Samples	Jul 09, 2008	V D A	3741	Case-cont
GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes	Jul 09, 2008	V D A	1825	Case-cont
GAIN: Whole Genome Association Study of Bipolar Disorder	Version 1: Nov 25, 2008 Version 2: Dec 01, 2008	V D A	3261	Case-cont
GAW16 Framingham and Simulated Data	Oct 19, 2008	V D A	7130	Longitudir population-b
Genome-wide Association Studies in the Hutterites		V D A	632	Population-b
Genome-wide Association Study of Neuroblastoma		V D A	1032	Case-cont
Genome-wide Study in Amyotrophic Lateral Sclerosis and Controls: First Stage Analysis	Jun 26, 2008	V D A	544	Case-cont
Ischemic Stroke Genetics Study (ISGS)	Jun 26, 2008	V D A	485	Case-cont
Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration	Mar 03, 2008	V D A	1550	Case-cont
NEI Age-Related Eye Disease Study (AREDS)	Jun 11, 2007	V D A	600	Case-cont
NINDS Parkinson's Disease	Oct 12, 2007	V D A	535	Case-cont
NINDS Parkinsonism Study	Oct 12, 2007	V D A	1283	Case-se
NINDS Repository Cerebrovascular Disease/Stroke Study	Jun 26, 2008	V D A	870	Case-se
NINDS Repository Motor Neuron Disease/ALS Study	Jun 26, 2008	V D A	1790	Case-se
NINDS Repository Neurologically Normal Control Collection	Oct 12, 2007	V D A	2723	Control-s
POPRES: Population Reference Sample		V D A	5919	Population sa Control-s
SEARCH GWA Study of Statin-Induced Myopathy		V D A	175	Case-cont
Study of Irish Amyotrophic Lateral Sclerosis (SIALS)		V D A	432	Case-cont
The Finland-United States Investigation of NIDDM Genetics (FUSION) study		V D A	2335	Case-cont
Whole Genome Association Study of Systemic Lupus Erythematosus		V D A	4651	Case-cont


PubChem
108 million substances x
650,000 assays
1 billion points of data
within a grid of
70 trillion cells
BioAssay
Compound
Substance
GO
Advanced search
NCBI Resources How To
PubChem Substance
all[filt]
Save search Limits
Summary, 20 per page, Sorted by Default order
Results: 1 to 20 of 108327716
Cadmium ion; CD
Source: MMDB (105286.3)
NCBI Resources How To
PubChem Compound
Results: 1 to 20 of 32454538
NCBI Resources How To
PubChem BioAssay
all[filt]
Save search Limits
Display Settings: Summary, 20 per page, Sorted by Default order
Results: 1 to 20 of 648590
TBK1 % inhibition at 1 uM [UNC Frye lab]
1. Source: ChEMBL
Protein Target: Serine/threonine-protein kinase TBK1; NF-kappa-B-activating kinase
Compound BioActivity: 366 Tested
All data
AID: 651546
Protein Target Related BioAssays by Target Related BioAssays by Depositor
PIP5K1 (Caliper assay) % inhibition at 5 uM [UNC Frye lab]

browse by disease

— A —

— I —

— R —

browse by disease

A

- Anal Cancer
- Anemia
- Asthma

B

- Bladder Cancer
- Brain Cancer
- Breast Cancer

C

- Carcinoid
- Cervical Cancer
- Chronic Obstructive Pulmonary Disease

I

- Idiopathic Pulmonary Fibrosis

K

- Kidney Cancer

L

- Leukemia
- Liver Cancer
- Lung Cancer

M

- Melanoma
- Monoclonal Gammopathy

R

- Rheumatoid Arthritis

S

- Sarcoidosis
- Scleroderma
- Systemic Lupus Erythematosus

T

- Testicular Cancer

U

- Uterine Cancer

Search Results

You've Selected:

Disease: Leukemia (X)

Clear All Selections

Category

Products (21)

Tissue

Bone Marrow (9)
Peripheral Blood (12)

Cell Type

B Cells CD19 (2)
B Cells Negative Selection (2)
Buffy Coat (1)
CD45 (2)
Fresh (2)
Mononuclear Cells (2)
Plasma (1)
Serum (1)
Special Processing (2)
T Cells CD3 (2)
T Cells Negative Selection (2)
Viable Plated Cells (2)

Units

0.3mL (1)
0.5 million cells (10)
0.5mL (2)
1 unit (2)
5.0 million cells (2)

Price

$0.00 - $1,000.00 (17)
$1,000.00 - $2,000.00 (2)

Leukemia

21 Items Previous | 1 | 2 | Next View as:

15 Items Per Page | Sort By...

bma

Bone Marrow | B Cells, Negative Selection | Leukemia
SKU: BMA-BCE-LE
$500.00

bma

Bone Marrow | B Cells, CD19 | Leukemia
SKU: BMA-CD19-LE
$500.00

bma

Bone Marrow | T Cells, CD3 | Leukemia
SKU: BMA-CD3-LE
$500.00

bma

Bone Marrow | CD45 | Leukemia
SKU: BMA-CD45-LE
$500.00

bma

Bone Marrow | Fresh | Leukemia
SKU: BMA-FRE-LE
$2,500.00

bma

Bone Marrow | Mononuclear Cells | Leukemia
SKU: BMA-MON-LE
$750.00

bma

Bone Marrow | Special Processing | Leukemia
SKU: BMA-SPE-LE
$500.00

bma

Bone Marrow | T Cells, Negative Selection | Leukemia
SKU: BMA-TCE-LE
$500.00

Search Results

You've Selected:

Disease: Leukemia (X)

Clear All Selections

Category

Products (21)

Tissue

Bone Marrow (9)
Peripheral Blood (12)

Cell Type

B Cells CD19 (2)
B Cells Negative Selection (2)
Buffy Coat (1)
CD45 (2)
Fresh (2)
Mononuclear Cells (2)
Plasma (1)
Serum (1)
Special Processing (2)
T Cells CD3 (2)
T Cells Negative Selection (2)
Viable Plated Cells (2)

Units

0.3mL (1)
0.5 million cells (10)
0.5mL (2)
1 unit (2)
5.0 million cells (2)

Price

$0.00 - $1,000.00 (17)

Leukemia

21 Items Previous 1 2 Next View as:

15 Items Per Page | Sort By...

pbl

Peripheral Blood | Mononuclear Cells | Leukemia
SKU: PBL-MON-LE
$500.00

pbl

Peripheral Blood | Plasma | Leukemia
SKU: PBL-PLA-LE

pbl

Peripheral Blood | Serum | Leukemia
SKU: PBL-SER-LE
$55.00

pbl

Peripheral Blood | Special Processing | Leukemia
SKU: PBL-SPE-LE
$500.00

pbl

Peripheral Blood | T Cells, Negative Selection | Leukemia
SKU: PBL-TCE-LE
$600.00

pbl

Peripheral Blood | Viable Plated Cells | Leukemia
SKU: PBL-VPC-LE
$1,000.00

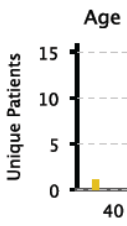
Previous 1 2 Next

Female

Summary of Available Inventory

Gender	500uL Plasma (K2EDTA) Unique Patients	500uL Plasma (K2EDTA) Available Samples	500uL Plasma (Lithium Heparin) Unique Patients	500uL Plasma (Lithium Heparin) Available Samples	300uL Buffy Coat (K2EDTA) Unique Patients	300uL Buffy Coat (K2EDTA) Available Samples
Female						
Male	31	498	11	92	31	210
Total	31	498	11	92	31	210

Male



Available Inventory by Patient ID

Patient ID	Age	Race	Sex	Alcohol Use	Tobacco Use
12416CF6D	73	White	M	Current Use	Previous Use

Drugs Administered
Cyclophosphamide
Dexamethasone Sodium
Docetaxel
Leuprolide Acetate
Palonosetron HCL
Pegfilgrastim
Zoledronic Acid

500uL Plasma (K2EDTA Tubes)	
Catalog Number (Unique Draws)	Available Samples
BBP0500-A112416CF6D011108P4	2
BBP0500-A112416CF6D020108P4	2
BBP0500-A112416CF6D020808P4	3
BBP0500-A112416CF6D021508P4	3
BBP0500-A112416CF6D022908P4	2
BBP0500-A112416CF6D031907P4	3
BBP0500-A112416CF6D032108P4	2
BBP0500-A112416CF6D032907P4	2
BBP0500-A112416CF6D040408P4	3
BBP0500-A112416CF6D041108P4	3
BBP0500-A112416CF6D041808P4	2
BBP0500-A112416CF6D042607P4	1
BBP0500-A112416CF6D050908P4	3
BBP0500-A112416CF6D051608P4	3
BBP0500-A112416CF6D052308P4	3
BBP0500-A112416CF6D071207P4	3
BBP0500-A112416CF6D071907P4	3
BBP0500-A112416CF6D072607P4	3
BBP0500-A112416CF6D080907P4	2
BBP0500-A112416CF6D090707P4	2
BBP0500-A112416CF6D101207P4	2
BBP0500-A112416CF6D110907P4	3
BBP0500-A112416CF6D111607P4	2
BBP0500-A112416CF6D120607P4	3
BBP0500-A112416CF6D121307P4	3
BBP0500-A112416CF6D122007P4	2

500uL Plasma (Lithium Heparin Tubes)	
Catalog Number (Unique Draws)	Available Samples
BBP0500-A112416CF6D011108G4	2
BBP0500-A112416CF6D020108G4	1
BBP0500-A112416CF6D020808G4	3
BBP0500-A112416CF6D031907G4	1
BBP0500-A112416CF6D032907G4	2
BBP0500-A112416CF6D042607G4	2
BBP0500-A112416CF6D053107G4	2
BBP0500-A112416CF6D071207G4	2
BBP0500-A112416CF6D071907G4	3
BBP0500-A112416CF6D072607G4	2
BBP0500-A112416CF6D080907G4	2
BBP0500-A112416CF6D090707G4	2
BBP0500-A112416CF6D101207G4	1
BBP0500-A112416CF6D110907G4	3
BBP0500-A112416CF6D111607G4	3
BBP0500-A112416CF6D120607G4	2
BBP0500-A112416CF6D121307G4	3
BBP0500-A112416CF6D122007G4	2

30
Catalog N
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A
BBB0300-A

pbl

Peripheral Blood | Plasma | Leukemia

SKU: PBL-PLA-LE

$55.00

Be the first to review this product

Quick Overview: 0.5mL Plasma specimen collected in K2EDTA tube and stored in 1.0mL cryovial. Sample stored at -80C and shipped on dry ice.

# of Samples per Patient

-- Please Select --
-- Please Select --
1
2 +$55.00
3 +$110.00
4 +$165.00
5 +$220.00
6 +$275.00
7 +$330.00

Disease subtype

-- Please Select --

Units

-- Please Select --

add to cart

Add to Saved Items | Email A Colleague

Overview › | Product Reviews ›

Additional Information

SKU	PBL-PLA-LE
Treatment Status	Any Treatment, Pre Treatment, Post Treatment, Active Treatment, Recurrent/Refractory Disease, Remission

assay depot

the marketplace for pharmaceutical research services

All the Services you need

Researchers around the world use Assay Depot to find the latest research services

Take the Tour

Search... Search or click on a button below to browse research services

biology chemistry dmpk pharmacology toxicology

assay depot
the marketplace for pharmaceutical research services

biology | chemistry | dmpk | pharmacology | toxicology

Bone Models

Bone Metastases
Osteoarthritis
Osteoporosis

Cardiovascular Models

Atrial Arrhythmias
Coronary Artery Disease
Hypertension
Ischemia
Myocardial Infarction
Restenosis
Ventricular Tachycardia

Dermatology Models

Acne
Atopic Dermatitis
Hair Growth
Lupus
Psoriasis
Rosacea
Skin Graft
Wound Healing

Diabetes Models

BB/W Rats
Food Intake
Goto-Kakizaki Rats
Non Obese Diabetic Mice
Obese Mice
Primate Diabetes
Streptozotocin Mice
Streptozotocin Rats
More...

Genitourinary Models

Chronic Kidney Disease
Cystometry
Endometriosis
IGA Glomerulonephritis
Interstitial Cystitis
Spinalized Rats

In Vitro Models

In Vitro Bone Models
In Vitro CVD Models
In Vitro Diabetes Models
In Vitro Eye Models
In Vitro Oncology Models
In Vitro Skin Models

In Vivo Technologies

Cognition
EEG
Electrophysiology
Imaging
Microdialysis

Infectious Disease

Bacterial Infection
Dengue Virus
Hepatitis C Virus
Influenza
LCMV Mouse
Malaria

Inflammation Models

Arthritis
Delayed Type Hypersens
Edema
Hemophilia
Irritable Bowel Disease
Irritant
LPS Acute Response
Mucositis
More...

Neurological Models

Alzheimer's Disease
Anxiety
Behavioral Tests
Cerebral Palsy
Circadian Profiling
Depression
Epilepsy
Olfactory Testing
More...

Oncology Models

Angiogenesis
Cachexia

Ophthalmic Models

Cataract
Corneal Dystrophy

Otology Models

Hearing Loss
Meniere's Disease

Pain Models

General Pain
Inflammatory Pain

Respiratory Models

Ascaris Lung Allergy
Cough

ob/ob Diabetes Model - 16 Mice

Service Description

Provider: Links Biosciences is a US company with laboratories in Hangzhou, China. The laboratory has been offering exploratory (non-GLP) pharmacology services to US and Chinese biopharma since 2004.

Background: The obese mutant mouse model was first reported by Ingalls A *et al* from the Jackson Laboratory in 1951 (Obese, a New Mutation in the House Mouse [164 KB]). The obese mouse resulted from a spontaneous mutation in a gene that was named *ob* in the V stock. Mice homozygous for the obese spontaneous mutation, (Lep^ob^; commonly referred to as *ob* or *ob/ob*), are first recognizable at about 4 weeks of age. Homozygous mutant mice gain weight rapidly and may reach three times the weight of wild-type controls. In addition to obesity, mutant mice exhibit hyperphagia, a diabetes-like syndrome of hyperglycemia, glucose intolerance, elevated plasma insulin, subfertility, impaired wound healing, and an increase in hormone production from both pituitary and adrenal glands. Friedman J *et al* reported leptin in 1994, and demonstrated that leptin, the product of the *ob* gene, was produced in white adipose tissue and served as the peripheral signal to the central nervous system of nutritional status.

