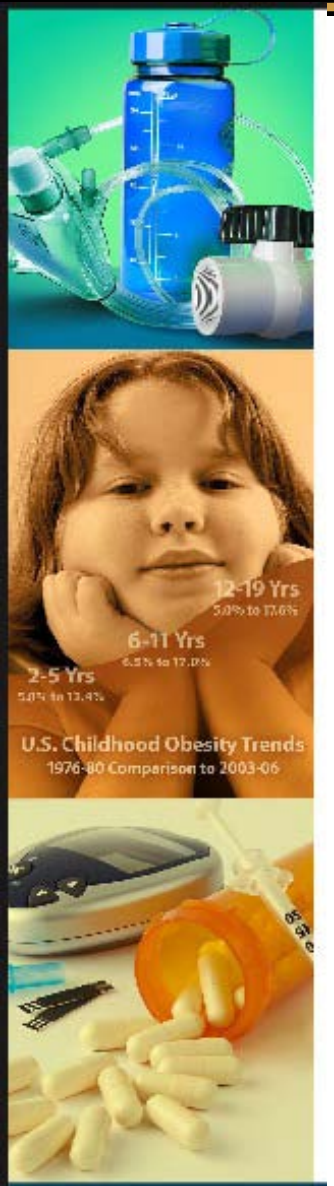
NTP
National Toxicology Program



NTP Workshop: Role of Environmental Chemicals in the Development of Diabetes and Obesity

Breakout Group on

Bioinformatics and High Throughput Screening (HTS)

William Janzen (chair)
Vickie Walker (rapporteur)

Crabtree Marriott Hotel
January 11-13, 20111

Bioinformatics and HTS Breakout Group Members

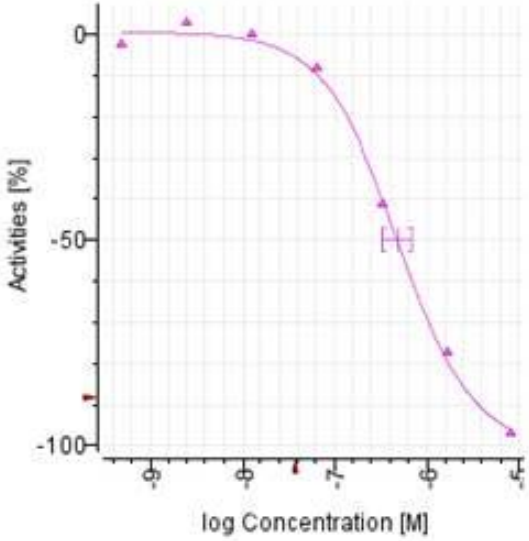
Scott Auerbach, NIH/NIEHS/NTP	Chirag Patel, Stanford
Bruce Blumberg, UC-Irvine	David Reif, US EPA
Deborah Clegg, U Texas Southwestern	Jennifer Schlezinger, Boston U
David Dix, US EPA	Supriya Srinivasan , Scripps Research Institute
Alison Holloway, McMaster U	Vickie Walker, NIH/NIEHS/NTP (rapporteur)
Keith Houck, US EPA	Morris White, Harvard
William Janzen, UNC – Chapel Hill (chair)	
Richard Judson, US EPA	
James Kaput, FDA/NCTR	
Franck Mauvais-Jarvis, Northwestern	
Angel Nadal, Miguel Hernandez U	

In response to the 2007 NAS report on "Toxicity Testing in the 21st Century: A Vision and a Strategy", a Memorandum of Understanding was signed in 2008 and expanded in 2010 to include the FDA

- **National Toxicology Program:** Linda S. Birnbaum, Ph.D., DABT, ATS Director, National Institute of Environmental Health Sciences, National Institutes of Health
- **NIH Chemical Genomics Center:** Eric D. Green, M.D., Ph.D. Director National Human Genome Research Institute
- **U.S. Environmental Protection Agency:** Paul T. Anastas, Ph.D. Assistant Administrator Office of Research and Development
- **Food and Drug Administration:** Janet Woodcock, M.D., Director Center for Drug Evaluation and Research

The NCGC

- Conducts quantitative high-throughput screening (qHTS)
  - >300,000 profiles/week
- qHTS profile
  - 1536-well plate format
  - 15-point concentration-response curve
  - DMSO soluble
  - 5 nM to 92 μM typical
  - ~5 μL assay volume
  - ~1000 cells/well



NCGC: Phase I qHTS screening

- ~2800 compounds provided by the NTP and the EPA

Apoptosis

- 3/7, 8, & 9

Cell Viability

- ATP
- LDH
- Protease release

DNA damage

- ELG1
- p53
- Multiple repair-deficient DT40 cell lines

Epigenetics

- LDR

Mitochondrial toxicity

Nuclear Receptors

- hAR
- hERα
- hFXR
- hGR
- hLXRβ
- hPPARα
- hPPARγ
- hPPARδ
- hPXR
- rPXR
- hROR
- hRXR
- hTRβ
- hVDR

