

GenMAPP and MAPPFinder:
Tools for Viewing and Analyzing
Microarray Data on Biological Pathways

Kam D. Dahlquist, Ph.D.

Gladstone Institute of Cardiovascular Disease
University of California, San Francisco

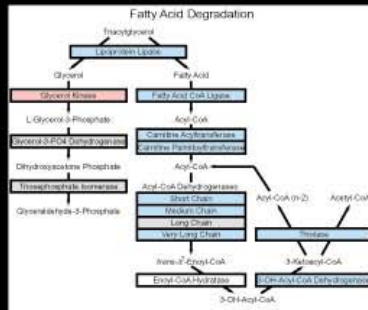
Advanced Topics in Microarray Analysis
January 22, 2003

Outline



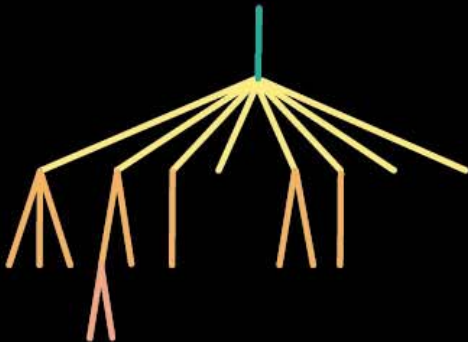
Need for GenMAPP

DNA microarray studies of a mouse model of dilated cardiomyopathy



GenMAPP

View data in the context of biological pathways



MAPPFinder

