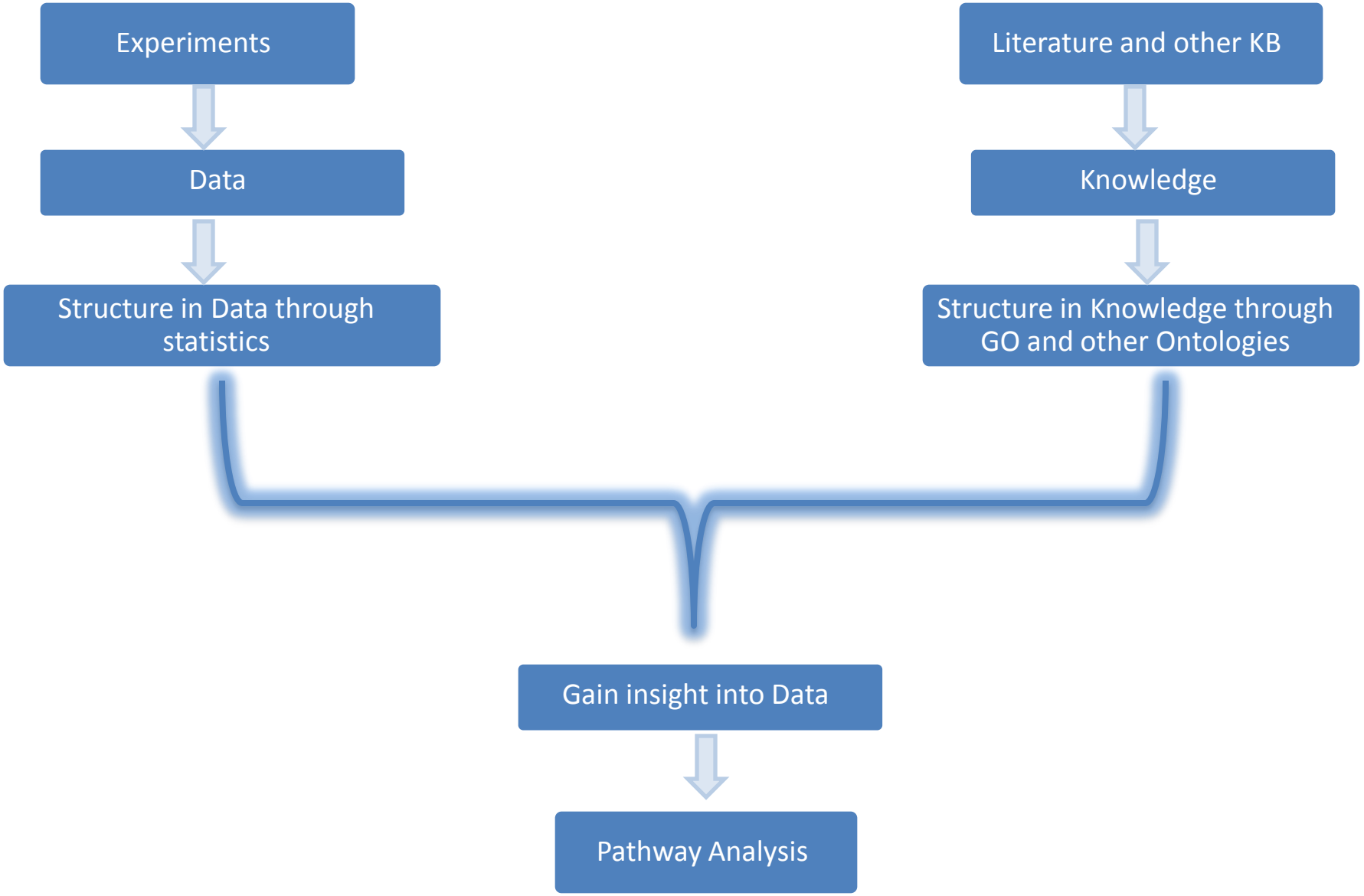
Pathway Analysis :
An Introduction


Experiments
Data
Structure in Data through statistics
Literature and other KB
Knowledge
Structure in Knowledge through GO and other Ontologies
Gain insight into Data
Pathway Analysis

Why Pathway Analysis?

- Logical next step in any high-throughput experiment
  - Treat samples
  - Collect mRNA
  - Label
  - Hybridize
  - Scan
  - Normalize
  - Select differentially regulated genes

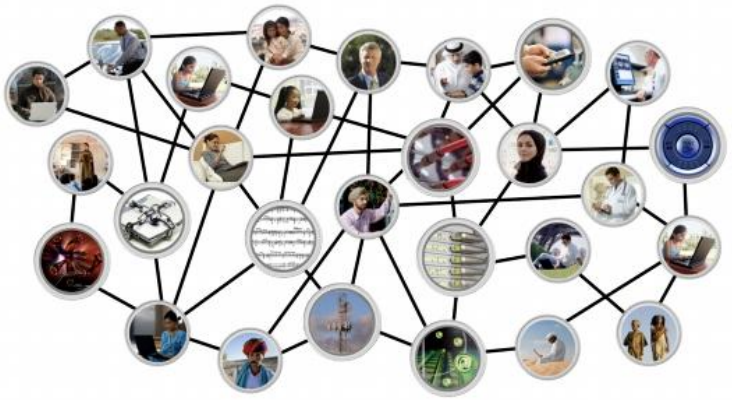
  Microarray Experiment
  - …
  - **Understand the biological phenomena involved**

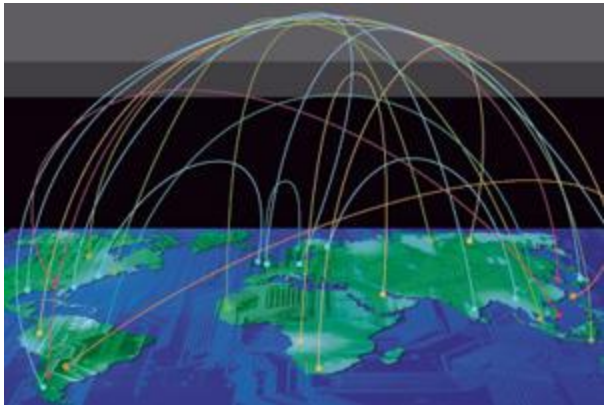
- High-throughput experiments *per se* do not produce biological findings
- Genes do not work alone, but in an intricate network of interactions
- Helps interpret the data in the context of biological processes, pathways and networks
- Global perspective on the data and problem at hand

