Beyond the Human Genome Project
Biology as Information
Human Genome Project
Human Genome Project

- Genetic map: Polymorphic landmarks to trace inheritance
- Physical map: Clones covering all chromosomal regions
- Sequence: DNA sequence (3 billion bases)
- Gene List: Identification of all genes

Information freely available without restriction
April 2003: FINISHED!

99.3% of genome
343 gaps
Human Gene Count: ~20-25,000
Beyond the Human Genome Project
Human Genome Project

Goal

• Know all sequence in human genome
Human Genome Project

**Goal**
- Know *all* sequence in human genome

**Next Challenge:**
- Understand sequence in human genome
Beyond the Human Genome Project

Goal

- Know all sequence in human genome
- Genome Content
  Know all functional elements in genome
- Genome Differences Among Individuals
  Know all human genetic variation
- Genome Expression
  Know all signatures of cellular response
- Genome Changes in Tumors
  Know all significant genomic changes
1. Functional Elements
Comparing Genomes
The Mouse Genome

Mouse Genome Sequencing Consortium, nature
Regions of conserved synteny: ~95% of genome
Regions of conserved synteny: ~95% of genome
Conserved sequence: Coding, Non-coding

PPARγ

Large elements (>100 bp)

Exons

Non-exons

75% 90%
~5% of Genome under selection

~500,000 conserved elements

1/3 = protein-coding genes

2/3 = ???
Some conserved sequence are regulatory elements

Delete conserved, non-coding segment

Causes downregulation of three nearby genes: IL-4, IL-13 and IL-5

Eddy Rubin and colleagues
~5% of Genome under Selection

Solution: More Mammalian Genomes
Mammalian Genomes

Human
Finished!

Mouse, Rat
Draft ~96%

Chimp
Draft ~92%

Dog
Draft ~95%
1. Functional Elements
Comparing Yeasts

Manolis Kellis
Bruce Birren
Comparative Genomics: Yeast

Total branch length:
0.83 substitutions per site
Yeast Genes

*S. cerevisiae*

*S. paradoxus*

*S. mikatae*

*S. bayanus*
Signature of Genes

Find All ‘Regions’
without
Frameshifting
Deletions

Gene

Intergenic

528 deleted
43 novel
34 merged
280 boundaries
Signature of Regulatory Elements

Find All ‘Motifs’ with Unusual Patterns of Genome-wide Conservation

Comprehensive Catalog of Regulatory Motifs

- **RTCAYT-4-ACGA**
  - **Motif**: ABF1
  - **Function**: origin, silencing
  - **Conservation Score**: $10^{-48}$

- **GTTACCCCGG**
  - **Motif**: REB1
  - **Function**: cell cycle control
  - **Conservation Score**: $10^{-31}$

- **AATGACTCAT**
  - **Motif**: GCN4/BAS1
  - **Function**: purine metabolism
  - **Conservation Score**: $10^{-29}$

- **TCGG-11-CCG**
  - **Motif**: GAL4
  - **Function**: carbohydrate metabolism
  - **Conservation Score**: $10^{-28}$

- **ATGAACACAA**
  - **Motif**: STE12/DIG1
  - **Function**: mating type determination
  - **Conservation Score**: $10^{-28}$

- **TTTGTTTAC**
  - **Motif**: FKH1,2/NDD1
  - **Function**: budding
  - **Conservation Score**: $10^{-27}$

- **WTATAATW**
  - **Motif**: TBP
  - **Function**: TATA box
  - **Conservation Score**: $10^{-27}$

- **GAAAATTTTTCA**
  - **Motif**: cx163,156,132
  - **Function**: rRNA complexes
  - **Conservation Score**: $10^{-18}$

- **GCGATGAGATG**
  - **Motif**: cx163,156,132
  - **Function**: rRNA complexes
  - **Conservation Score**: $10^{-14}$

- **GTCACGTG**
  - **Motif**: CBF1
  - **Function**: phosphate NS metabolism
  - **Conservation Score**: $10^{-14}$

- **GGTGGAACAA**
  - **Motif**: cx157,148
  - **Function**: protein synthesis & degradation
  - **Conservation Score**: $10^{-22}$

- **TTWCGCGT**
  - **Motif**: SWI4,6/MBP1
  - **Function**: DNA replication
  - **Conservation Score**: $10^{-15}$
Implications for Human Genome

S. cerevisiae

S. paradoxus

S. mikatae

S. bayanus

Total branch length:
0.83 substitutions per site

Signal to Noise is ~20-fold lower
Branch Length must be increased (by $\log_e(20) \approx 3$)
Sequencing mammalian genomes
Sequencing: More Species

Elephant

Cat

Armadillo

Shrew

Bat

Tenrec

Rabbit

Hedgehog
Four-species Alignment

-902
Human  CTGCCT----AAGTAGCCTAGACGCTCCGTGC-GGCTCCCGGGGCGG- TAG
Mouse  CGCCGC----CTGCAATATTTTAC--
Rat    CTGCTC----ATGCCGATTTTCAC--
Dog    CTGCTTTTCAACAGTGGGGCAGACGTCCCGCCGCCCACAGGCAGGGCCG
*  *        *        **

Human  GCCTGGCCGAAAATCTCTCCCGCGCCTGACCTTGGGTTGCCCCAGCCA
Mouse  ---------AAGCCTGTGGCGCGC-CGTGACCTTGGGCTGCCCCAGGCG
Rat    ---------AAGTTTCT---CTGC-CCTGACCTTGGGCTGCCCCAGGCG
Dog    GGCTGC----AGACCTGCCCCGAGGACAATGACCTTGGGCGGCCG
*  *       *  **********  *** ***

Human  GGCTGCGG GCCCGAGACCCCCG-------------------GG CCTCCCT
Mouse  GGCTGCAG GCTCACCACCCC---------------------GTCTTTTCT
Rat    AG--GCATACACCCCG CCTT---------------------TTTTTTTTT
Dog    GGCCGCGGGCCCAGCCCGGGCCTCGCCCTCCTCCCTCCTCCCTCCCTCCCT
*  **   * *    **                           *   *

Human  GCCCCCCG-----------CGCCGCCCCG ATTTGCCCTCAGAGAGGTAT
Mouse  GCTTTTCG-----------AGTCGGCCCGCTCTGCTCCCAG-GAGAGCAT
Rat    TTTTTTTTTTTGCCGTTCAA GAGCCCTGTTCTGCTCTCAA-AAGGGTA
Dog    GCCCCC CG-----------GACCGCCCC GCTTCACCCTCCCAGCTGGGAA
* ** * *   * * *       * *

-693
Human  ---CGATCTTATTT-CTGGGTCTACGGCAAACTCCAAGGCTACAAAC
Mouse  TCACGGTCTTATTTAGTGAGCGTAAGGCAAATCTGAATACCCAGCA
Rat    TAACGGTCTTATTTATTGGGCGCAAAGCAAACTTTAATACCCAGCAGG
Dog    CCCCGG GCCTGAATACGGAGTCAGCCGCACACTTCACGGCCCAAACGCGG
**  * *   *   * *      *** *    *    * *     *
Beyond the Human Genome Project

Goal

- **Know all** sequence in human genome

- **Genome Content**
  
  *Know all* functional elements in genome

- **Genome Differences Among Individuals**
  
  *Know all* human genetic variation

- **Genome Expression**
  
  *Know all* signatures of cellular response

- **Genome Changes in Tumors**
  
  *Know all* significant genomic changes
2. Human Genetic Variation
Comparing Individuals

David Altshuler
Stacey Gabriel
Single Nucleotide Polymorphisms (SNPs): 1 per 1300 bases
Humans Have Little Genetic Variation

- ~3,000 generations
- ~6,000,000,000,000 humans around world
- ~10,000 humans in Africa

Most genetic variation is common
Genetic Variation Can ‘Cause’ Genetic Disease

chromosome 19

T T T E2
T C C E3
C C C E4

Alzheimer's
# Common Variant Important in Risk of Common Disease

<table>
<thead>
<tr>
<th>Common Variant</th>
<th>Associated Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoE4</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>Factor V&lt;sup&gt;Leiden&lt;/sup&gt;</td>
<td>Venous thrombosis</td>
</tr>
<tr>
<td>HFE</td>
<td>Hemachromatosis</td>
</tr>
<tr>
<td>PPAR&lt;sub&gt;γ&lt;/sub&gt;</td>
<td>Type 2 Diabetes</td>
</tr>
<tr>
<td>MTHFR&lt;sup&gt;667T&lt;/sup&gt;</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>CCR5</td>
<td>HIV resistance</td>
</tr>
<tr>
<td>HLA-DQ&lt;sub&gt;α&lt;/sub&gt;</td>
<td>Type 1 Diabetes</td>
</tr>
</tbody>
</table>
Finding Susceptibility Genes

Paradigm for Human Genetics
• Enumerate all common variants
• Correlate with disease

Variants

<table>
<thead>
<tr>
<th>Disease</th>
<th>Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td></td>
</tr>
<tr>
<td>Arthritis</td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
</tr>
</tbody>
</table>
Finding All Human Variation

1998  Initial SNP Map  4,000 SNPs
1999  TheSNP Consortium (TSC)  Goal:  300,000 SNPs
2001  TSC & HGP  2,000,000 SNPs
2004  TSC, HGP, HapMap  7,800,000 SNPs
DNA Variation Occurs in ‘Blocks’

**Chrom 5q22**

<table>
<thead>
<tr>
<th>111 kb</th>
<th>13 kb</th>
<th>14 kb</th>
<th>12 kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGGGGGGACAAAC</td>
<td>GCATTC</td>
<td>CGGAGTACGGA</td>
<td>TGC GCCC</td>
</tr>
<tr>
<td>GCATTTC</td>
<td>GACTGC TTCG</td>
<td>TTGCC CCC</td>
<td></td>
</tr>
<tr>
<td>AACAAATTCGGG</td>
<td>AACCCC</td>
<td>CGCAGCACGGA</td>
<td>CTGCTAT</td>
</tr>
</tbody>
</table>

98%  97%  91%  98%
Next Goal: Human Haplotype Map
Haplotype Map of Human Genome

Chromosome 7q31: >1 SNPs over 1000 bases
Beyond the Human Genome Project

Goal

• Know all sequence in human genome

• Genome Content
  Know all functional elements in genome

• Genome Differences Among Individuals
  Know all human genetic variation

• Genome Expression
  Know all signatures of cellular response

• Genome Changes in Tumors
  Know all significant genomic changes
3. Cell Signatures:
Cancer

Todd Golub
Leukemias
Leukemias

AML

ALL
Other Recent Examples

- Further split of ALL
- Lymphoma
- Lung
- Breast
- Prostate
- Medulloblastoma
Splitting Leukemia: MLL

