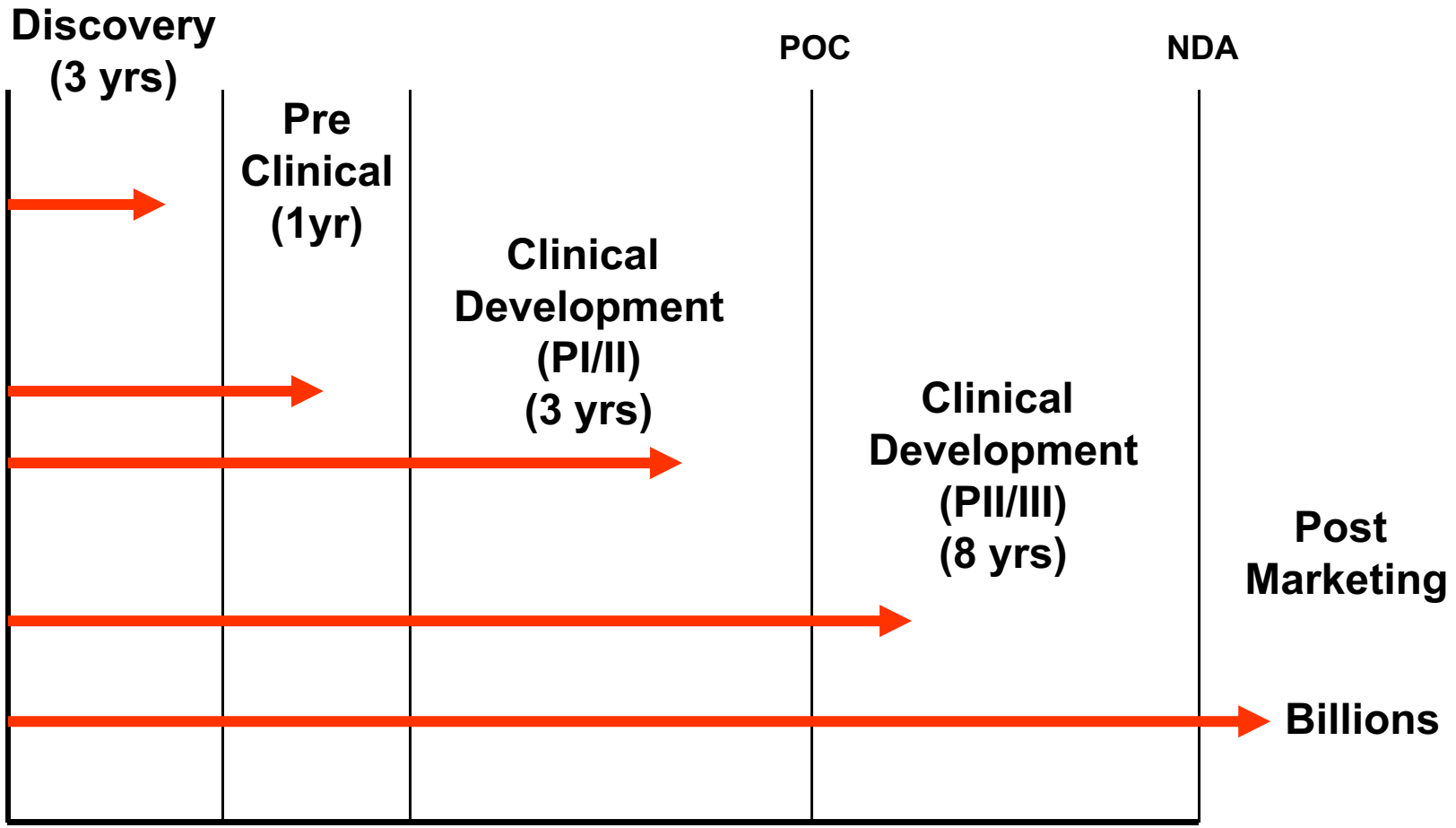
Newer Methods of Predicting Toxicity: Characterizing Mechanisms of Toxicity

Designing Safer Medicines in Discovery
SCI, London
March 2011

Mark Gosink, Ph.D.
Systems Toxicology, Investigative Toxicology
Groton, CT, USA

Fail Early – Saves Time and Money



Difficulties in Predictive Drug Safety

- Chemotype risk prediction involves the identification of chemical sub-structures or features which are common to an adverse event
- Standard methods work well for the prediction of AE with a general cellular mechanism
  - Genotoxicity – polycyclic aromatics, free radicals, etc.
  - Phospholipidosis - cationic amphiphilic
- Affinity to known high-risk targets is also used to improve safety prediction
  - hERG – QT prolongation
  - Cyp450 – drug-drug interactions