Pathways

- AP1
- ARE/Nrf2
- CRE
- hERG
- HRE
- HSP
- JNK3
- NFκB

NCGC: Phase II

- Will be screening >10,000 compounds, with 3X concentration response curves
- Initial screening focus on nuclear receptors and stress response pathways; then will focus on disease oriented pathways

NCGC

- Drugs
- Drug-like compounds
- Active pharmaceutical ingredients

EPA

- ToxCast I and II compounds
- Antimicrobial Registration Program
- Endocrine Disruptor Screening Program
- OECD Molecular Screening Working Group List
- Failed Drugs

NTP

- NTP-studied compounds of all types
- NTP nominations and related compounds
- ICCVAM and NICEATM validation and reference compounds
- Outside collaborators (e.g., U.S. Army Public Health Command)

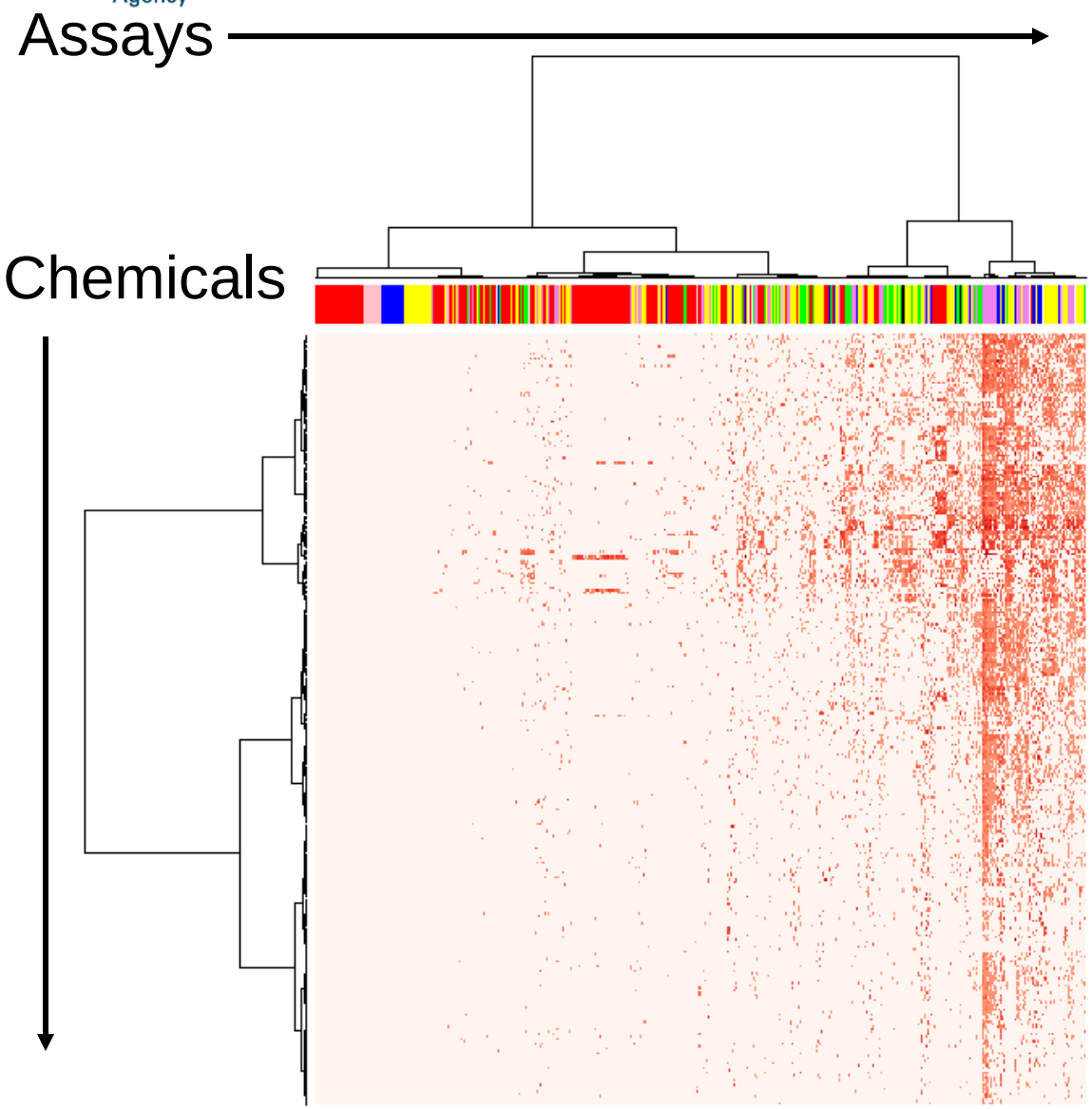
EPA's ToxCast™ Program

- Research program of EPA's National Center for Computational Toxicology (NCCT) (see http://www.epa.gov/ncct/toxcast)
- Addresses chemical screening and prioritization needs for EPA
- Comprehensive use of HTS technologies to generate biological fingerprints and predictive signatures
- Data released via ACToR (Aggregated Computational Toxicology Resource) (http://epa.gov/actor)
  - Contains data on ~500,000 environmental chemicals
  - Multiple Domains - Physchem, biological, use levels, regulations
  - Brings together data from >200 sources