Service Details: This service offers a 28 day db/db mouse model of T2DM and obesity. Customer has various options that are conveyed to Links Biosciences using a Service Order Form. Customer assigns up to 16 mice to

assay depot

biology | chemistry | dmpk | pharmacology | toxicology

Home » Pharmacology » Diabetes and Obesity » Obese Mice

ob/ob Diabetes Model - 16 Mice

$9,000.00 USD
per service

9 week
turn around time

Provided By
Links Biosciences

Request Info

Add to Cart

SHARE

Service Description

Provider: Links Biosciences is a US company with laboratories in Hangzhou, China. The laboratory has been offering exploratory (non-GLP) pharmacology services to US and Chinese biopharma since 2004.

Background: The obese mutant mouse model was first reported by Ingalls A *et al* from the Jackson Laboratory in 1951 (Obese, a New Mutation in the House Mouse [164 KB]). The obese mouse resulted from a spontaneous mutation in a gene that was named *ob* in the V stock. Mice homozygous for the obese spontaneous mutation, (Lep^ob^; commonly referred to as *ob* or *ob/ob*), are first recognizable at about 4 weeks of age. Homozygous mutant mice gain weight rapidly and may reach three times the weight of wild-type controls. In addition to obesity, mutant mice exhibit hyperphagia, a diabetes-like syndrome of hyperglycemia, glucose intolerance, elevated plasma insulin, subfertility, impaired wound healing, and an increase in hormone production from both pituitary and adrenal glands. Friedman J *et al* reported leptin in 1994, and demonstrated that leptin, the product of the *ob* gene, was produced in white adipose tissue and served as the peripheral signal to the central nervous system of nutritional status.

Service Details: This service offers a 28 day db/db mouse model of T2DM and obesity. Customer has various options that are conveyed to Links Biosciences using a Service Order Form. Customer assigns up to 16 mice to

Validation methods are increasingly commoditized

Scroll down to browse a list of available research models for **Type I and Type II diabetes, hyperglycemia, insulin resistance, diet-induced obesity and related diseases**. Use the filters on the left to refine the list and then click on any listing to view technical information or to ask a question.

Click on the Vendors tab to view a complete list of CROs that offer diabetes and obesity pharmacology models.

VIEW SERVICES | VIEW VENDORS | **133 results** get help

Search Filters

Diabetes and Obesity

- BB/W Rats
- Food Intake
- Goto-Kakizaki Rats
- Non Obese Diabetic Mice
- Obese Mice
- Obese Primates
- Primate Diabetes
- Streptozotocin Mice
- Streptozotocin Rats
- db/db Diabetic Mice
- fa/fa Zucker Diabetic Rats

Certifications help

- GLP (48)
- AAALAC (28)
- GMP (20)
- ISO 9001 (7)
- GCP (7)
- FDA (5)
- USDA (4)

more

Locations

United States (64)

Univ. of Maryland School of Medicine Obesity and Diabetes Research Center

University of Maryland School of Medicine Obesity and Diabetes Research Center focuses on research of obesity, diabetes, and aging in nonhuman primates.

vendor info | Add

Transgenic Rabbit Models

Transgenic Rabbit Models offers transgenic rabbit models for the study of atherosclerosis, ophtalmology, hypertrophic myopathies, diabetes, obesity, hemostasis, respiratory diseases, AIDS, and cancer.

vendor info | Add

Ophthy-DS

Ophthy-DS offers ophthalmic model services for macular degeneration, diabetes, uveitis, and dry eye.

vendor info | Add

PharmaNess

PharmaNess offers pharmacokinetics, pharmacodynamics, formulations, behavioral assay, in vivo screening, ex vivo screening, microscopy, stereology and histology staining services.

vendor info | Add

Wisconsin National Primate Research Center

Wisconsin National Primate Research Center focuses on research of regenerative medicine, reproduction, immunology, virology, aging, and metabolic diseases.

vendor info | Add

Selected Vendors

Scroll down to browse a list of available research models for **Type I and Type II diabetes, hyperglycemia, insulin resistance, diet-induced obesity and related diseases**. Use the filters on the left to refine the list and then click on any listing to view technical information or to ask a question.

Click on the Vendors tab to view a complete list of CROs that offer diabetes and obesity pharmacology models.

VIEW SERVICES | VIEW VENDORS | 133 results | get help

Search Filters

Diabetes and Obesity

- BB/W Rats
- Food Intake
- Goto-Kakizaki Rats
- Non Obese Diabetic Mice
- Obese Mice
- Obese Primates
- Primate Diabetes
- Streptozotocin Mice
- Streptozotocin Rats
- db/db Diabetic Mice
- fa/fa Zucker Diabetic Rats

Certifications help

- GLP (48)
- AAALAC (28)
- GMP (20)
- ISO 9001 (7)
- GCP (7)
- FDA (5)
- USDA (4)

more

Locations

United States (64)

Univ. of Maryland School of Medicine Obesity and Diabetes Research Center

University of Maryland School of Medicine Obesity and Diabetes Research Center focuses on research of obesity, diabetes, and aging in nonhuman primates.

vendor info | Add

Transgenic Rabbit Models

Transgenic Rabbit Models offers transgenic rabbit models for the study of atherosclerosis, ophtalmology, hypertrophic myopathies, diabetes, obesity, hemostasis, respiratory diseases, AIDS, and cancer.

vendor info | Add

Ophthy-DS

Ophthy-DS offers ophthalmic model services for macular degeneration, diabetes, uveitis, and dry eye.

vendor info | Add

PharmaNess

PharmaNess offers pharmacokinetics, pharmacodynamics, formulations, behavioral assay, in vivo screening, ex vivo screening, microscopy, stereology and histology staining services.

vendor info | Add

Wisconsin National Primate Research Center

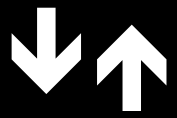
Wisconsin National Primate Research Center focuses on research of regenerative medicine, reproduction, immunology, virology, aging, and metabolic diseases.

vendor info | Add

Selected Vendors

Translational Pipeline

Clinical and Molecular Measurements



Translational Question or Trial

Statistical/Computational methods

Validating drug or biomarker

Translational Pipeline

Clinical and Molecular Measurements

Commodity

↓↑

Translational Question or Trial

↓

Statistical/Computational methods

↓

Validating drug or biomarker

Commodity

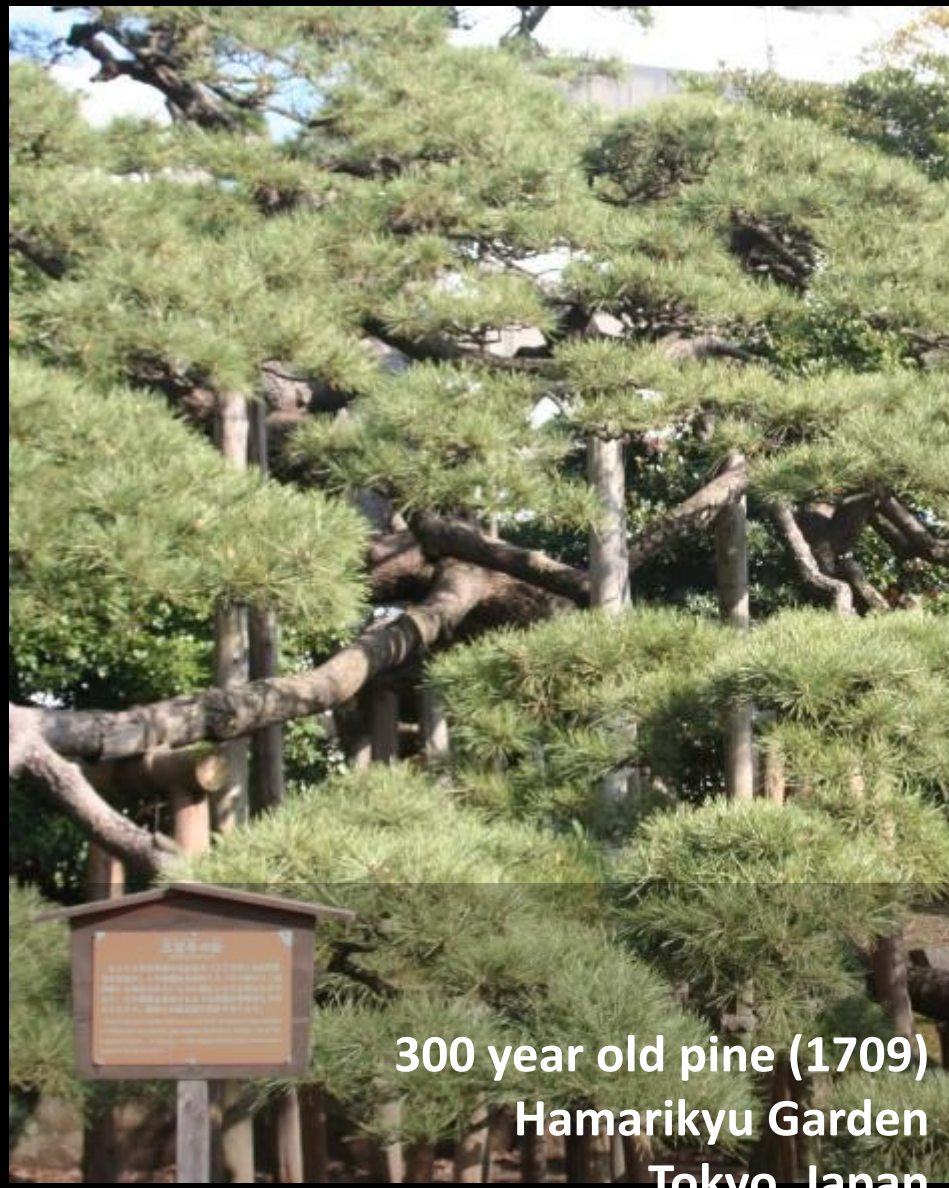
We are used to starting computer, IT, and Internet companies in garages...

We are used to starting computer, IT, and Internet companies in garages...

Potentials for starting a "garage biotech"?

Trees in Biomedicine

- Linnaeus 1707-1778
- Promoted binomial nomenclature for taxonomy
  - *Homo sapiens*, *Mus musculus*
- But 300 year old trees need crutches!
- The species taxonomy is commonly rearranged based on DNA
  - *Pneumocystis jiroveci* and *Pneumocystis carinii*



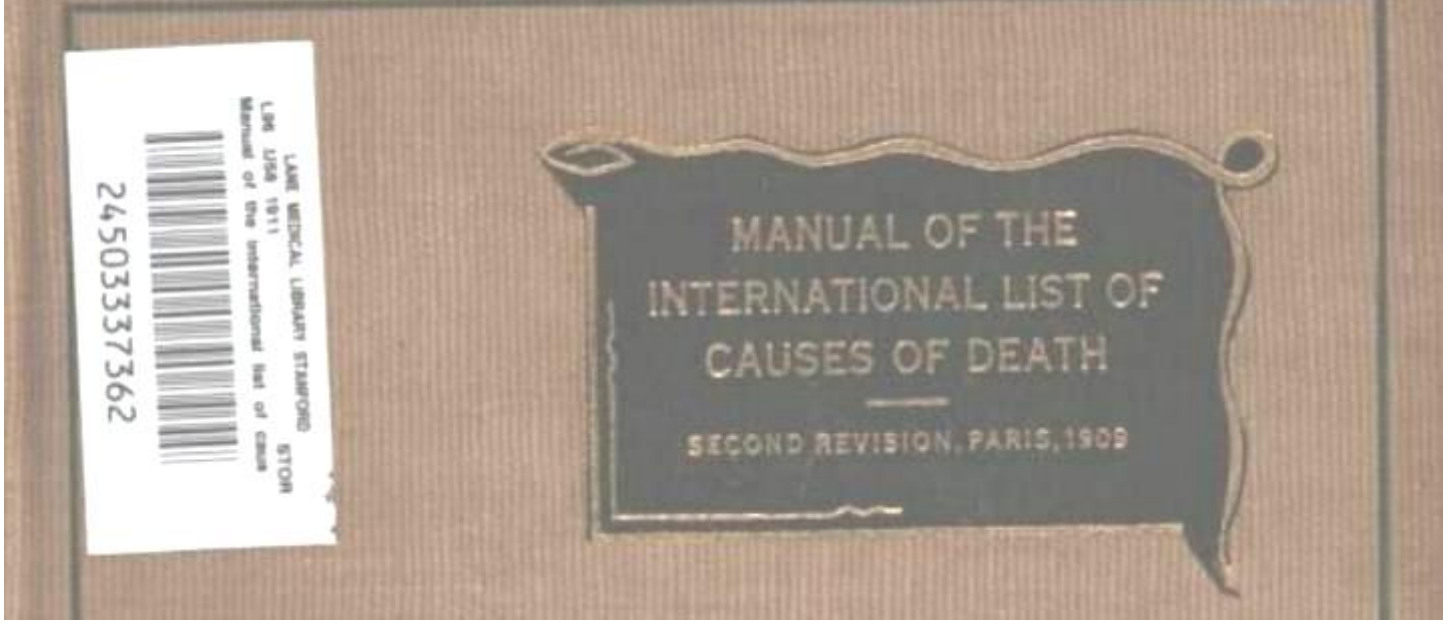
300 year old pine (1709)
Hamarikyu Garden
Tokyo, Japan

Trees of disease: Nosology

- Linnaeus also co-founder of systematic nosology
  - Nosology = classification of disease
  - *Genera Morborum* (1763)
- Why not classify diseases based on genomics?
  - Could reshuffle thinking about diseases and drugs
  - Public molecular data: 1 million+ microarrays, grows 2-3x/yr