- Some types of adverse reactions remain difficult to predict
  - Liver and cardiac injury still account for the majority of drug withdrawals
- Difficulty may lie with the multi-gene-ic nature of these AE's
  - Complex nature of the AE (cardiomyopathy, altered heart rate, and hypertension) can all lead to cardiac arrest

Complex Phenotypes Makes the Development of Predictive Models Difficult

- Identical phenotypic changes can result from multiple genes
- Mouse mutations which result in an 'Increased Systolic Blood Pressure'
  - 52 genes spanning multiple protein classes (ie Ca channels, cyclase's, synthase's, etc.)
  - Other genes in similar pathways also linked to BP

Entrez	Description	Entrez	Description
20928	Abcc9 - ATP-binding cassette, sub-family C (CFTR/MRP), member 9	15452	Hprt - hypoxanthine guanine phosphoribosyl transferase
70008	Ace2 - angiotensin I converting enzyme (peptidyl-dipeptidase A) 2	15461	Hras1 - Harvey rat sarcoma virus oncogene 1
11519	Add2 - adducin 2 (beta)	15932	Idua - iduronidase, alpha-L-
11450	Adipoq - adiponectin, C1Q and collagen domain containing	16000	Igf1 - insulin-like growth factor 1
11607	Agtr1a - angiotensin II receptor, type 1a	16367	Irs1 - insulin receptor substrate 1
11609	Agtr2 - angiotensin II receptor, type 2	384783	Irs2 - insulin receptor substrate 2
11816	Apoe - apolipoprotein E	16534	Kcnn4 - potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4
12297	Cacnb3 - calcium channel, voltage-dependent, beta 3 subunit	16653	Kras - v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
12577	Cdkn1c - cyclin-dependent kinase inhibitor 1C (P57)	16847	Lepr - leptin receptor
53419	Corin - corin	18125	Nos1 - nitric oxide synthase 1, neuronal
19025	Ctsa - cathepsin A	18126	Nos2 - nitric oxide synthase 2, inducible
110115	Cyp11b1 - cytochrome P450, family 11, subfamily b, polypeptide 1	18127	Nos3 - nitric oxide synthase 3, endothelial cell
13109	Cyp2j5 - cytochrome P450, family 2, subfamily j, polypeptide 5	18160	Npr1 - natriuretic peptide receptor 1
13117	Cyp4a10 - cytochrome P450, family 4, subfamily a, polypeptide 10	14815	Nr3c1 - nuclear receptor subfamily 3, group C, member 1
13119	Cyp4a14 - cytochrome P450, family 4, subfamily a, polypeptide 14	20028	Pdc - phosducin
13388	Dll1 - delta-like 1 (Drosophila)	18976	Pomc - pro-opiomelanocortin-alpha
13489	Drd2 - dopamine receptor D2	19091	Prkg1 - protein kinase, cGMP-dependent, type I
13492	Drd5 - dopamine receptor D5	19224	Ptgs1 - prostaglandin-endoperoxide synthase 1
13614	Edn1 - endothelin 1	19225	Ptgs2 - prostaglandin-endoperoxide synthase 2
13618	Ednrb - endothelin receptor type B	51801	Ramp1 - receptor (calcitonin) activity modifying protein 1
13717	Eln - elastin	54409	Ramp2 - receptor (calcitonin) activity modifying protein 2
13857	Epor - erythropoietin receptor	20277	Scnn1b - sodium channel, nonvoltage-gated 1 beta
14226	Fkbp1b - FK506 binding protein 1b	64384	Sirt3 - sirtuin 3 (silent mating type information regulation 2, homolog) 3 (S. cerevisiae)
83554	Fstl3 - follistatin-like 3	22598	Slc6a18 - solute carrier family 6 (neurotransmitter transporter), member 18
60596	Gucy1a3 - guanylate cyclase 1, soluble, alpha 3	53791	Tlr5 - toll-like receptor 5
54195	Gucy1b3 - guanylate cyclase 1, soluble, beta 3	22337	Vdr - vitamin D receptor