Todd, Golub, Scott Armstrong, Stan Korsmeyer, Dana-Farber Cancer Institute

**ALL patients**
Some have MLL translocation
Common in infants
Poor prognosis

*Is it a distinct type of cancer?*
Splitting Leukemia: MLL

Todd, Golub, Scott Armstrong, Stan Korsmeyer, Dana-Farber Cancer Institute

ALL patients
Some have MLL translocation
Common in infants
Poor prognosis

Is it a distinct type of cancer?
Treating MLL with anti-FLT3 drug

Todd, Golub, Scott Armstrong, Stan Korsmeyer, Dana-Farber Cancer Institute

Control Mice

Week 1  Week 2  Week 3  Week 4

Oral FLT3 Drug

Week 1  Week 2  Week 3  Week 4
3. Cellular Signatures: Connectivity Map

Todd Golub
Justin Lamb
genes connect genes with diseases and the drugs that treat them
a connectivity map

- BCR-ABL: chronic myeloid leukemia
- imatinib
- EWS-FLI: Ewing's sarcoma
- PDE5: sildenafil
- erectile dysfunction
- COX2
- LTD4: montelukast
- asthma
- rosiglitazone
- type II diabetes
- PPARγ
- colon cancer
- HDAC
- ATOX1
- SNF5/INI1
- SIX1
- TEPA
- steroid-resistant ALL
- SIX1
- colon cancer
- kidney cancer
- VHL
- obesity
- epilepsy
- valproate
- compound ‘X’

Questions marks indicate unknown connections.
a common global analytical space

state space = RNA expression
A Universal Functional Bio-assay

Functional Discovery via a Compendium of Expression Profiles


yeast knockout strains \((n = 287)\)

pharmaceuticals \((n = 13)\)

differentially expressed genes

perturbations (gene ablation or small molecule)

discover functional relatedness

high-density microarrays
Challenges to Building a Connectivity Map

limitations of cell culture
  • modifying influence of *inter- / extra- cellular environment*
  • appropriate *target cell population* (200 distinct cell-types)
  • *immortalized cell lines in rich culture medium*

treatment variables *(i.e. concentration and duration)*
  • drugs can have very *different effects at different concentrations*
  • the *extent of ablation / overexpression* will be different for each gene
  • widely varying *action kinetics*
  • many treatments may result in a *switch to a common cellular state*
  • *compensatory effects*

scale and throughput
  • the *number of determinations required* will be large

informatics and analytics
Pilot Experiment

100 perturbagens (chemicals + RNAi)
- estrogen receptor signalling (SERMs)
- PPAR signalling (diabetes)
- Protein Kinase Inhibitors
- Immunomodulators
- NSAIDs
- Anti-depressants

x 10 cell lines
- 5 cell types: myeloblast, astrocyte, melanocyte, fibroblast, primary fibroblast
- 5 epithelial lines: breast, prostate, colon, lung, uterus

x 4 dosages

x 4 time points

x 2 duplicates (~32,000)

Current: 650 experiments
Given a cell signature: What treatments trigger it?

Define signature (gene set)  Genes up- and down-regulated by perturbation

Compare signature to profile for each perturbagen

Rank perturbagens by signature strength

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>drug Y</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>drug e</td>
<td>0.993</td>
</tr>
<tr>
<td>3</td>
<td>gene S</td>
<td>0.791</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>gene n</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>drug I</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>drug L</td>
<td>0</td>
</tr>
<tr>
<td>997</td>
<td>drug N</td>
<td>-0.877</td>
</tr>
<tr>
<td>998</td>
<td>gene E</td>
<td>-0.945</td>
</tr>
<tr>
<td>999</td>
<td>drug G</td>
<td>-1</td>
</tr>
</tbody>
</table>

positive connectivity

no connectivity

negative connectivity
Example 1: Histone Deacetylase Signature

**Signature:**

![Structures of TSA, SAHA, and MS-275](image)

10 up- and 12 down-regulated genes (Glaser et al. 2003)

**Test signature in profile of each perturbagen**

**sodium valproate**

- up-regulated
- down-regulated

**metformin**

**Rank perturbagens**

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>cell line</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>trichostatin A-b</td>
<td>MCF-7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>trichostatin A-a</td>
<td>MCF-7</td>
<td>0.993</td>
</tr>
<tr>
<td>3</td>
<td>sodium valproate-a</td>
<td>MCF-7</td>
<td>0.791</td>
</tr>
<tr>
<td>4</td>
<td>sodium valproate-b</td>
<td>MCF-7</td>
<td>0.701</td>
</tr>
<tr>
<td>5</td>
<td>trichostatin A-c</td>
<td>HL-60</td>
<td>0.596</td>
</tr>
<tr>
<td>6</td>
<td>phenyl butyrate</td>
<td>HL-60</td>
<td>0.542</td>
</tr>
</tbody>
</table>

three different HDAC inhibitors are top ranked
Example 2: Estrogen treatment of Rats

**Signature:**

48 up- and 5 down-regulated genes  
(Rhen et al. 2003)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>amitriptyline</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>estrogen-c</td>
<td>0.921</td>
</tr>
<tr>
<td>3</td>
<td>NDGA</td>
<td>0.517</td>
</tr>
<tr>
<td>4</td>
<td>15Δ-PGJ2</td>
<td>0.497</td>
</tr>
<tr>
<td>5</td>
<td>estrogen-a</td>
<td>0.470</td>
</tr>
<tr>
<td>6</td>
<td>estrogen-b</td>
<td>0.435</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>tamoxifen</td>
<td>-0.733</td>
</tr>
<tr>
<td>119</td>
<td>rofecoxib</td>
<td>-0.784</td>
</tr>
<tr>
<td>120</td>
<td>monastrol</td>
<td>-0.814</td>
</tr>
<tr>
<td>121</td>
<td>MK-886</td>
<td>-0.849</td>
</tr>
<tr>
<td>122</td>
<td>raloxifene</td>
<td>-0.861</td>
</tr>
<tr>
<td>123</td>
<td>diclofenac</td>
<td>-0.879</td>
</tr>
<tr>
<td>124</td>
<td>LY294002</td>
<td>-0.884</td>
</tr>
<tr>
<td>125</td>
<td>novobiocin</td>
<td>-0.894</td>
</tr>
<tr>
<td>126</td>
<td>SC-58125</td>
<td>-0.907</td>
</tr>
<tr>
<td>127</td>
<td>phenyl_butyrate</td>
<td>-0.917</td>
</tr>
<tr>
<td>128</td>
<td>sulindac_sulfone</td>
<td>-0.955</td>
</tr>
<tr>
<td>129</td>
<td>rapamycin</td>
<td>-0.965</td>
</tr>
<tr>
<td>130</td>
<td>sulindac</td>
<td>-0.973</td>
</tr>
<tr>
<td>131</td>
<td>fulvestrant</td>
<td>-1</td>
</tr>
</tbody>
</table>

three independent estrogen instances

three different SERMs
Example 2: Estrogen treatment of Rats

**Signature**

48 up- and 5 down-regulated genes

(Rhen et al. 2003)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>amitriptyline</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>estrogen-c</td>
<td>0.921</td>
</tr>
<tr>
<td>3</td>
<td>NDGA</td>
<td>0.517</td>
</tr>
<tr>
<td>4</td>
<td>15Δ-PGJ2</td>
<td>0.497</td>
</tr>
<tr>
<td>5</td>
<td>estrogen-a</td>
<td>0.470</td>
</tr>
<tr>
<td>6</td>
<td>estrogen-b</td>
<td>0.435</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>tamoxifen</td>
<td>-0.733</td>
</tr>
<tr>
<td>119</td>
<td>rofecoxib</td>
<td>-0.784</td>
</tr>
<tr>
<td>120</td>
<td>monastrol</td>
<td>-0.814</td>
</tr>
<tr>
<td>121</td>
<td>MK-886</td>
<td>-0.849</td>
</tr>
<tr>
<td>122</td>
<td>raloxifene</td>
<td>-0.861</td>
</tr>
<tr>
<td>123</td>
<td>diclofenac</td>
<td>-0.879</td>
</tr>
<tr>
<td>124</td>
<td>LY294002</td>
<td>-0.884</td>
</tr>
<tr>
<td>125</td>
<td>novobiocin</td>
<td>-0.894</td>
</tr>
<tr>
<td>126</td>
<td>SC-58125</td>
<td>-0.907</td>
</tr>
<tr>
<td>127</td>
<td>phenyl_butyrate</td>
<td>-0.917</td>
</tr>
<tr>
<td>128</td>
<td>sulindac_sulfone</td>
<td>-0.955</td>
</tr>
<tr>
<td>129</td>
<td>rapamycin</td>
<td>-0.965</td>
</tr>
<tr>
<td>130</td>
<td>sulindac</td>
<td>-0.973</td>
</tr>
<tr>
<td>131</td>
<td>fulvestrant</td>
<td>-1</td>
</tr>
</tbody>
</table>

Recently discovered to be estrogenic

(Fujimoto et al. 2004)
Example 3: Human Diabetic Muscle

**Signature**

PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes

55 oxidative phosphorylation genes

(Mootha et al. 2003)

- **metformin**
  - (Glucophage®)
  - clinically-relevant oral antidiabetic

- **sodium valproate**
  - (clinically-relevant anticonvulsant)

False Positive?
Example 4: Ablation of ATOX1 copper transporter

**Signature**

![Expression level (Affy) and knock-down (Taqman) graph with 22 up- and 7 down-regulated genes.]

