Bioinformatics-based identification of assays that inform on disease hazard

Scott Auerbach, Raymond Tice and Kristine Thayer

National Toxicology Program, National Institute of Environmental Health Sciences, NIH,DHHS, Research Triangle Park, NC 27709

NIEHS National Institute of Environmental Health Sciences

Abstract

Animal-based hazard characterization, although notably comprehensive in its coverage of disease space, is incapable of detecting disease hazards where animal models are underdeveloped (e.g., autism). In order to identify disease-centric molecular targets that can be developed into high throughput screens for the detection of disease-specific hazard, we have performed a bioinformatic analysis that employs multi-database/gene-centric literature mining, gene class information, and expression. First disease-gene associations were mined from a number of resources including the Comparative Toxicogenomics Database, Ingenuity Pathways Knowledgebase, GeneGo Knowledgebase, OMIM, Entrez Gene, CoPub, GeneCards, and Phenopedia. Associated genes were then filtered based on whether they encode a protein with a 'druggable domain' meaning there is a high probability that they will interact with small molecules. Using this approach, we have identified screening targets for autism (e.g., GLO1, OXTR, GABRA5, SLC6A4, CHRNA4), Type 1 diabetes (e.g.,CTLA4, PTPN22, IL2RA, ITPR3), Type 2 diabetes (e.g., PPARG, KCNJ11, HNF4A, ABCC8) and obesity (e.g., PPARG, PPARA, SERPINE1, PPARD). Gene-linked assay data from phase 1 of the EPA's ToxCast program was then used in combination with our bioinformatic-based analysis of disease to identify hypothetical relationships between chemical exposure and disease hazard.

Introduction

Some of the most common causes of morbidity and mortality in the developed world, such as diabetes, obesity, heart disease, and neurological disease, are not comprehensively queried in standard chemical toxicity studies. Complex diseases are often expensive to evaluate and specifically designed studies are needed to capture chemical hazards related to these diseases. For this reason, in vitro assays that can provide signals to justify if such studies would be of significant value.

As genomics has advanced, the scientific community has begun to better understand the molecular level events that lead to a wide-variety of human health effects. Such discoveries have allowed for the in vitro modeling of disease-related pathophysiological processes and such models that can be used to determine the impact of environmental agents on such processes. Before a large scale endeavor can be undertaken to model and query human disease processes in vitro, the relationships between disease biology and testable biological space needs to be established.

The goal of this project is to create a meta-database that relates genes, pathways, and biological processes to human disease and subsequently to identify the chemical genomic space within these relationships that can be exploited to query the effects of chemicals on molecular processes related to human disease.

Methods

Creation of Disease to Gene Data Resource. A list of all known human genes was derived through the Entrez Gene Database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene). Disease-gene relationships were then extracted from a variety of hand curated (Comparative Toxicogenomics Database (http://ctd.mdibl.org/), Ingenuity Pathways Knowledgebase (http://ingenuity.com/), GeneGo Knowledgebase (http://www.genego.com/), OMIM (http://www.ncbi.nlm.nih.gov/omim) and autocurated (Entrez Gene, CoPub (http://services.nbic.nl/copub/portal/) , GeneCards (http://www.genecards.org/) , and Phenopedia (http://www.hugenavigator.net/HuGENavigator/startPagePhenoPedia.do) databases. All databases were accessed in November 2011. Gene found to be associated with diseases in these databases were mapped to the list of human genes. If an association came from a hand-curated database it was assigned a score of "2". Associations from auto-curated databases were assigned a score of "1". A cumulative score for all genes for a given disease was then calculated by summing the scores for the genes across all databases. In order to gauge the plausibility that a given gene would be likely to serve as a molecular target of a small molecule, a 'druggability' score was determined by gene searching for the presence of a gene on six different druggable genome lists downloaded from Sophic Alliance (Ensembl, DrugBank, InterPro, BioLT, Novartis, and Qiagen). Each time a gene was found on a list, a score of "1" was assigned and these scores were summed over all six druggable genomes. Hence, the highest druggability score was "6". The score cutoffs for reporting a gene-disease linkage was an "8" for disease-gene and "3" for druggability.

Identification of chemicals plausibly associated with disease phenotypes. In order to identify chemicals that may impact the disease phenotypes, the above described analysis was used to evaluate ToxCast Phase 1 AC50 data downloaded from ToxCastDB (http://actor.epa.gov/actor/faces/ToxCastDB/Home.jsp). All chemicals that did not exhibit activity in a given ToxCast assay were assigned an AC50 of 10000 μM. A subset of the Toxcast assays have been mapped to genomic targets (genes). Only results from gene-linked assays were considered for identifying chemical-disease relationships. Genes with existing assay data are highlighted in bold in the tables. Disease scores for individual assay/chemical combinations were calculated as follows:

Disease score for single assay/chemical =1/(AC50)/disease-gene score^{2})

Cumulative disease scores for all chemicals were calculated by identifying those chemicals that exhibited the highest average disease score over all assays associated with a given disease.

Results

Table 1.

Autism

Hand curated: CTD, Ingenuity, GeneGo, OMIM; Automated curation: Entrez Gene, CoPub, GeneCards, Phenopedia

Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Disease-Gene Score	Druggability Score	Existing Assays
glyoxalase I	2	2	2	2	1	1	1	1	12	3	0
gamma-aminobutyric acid (GABA) A receptor, alpha 5	2	2	2	0	1	1	1	1	10	5	1
oxytocin receptor	2	2	2	0	1	1	1	1	10	6	1
solute carrier family 6 (neurotransmitter transporter, serotonin),	2	2	2	0	1	1	1	1	10	5	1
gamma-aminobutyric acid (GABA) A receptor, alpha 2	2	2	2	0	1	1	1	1	10	5	0
gamma-aminobutyric acid (GABA) A receptor, alpha 4	2	2	2	0	1	1	1	1	10	5	0
gamma-aminobutyric acid (GABA) A receptor, beta 3	2	2	2	0	1	1	1	1	10	5	0
phosphatase and tensin homolog	2	2	2	0	1	1	1	1	10	4	0
cholinergic receptor, nicotinic, alpha 4	2	2	2	0	0	1	1	1	9	6	1
cholinergic receptor, nicotinic, alpha 7	2	2	2	0	0	1	1	1	9	5	1
gamma-aminobutyric acid (GABA) A receptor, alpha 1	2	2	2	0	0	1	1	1	9	5	1
5-hydroxytryptamine (serotonin) receptor 2A	2	2	2	0	0	1	1	1	9	6	1
met proto-oncogene (hepatocyte growth factor receptor)	2	0	2	2	0	1	1	1	9	5	1
4-aminobutyrate aminotransferase	2	2	2	0	1	0	1	1	9	5	0
arginine vasopressin receptor 1A	2	0	2	0	1	1	1	1	8	5	1
glutamate receptor, ionotropic, N-methyl D-aspartate 2A	2	2	2	0	0	0	1	1	8	4	1
5-hydroxytryptamine (serotonin) receptor 1A	2	2	2	0	0	0	1	1	8	6	1
monoamine oxidase A	2	0	2	0	1	1	1	1	8	6	1
cholinergic receptor, nicotinic, beta 2 (neuronal)	2	2	2	0	0	1	1	0	8	5	0
dopamine receptor D3	2	2	0	0	1	1	1	1	8	6	0
fragile X mental retardation 1	2	0	2	0	1	1	1	1	8	3	0
glutamate receptor, ionotropic, kainate 2	2	0	2	0	1	1	1	1	8	4	0
integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	2	0	2	0	1	1	1	1	8	6	0
reelin	2	0	2	0	1	1	1	1	8	3	0
tryptophan hydroxylase 2	2	0	2	0	1	1	1	1	8	3	0

Figure 1.

Autism – Top ToxCast Hits

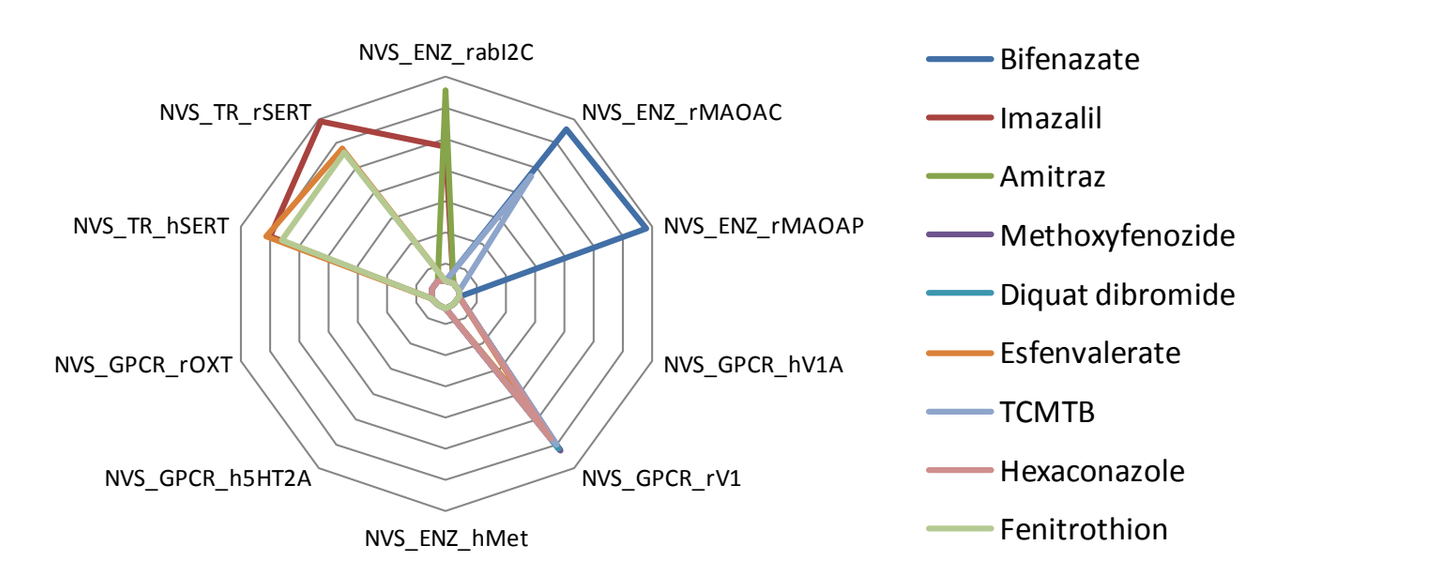


Table 2.

Type 1 Diabetes

Hand curated: CTD, Ingenuity, GeneGo, OMIM; Automated curation: Entrez Gene, CoPub, GeneCards, Phenopedia

Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Disease-Gene Score	Druggability Score	Existing Assays
cytotoxic T-lymphocyte-associated protein 4	2	2	2	2	1	1	1	1	12	4	0
insulin	2	2	2	2	1	1	1	1	12	3	0
protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	2	2	2	2	1	0	1	1	11	5	0
interleukin 2 receptor, alpha	2	2	2	2	1	0	1	1	11	4	0
inositol 1,4,5-triphosphate receptor, type 3	2	2	2	2	1	0	1	1	11	3	0
peroxisome proliferator-activated receptor gamma	2	2	2	0	1	1	1	1	10	6	1
insulin-like growth factor 1 (somatomedin C)	2	2	2	0	1	1	1	1	10	4	0
2',5'-oligoadenylate synthetase 1, 40/46kDa	2	2	2	2	0	0	1	1	10	3	0
vitamin D (1,25- dihydroxyvitamin D3) receptor	2	2	2	0	1	0	1	1	9	6	1
potassium inwardly-rectifying channel, subfamily J, member 11	2	2	2	0	1	0	1	1	9	6	1
aldehyde dehydrogenase 2 family (mitochondrial)	2	2	2	0	1	0	1	1	9	6	0
tumor necrosis factor	2	2	2	0	1	0	1	1	9	5	1
transforming growth factor, beta 1	2	2	2	0	1	0	1	1	9	5	1
paraoxonase 1	2	2	2	0	0	1	1	1	9	5	0
nuclear factor of kappa light polypeptide gene enhancer in B-ce	2	2	2	0	1	0	1	1	9	5	1
insulin receptor	2	2	2	0	0	1	1	1	9	5	1
interleukin 1, beta	2	2	2	0	1	0	1	1	9	5	0
chemokine (C-C motif) receptor 5	2	0	2	2	1	0	1	1	9	5	0
interleukin 12B (natural killer cell stimulatory factor 2, cytotoxic	2	2	2	0	1	0	1	1	9	4	1
interleukin 18 (interferon-gamma-inducing factor)	2	2	2	0	1	0	1	1	9	3	0
lymphotoxin alpha (TNF superfamily, member 1)	0	2	2	0	1	1	1	1	8	5	0
angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	0	2	2	0	1	1	1	1	8	5	0
interferon, gamma	2	2	2	0	0	0	1	1	8	5	0
retinoid X receptor, beta	2	2	2	0	0	0	1	1	8	4	1
transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	0	2	2	0	1	1	1	1	8	3	0
MHC class I polypeptide-related sequence A	0	2	2	0	1	1	1	1	8	3	0
leptin	2	0	2	0	1	1	1	1	8	3	0
insulin receptor substrate 1	2	0	2	0	1	1	1	1	8	3	0
CD69 molecule	2	2	2	0	1	0	1	0	8	3	1

Figure 2.

Type 1 Diabetes – Top ToxCast Hits

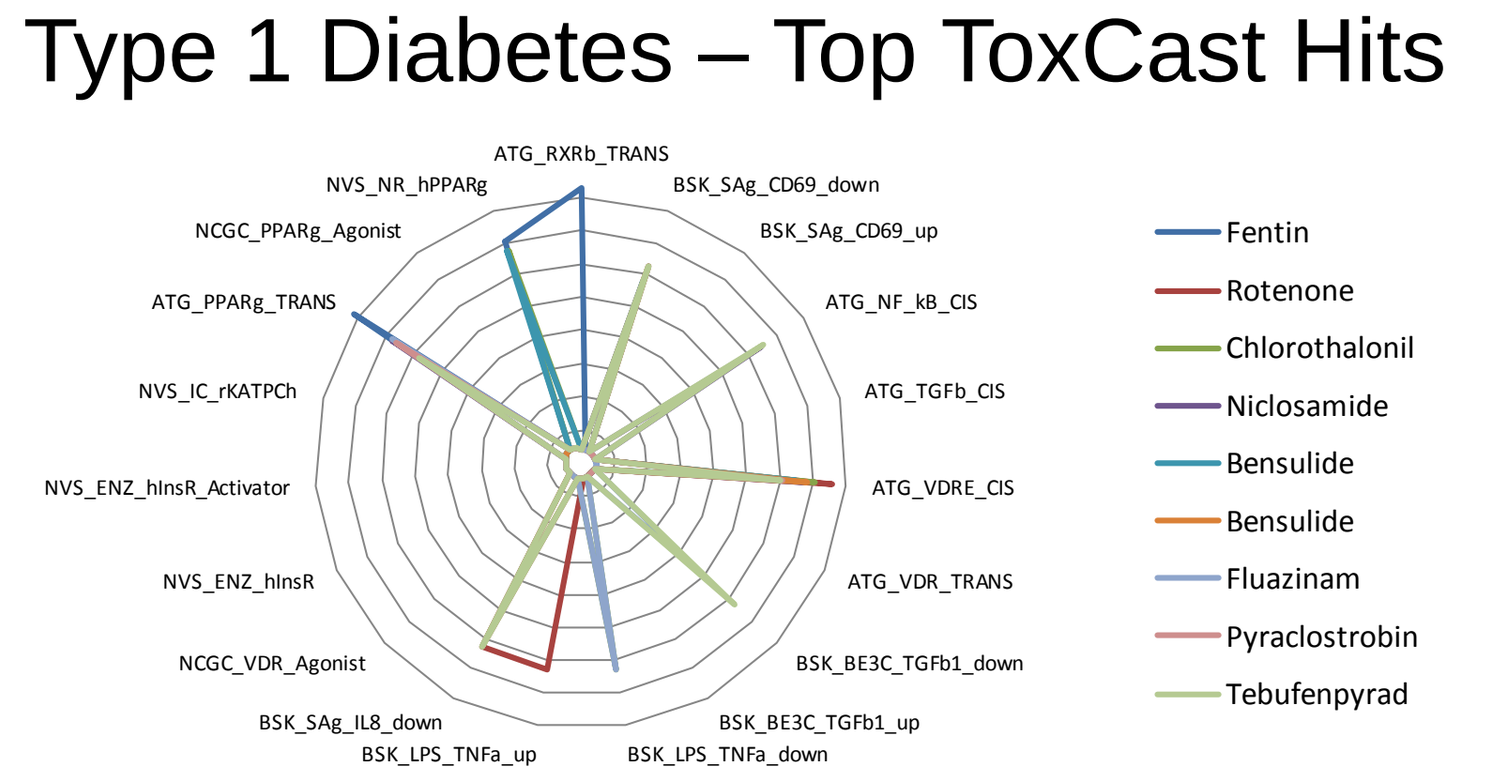


Table 3.