How can Systems Biology Approaches be Used to Determine Risk

- Compound-driven approaches
  - Multi-endpoint / High-throughput bioassays are gaining favor as a means to develop toxicity prediction models
  - 'Omics technologies offer the ability to assay 35,000 endpoints simulataneously
  - Like bioassays, 'omics techniques can be to identify 'signatures' of toxicity
    - Can also be used to elucidate mechanism
- Target-driven approaches
  - Can be employed before a compound is developed to identify on-target (or target-class) risks
  - Using pathway analyses, even poorly annotated targets can be evaluated for risk
  - Collections of single gene mutation phenotypes can be used 'in reverse' to identify a likely mechanism for an observed adverse effect

Toxicogenomics Workflow

Transcriptomics/Proteomics/Metabolomics

Compound-driven

Experimental design
- Multiple replicates
- Multiple doses
- Multiple timepoints
- Sample pooling

Data generation
- Signal generation
- Normalization

Identification of differentially expressed genes
- May entail conversion of protein or other IDs to gene IDs

Target-driven

Direct gene information
- Literature mining
- Mutational phenotypes
- Tissue expression

Informatic identification of related genes
- Gene in same pathway
- Genes with similar active site
- Genes with similar phenotype

Multi-Gene Analysis

Toxicogenomics / Systems Biology
- ***Categorical Analysis***
- ***Pathway Analysis***
- ***De novo Networks***
- ***Casual Reasoning***

Toxicogenomics Workflow – Experiment Driven

Transcriptomics/Proteomics/Metabolomics

Compound-driven

Experimental design
- Multiple replicates
- Multiple doses
- Multiple timepoints
- Sample pooling

Data generation
- Signal generation
- Normalization

Identification of differentially expressed genes
- May entail conversion of protein or other IDs to gene IDs

Target-driven

Informatic identification of related genes
- Gene in same pathway
- Genes with similar active site
- Genes with similar phenotype

Multi-Gene Analysis

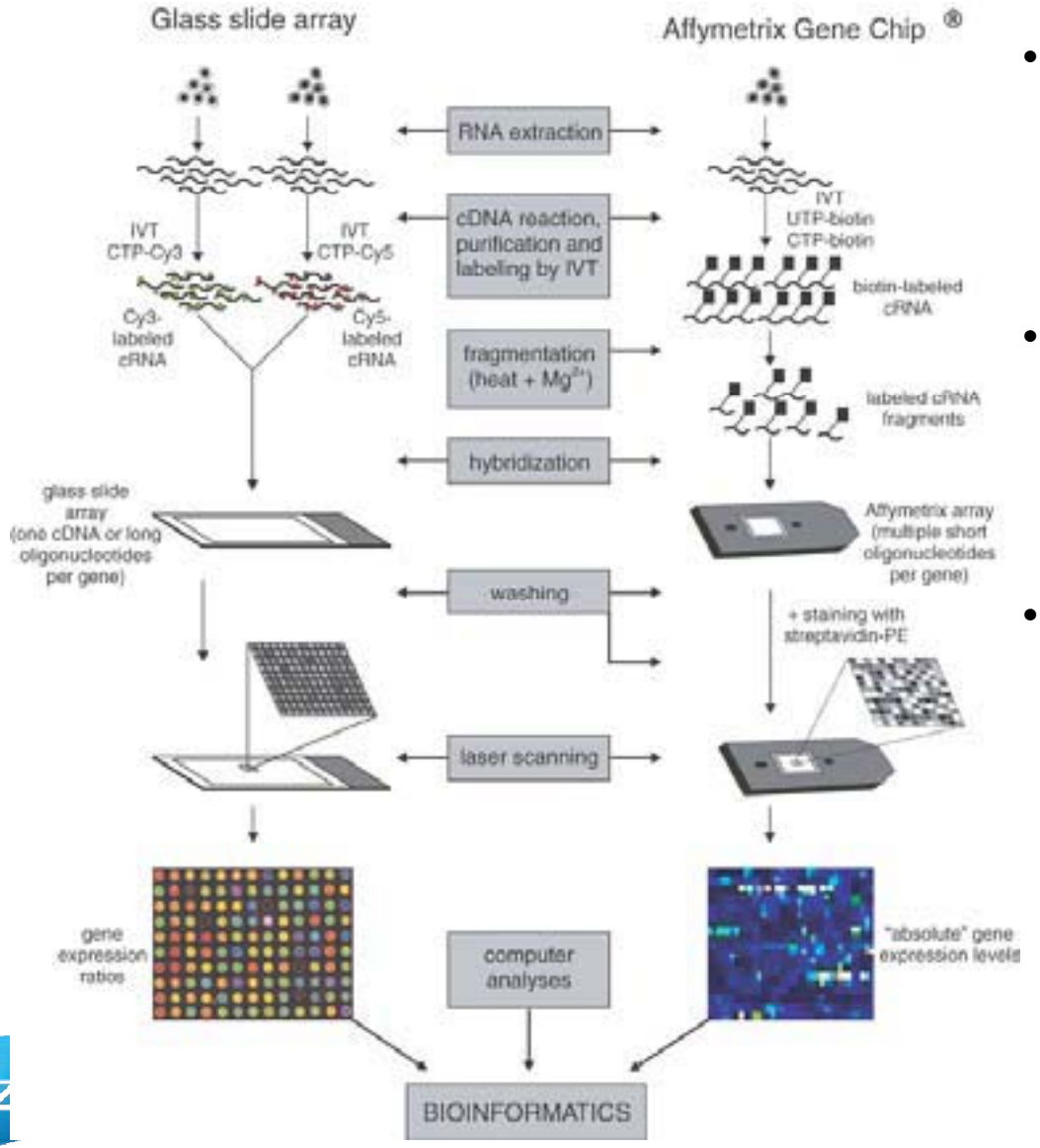
Toxicogenomics / Systems Biology
- *Categorical Analysis*
- *Pathway Analysis*
- *De novo Networks*
- *Casual Reasoning*

