Discovering new drugs and diagnostics from 300 billion points of data

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

Lucile Packard Children's Hospital AT STANFORD

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

Disclosures

- Scientific founder and advisory board membership
  - Genstruct
  - NuMedii
  - Personalis
  - Carmenta
  - Assay Depot
- Past or present consultancy
  - Lilly
  - Johnson and Johnson
  - Roche
  - NuMedii
  - Genstruct
  - Tercica
  - Ansh Labs
  - Prevendia

- Honoraria
  - Lilly
  - Pfizer
  - Siemens
  - Bristol Myers Squibb
- Speakers' bureau
  - None

- Companies started by students
  - Carmenta
  - NuMedii
  - Stimulomics
  - NunaHealth
  - Praedicat
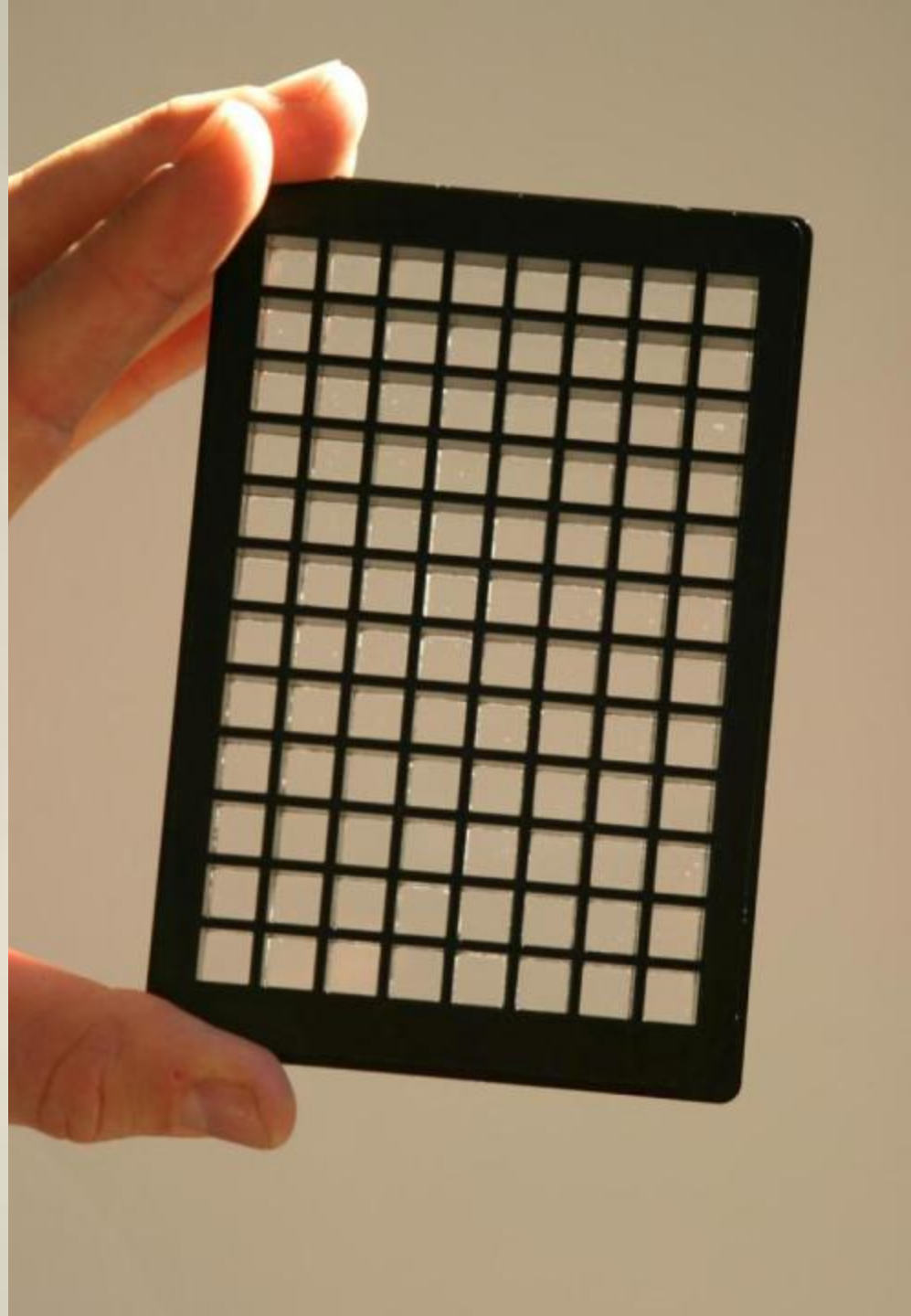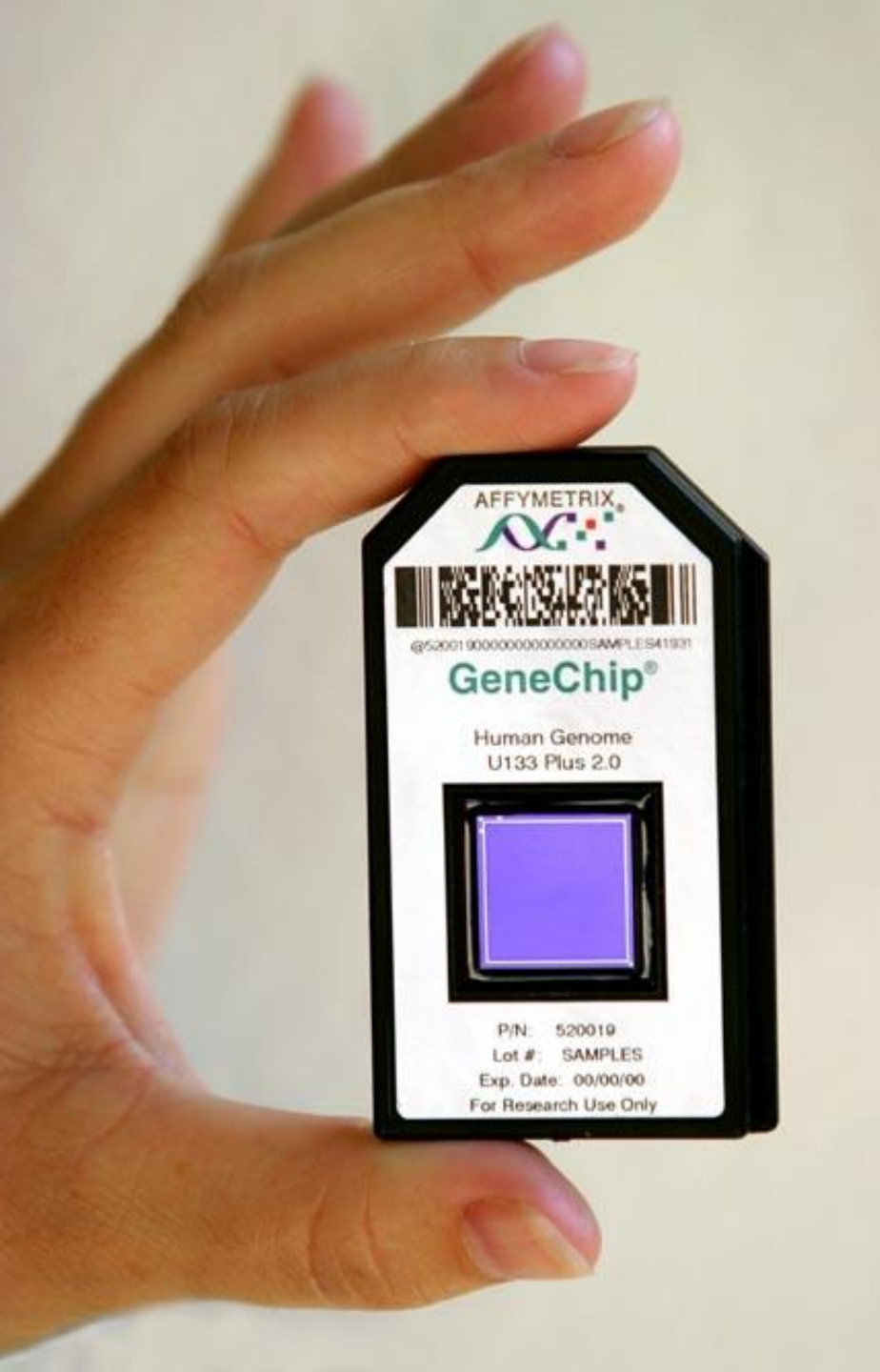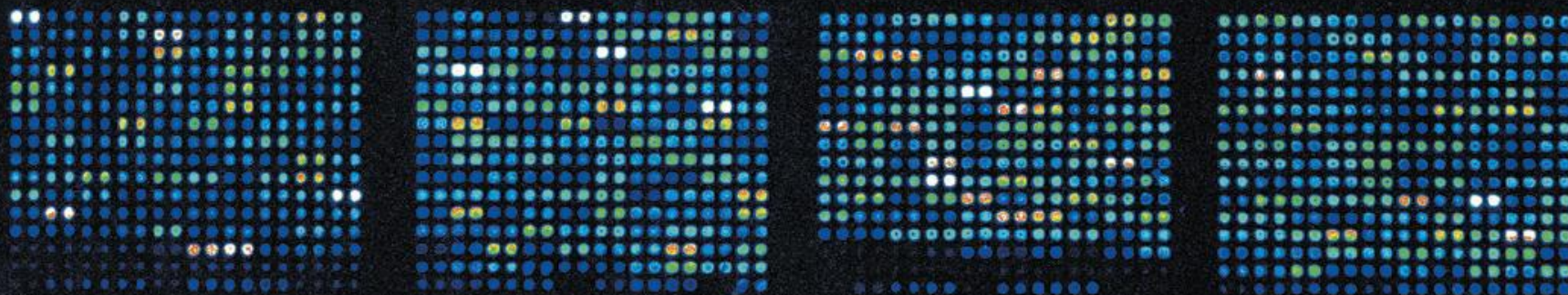
DNA microarrays allow researchers to analyse the expression of a huge number of genes simultaneously.