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATOX1_retroRNAi R1+R2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>ATOX1_retroRNAi R2</td>
<td>0.998</td>
</tr>
<tr>
<td>3</td>
<td>ATOX1_retroRNAi R1</td>
<td>0.996</td>
</tr>
<tr>
<td>4</td>
<td>ATOX1_retroRNAi R3</td>
<td>0.558</td>
</tr>
<tr>
<td>5</td>
<td>tetraethylenepentamine</td>
<td>0.521</td>
</tr>
</tbody>
</table>

another ATOX1 RNAi
Example 4: Ablation of ATOX1 copper transporter

**Signature**

22 up- and 7 down-regulated genes

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATOX1_retroRNAi R1+R2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>ATOX1_retroRNAi R2</td>
<td>0.998</td>
</tr>
<tr>
<td>3</td>
<td>ATOX1_retroRNAi R1</td>
<td>0.996</td>
</tr>
<tr>
<td>4</td>
<td>ATOX1_retroRNAi R3</td>
<td>0.558</td>
</tr>
<tr>
<td>5</td>
<td>tetraethylenepentamine</td>
<td>0.521</td>
</tr>
</tbody>
</table>

A copper chelator
Example 5: Drug resistant ALL

transcriptional analysis

acute lymphoblastic leukemia ($n=36$)

marker selection

ex vivo steroid response

gene expression signature

Scott Armstrong, DFCI
Todd Golub, DFCI
Example 5: Reversing drug resistance

**signature:** glucocorticoid resistant acute lymphoblastic leukemia

50 ‘sensitive’ and 50 ‘resistant’ markers

(David Twomey and Scott Armstrong)

<table>
<thead>
<tr>
<th>state</th>
<th>rank</th>
<th>perturbagen</th>
<th>cell line</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensitive</td>
<td>1</td>
<td>37-staurosporine</td>
<td>HL60</td>
<td>1</td>
</tr>
<tr>
<td>sensitive</td>
<td>2</td>
<td>37-rofecoxib</td>
<td>HL60</td>
<td>0.838</td>
</tr>
<tr>
<td>sensitive</td>
<td>3</td>
<td>35-rapamycin</td>
<td>HL60</td>
<td>0.804</td>
</tr>
<tr>
<td>sensitive</td>
<td>4</td>
<td>42-rapamycin</td>
<td>ssMCF7</td>
<td>0.789</td>
</tr>
<tr>
<td>sensitive</td>
<td>5</td>
<td>60-tretinoin</td>
<td>PC3</td>
<td>0.708</td>
</tr>
<tr>
<td>sensitive</td>
<td>6</td>
<td>42-LY294002</td>
<td>ssMCF7</td>
<td>0.703</td>
</tr>
<tr>
<td>sensitive</td>
<td>7</td>
<td>60-diclofenac</td>
<td>PC3</td>
<td>0.677</td>
</tr>
<tr>
<td>sensitive</td>
<td>8</td>
<td>6-chlorpropamide</td>
<td>MCF7</td>
<td>0.617</td>
</tr>
<tr>
<td>sensitive</td>
<td>9</td>
<td>35-LY294002</td>
<td>HL60</td>
<td>0.607</td>
</tr>
<tr>
<td>sensitive</td>
<td>10</td>
<td>65-rofecoxib</td>
<td>PC3</td>
<td>0.590</td>
</tr>
<tr>
<td>sensitive</td>
<td>11</td>
<td>55-LY294002</td>
<td>PC3</td>
<td>0.558</td>
</tr>
<tr>
<td>sensitive</td>
<td>12</td>
<td>26-rapamycin</td>
<td>MCF7</td>
<td>0.544</td>
</tr>
</tbody>
</table>

... 186

**hypothesis:** rapamycin induces glucocorticoid sensitivity

Armstrong *et al* (unpublished)
Example 5: Rapamycin induces dex sensitivity

Armstrong et al (unpublished)
Toward a first generation Connectivity Map

- **twenty thousand diverse perturbations**...
  - ever-approved pharmaceuticals and ‘landmark compounds’ (~2000)
  - other bioactive small molecules (~7000)
    Institute for Chemistry and Chemical Biology (ICCB)
  - RNAi-mediated gene ablation (~5000)
    The RNAi Consortium (TRC)
  - ORF-mediated ectopic expression (~5000)
    TRC and Harvard Institute of Proteomics (HIP)
  - clinically annotated human tissue specimens (~1000)
    The Harvard Hospitals
Beyond the Human Genome Project

Goal

- Know all sequence in human genome

- Genome Content
  Know all functional elements in genome

- Genome Differences Among Individuals
  Know all human genetic variation

- Genome Expression
  Know all signatures of cellular response

- Genome Changes in Tumors
  Know all significant genomic changes
4. Mutations in Cancer

Matt Meyerson
Bill Sellers
Systematic Search for Mutations in Lung Cancer

Matt Meyerson, Bill Sellers, DFCI

Resequencing ~50 Kinase Genes in Tumors

Epidermal Growth Factor Receptor (EGFR)
Mutations in Tumors:
- Japanese
- Non-smokers
- Women
- Adenocarcinomas

Matches Response Profile of New Drug, Iressa
Patient Response to Iressa

Iressa Responders ~ EGRF$^+$ tumors

Before treatment, Oct 2003

After treatment, Jan 2004

Bruce Johnson, DFCI
Beyond the Human Genome Project

Goal

• Know all sequence in human genome

• Genome Content
  Know all functional elements in genome

• Genome Differences Among Individuals
  Know all human genetic variation

• Genome Expression
  Know all signatures of cellular response

• Genome Changes in Tumors
  Know all significant genomic changes
The Eli and Edythe L. Broad Institute
A Collaboration of Massachusetts Institute of Technology, Harvard University and affiliated Hospitals, and Whitehead Institute for Biomedical Research
Initial Core Members
Stuart Schrieber, FAS Chem
David Altshuler, HMS Genetics
Todd Golub, DFCI
Eric Lander, MIT/HMS/WIBR

60 Associate Members:
FAS, HMS, HSPH, MIT, WIBR
MGH, DFCI, BWH, BI, TCH

The Eli and Edythe L. Broad Institute
A Collaboration of Massachusetts Institute of Technology, Harvard University and affiliated Hospitals, and Whitehead Institute for Biomedical Research
Acknowledgements

Genome Analysis
Kerstin Lindblad-Toh
Bruce Birren
Chad Nusbaum
Michele Clamp
Jean Chang
Rob Nicol
Jen Baldwin
Manolis Kellis
Xiaohui Xie

Human Variation
David Altshuler
Stacey Gabriel
Mark Daly

Cancer Classification
Todd Golub
Christine Ladd-Acosta
Scott Armstrong
Stan Korsmeyer
David Twomey
Ben Ebert

Connectivity Map
Justin Lamb
Todd Golub
Emily Crawford
Jean-Philippe Brunet
Dave Peck
Michael Reich

Cancer Mutations
Matt Meyerson
Bill Sellers
Sequencing: Current Species

- Platypus
- Opossum
- Tenrec
- Short-eared elephant shrew
- Phyllostomid microbat
- Free-tailed bat
- False vampire bat
- Flying fox
- Roussette fruit bat - megabat
- Whale
- Dolphin
- Hippo
- Cow
- Pig
- Llama
- Horse
- Rhinoceros
- Tapir
- Cat
- Dog
- Pangolin
- Sciurid
- Mouse
- Rat
- Hystricid
- Caviomorph
- Rabbit
- Pika
- Tree shrew tupaia
- Tree shrew urogale
- Flying lemur variegatus
- Flying lemur volans
- Lemur
- Mouse lemur
- Galago
- Bushbaby
- Tarsier banaeus
- Tarsier syrichta
- Spider monkey
- Goeldi monkey
- Marmoset
- Macaque
- Baboon
- Vervet
- Human
- Chimpanzee
- Gorilla
- Orangutan
- Gibbon
- Tenrec
- Golden mole
- Short-eared elephant shrew
- Long-eared elephant shrew
- Aardvark
- Sirenian
- Hyrax
- Elephant
- Marusupial
- Cow
Next Phase of Human Genome Project: A Transcriptional Connectivity Map

Justin Lamb
Todd Golub
Two common variants reproducibly associated with risk of type 2 diabetes are in drug targets.

- $K_{ir}6.2$ E23K
- Sulfonylurea
- $PPAR_\gamma$ P12A thiazoladinedione
### Dog Genome Sequence

**Genome assembly much better than mouse**

<table>
<thead>
<tr>
<th></th>
<th>Dog</th>
<th>Mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig N50</td>
<td>122 kb</td>
<td>25 kb</td>
</tr>
<tr>
<td>Supercontig N50</td>
<td>42 Mb</td>
<td>17 Mb</td>
</tr>
<tr>
<td>Coverage</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>Genome size</td>
<td>2.5 Gb</td>
<td>2.7 Gb</td>
</tr>
</tbody>
</table>
Huge blocks (2-10 Mb) across 50% of genome are effectively homozygous, with 1000-fold lower SNP rate.
10 Breeds screened for Polymorphisms

600,000 SNPs identified

Poodle
Boxer
Beagle
Alaskan malamute
Australian Terrier
English Shepherd
German Shepherd
Italian Greyhound
Labrador retriever
Portuguese waterdog
Rottweiler
Dog Genome

Extraordinary Haplotype Structure: Local

Number of Local Haplotypes:

All Breeds ~7
Single Breed ~3
Dog Genome
Extraordinary Haplotype Structure: Long-range

Long Range Correlation (LD): Biphasic

Implies:
Association studies within-breed should be easy (~10,000 SNPs)
Examples of *common* variants that influence risk of common human diseases

ApoE4: Alzheimers
Factor V*leiden*: DVT
HLA: other autoimmune dz.
HFE: Hemachromatosis
CCR5: HIV infection
CTLA4: Graves disease
NOD2: IBD
5q31 haplotype: IBD
PPARG: T2DM
Kir6.2: T2D
Neuregulin: schizophrenia
ApoA5: triglycerides
LPL: lipid levels
G6PD: malaria
Example 5: Obesity

**signature: 'cafeteria diet'**

- 163 up- and 184 down-regulated genes
  - (López et al. 2003)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>cell line</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41-tretinoin</td>
<td>HL60</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>55-rosiglitazone</td>
<td>PC3</td>
<td>0.946</td>
</tr>
<tr>
<td>3</td>
<td>37-troglitazone</td>
<td>HL60</td>
<td>0.753</td>
</tr>
<tr>
<td>4</td>
<td>36-WY-14643</td>
<td>MCF7</td>
<td>0.739</td>
</tr>
<tr>
<td>5</td>
<td>55-troglitazone</td>
<td>PC3</td>
<td>0.735</td>
</tr>
<tr>
<td>6</td>
<td>88-tretinoin</td>
<td>HL60</td>
<td>0.724</td>
</tr>
<tr>
<td>7</td>
<td>36-estrogen</td>
<td>MCF7</td>
<td>0.724</td>
</tr>
<tr>
<td>8</td>
<td>17-tretinoin</td>
<td>MCF7</td>
<td>0.708</td>
</tr>
<tr>
<td>9</td>
<td>56-fasudil</td>
<td>PC3</td>
<td>0.700</td>
</tr>
<tr>
<td>10</td>
<td>31-2-deoxyglucose</td>
<td>MCF7</td>
<td>0.693</td>
</tr>
<tr>
<td>11</td>
<td>40-genistein</td>
<td>MCF7</td>
<td>0.665</td>
</tr>
<tr>
<td>12</td>
<td>40-tretinoin</td>
<td>MCF7</td>
<td>0.658</td>
</tr>
<tr>
<td>13</td>
<td>39-cobalt_chloride</td>
<td>HL60</td>
<td>0.632</td>
</tr>
<tr>
<td>14</td>
<td>9-rosiglitazone</td>
<td>MCF7</td>
<td>0.601</td>
</tr>
<tr>
<td>15</td>
<td>43-phenyl_butyrate</td>
<td>MCF7</td>
<td>0.600</td>
</tr>
<tr>
<td>16</td>
<td>17-Ro_13-7410</td>
<td>MCF7</td>
<td>0.596</td>
</tr>
</tbody>
</table>

- thiazoledinediones
- RAR agonists
Example 5: Obesity - Clinical connections

weight gain is an established consequence of thiazolidinedione therapy...