Phased Development of ToxCast™

Phase	Number of Chemicals	Chemical Criteria	Purpose	Number of Assays	Target Date
I	***320***	***Data Rich (pesticides)***	***Signature Development***	***554***	***FY08***
Ib	15	Nanomaterials	Pilot	166	FY09
IIa	>300	Data Rich Chemicals	Validation	>400	FY10
IIb	**>100**	**Known Human Toxicants**	**Extrapolation**	**>400**	**FY10**
IIc	>300	Expanded Structure and Use Diversity	Extension	>400	FY10
IId	>12	Nanomaterials	PMN	>200	FY10
Iie	50 NTP	Immunotoxicants	Signature development	>400	FY10
III	Thousands	Data poor	Prediction and Prioritization	>300	FY 12

The ToxCast™ In Vitro Data Set



Many hits – median value=50

Fewer in cell-free HTS (Novascreen, red)

Many hits are at or near top of tested concentration range

A few are in nanomolar range

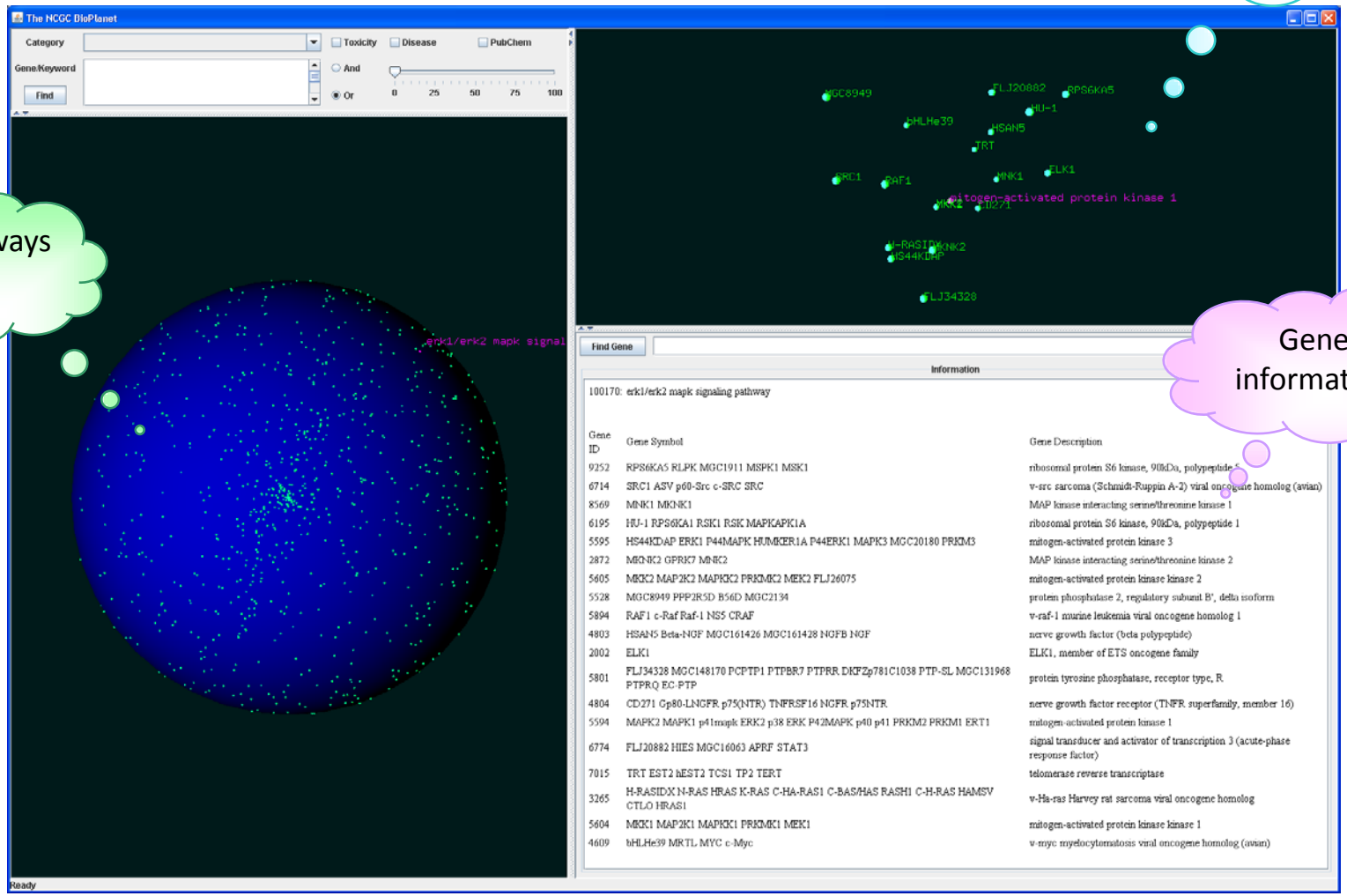
The NCGC BioPlanet

- Hosts the universe of pathways
- All pathway annotations from manually curated, public sources
- Integrates pathways from different data sources
- Annotates pathways by source, species, biological function/process, disease/toxicity relevance, assay availability
- Easy visualization, browsing, analysis of pathways
- Facilitates pathway assay selection/prioritization for Tox21 production phase
- Will be publicly available