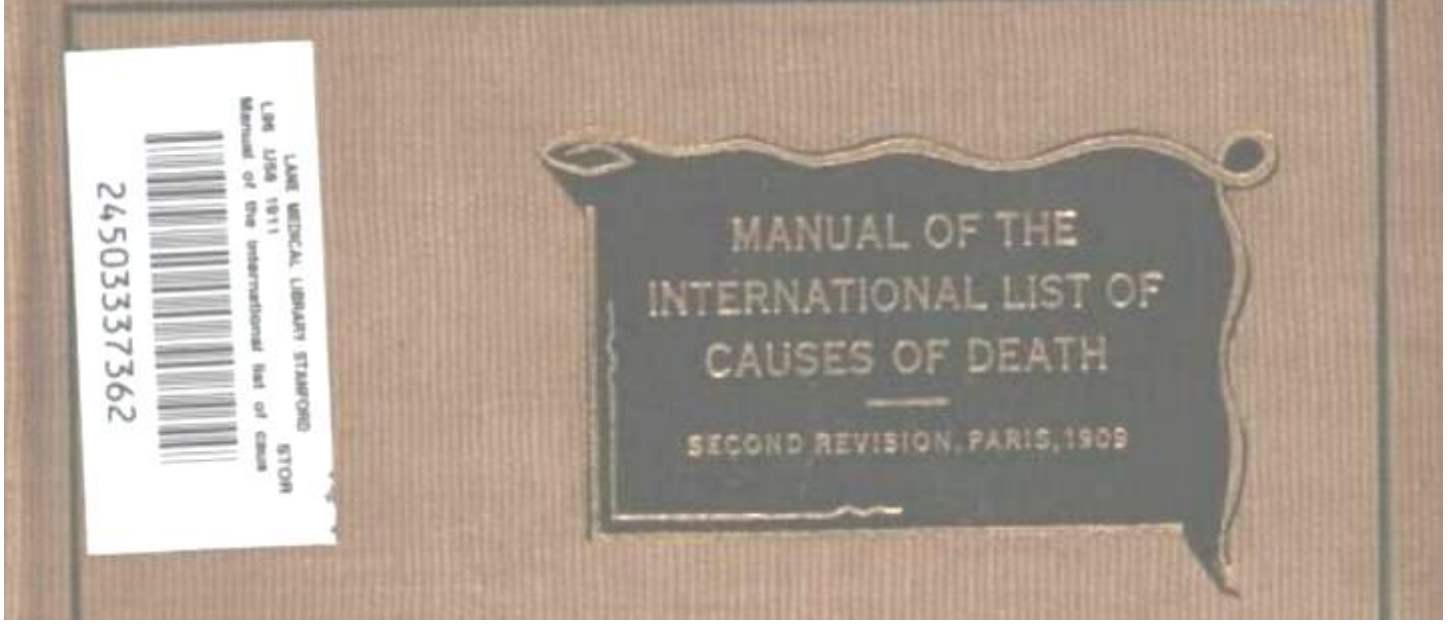
Exanthematic	Feverish, with skin eruptions
Critical	Feverish, with urinary problems
Phlogistic	Feverish, with heavy pulse and topical pain
Dolorous	Painful
Mental	With alienation of judgment
Quietal	With loss of movement
Motor	With involuntary motion
Suppressorial	With impeded motions
Evacuatorial	With evacuation of liquids
Deformities	Changed appearance of solid parts
Blemishes	External and palpable

Bramley M. Coding Matters 2001, 8:1.

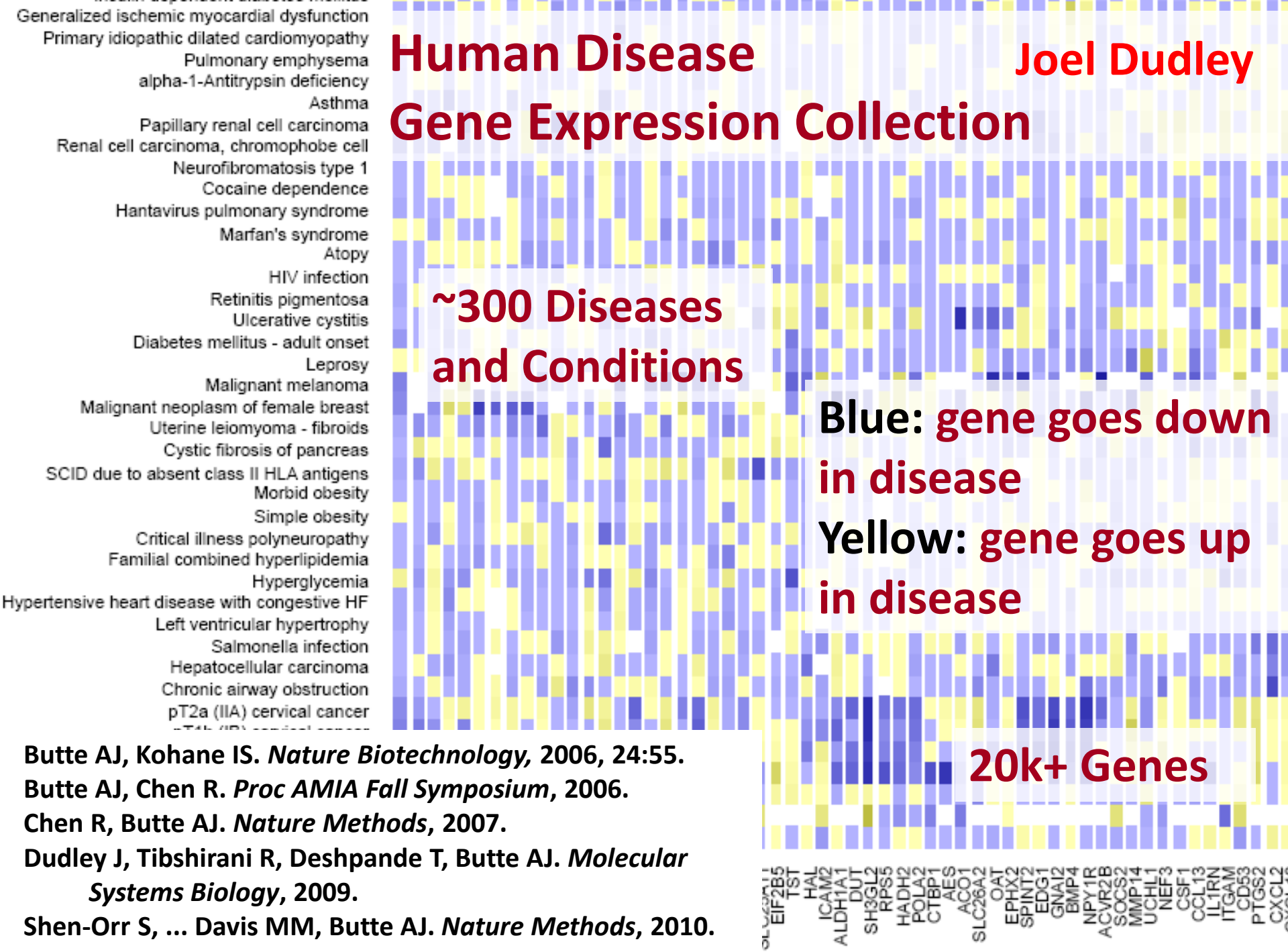


- 39 Cancer of the buccal cavity
- 40 Cancer of stomach and liver
- 41 Cancer of peritoneum, intestines, rectum
- 42 Cancer of female genital organs
- 43 Cancer of breast
- 44 Cancer of skin
- 45 Cancer of other organs or not specified

Lung is an "other organ“; Brain is an "other organ"



- 50 Diabetes
  - *No type 1 or type 2*
- *Endocrine diseases were under General Diseases*
- 88 Disease of the thyroid body
  - *Under Disease of the Respiratory System*
- 5 Smallpox, 13 Cholera, 15 Plague, 21 Glanders, 22 Anthrax
  - *All bioterroristic today*
- 189 Visitation from God



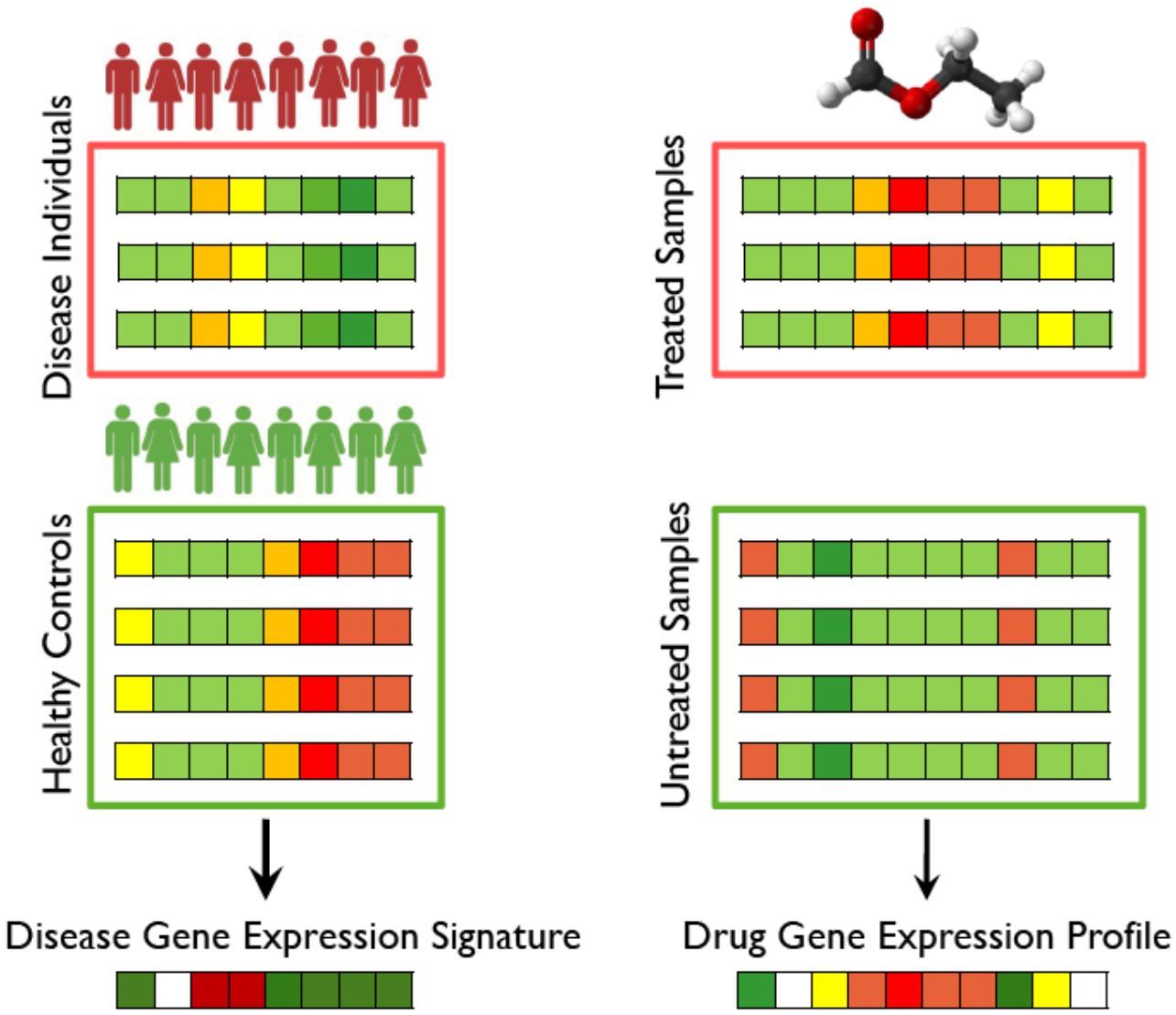
Butte AJ, Kohane IS. *Nature Biotechnology,* 2006, 24:55.

Butte AJ, Chen R. *Proc AMIA Fall Symposium*, 2006.

Chen R, Butte AJ. *Nature Methods*, 2007.

Dudley J, Tibshirani R, Deshpande T, Butte AJ. *Molecular Systems Biology*, 2009.

Shen-Orr S, ... Davis MM, Butte AJ. *Nature Methods*, 2010.

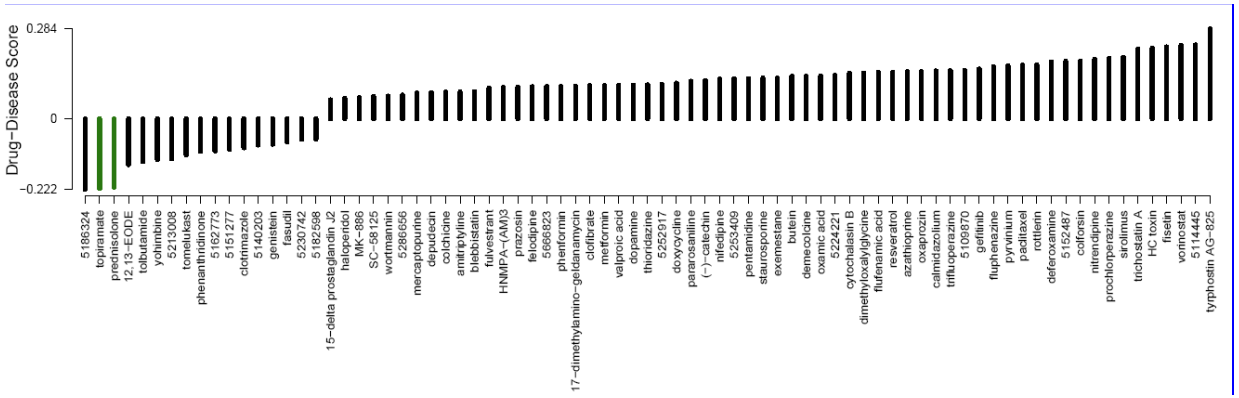


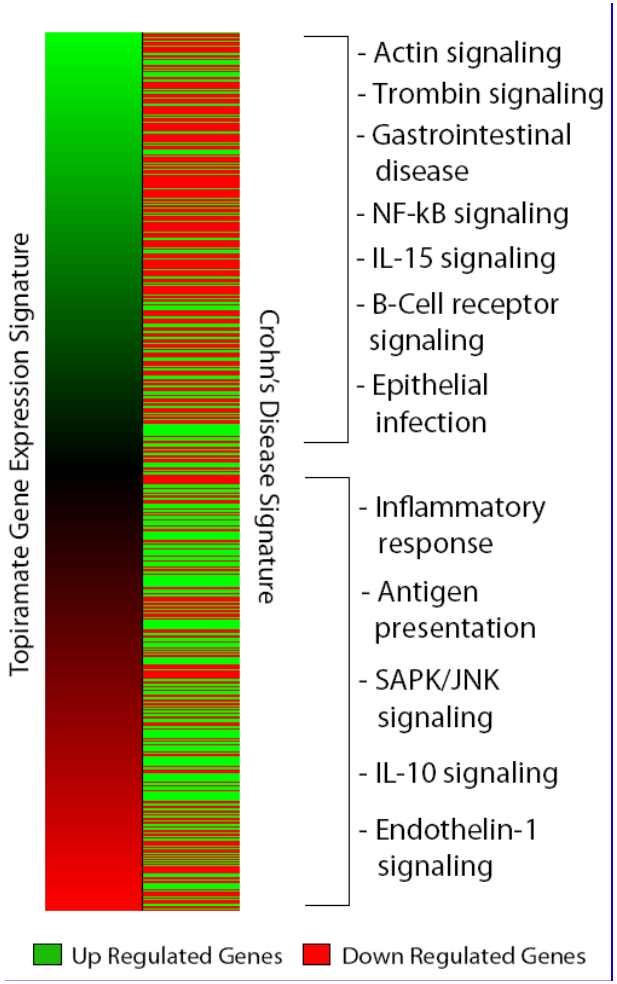
Marina Sirota
Joel Dudley

Lamb J, ..., Golub TR. *Science*, 2006.
Sirota M, Dudley JT, ..., Sweet-Cordero A, Sage J, Butte AJ. *Science Translational Medicine*, 2011.