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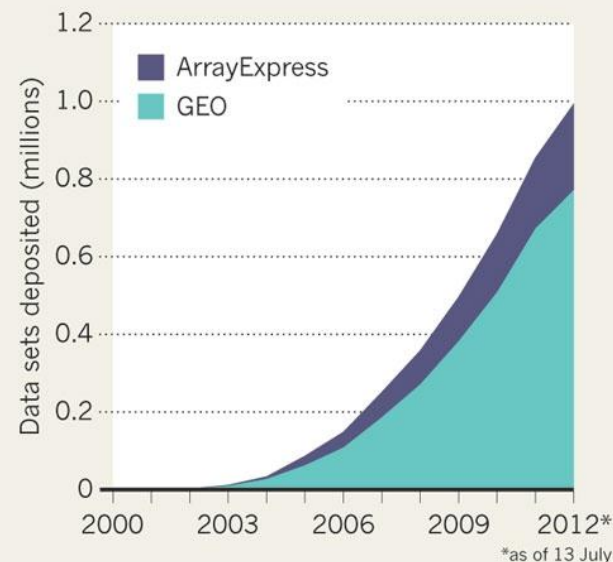
That is exactly the kind of search that led Khatri's boss, Atul Butte, a bioinformatician at the Stanford School of Medicine in California, to identify a new drug target for diabetes. After downloading data from 130 gene-expression

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

DATA DUMP

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

Total 1.07 million microarrays available

Doubles every 2-3 years

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

GEO DataSets | GEO DataSets ▼ | breast cancer | ⊗ | Sear

Save search   Limits   Advanced

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

Filter your results:

All (35583)

Results: 1 to 20 of 35583   << First < Prev Page [1] of 1780 Next > Last >>

DataSets (90)

Platforms (27)

Breast cancer: histologically normal breast epithelium

1. 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.

Samples (34162)

Series (1304)

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

▼ Top Organisms [

Homo sapiens (330
Mus musculus (242
Rattus norvegicus (
Canis lupus familia
Human herpesvirus
More...

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

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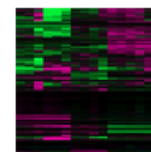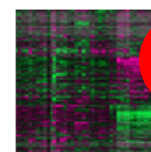
Organism: Homo sapiens
Type: Expression profiling by array, transformed count, 2 agent, 3 dose, 2 time sets
Platform: GPL571   Series: GSE7848 16 Samples
Download data: GEO (CEL)

DataSet   Accession: GDS3638   ID: 3638

PubMed   Similar studies   GEO Profiles   Analyze DataSet

Find related data

Database: | Select ▼

Find items

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 >>

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

☐ Actein effect on **breast cancer** cell line: dose response and time course
2. Analysis of
   from the her
   inhibitory eff
   Organism:
   Type:
   Platform: GP
   Download dat
   DataSet   Ac
   PubMed   S

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

Intel and Siemens Competition finalist
• **Andrew Liu** (2010)
Intel Science Talent Search semi-finalists
• **Rohan Chakicherla** (2009)
• **Denzil Sikka** (2009)
• **Tony Ho** (2010)
• **Irving Hsu** (2011)

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	Longitudin
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-offsprin
GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen		V D A	-	Parent-offsprin
GAIN: International Multi-Center ADHD Genetics Project	Mar 26, 2008	V D A	2835	Parent-offsprin
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	Longitudin 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

About ImmPort

ImmPort, the Immunology Database and Analysis Portal, is a one stop shop to access reference and experiment data for immunologists. ImmPort provides advanced information technology support in the production, analysis, archiving, and exchange of scientific data for the diverse community of life science researchers supported by NIAID/DAIT.

▶ What is ImmPort

DATA SOURCES → **IMMPORT DATABASE** → **IMMPORT TOOLS**

NIAID/DAIT Investigators
Experimental Data
Clinical Study Data

Public Reference Databases
Examples:

Data Standardization
Quality Control
Data Curation
Maps to Ontologies

Search Data
Visualization
Data Analysis

Important Notice

We have completed the migration of ImmPort to the NIAID hosted facility. Thank you for your patience during the transition and for finding us at our new home, immport.niaid.nih.gov

Sign In

User Name:

Password:

SIGN IN   Forgot Password?

Register for ImmPort ❓
Benefits of Registration:
» Access to data visualization tools
» Compare data sets to other researchers
» Access to data analysis tools

FlowCAP

Flow Cytometry: Critical Assessment of Population Identification Methods project.

The goal of FlowCAP

What You Can Do:

Search Data | **Visualize Data** | **Analyze Data**

Genes
Access integrated information about genes and their protein products, including structural and functional attributes, biological pathway membership, protein-protein interactions and more.

SNPs
Search the ImmPort Single Nucleotide Polymorphism (SNP) database, a collection of NCBI dbSNP, Entrez Gene and HapMap data.

Pathways
Explore the biological pathways associated with particular genes or proteins using pathway information from KEGG. (BioCarta, Reactome, MGI

MHC Alleles
Find complete DNA and protein sequences, sequence features, and population frequencies of MHC, MIC and TAP alleles. Align MHC sequences to visualize extent of polymorphisms.

ImmPort Gene Lists
View detailed information about immunologically-relevant genes by category. Examine differential gene expression information from immunologically-relevant microarray data sets.

Immune Cells
Learn about immunologically relevant cell types and the surface markers expressed and the cytokines

PubChem

108 million substances x 650,000 assays

1 billion points of data within a grid of 70 trillion cells

BioAssay · Compound · Substance

Advanced search

Chemical structure search | BioActivity analysis

NCBI Resources How To

Chem PubChem Substance all[filt]

stance Save search Limits

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

Results: 1 to 20 of 108327716

Cadmium ion: CD
Source: MMDB (105286.3)
SID: 15
Summary

MANG
Source:
SID: 15
Summary

NCBI Resources How To

PubChem PubChem C

Compound

Display Settings: Summary, 20 per

Results: to 20 of 32454538

NCBI Resources How To

PubChem PubChem BioAssay all[filt]

BioAssay Save search Limits Advance

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]

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

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

$9,000.00 USD
per service

9 week
turn around time

Provided By
Links Biosciences

Request Info

Add to Cart

SHARE

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

Ask An Expert

Use our free service locator program to find the research services you need.

Search PubMed

Search PubMed for "Diabetes and Obesity" using BioWizard.

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

Translational Pipeline

Clinical and Molecular Measurements

⬇⬆

Translational Question or Trial

⬇

Statistical/Computational methods

⬇

Validating drug or biomarker

Translational Pipeline

Clinical and Molecular Measurements

⬇⬆

Translational Question or Trial

⬇

Statistical/Computational methods

⬇

Validating drug or biomarker

Commodity

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"?

Type 2 Diabetes Mellitus

- Affects 20 million in US, 170 million world-wide
- Leading cause of kidney failure, blindness, amputation
- Major risk factor for heart disease, stroke, birth defects
- 12% of all US health-care dollars

- Prevalence in children born after the year 2000 expected to reach 30%

- Many drugs available to elicit more insulin secretion, heighten insulin response, lower glucagon secretion
- New drugs still needed and used: DPP-4 inhibitors (2008)

Any one experiment does not yield clear disease-causal factors

Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1

Mary Elizabeth Patti*[†‡], Atul J. Butte*[†§], Sarah Crunkhorn*, Kenneth Cusi[¶], Rachele Berria[¶], Sangeeta Kashyap[¶], Yoshinori Miyazaki[¶], Isaac Kohane[§], Maura Costello*, Robert Saccone*, Edwin J. Landaker*, Allison B. Goldfine*, Edward Mun[∥], Ralph DeFronzo[¶], Jean Finlayson[¶], C. Ronald Kahn*, and Lawrence J. Mandarino[¶]

*Research Division, Joslin Diabetes Center, Boston, MA 02215; §Division of Informatics and Endocrinology, Children's Hospital, Boston, MA 02115; ∥Department of Surgery, Beth Israel Deaconess Medical Center, Boston, MA 02215; and ¶University of Texas Health Science Center, San Antonio, TX 78229

Contributed by C. Ronald Kahn, May 14, 2003

Type 2 diabetes mellitus (DM) is characterized by insulin resistance and pancreatic β cell dysfunction. In high-risk subjects, the earliest detectable abnormality is insulin resistance in skeletal muscle. Impaired insulin-mediated signaling, gene expression, glycogen synthesis, and accumulation of intramyocellular triglycerides have all been linked with insulin resistance, but no specific defect responsible for insulin resistance and DM has been identified in humans. To identify genes potentially important in the pathogenesis of DM, we analyzed gene expression in skeletal muscle from