Transcriptomics

(Global mRNA Profiling)

- *SAGE – serial analysis of gene expression (outdated)*
- Deep-sequencing / next-generation sequencing
  - Parallel sequencing techniques allow simultaneous sequencing of thousands of samples
- qRT-PCR panels
  - Multiplexed panels (typically focused)
  - Limited numbers of targets but growing
- Microarray Formats
  - Standard
    - Complementary DNA (cDNA)
    - Oligonucleotide (i.e. Affymetrix, & others)
  - Other formats
    - Electrokinetic, fiberoptic, micro/nano-fluidic, bead arrays
      - Some allow multiple samples to be run at the same time
    - miRNA chips

Primary Microarray Formats



- Two color chips tend to be more sensitive; however multiple sample comparisons are difficult
- A number of companies such as Illumina, Agilent, others also make oligonucleotide microarrays
- Both chip types are limited by genes on chip
  - Deep sequencing more useful when genome is not well characterized

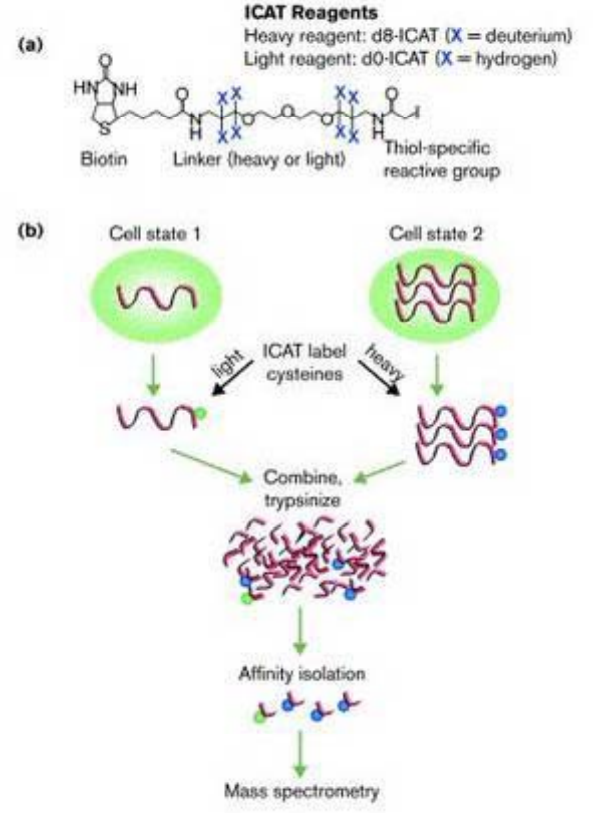
From: Leukemia (2003) 17, 1324–1332. / http://www.nature.com/leu/journal/v17/n7/fig_tab/2402974f1.html