Trends in Bioinformatics

Sequence Comparison	Today
Functional Comparison	Tomorrow
Pathway Discovery	Bridge to the Future and fully understanding of molecular basis of disease

Remember everything is a *relationship (connected)*… what we are trying to do here is find that relationship (connection)



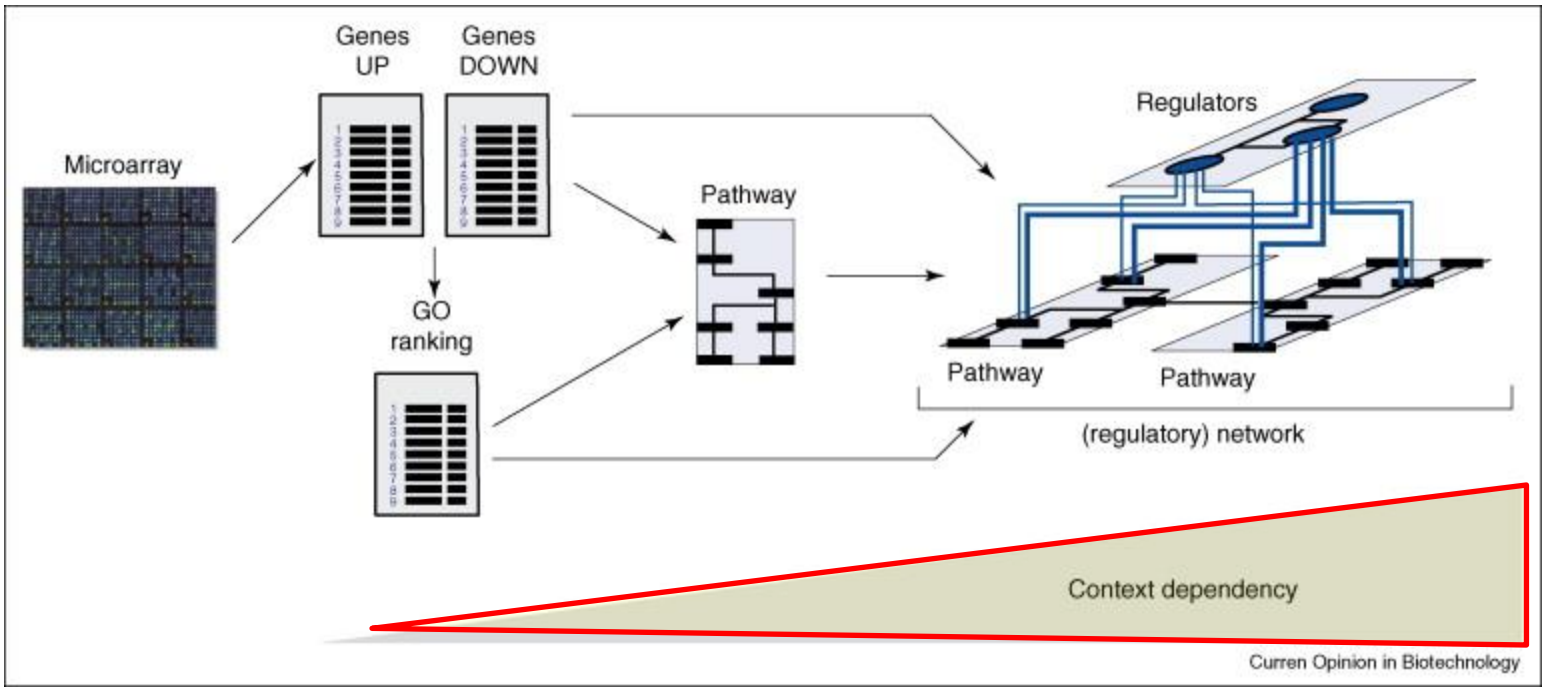


What do we get out of PA?

- In-depth and contextualized findings to help understand the mechanisms of disease in question
- Identification of genes and proteins associated with the etiology of a specific disease
- Prediction of drug targets
- Understand how to intervene therapeutically in disease processes
- Conduct targeted literature searches

What do we get out of PA? …*cont*

- Data integration: integrate diverse biological information
  - Scientific literature, knowledge databases
  - Genome sequences
  - Protein sequences, motifs and structures
- Functional discovery: assign function to genes
  - Only 5% of known genes have assigned functions
  - Without understanding the function, no drug discovery can be done



Werner T. Curr Opion Biotechnology 2008

Available Tools for Pathway Analysis *(non-exhaustive list)*

- GeneGo/MetaCore (www.genego.com)
- Ingenuity Pathway Analysis (www.ingenuity.com)
- Pathway Studio (www. ariadnegenomics.com)
- GenMAPP (www. genmapp.com)
- WikiPathways (www. wikipathways.org)
- cPath (cbio.mskcc.org/cpath)
- BioCyc (www.biocyc.org)
- Pubgene (www.pubgene.org)
- PANTHER (www. pantherdb.org)
- WebGestalt (bioinfo.vanderbilt.edu/webgestalt/)
- ToppGene Suite(/toppgene.cchmc.org/)
- DAVID (david.abcc.ncifcrf.gov/)
- Pathway Painter(pathway.painter.gsa-online.de/)

Available Databases *(non-exhaustive list)*

Database		Description
Protein–Protein Interaction Databases: Organize experimental and/or in silico interactions	BIND	200,000 documented biomolecular interactions and complexes
	MINT	Experimentally verified interactions
	HPRD	Elegant and comprehensive presentation of the interactions, entities, and evidences
	MPact	Yeast interactions. A part of MIPS
	DIP	Experimentally determined interactions
	IntAct	Database and analysis system of binary and multiprotein interactions
	PDZBase	PDZ Domain containing proteins
	GNPV	Based on specific experiments and literature
	BioGrid	Physical and genetic interactions
	UniHi	Comprehensive human protein interactions
	OPHID	Combines PPI from BIND, HPRD, and MINT
Metabolic Pathways Databases: Compendium of pathways describing metabolic and physical processes (Primary source for metabolic information initiated by Stanford Research Initiative)	EcoCyc	Entire genome and biochemical machinery of *E. coli*
	MetaCyc	Pathways of more than 165 species
	HumanCyc	Human metabolic pathways and the human genome
	BioCyc	Collection of databases for several organism
Signaling Pathways Databases: Pathways pertaining to signal transduction	KEGG	Comprehensive. Links to several useful databases
	PANTHER	Compendium of pathways built using CellDesigner
	Reactome	Hierarchical layout. Extensive links to relevant databases
	Biomodels	Domain experts curated pathways and associated mathematical models
	STKE	Repository of canonical pathways
	Ingenuity Systems	Commercial mammalian biological knowledgebase
	PID	Compendium of several assembled signaling pathways
	BioPP	Repository of biological pathways built using CellDesigner