Type 2 Diabetes

Hand curated: CTD, Ingenuity, GeneGo, OMIM; Automated curation: Entrez Gene, CoPub, GeneCards, Phenopedia

Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Disease-Gene Score	Druggability Score	Existing Assays
peroxisome proliferator-activated receptor gamma	2	2	2	2	1	1	1	1	12	6	1
potassium inwardly-rectifying channel, subfamily J, member 11	2	2	2	2	1	1	1	1	12	6	1
hepatocyte nuclear factor 4, alpha	2	2	2	2	1	1	1	1	12	6	1
ATP-binding cassette, sub-family C (CFTR/MRP), member 8	2	2	2	2	1	1	1	1	12	6	0
ectonucleotide pyrophosphatase/phosphodiesterase 1	2	2	2	2	1	1	1	1	12	5	0
glucagon receptor	2	2	2	2	1	1	1	1	12	4	0
neurogenic differentiation 1	2	2	2	2	1	1	1	1	12	3	0
lipase, hepatic	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 2	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 1	2	2	2	2	1	1	1	1	12	3	0
glucokinase (hexokinase 4)	2	2	2	2	1	1	1	1	12	3	0
interleukin 6 (interferon, beta 2)	2	2	2	2	1	0	1	1	11	5	1
vascular endothelial growth factor A	2	2	2	2	1	0	1	1	11	3	0
glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	2	2	2	2	0	1	1	1	11	3	0
peroxisome proliferator-activated receptor alpha	2	2	2	0	1	1	1	1	10	6	1
lipoprotein lipase	2	2	2	0	1	1	1	1	10	6	0
apolipoprotein B (including Ag(x) antigen)	2	2	2	0	1	1	1	1	10	6	0
ATP-binding cassette, sub-family A (ABC1), member 1	2	2	2	0	1	1	1	1	10	6	0
serpin peptidase inhibitor, clade E (nexin, plasminogen activato	2	2	2	0	1	1	1	1	10	5	1
protein tyrosine phosphatase, non-receptor type 1	2	0	2	2	1	1	1	1	10	5	1
insulin receptor	2	2	2	0	1	1	1	1	10	5	1
v-akt murine thymoma viral oncogene homolog 2	2	2	2	2	0	0	1	1	10	5	1
angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	2	2	2	0	1	1	1	1	10	5	0
angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	2	2	2	0	1	1	1	1	10	5	0
uncoupling protein 3 (mitochondrial, proton carrier)	2	2	2	0	1	1	1	1	10	4	0
solute carrier family 2 (facilitated glucose transporter), member 1	2	2	2	0	1	1	1	1	10	4	0
cholesteryl ester transfer protein, plasma	2	2	2	0	1	1	1	1	10	4	0
transcription factor 7-like 2 (T-cell specific, HMG-box)	2	0	2	2	1	1	1	1	10	3	0
phosphoenolpyruvate carboxykinase 1 (soluble)	2	2	2	0	1	1	1	1	10	3	0
leptin receptor	2	2	2	0	1	1	1	1	10	3	0
leptin	2	2	2	0	1	1	1	1	10	3	0
insulin	2	2	2	0	1	1	1	1	10	3	0
mitogen-activated protein kinase 8 interacting protein 1	2	2	2	2	1	0	1	0	10	3	0
vitamin D (1,25- dihydroxyvitamin D3) receptor	2	2	2	0	1	0	1	1	9	6	1
prostaglandin-endoperoxide synthase 2 (prostaglandin G/H syn	2	2	2	0	1	0	1	1	9	6	1
androgen receptor	2	2	2	0	1	0	1	1	9	6	1
v-akt murine thymoma viral oncogene homolog 1	2	2	2	0	1	0	1	1	9	6	1
tumor necrosis factor	2	2	2	0	1	0	1	1	9	5	1
transforming growth factor, beta 1	2	2	2	0	1	0	1	1	9	5	1
nuclear receptor subfamily 3, group C, member 1 (glucocorticoi	2	2	2	0	1	0	1	1	9	5	1
nuclear factor of kappa light polypeptide gene enhancer in B-ce	2	2	2	0	1	0	1	1	9	5	1
low density lipoprotein receptor	2	2	2	0	1	0	1	1	9	5	1
intercellular adhesion molecule 1	2	2	2	0	1	0	1	1	9	5	1
cholinergic receptor, muscarinic 3	2	2	2	0	1	0	1	1	9	5	1
angiotensin II receptor, type 1	2	2	2	0	0	1	1	1	9	5	1
transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamy	2	2	2	0	1	0	1	1	9	5	0
nitric oxide synthase 3 (endothelial cell)	2	2	2	0	1	0	1	1	9	5	0
interleukin 1, beta	2	2	2	0	1	0	1	1	9	5	0
carnitine palmitoyltransferase 1A (liver)	2	2	2	0	1	0	1	1	9	5	0
solute carrier family 2 (facilitated glucose transporter), member 4	2	0	2	2	1	1	1	0	9	5	0
selectin E	2	2	2	0	1	0	1	1	9	4	1
phosphatase and tensin homolog	2	2	2	0	1	0	1	1	9	4	0
poly (ADP-ribose) polymerase 1	2	2	2	0	1	0	1	1	9	4	0
kinesin family member 11	2	2	2	0	1	0	1	1	9	4	0
CD36 molecule (thrombospondin receptor)	2	2	2	0	1	0	1	1	9	3	0
ATP-binding cassette, sub-family G (WHITE), member 1	2	2	2	0	1	1	1	0	9	3	0
insulin-degrading enzyme	0	2	2	0	1	1	1	1	8	6	0
peroxisome proliferator-activated receptor delta	2	0	2	0	1	1	1	1	8	5	1
methylenetetrahydrofolate reductase (NAD(P)H)	2	0	2	0	1	1	1	1	8	5	1
adrenergic, beta-3-, receptor	0	2	2	0	1	1	1	1	8	5	1
sex hormone-binding globulin	2	0	2	0	1	1	1	1	8	5	0
scavenger receptor class B, member 1	2	0	2	0	1	1	1	1	8	5	0
paraoxonase 1	2	0	2	0	1	1	1	1	8	5	0
potassium voltage-gated channel, KQT-like subfamily, member 1	2	0	2	0	1	1	1	1	8	5	0
islet amyloid polypeptide	0	2	2	0	1	1	1	1	8	5	0
glucagon-like peptide 1 receptor	0	2	2	0	1	1	1	1	8	5	0
apolipoprotein E	2	0	2	0	1	1	1	1	8	5	0
protein kinase C, zeta	2	0	2	0	1	1	1	1	8	4	1
uncoupling protein 1 (mitochondrial, proton carrier)	2	0	2	0	1	1	1	1	8	4	0
solute carrier family 2 (facilitated glucose transporter), member 2	2	0	2	0	1	1	1	1	8	4	0
retinol binding protein 4, plasma	2	0	2	0	1	1	1	1	8	4	0
paraoxonase 2	2	0	2	0	1	1	1	1	8	4	0
insulin-like growth factor 1 (somatomedin C)	2	0	2	0	1	1	1	1	8	4	0
albumin	2	2	0	0	1	1	1	1	8	4	0
aldo-keto reductase family 1, member B1 (aldose reductase)	2	0	2	0	1	1	1	1	8	4	0
microsomal triglyceride transfer protein	2	0	2	0	1	1	1	1	8	3	0
hydroxysteroid (11-beta) dehydrogenase 1	2	0	2	0	1	1	1	1	8	3	0
guanine nucleotide binding protein (G protein), beta polypeptide 3	2	0	2	0	1	1	1	1	8	3	0
ghrelin/obestatin prepropeptide	2	0	2	0	1	1	1	1	8	3	0
apolipoprotein A-I	2	0	2	0	1	1	1	1	8	3	0

