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

Atul Butte, MD, PhD
Chief, Division of Systems Medicine,
Department of Pediatrics,
Department of Medicine, and, by courtesy,
Computer Science
Center for Pediatric Bioinformatics, LPCH
Stanford University

abutte@stanford.edu
@atulbutte
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

The data deluge

The Economist

Overload
Global information created and available storage
Exabytes

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

By Chris Anderson 06.23.08

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

The Data Deluge

And How To Handle It. A 14-Page Special Report
Data’s shameful neglect

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 health outcomes. Moreover, the principle 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.
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 public databases, researchers can identify disease trends without ever having to examine samples.

BY MONYA BAKER

Pravesh 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 experiments in parallel, Butte noticed a common feature among samples from type 2 diabetes patients: genes involved in the metabolism of glucose were down-regulated. This led to the discovery of an antibiotic that can normalize blood glucose levels.

That type of discovery is now routine, according to Butte. “We can find interesting genes on the fly,” he says. “We can rapidly pull out any gene that appears interesting. There is no need to go through the literature. We can just find the gene. If it is interesting, we will use it as a starting point 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 database started by the European Bioinformatics Institute in Cambridge.

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

Data sets deposited (millions)

![Graph showing the increase in data sets deposited over time](image_url)
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 »
Total 1.1 million microarrays available
Doubles every 2-3 years

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)