Most databases have a graphics viewer for displaying entities and interactions. Refer to Table S1 for a more detailed description and URLs of these databases.
BIND, Biomolecular Interaction Network Database; BioPP, Biological Pathway Publisher; DIP, Database of Interacting Proteins; EcoCyc, Encyclopaedia of *E. coli* Genes and Metabolism; GNPV, Genome Network Platform Viewer; HPRD, Human Protein Reference Database; KEGG, Kyoto Encyclopedia of Genes and Genomes; MetaCyc, a Metabolic Pathway database; MINT, Molecular INTeration database; MIPS, Munich Information center for Protein Sequences; OPHID, Online Predicted Human Interaction Database; PANTHER, Protein Analysis through Evolutionary Relationship database; PID, The Pathway Interaction Database; STKE, Signal Transduction Knowledge Environment; UNIHI, Unified Human Interactome.
doi:10.1371/journal.pcbi.0040016.t001

Why Pathway Analysis Software?

- A learning tool
  - Study a group of gene products.
- A data analysis tool.
  - Which pathways are particularly affected?
  - What disease has similar biomarkers?
- A hypothesis generation tool
  - Can provide insight into mechanisms of regulation of your genes.
    Which is the likely causative agent for the observed changes?
    What is likely to happen as a result of these changes?
  - Suggest effects of gene knock-in or knock-outs.
  - Suggest side-effects of drugs.
  - Can highlight new phenomena that needs further investigation.
    What does the program <u>not</u> explain?

Caveat *or* how far the tools will take you in your quest for knowledge

- Tools are new
- Databases always evolving
- New Discoveries happen all the time

Caveats : *Application used*

KEGG pathway from Webgestalt	*P*-value	KEGG pathway from GATHER	*P*-value	KEGG pathway from DAVID	*P*-value
Cytokine-cytokine receptor interaction	8.08×10^{-6}	Huntington's disease	0.60	Jak-STAT signaling pathway	5.24×10^{-2}
Calcium signaling pathway	1.07×10^{-5}	Fatty acid biosynthesis (path 2)	0.61	Cytokine-cytokine receptor interaction	6.19×10^{-2}
Huntington's disease	1.08×10^{-5}	Benzoate degradation via hydroxylation	0.68	Fatty acid elongation in mitochondria	8.55×10^{-2}
Cell adhesion molecules (CAMs)	1.74×10^{-5}	Valine, leucine and isoleucine degradation	0.68	Valine, leucine and isoleucine degradation	0.21
Jak-STAT signaling pathway	5.87×10^{-5}	Vitamin B6 metabolism	0.68	Bile acid biosynthesis	0.23
MAPK signaling pathway	8.59×10^{-5}	d-arginine and d-ornithine metabolism	0.68	Fc epsilon ri signaling pathway	0.24
Long-term potentiation	2.16×10^{-4}	Calcium signaling pathway	0.72	Tight junction	0.25
ECM-receptor interaction	7.29×10^{-4}	Cytokine-cytokine receptor interaction	0.72	Hematopoietic cell lineage	0.28
Regulation of actin cytoskeleton	7.92×10^{-4}	Jak-STAT signaling pathway	0.72	Complement and coagulation cascade	0.32
Leukocyte transendothelial migration	9.30×10^{-4}	Toll-like receptor signaling pathway	0.72	Fatty acid metabolism	0.37

SNPs which showed association with T2D (Po0.003) were included in this study and were mapped backed to regions on the genome, and the predicted candidate genes were used for analysis.The top-10 ranking KEGG pathways per method are shown.

Elbers et. al 2009

Caveats : *Pathway DB used*

Panther pathways	*P*-value	BioCarta pathways from Webgestalt	*P*-value
Ionotropic glutamate receptor pathway	3.77×10^{-4}	Dendritic cells in regulating TH1 and TH2 Development	6.21×10^{-5}
Wnt signaling pathway	7.00×10^{-4}	Cytokines and inflammatory response	1.50×10^{-4}
Inflammation mediated by chemokine and cytokine signaling pathway	1.63×10^{-3}	Th1/Th2 Differentiation	6.22×10^{-4}
Angiogenesis	1.68×10^{-3}	The role of eosinophils in the chemokine network of allergy	6.87×10^{-4}
Oxidative stress response	4.17×10^{-3}	GATA3 participate in activating the Th2 cytokine genes expression	1.16×10^{-3}
Huntington disease	4.69×10^{-3}	Monocyte and its surface molecules	1.46×10^{-3}
Heterotrimeric G-protein signaling pathway-Gq α- and Go α-mediated pathway	5.61×10^{-3}	Selective expression of chemokine receptors during T-cell polarization	1.48×10^{-3}
Heterotrimeric G-protein signaling pathway-Gi α- and Gs α-mediated pathway	6.13×10^{-3}	Antigen dependent B cell activation	1.80×10^{-3}
Axon guidance mediated by Slit/Robo	7.33×10^{-3}	FAS signaling pathway (CD95)	2.13×10^{-3}
O-antigen biosynthesis	7.63×10^{-3}	Acute myocardial infarction	2.19×10^{-3}