Figure 3.

Type 2 Diabetes – Top ToxCast Hits

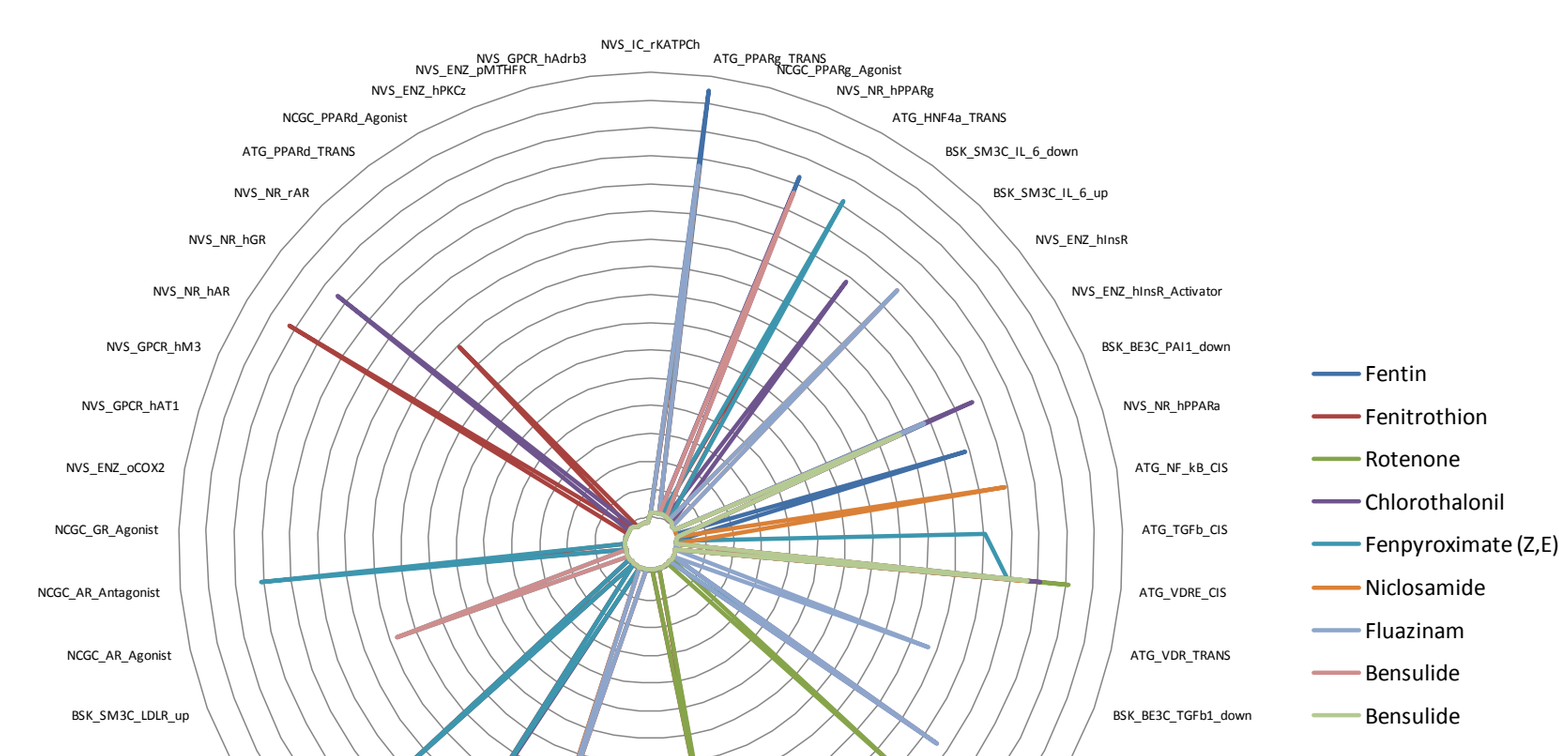


Table 4.