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)
<table>
<thead>
<tr>
<th>Study</th>
<th>Embargo Release</th>
<th>Details</th>
<th>Participants</th>
<th>Type of Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)</td>
<td>Feb 13, 2009</td>
<td>V D A A</td>
<td>1991</td>
<td>Case-control</td>
</tr>
<tr>
<td>Framingham SHARE</td>
<td>Version 1: Oct 19, 2008</td>
<td>V D</td>
<td>14277</td>
<td>Longitudinal</td>
</tr>
<tr>
<td>GAIN: Collaborative Association Study of Psoriasis</td>
<td>Aug 13, 2008</td>
<td>V D A A</td>
<td>2675</td>
<td>Case-control</td>
</tr>
<tr>
<td>GAIN: Genotyping the 270 HapMap samples for GAIN by Broad</td>
<td></td>
<td>V D A A</td>
<td>-</td>
<td>Parent-offspring</td>
</tr>
<tr>
<td>GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen</td>
<td></td>
<td>V D A A</td>
<td>-</td>
<td>Parent-offspring</td>
</tr>
<tr>
<td>GAIN: International Multi-Center ADHD Genetics Project</td>
<td>Mar 26, 2008</td>
<td>V D A A</td>
<td>2635</td>
<td>Parent-offspring</td>
</tr>
<tr>
<td>GAIN: Linking Genome-Wide Association Study of Schizophrenia</td>
<td>Version 1: Nov 07, 2008</td>
<td>V D A A</td>
<td>5068</td>
<td>Case-control</td>
</tr>
<tr>
<td>GAIN: Major Depression: Stage 1 Genome-wide Association in Population-Based Samples</td>
<td>Jul 09, 2008</td>
<td>V D A A</td>
<td>3741</td>
<td>Case-control</td>
</tr>
<tr>
<td>GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes</td>
<td>Jul 09, 2008</td>
<td>V D A A</td>
<td>1625</td>
<td>Case-control</td>
</tr>
<tr>
<td>GAIN: Whole Genome Association Study of Bipolar Disorder</td>
<td>Version 1: Nov 25, 2008</td>
<td>V D A A</td>
<td>3261</td>
<td>Case-control</td>
</tr>
<tr>
<td>GAW16 Framingham and Simulated Data</td>
<td>Oct 19, 2008</td>
<td>V D A A</td>
<td>7139</td>
<td>Longitudinal population-based</td>
</tr>
<tr>
<td>Genome-wide Association Studies in the Hutterites</td>
<td></td>
<td>V D</td>
<td>632</td>
<td>Population-based</td>
</tr>
<tr>
<td>Genome-wide Association Study of Neuroblastoma</td>
<td></td>
<td>V D</td>
<td>1032</td>
<td>Case-control</td>
</tr>
<tr>
<td>Genome-wide Study in Amyotrophic Lateral Sclerosis and Controls: First Stage Analysis</td>
<td>Jun 26, 2008</td>
<td>V D A A</td>
<td>544</td>
<td>Case-control</td>
</tr>
<tr>
<td>Ischemic Stroke Genetics Study (SGS)</td>
<td>Jun 26, 2008</td>
<td>V D A A</td>
<td>495</td>
<td>Case-control</td>
</tr>
<tr>
<td>Mayo-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson's Solutions) Collaboration</td>
<td>Mar 03, 2008</td>
<td>V D A A</td>
<td>1550</td>
<td>Case-control</td>
</tr>
<tr>
<td>NEI Age-Related Eye Disease Study (AREDS)</td>
<td>Jun 11, 2007</td>
<td>V D A A</td>
<td>600</td>
<td>Case-control</td>
</tr>
<tr>
<td>NINDS Parkinson's Disease</td>
<td>Oct 12, 2007</td>
<td>V D A A</td>
<td>535</td>
<td>Case-control</td>
</tr>
<tr>
<td>NINDS Parkinsonism Study</td>
<td>Oct 12, 2007</td>
<td>V D A A</td>
<td>1283</td>
<td>Case-control</td>
</tr>
<tr>
<td>NINDS Repository Cerebrovascular Disease/Stroke Study</td>
<td>Jun 26, 2008</td>
<td>V D A A</td>
<td>870</td>
<td>Case-control</td>
</tr>
<tr>
<td>NINDS Repository Motor Neuron Disease/ALS Study</td>
<td>Jun 26, 2008</td>
<td>V D A A</td>
<td>1790</td>
<td>Case-control</td>
</tr>
<tr>
<td>NINDS Repository Neurologically Normal Control Collection</td>
<td>Oct 12, 2007</td>
<td>V D A A</td>
<td>2723</td>
<td>Control-s</td>
</tr>
<tr>
<td>POPRES: Population Reference Sample</td>
<td></td>
<td>V D</td>
<td>5819</td>
<td>Population-s</td>
</tr>
<tr>
<td>SEARCH GWA Study of Statin-Induced Myopathy</td>
<td></td>
<td>V D A A</td>
<td>175</td>
<td>Case-control</td>
</tr>
<tr>
<td>Study of Irish Amyotrophic Lateral Sclerosis (SIALS)</td>
<td></td>
<td>V D A A</td>
<td>432</td>
<td>Case-control</td>
</tr>
<tr>
<td>The Finland-United States Investigation of NIDDM Genetics (FUSION) study</td>
<td></td>
<td>V D A A</td>
<td>2335</td>
<td>Case-control</td>
</tr>
<tr>
<td>Whole Genome Association Study of Systemic Lupus Erythematosus</td>
<td></td>
<td>V D A A</td>
<td>4651</td>
<td>Case-control</td>
</tr>
<tr>
<td>Study</td>
<td>Embargo Release</td>
<td>Details</td>
<td>Participants</td>
<td>Type of Study</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-----------------</td>
<td>---------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD)</td>
<td>Feb 13, 2009</td>
<td></td>
<td>1991</td>
<td>Case-control</td>
</tr>
</tbody>
</table>
Version 2: Feb 01, 2009  
Version 3: Jul 06, 2009   |         | 14277      | Longitudinal  |
| GAIN: Collaborative Association Study of Psoriasis                  | Aug 13, 2008     |         | 2675         | Parent-offspring |
| GAIN: Genotyping the 270 HapMap samples for GAIN by Broad            |                 |         | -            | Parent-offspring |
| GAIN: Genotyping the 270 HapMap samples for GAIN by Perlegen         |                 |         | -            | Parent-offspring |
| GAIN: International Multi-Center ADHD Genetics Project              | Mar 26, 2008     |         | 2635         | Case-control   |
Version 2: Dec 03, 2008   |         | 5060         | Case-control   |
| GAIN: Major Depression: Stage 1 Genomewide Association in Population-Based Samples | Jul 09, 2008 |         | 3741         | Case-control   |
| GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes | Jul 09, 2008 |         | 1625         | Case-control   |
| GAIN: Whole Genome Association Study of Bipolar Disorder            | Version 1: Nov 25, 2009  
Version 2: Dec 01, 2008   |         | 3261         | Case-control   |
| GAW16 Framingham and Simulated Data                                 | Oct 19, 2008     |         | 7139         | Longitudinal population-based |
| Genome-wide Association Studies in the Hutterites                   |                 |         | 632          | Population-based |
| Genome-wide Association Study of Neuroblastoma                      |                 |         | 1032         | Case-control   |
| Genome-wide Study in Amyotrophic Lateral Sclerosis and Controls: First Stage Analysis | Jun 26, 2008 |         | 544          | Case-control   |
| Ischemic Stroke Genetics Study (SGS)                                | Jun 26, 2008     |         | 495          | Case-control   |
| Nano-Perlegen LEAPS (Linked Efforts to Accelerate Parkinson’s Solutions) Collaboration | Mar 03, 2008 |         | 1550         | Case-control   |
| NEI: Age-Related Eye Disease Study (AREDS)                         | Jun 11, 2007     |         | 600          | Case-control   |
| NINDS Parkinson's Disease                                           | Oct 12, 2007     |         | 535          | Case-control   |
| NINDS Parkinsonism Study                                            | Oct 12, 2007     |         | 1283         | Case-control   |
| NINDS Repository Cerebrovascular Disease/Stroke Study               | Jun 26, 2008     |         | 870          | Case-control   |
| NINDS Repository Motor Neuron Disease/ALS Study                     | Jun 26, 2008     |         | 1790         | Case-control   |
| NINDS Repository Neurologically Normal Control Collection           | Oct 12, 2007     |         | 2723         | Control-sample |
| POPRES: Population Reference Sample                                |                 |         | 5819         | Population-sample |
| SEARCH GWA Study of Statin-Induced Myopathy                        |                 |         | 175          | Case-control   |
| Study of Irish Amyotrophic Lateral Sclerosis (SIALS)                |                 |         | 432          | Case-control   |
| The Finland-United States Investigation of NIDDM Genetics (FUSION) study |                 |         | 2335         | Case-control   |
| Whole Genome Association Study of Systemic Lupus Erythematosus       |                 |         | 4651         | Case-control   |
108 million substances x 650,000 assays
1 billion points of data within a grid of 70 trillion cells
browse by disease