The Universe of Pathways

Detailed view of a pathway

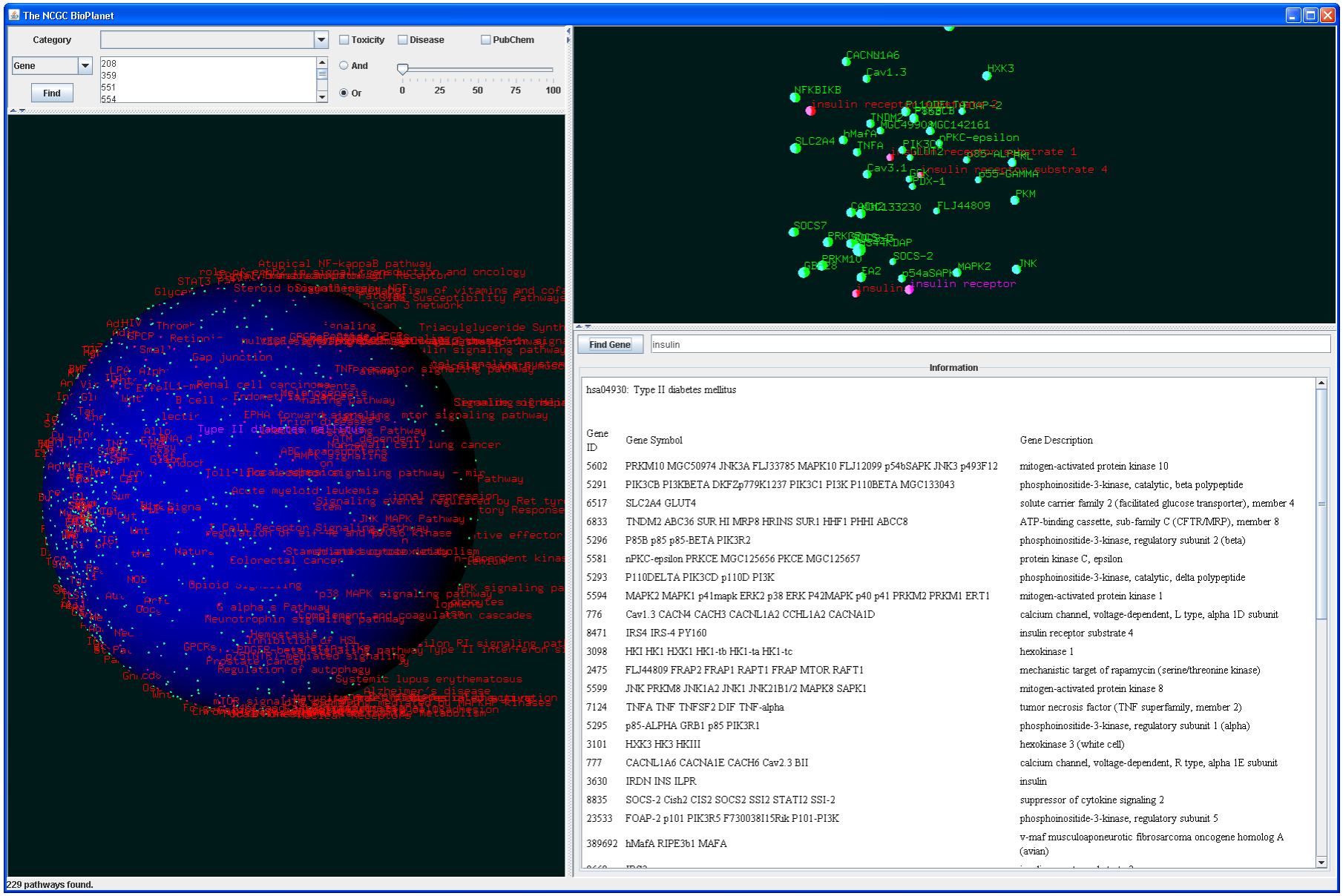
Pathways