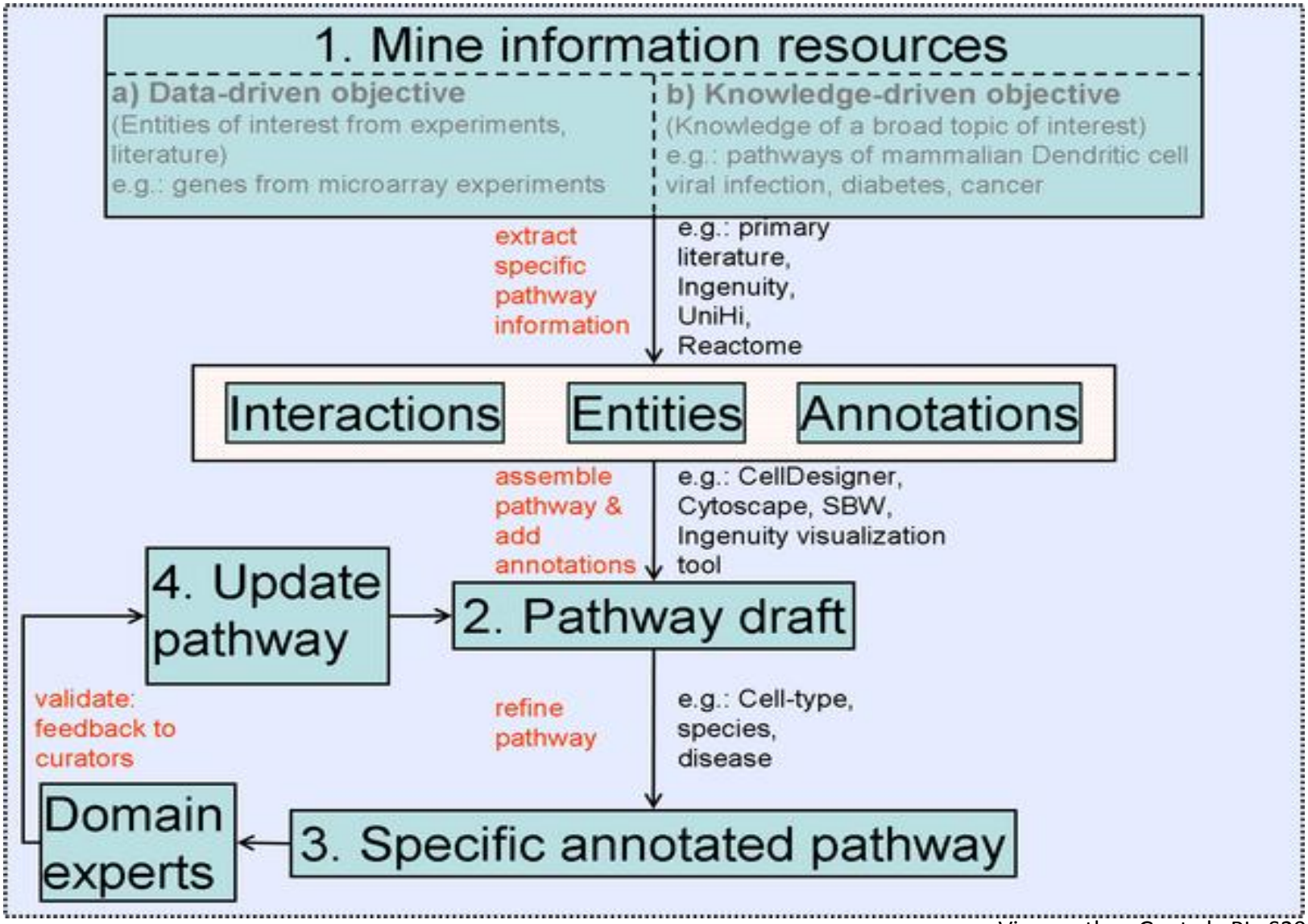
SNPs which showed association with T2D (Po0.003) were included in this study and were mapped backed to regions on the genome and the predicted candidate genes were used for analysis. The highest 10 ranking pathways per method are shown for Webgestalt ‘‘BioCarta’’ and PANTHER.

Elbers et. al 2009

Caveats: *Why*

- Use of different databases
  - Eg. KEGG, BioCarta, Properietary
- Use of different updates
  - Use of different database updates
- Use of different statistical tests
- Use of different definitions/classification
  - Ex. Some use “inflammation” while in others pathway is divided into inflammation related pathways like “Jak-STAT signaling” and “cytokine-cytokine receptor interaction” pathways. While some use hybrid models like GO hybrid (IPA) and others use GO (Metacore)

Biological Pathway Building Process



Viswanathan G, et al. PLoS2008

Stages in Pathway Analysis

- 1st Stage Analysis
  - Data Driven Objective (DDO)
  - Used mainly in determining relationship information of genes or proteins identified in a specific experiment (e.g. microarray study)
  - Focused
- 2nd Stage Analysis
  - Knowledge Driven Objective (KDO)
  - Used mainly in developing a comprehensive pathway knowledge base for a particular domain of interest (e.g. cell type, disease, system)
  - Intergration
- Repeat 1st Stage after generating new leads and hypothesis

Basic Concepts

- Node
  - Symbolizes a list of, for example, genes.
  - This is essentially a one-dimensional representation of the data
- Pathway
  - Linked list of interconnected nodes.
  - This is essentially a two-dimensional representation of the data
- Network
  - A network of cellular functions and regulations involving interconnected pathways
  - This is essentially a multi-dimensional representation of the data

Pathway Creation Algorithms in MetaCore

- **Analyze Network**: Creates a list of possible networks, ranked according to how many objects in the network correspond to the user's list of genes, how many nodes are in the network, how many nodes are in each smaller network.
- **Analyze Networks (Transcription Factors):** For every transcription factor (TF) with direct target(s) in the root list, this algorithm generates a sub-network consisting of all shortest paths to this TF from the closest receptor with direct ligand(s) in the root list.
- **Shortest paths:** Uses Dijkstra's shortest paths algorithm to find the shortest directed paths between the selected objects.
- **Self regulation :**Finds the shortest directed paths containing transcription factors between the selected objects

- **Direct interactions:** Draws direct interactions between selected objects.
  No additional objects are added to the network
- **Auto expand :** Draws sub-networks around the selected objects, stopping the expansion when the sub-networks intersect.
- **Transcription regulation :** Generates sub-networks centered on transcription factors. Sub-networks are ranked by a P-value and interpreted in terms of Gene Ontology.
- **Analyze network (receptors) :**For every receptor with direct ligand(s) in the root list, this algorithm generates a sub-network consisting of all shortest paths from that receptor to the closest TF with direct target(s) in the root list.