--- A ---
› Anal Cancer
› Anemia
› Asthma

--- B ---
› Bladder Cancer
› Brain Cancer
› Breast Cancer

--- C ---
› Carcinoid
› Cervical Cancer
› Chronic Obstructive Pulmonary Disease

--- D ---

--- E ---

--- F ---

--- G ---

--- H ---

--- I ---
› Idiopathic Pulmonary Fibrosis

--- K ---
› Kidney Cancer

--- L ---
› Leukemia
› Liver Cancer
› Lung Cancer

--- M ---
› Melanoma
› Monoclonal Gamopathy

--- N ---

--- O ---

--- P ---

--- Q ---

--- R ---
› Rheumatoid Arthritis

--- S ---
› Sarcoidosis
› Scleroderma
› Systemic Lupus Erythematosus

--- T ---
› Testicular Cancer

--- U ---
› Uterine Cancer
Leukemia

**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)

21 Items

<table>
<thead>
<tr>
<th>Item Description</th>
<th>SKU</th>
<th>Quantity</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone Marrow</td>
<td>B Cells, Negative Selection</td>
<td>BMA-BCE-LE</td>
<td>15 Items Per Page</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>B Cells, CD19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>T Cells, CD3</td>
<td>BMA-CD3-LE</td>
<td>$500.00</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>CD45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>Fresh</td>
<td>BMA-FRE-LE</td>
<td>$2,500.00</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>Mononuclear Cells</td>
<td>BMA-MON-LE</td>
<td>$750.00</td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>Special Processing</td>
<td>BMA-SPE-LE</td>
<td></td>
</tr>
<tr>
<td>Bone Marrow</td>
<td>T Cells, Negative Selection</td>
<td>BMA-TCE-LE</td>
<td></td>
</tr>
<tr>
<td>Product Description</td>
<td>SKU</td>
<td>Price</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>----------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>Mononuclear Cells</td>
<td>Leukemia</td>
<td>SKU: PBL-MON-LE</td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>Plasma</td>
<td>Leukemia</td>
<td>SKU: PBL-PLA-LE</td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>Serum</td>
<td>Leukemia</td>
<td>SKU: PBL-SER-LE</td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>Special Processing</td>
<td>Leukemia</td>
<td>SKU: PBL-SPE-LE</td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>T Cells, Negative Selection</td>
<td>Leukemia</td>
<td>SKU: PBL-TCE-LE</td>
</tr>
<tr>
<td>Peripheral Blood</td>
<td>Viable Plated Cells</td>
<td>Leukemia</td>
<td>SKU: PBL-VPC-LE</td>
</tr>
</tbody>
</table>
### Summary of Available Inventory