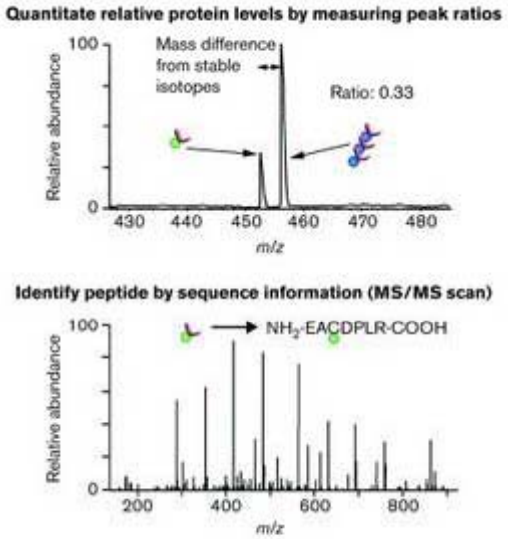
Proteomics

(Global Protein Profiling)

- 1D / 2D PAGE – polyacrylamide gel electrophoresis
  - Sometimes used together with MS for spot identification
- Protein / antibody microarrays
  - Multiple antibodies or samples spotted on array
- ChIP-on-chip – DNA-protein binding arrays
- LC-MS
  - Works best with sequenced genomes for protein database
  - Complexity / post-translation modifications and / or dynamic range of protein concentration in biological samples limits mixture analyses to a few hundred proteins
    - Albumin makes up ~50% of the total protein in normal human plasma or ~10 orders of magnitude greater than the concentration of some cytokines such as IL-6 and IL-8
    - Partial purification techniques: MudPIT (Multidimensional Protein Identification Technology), nanobeads
  - Quantification is difficult due to ionization efficiency differences
    - Stable isotope labeling (can be multiplexed)
      - ICAT, iTRAQ, MeCATs, others

Stable Isotope Labeling





- Because of the difficultly in quantitating proteins from inter-protein MS peak heights, isotopic labeling of proteins is sometimes used to separate treated proteins from controls

Images from: http://www.immun.lth.se/research/protein_technology/what_is_proteomics%3F/peptide_based_proteomics/

Metabolomics

(Global Metabolite Profiling)

- Perhaps the oldest of ‘Omics technologies
  - Urine color & taste in medical diagnoses dates back ~2000 years

- Metabolite Identification
  - Chromatography
    - Gas, HPLC, capillary
  - MS
    - Typically combined with chromatography
  - NMR
    - NMR is non-destructive and can be used in intact tissues

- MS has better sensitivity while NMR gives more liable quantitation
- Metabolomics particularly useful in pharmacogenomics analysis
- Pathway databases sometimes incorporate metabolite information making direct pathway analysis easier

Toxicogenomics Workflow - Analyses

Transcriptomics/Proteomics/Metabolomics

Compound-driven

Experimental design
- Multiple replicates
- Multiple doses
- Multiple timepoints
- Sample pooling

Data generation
- Signal generation
- Normalization

Identification of differentially expressed genes
- May entail conversion of protein or other IDs to gene IDs

Target-driven

Informatic identification of related genes
- Gene in same pathway
- Genes with similar active site
- Genes with similar phenotype

Multi-Gene Analysis

Toxicogenomics / Systems Biology
- ***Categorical Analysis***
- ***Pathway Analysis***
- ***De novo Networks***
- ***Casual Reasoning***

Analysis of Multiple Genes

- ‘Omics approaches can yield 10’s to 1000’s of treatment-related differentially expressed genes
  - Individual gene assessment is difficult

- Categorical Analysis
- Pathway Analysis
- Casual Reasoning
- De novo pathway / network analysis

- Gene set enrichment analysis is commonly used in pathway analyses
- Approach originally developed to compare groups of genes to categorical information
  - Gene Ontology categories most commonly used