Gene information

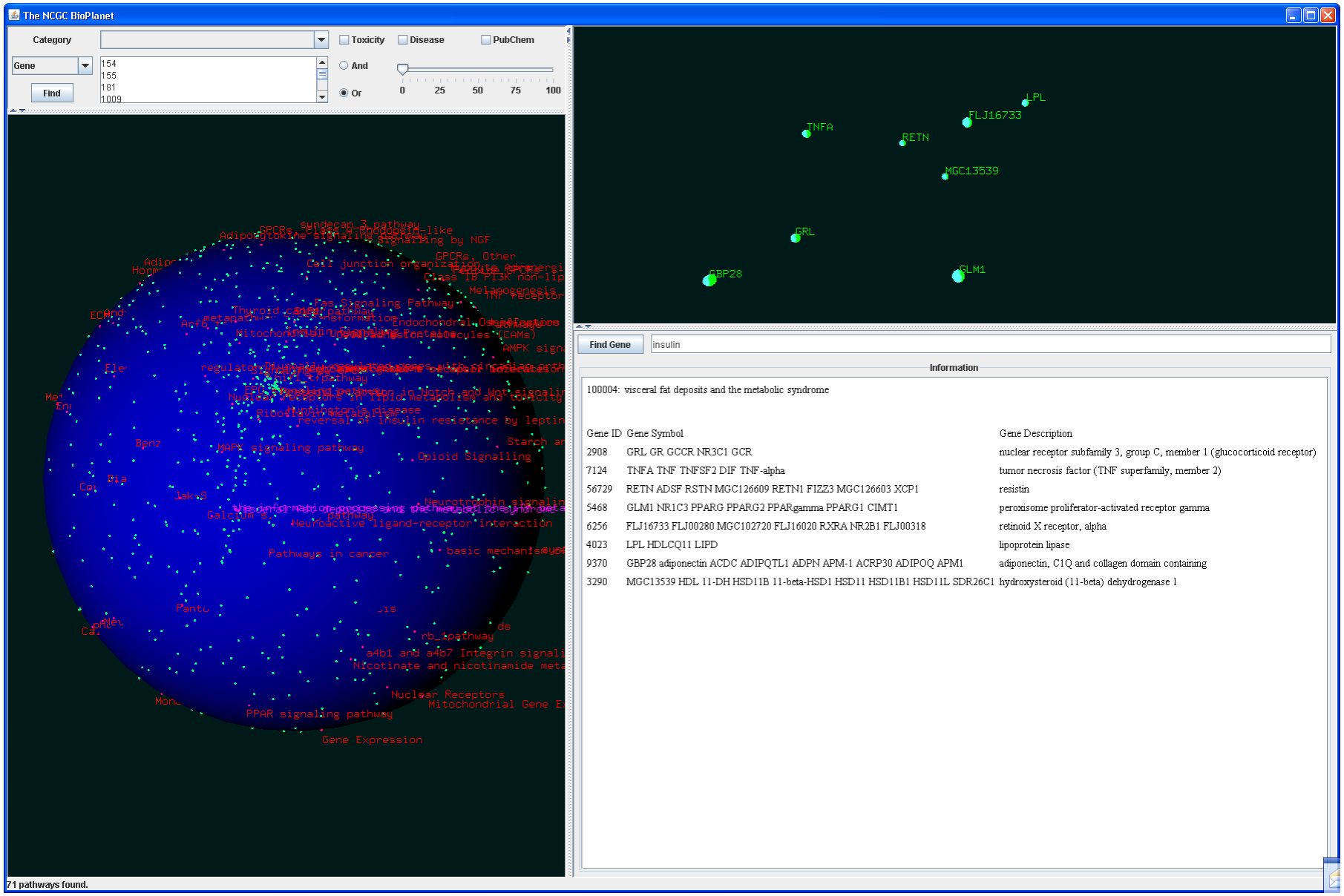
~1100 human pathways mapped to the pathway globe