Candidate anti-seizure drug against inflammatory bowel disease

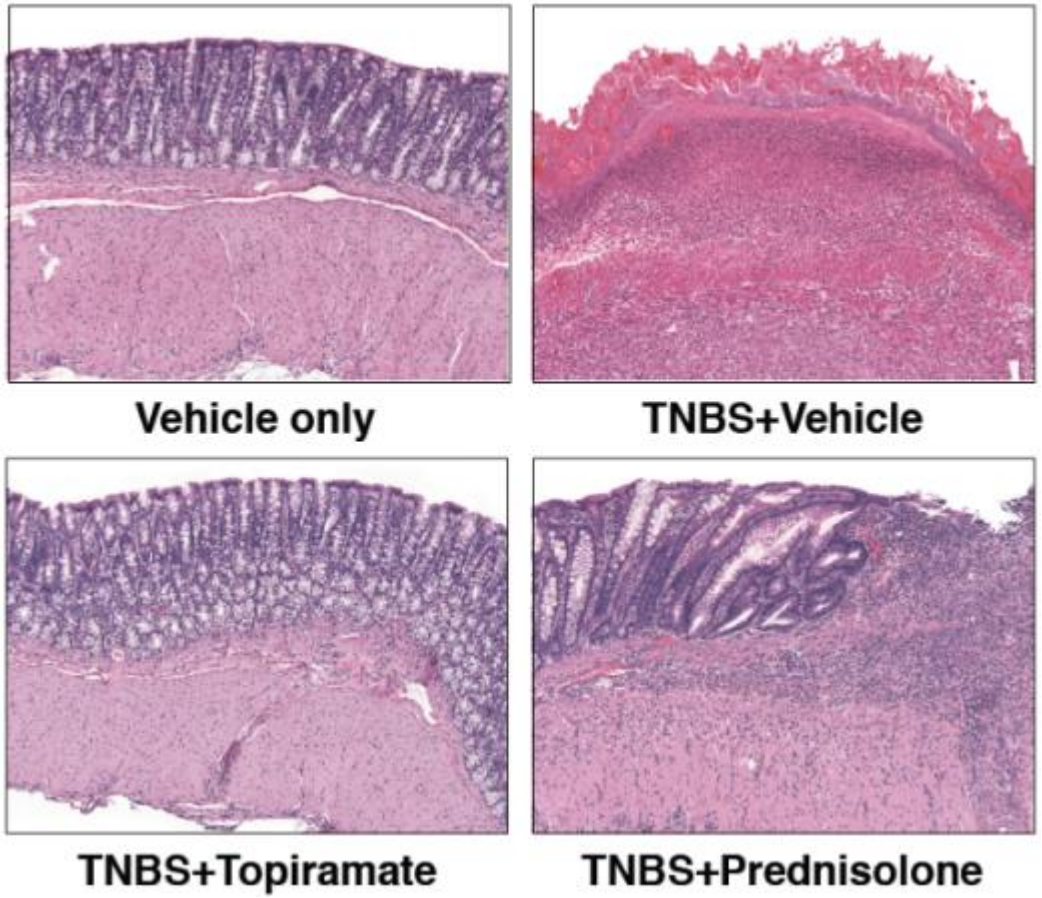
Marina Sirota
Joel Dudley

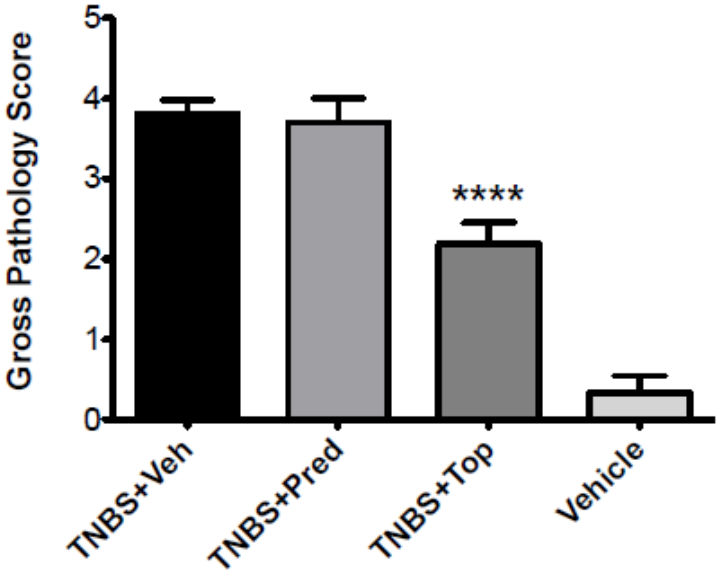




Sirota M, Dudley JT, ..., Sweet-Cordero A, Sage J, Butte AJ.
***Science Translational Medicine*, 2011.**

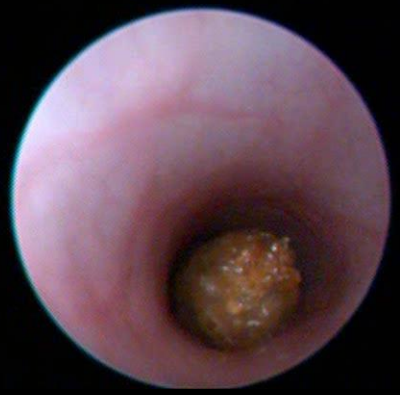
Anti-seizure drug works against a rat model of inflammatory bowel disease



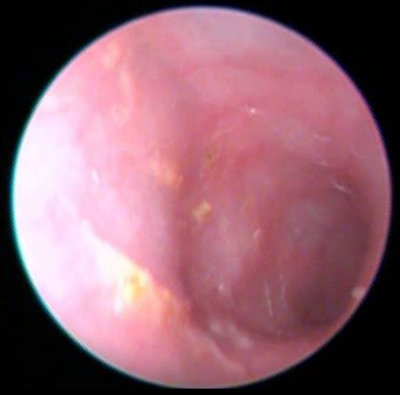


Marina Sirota
Joel Dudley
Mohan M Shenoy
Jay Pasricha

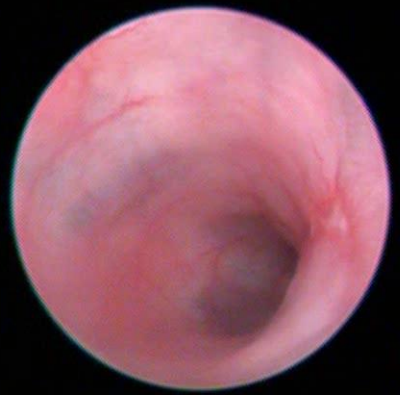
Dudley JT, Sirota M, ..., Pasricha J, Butte AJ. *Science Translational Medicine*, 2011.



Rat colonoscopy



Rat with Inflammatory Bowel Disease

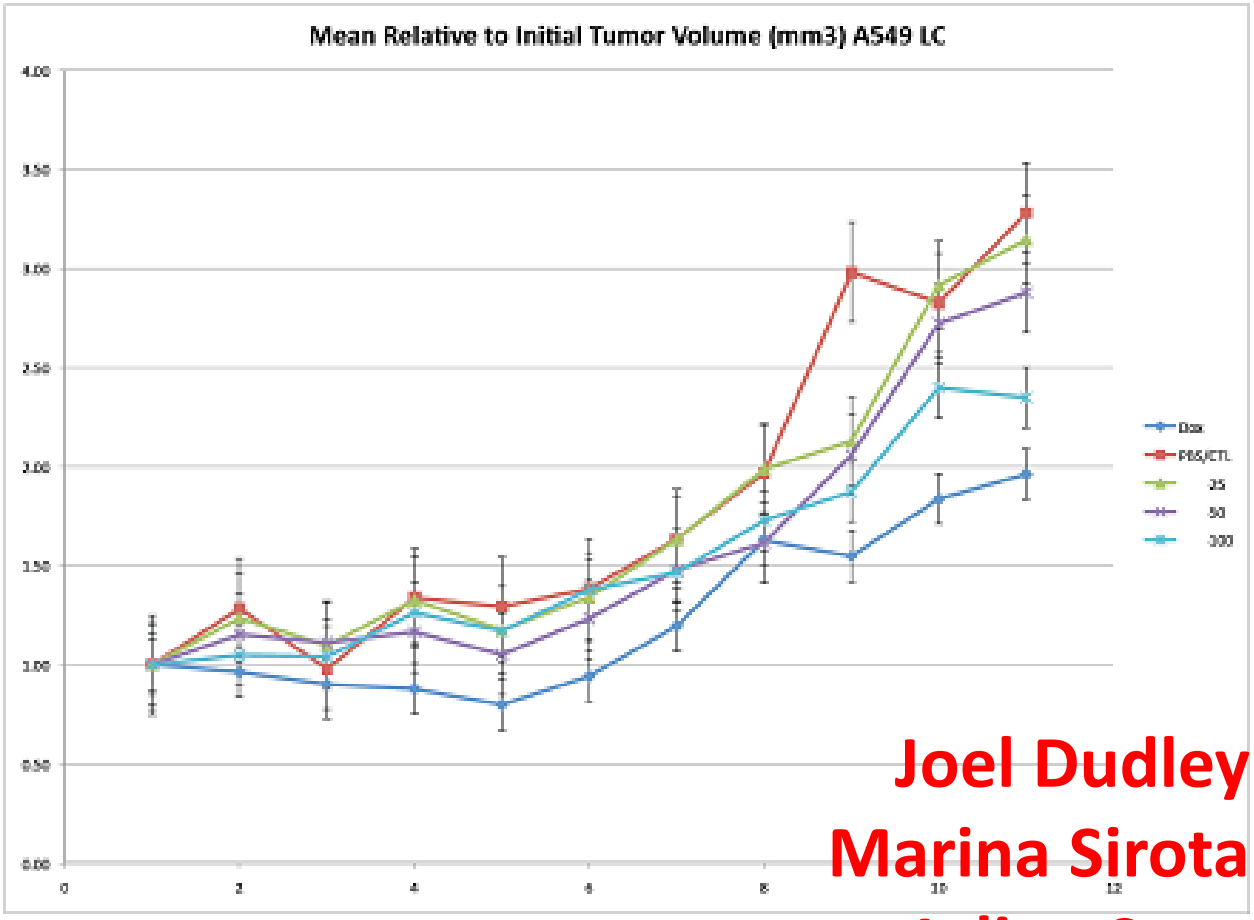


Inflammatory Bowel Disease After Anti-seizure Drug

Dudley JT, Sirota M, ..., Pasricha J, Butte AJ. *Science Translational Medicine*, 2011.

Anti-ulcer drug works for lung adenocarcinoma

- Human lung adenocarcinoma cell lines explanted into mouse models
- Followed growth 11 days
- Positive-control doxorubicin grew to 2x original volume
- Tumors in mice treated with vehicle grew to 3.25x original volume
- Not only did our compound work statistically better than control, it worked in a dose-dependent manner
- Tumors in mice treated with 50 mg/kg/day grew 2.8x
- Those treated with 100 mg/kg/day grew only 2.3x.



Joel Dudley
Marina Sirota
Julien Sage

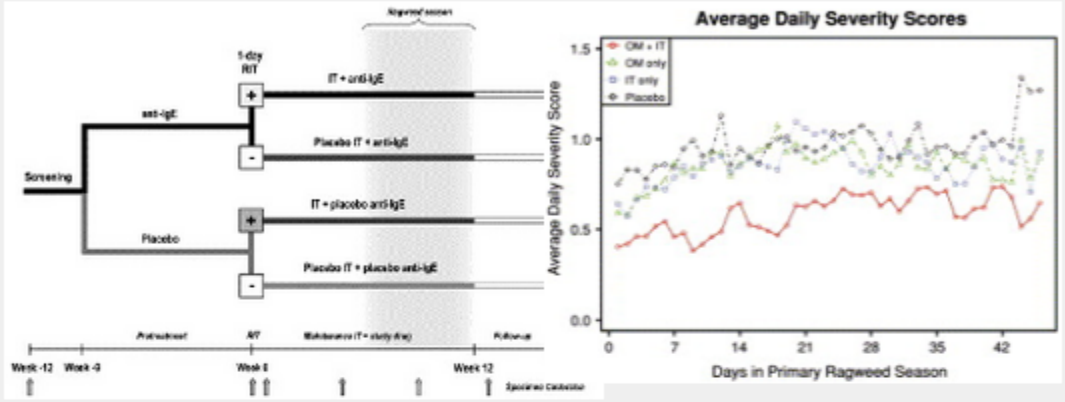
Sirota M, Dudley JT,..., Sage J, Butte AJ. *Science Translational Medicine*, 2011,

Study: Efficacy and Safety Evaluation of Allergen Immunotherapy Co-Administered with Omalizumab

Combination treatment with omalizumab (recombinant humanized monoclonal anti-IgE antibody) and rush immunotherapy (RIT) for ragweed-induced allergic rhinitis. Omalizumab pretreatment enhances the safety of RIT for ragweed allergic rhinitis. The combination of ragweed immunotherapy and anti-IgE resulted in prolonged inhibition of allergen-IgE binding compared with either treatment alone.