- One example of a microarray experiment with diabetes and control samples

- 187 genes differentially expressed
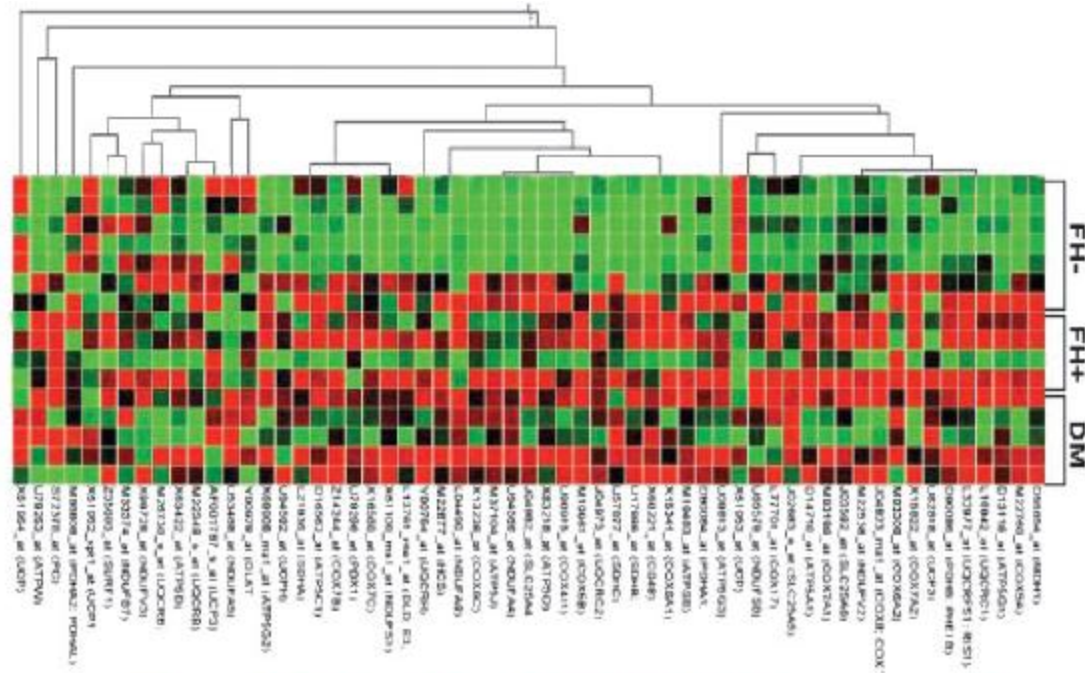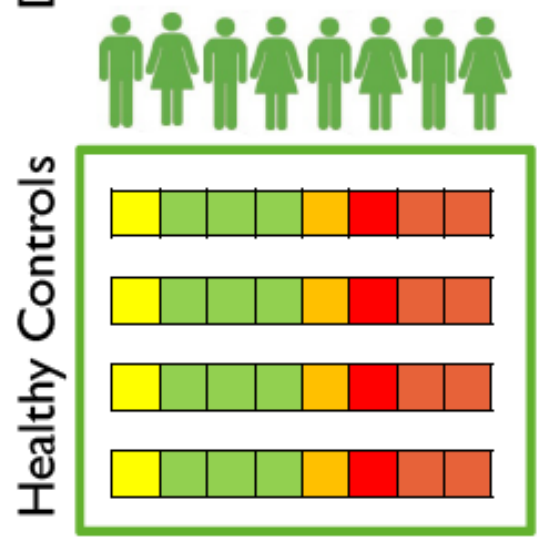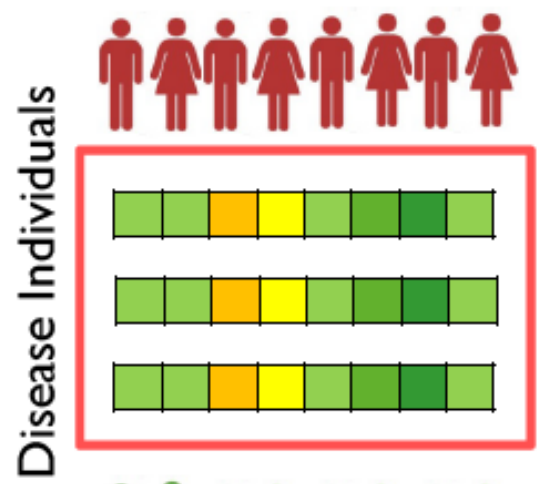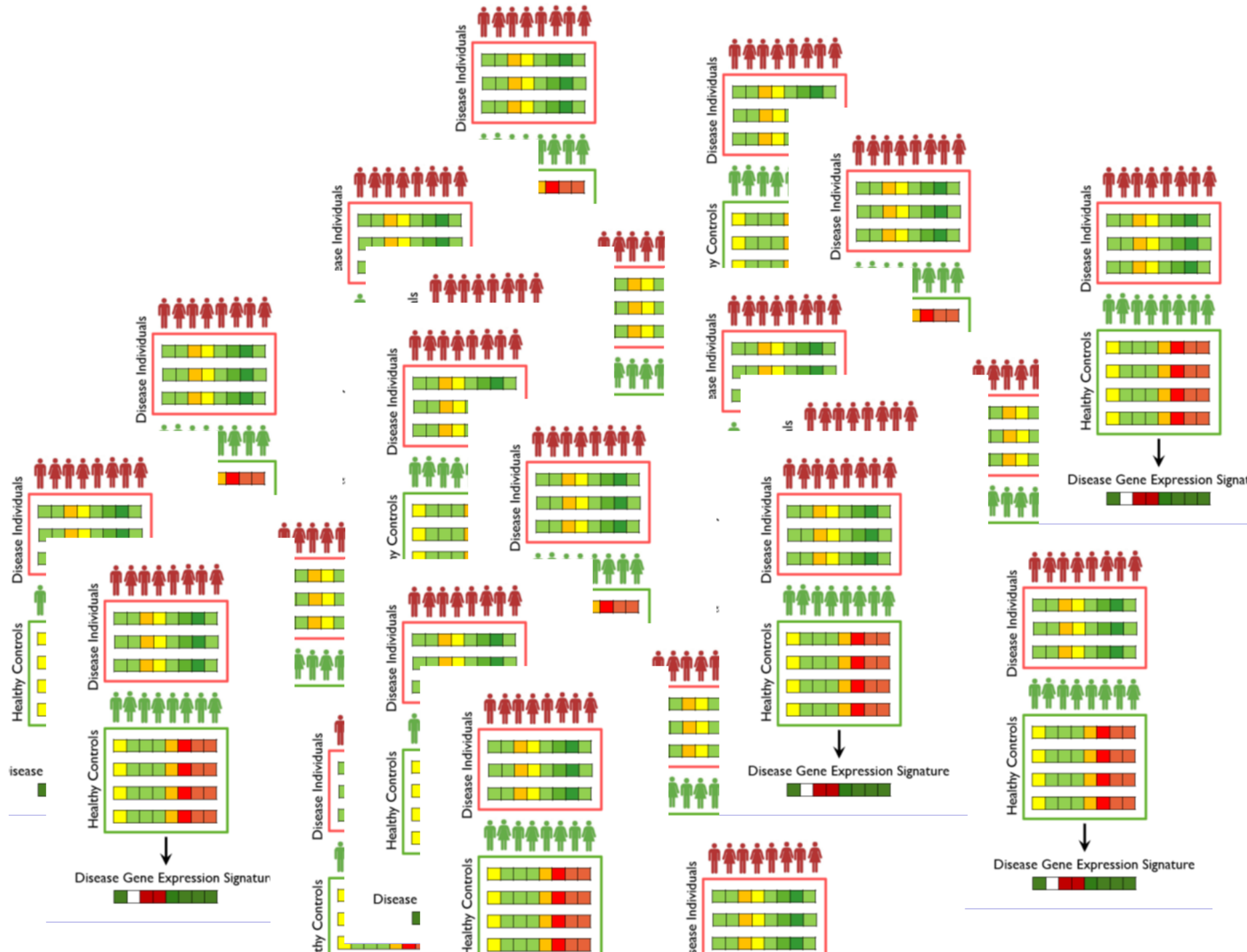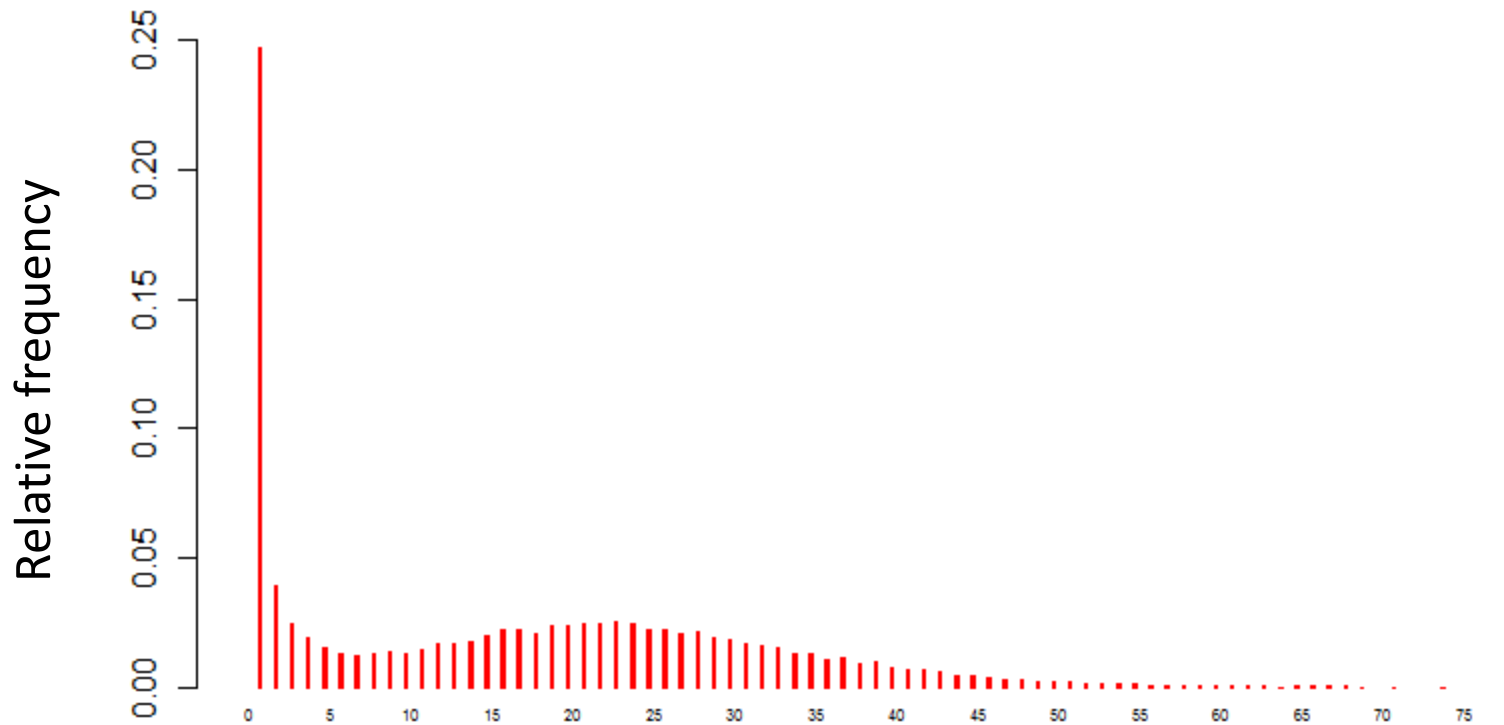
Disease Individuals