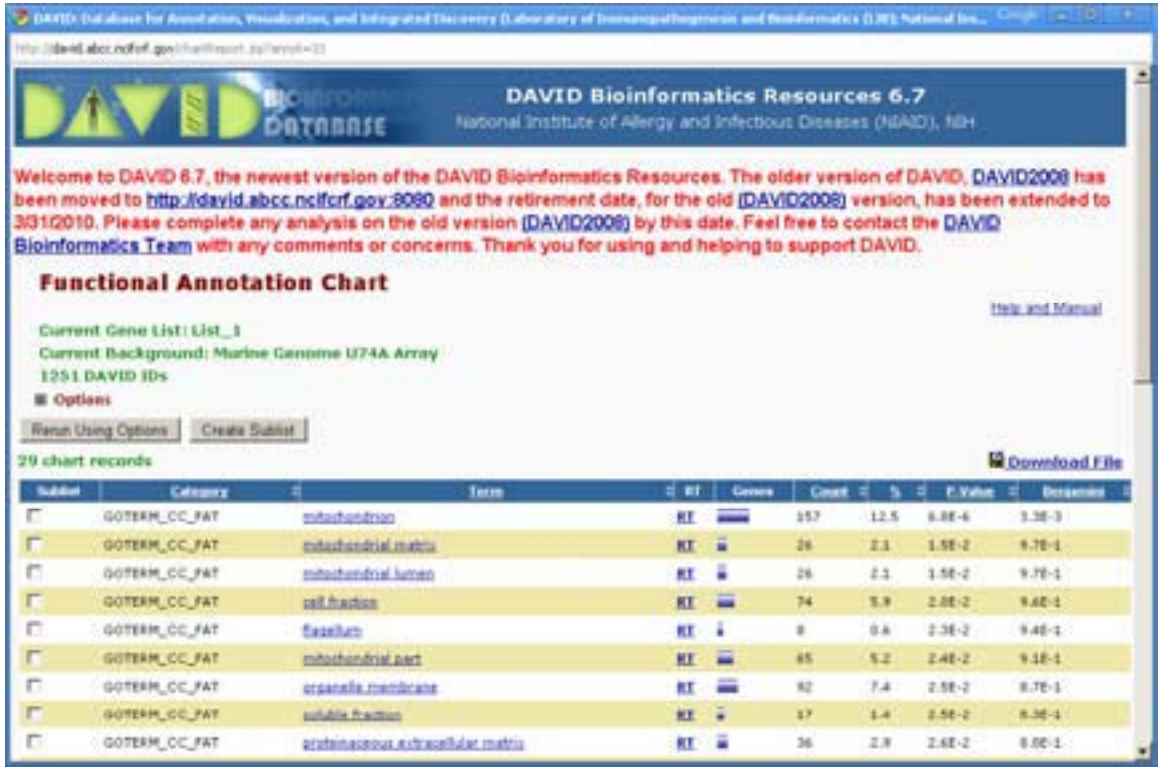
Gene Set Enrichment Analysis (GSEA)

- GSEA is a statistical approach to compare an input list of genes to a collection of gene-sets (i.e. genes in a pathway)

- GSEA identifies statistical over-representation by genes in gene-set collections to genes in your input list
  - Statistical tests vary between tools, some use parametric and others non-parametric test
  - Some tools also consider directionality of change

- A number of gene set enrichment tools are available
  - David, GSEA, Ingenuity, …
  - Each tool may have a number of gene set categories available for analysis

Using Transcriptomics to Identify Potential Toxicity Mechanism



- 3T3-L1 cells treated with 20μM troglitazone
- Expression analyses performed using Affymetrix U74 chips
- Differentially expressed genes analyzed using DAVID
- Most significant GO ‘cellular component’ is mitochondion
- Recent reports suggest mitochondrial damage may play a role in troglitazone liver effects

1: Lim PL, Liu J, Go ML, Boelsterli UA. The mitochondrial superoxide/thioredoxin-2/Ask1 signaling pathway is critically involved in troglitazone-induced cell injury to human hepatocytes. Toxicol Sci. 2008 Feb;101(2):341-9. Epub 2007 Nov 1. PubMed PMID: 17975114.