Weight changes during long-term treatment with Pioglitazone as monotherapy.

Example 5: Obesity - Cellular correlates

Thiazolinediones are adipogenic and PPARγ agonists...

Fajas et al. (2002)

Tretinoin is adipogenic in human promyelocytic cells...

Inazawa et al. (2003)
Example 5: Potential anti-obesity agents

**signature: ‘cafeteria diet’**

163 up- and 184 down- regulated genes (López et al 2003)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>152</td>
<td>16-celecoxib</td>
<td>-0.514</td>
</tr>
<tr>
<td>153</td>
<td>39-celecoxib</td>
<td>-0.518</td>
</tr>
<tr>
<td>154</td>
<td>42-alpha-estradiol</td>
<td>-0.519</td>
</tr>
<tr>
<td>155</td>
<td>13-NDGA</td>
<td>-0.520</td>
</tr>
<tr>
<td>156</td>
<td>54-17-AAG</td>
<td>-0.533</td>
</tr>
<tr>
<td>157</td>
<td>33-valproate_10mM</td>
<td>-0.551</td>
</tr>
<tr>
<td>158</td>
<td>20-clofibrate</td>
<td>-0.567</td>
</tr>
<tr>
<td>159</td>
<td>18-rofecoxib</td>
<td>-0.591</td>
</tr>
<tr>
<td>160</td>
<td>73-SC-58125</td>
<td>-0.611</td>
</tr>
<tr>
<td>161</td>
<td>18-celecoxib</td>
<td>-0.617</td>
</tr>
<tr>
<td>162</td>
<td>7b-exemestane</td>
<td>-0.618</td>
</tr>
<tr>
<td>163</td>
<td>5-estrogen</td>
<td>-0.644</td>
</tr>
<tr>
<td>164</td>
<td>33-valproate_500uM</td>
<td>-0.657</td>
</tr>
<tr>
<td>165</td>
<td>18-SC-58125</td>
<td>-0.667</td>
</tr>
<tr>
<td>166</td>
<td>53-trifluoperazine</td>
<td>-0.718</td>
</tr>
<tr>
<td>167</td>
<td>26-rapamycin</td>
<td>-0.725</td>
</tr>
<tr>
<td>168</td>
<td>6-tamoxifene</td>
<td>-0.753</td>
</tr>
<tr>
<td>169</td>
<td>18-LM-1685</td>
<td>-0.849</td>
</tr>
<tr>
<td>170</td>
<td>33-valproate_2mM</td>
<td>-1</td>
</tr>
</tbody>
</table>

**COX-2 inhibitors**

![Chemical structures](image)

SC-58125, LM-1685, rofecoxib, celecoxib
Example 5: COX-2 inhibitors and adipogenesis

COX-2 inhibitors block adipocyte differentiation...

3T3-L1 cells

0 1 µM 10 µM SC-58236

A Universal Functional Bio-assay… Drugs

PNAS | August 5, 2003 | vol. 100 | no. 16

Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro

Erik C. Gunther*, David J. Stone*, Robert W. Gerwien, Patricia Bento, and Melvyn P. Heyes

psychoactives (n = 36)
cultured neurons

high-density microarrays

classification by indication through transitory features
A Universal Functional Bio-assay... Genes

A Mechanism of Cyclin D1 Action Encoded in the Patterns of Gene Expression in Human Cancer

Justin Lamb, Sridhar Rameshwar, Heide L. Ford, Bernardo Contreras, Robert V. Martinez, Frances S. Kittrell, Cynthia A. Zahnkow, Nick Patterson, Todd R. Gokub, and Mark E. Ewen

...
GSEA: Functional Sets

PGC1α timepoint

Control

PGC1α + vs -

Sorted Gene List

Gene Set

1

2
Searching Promoters for Regulatory Motifs

5000 genes with ‘known’ transcription starts
Compare Human, Mouse, Rat, Dog

Transcription Start
-1 kb  +1 kb

Human
Mouse
Rat
Dog
Conserved Sequence
Global Cancer Map

Tissue Types
1. Breast Adeno
2. Prostate Adeno
3. Lung Adeno
4. Colorectal Adeno
5. Lymphoma
6. Bladder
7. Melanoma
8. Uterine Adeno
9. Leukemia (AML)
10. Leukemia (ALL-T)
11. Leukemia (ALL-B)
12. Renal
13. Pancreatic Adeno
14. Ovary
15. Mesothelioma
16. CNS

Normal Tissue Types
1. Breast
2. Prostate
3. Lung
4. Colon
5. Lymphoid GC
6. Bladder
7. Uterus
8. Kidney
9. Pancreas
10. Ovary
11. Whole Brain
12. Cerebellum
13. Muscle
14. Stomach
All Cellular Signatures:

GSEA
Regions of conserved synteny: ~95% of genome
Regions of conserved synteny: ~95% of genome
Protein-coding gene count falling

Human protein-coding genes: 20,000-25,000
Large conserved elements: Coding, Non-coding

PPARγ

Large conserved elements (>100 bp)

Exons

Non-exons

75% 90%
~5% of Genome under selection
~5% of Genome under selection

~500,000 conserved elements

33% = protein coding genes
67% = ???
Mammalian Genomes

- Human
  Finished!
- Mouse, Rat
  Draft ~96%
- Chimp
  Draft ~92%
- Dog
  Draft ~95%

Also Done: Marsupial (Monodelphis)
Comparison between Species:
Finding functional elements in Yeast
Comparing many species

- S. cerevisiae
- S. paradoxus (6X)
- S. mikatae (6X)
- S. bayanus (6X)
Multiple species comparison allows extraction of signal from noise . . .

. . . to identify small (~10 bp) conserved islands covering ~13% of genome

**GAL1-10:**

- Islands
- Known elements
Identifying regulatory signals

Gal4: CGG(N11)CCG
Testing all motifs for genic, non-genic conservation

- CGG(N11)CCG

Patterns with potential intergenic function

Most patterns

Conservation in coding regions

Intergenic Conservation
Genome-wide motif discovery

Test 1
Test 2
Test 3

2,000 Mini-motifs

Extend
Extend
Extend

Collapse
Collapse
Collapse

Merge

Full Motifs

72 Full motifs
<table>
<thead>
<tr>
<th>Regulatory motifs</th>
<th>Protein(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTCAYT-4-ACGA</td>
<td>ABF1</td>
</tr>
<tr>
<td>GTTACCCCGG</td>
<td>REB1</td>
</tr>
<tr>
<td>AATGACTCAT</td>
<td>GCN4/BAS1</td>
</tr>
<tr>
<td>TCGG-11-CCG</td>
<td>GAL4</td>
</tr>
<tr>
<td>ATGAAACAA</td>
<td>STE12/DIG1</td>
</tr>
<tr>
<td>TTTGTTTAC</td>
<td>FKH1,2/NDD1</td>
</tr>
<tr>
<td>WTATAATW</td>
<td>TBP</td>
</tr>
<tr>
<td>GAAAAATTTCAC</td>
<td>cx163,156,132</td>
</tr>
<tr>
<td>GCAGATGAGATG</td>
<td>cx163,156,132</td>
</tr>
<tr>
<td>GTCACGTG</td>
<td>CBF1</td>
</tr>
<tr>
<td>GGTGGCAA</td>
<td>cx157,148</td>
</tr>
<tr>
<td>TTWCACGCGTT</td>
<td>SWI4,6/MBP1</td>
</tr>
<tr>
<td>ATGATGGGTG</td>
<td>RAP1/FHL1</td>
</tr>
<tr>
<td>TTC . GAA . TTC</td>
<td>HSF1</td>
</tr>
<tr>
<td>GATAAAG</td>
<td></td>
</tr>
<tr>
<td>CCGG . CCGG</td>
<td>LEU3</td>
</tr>
<tr>
<td>TTT – 4,7,11 – TTT</td>
<td>FKH2</td>
</tr>
</tbody>
</table>
Intersecting with Functional Categories

MIPS
CGG-11-CCG

Transport
Nucleus
Transcription
Cell Cycle
Energy
Cell fate

Carbohydrate metabolism

Gal genes
Gal1, Gal2, Gal3, Gal7, Gal10, Hxt3, Mth1, Pcl10, Gcy1

Significance of overlap: $10^{-28}$
Intersecting with...

- **Functional Classification**
  MIPS: CGG-11-CCG

- **Chromatin IP**
  Gal4: TCA-6-ACG

- **Protein Complexes by MassSpec**
  GATGAG

- **Expression Clusters**
  ACGCGT

**Specificity P-value**

- MIPS: $10^{-28}$
- TCA-6-ACG: $10^{-48}$
- GATGAG: $10^{-18}$
- ACGCGT: $10^{-29}$
<table>
<thead>
<tr>
<th>Regulatory motifs</th>
<th>Factor</th>
<th>Function</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTCA YT-4-ACGA</td>
<td>ABF1</td>
<td>origin, silencing</td>
<td>10^{-48}</td>
</tr>
<tr>
<td>GTTACCCCGG</td>
<td>REB1</td>
<td>cell cycle control</td>
<td>10^{-31}</td>
</tr>
<tr>
<td>AATGACTCAT</td>
<td>GCN4/BAS1</td>
<td>purine metabolism</td>
<td>10^{-29}</td>
</tr>
<tr>
<td>TCGG-11-CCG</td>
<td>GAL4</td>
<td>carbohydrate metabolism</td>
<td>10^{-28}</td>
</tr>
<tr>
<td>ATGAAACAAA</td>
<td>STE12/DIG1</td>
<td>mating type determination</td>
<td>10^{-28}</td>
</tr>
<tr>
<td>TTTGTTTAC</td>
<td>FKH1,2/NDD1</td>
<td>budding</td>
<td>10^{-27}</td>
</tr>
<tr>
<td>WTATAATW</td>
<td>TBP</td>
<td>TATA box</td>
<td></td>
</tr>
<tr>
<td>GAAAATTTTTTCA</td>
<td>cx163,156,132</td>
<td>rRNA complexes</td>
<td>10^{-18}</td>
</tr>
<tr>
<td>GC GATGAGATG</td>
<td>cx163,156,132</td>
<td>rRNA complexes</td>
<td>10^{-14}</td>
</tr>
<tr>
<td>GTCACGTG</td>
<td>CBF1</td>
<td>phosphate NS metabolism</td>
<td>10^{-14}</td>
</tr>
<tr>
<td>GGTGGCAA</td>
<td>cx157,148</td>
<td>protein synthesis &amp; degradation</td>
<td>10^{-22}</td>
</tr>
<tr>
<td>TTTWCGCGTT</td>
<td>SWI4,6/MBP1</td>
<td>DNA replication</td>
<td>10^{-15}</td>
</tr>
<tr>
<td>ATGTATGGGTG</td>
<td>RAP1/FHL1</td>
<td>ribosome biogenesis</td>
<td>10^{-15}</td>
</tr>
<tr>
<td>TTC . GAA . TTC</td>
<td>HSF1</td>
<td>stress response</td>
<td>10^{-13}</td>
</tr>
<tr>
<td>GATAAAG</td>
<td>HSF1</td>
<td>stress response</td>
<td>10^{-12}</td>
</tr>
<tr>
<td>CCGG . CCGG</td>
<td>LEU3</td>
<td>nitrogen and sulfur metabolism</td>
<td>10^{-9}</td>
</tr>
<tr>
<td>TTT – 4,7,11 – TTT</td>
<td>FKH2</td>
<td>budding, cell polarity, filament</td>
<td>10^{-8}</td>
</tr>
</tbody>
</table>
Genome-wide co-occurrence map