Healthy Controls

Disease Gene Expression Signature

Disease Individuals

Healthy Controls

Disease Gene Expression Signature

Intersect 130 T2D microarray experiments

# of positive RNA microarray experiments (out of 130)

Keiichi Kodama

Intersect 130 T2D microarray experiments

Most of the 25000 genes in the genome are positive in few T2D microarray experiments

# of positive RNA microarray experiments (out of 130)

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Keiichi Kodama
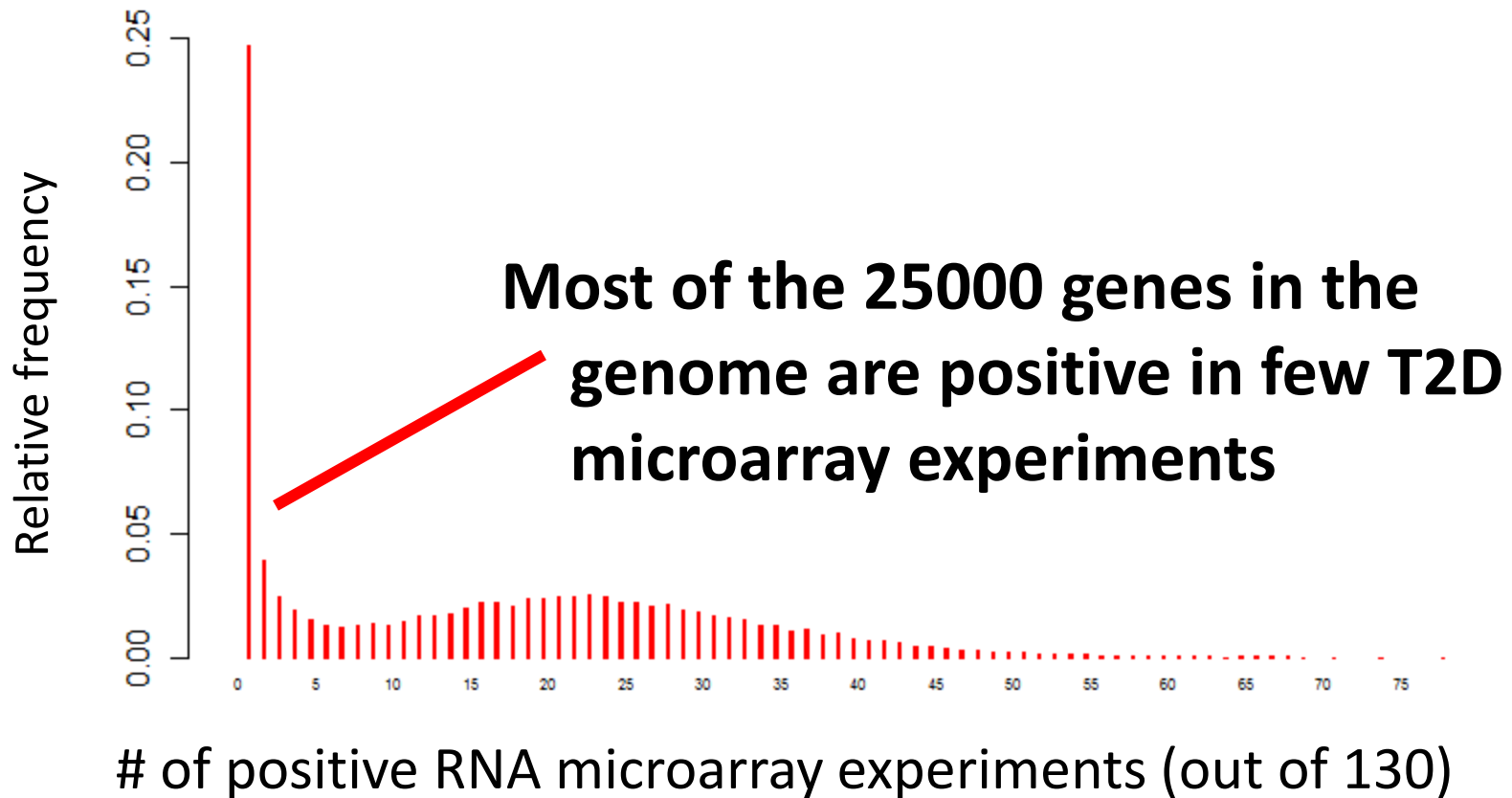
Intersect 130 T2D microarray experiments

The 186 best known drug targets or genes with DNA variants (from GWAS) are positive in more experiments
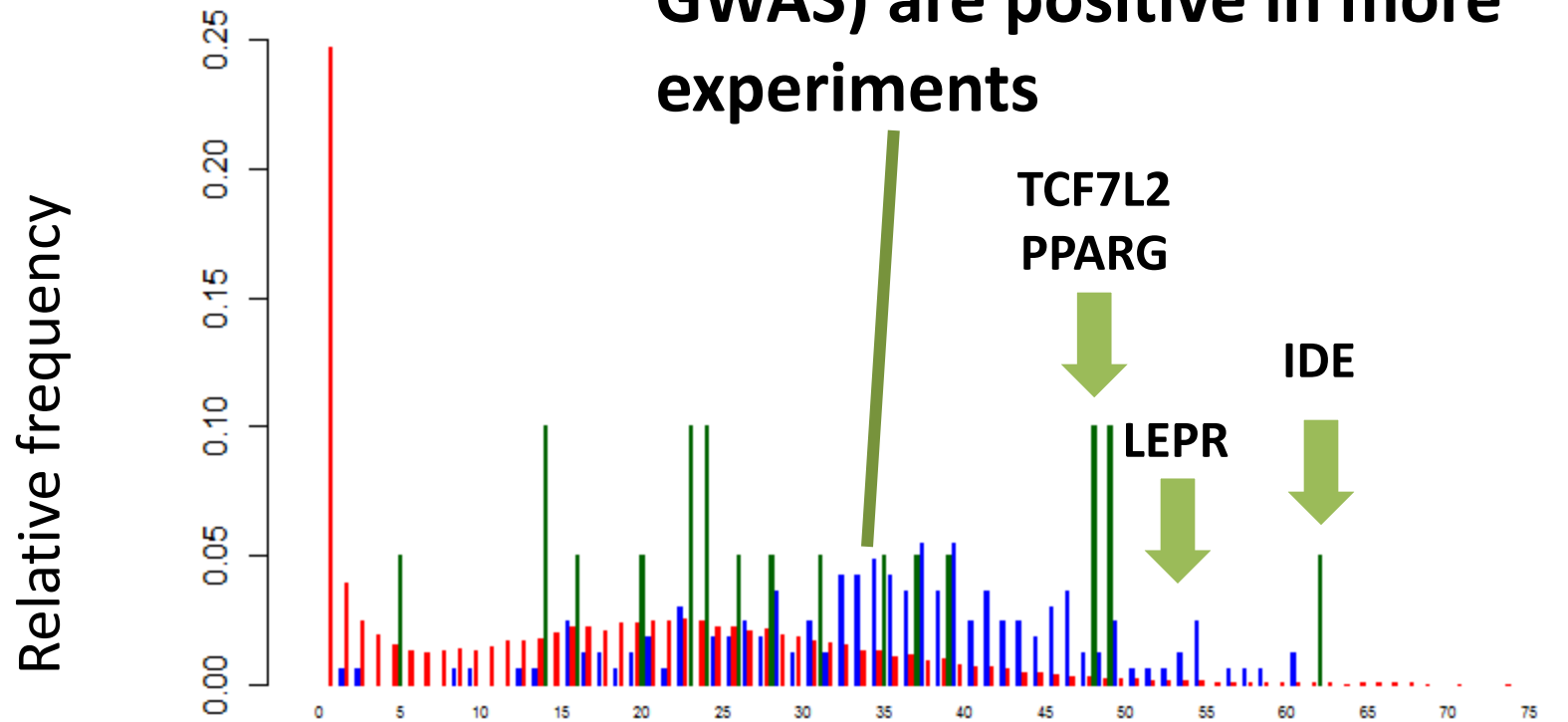
TCF7L2
PPARG