An Example to illustrate the Stages in Pathway Analysis

- 1st Stage Analysis
  - Data Driven Objective (DDO)
  - Used mainly in determining relationship information of genes or proteins identified in a specific experiment (e.g. microarray study)
  - Focused topic of interest
- 2nd Stage Analysis
  - Knowledge Driven Objective (KDO)
  - Used mainly in developing a comprehensive pathway knowledge base for a particular domain of interest (e.g. cell type, disease, system)
  - Broad topic of interest
- Repeat 1st Stage after generating new leads and hypothesis

Example

MicroRNA network interactions in REH/MSC cells

miRNA's are 22-nucleotide non-coding RNAs that regulate gene expression through base pairing with target mRNA

Endogenous Regulatory Functions

Invertebrates

developmental timing
neuronal differentiation
cell proliferation, growth control, programmed cell death

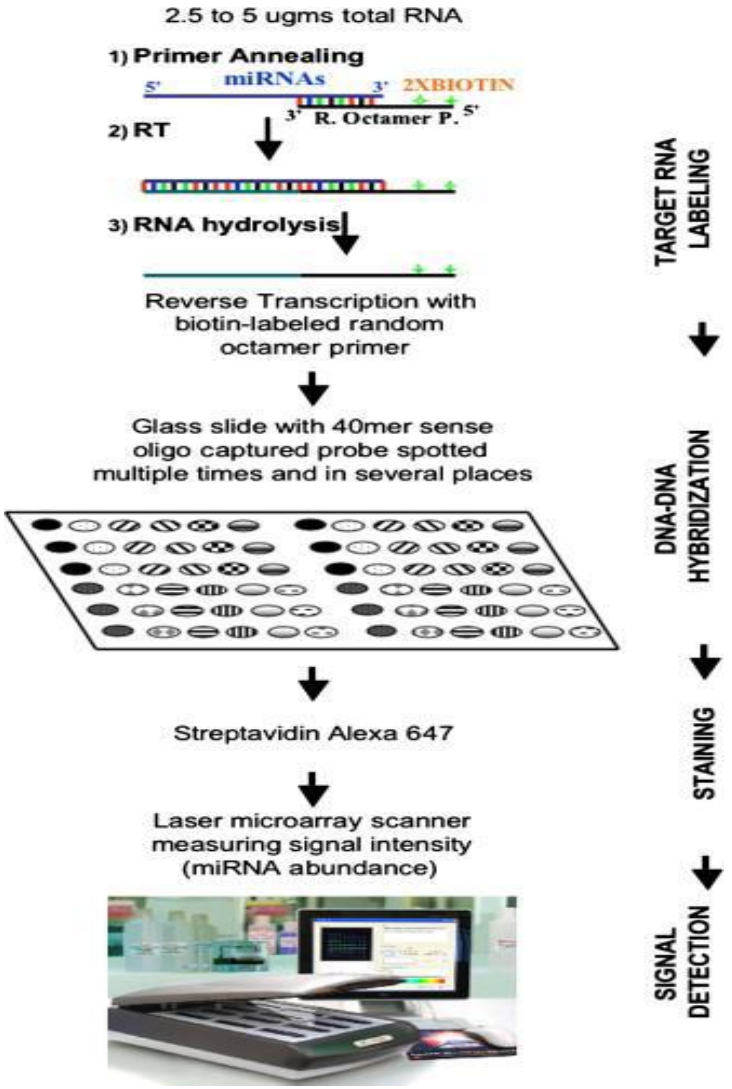
Mammals

embryogenesis
stem cell maintenance
hematopoietic cell differentiation
brain development

Background

- Experiments were performed to analyze the effect of low oxygen conditions and the interaction with the microenvironment in the expression pattern of microRNA's in REH cells.
- In this project we look at the possible interactions between the measured microRNA's with other molecules related to the pathogenesis of lymphocytic leukemia.
- In this initial stage we plan to put this complex system of microRNA interactions in context with its surrounding interactions.
- It was discerned from the current analysis that the anti-apoptotic action of microRNA 21 may be due to its interaction with Bcl-2 and MCL-1 in MSC cells. However, this needs to be further explored.
- In this initial report we looked at all the tested microRNAs in the context of its associated biological networks.