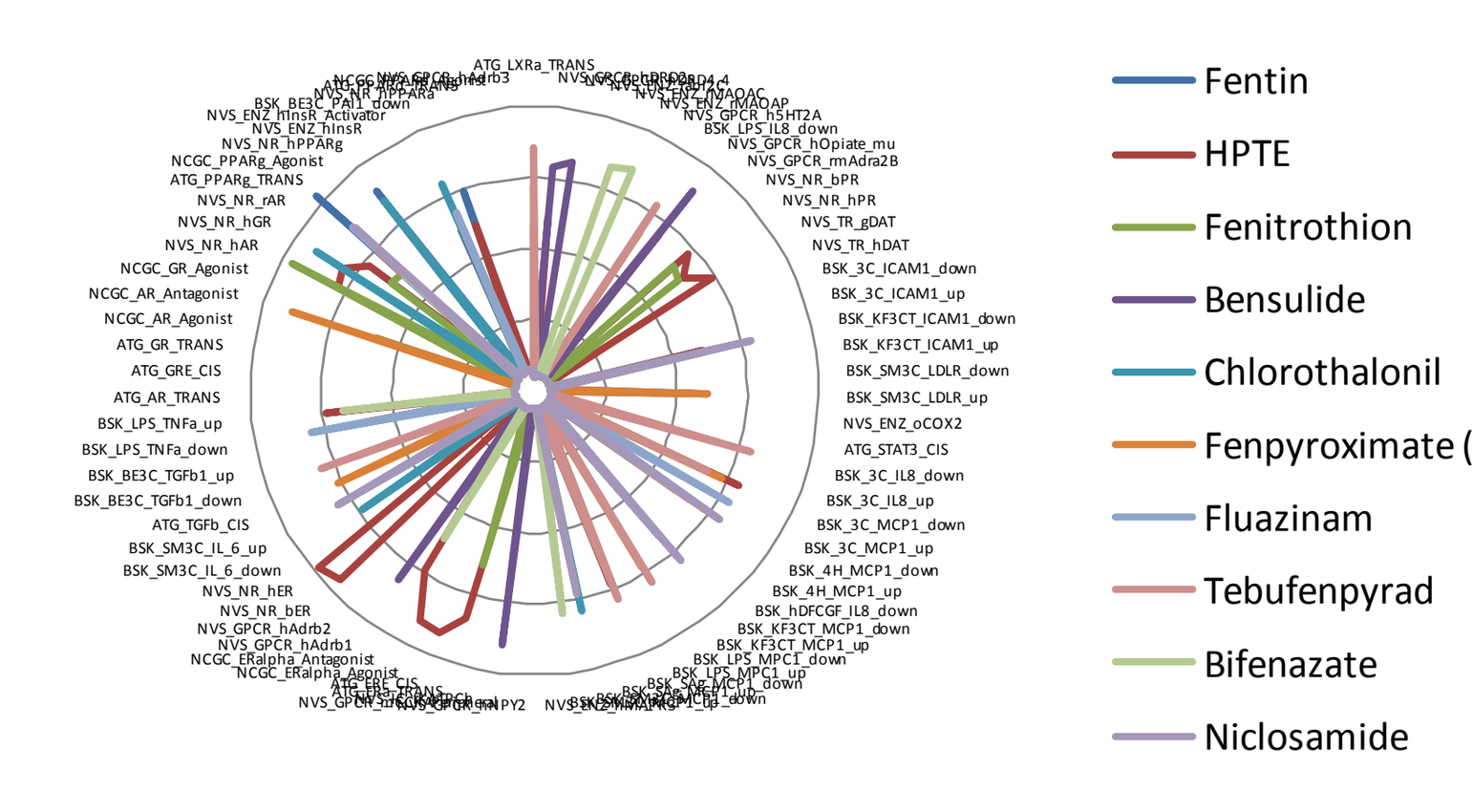
Obesity

Hand curated: CTD, Ingenuity, GeneGo, OMIM; Automated curation: Entrez Gene, CoPub, GeneCards, Phenopedia

Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Disease-Gene Score	Druggability Score	Existing Assays
peroxisome proliferator-activated receptor gamma	2	2	2	2	1	1	1	1	12	6	1
peroxisome proliferator-activated receptor alpha	2	2	2	2	1	1	1	1	12	6	1
serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibit	2	2	2	2	1	1	1	1	12	5	1
peroxisome proliferator-activated receptor delta	2	2	2	2	1	1	1	1	12	5	1
insulin receptor	2	2	2	2	1	1	1	1	12	5	1
adrenergic, beta-3-, receptor	2	2	2	2	1	1	1	1	12	5	1
proprotein convertase subtilisin/kexin type 1	2	2	2	2	1	1	1	1	12	5	0
ectonucleotide pyrophosphatase/phosphodiesterase 1	2	2	2	2	1	1	1	1	12	5	0
cannabinoid receptor 1 (brain)	2	2	2	2	1	1	1	1	12	5	0
neuropeptide Y	2	2	2	2	1	1	1	1	12	4	1
melanocortin 4 receptor	2	2	2	2	1	1	1	1	12	4	0
melanocortin 3 receptor	2	2	2	2	1	1	1	1	12	4	0
fatty acid binding protein 4, adipocyte	2	2	2	2	1	1	1	1	12	4	0
pro-melanin-concentrating hormone	2	2	2	2	1	1	1	1	12	3	0
neuropeptide Y receptor Y5	2	2	2	2	1	1	1	1	12	3	0
lipase, hepatic	2	2	2	2	1	1	1	1	12	3	0
leptin receptor	2	2	2	2	1	1	1	1	12	3	0
leptin	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 2	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 1	2	2	2	2	1	1	1	1	12	3	0
insulin	2	2	2	2	1	1	1	1	12	3	0
hydroxysteroid (11-beta) dehydrogenase 1	2	2	2	2	1	1	1	1	12	3	0
guanine nucleotide binding protein (G protein), beta polypeptide 3	2	2	2	2	1	1	1	1	12	3	0
GNAS complex locus	2	2	2	2	1	1	1	1	12	3	0
ghrelin/obestatin prepropeptide	2	2	2	2	1	1	1	1	12	3	0
aquaporin 7	2	2	2	2	1	1	1	1	12	3	0
agouti related protein homolog (mouse)	2	2	2	2	1	1	1	1	12	3	0
5-hydroxytryptamine (serotonin) receptor 2C	2	2	2	2	1	0	1	1	11	6	1
estrogen receptor 1	2	2	2	2	1	0	1	1	11	6	1
androgen receptor	2	2	2	2	1	0	1	1	11	6	1
adrenergic, beta-1-, receptor	2	2	2	2	1	0	1	1	11	6	1
tumor necrosis factor	2	2	2	2	1	0	1	1	11	5	1
transforming growth factor, beta 1	2	2	2	2	1	0	1	1	11	5	1
nuclear receptor subfamily 3, group C, member 1 (glucocorticoid recep	2	2	2	2	1	0	1	1	11	5	1
interleukin 6 (interferon, beta 2)	2	2	2	2	1	0	1	1	11	5	1
proopiomelanocortin	2	2	2	2	1	0	1	1	11	5	0
nuclear receptor subfamily 0, group B, member 2	2	2	2	2	1	0	1	1	11	5	0
fatty acid synthase	2	2	2	2	1	0	1	1	11	5	0
cytochrome P450, family 19, subfamily A, polypeptide 1	2	2	2	2	1	0	1	1	11	5	0
apolipoprotein E	2	2	2	2	1	0	1	1	11	5	0
angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	2	2	2	2	1	0	1	1	11	5	0
adrenergic, beta-2-, receptor, surface	2	2	2	2	1	0	1	1	11	4	1
solute carrier family 6 (amino acid transporter), member 14	2	2	2	2	1	0	1	1	11	4	0
dipeptidyl-peptidase 4	2	2	2	2	1	0	1	1	11	4	0
toll-like receptor 4	2	2	2	2	1	0	1	1	11	3	1
growth hormone 1	2	2	2	2	1	0	1	1	11	3	0
CART prepropeptide	2	2	2	2	1	0	1	1	11	3	0
brain-derived neurotrophic factor	2	2	2	2	1	0	1	1	11	3	0
hypocretin (orexin) neuropeptide precursor	2	2	2	2	1	1	1	0	11	3	0
potassium inwardly-rectifying channel, subfamily J, member 11	2	0	2	2	1	1	1	1	10	6	1
cholecystokinin A receptor	2	0	2	2	1	1	1	1	10	6	1
lipoprotein lipase	2	0	2	2	1	1	1	1	10	6	0
apolipoprotein B (including Ag(x) antigen)	2	0	2	2	1	1	1	1	10	6	0
protein tyrosine phosphatase, non-receptor type 1	2	0	2	2	1	1	1	1	10	5	1
chemokine (C-C motif) ligand 2	2	0	2	2	1	1	1	1	10	5	1
mitogen-activated protein kinase 3	2	2	2	2	1	0	1	0	10	5	1
interleukin 8	2	2	2	2	0	0	1	1	10	5	1
solute carrier family 2 (facilitated glucose transporter), member 4	2	0	2	2	1	1	1	1	10	5	0
glucagon-like peptide 1 receptor	2	2	2	0	1	1	1	1	10	5	0
gastric inhibitory polypeptide receptor	0	2	2	2	1	1	1	1	10	5	0
cathepsin S	2	2	2	0	1	1	1	1	10	5	0
neurotrophic tyrosine kinase, receptor, type 2	2	2	2	2	1	0	1	0	10	5	0
signal transducer and activator of transcription 3 (acute-phase respon	2	2	2	0	1	1	1	1	10	4	1
uncoupling protein 3 (mitochondrial, proton carrier)	2	0	2	2	1	1	1	1	10	4	0
uncoupling protein 1 (mitochondrial, proton carrier)	2	0	2	2	1	1	1	1	10	4	0
retinol binding protein 4, plasma	2	0	2	2	1	1	1	1	10	4	0
growth hormone secretagogue receptor	2	0	2	2	1	1	1	1	10	4	0
mitogen-activated protein kinase 8	2	2	2	2	1	0	1	0	10	4	0
neuropeptide Y receptor Y2	2	2	2	0	1	1	1	1	10	3	1
transcription factor 7-like 2 (T-cell specific, HMG-box)	2	0	2	2	1	1	1	1	10	3	0
TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	2	0	2	2	1	1	1	1	10	3	0
suppressor of cytokine signaling 3	2	2	2	0	1	1	1	1	10	3	0
sirtuin (silent mating type information regulation 2 homolog) 1 (S. cerevisiae)	2	0	2	2	1	1	1	1	10	3	0
neuromedin U	0	2	2	2	1	1	1	1	10	3	0
microsomal triglyceride transfer protein	2	0	2	2	1	1	1	1	10	3	0
glucokinase (hexokinase 4)	0	2	2	2	1	1	1	1	10	3	0
diacylglycerol O-acyltransferase homolog 1 (mouse)	2	2	0	2	1	1	1	1	10	3	0
ciliary neurotrophic factor	2	0	2	2	1	1	1	1	10	3	0
angiopoietin-like 6	2	2	0	2	1	1	1	1	10	3	0
complement component 3	2	2	2	2	1	0	1	0	10	3	0
solute carrier family 6 (neurotransmitter transporter, dopamine), meml	2	2	2	0	1	0	1	1	9	6	1
prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase a	2	2	0	2	1	0	1	1	9	6	1
progesterone receptor	2	2	2	0	1	0	1	1	9	6	1
monoamine oxidase A	2	2	2	0	1	0	1	1	9	6	1
5-hydroxytryptamine (serotonin) receptor 2A	0	2	2	2	1	0	1	1	9	6	1
opioid receptor, mu 1	2	2	2	0	1	0	1	1	9	5	1
low density lipoprotein receptor	2	2	2	0	1	0	1	1	9	5	1
intercellular adhesion molecule 1	2	2	2	0	1	0	1	1	9	5	1
nitric oxide synthase 3 (endothelial cell)	2	2	0	2	1	0	1	1	9	5	0
glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa)	0	2	2	2	1	0	1	1	9	5	0
carboxypeptidase E	2	2	0	2	1	0	1	1	9	5	0
angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	2	0	2	2	1	0	1	1	9	5	0
hydroxysteroid (11-beta) dehydrogenase 2	2	2	2	0	1	1	1	0	9	5	0
adrenergic, alpha-2B-, receptor	0	2	2	2	1	0	1	1	9	4	1
uncoupling protein 2 (mitochondrial, proton carrier)	2	0	2	2	1	0	1	1	9	4	0
tumor necrosis factor receptor superfamily, member 1A	2	2	2	0	1	0	1	1	9	4	0
cholesteryl ester transfer protein, plasma	2	0	2	2	1	0	1	1	9	4	0
interleukin 10	2	2	2	0	1	0	1	1	9	3	1
sorbin and SH3 domain containing 1	0	2	2	2	0	1	1	1	9	3	0
protein kinase, cAMP-dependent, regulatory, type II, beta	2	2	0	2	0	1	1	1	9	3	0
interleukin 1 receptor antagonist	2	2	0	2	1	0	1	1	9	3	0
interleukin 18 (interferon-gamma-inducing factor)	2	0	2	2	1	0	1	1	9	3	0
delta-like 1 homolog (Drosophila)	0	2	2	2	1	0	1	1	9	3	0
corticotropin releasing hormone receptor 1	2	2	2	0	1	0	1	1	9	3	0
clock homolog (mouse)	0	2	2	2	1	0	1	1	9	3	0
CD14 molecule	2	2	2	0	1	0	1	1	9	3	0
dopamine receptor D2	0	2	2	0	1	1	1	1	8	6	1
v-akt murine thymoma viral oncogene homolog 1	2	0	2	2	1	0	1	0	8	6	1
v-akt murine thymoma viral oncogene homolog 2	2	0	2	2	0	1	1	0	8	5	1
acetyl-CoA carboxylase beta	2	2	0	0	1	1	1	1	8	5	0
acetyl-CoA carboxylase alpha	2	0	0	2	1	1	1	1	8	5	0
tyrosine hydroxylase	2	0	2	2	0	0	1	1	8	5	0
nuclear receptor subfamily 1, group H, member 3	2	0	2	0	1	1	1	1	8	4	1
dopamine receptor D4	0	2	2	2	0	0	1	1	8	4	1
insulin-like growth factor binding protein 3	2	0	2	0	1	1	1	1	8	4	0
cholecystokinin	2	0	2	2	0	0	1	1	8	4	0
ribosomal protein S6 kinase, 70kDa, polypeptide 1	2	2	0	2	1	0	1	0	8	4	0
serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), mem	2	0	2	0	1	1	1	1	8	3	0
inositol polyphosphate phosphatase-like 1	0	2	0	2	1	1	1	1	8	3	0
CD36 molecule (thrombospondin receptor)	2	0	2	2	0	0	1	1	8	3	0
prohibitin	2	0	2	2	1	0	1	0	8	3	0
high mobility group AT-hook 2	0	2	2	2	1	0	1	0	8	3	0