Pathways containing obesity genes



Gene ID	Gene Symbol	Gene Description
5602	PRKM10 MGC50974 JNK3A FLJ33785 MAPK10 FLJ12099 p54bSAPK JNK3 p493F12	mitogen-activated protein kinase 10
5291	PIK3CB PI3KBETA DKFZp779K1237 PIK3C1 PI3K P110BETA MGC133043	phosphoinositide-3-kinase, catalytic, beta polypeptide
6517	SLC2A4 GLUT4	solute carrier family 2 (facilitated glucose transporter), member 4
6833	TNDM2 ABC36 SUR HI MRP8 HRINS SUR1 HHF1 PHHI ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
5296	P85B p85 p85-BETA PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta)
5581	nPKC-epsilon PRKCE MGC125656 PKCE MGC125657	protein kinase C, epsilon
5293	P110DELTA PIK3CD p110D PI3K	phosphoinositide-3-kinase, catalytic, delta polypeptide
5594	MAPK2 MAPK1 p41mapk ERK2 p38 ERK P42MAPK p40 p41 PRKM1 ERT1	mitogen-activated protein kinase 1
776	Cav1.3 CACN4 CACH3 CACNL1A2 CCHL1A2 CACNA1D	calcium channel, voltage-dependent, L type, alpha 1D subunit
8471	IRS4 IRS-4 PY160	insulin receptor substrate 4
3098	HKI HK1 HXK1 HK1-tb HK1-ta HK1-tc	hexokinase 1
2475	FLJ44809 FRAP2 FRAP1 RAPT1 FRAP MTOR RAFT1	mechanistic target of rapamycin (serine/threonine kinase)
5599	JNK PRKM8 JNK1A2 JNK1 JNK21B1/2 MAPK8 SAPK1	mitogen-activated protein kinase 8
7124	TNFA TNF TNFSF2 DIF TNF-alpha	tumor necrosis factor (TNF superfamily, member 2)
5295	p85-ALPHA GRB1 p85 PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
3101	HXK3 HK3 HKIII	hexokinase 3 (white cell)
777	CACNL1A6 CACNA1E CACH6 Cav2.3 BII	calcium channel, voltage-dependent, R type, alpha 1E subunit
3630	IRDN INS ILPR	insulin
8835	SOCS-2 Cish2 CIS2 SOCS2 SSI2 STATI2 SSI-2	suppressor of cytokine signaling 2
23533	FOAP-2 p101 PIK3R5 F730038I15Rik P101-PI3K	phosphoinositide-3-kinase, regulatory subunit 5
389692	hMafA RIPE3b1 MAFA	v-maf musculoaponeurotic fibrosarcoma oncogene homolog A (avian)