Experimental Details



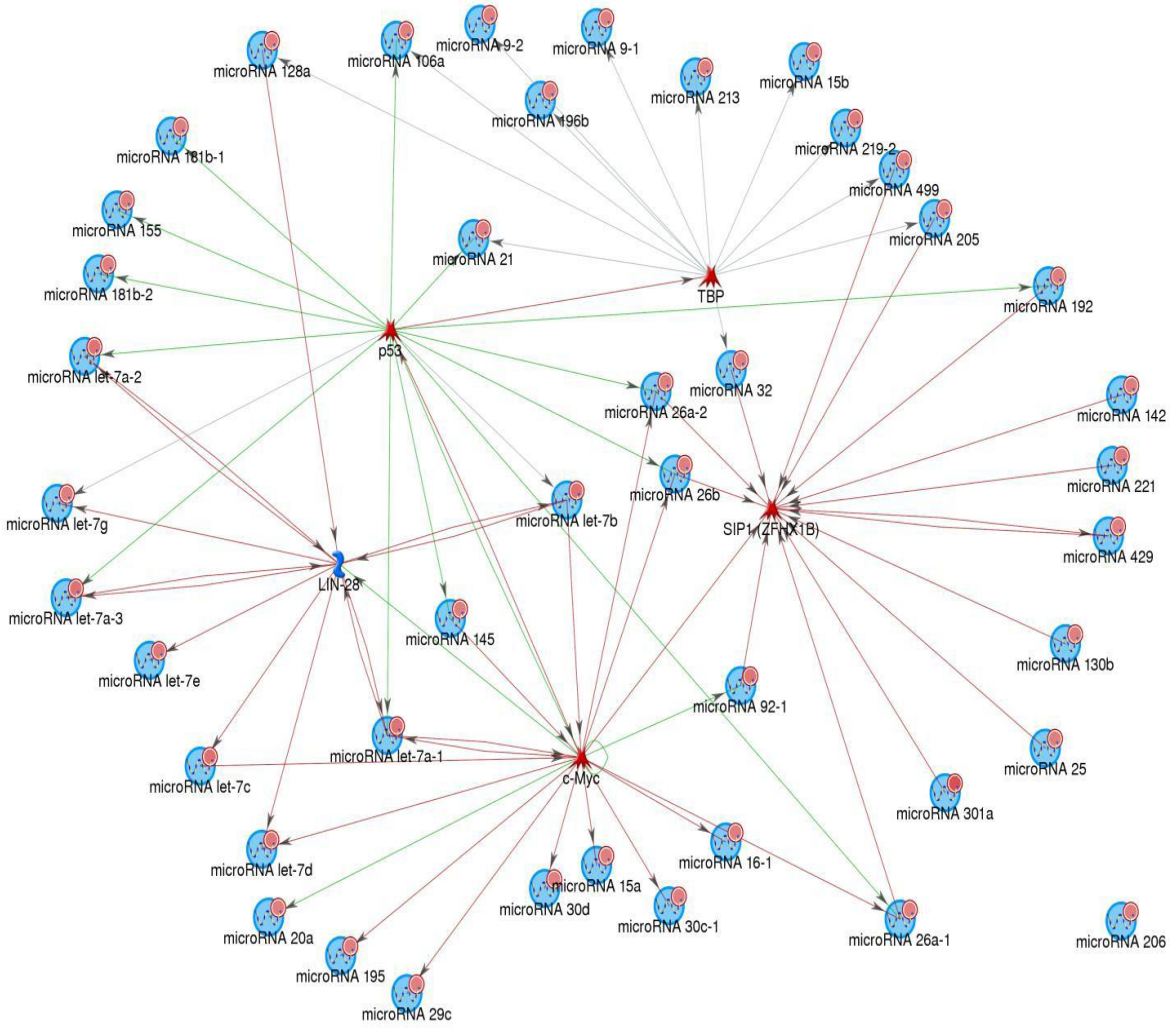
- Briefly, REH cells were cultured alone or co-cultured with N.BM MSC (or H-Tert immortalized MCS) for 24 h and 48 h under different pO2 conditions. At the end of 24 h and 48 h REH cells and MCS cells were sorted by Flow Cytometry and each cell population was lysed with TRIZOL to extract total RNA separately.
- MicroRNA assay details:
- The biotin-labeled cDNA targets are prepared by a simple reverse transcription into first strand cDNA. Total RNA is primed for reverse transcription by a random Octomer conjugated with two biotins and a 5′ poly (A) tail. This procedure results in an equal copy number of biotin–cDNA target to the templates of miRNA (see figure on left).
- Two 40 mer oligo probes, one for the mature miRNA and the other for precursor oligo, were designed from the sense strand of both arms of the hairpin structure of the microRNA precursor sequence collected from the Sanger Database. The oligo probes were modified at the 5′ end with Amine-C6 linker and ordered from Integrated DNA technology (IDT) (Coralville, IA, USA) at 50 or 100 µM stock concentration in H2O (see figure on left).

- REH: pre-B Acute lymphoblastic leukemia cell line
- MSC: Mesenchymal stroma cell line