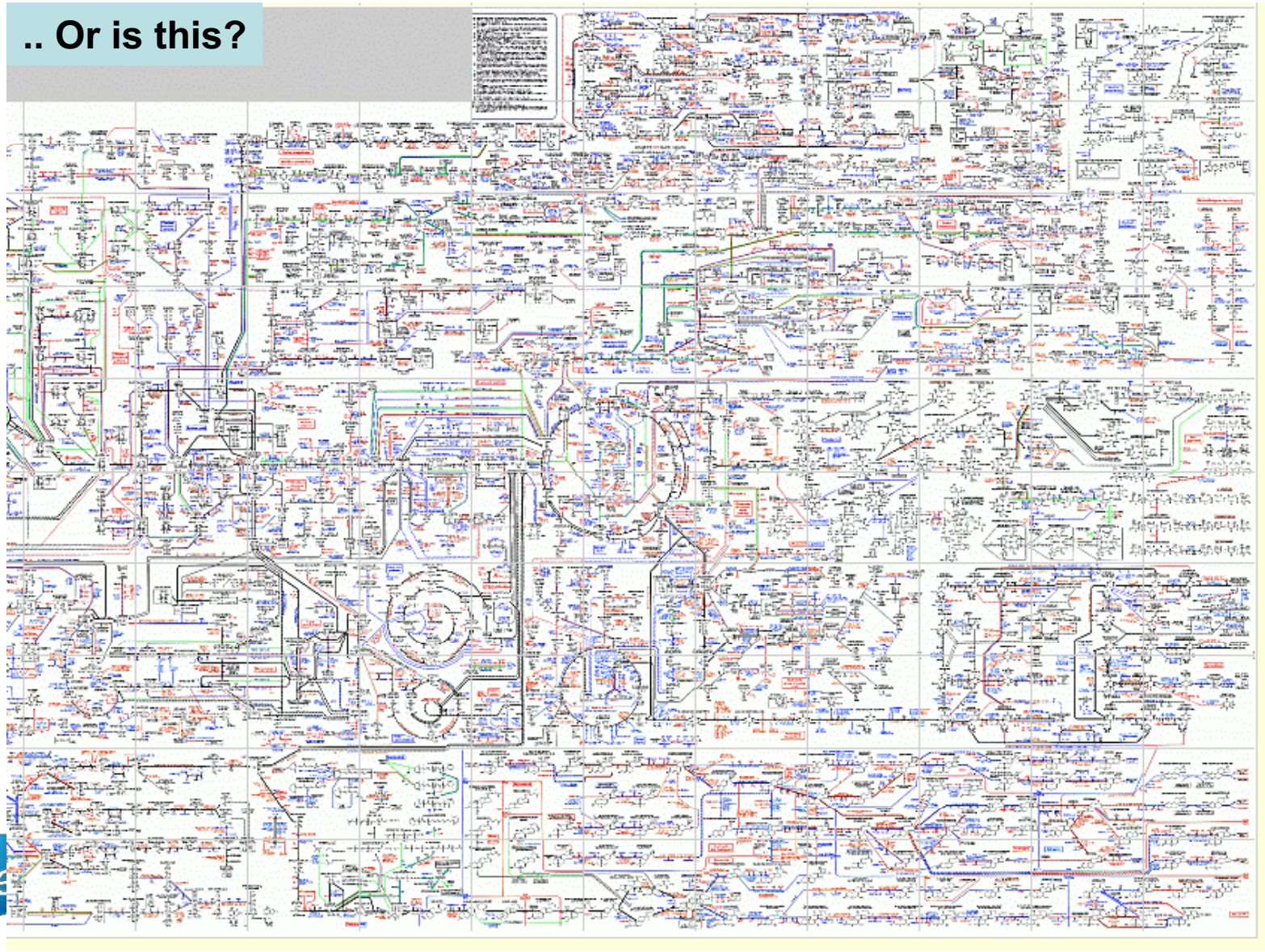
2: Rachek LI, Yuzefovych LV, Ledoux SP, Julie NL, Wilson GL. Troglitazone, but not rosiglitazone, damages mitochondrial DNA and induces mitochondrial dysfunction and cell death in human hepatocytes. Toxicol Appl Pharmacol. 2009 Nov 1;240(3):348-54. Epub 2009 Jul 24. PubMed PMID: 19632256.

Pathway Analysis

- Genes in a pathway are often coordinately regulated
- Most analysis tools treat pathways as another type of category
  - Genes in a single pathway would constitute a gene set
  - Standard gene set enrichment analysis methods can be used to analyze for pathways over-represented in an experiment
- Pathways from different providers general show considerable over lap but individual genes may vary
- Pathways (within a single provider) exhibit a significant amount of overlap
  - Since multiple testing corrections assume independence, they tend to be overly restrictive

What Is A Pathway?

.. Or is this?