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unique Patients</th>
<th>Available Samples</th>
<th>Unique Patients</th>
<th>Available Samples</th>
<th>Unique Patients</th>
<th>Available Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>31</td>
<td>498</td>
<td>11</td>
<td>92</td>
<td>31</td>
<td>210</td>
</tr>
<tr>
<td>Male</td>
<td>31</td>
<td>498</td>
<td>11</td>
<td>92</td>
<td>31</td>
<td>210</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>996</td>
<td>22</td>
<td>184</td>
<td>62</td>
<td>420</td>
</tr>
</tbody>
</table>

### Available Inventory by Patient ID

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age</th>
<th>Race</th>
<th>Sex</th>
<th>Alcohol Use</th>
<th>Tobacco Use</th>
<th>Drugs Administered</th>
<th>500ul Plasma (K2EDTA Tubes)</th>
<th>500ul Plasma (Lithium Heparin Tubes)</th>
<th>300ul Buffy Coat (K2EDTA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12416CFED</td>
<td>73</td>
<td>White</td>
<td>M</td>
<td>Current Use</td>
<td>Previous Use</td>
<td>Cyclophosphamide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dexamethasone Sodium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Docetaxel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leuprolide Acetate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Paclitaxel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pegfilgrastim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Zoledronic Acid</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Go to www.openbiosystems.com to place an order, or call 1-888-412-2225 to speak directly with a representative.
Peripheral Blood | Plasma | Leukemia
SKU: PBL-PLA-LE
$55.00

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

# of Samples per Patient

<table>
<thead>
<tr>
<th>Number</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$55.00</td>
</tr>
<tr>
<td>2</td>
<td>$110.00</td>
</tr>
<tr>
<td>3</td>
<td>$165.00</td>
</tr>
<tr>
<td>4</td>
<td>$220.00</td>
</tr>
<tr>
<td>5</td>
<td>$275.00</td>
</tr>
<tr>
<td>6</td>
<td>$330.00</td>
</tr>
<tr>
<td>7</td>
<td>$385.00</td>
</tr>
</tbody>
</table>

Disease subtype

<table>
<thead>
<tr>
<th>Disease subtype</th>
<th>Selected</th>
</tr>
</thead>
</table>

Units

<table>
<thead>
<tr>
<th>Units</th>
<th>Selected</th>
</tr>
</thead>
</table>

add to cart

Overview

Product Reviews

Additional Information

<table>
<thead>
<tr>
<th>SKU</th>
<th>PBL-PLA-LE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Status</td>
<td>Any Treatment, Pre Treatment, Post Treatment, Active Treatment, Recurrent/Refractory Disease, Remission</td>
</tr>
</tbody>
</table>
All the Services you need
Researchers around the world use Assay Depot to find the latest research services
Take the Tour