Figure 4.

Obesity– Top ToxCast Hits



Discussion

The vision for toxicology in the 21st century is to be able to use in vitro assay data to predict in vivo phenotypes such as the ones highlighted here. At this current juncture, the primary roadblock to achieving this goal is a molecular/systems level understanding of disease phenotype and its causes. As our understanding of complex disease progresses, it is likely that in vitro assays will be developed that have the capacity to predict specific disease hazards associated with chemical exposure. However, until this point is reached, the field will be bound to what is already been established by in vivo toxicity assessments, as it will be limited to identifying patterns of response in in vitro assays elicited by phenotypically characterized chemicals and then using these patterns to evaluate the chemicals with unknown toxicological properties. This approach will likely be quite effective for predicting the most common toxicities (e.g. liver toxicity) because they are typically the best studied and there are a large number of molecules that can be used to train predictive models. Toxicological assessments are, however, done to identify both common and uncommon toxicities and it is often the uncommon toxicities that are of greatest concern. An example of an uncommon toxicity would be chemical-induced autism. As it stands now, there are very few if any chemicals that have been causally linked to autism, hence it is largely unknown how a chemical would elicit this phenotype. Therefore the question arises, what assays could be used to predict risk of such a phenotype? In recent years molecular genetics has revealed certain causes of syndromic autism. It is plausible that chemicals which perturbed the molecular systems identified by these studies may lead to autism. However, it is also just as likely the biological processes that are completely independent of the molecular networks defined by genetic studies can be altered by chemical exposure and lead to an autistic phenotype. Still a further challenge to a molecular genetics approach to assay identification relates to our limited understanding of the differences in phenotypic outcomes that manifest due to pharmacological alteration as opposed to genetic perturbation. Bearing these limitations in mind, the analysis presented in this poster attempts to leverage data from molecular level phenotyping of disease with the goal of identifying assays that plausibly predict the interaction between chemicals and complex in vivo phenotypes. Considering the noted caveats of the informatics-based approaches used in this analysis, all findings as it relates to chemical-induced disease should be treated purely as hypothetical until validated in an appropriate in vivo model.

Conclusions

- Gene-to-disease data resources have potential utility in the identification of assays that can inform on the hypothetical disease-related chemical perturbations.
- The molecular causes of disease are poorly defined for most complex diseases, therefore making biologically anchored assay development a significant challenge.
- The approach described here identifies disease-gene relationships that are well established in the literature; however the approach lacks the ability to synthesize data and therefore identify novel targets not explicitly identified in the primary information resources and does not identify hierarchical causality or yet differentiate between cause and effect in disease-gene relationships.

Acknowledgements

The authors would like to thank the members of the EPA's National Center for Computational Toxicology for providing access to the assay data necessary for our analysis. We would also like to thank Alex Merrick for critical review of poster presentation. The information presented in this poster has not been peer reviewed and does not necessarily represent the views of the NTP, NIEHS or the United States Government.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health