IDE

LEPR

Relative frequency

\# of positive RNA microarray experiments (out of 130)

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Keiichi Kodama

Intersect 130 T2D microarray experiments

# of positive RNA microarray experiments (out of 130)

Close collaboration with Dr. Takashi Kadowaki, Momoko Horikoshi, Kazuo Hara, University of Tokyo

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Keiichi Kodama

Intersect 130 T2D microarray experiments

B

A

Relative frequency

# of positive RNA microarray experiments (out of 130)

Close collaboration with Dr. Takashi Kadowaki, Momoko Horikoshi, Kazuo Hara, University of Tokyo

Keiichi Kodama
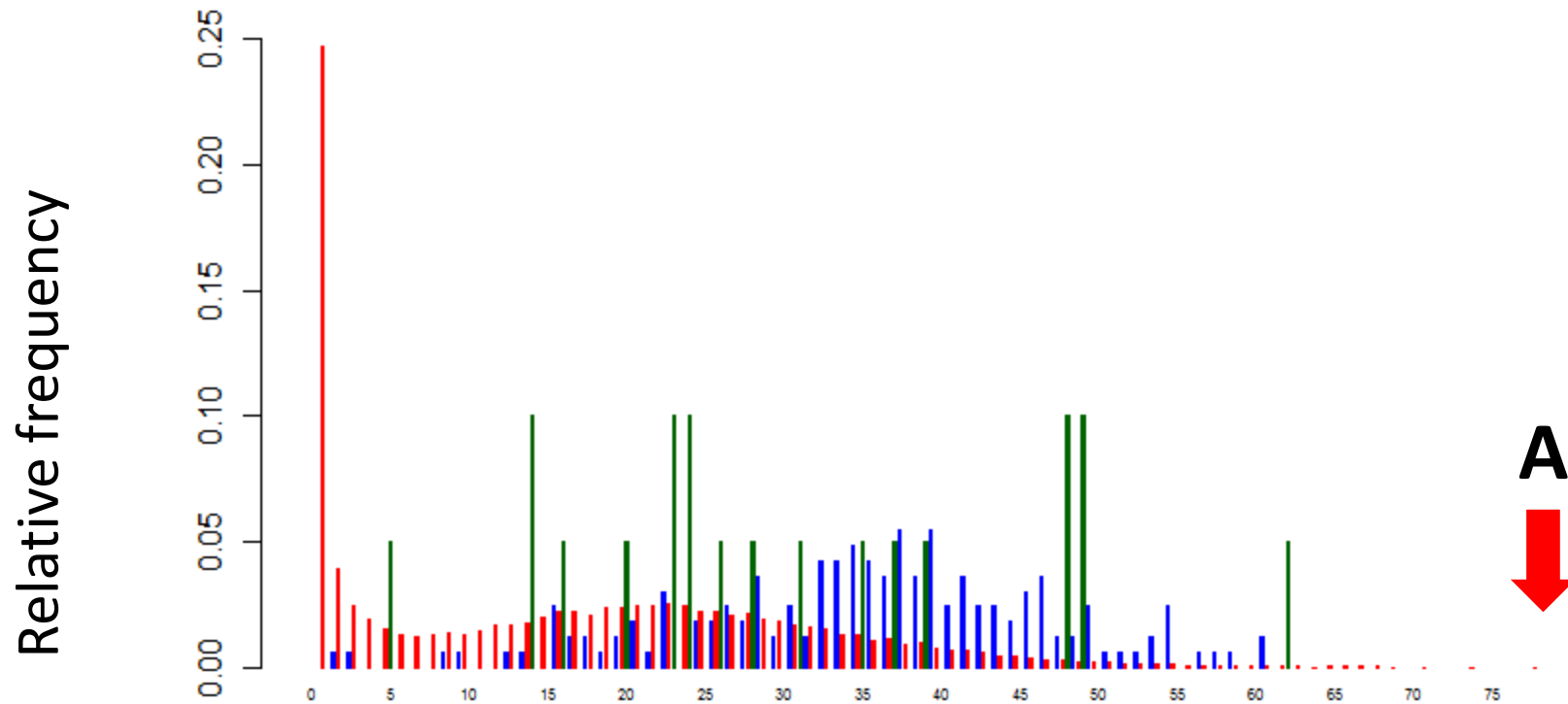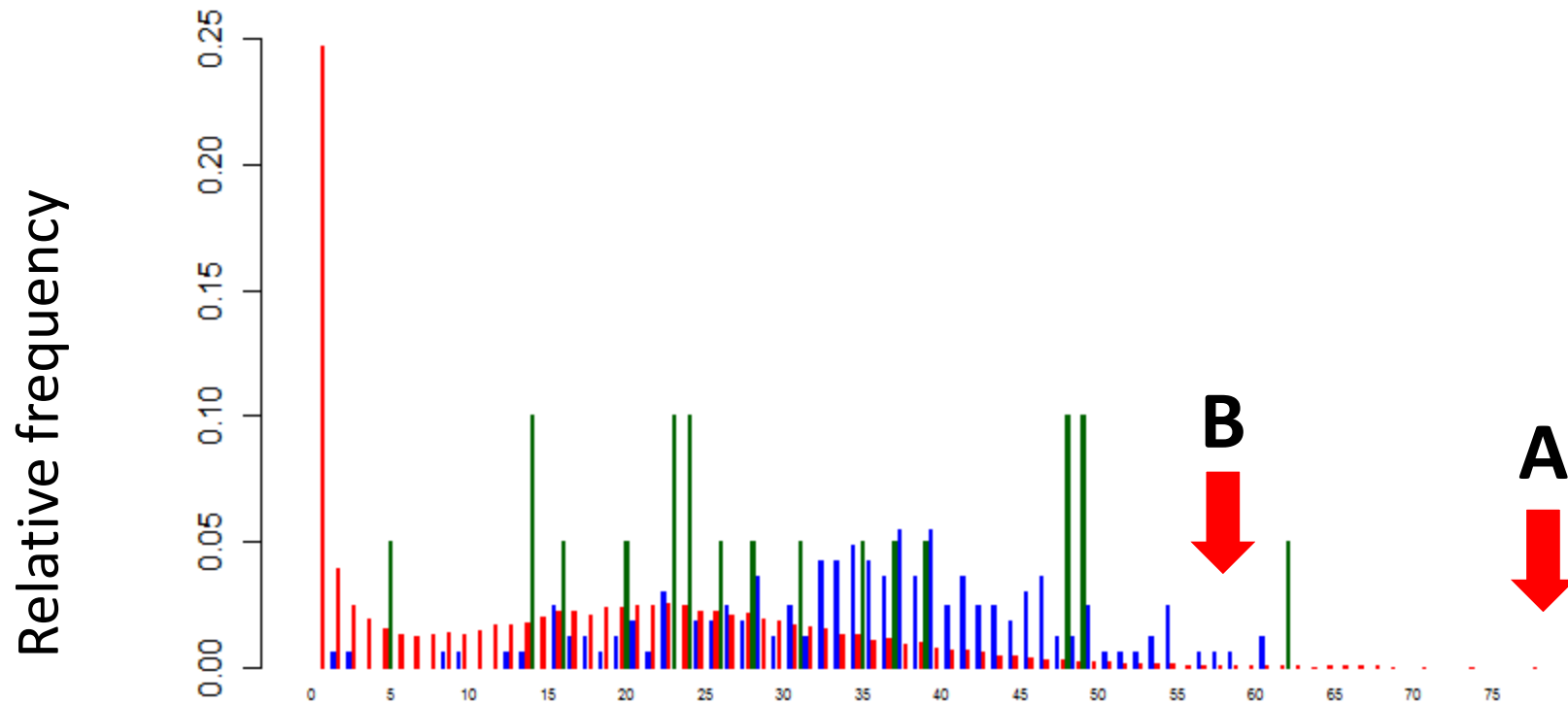
Gene A is higher in high fat diet
Gene A is expressed in mouse fat infiltrate
Gene A and B are co-expressed in fat

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.
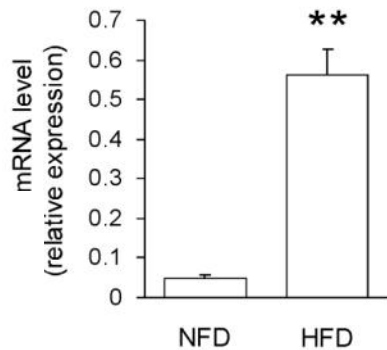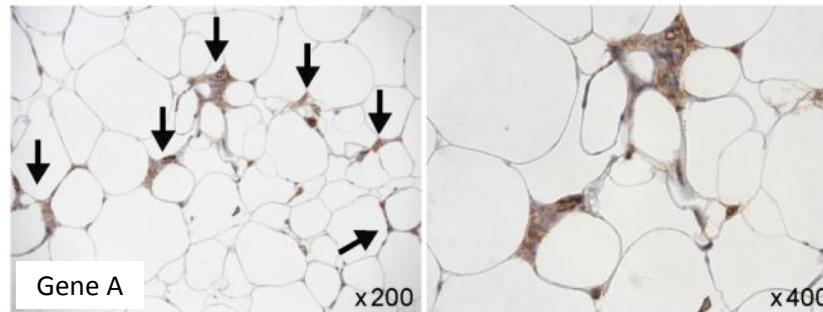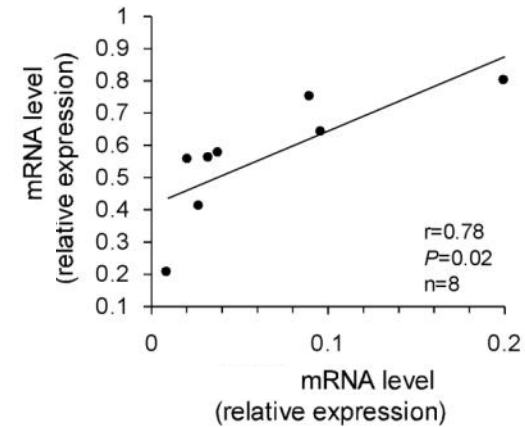
Keiichi Kodama