Pathways containing obesity genes

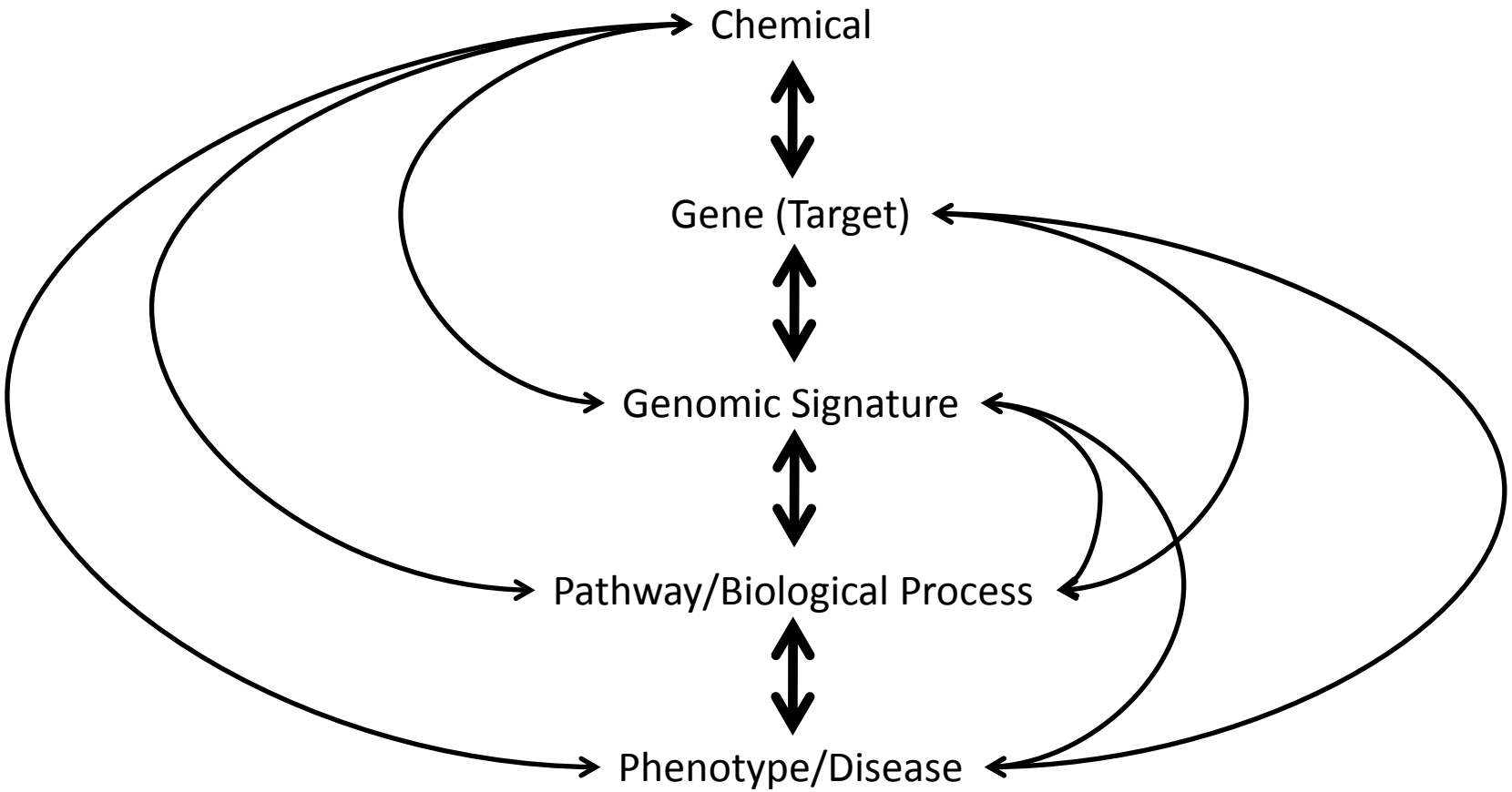


Data Resources

Tox21 Data

- ToxCastDB contains the ToxCast and ToxRefDB data for browsing
  - http://actor.epa.gov/actor/faces/ToxMiner/Home.jsp
- ACToR contains a much large amount of data on other types of toxicity and exposure
  - http://actor.epa.gov
- NTP CEBS (Chemical Effects in Biological Systems) – NTP historical data, DrugMatrix, HTS, *C. elegans*, etc.
  - http://www.niehs.nih.gov/research/resources/databases/cebs/
- PubChem
  - http://pubchem.ncbi.nlm.nih.gov/

Relationships Abound



Database Resources: Chemicals to Genes (Targets)

- DrugBank
  - http://www.drugbank.ca/
- FooDB
  - http://www.drugbank.ca/
- Small Molecule Pathway Database (SMPDB)
  - http://www.smpdb.ca/
- Toxin and Toxin Target Database
  - http://www.t3db.org/
- Potential Drug Target Database
  - http://www.dddc.ac.cn/pdtd/
- KEGG Compound (Ligand)
  - http://www.genome.jp/kegg/compound/
  - http://www.genome.jp/ligand/
- STITCH
  - http://stitch.embl.de/
- PharmGKB
  - http://www.pharmgkb.org/
- PubChem
  - http://pubchem.ncbi.nlm.nih.gov/
- SuperTarget
  - http://bioinf-tomcat.charite.de/supertarget/
- SuperToxic
  - http://bioinf-services.charite.de/supertoxic/
- ChEBML
  - https://www.ebi.ac.uk/chembldb/index.php
- Noted earlier: Ingenuity, GeneGo CoPub, CTD

- PubChem
  - http://pubchem.ncbi.nlm.nih.gov/
- SuperTarget
  - http://bioinf-tomcat.charite.de/supertarget/
- ChEBML
  - https://www.ebi.ac.uk/chembldb/index.php
- GeneGo
  - Hand curated database with variety of gene ontologies
  - http://www.genego.com/
- Ingenuity
  - Hand curated database with variety of gene ontologies
  - http://www.ingenuity.com/
- NextBio
  - Integrated public data mining resource for transcriptomics, CGH, genome-wide methylation
  - http://www.nextbio.com/b/nextbio.nb
- Omic Browser
  - http://omicspace.riken.jp/
- Comparative Toxicogenomics Database (CTD)
  - Text mining (hand curation)
  - http://ctd.mdibl.org/
- Ki Database
  - http://pdsp.med.unc.edu/kidb.php

Database Resource: Chemicals to Genomic Signatures

- ArrayExpress
  - http://www.ebi.ac.uk/arrayexpress/
- Connectivity Map
  - http://www.broadinstitute.org/genome_bio/connectivitymap.html
- DrugMatrix (freely available through NIEHS by mid 2011)
- GEO
  - http://www.ncbi.nlm.nih.gov/geo/
- NextBio
  - http://www.nextbio.com/b/nextbio.nb

Database Resource: Chemicals to Pathways

- CoPub
  - http://services.nbic.nl/copub/portal/
- Comparative Toxicogenomics Database (CTD)
  - http://ctd.mdibl.org/
- GeneGo
  - http://www.genego.com/
- Ingenuity
  - http://www.ingenuity.com/
- KEGG Compound/Ligand
  - http://www.genome.jp/kegg/compound/
  - http://www.genome.jp/ligand/
- Omic Browser
  - http://omicspace.riken.jp/