1: 1st Stage Analysis

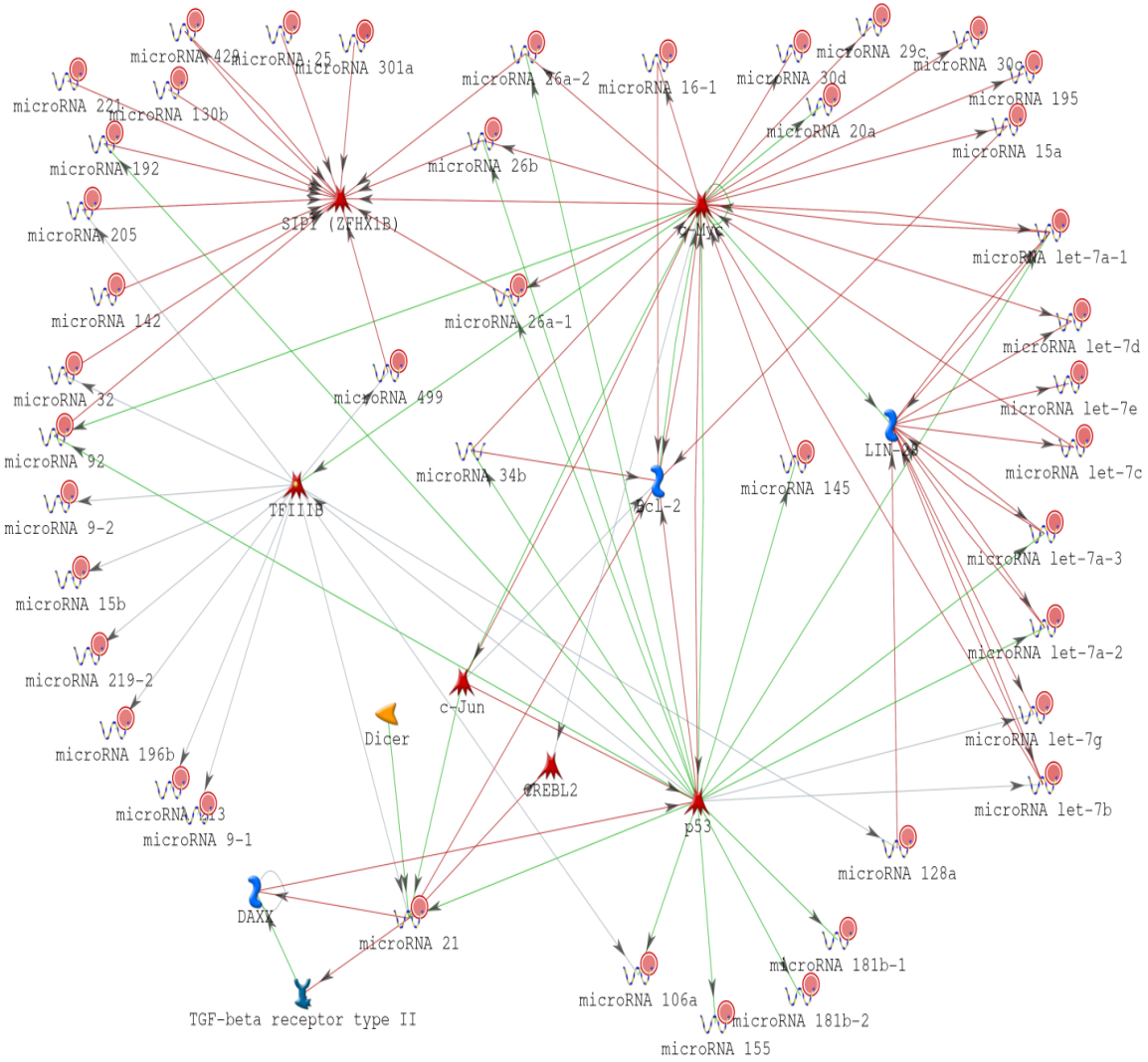
NETWORK RELATIONSHIPS BASED ON THE EXPERIMENTAL INPUT

microRNA interaction partners



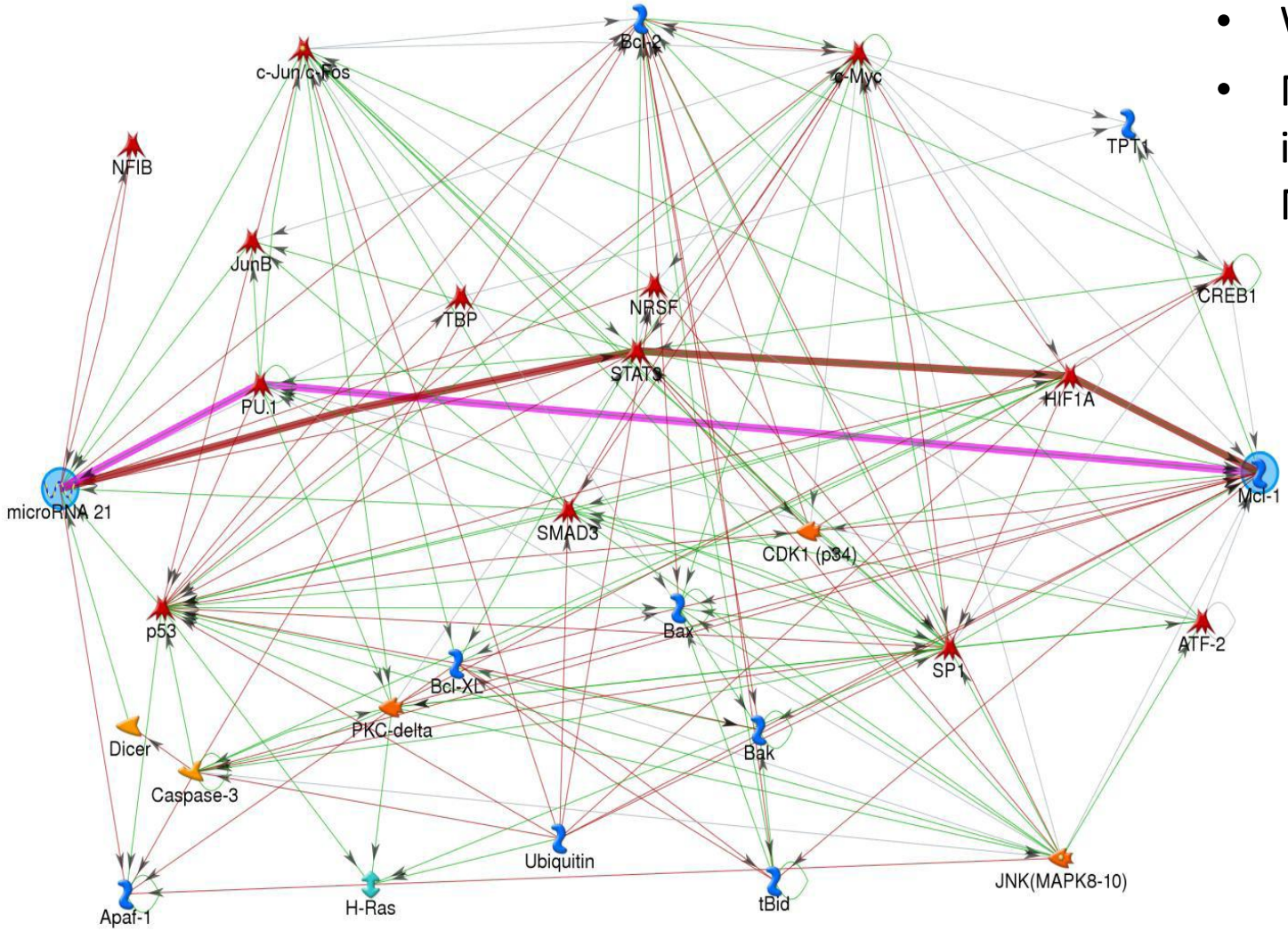
- Hub interactions

microRNA 21 and its relationship to Bcl-2



- Why BCL-2
- Inhibits BCL-2 activity

microRNA 21, Mcl-1, interactions



- Why Mcl-1
- Mir21 indirectly interacts with Mcl-1

2: 1st Stage Analysis

GO ENRICHMENT FOR THE MEASURED MICRORNA

GO enrichment for the measured microRNA

#	Process	%	p-Value
1	release of cytochrome c from mitochondria	30.77	6.35E-09
2	general transcription from RNA polymerase II promoter	23.08	7.7E-09
3	cell death	61.54	2.47E-08
4	death	61.54	2.67E-08
5	regulation of gene expression	92.31	3.29E-08
6	apoptotic mitochondrial changes	30.77	3.43E-08
7	multi-organism process	61.54	1.09E-07
8	regulation of macromolecule metabolic process	92.31	1.78E-07
9	regulation of metabolic process	92.31	2.81E-07
10	regulation of macromolecule biosynthetic process	84.62	7.95E-07
11	regulation of cellular biosynthetic process	84.62	1.03E-06
12	regulation of biosynthetic process	84.62	1.06E-06

3: 1st Stage Analysis

DISEASE REPRESENTATION FOR THE MEASURED MICRORNA

Network-disease associations

#	Disease	%	p-Value
1	Lymphoma, Intermediate-Grade	23.81	5.59E-09
2	Lymphoma, Diffuse	26.19	9.94E-09
3	Lymphoma, Small-Cell	23.81	5.7E-08
4	Lymphoma, Mantle-Cell	21.43	7.72E-08
5	Lymphoma, Small Cleaved-Cell, Diffuse	21.43	7.72E-08
6	Decompression Sickness	7.14	9.71E-08
7	Barotrauma	7.14	9.71E-08
8	Vaccinia	14.29	2.05E-07
9	Poxviridae Infections	14.29	4.19E-07
10	Leukemia, B-Cell	14.29	8.59E-07
11	Lymphoma, B-Cell	26.19	9.68E-07
12	Lymphoma, High-Grade	14.29	3.27E-06

4: 1st Stage Analysis

POSSIBLE THERAPEUTIC TARGETS AND OTHER INTERACTIONS FOR THE MICRORNA NETWORK

- No direct activators/ inhibitors for miR21
- However, indirect activators /inhibitors for miR21

Questions…

- What new things have we learned?
- What type of things should we expect to learn in general?
- What new experiments do these suggest?
- How can this new knowledge be exploited?