PubMed ID: 16387596

DOWNLOAD DATA PACKAGE



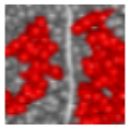
Flow Cytometry Analysis (FLOCK)

Flow cytometry analysis component includes:

- Automated cell population identification
- Result visualization in 2D and 3D
- Statistical analysis of population characteristics
- Automated mapping of populations across multiple samples

flock
Flow cytometry clustering without K

MHC Validation and Analysis



MHC Sequence Feature Variant Type (SFVT) Analysis enables genetic association analysis of classical HLA protein sub-regions defined with structural (e.g. helix) and functional (e.g. binding site) information.

MHC Alleles

Complete DNA and protein sequences, sequence features, and population frequencies of MHC, MIC and TAP alleles. Align MHC sequences horizontally to visualize extent of polymorphisms across all alleles in a locus.

Research Programs

	Study Title	PI	Type of Ex...	Public Rel...

Data Summary

Studies	32
Subjects	11351
Biological Samples	140949
Experiments	145
ELISA Results	82558

Supported NIAID programs

The BISC provides bioinformatics support to the following DAIT-funded networks and research consortia (participating centers); in the future additional networks and/or consortia may be added or current networks and/or consortia removed to reflect changing research priorities of the Institute:

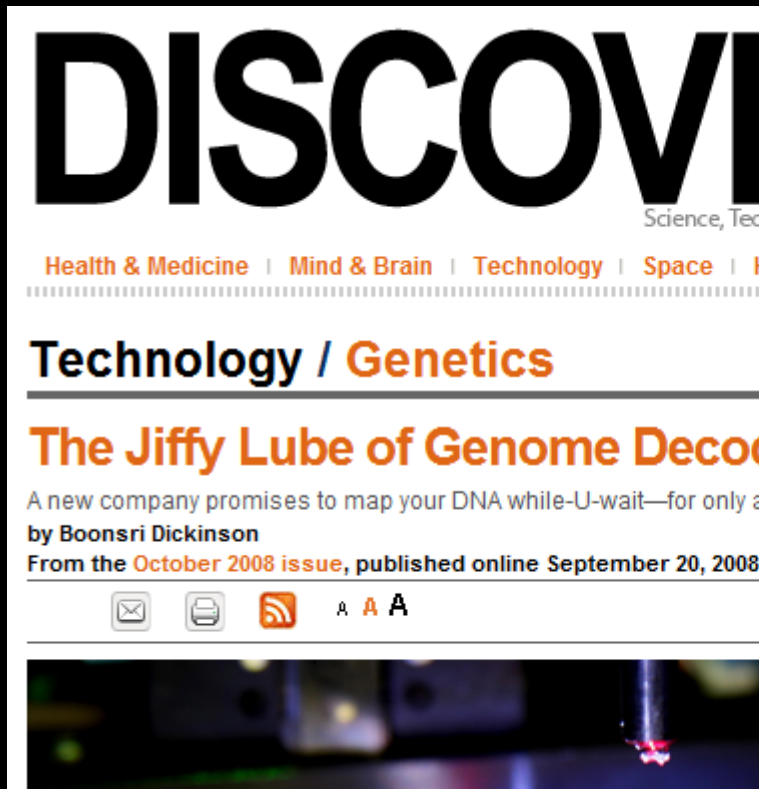
- Collaborative Network for Clinical Research on Immune Tolerance Network
- Atopic Dermatitis Research Network (ADRN)
- Clinical Trials in Organ Transplantation (CTOT)
- Clinical Trials in Organ Transplantation in Children (CTOT-C)
- Population Genetics Analysis Program
- Protective Immunity for Special Populations
- HLA Region Genomics in Immune-mediated Diseases
- Maintenance of Macaque Specific Pathogen-Free Breeding Colonies
- Modeling Immunity for Biodefense
- Reagent Development for Toll-like and other Innate Immune Receptors
- Adjuvant Development Program
- Innate Immune Receptors and Adjuvant Discovery Program
- Human Immunology Project Consortium
- Non-human Primate Transplantation Tolerance Cooperative Study Group

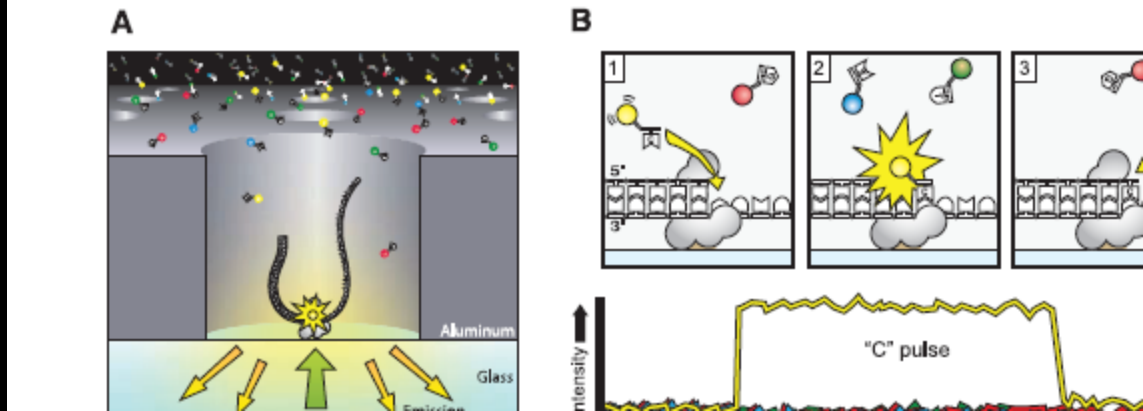
Public release of raw individual-level clinical trials data

- Reproducibility
- Transparency
- Enable learning
- Return data to the community
- New science
- Enable new ventures

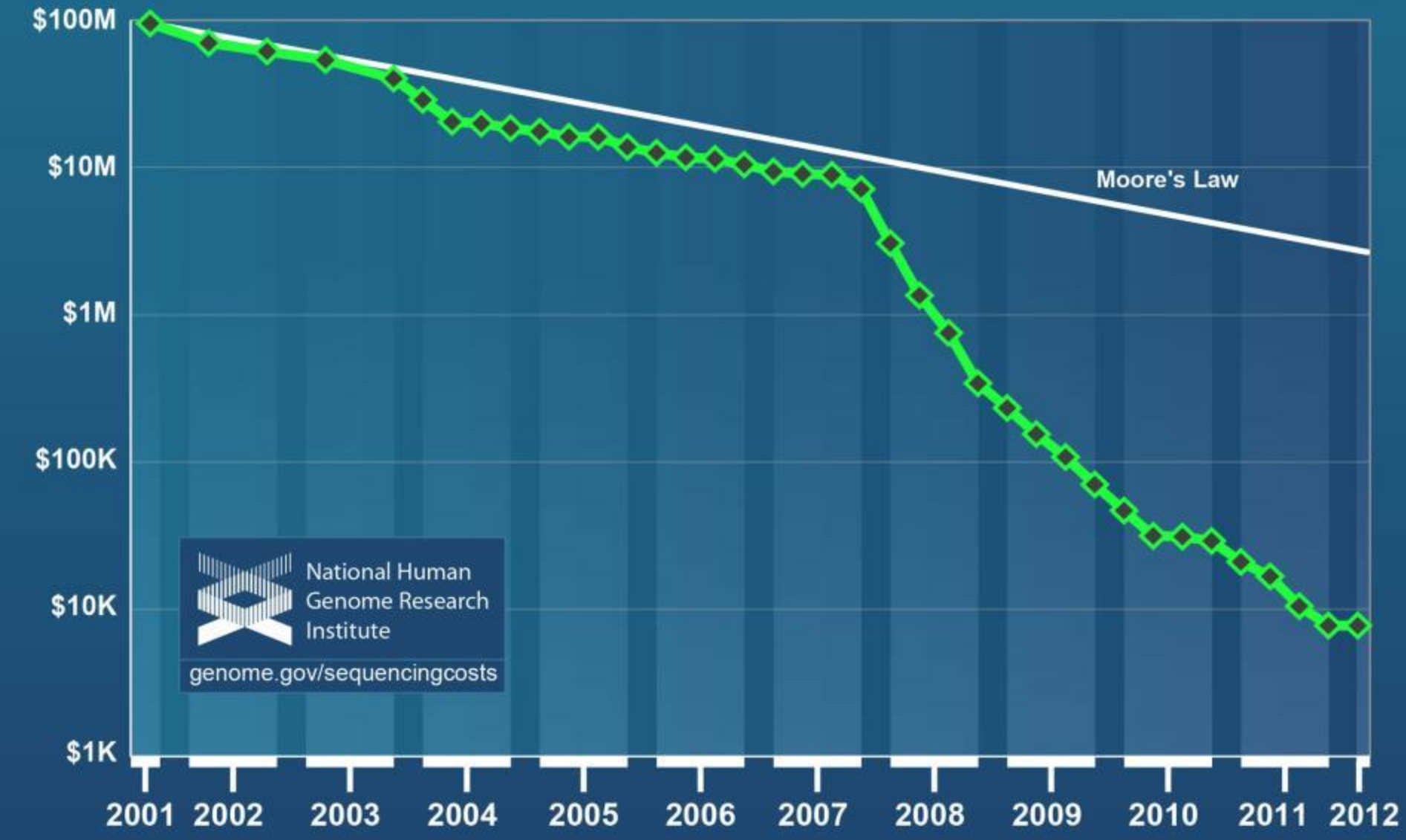
Sequencing Excitement

- 454/Roche, Life Technologies
- Helicos: $30k genome
- Pacific Biosystems: sequence human genome in 15 minutes
- Run times in minutes at a cost of hundreds of dollars
- 20 TB in 15 minutes
- $~1000 genomes: Illumina, Ion Torrent
- Complete Genomics: towards 80 genomes/day


DISCOVE
Science, Tec
Health & Medicine | Mind & Brain | Technology | Space |
Technology / Genetics
The Jiffy Lube of Genome Deco
A new company promises to map your DNA while-U-wait—for only a
by Boonsri Dickinson
From the October 2008 issue, published online September 20, 2008
A A A



Cost per Genome



September 28, 2011

How Low Can We Go? Molecules, Photons, and Bits

Photons. The cost of photons is the cost of the optical and fluidic instrument designed to generate and capture photons from the fluorescent molecules. We can reduce the instrument cost per genome by successfully using more, faster cameras. Our current instruments are equipped with two electron multiplying charge coupled device (EMCCD) cameras. There is a new generation of fast complementary metal oxide semiconductor (CMOS) cameras, developed for other industries that are about 15 times faster than our current cameras (and also less expensive). New sequencing instruments that successfully use four of these fast new cameras could reduce the instrument cost per genome by about a factor of 30, from < $1,000 to $1,000/(2 x 15) or approximately $33 per genome.

Complete genomics

Revolutionizing human genome discovery

About Us | Services | Data & Analysis | Knowledge Center | Support & Community

Home » Data & Analysis » Downloads

Sample Sequence Data

Complete Genomics has recently made several complete human genome data sets available. The genomes were sequenced at the Complete Genomics commercial genome sequencing center in Mountain View, California as part our Complete Genomics Analysis Service (CGA™ Service). These data are largely consistent with the quality and attributes of other data provided to Complete Genomics customers.

When using these data in your research please cite the Complete Genomics website and our publication "Human Genome Sequencing Using Unchained Base Reads on Self-assembling DNA Nanoarrays." Science 1 January 2010 Vol. 327, no. 5961, pp. 78 - 81 DOI: 10.1126/science.1181498

69 Genome Data Set | Documentation

Overview

Complete Genomics is releasing a set of public genome sequences on its FTP server (ftp2.completegenomics.com). There are four sets of data: a Yoruba trio; a Puerto Rican trio; a 17-member 3-generation pedigree; and a diversity panel representing 9 different populations. The CEPH samples within

nature biotechnology

LETTERS

Published online August 10, 2009

Single-molecule sequencing of an individual human genome

Dmitry Pushkarev[1,2], Norma F Neff[1,2] & Stephen R Quake[1]

Recent advances in high-throughput DNA sequencing technologies have enabled order-of-magnitude improvements in both cost and throughput. Here we report the use of single-molecule methods to sequence an individual human genome. We aligned billions of 24- to 70-bp reads (32 bp average) to ~90% of the National Center for Biotechnology Information (NCBI) reference genome, with 28× average coverage. Our results were obtained on one sequencing instrument by a single operator with four data collection runs. Single-molecule sequencing enabled analysis of human genomic information without the need for cloning, amplification or ligation. We determined ~2.8 million single nucleotide polymorphisms (SNPs) with a false-positive rate of less than 1% as validated by Sanger sequencing and 99.8% concordance with SNP genotyping arrays. We identified 752 regions of copy number variation by analyzing coverage depth alone and validated 27 of these using digital PCR. This milestone should allow widespread application of genome sequencing to many aspects of genetics and human health, including personal genomics.

on a surface can be extended asynchronously, thereby allowing substantial flexibility in the kinetics of sequencing chemistry. Previous reports of single-molecule sequencing have been proofs of principle[11–13], and their sequencing throughput has not been competitive with alternative approaches. Generally, read lengths have been relatively short and error rates have been dominated by deletions; it has not been clear whether the resulting sequence quality is suitable for human genome sequencing applications.

The Heliscope Single Molecule Sequencer (Helicos Biosciences) is the first commercial release of a single-molecule sequencing instrument. It allows one to follow ~1 billion individual molecules as they are sequenced over the course of a week—a throughput that is practical for human genome sequencing. There have been several technical improvements to the platform since the reported sequencing of a viral genome[12], including more than a 1,000-fold improvement in parallelism, a new generation of sequencing reagents that allows digital measurement of homopolymer sequences, and a new software algorithm, IndexDP, for performing alignments to the entire human genome.