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

biology  chemistry  dmpk  pharmacology  toxicology
<table>
<thead>
<tr>
<th>Assay Depots</th>
<th>Pharmacology Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bone Models</strong></td>
<td>Bone Metastases</td>
</tr>
<tr>
<td><strong>Cardiovascular Models</strong></td>
<td>Atrial Arrhythmias</td>
</tr>
<tr>
<td><strong>Dermatology Models</strong></td>
<td>Acne</td>
</tr>
<tr>
<td><strong>Diabetes Models</strong></td>
<td>BB/W Rats</td>
</tr>
<tr>
<td><strong>Genitourinary Models</strong></td>
<td>Chronic Kidney Disease</td>
</tr>
<tr>
<td><strong>In Vitro Models</strong></td>
<td>In Vitro Bone Models</td>
</tr>
<tr>
<td><strong>In Vivo Technologies</strong></td>
<td>Cognition</td>
</tr>
<tr>
<td><strong>Infectious Disease</strong></td>
<td>Bacterial Infection</td>
</tr>
<tr>
<td><strong>Inflammation Models</strong></td>
<td>Arthritis</td>
</tr>
<tr>
<td><strong>Neurological Models</strong></td>
<td>Alzheimer's Disease</td>
</tr>
<tr>
<td><strong>Oncology Models</strong></td>
<td>Angiogenesis</td>
</tr>
<tr>
<td><strong>Ophthalmic Models</strong></td>
<td>Cataract</td>
</tr>
<tr>
<td><strong>Otology Models</strong></td>
<td>Hearing Loss</td>
</tr>
<tr>
<td><strong>Pain Models</strong></td>
<td>General Pain</td>
</tr>
<tr>
<td><strong>Respiratory Models</strong></td>
<td>Ascaris Lung Allergy</td>
</tr>
</tbody>
</table>
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
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.
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. 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.
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
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

<table>
<thead>
<tr>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exanthematic</td>
<td>Feverish, with skin eruptions</td>
</tr>
<tr>
<td>Critical</td>
<td>Feverish, with urinary problems</td>
</tr>
<tr>
<td>Phlogistic</td>
<td>Feverish, with heavy pulse and topical pain</td>
</tr>
<tr>
<td>Dolorous</td>
<td>Painful</td>
</tr>
<tr>
<td>Mental</td>
<td>With alienation of judgment</td>
</tr>
<tr>
<td>Quietal</td>
<td>With loss of movement</td>
</tr>
<tr>
<td>Motor</td>
<td>With involuntary motion</td>
</tr>
<tr>
<td>Suppressorial</td>
<td>With impeded motions</td>
</tr>
<tr>
<td>Evacuatorial</td>
<td>With evacuation of liquids</td>
</tr>
<tr>
<td>Deformities</td>
<td>Changed appearance of solid parts</td>
</tr>
<tr>
<td>Blemishes</td>
<td>External and palpable</td>
</tr>
</tbody>
</table>

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
Human Disease Gene Expression Collection

~300 Diseases and Conditions

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

20k+ Genes

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

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

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

Marina Sirota
Joel Dudley
Mohan M Shenoy
Jay Pasricha
Rat colonoscopy

Rat with Inflammatory Bowel Disease

Inflammatory Bowel Disease After Anti-seizure Drug

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

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

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.
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
- $\sim$1000 genomes: Illumina, Ion Torrent
- Complete Genomics: towards 80 genomes/day
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.
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 of 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.1184198
Single-molecule sequencing of an individual human genome

Dmitry Pushkarev¹,², Norma F Neff¹,² & Stephen R Quake¹

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 28x 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 principle¹¹–¹³, 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 genome¹², 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 the first human genome using a calibration of sequencing from the first to the 27th read position.
Clinical assessment incorporating a personal genome


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.
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
Variants predisposing to cardiac risk