[Diagram of gene interactions]
Motif combinations change specificity

Conserved occurrences of Ste12, Tec1

Ste12  Tec1

Mating  Budding  Filamentation
Comparison between Species: Finding functional elements in Mammals
Digression: Evolution of Yeast
2. All Cellular Signatures
Diabetes: Serious Side Effects, Worldwide Epidemic

- Diabetes Mellitus Type 2 (DM2)
- 100M affected now, ~200M by 2010
- Affects 7% of American population
- Major contributor to:
  - coronary artery disease
  - stroke
  - blindness
  - amputation
  - kidney failure
- Concordance rate in MZ twins: 80%
Type 2 Diabetes (DM2) Characterized by Insulin Resistance

Pancreas

Insulin production

Muscle, Fat

Insulin-mediated glucose uptake

Blood Glucose
Many Pathways Invoked to Explain Insulin Resistance

- insulin signaling
- PI-3 kinase signaling
- TNF signaling
- Glut 4 transporters
- glycogen production
- glycogen breakdown
- β-oxidation
- glucosamine pathway
- beta cell development
- insulin production
- insulin receptor signaling
- mitochondrial metabolism
- cytokine signaling
- adipogenesis
- neuronal control of appetite
- adrenergic signaling

Persegheni et al. NEJM 1996
Genomic Approaches to Understanding DM2

*Human physiology* → ? → *Human genetics* → ? → *Cell-based models*

*Human physiology*

*Cell-based models*
Genomic Approaches to Understanding DM2

*Human physiology*  

*Cell-based models*
• 17 diabetics, 18 normal, 8 IGT
• European males, age matched ~65 yr
• Muscle biopsy: 2-hr euglycemic, hyperinsulinemic clamp (45 mU/m²)
Gene Expression in Diabetic and Normal Muscle

Diabetic vs. Normal

Sorted Gene List

18 Diabetic

17 Normal

Genes with significant differences?

NONE!
Integrative Genomics: Study Gene Sets (GSEA)

18 Diabetic vs. Normal

Gene 1
Gene 2
...

Gene List
Gene 345
Gene 28
...

Sorted Gene List

Urea cycle
28 genes
Integrative Genomics: Study Gene Sets (GSEA)

INS signaling:

42 genes

Diabetic vs. Normal

Sorted Gene List

18 Diabetic

17 Normal
## 150 Gene Sets

### Pathways Curated at WICGR
- FFA Oxidation
- Gluconeogenesis
- Glycolysis
- Glycogen metabolism
- GO:0005739 (Insulin signaling)
- Ketone body metabolism
- Pyruvate metabolism
- Reactive oxygen species
- Kreb's cycle
- Oxidative phosphorylation (OXPHOS)

### 36 GNF Mouse Expression Clusters
- Cluster 00...

### Pathways from NetAFFX (October 2002)

<table>
<thead>
<tr>
<th>Pathway ID</th>
<th>Pathway Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP00010</td>
<td>Glycolysis_Gluconeogenesis</td>
</tr>
<tr>
<td>MAP00020</td>
<td>Citrate cycle_TCA cycle</td>
</tr>
<tr>
<td>MAP00030</td>
<td>Pentose phosphate pathway</td>
</tr>
<tr>
<td>MAP00040</td>
<td>Pentose and glucuronate interconversions</td>
</tr>
<tr>
<td>MAP00051</td>
<td>Fructose and mannose metabolism</td>
</tr>
<tr>
<td>MAP00052</td>
<td>Galactose metabolism</td>
</tr>
<tr>
<td>MAP00053</td>
<td>Ascorbate and aldarate metabolism</td>
</tr>
<tr>
<td>MAP00061</td>
<td>Fatty acid biosynthesis_path_1</td>
</tr>
<tr>
<td>MAP00062</td>
<td>Fatty acid biosynthesis_path_2</td>
</tr>
<tr>
<td>MAP00071</td>
<td>Fatty acid metabolism</td>
</tr>
<tr>
<td>MAP00072</td>
<td>Synthesis and degradation of ketone bodies</td>
</tr>
<tr>
<td>MAP00090</td>
<td>Steroid biosynthesis</td>
</tr>
<tr>
<td>MAP00120</td>
<td>Bile acid biosynthesis</td>
</tr>
<tr>
<td>MAP00130</td>
<td>Ubiquinone biosynthesis</td>
</tr>
<tr>
<td>MAP00140</td>
<td>C21 Steroid-hormone metabolism</td>
</tr>
<tr>
<td>MAP00150</td>
<td>Androgen and estrogen metabolism</td>
</tr>
<tr>
<td>MAP00190</td>
<td>Cis-diol dehydrogenase/phosphatase</td>
</tr>
<tr>
<td>MAP00193</td>
<td>ATP synthase</td>
</tr>
<tr>
<td>MAP00195</td>
<td>Photosynthesis</td>
</tr>
<tr>
<td>MAP00200</td>
<td>Urea cycle and metabolism of amino groups</td>
</tr>
<tr>
<td>MAP00230</td>
<td>Purine metabolism</td>
</tr>
<tr>
<td>MAP00240</td>
<td>Pyrimidine metabolism</td>
</tr>
<tr>
<td>MAP00251</td>
<td>Glutamate metabolism</td>
</tr>
<tr>
<td>MAP00252</td>
<td>Alanine and aspartate metabolism</td>
</tr>
<tr>
<td>MAP00253</td>
<td>Tryptophane, biosynthesis</td>
</tr>
<tr>
<td>MAP00260</td>
<td>Glycine, serine and threonine metabolism</td>
</tr>
<tr>
<td>MAP00271</td>
<td>Methionine metabolism</td>
</tr>
<tr>
<td>MAP00272</td>
<td>Cysteine metabolism</td>
</tr>
<tr>
<td>MAP00280</td>
<td>Valine, leucine and isoleucine degradation</td>
</tr>
<tr>
<td>MAP00290</td>
<td>Valine, leucine and isoleucine biosynthesis</td>
</tr>
<tr>
<td>MAP00300</td>
<td>Lysine biosynthesis</td>
</tr>
<tr>
<td>MAP00310</td>
<td>Lysine degradation</td>
</tr>
<tr>
<td>MAP00330</td>
<td>Arginine and proline metabolism</td>
</tr>
<tr>
<td>MAP00340</td>
<td>Histidine metabolism</td>
</tr>
<tr>
<td>MAP00350</td>
<td>Tyrosine metabolism</td>
</tr>
<tr>
<td>MAP00360</td>
<td>Phenylalanine metabolism</td>
</tr>
<tr>
<td>MAP00361</td>
<td>gamma-Hexachlorocyclohexane degradation</td>
</tr>
<tr>
<td>MAP00380</td>
<td>Tryptophan metabolism</td>
</tr>
<tr>
<td>MAP00400</td>
<td>Phenylalanine, tyrosine and tryptophan biosynthesis</td>
</tr>
<tr>
<td>MAP00410</td>
<td>beta-Alanine metabolism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathway ID</th>
<th>Pathway Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP00430</td>
<td>Taurine and hypouracil metabolism</td>
</tr>
<tr>
<td>MAP00440</td>
<td>Aminophosphonate metabolism</td>
</tr>
<tr>
<td>MAP00450</td>
<td>Seleniumic acid metabolism</td>
</tr>
<tr>
<td>MAP00460</td>
<td>Cyanophosphonate metabolism</td>
</tr>
<tr>
<td>MAP00471</td>
<td>D-Glutamine and D-glutamate metabolism</td>
</tr>
<tr>
<td>MAP00472</td>
<td>D-Arginine and D-ornithine metabolism</td>
</tr>
<tr>
<td>MAP00480</td>
<td>Glicophosphate metabolism</td>
</tr>
<tr>
<td>MAP00500</td>
<td>Starch and sucrose metabolism</td>
</tr>
<tr>
<td>MAP00510</td>
<td>N-Glycans biosynthesis</td>
</tr>
<tr>
<td>MAP00511</td>
<td>N-Glycan degradation</td>
</tr>
<tr>
<td>MAP00512</td>
<td>O-Glycans biosynthesis</td>
</tr>
<tr>
<td>MAP00520</td>
<td>Nucleotide sugars metabolism</td>
</tr>
<tr>
<td>MAP00521</td>
<td>Streptococcal biosynthesis</td>
</tr>
<tr>
<td>MAP00522</td>
<td>Erythromycin biosynthesis</td>
</tr>
<tr>
<td>MAP00530</td>
<td>Aminosugar metabolism</td>
</tr>
<tr>
<td>MAP00531</td>
<td>Glycosaminoglycan degradation</td>
</tr>
<tr>
<td>MAP00532</td>
<td>Chondroitin, Heparin, sulfatide biosynthesis</td>
</tr>
<tr>
<td>MAP00533</td>
<td>Keratan sulfate biosynthesis</td>
</tr>
<tr>
<td>MAP00550</td>
<td>Peptidoglycan biosynthesis</td>
</tr>
<tr>
<td>MAP00561</td>
<td>Glycosphingolipid metabolism</td>
</tr>
<tr>
<td>MAP00562</td>
<td>Inositol phosphates metabolism</td>
</tr>
<tr>
<td>MAP00570</td>
<td>Sphingosinephospholipid biosynthesis</td>
</tr>
<tr>
<td>MAP00580</td>
<td>Phospholipid degradation</td>
</tr>
<tr>
<td>MAP00590</td>
<td>Prostaglandin_and_ leukotriene metabolism</td>
</tr>
<tr>
<td>MAP00600</td>
<td>Sphingomyelin biosynthesis</td>
</tr>
<tr>
<td>MAP00601</td>
<td>Blood group glycolipid biosynthesis last series</td>
</tr>
<tr>
<td>MAP00602</td>
<td>Blood group glycolipid biosynthesis_neolact series</td>
</tr>
<tr>
<td>MAP00603</td>
<td>Globoside metabolism</td>
</tr>
<tr>
<td>MAP00620</td>
<td>Pyruvate metabolism</td>
</tr>
<tr>
<td>MAP00625</td>
<td>Tetrachloroethene degradation</td>
</tr>
<tr>
<td>MAP00630</td>
<td>Glyoxylate and glycerate metabolism</td>
</tr>
<tr>
<td>MAP00631</td>
<td>1,2-Dicarboxylic acid metabolism</td>
</tr>
<tr>
<td>MAP00640</td>
<td>Propanoate metabolism</td>
</tr>
<tr>
<td>MAP00643</td>
<td>Styrene degradation</td>
</tr>
<tr>
<td>MAP00650</td>
<td>Butanoate metabolism</td>
</tr>
<tr>
<td>MAP00670</td>
<td>Oxa carbon pool by bile</td>
</tr>
<tr>
<td>MAP00680</td>
<td>Methane metabolism</td>
</tr>
<tr>
<td>MAP00710</td>
<td>Carbon fixation</td>
</tr>
<tr>
<td>MAP00720</td>
<td>Reductive carbohydrate cycle_CO2 fixation</td>
</tr>
<tr>
<td>MAP00740</td>
<td>Ribofuranosyl metabolism</td>
</tr>
<tr>
<td>MAP00750</td>
<td>Vitamin B6 metabolism</td>
</tr>
<tr>
<td>MAP00760</td>
<td>Nicotinamide and nicotinate metabolism</td>
</tr>
<tr>
<td>MAP00770</td>
<td>Pantethenate and CoA biosynthesis</td>
</tr>
<tr>
<td>MAP00780</td>
<td>Biotin metabolism</td>
</tr>
<tr>
<td>MAP00790</td>
<td>Folate biosynthesis</td>
</tr>
<tr>
<td>MAP00830</td>
<td>Retinol metabolism</td>
</tr>
<tr>
<td>MAP00860</td>
<td>Porphyrin and chlorophyll metabolism</td>
</tr>
<tr>
<td>MAP00900</td>
<td>Tryptophan biosynthesis</td>
</tr>
<tr>
<td>MAP00910</td>
<td>Nitrogen metabolism</td>
</tr>
<tr>
<td>MAP00920</td>
<td>Sulfur metabolism</td>
</tr>
<tr>
<td>MAP00940</td>
<td>Xanthosin and xanthine biosynthesis</td>
</tr>
<tr>
<td>MAP00950</td>
<td>Alkaloid biosynthesis_1</td>
</tr>
<tr>
<td>MAP00960</td>
<td>Alkaloid biosynthesis_2</td>
</tr>
<tr>
<td>MAP00970</td>
<td>Aminocyclopentadiene biosynthesis</td>
</tr>
<tr>
<td>MAP00980</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>MAP00990</td>
<td>DNA polymerase</td>
</tr>
<tr>
<td>MAP01000</td>
<td>Type III secretion system</td>
</tr>
<tr>
<td>MAP01006</td>
<td>Type II secretion system</td>
</tr>
</tbody>
</table>