Kyoko Toda

Gene A knockout has reduced infiltrate in fat

- **Mac-2 stain**

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.
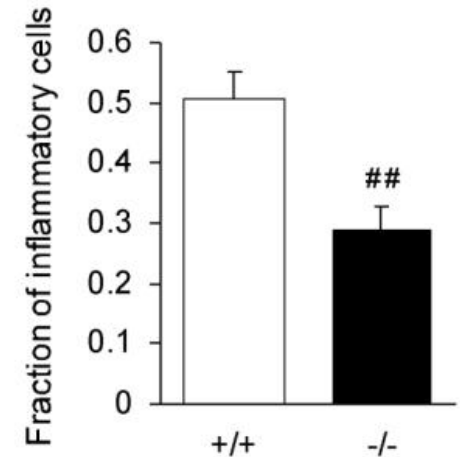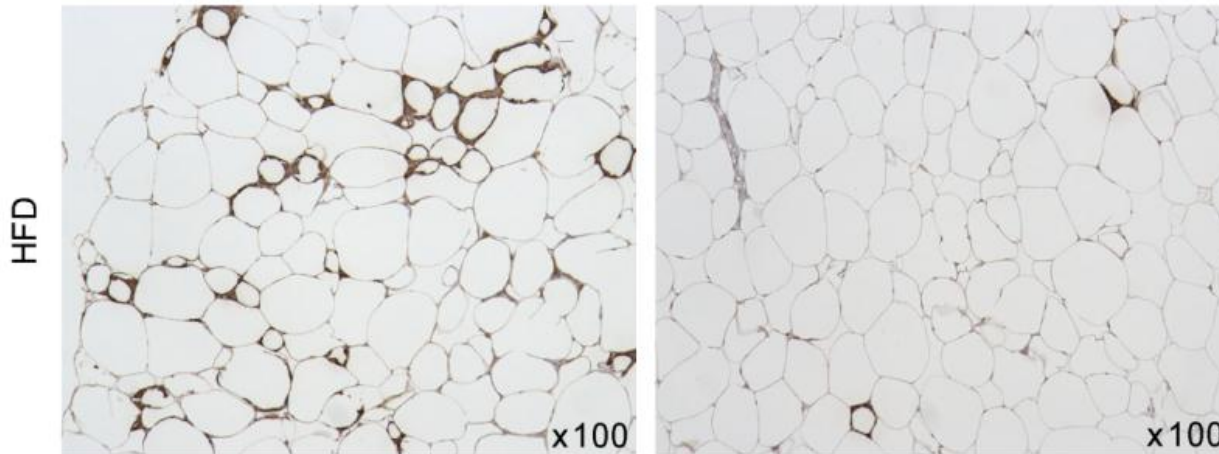
Keiichi Kodama

Kyoko Toda

Gene A knockout has increased insulin sensitivity
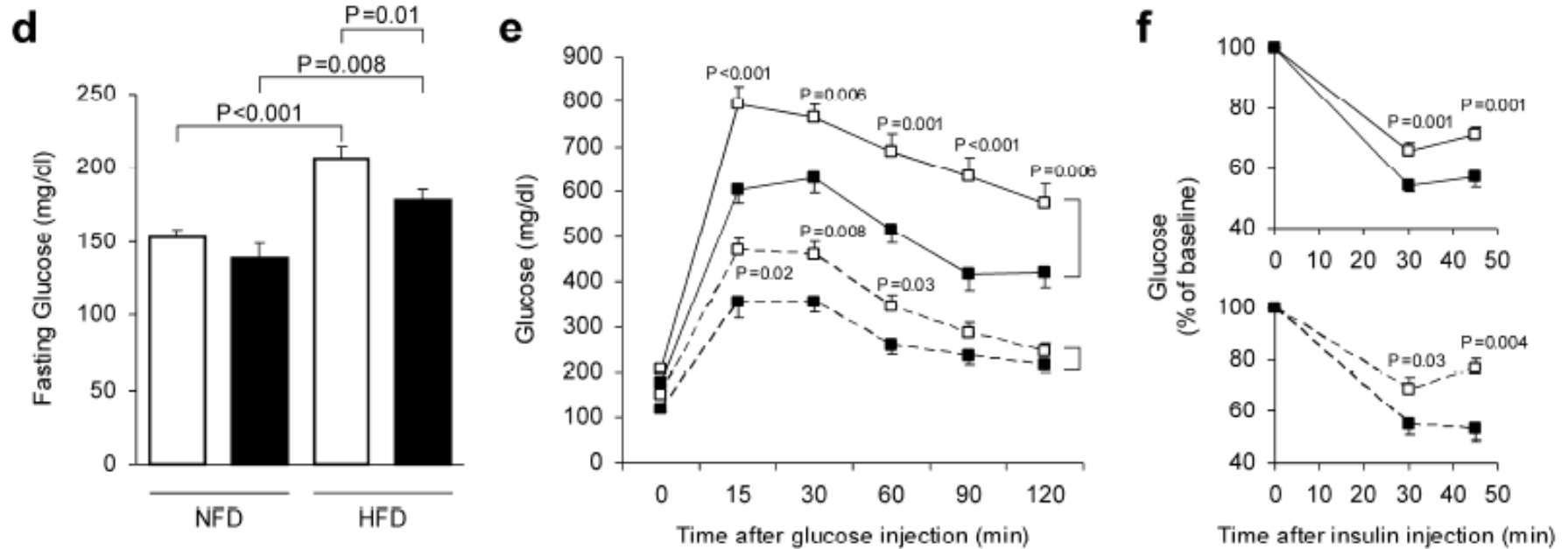
- **No change in weight gain**

Kodama K, Horikoshi M, …, Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Keiichi Kodama

Kyoko Toda

Inflammatory infiltrate in human fat → Protein of Gene A
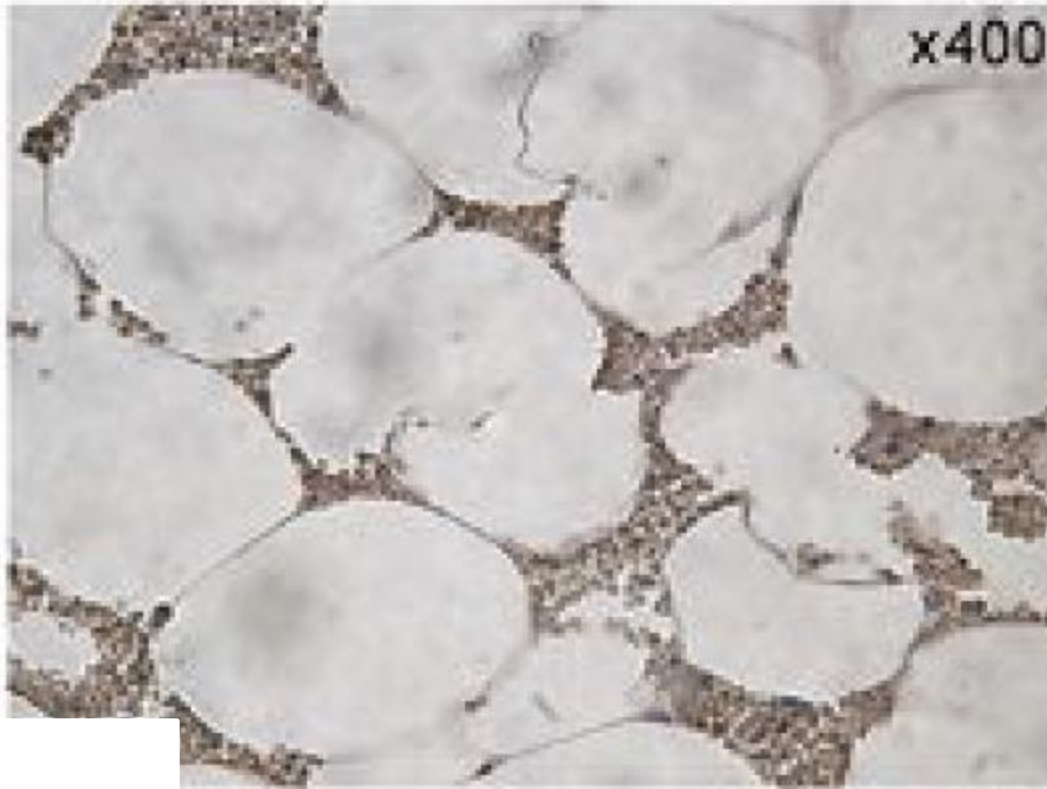
- Paraffin-embedded omental adipose tissue from an obese 57 year woman, BMI 36.9 kg/m^2
- Analyzed for Protein A immunoreactivity