- 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

## 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

<table>
<thead>
<tr>
<th>Gene name</th>
<th>SNP location</th>
<th>Patient genotype</th>
<th>Drug(s) affected</th>
<th>Summary of effects</th>
<th>Level of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC01B1 Solute carrier organic anion transporter family, member 1B1</td>
<td>rs4149056</td>
<td>T/T</td>
<td>HMG-CoA reductase inhibitors (statins)</td>
<td>No increased risk of myopathy</td>
<td>High[33][34]</td>
</tr>
<tr>
<td>CYP2C19 Cytochrome P450, family 2, subfamily C, polypeptide 19</td>
<td>rs4244285</td>
<td>A/G</td>
<td>Clopidogrel and CYP2C19 substrates</td>
<td>CYP2C19 poor metaboliser; many drugs might need adjustment</td>
<td>High[35]</td>
</tr>
<tr>
<td>VKORC1 Vitamin K epoxide reductase complex, subunit 1</td>
<td>rs9932331</td>
<td>C/T</td>
<td>Warfarin</td>
<td>Reduced dose needed</td>
<td>High[36]</td>
</tr>
<tr>
<td>CYP4F2 Cytochrome P450, family 4, subfamily F, polypeptide 2</td>
<td>rs2108622</td>
<td>C/C</td>
<td>Warfarin</td>
<td>Reduced dose needed</td>
<td>High[37]</td>
</tr>
<tr>
<td>ADRB1 β1 adrenergic receptor</td>
<td>rs1801252</td>
<td>A/A</td>
<td>Atenolol, metoprolol</td>
<td>Might be preferable to calcium-channel blockers</td>
<td>High[38][39]</td>
</tr>
<tr>
<td>SLC01B1 Solute carrier organic anion transporter family, member 1B1</td>
<td>rs11045819</td>
<td>A/C</td>
<td>Fluvastatin</td>
<td>Good response</td>
<td>Medium[40]</td>
</tr>
<tr>
<td>HMGCR HMG-CoA reductase</td>
<td>rs1723540</td>
<td>T/T</td>
<td>Pravastatin</td>
<td>Patient might have good response</td>
<td>Medium[41]</td>
</tr>
<tr>
<td>HMGCR HMG-CoA reductase</td>
<td>rs17244841</td>
<td>A/A</td>
<td>Pravastatin, simvastatin</td>
<td>No reduced efficacy</td>
<td>Medium[42]</td>
</tr>
<tr>
<td>ADRB2 β2 adrenergic receptor, surface</td>
<td>rs1042713</td>
<td>A/G</td>
<td>β blockers</td>
<td>Other treatment options might be preferable</td>
<td>Medium[43]</td>
</tr>
<tr>
<td>CYP2D6 Cytochrome P450, family 2, subfamily D, polypeptide 6</td>
<td>rs3892097</td>
<td>C/C</td>
<td>Metoprolol and other CYP2D6 substrates</td>
<td>Normal CYP2D6 metaboliser</td>
<td>Medium[44]</td>
</tr>
<tr>
<td>CDKN2A/B Cyclin-dependent kinase inhibitor 2A/28</td>
<td>rs10811661</td>
<td>T/T</td>
<td>Metformin</td>
<td>Reduced likelihood of response</td>
<td>Medium[45]</td>
</tr>
<tr>
<td>CDKN2A/B Cyclin-dependent kinase inhibitor 2A/28</td>
<td>rs10811661</td>
<td>T/T</td>
<td>Troglitazone</td>
<td>Reduced likelihood of response</td>
<td>Medium[46]</td>
</tr>
</tbody>
</table>

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**


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
• 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

<table>
<thead>
<tr>
<th>Number of papers curated</th>
<th>Distinct SNPs</th>
<th>Diseases and phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>~11,250</td>
<td>~192,000</td>
<td>~4,400</td>
</tr>
</tbody>
</table>

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...

Post-test probability is calculated with likelihood ratio

Pre-test odds $\times$ likelihood ratio $\rightarrow$ Post-test odds

Pre-test odds $\times$ LR1 $\times$ LR2 $\times$ LR3 $\rightarrow$ Post-test odds

Can chain likelihood ratios from independent tests


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.

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.