### Additional Information
- 113 locally and publicly curated pathways and gene sets
- 36 gene expression clusters based on public mouse data
OXPHOS gene set shows strong correlation

18 Diabetic

17 Normal

Diabetic vs. Normal

Sorted Gene List

Gene 1
Gene 2
...

Gene 345
Gene 28
...

p = 0.00027

OXPHOS: 106 genes
Oxidative Phosphorylation (OXPHOS) Genes
OXPHOS genes show modest decrease (~20%), but highly consistent across set.
An Expression Cluster also shows strong correlation

Diabetic vs. Normal

Sorted Gene List

OXPHOS 106 genes
Cluster c20 225 genes

18 Diabetic

17 Normal
Diabetes Association Due to Tightly Co-Regulated Subset: OXPHOS-CR

$r=0.61$
What Co-Regulates OXPHOS-CR Genes?

Hypothesis:

**PGC-1α**, Regulator implicated in Mitochondrial Biogenesis

Wu et al. Cell 1999
Hypothesis confirmed:

PGC-1α Transgene Induces OXPHOS-CR Genes

$r = 0.61$
Models of insulin resistance point to mitochondria

Cell-based models

mitochondrion

ROS
Cellular Models of Insulin Resistance

Various Treatments Produce Insulin Resistance (e.g., TNFa, Dex)

3T3L1 adipocytes

Glucose uptake

Insulin

Insulin Resistance

Insulin Resistance

TNF

Dex

??

Insulin Resistance
Gene Expression in Cell Models of Insulin Resistance

Sorted Gene List

Genes with significant differences?
Gene Expression in Cell Models of Insulin Resistance

Insulin Resistant vs. Normal

Sorted Gene List

ROS Genes
- Metallothionein 1
- Cytochrome P450, 1b1
- Xanthine Dehydrogenase
- Delta Aminolev. Dehydratase
- Thiosulfate Sulfurtransferase
- Spermidine Acetyltransferase
- Thioredoxin, Mitochondrial
- Hemopexin
- Haptoglobin
- Ceruloplasmin
- Cytochrome C oxidase, VIIIb
- Metallothionein 2
What is ROS?

Three Key Reactive Oxygen Species: $\text{O}_2^-$  $\text{H}_2\text{O}_2$  $\text{OH}^\cdot$

- Mitochondria are principal source of ROS
- $\text{O}_2^-$ results from imbalance in substrate supply and electron transport chain capacity
- $\text{H}_2\text{O}_2$ and $\text{OH}^\cdot$ result from $\text{O}_2^-$ and $\text{Fe}^{++}$
- All are capable of damaging cellular proteins
Is ROS increased in models of insulin resistance? Is ROS just a marker or a cause of insulin resistance?

TNF

Dex

ROS ?

Insulin Resistance

**ROS assays**
- $\text{H}_2\text{O}_2$ release (Amplex Red)
- Intracellular oxidizing environ (DCF)
- Intracellular $\text{O}_2^-$ (HydroEthidine)

**Small molecules and transgenes**
- N-acetylcysteine
- MnTBAP
- Catalase transgene
- Frataxin transgene

Restores ~50% of insulin sensitivity
Genomic Approaches to Understanding DM2

Human physiology

Cell-based models

Human genetics
Association Study: Type 2 Diabetes

Initial screen

Re-test

Confirmation

4,500 DNAs from two populations

\( \text{PPAR}_\gamma \) (\( p < 0.001 \))
**PPARγ**

**PPARγ: Nuclear Hormone Receptor**
- implicated in mitochondrial biogenesis
- co-activated by PGC1α

**Variant:**
- Pro12Ala (Freq = 16%)
- Affects receptor activity
- Pro allele increases risk by 30%
- All reports together p = $10^{-8}$

Population attributable risk = 25%

**Clinical Connection:**
- PPARγ is target of troglitazones, which restore insulin sensitivity
Three approaches all point to mitochondria

*Human physiology*

- OXPHOS
- PGC1a

*Cell-based models*

- ROS
- ALS + FA

*Human genetics*

- PPARγ

Diagram showing the interconnection between human physiology, cell-based models, and human genetics, focusing on mitochondria, OXPHOS, PGC1a, ROS, ALS + FA, and PPARγ.
2. Cellular Signatures: Connectivity Map
Beyond the Human Genome Project

Goal
- Know *all* sequence in human genome
- Know *all* functional elements in genome
- Know *all* human genetic variation
- Know *all* signatures of cellular response
- Know *all* mechanisms of cancer
Acknowledgements

**Functional Elements, Yeast Duplication**
Manolis Kellis
Bruce Birren
Xiaohui Xie

**Diabetes**
David Altshuler
Nick Houstis
Vamsi Mootha
Aravind Subramanian
Jill Mesirov
Pablo Tamayo

**Connectivity Map**
Todd Golub
Justin Lamb
Gene Set Enrichment Analysis (GSEA)

Vamsi Mootha
Aravind Subramanian
Pablo Tamayo
Jill Mesirov
Gene Set Database

Curated: 565 sets
Biocarta, Signaling Gateway, GenMAPP … (11 public sources)

Computed: 6120 sets
3 Human cancer compendia 2000 conditions
3 Human normal tissue compendia 70 tissue types, 500 conditions
1 Mouse normal tissue compendium 34 tissue types, 100 conditions

Positional: 311 sets
Cytogenetic bands

Experimental: 125 sets
Transcriptional Profiles & Target Lists

Many sets are closely related…
GSEA with Positional Sets: Downs Syndrome

Downs vs. Normal Sorted Gene List

P = 0.012
### GSEA with Known Pathways: Medulloblastomas

**Classic**  
**Desmoplastic**

**Sonic Hedgehog Pathway**

<table>
<thead>
<tr>
<th>Enriched in <strong>Desmoplastic</strong> Tumors</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shh Pathway</td>
<td>0.012</td>
</tr>
<tr>
<td>Melanocyte Pathway</td>
<td>0.015</td>
</tr>
<tr>
<td>Keratinocyte Pathway</td>
<td>0.018</td>
</tr>
<tr>
<td>Erbb4 Pathway</td>
<td>0.020</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enriched in <strong>Classic</strong> Tumors</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrin in Post-Synaptic Differentiation</td>
<td>0.004</td>
</tr>
<tr>
<td>p35 Pathway</td>
<td>0.009</td>
</tr>
<tr>
<td>Lissencephaly:Neuronal Migration and Development</td>
<td>0.012</td>
</tr>
<tr>
<td>Wnt Pathway</td>
<td>0.033</td>
</tr>
</tbody>
</table>
‘Stemness’ Signature: Individual Genes