Keiichi Kodama

Serum soluble Gene A protein
correlates with human HbA1c and insulin resistance
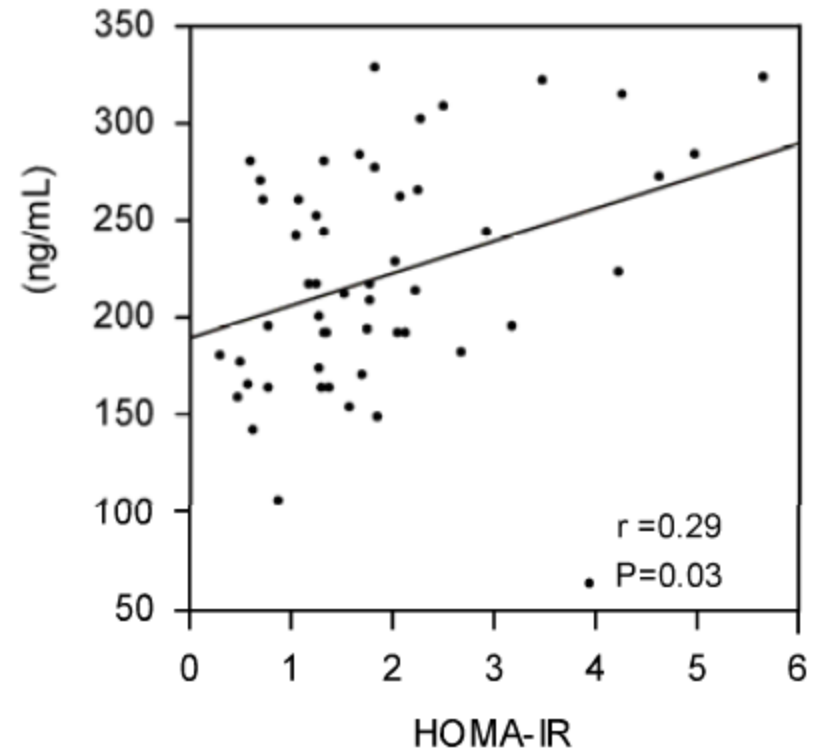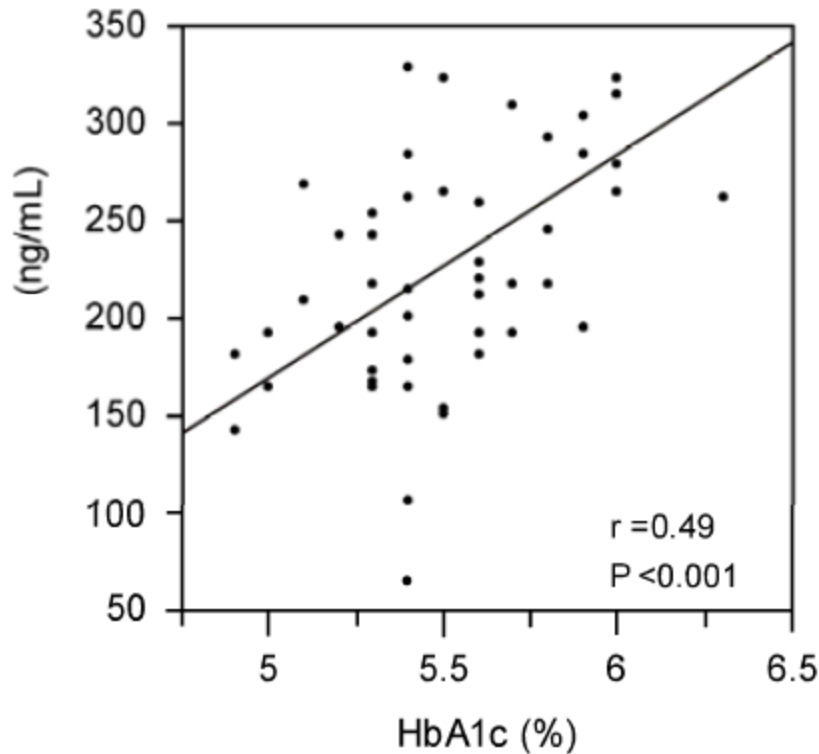
- n = 55 non-diabetics
- 60.3 years of age ± 15, 36 males, 19 females
- BMI 23.2 ± 4.3 kg/m^2

Keiichi Kodama

Momoko Horikoshi

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Therapeutic antibody against Gene A → reduces glucose
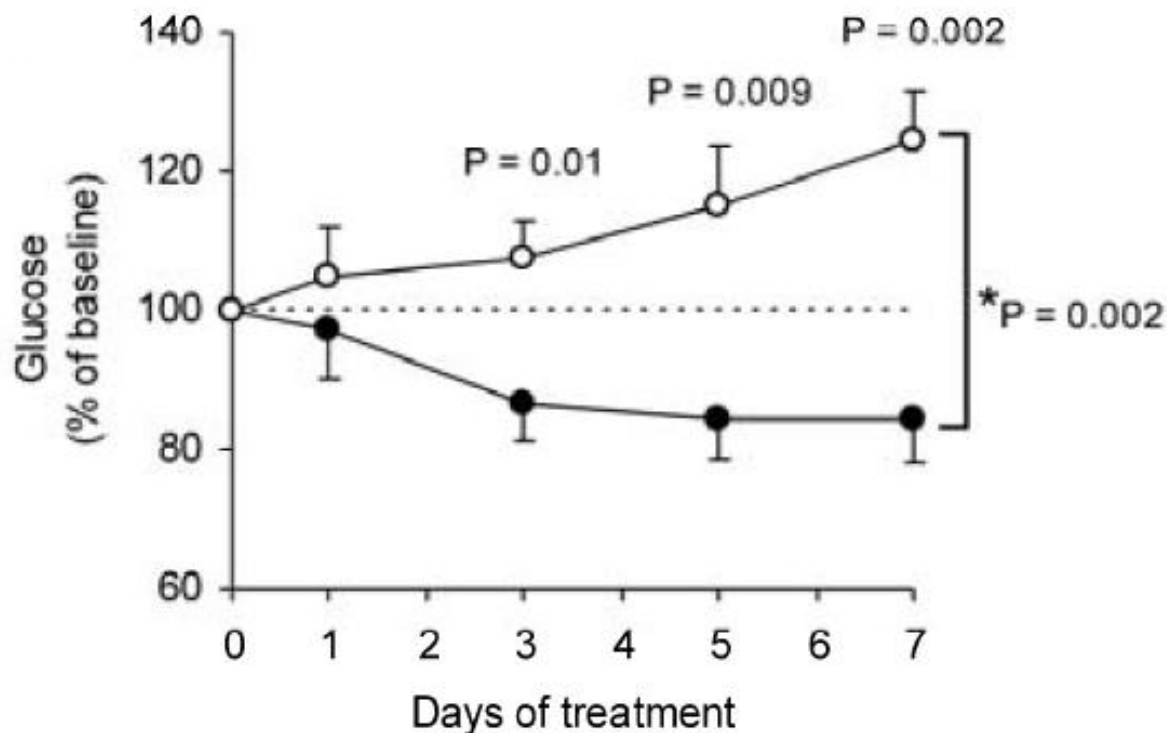
- C57BL6/6J fed high-fat diet for 18 weeks
- Intraperitoneal injection of rat anti-mouse anti-A antibody (n=8) or isotype control (n=8)
- 100 µg at day 0 and 50 µg at day 1-7

Keiichi Kodama

Kodama K, Horikoshi M, ..., Maeda S, Kadowaki T, Butte AJ. *PNAS*, 2012.

Human Disease
Gene Expression Collection

Joel Dudley

~300 Diseases
and Conditions

Blue: gene goes down
in disease
Yellow: gene goes up
in disease

20k+ Genes

Generalized ischemic myocardial dysfunction
Primary idiopathic dilated cardiomyopathy
Pulmonary emphysema
alpha-1-Antitrypsin deficiency
Asthma
Papillary renal cell carcinoma
Renal cell carcinoma, chromophobe cell
Neurofibromatosis type 1
Cocaine dependence
Hantavirus pulmonary syndrome
Marfan's syndrome
Atopy
HIV infection
Retinitis pigmentosa
Ulcerative cystitis
Diabetes mellitus - adult onset
Leprosy
Malignant melanoma
Malignant neoplasm of female breast
Uterine leiomyoma - fibroids
Cystic fibrosis of pancreas
SCID due to absent class II HLA antigens
Morbid obesity
Simple obesity
Critical illness polyneuropathy
Familial combined hyperlipidemia
Hyperglycemia
Hypertensive heart disease with congestive HF
Left ventricular hypertrophy
Salmonella infection
Hepatocellular carcinoma
Chronic airway obstruction
pT2a (IIA) cervical cancer
pT1b (IB) cervical cancer

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.