Database Resources: Chemicals to Phenotype/Disease

- Side Effect Resource (SIDER)
  - http://sideeffects.embl.de/
- Drug Adverse Reaction Database (DART)
  - http://xin.cz3.nus.edu.sg/group/drt/dart.asp
- NTP Database
  - http://ntp.niehs.nih.gov/?objectid=72016020-BDB7-CEBA-F3E5A7965617C1C1
- ToxRefDB
  - http://actor.epa.gov/toxrefdb/faces/Home.jsp
- Carcinogen Potency Database
  - http://potency.berkeley.edu/
- REPDOSE
  - http://www.fraunhofer-repdose.de/
- CHE Toxicant and Disease Database
  - http://www.healthandenvironment.org/tddb
- IARC Monographs
  - http://monographs.iarc.fr/
- ToxNet
  - http://toxnet.nlm.nih.gov/index.html
- ACToR
  - http://actor.epa.gov/actor/faces/ACToRHome.jsp
- DSSTox
  - http://www.epa.gov/ncct/dsstox/
- ATSDR - Health Effects of Exposure to Substances and Carcinogens
  - http://www.atsdr.cdc.gov/substances/ToxOrganSystems.asp
- Comparative Toxicogenomics DataBase
  - http://ctd.mdibl.org/
- Omic Browser
  - http://omicspace.riken.jp/
- eChemPortal
  - http://www.echemportal.org/echemportal/index?pageID=0&request_locale=en

Data Resources: Genes (Targets) to Genomic Signatures

- NextBio
- DrugMatrix (freely available through NIEHS by mid 2011)
- Connectivity Map
  - http://www.broadinstitute.org/genome_bio/connectivitymap.html
- ArrayExpress
  - http://www.ebi.ac.uk/arrayexpress/
- GEO
  - http://www.ncbi.nlm.nih.gov/geo/

Database Resources: Genes (Targets) to Pathways & Biological Processes

- CoPub
  - http://services.nbic.nl/copub/portal/
- Comparative Toxicogenomics Database (CTD)
  - http://ctd.mdibl.org/
- GeneGo
  - http://www.genego.com/
- Ingenuity
  - http://www.ingenuity.com/
- NextBio
  - http://www.nextbio.com/b/nextbio.nb
- Omic Browser
  - http://omicspace.riken.jp/

Database Resources: Genes (Targets) to Disease

- Catalog of Published Genome-Wide Association Studies
  - http://www.genome.gov/gwastudies/#1
- CoPub
  - http://services.nbic.nl/copub/portal/
- Comparative Toxicogenomics Database (CTD)
  - http://ctd.mdibl.org/
- dbGAP
  - http://www.ncbi.nlm.nih.gov/gap
- Entrez Gene
  - http://www.ncbi.nlm.nih.gov/gene
- GeneCards
  - http://www.genecards.org/
- GeneGo
  - http://www.genego.com/
- Human Gene Mutation Database
  - http://www.hgmd.cf.ac.uk/ac/index.php
- Ingenuity
  - http://www.ingenuity.com/
- NextBio
  - http://www.nextbio.com/b/nextbio.nb
- OMIM
  - http://www.ncbi.nlm.nih.gov/omim
- Omic Browser
  - http://omicspace.riken.jp/
- Phenopedia
  - http://hugenavigator.net/HuGENavigator/startPagePhenoPedia.do

Database Resources: Genomic Signatures to Pathways

- DAVID Bioinformatics
  - http://david.abcc.ncifcrf.gov/
- ToppGene
  - http://toppgene.cchmc.org/
- CoPub
  - http://services.nbic.nl/copub/portal/
- GeneGo
  - http://www.genego.com/
- Ingenuity
  - http://www.ingenuity.com/
- GSEA
  - http://www.broadinstitute.org/gsea/index.jsp

Database Resources: Genomic Signatures to Disease

- CoPub
  - http://services.nbic.nl/copub/portal/
- Comparative Toxicogenomics Database (CTD)
  - http://ctd.mdibl.org/
- GeneGo
  - http://www.genego.com/
- Ingenuity
  - http://www.ingenuity.com/
- NextBio
  - http://www.nextbio.com/b/nextbio.nb
- Omic Browser
  - http://omicspace.riken.jp/

Database Resources: Pathways to Disease

- CoPub
  - http://services.nbic.nl/copub/portal/
- Comparative Toxicogenomics Database (CTD)
  - http://ctd.mdibl.org/
- GeneGo
  - http://www.genego.com/
- Ingenuity
  - http://www.ingenuity.com/
- NextBio
  - http://www.nextbio.com/b/nextbio.nb
- Omic Browser
  - http://omicspace.riken.jp/

Database Resources: Chemicals to Genomic Signatures

- NextBio
- DrugMatrix (freely available through NIEHS by mid 2011)
- Connectivity Map
  - http://www.broadinstitute.org/genome_bio/connectivitymap.html
- ArrayExpress
  - http://www.ebi.ac.uk/arrayexpress/
- GEO
  - http://www.ncbi.nlm.nih.gov/geo/

Discussion Points & Recommendations

- Use *in* vitro assays to prototype the process instead of expansively looking at gene disease relationships - go with what we know
- Other methods for approaching identification of assays
  - Identify compounds with known molecular pharmacology (e.g. weight gain)
  - Identify receptor or gene for further screening

Discussion Points & Recommendations

- Proceed With Caution - Can we use a chemical biology approach to predict an effect? More to it than *in vitro* screening
- Bioinformatics approach is not complete
- Cellular phenotype assays rather than gene centric