Rimalho-Santos

Ivanova

Fortunel
‘Stemness’ Signature: GSEA

Ramalho-Santos

Ivanova

Fortunel

133
Example 5: PI3K/mTOR inhibition

**signature: LY294002 (PI3K/mTOR inhibitor)**

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25-LY294002</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>19-LY294002</td>
<td>0.988</td>
</tr>
<tr>
<td>3</td>
<td>26-LY294002</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>42-LY294002</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>35-LY294002</td>
<td>0.708</td>
</tr>
<tr>
<td>6</td>
<td>28-trichostatin_A-b</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>26-rapamycin</td>
<td>0.547</td>
</tr>
<tr>
<td>8</td>
<td>28-trichostatin_A-a</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>53-trifluoperazine</td>
<td>0.468</td>
</tr>
<tr>
<td>10</td>
<td>38-raloxifene</td>
<td>0.462</td>
</tr>
<tr>
<td>11</td>
<td>23-fulvestrant</td>
<td>0.461</td>
</tr>
<tr>
<td>12</td>
<td>53-stauroporine</td>
<td>0.442</td>
</tr>
<tr>
<td>13</td>
<td>13-raloxifene</td>
<td>0.414</td>
</tr>
<tr>
<td>14</td>
<td>46-thioridazine</td>
<td>0.412</td>
</tr>
<tr>
<td>15</td>
<td>38-tamoxifen</td>
<td>0.404</td>
</tr>
<tr>
<td>16</td>
<td>42-rapamycin</td>
<td>0.387</td>
</tr>
<tr>
<td>17</td>
<td>42-wortmannin</td>
<td>0.383</td>
</tr>
<tr>
<td>18</td>
<td>19-wortmannin</td>
<td>0.381</td>
</tr>
<tr>
<td>19</td>
<td>35-trichostatin_A</td>
<td>0.364</td>
</tr>
<tr>
<td>20</td>
<td>35-rapamycin</td>
<td>0.363</td>
</tr>
<tr>
<td>21</td>
<td>35-phenyl_butyrate</td>
<td>0.360</td>
</tr>
<tr>
<td>22</td>
<td>6-tamoxifen</td>
<td>0.359</td>
</tr>
<tr>
<td>23</td>
<td>33-valproate_2mM</td>
<td>0.355</td>
</tr>
<tr>
<td>24</td>
<td>33-valproate_10mM</td>
<td>0.345</td>
</tr>
<tr>
<td>25</td>
<td>31-stauroporine</td>
<td>0.341</td>
</tr>
<tr>
<td>26</td>
<td>17-17-AAG</td>
<td>0.338</td>
</tr>
<tr>
<td>27</td>
<td>1-valproate_1mM</td>
<td>0.316</td>
</tr>
<tr>
<td>28</td>
<td>41-wortmannin</td>
<td>0.301</td>
</tr>
</tbody>
</table>

...
Example 5: PI3K/mTOR inhibition

**signature:** LY294002 (PI3K/mTOR inhibitor)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25-LY294002</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>19-LY294002</td>
<td>0.988</td>
</tr>
<tr>
<td>3</td>
<td>26-LY294002</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>42-LY294002</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>35-LY294002</td>
<td>0.708</td>
</tr>
<tr>
<td>6</td>
<td>28-trichostatin_A-b</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>26-rapamycin</td>
<td>0.547</td>
</tr>
<tr>
<td>8</td>
<td>28-trichostatin_A-a</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>53-trifluoperazine</td>
<td>0.468</td>
</tr>
<tr>
<td>10</td>
<td>38-raloxifene</td>
<td>0.462</td>
</tr>
<tr>
<td>11</td>
<td>23-fulvestrant</td>
<td>0.461</td>
</tr>
<tr>
<td>12</td>
<td>53-staurosporine</td>
<td>0.442</td>
</tr>
<tr>
<td>13</td>
<td>13-raloxifene</td>
<td>0.414</td>
</tr>
<tr>
<td>14</td>
<td>46-thioridazine</td>
<td>0.412</td>
</tr>
<tr>
<td>15</td>
<td>38-tamoxifen</td>
<td>0.404</td>
</tr>
<tr>
<td>16</td>
<td>42-rapamycin</td>
<td>0.387</td>
</tr>
<tr>
<td>17</td>
<td>42-wortmannin</td>
<td>0.383</td>
</tr>
<tr>
<td>18</td>
<td>19-wortmannin</td>
<td>0.381</td>
</tr>
<tr>
<td>19</td>
<td>35-trichostatin_A</td>
<td>0.384</td>
</tr>
<tr>
<td>20</td>
<td>35-rapamycin</td>
<td>0.363</td>
</tr>
<tr>
<td>21</td>
<td>35-phenyl_butyrate</td>
<td>0.360</td>
</tr>
<tr>
<td>22</td>
<td>6-tamoxifen</td>
<td>0.359</td>
</tr>
<tr>
<td>23</td>
<td>33-valproate_2mM</td>
<td>0.355</td>
</tr>
<tr>
<td>24</td>
<td>33-valproate_10mM</td>
<td>0.345</td>
</tr>
<tr>
<td>25</td>
<td>31-staurosporine</td>
<td>0.341</td>
</tr>
<tr>
<td>26</td>
<td>17-17-AAG</td>
<td>0.338</td>
</tr>
<tr>
<td>27</td>
<td>1-valproate_1mM</td>
<td>0.316</td>
</tr>
<tr>
<td>28</td>
<td>41-wortmannin</td>
<td>0.301</td>
</tr>
</tbody>
</table>

*Cell Death and Differentiation (2001) 8, 841–849*

Staurosporine inhibits phosphorylation of translational regulators linked to mTOR

AR Tee and CG Proud
Example 5: PI3K/mTOR inhibition

**signature:** LY294002 (PI3K/mTOR inhibitor)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25-LY294002</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>19-LY294002</td>
<td>0.988</td>
</tr>
<tr>
<td>3</td>
<td>26-LY294002</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>42-LY294002</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>35-LY294002</td>
<td>0.708</td>
</tr>
<tr>
<td>6</td>
<td>28-trichostatin_A-b</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>26-rapamycin</td>
<td>0.547</td>
</tr>
<tr>
<td>8</td>
<td>28-trichostatin_A-a</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>53-trifluoperazine</td>
<td>0.468</td>
</tr>
<tr>
<td>10</td>
<td>38-raloxifene</td>
<td>0.462</td>
</tr>
<tr>
<td>11</td>
<td>23-fulvestrant</td>
<td>0.461</td>
</tr>
<tr>
<td>12</td>
<td>53-stauroporine</td>
<td>0.442</td>
</tr>
<tr>
<td>13</td>
<td>13-raloxifene</td>
<td>0.414</td>
</tr>
<tr>
<td>14</td>
<td>46-thioridazine</td>
<td>0.412</td>
</tr>
<tr>
<td>15</td>
<td>38-tamoxifen</td>
<td>0.404</td>
</tr>
<tr>
<td>16</td>
<td>42-rapamycin</td>
<td>0.387</td>
</tr>
<tr>
<td>17</td>
<td>42-wortmannin</td>
<td>0.383</td>
</tr>
<tr>
<td>18</td>
<td>19-wortmannin</td>
<td>0.381</td>
</tr>
<tr>
<td>19</td>
<td>35-trichostatin_A</td>
<td>0.364</td>
</tr>
<tr>
<td>20</td>
<td>35-rapamycin</td>
<td>0.363</td>
</tr>
<tr>
<td>21</td>
<td>35-phenyl_butyrate</td>
<td>0.360</td>
</tr>
<tr>
<td>22</td>
<td>6-tamoxifen</td>
<td>0.359</td>
</tr>
<tr>
<td>23</td>
<td>33-valproate_2mM</td>
<td>0.355</td>
</tr>
<tr>
<td>24</td>
<td>33-valproate_10mM</td>
<td>0.345</td>
</tr>
<tr>
<td>25</td>
<td>31-stauroporine</td>
<td>0.341</td>
</tr>
<tr>
<td>26</td>
<td>17-17-AAG</td>
<td>0.338</td>
</tr>
<tr>
<td>27</td>
<td>1-valproate_1mM</td>
<td>0.316</td>
</tr>
<tr>
<td>28</td>
<td>41-wortmannin</td>
<td>0.301</td>
</tr>
</tbody>
</table>

…

three different SERMs

---

**Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase**

Tommaso Simoncini*, Ali Hafezi-Moghadam†, Derek P. Brazil‡, Klaus Ley†, William W. Chin‡ & James K. Liao*
**Example 5: PI3K/mTOR inhibition**

**signature:** LY294002 (PI3K/mTOR inhibitor)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25-LY294002</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>19-LY294002</td>
<td>0.988</td>
</tr>
<tr>
<td>3</td>
<td>26-LY294002</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>42-LY294002</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>35-LY294002</td>
<td>0.708</td>
</tr>
<tr>
<td>6</td>
<td>28-trichostatin_A-b</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>26-rapamycin</td>
<td>0.547</td>
</tr>
<tr>
<td>8</td>
<td>28-trichostatin_A-a</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>53-trifluoperazine</td>
<td>0.468</td>
</tr>
<tr>
<td>10</td>
<td>38-raloxifene</td>
<td>0.462</td>
</tr>
<tr>
<td>11</td>
<td>23-fulvestrant</td>
<td>0.461</td>
</tr>
<tr>
<td>12</td>
<td>53-stauroporine</td>
<td>0.442</td>
</tr>
<tr>
<td>13</td>
<td>13-raloxifene</td>
<td>0.414</td>
</tr>
<tr>
<td>14</td>
<td>46-thioridazine</td>
<td>0.412</td>
</tr>
<tr>
<td>15</td>
<td>38-tamoxifen</td>
<td>0.404</td>
</tr>
<tr>
<td>16</td>
<td>42-rapamycin</td>
<td>0.387</td>
</tr>
<tr>
<td>17</td>
<td>42-wortmannin</td>
<td>0.383</td>
</tr>
<tr>
<td>18</td>
<td>19-wortmannin</td>
<td>0.381</td>
</tr>
<tr>
<td>19</td>
<td>35-trichostatin_A</td>
<td>0.364</td>
</tr>
<tr>
<td>20</td>
<td>35-rapamycin</td>
<td>0.363</td>
</tr>
<tr>
<td>21</td>
<td>35-phenyl_butyrate</td>
<td>0.360</td>
</tr>
<tr>
<td>22</td>
<td>6-tamoxifen</td>
<td>0.359</td>
</tr>
<tr>
<td>23</td>
<td>33-valproate_2mM</td>
<td>0.355</td>
</tr>
<tr>
<td>24</td>
<td>33-valproate_10mM</td>
<td>0.345</td>
</tr>
<tr>
<td>25</td>
<td>31-stauroporine</td>
<td>0.341</td>
</tr>
<tr>
<td>26</td>
<td>17-17-AAG</td>
<td>0.338</td>
</tr>
<tr>
<td>27</td>
<td>1-valproate_1mM</td>
<td>0.316</td>
</tr>
<tr>
<td>28</td>
<td>41-wortmannin</td>
<td>0.301</td>
</tr>
</tbody>
</table>

**CANCER CELL: DECEMBER 2003**

A chemical genetic screen identifies inhibitors of regulated nuclear export of a Forkhead transcription factor in PTEN-deficient tumor cells

Tweeny R. Kau,1,4 Frank Schroeder,4 Shriapriya Ramaswamy,3,5 Cheryl L. Wojcilewski,1,4 Jean J. Zhao,4 Thomas M. Roberts,2,4 Jon Clardy,1 William R. Goodwin,2,4 and Pamela A. Silver1,4,4
### Example 5: PI3K/mTOR inhibition