Disease Individuals

Healthy Controls

Disease Gene Expression Signature

Treated Samples

Untreated Samples

Drug Gene Expression Profile

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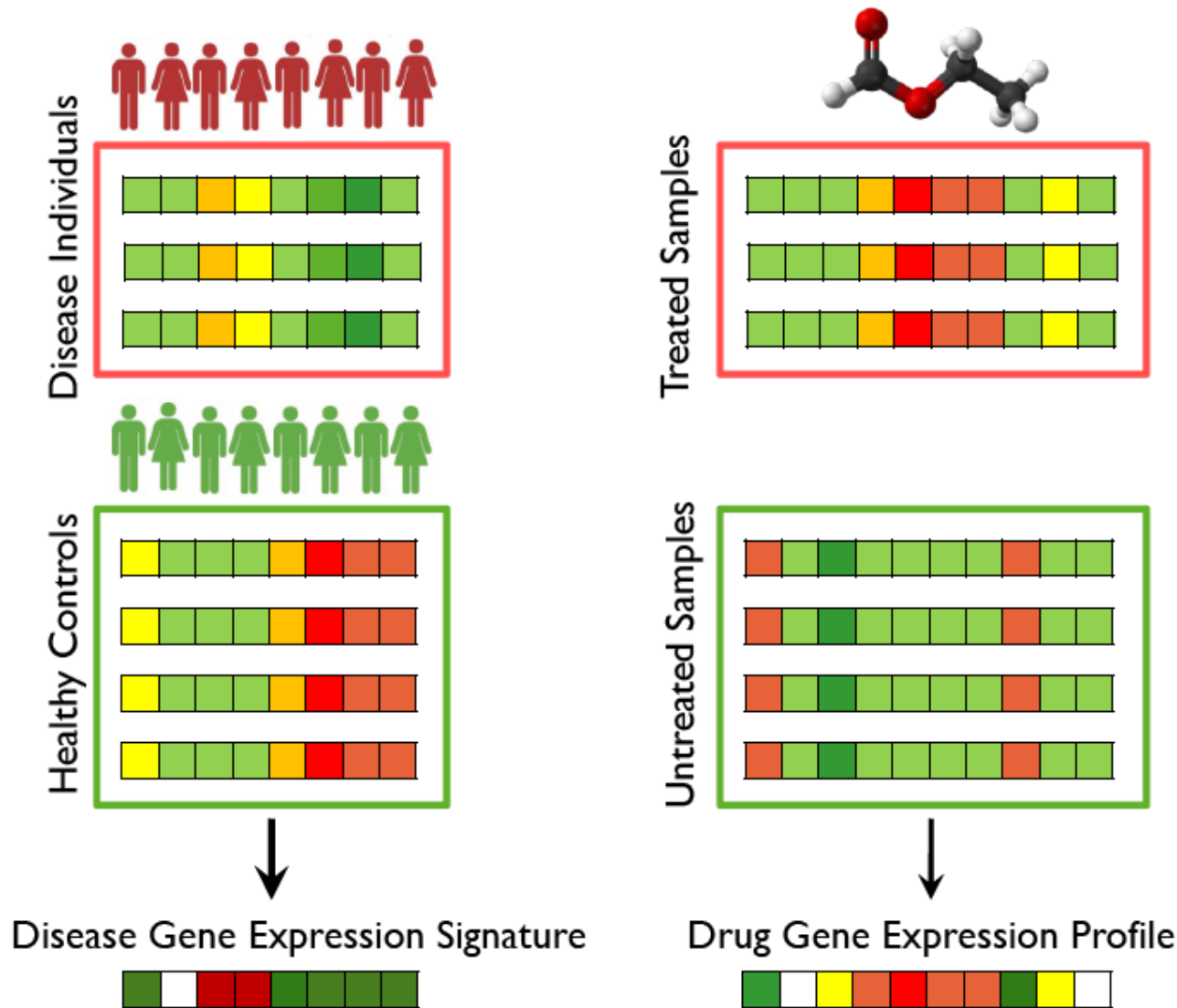
Lamb J, …, Golub TR. *Science*, 2006.
Sirota M, Dudley JT, …, Sweet-Cordero A, Sage J, Butte AJ.
Science Translational Medicine, 2011.

Anti-seizure drug works against a rat model of inflammatory bowel disease
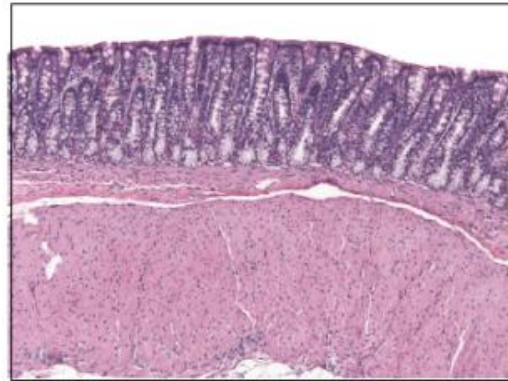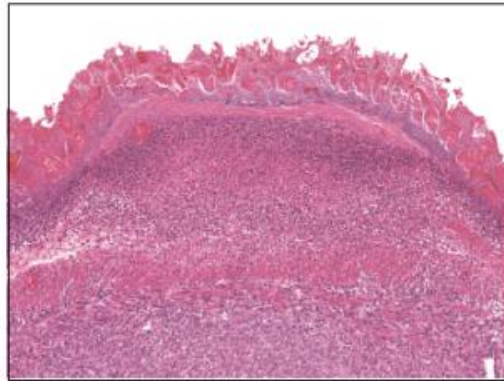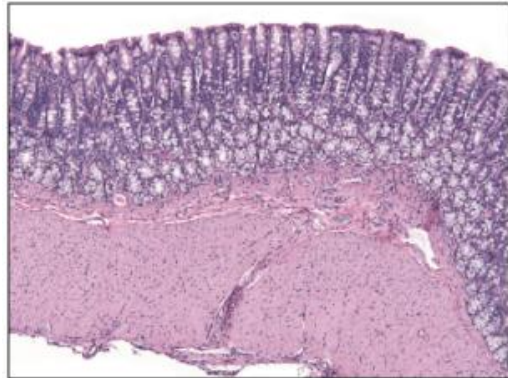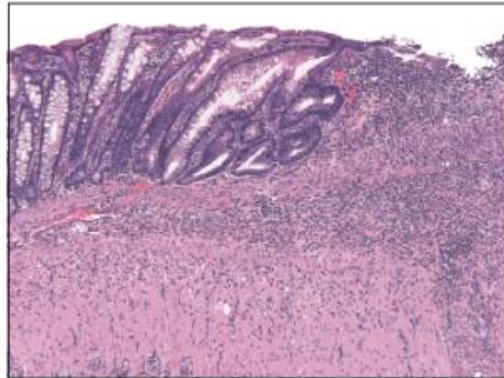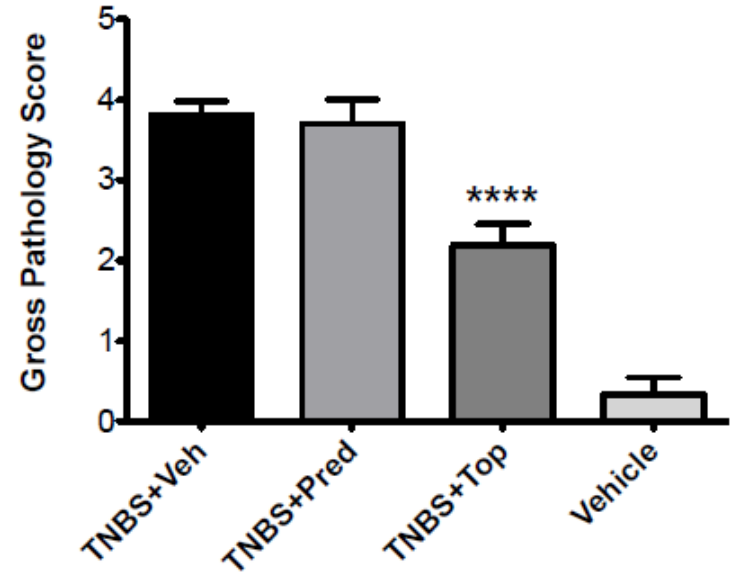
Marina Sirota
Joel Dudley
Mohan M Shenoy
Jay Pasricha

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.

Six Lessons Learned

- Public molecular data has incredible utility
  - All public-funded data should eventually become publicly-available
  - Consider mechanisms to promote secondary uses and computation
- Sufficient high quality data already exists to impact medicine
  - More is better, but no reason to wait for more data
  - Should never wait for perfect data, experiment, conditions
  - Requiring perfection can even slow secondary use
- It's not just about infrastructure, it's about using it
  - Too many tools. Those who build platforms → use them too!
- Sticks seem to work better than carrots
  - Continue exponential growth, more transparency
- Need to train students to initiate science with data
  - High school → higher education → career changers
  - Scaling through students

Medical Entrepreneurship at Stanford

- Help academicians overcome obstacles in **moving research innovations from bench to bedside** for real clinical needs.
- Educate faculty, postdoc fellows and graduate students on translational research process and path to clinic so **developing promising discoveries becomes second nature**

- Accelerate the development of Stanford's top entrepreneurs through **experimental education**. Community for founders.
- 12% of Stanford has applied. 650 companies applied. 10% accepted; of those 85% funded.

- Train students, fellows and faculty in systematic approach to **needs finding and invention** of new technologies (devices).
- Fellowships, classes, mentoring, career services.

- Entrepreneurship center for engineering school. Provides students from all majors with **entrepreneurial skills** needed to use innovations to solve major world problems.
- 2000 free videos and podcasts, featuring entrepreneurship and innovation thought leaders.

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
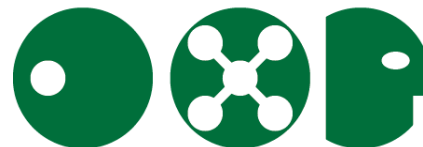- Minnie Sarwal / Nephrology
- David Miklos / Oncology

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

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

Admin and Tech Staff
- Susan Aptekar
- Camilla Morrison
- Alex Skrenchuk

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