We used two of the instrument's 50 flow-cell channels to resequence

Clinical assessment incorporating a personal genome

Euan A Ashley, Atul J Butte, Matthew T Wheeler, Rong Chen, Teri E Klein, Frederick E Dewey, Joel T Dudley, Kelly E Ormond, Aleksandra Pavlovic, Alexander A Morgan, Dmitry Pushkarev, Norma F Neff, Louanne Hudgins, Li Gong, Laura M Hodges, Dorit S Berlin, Caroline F Thorn, Katrin Sangkuhl, Joan M Hebert, Mark Woon, Hersh Sagreiya, Ryan Whaley, Joshua W Knowles, Michael F Chou, Joseph V Thakuria, Abraham M Rosenbaum, Alexander Wait Zaranek, George M Church, Henry T Greely, Stephen R Quake, Russ B Altman

Summary

Background The cost of genomic information has fallen steeply, but the clinical translation of genetic risk estimates remains unclear. We aimed to undertake an integrated analysis of a complete human genome in a clinical context.

Methods We assessed a patient with a family history of vascular disease and early sudden death. Clinical assessment included analysis of this patient's full genome sequence, risk prediction for coronary artery disease, screening for causes of sudden cardiac death, and genetic counselling. Genetic analysis included the development of novel methods for the integration of whole genome and clinical risk. Disease and risk analysis focused on prediction of genetic risk of variants associated with mendelian disease, recognised drug responses, and pathogenicity for novel variants. We queried disease-specific mutation databases and pharmacogenomics databases to identify genes and mutations with known associations with disease and drug response. We estimated post-test probabilities of disease by applying likelihood ratios derived from integration of multiple common variants to age-appropriate and sex-appropriate pre-test probabilities. We also accounted for gene-environment interactions and conditionally dependent risks.

***Lancet*, 375:1525, May 1, 2010.**

Patient zero

- 40 year old male in good health presents to his doctor with his whole genome
- No symptoms
- Exercises regularly
- Takes no medications
- Family history of aortic aneurysm
- Family history of sudden death

- Presents with 2.8 million SNPs
- 752 copy number variants

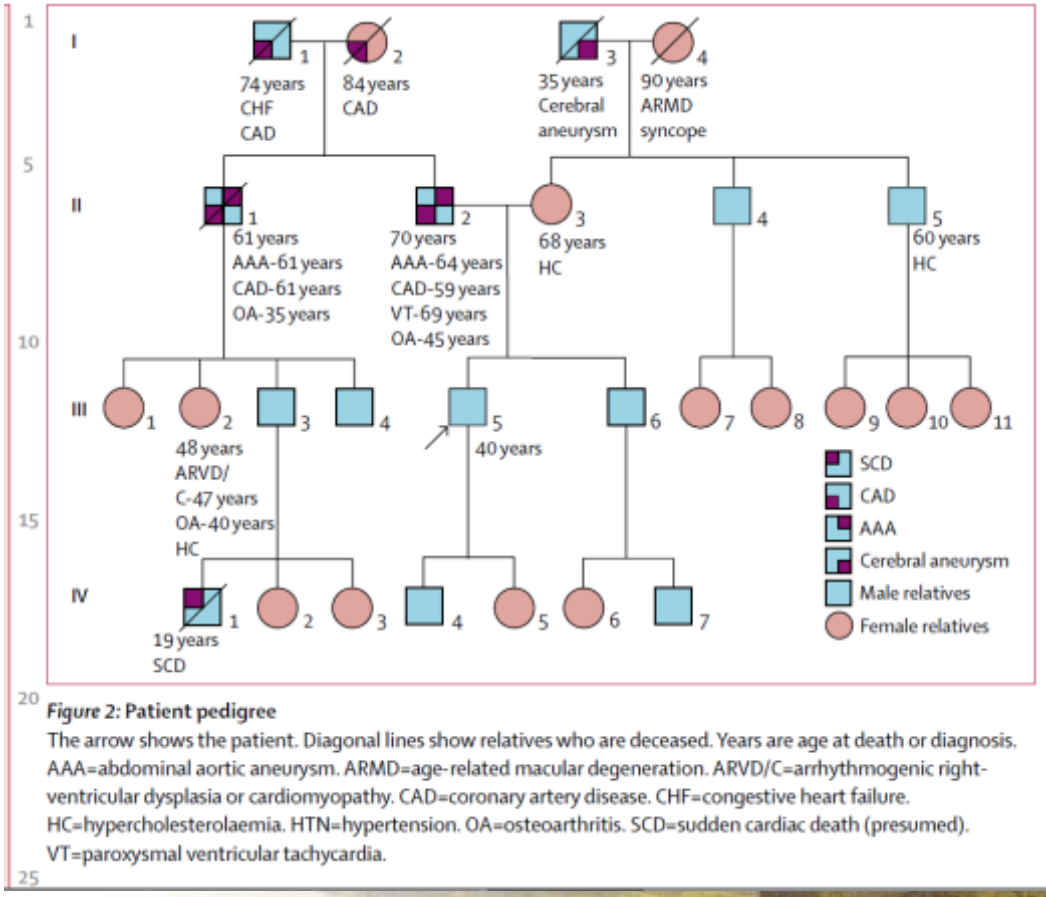


Figure 2: **Patient pedigree**
The arrow shows the patient. Diagonal lines show relatives who are deceased. Years are age at death or diagnosis. AAA=abdominal aortic aneurysm. ARMD=age-related macular degeneration. ARVD/C=arrhythmogenic right-ventricular dysplasia or cardiomyopathy. CAD=coronary artery disease. CHF=congestive heart failure. HC=hypercholesterolaemia. HTN=hypertension. OA=osteoarthritis. SCD=sudden cardiac death (presumed). VT=paroxysmal ventricular tachycardia.

STANFORD
SCHOOL OF MEDICINE

Variants predisposing to cardiac risk

Previously described variants of unknown importance in disease-associated genes						
TMEM43[24]	M41V	Transmembrane protein 43	3	14146021	None	A
MYBPC3[25]	R326Q	Myosin-binding protein C, cardiac-type	11	47324447	rs34580776	C
Novel variants potentially associated with rare disease						
DSP[11]	R1838H	Desmoplakin	6	7528007	Novel	G

- **Rare variants in 3 genes clinically associated with sudden cardiac death: *TMEM43*, *DSP*, and *MYBPC3***
- **Variant in *LPA* consistent with a family history of coronary artery disease**

Euan Ashley and team

Ashley et al (2010), *Lancet* 375:1525

Pharmacogenomics predictions

- **Heterozygous null mutation in *CYP2C19* → clopidogrel resistance?**
- **Variants associated with positive response to lipid-lowering therapy**
- ***CYP4F2* and *VKORC1* variants → low initial warfarin dose**

	Gene name	SNP location	Patient genotype	Drug(s) affected	Summary of effects	Level of evidence
SLCO1B1	Solute carrier organic anion transporter family, member 1B1	rs4149056	T/T	HMG-CoA reductase inhibitors (statins)	No increased risk of myopathy	High[32-34]
CYP2C19	Cytochrome P450, family 2, subfamily C, polypeptide 19	rs4244285	A/G	Clopidogrel and CYP2C19 substrates	CYP2C19 poor metaboliser; many drugs might need adjustment	High[35]
VKORC1	Vitamin K epoxide reductase complex, subunit 1	rs9923231	C/T	Warfarin	Reduced dose needed	High[36]
CYP4F2	Cytochrome P450, family 4, subfamily F, polypeptide 2	rs2108622	C/C	Warfarin	Reduced dose needed	High[37]
ADRB1	β1 adrenergic receptor	rs1801252	A/A	Atenolol, metoprolol	Might be preferable to calcium-channel blockers	High[38,39]
SLCO1B1	Solute carrier organic anion transporter family, member 1B1	rs11045819	A/C	Fluvastatin	Good response	Medium[40]
HMGCR	HMG-CoA reductase	rs17238540	T/T	Pravastatin	Patient might have good response	Medium
HMGCR	HMG-CoA reductase	rs17244841	A/A	Pravastatin, simvastatin	No reduced efficacy	Medium
ADRB2	β2 adrenergic receptor, surface	rs1042713	A/G	β blockers	Other treatment options might be preferable	Medium[41]
ADRB2	β2 adrenergic receptor, surface	rs1042714	C/C	β blockers	Other treatment options might be preferable	Medium[41,42]
CYP2D6	Cytochrome P450, family 2, subfamily D, polypeptide 6	rs3892097 rs1800716	C/C	Metoprolol and other CYP2D6 substrates	Normal CYP2D6 metaboliser	Medium[43]
CDKN2A/B	Cyclin-dependent kinase inhibitor 2A/2B	rs10811661	T/T	Metformin	Reduced likelihood of response	Medium[44]
CDKN2A/B	Cyclin-dependent kinase inhibitor 2A/2B	rs10811661	T/T	Troglitazone	Reduced likelihood of response	Medium[44]

SNP=single nucleotide polymorphism. HMG-CoA=3-hydroxy-3-methylglutaryl-coenzyme A.

***Table 3*: Pharmacogenomic variants with summary of effects and level of evidence**

Russ Altman and team

Ashley EA*, Butte AJ*, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Morgan AA, Pushkarev D, Neff NF, Knowles W, Chou M, Thakuria J, Rosenbaum A, Zaranek AW, Church G, Greely HT*, Quake SR*, Altman RB*. Clinical evaluation incorporating a personal genome. *Lancet*, 2010.

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 Färkkilä, 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),[1–3] 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)[4] 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.[6–13] 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.[13,14,16] 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.20423
Published online 13 March 2008 in Wiley InterScience (www.interscience.wiley.com)

Rong Chen
Optra Systems

- Study published in 2008 in *Inflammatory Bowel Disease*
- Crohn's Disease and Ulcerative Colitis
- Investigated 9 loci in 700 Finnish IBD patients
- We record 100+ items
  - GWAS, non-GWAS papers
  - Disease, Phenotype
  - Population, Gender
  - Alleles and Genotypes
  - p-value (and confidence)
  - Odds ratio (and confidence)
  - Technology, Study design
  - Genetic model
- Mapped to UMLS concepts

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 Färkkilä, 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.

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),[1–3] sev-

sinki, Haartmaninkatu 4, FIN-00290 Helsinki, Finland (e-mail: kimmo.kontula@hus.fi).

DOI 10.1002/ibd.20431

Published online13 March 2008 in Wiley InterScience (www.interscience.wiley.com).

port.[13,14,16] 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

TABLE 1. Case-control Analysis of the *IL23R* Gene Including 8 SNPs

dbSNP ID	Allele	Location	Controls $n = 292$	IBD $n = 697$	*P* value	CD $n = 238$	*P* value	UC $n = 459$	*P* value
rs1004819	C	Intron 5	0.751	0.704	**0.037**	0.671	**0.005**	0.721	0.215
	T		0.249	0.296		0.329		0.279	

- Study published in 2008 in *Inflammatory Bowel Disease*
- Crohn's Disease and Ulcerative Colitis
- Investigated 9 loci in 700 Finnish IBD patients
- We record 100+ items
  - GWAS, non-GWAS papers
  - Genetic model
- Mapped to UMLS concepts

VARIMED: Variants Informing Medicine

Number of papers curated	Distinct SNPs	Diseases and phenotypes
~11,250	~192,000	~4,400

Chen R, Davydov EV, Sirota M, Butte AJ.
PLoS One.
2010 October: 5(10): e13574.

Rong Chen
Optra Systems

Moving from OR to LR

Odds ratio

Ratio of odds of test positivity in cases over odds of test positivity in non-cases

Likelihood ratio (+)

The probability of test positive in cases, over the probability of test positive in non-cases

Sensitivity / (1 – Specificity)

Very similar, but different...

Morgan A, Chen R, Butte AJ. *Genomic Medicine*, 2010.

Post-test probability is calculated with likelihood ratio

Pre-test odds x likelihood ratio → Post-test odds

Pre-test odds x LR1 x LR2 x LR3 → Post-test odds

Can chain likelihood ratios from independent tests

Morgan A, Chen R, Butte AJ. *Genomic Medicine*, 2010.

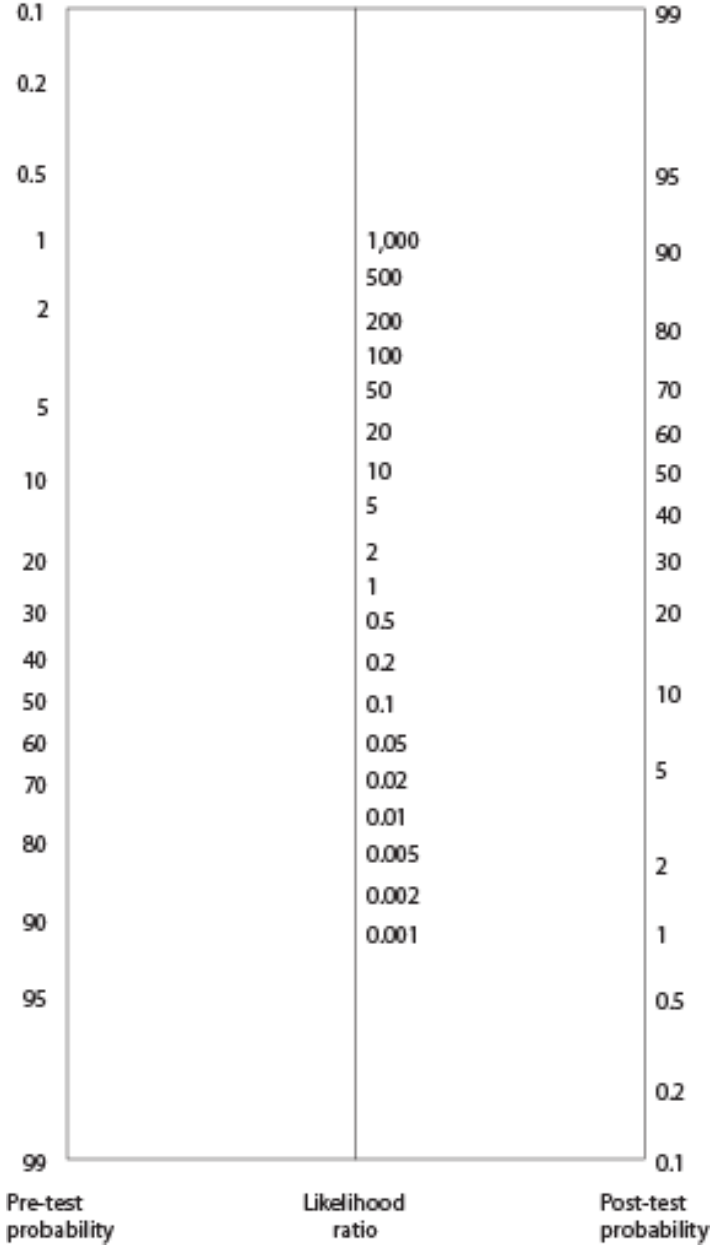


Figure 1. Nomogram for likelihood ratios. The pre-test and post-test probabilities and likelihood ratios of any diagnostic test, including a genetic test, can be visualized using a nomogram familiar to most physicians and medical students. The nomogram shown is derived from the Fagan nomogram [14], and modified from one generated using a web-based tool [28]. The left side of the figure indicates a hypothetical pre-test probability of disease of 27%. Three lines represent the three possible genotypes, from top to bottom: homozygous risk alleles with a likelihood ratio of 1.61, heterozygous alleles with a likelihood ratio of 1.26, and homozygous protective alleles with a likelihood ratio of 0.83. The right side of the figure indicates three possible post-test probabilities resulting from the three genotypes. Multiple such tests can be 'chained' together serially, if they describe independent risks and cover the same pre-test assumptions.

Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5): 257.

Morgan, Chen, Butte. Likelihood ratios for genomic medicine. *Genome Medicine*. 2010; 2:30.

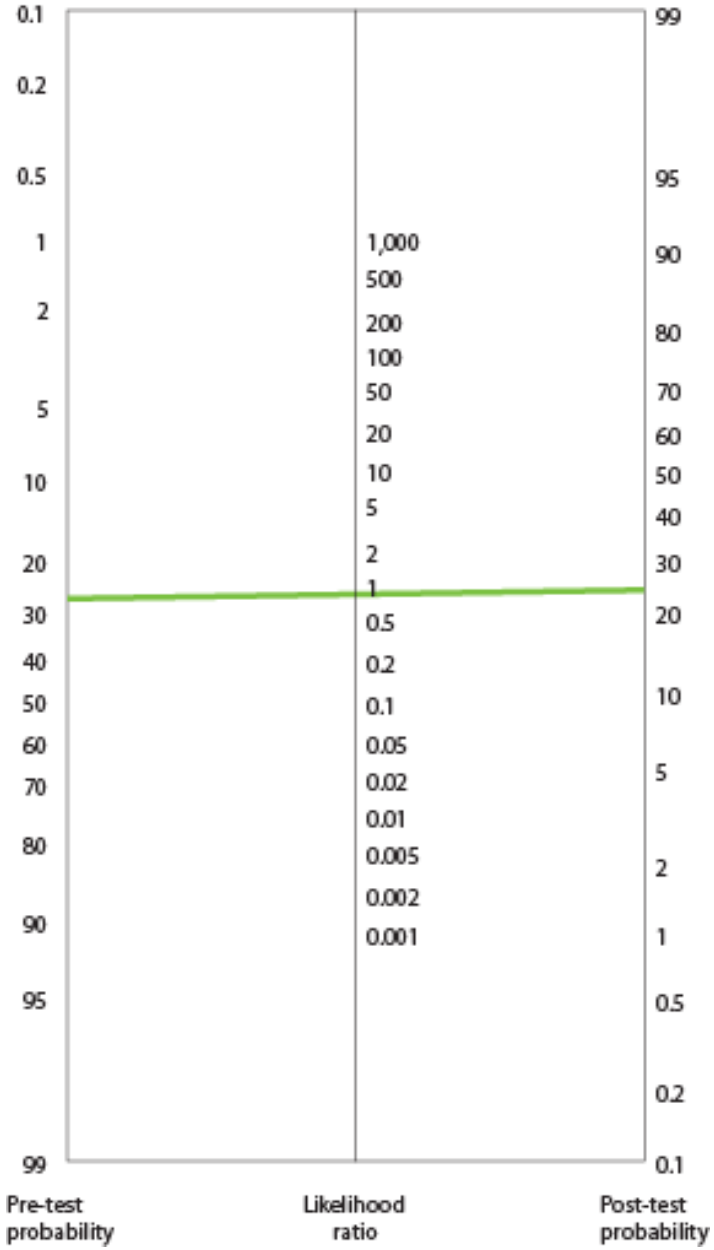


Figure 1. Nomogram for likelihood ratios. The pre-test and post-test probabilities and likelihood ratios of any diagnostic test, including a genetic test, can be visualized using a nomogram familiar to most physicians and medical students. The nomogram shown is derived from the Fagan nomogram [14], and modified from one generated using a web-based tool [28]. The left side of the figure indicates a hypothetical pre-test probability of disease of 27%. Three lines represent the three possible genotypes, from top to bottom: homozygous risk alleles with a likelihood ratio of 1.61, heterozygous alleles with a likelihood ratio of 1.26, and homozygous protective alleles with a likelihood ratio of 0.83. The right side of the figure indicates three possible post-test probabilities resulting from the three genotypes. Multiple such tests can be 'chained' together serially, if they describe independent risks and cover the same pre-test assumptions.

Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5): 257.

Morgan, Chen, Butte. Likelihood ratios for genomic medicine. *Genome Medicine*. 2010; 2:30.

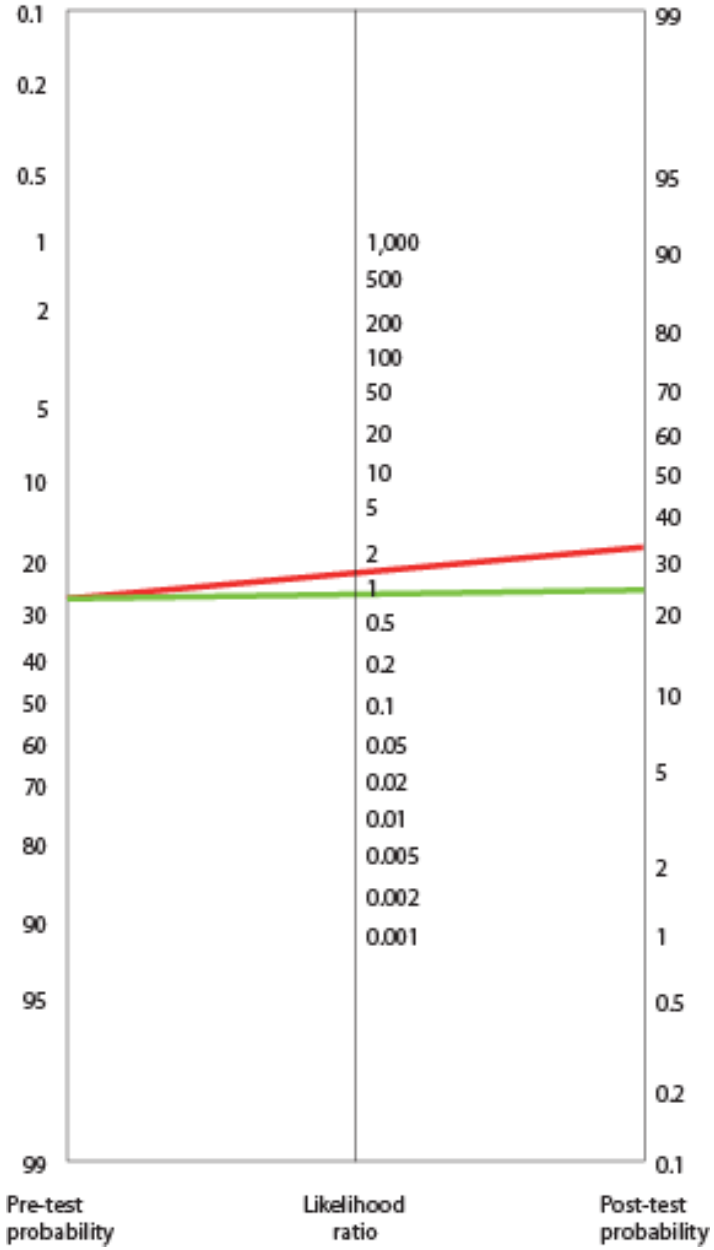


Figure 1. Nomogram for likelihood ratios. The pre-test and post-test probabilities and likelihood ratios of any diagnostic test, including a genetic test, can be visualized using a nomogram familiar to most physicians and medical students. The nomogram shown is derived from the Fagan nomogram [14], and modified from one generated using a web-based tool [28]. The left side of the figure indicates a hypothetical pre-test probability of disease of 27%. Three lines represent the three possible genotypes, from top to bottom: homozygous risk alleles with a likelihood ratio of 1.61, heterozygous alleles with a likelihood ratio of 1.26, and homozygous protective alleles with a likelihood ratio of 0.83. The right side of the figure indicates three possible post-test probabilities resulting from the three genotypes. Multiple such tests can be 'chained' together serially, if they describe independent risks and cover the same pre-test assumptions.

Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5): 257.

Morgan, Chen, Butte. Likelihood ratios for genomic medicine. *Genome Medicine*. 2010; 2:30.

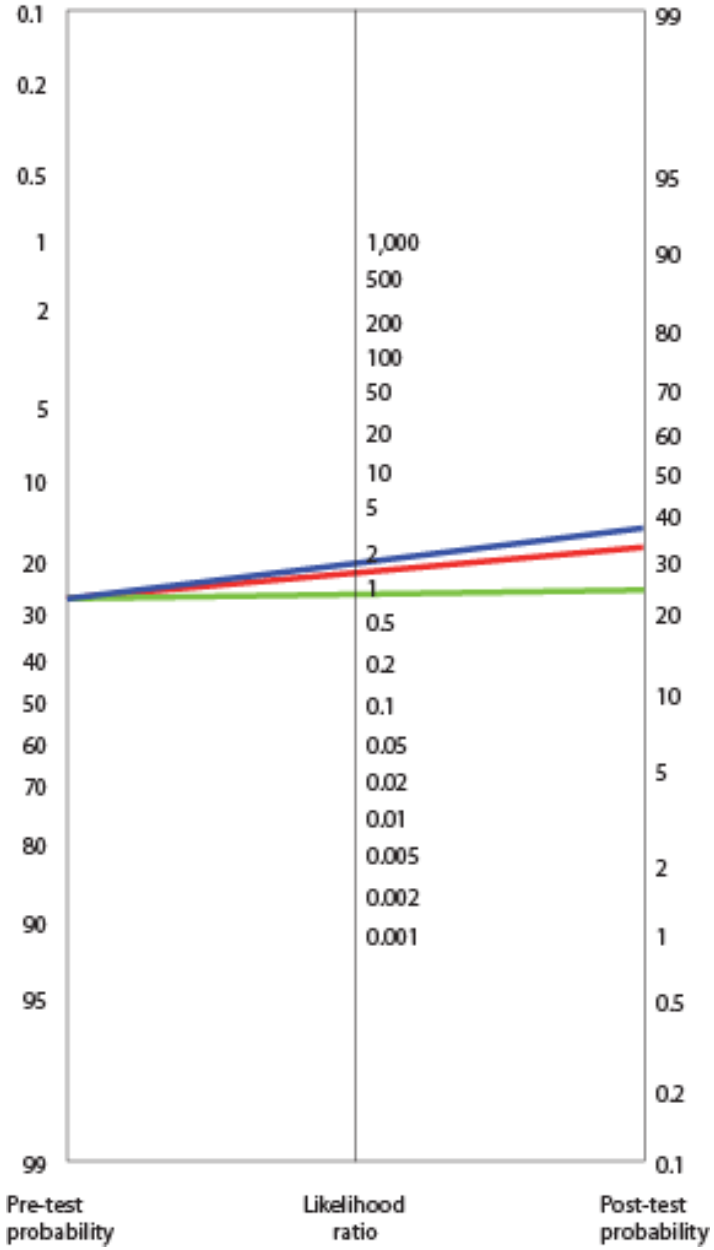


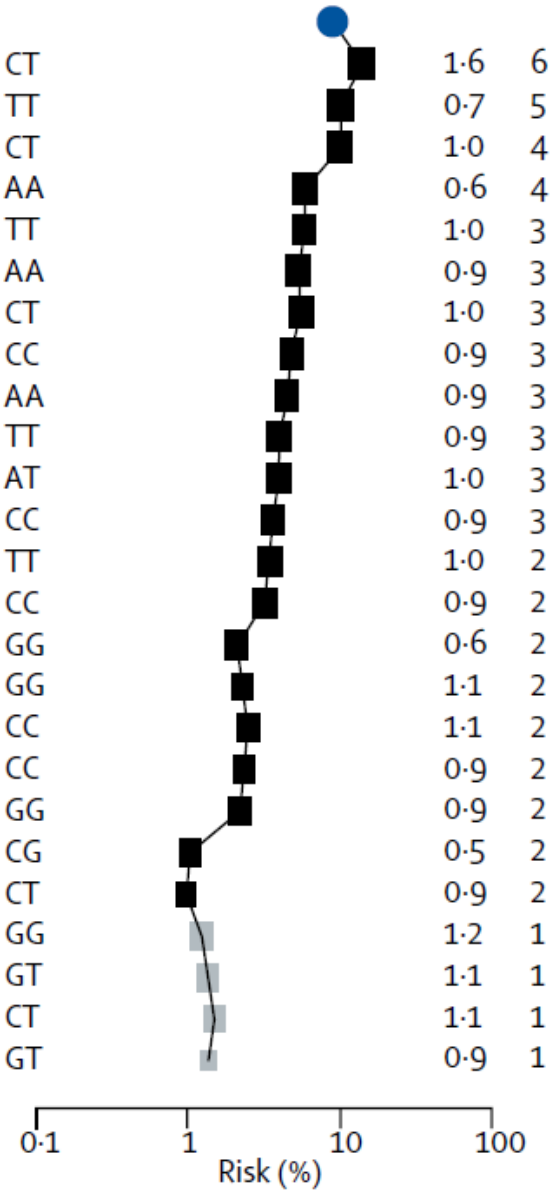
Figure 1. Nomogram for likelihood ratios. The pre-test and post-test probabilities and likelihood ratios of any diagnostic test, including a genetic test, can be visualized using a nomogram familiar to most physicians and medical students. The nomogram shown is derived from the Fagan nomogram [14], and modified from one generated using a web-based tool [28]. The left side of the figure indicates a hypothetical pre-test probability of disease of 27%. Three lines represent the three possible genotypes, from top to bottom: homozygous risk alleles with a likelihood ratio of 1.61, heterozygous alleles with a likelihood ratio of 1.26, and homozygous protective alleles with a likelihood ratio of 0.83. The right side of the figure indicates three possible post-test probabilities resulting from the three genotypes. Multiple such tests can be 'chained' together serially, if they describe independent risks and cover the same pre-test assumptions.

Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med*. 1975 Jul 31;293(5): 257.

Morgan, Chen, Butte. Likelihood ratios for genomic medicine. *Genome Medicine*. 2010; 2:30.

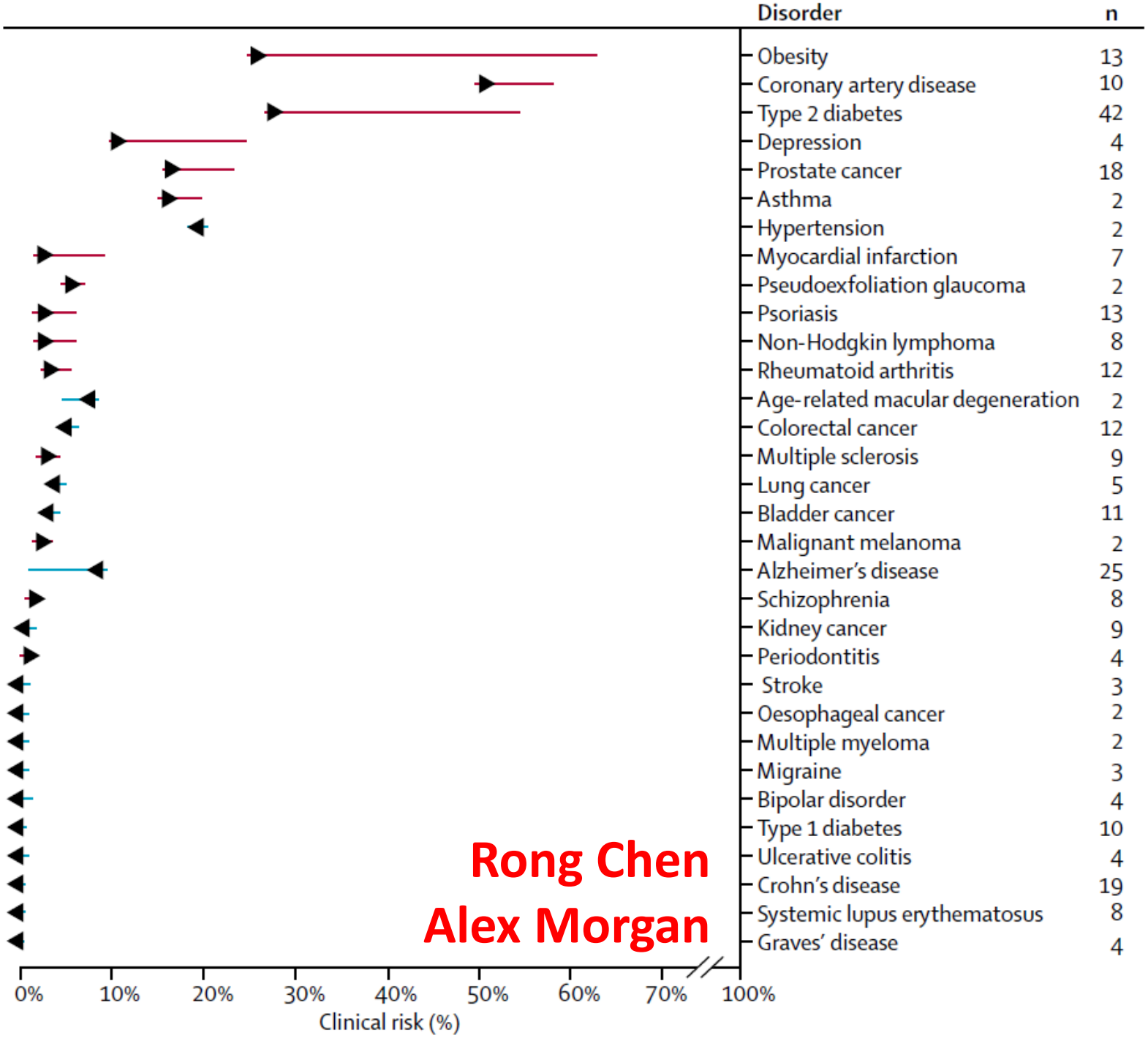
D Alzheimer's disease

Gene*	SNP location	Patient genotype	LR	Studies†	Samples‡	Post-test probability (%)
						9·0%
TOMM40	rs157581	CT	1·6	6	7740	13·90%
DAPK1	rs4878104	TT	0·7	5	10397	10·19%
TRAK2	rs13022344	CT	1·0	4	6512	10·12%
DAPK1	rs4877365	AA	0·6	4	4841	5·89%
E8F3	rs11016976	TT	1·0	3	5736	5·87%
TNK1	rs1554948	AA	0·9	3	5736	5·32%
MYH13	rs2074877	CT	1·0	3	5366	5·55%
GALP	rs3745833	CC	0·9	3	5366	4·82%
PCK1	rs8192708	AA	0·9	3	5366	4·47%
	rs1859849	TT	0·9	3	5304	4·02%
	rs11622883	AT	1·0	3	5248	3·97%
WWC1	rs17070145	CC	0·9	3	2545	3·65%
LMNA	rs505058	TT	1·0	2	4646	3·49%
ACAN	rs2882676	CC	0·9	2	4590	3·22%
PGBD1	rs3800324	GG	0·6	2	4590	2·11%
GOLM1	rs10868366	GG	1·1	2	2156	2·30%
GOLM1	rs7019241	CC	1·1	2	2156	2·49%
	rs9886784	CC	0·9	2	2156	2·36%
	rs10519262	GG	0·9	2	2156	2·22%
	rs463946	CG	0·5	2	1922	1·04%
PLAU	rs2227564	CT	0·9	2	956	0·98%
ADAM12	rs1278279	GG	1·2	1	2320	1·23%
SORL1	rs2070045	GT	1·1	1	2031	1·36%
ABCA1	rs2230806	CT	1·1	1	1691	1·50%
PSEN1	rs165932	GT	0·9	1	170	1·37%



Rong Chen
Alex Morgan

Ashley EA*, Butte AJ*, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Morgan AA, Pushkarev D, Neff NF, Knowles W, Chou M, Thakuria J, Rosenbaum A, Zaranek AW, Church G, Greely HT*, Quake SR*, Altman RB*. Clinical evaluation incorporating a personal genome. *Lancet*, 2010.


Disorder
n
Obesity 13
Coronary artery disease 10
Type 2 diabetes 42
Depression 4
Prostate cancer 18
Asthma 2
Hypertension 2
Myocardial infarction 7
Pseudoexfoliation glaucoma 2
Psoriasis 13
Non-Hodgkin lymphoma 8
Rheumatoid arthritis 12
Age-related macular degeneration 2
Colorectal cancer 12
Multiple sclerosis 9
Lung cancer 5
Bladder cancer 11
Malignant melanoma 2
Alzheimer's disease 25
Schizophrenia 8
Kidney cancer 9
Periodontitis 4
Stroke 3
Oesophageal cancer 2
Multiple myeloma 2
Migraine 3
Bipolar disorder 4
Type 1 diabetes 10
Ulcerative colitis 4
Crohn's disease 19
Systemic lupus erythematosus 8
Graves' disease 4
Rong Chen
Alex Morgan
0%
10%
20%
30%
40%
50%
60%
70%
100%
Clinical risk (%)

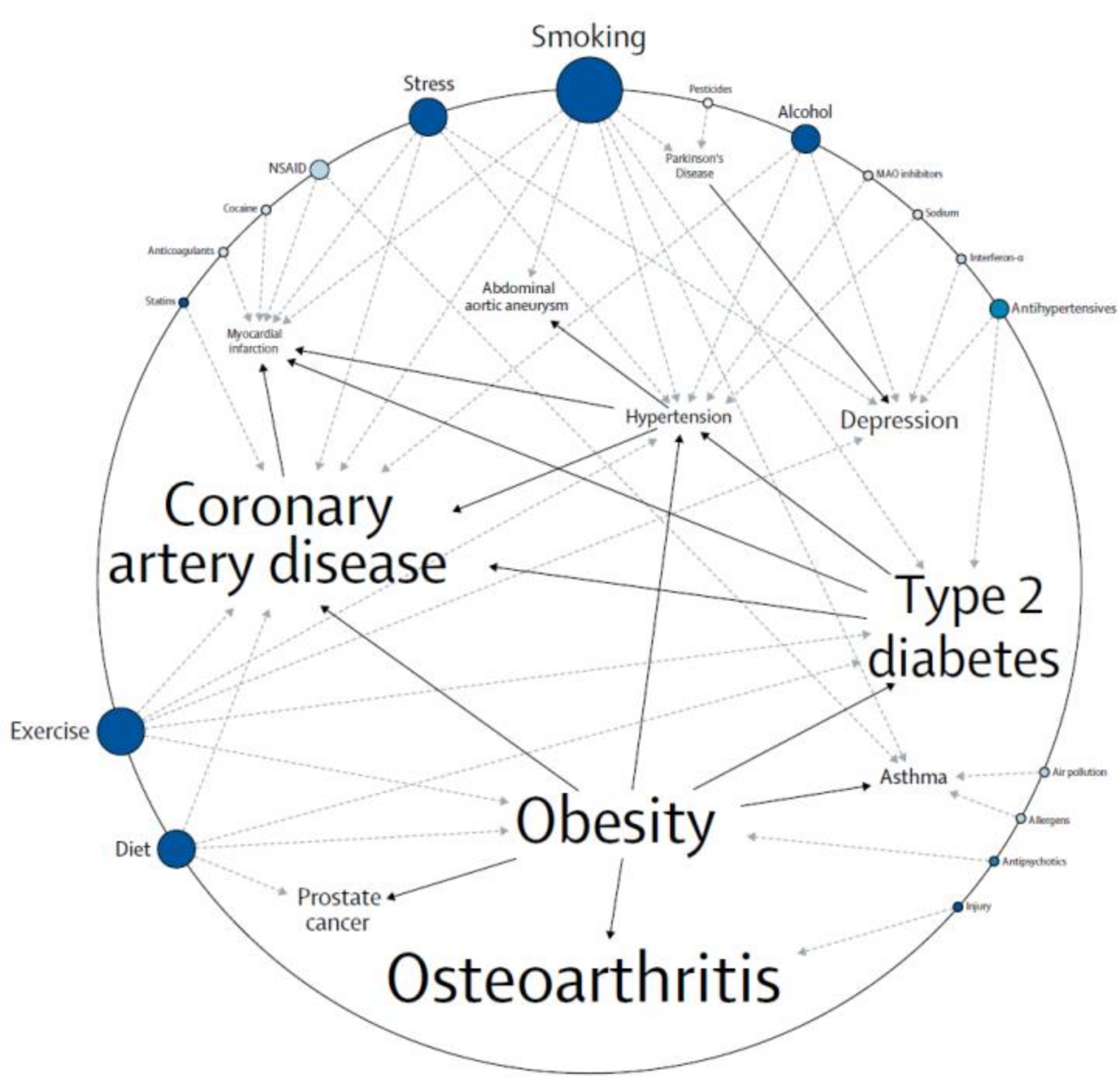
So what can we do about the risk?

- Diseases with higher post-test probabilities
- How to alter the influence of genetics?

- Diseases are caused by genes and environment

- We need a simple "prescription" for environmental change for a genome-enabled patient

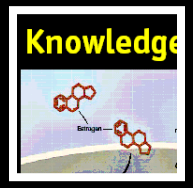
- How do we compensate for our genomes?



Rong Chen
Alex Morgan
Joel Dudley

Take Home Points

- Molecular, clinical, trials, and epidemiological data and tools already exist → diagnostics and therapeutics.

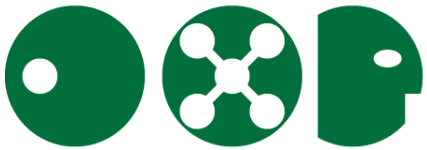


- Public big data is highly enabling.
  Use it, and share your data after publication.

- Personalized medicine $\geq$ DNA. Needs to include other clinical, molecular, and environment measures.

Collaborators

- Takashi Kadowaki, Momoko Horikoshi, Kazuo Hara, Hiroshi Ohtsu / U Tokyo
- Kyoko Toda, Satoru Yamada, Junichiro Irie / Kitasato Univ and Hospital
- Shiro Maeda / RIKEN
- Alejandro Sweet-Cordero, Julien Sage / Pediatric Oncology
- Mark Davis, C. Garrison Fathman / Immunology
- Russ Altman, Steve Quake / Bioengineering
- Euan Ashley, Joseph Wu, Tom Quertermous / Cardiology
- Mike Snyder, Carlos Bustamante, Anne Brunet / Genetics
- Jay Pasricha / Gastroenterology
- Rob Tibshirani, Brad Efron / Statistics
- Hannah Valantine, Kiran Khush/ Cardiology
- Ken Weinberg / Pediatric Stem Cell Therapeutics
- Mark Musen, Nigam Shah / National Center for Biomedical Ontology
- Minnie Sarwal / Nephrology
- David Miklos / Oncology



Butte Lab

Systems Medicine • Stanford Pediatrics • Packard Children's Hospital

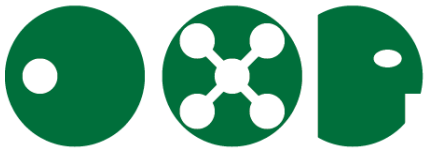
Support

- Lucile Packard Foundation for Children's Health
- NIH: NIAID, NLM, NIGMS, NCI; NIDDK, NHGRI, NIA, NHLBI, NCATS
- March of Dimes
- Hewlett Packard
- Howard Hughes Medical Institute
- California Institute for Regenerative Medicine
- Scleroderma Research Foundation
- Clayville Research Fund
- PhRMA Foundation
- Stanford Cancer Center, Bio-X

Admin and Tech Staff

- Susan Aptekar
- Camilla Morrison
- Alex Skrenchuk

- Tarangini Deshpande
- Alan Krensky, Harvey Cohen
- Hugh O'Brodovich
- Isaac Kohane



Butte Lab

Systems Medicine • Stanford Pediatrics • Packard Children's Hospital