**signature:** LY294002 (PI3K/mTOR inhibitor)

<table>
<thead>
<tr>
<th>rank</th>
<th>perturbagen</th>
<th>KS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25-LY294002</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>19-LY294002</td>
<td>0.988</td>
</tr>
<tr>
<td>3</td>
<td>26-LY294002</td>
<td>0.982</td>
</tr>
<tr>
<td>4</td>
<td>42-LY294002</td>
<td>0.944</td>
</tr>
<tr>
<td>5</td>
<td>35-LY294002</td>
<td>0.708</td>
</tr>
<tr>
<td>6</td>
<td>28-trichostatin_A-b</td>
<td>0.558</td>
</tr>
<tr>
<td>7</td>
<td>26-rapamycin</td>
<td>0.547</td>
</tr>
<tr>
<td>8</td>
<td>28-trichostatin_A-a</td>
<td>0.547</td>
</tr>
<tr>
<td>9</td>
<td>53-trifluoperazine</td>
<td>0.468</td>
</tr>
<tr>
<td>10</td>
<td>38-raloxifene</td>
<td>0.462</td>
</tr>
<tr>
<td>11</td>
<td>23-fulvestrant</td>
<td>0.461</td>
</tr>
<tr>
<td>12</td>
<td>53-staurosporine</td>
<td>0.442</td>
</tr>
<tr>
<td>13</td>
<td>13-raloxifene</td>
<td>0.414</td>
</tr>
<tr>
<td>14</td>
<td>46-thioridazine</td>
<td>0.412</td>
</tr>
<tr>
<td>15</td>
<td>38-tamoxifen</td>
<td>0.404</td>
</tr>
<tr>
<td>16</td>
<td>42-rapamycin</td>
<td>0.387</td>
</tr>
<tr>
<td>17</td>
<td>42-wortmannin</td>
<td>0.383</td>
</tr>
<tr>
<td>18</td>
<td>19-wortmannin</td>
<td>0.381</td>
</tr>
<tr>
<td>19</td>
<td>35-trichostatin_A</td>
<td>0.364</td>
</tr>
<tr>
<td>20</td>
<td>35-rapamycin</td>
<td>0.363</td>
</tr>
<tr>
<td>21</td>
<td>35-phenyl_butyrate</td>
<td>0.360</td>
</tr>
<tr>
<td>22</td>
<td>6-tamoxifen</td>
<td>0.359</td>
</tr>
<tr>
<td>23</td>
<td>33-valproate_2mM</td>
<td>0.355</td>
</tr>
<tr>
<td>24</td>
<td>33-valproate_10mM</td>
<td>0.345</td>
</tr>
<tr>
<td>25</td>
<td>31-staurosporine</td>
<td>0.341</td>
</tr>
<tr>
<td>26</td>
<td>17-17-AAG</td>
<td>0.338</td>
</tr>
<tr>
<td>27</td>
<td>1-valproate_1mM</td>
<td>0.316</td>
</tr>
<tr>
<td>28</td>
<td>41-wortmannin</td>
<td>0.301</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **HDAC inhibitors**
  - sodium valproate
  - phenyl butyrate
  - trichostatin A

**PI3K/mTOR inhibitor**

- LY294002

**Notes:**

- The KS score represents the degree of inhibition compared to the LY294002 reference.
Beyond the Human Genome Project

Audacious Goal

• Know all sequence in human genome

• Know all functional elements in genome  
  Comparing Organisms

• Know all human genetic variation  
  Comparing Humans

• Know all signatures of cellular responses  
  Comparing Cells
2. Comparing Individuals
Single Nucleotide Polymorphisms (SNPs): 1 per 1300 bases
Humans Have Little Genetic Variation

- 3,000 generations
- ~6,000,000,000 humans around world
- ~10,000 humans in Africa

Most genetic variation is common
Genetic Variation Can ‘Cause’ Genetic Disease

chromosome 19

TTT E2
TTC E3
CCC E4

Alzheimer’s
<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoE4</td>
<td>Alzheimer’s disease</td>
</tr>
<tr>
<td>Factor V&lt;sup&gt;Leiden&lt;/sup&gt;</td>
<td>Venous thrombosis</td>
</tr>
<tr>
<td>HFE</td>
<td>Hemachromatosis</td>
</tr>
<tr>
<td>PPARγ</td>
<td>Type 2 Diabetes</td>
</tr>
<tr>
<td>MTHFR&lt;sup&gt;667T&lt;/sup&gt;</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>CCR5</td>
<td>HIV resistance</td>
</tr>
<tr>
<td>HLA-DQ&lt;sub&gt;α&lt;/sub&gt;</td>
<td>Type 1 Diabetes</td>
</tr>
</tbody>
</table>
Finding Susceptibility Genes

Paradigm for Human Genetics
• Enumerate all common variants
• Correlate with disease

Variants

- Diabetes
- Arthritis
- Stroke
- Asthma
- Prostate
implications for c-map utility

with an *in vivo* phenotype or disease query signature…

**annotate small molecules**

- identify inducers and repressors of the phenotype…
  - lead/tool compound discovery
  - prediction of adverse effects

**annotate diseases**

- exploit knowledge of landmark compounds…
  - generate mechanistic hypotheses
  - identification of therapeutic targets
ALL: Signature of drug resistance

- Acute lymphoblastic leukemia ($n=36$)
- Transcriptional analysis
- Marker selection
- Ex vivo steroid response
- Gene expression signature

- Resistant
- Sensitive

- Viable cells over concentration graph
**ALL: Reversing drug resistance**

*signature: glucocorticoid resistant acute lymphoblastic leukemia*

50 ‘sensitive’ and 50 ‘resistant’ markers

(David Twomey and Scott Armstrong)

---

<table>
<thead>
<tr>
<th>state</th>
<th>rank</th>
<th>perturbagen</th>
<th>cell line</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensitive</td>
<td>1</td>
<td>37-staurosporine</td>
<td>HL60</td>
<td>1</td>
</tr>
<tr>
<td>sensitive</td>
<td>2</td>
<td>37-rofecoxib</td>
<td>HL60</td>
<td>0.838</td>
</tr>
<tr>
<td>sensitive</td>
<td>3</td>
<td>35-rapamycin</td>
<td>HL60</td>
<td>0.804</td>
</tr>
<tr>
<td>sensitive</td>
<td>4</td>
<td>42-rapamycin</td>
<td>ssMCF7</td>
<td>0.789</td>
</tr>
<tr>
<td>sensitive</td>
<td>5</td>
<td>60-tretinoin</td>
<td>PC3</td>
<td>0.708</td>
</tr>
<tr>
<td>sensitive</td>
<td>6</td>
<td>42-LY294002</td>
<td>ssMCF7</td>
<td>0.703</td>
</tr>
<tr>
<td>sensitive</td>
<td>7</td>
<td>60-diclofenac</td>
<td>PC3</td>
<td>0.677</td>
</tr>
<tr>
<td>sensitive</td>
<td>8</td>
<td>6-chlorpropamide</td>
<td>MCF7</td>
<td>0.617</td>
</tr>
<tr>
<td>sensitive</td>
<td>9</td>
<td>35-LY294002</td>
<td>HL60</td>
<td>0.607</td>
</tr>
<tr>
<td>sensitive</td>
<td>10</td>
<td>65-rofecoxib</td>
<td>PC3</td>
<td>0.590</td>
</tr>
<tr>
<td>sensitive</td>
<td>11</td>
<td>55-LY294002</td>
<td>PC3</td>
<td>0.558</td>
</tr>
<tr>
<td>sensitive</td>
<td>12</td>
<td>26-rapamycin</td>
<td>MCF7</td>
<td>0.544</td>
</tr>
</tbody>
</table>

---

**hypothsis:**
rapamycin induces glucocorticoid sensitivity
ALL: Rapamycin induces glucocorticoid sensitivity

Armstrong et al (unpublished)
an automated profiling production pipeline

- a fully integrated gene expression data generation solution
- cell culture, RNA isolation, target preparation and profiling
- process tracking/management, quality control and assurance
- seamless informatics and analytics
- interfaces and visualizers for the entire research community

Affymetrix HTA PEGarray
- oligonucleotide microarray
- 96 samples × >22,000 probe-sets
- integrated automated solution
our multidisciplinary team, and collaborators

Emily Crawford
Jean-Philippe Brunet
Dave Peck
Michael Reich
Christine Ladd-Acosta

Scott Armstrong
David Twomey
Ben Ebert
Jen Pretz
Michele Lee

Paul Clemons
Steve Haggarty

Todd Golub
Stuart Schreiber
Eric Lander
The Connectivity Map Project

Justin Lamb
Todd Golub
The Next Phase of the Human Genome Project

**Phase I:** define the elements
the genes, transcripts, proteins, regulatory elements, etc

*The Human Genome Sequencing Project*

**Phase II:** establish the connections between the elements
which elements operate together

catalog the effects of chemical intervention
connect small molecules (‘drugs’) with genes

*A Connectivity Map of the Human Genome*
A Connectivity Map of the Human Genome

“… connect genes with genes, drugs with drugs, and drugs with genes in a global map of functional relatedness…”

Statement of the Problem
conventional approaches aren’t up to the job

The Connectivity Map Solution
a universal functional bio-assay

Design and Objectives for a Pilot Study
addressing the challenges… feasibility and scale

Stories from the Pilot
connections, reproducibility and robustness

An Automated Profiling Production Pipeline
A3P… building a first generation Connectivity Map

What We Can Do When We’re Done
functional screens in silico… and rational drug design
Statement of the Problem—Connecting Genes

“...defining groups of genes that operate together is hard...”

cell-based phenotypic screens

- tedious and labor-intensive
- low sensitivity
- knowledge of the appropriate biological parameter to measure
- availability of a convenient assay

physical interaction mapping (ie Y2H, Y1H, ‘ChIP-on-chip’)

- broad scope and higher throughput
- rely upon secondary phenotypic analyses to assess functional relevance
Statement of the Problem—Drugs and Genes

“...pharmaceutical discovery relies upon screen after screen after screen...”

**cell-based phenotypic screens**
- knowledge of the appropriate biological parameter to measure
- availability of a convenient assay
- low sensitivity
- doesn’t identify the therapeutic target

**in vitro screens** (*ie* disrupting interactions, affecting activity)
- rapid and sensitive
- requires knowledge and functional characterization of the target
- development of a bespoke biochemical assay
- unable to detect lack of specificity or gross cellular toxicity
- requires secondary screening step
The Connectivity Map Approach

“…beyond the ‘case-by-case assay paradigm’…”

principle

use gene expression profiling as a universal functional bio-assay

capture the consequences of all possible genetic and chemical insults in this common global analytical space

match the patterns of gene expression associated with each insult to connect genes with genes, drugs with drugs, and drugs with genes in a genome- and library- scale map of functional relatedness

practice

acute genome-wide transcriptional responses of cultured human cells

ablation of every gene in the genome with RNA-interference (RNAi)

over-expression of every gene in the genome using ORF collections

treatment with a myriad of small molecules