<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP location</th>
<th>Patient genotype</th>
<th>LR</th>
<th>Studies†</th>
<th>Samples‡</th>
<th>Post-test probability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOMM40</td>
<td>rs157581</td>
<td>CT</td>
<td>1.6</td>
<td>6</td>
<td>7740</td>
<td>13.90%</td>
</tr>
<tr>
<td>DAPK1</td>
<td>rs487104</td>
<td>TT</td>
<td>0.7</td>
<td>5</td>
<td>10397</td>
<td>10.19%</td>
</tr>
<tr>
<td>TRAK2</td>
<td>rs13022344</td>
<td>CT</td>
<td>1.0</td>
<td>4</td>
<td>6512</td>
<td>10.12%</td>
</tr>
<tr>
<td>DAPK1</td>
<td>rs4877365</td>
<td>AA</td>
<td>0.6</td>
<td>4</td>
<td>4841</td>
<td>5.89%</td>
</tr>
<tr>
<td>E8F3</td>
<td>rs11016976</td>
<td>TT</td>
<td>1.0</td>
<td>3</td>
<td>5736</td>
<td>5.87%</td>
</tr>
<tr>
<td>TNK1</td>
<td>rs1554948</td>
<td>AA</td>
<td>0.9</td>
<td>3</td>
<td>5736</td>
<td>5.32%</td>
</tr>
<tr>
<td>MYH13</td>
<td>rs2074877</td>
<td>CT</td>
<td>1.0</td>
<td>3</td>
<td>5366</td>
<td>5.55%</td>
</tr>
<tr>
<td>GALP</td>
<td>rs3745833</td>
<td>CC</td>
<td>0.9</td>
<td>3</td>
<td>5366</td>
<td>4.82%</td>
</tr>
<tr>
<td>PCK1</td>
<td>rs8192708</td>
<td>AA</td>
<td>0.9</td>
<td>3</td>
<td>5304</td>
<td>4.47%</td>
</tr>
<tr>
<td></td>
<td>rs1859849</td>
<td>TT</td>
<td>0.9</td>
<td>3</td>
<td>5248</td>
<td>4.02%</td>
</tr>
<tr>
<td></td>
<td>rs11622883</td>
<td>AT</td>
<td>1.0</td>
<td>3</td>
<td>2545</td>
<td>3.97%</td>
</tr>
<tr>
<td>WWC1</td>
<td>rs17070145</td>
<td>CC</td>
<td>0.9</td>
<td>3</td>
<td>2545</td>
<td>3.65%</td>
</tr>
<tr>
<td>LMNA</td>
<td>rs505058</td>
<td>TT</td>
<td>1.0</td>
<td>2</td>
<td>4646</td>
<td>3.49%</td>
</tr>
<tr>
<td>ACAN</td>
<td>rs2882676</td>
<td>CC</td>
<td>0.9</td>
<td>2</td>
<td>4590</td>
<td>3.22%</td>
</tr>
<tr>
<td>PGBD1</td>
<td>rs3800324</td>
<td>GG</td>
<td>0.6</td>
<td>2</td>
<td>4590</td>
<td>2.11%</td>
</tr>
<tr>
<td>GOLM1</td>
<td>rs10868366</td>
<td>GG</td>
<td>1.1</td>
<td>2</td>
<td>2156</td>
<td>2.30%</td>
</tr>
<tr>
<td>GOLM1</td>
<td>rs7019241</td>
<td>CC</td>
<td>1.1</td>
<td>2</td>
<td>2156</td>
<td>2.49%</td>
</tr>
<tr>
<td></td>
<td>rs9886784</td>
<td>CC</td>
<td>0.9</td>
<td>2</td>
<td>2156</td>
<td>2.36%</td>
</tr>
<tr>
<td></td>
<td>rs10519262</td>
<td>GG</td>
<td>0.9</td>
<td>2</td>
<td>2156</td>
<td>2.22%</td>
</tr>
<tr>
<td></td>
<td>rs463946</td>
<td>CG</td>
<td>0.5</td>
<td>2</td>
<td>1922</td>
<td>1.04%</td>
</tr>
<tr>
<td>PLAU</td>
<td>rs2227564</td>
<td>CT</td>
<td>0.9</td>
<td>2</td>
<td>956</td>
<td>0.98%</td>
</tr>
<tr>
<td>ADAM12</td>
<td>rs1278279</td>
<td>GG</td>
<td>1.2</td>
<td>1</td>
<td>2320</td>
<td>1.23%</td>
</tr>
<tr>
<td>SORL1</td>
<td>rs2070045</td>
<td>GT</td>
<td>1.1</td>
<td>1</td>
<td>2031</td>
<td>1.36%</td>
</tr>
<tr>
<td>ABCA1</td>
<td>rs2230806</td>
<td>CT</td>
<td>1.1</td>
<td>1</td>
<td>1691</td>
<td>1.50%</td>
</tr>
<tr>
<td>PSEN1</td>
<td>rs165932</td>
<td>GT</td>
<td>0.9</td>
<td>1</td>
<td>170</td>
<td>1.37%</td>
</tr>
</tbody>
</table>

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?
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 ≥ 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
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