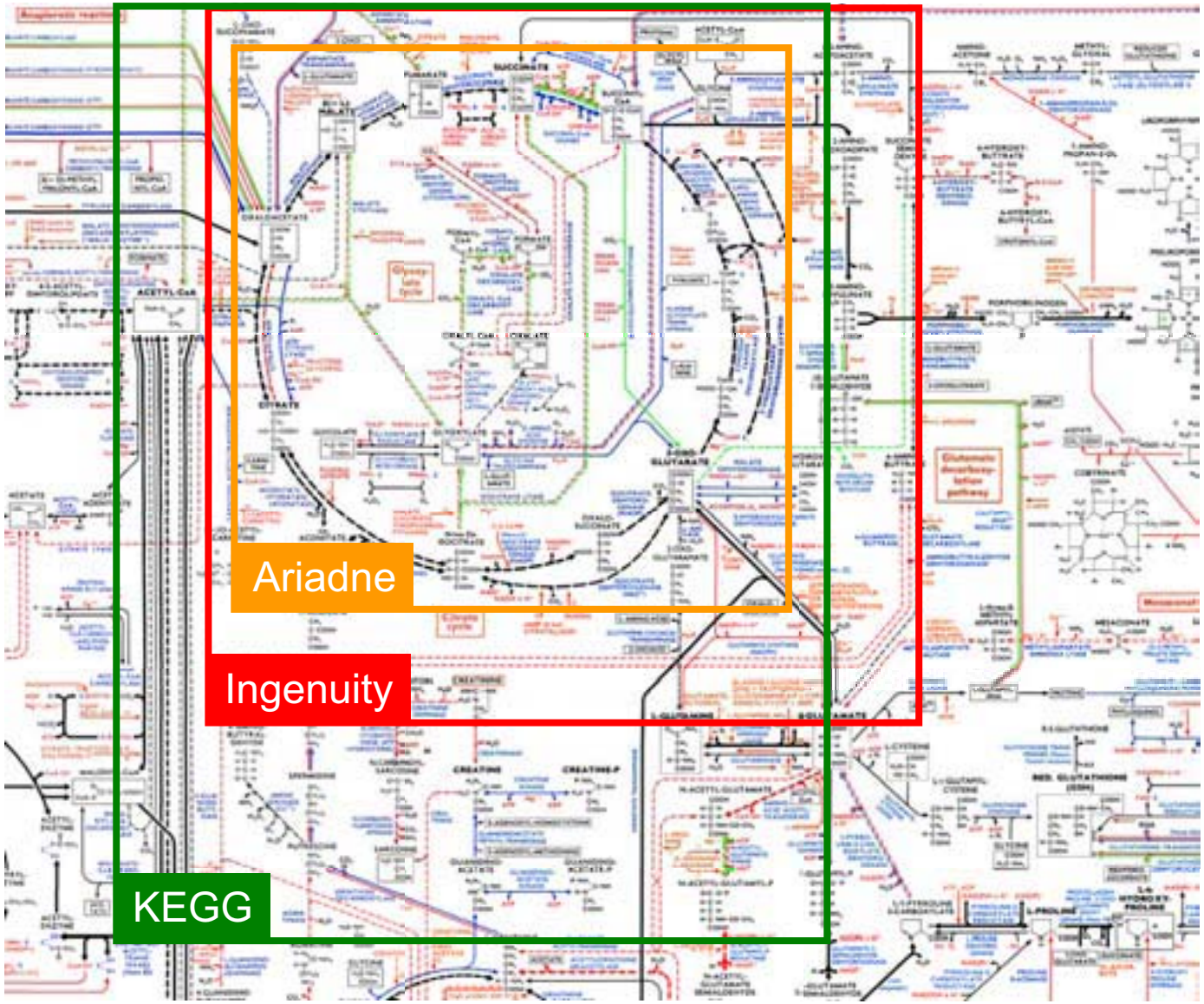
Considerations Before Pathway Analyses

- The ‘true pathway’ contains every gene/protein in the genome as well as all their interactions
  - ‘True pathway’ also varies due to cell-type or developmental stage

- For practical considerations, pathways are a somewhat arbitrary boundary drawn around a group of functionally related proteins/genes

- There are numerous public and private collections of pathway sets
  - Ingenuity, KEGG, BioCarta, Ariandne, Reactome, etc. each with its own definition of genes in a pathway

Considerations Before Pathway Analyses



Considerations Before Pathway Analyses

- The ‘true pathway’ contains every gene/protein in the genome as well as all their interactions
  - ‘True pathway’ also varies due to cell-type or developmental stage
- For practical considerations, pathways are a somewhat arbitrary boundary drawn around a group of functionally related proteins/genes
- There are numerous public and private collections of pathway sets
  - Ingenuity, KEGG, BioCarta, Ariadne, Reactome, etc. each with its own definition of genes in a pathway

- Accurate predictive analyses requires consistent definition of pathways