Stages in Pathway Analysis

- 1st Stage Analysis
  - Data Driven Objective (DDO)
  - Used mainly in determining relationship information of genes or proteins identified in a specific experiment (e.g. microarray study)
  - Focused topic of interest
- 2nd Stage Analysis
  - Knowledge Driven Objective (KDO)
  - Used mainly in developing a comprehensive pathway knowledge base for a particular domain of interest (e.g. cell type, disease, system)
  - Broad topic of interest
- Repeat 1st Stage after generating new leads and hypothesis
  - On and on we go … ☺

Demo

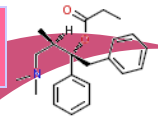
- Search and Explore
- Building pathways
- Analysis
- Canonical Pathways

A caveat

- Not every gene belongs to a pathway in the database…
- Pathways generated are a statistical probability rather than a biological certainty
- Context…context…context… it matters a lot in pathway analysis
- Findings need to be proved experimentally

Putting it together

Compound



Chemical properties, QSAR, etc…

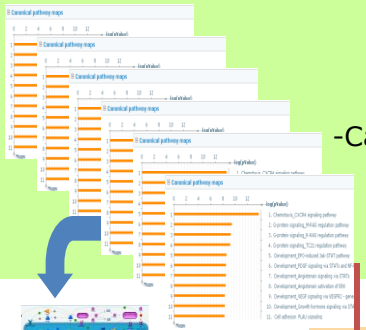
Predicted Metabolites

DB Similarity Search

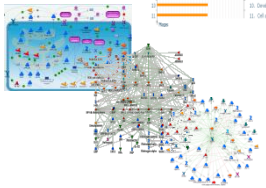
Similar DB Compounds

Lists of Targets (genes/proteins)
for the input molecule and for metabolites
Based on the **known targets for the Similar Compounds**

Functional Enrichment Analysis Based on the Predicted Targets:

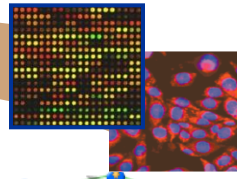


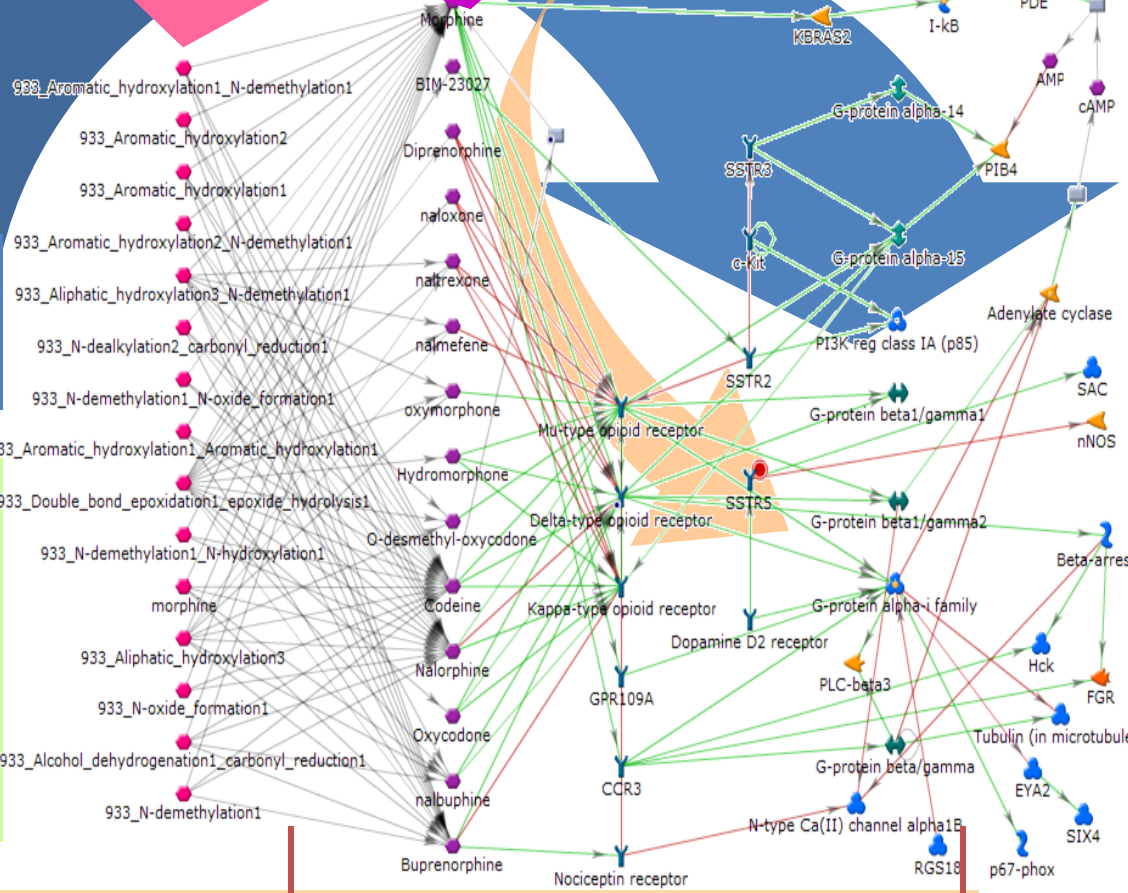
-Drug target networks
-Toxicity networks
-Canonical Pathway Maps
-Process networks
-GO processes
-Disease Networks



Network Analysis
Interconnecting compound/metabolites with their predicted targets and other associated network objects.

Experimental Data (Optional)
Import and overlay any experimental data





Integrative Systems-Level Summary of Predicted Primary and Secondary Effects of a query Compound