Combine GenMAPP and Gene Ontology to analyze global trends in the data

Creation and Analysis of the Ro1 Model of Dilated Cardiomyopathy



Engineered G protein-coupled receptor

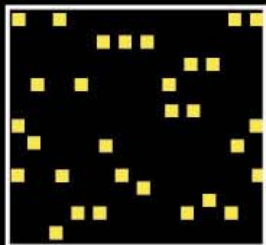
G_i



Overexpression in the hearts of mice using an inducible expression system

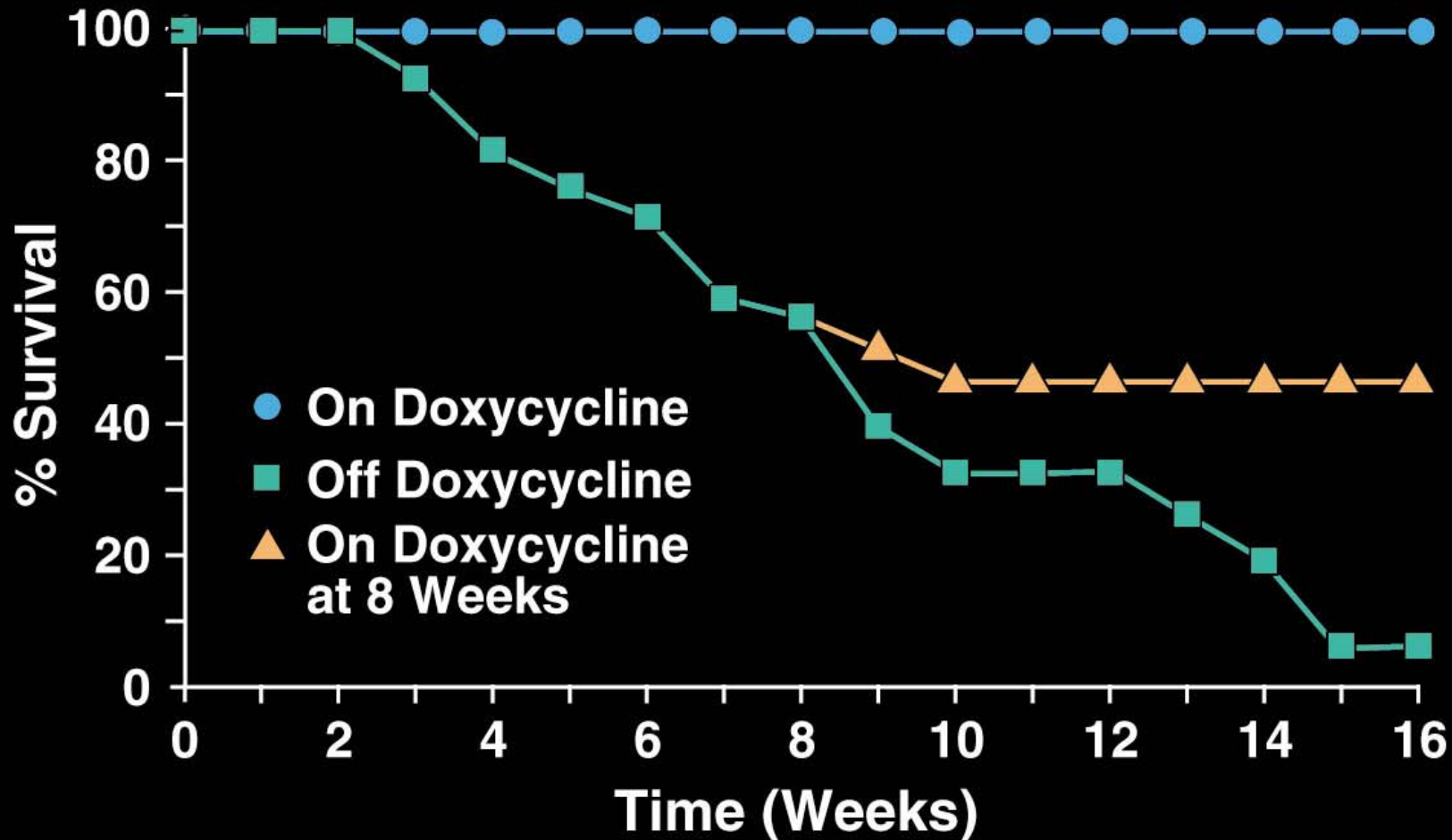


Hyperactive G_i signaling leads to dilated cardiomyopathy

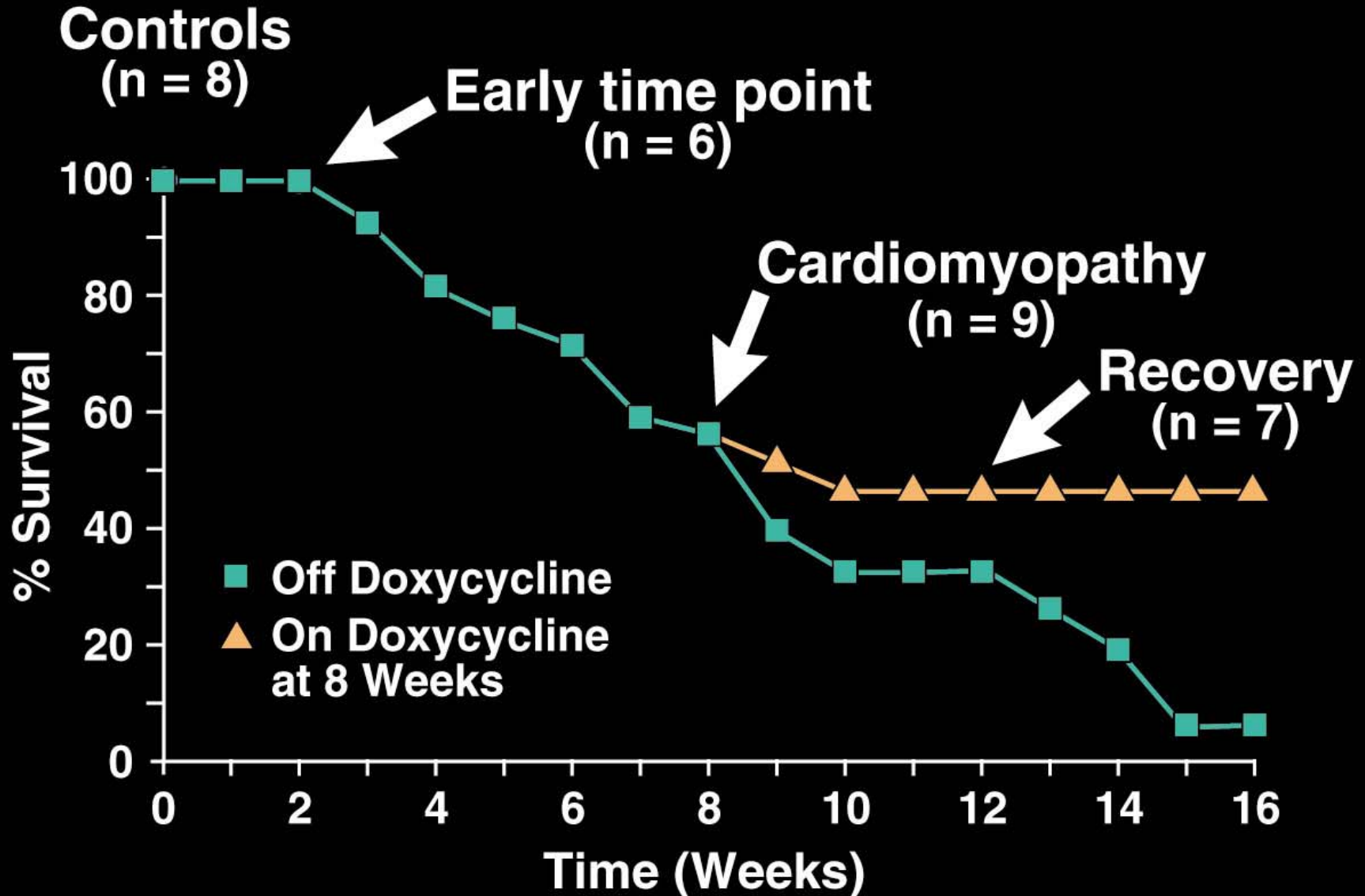


Measure gene expression on arrays

Repression of Receptor Expression Rescues Lethal Cardiomyopathy



Timing of Microarray Studies

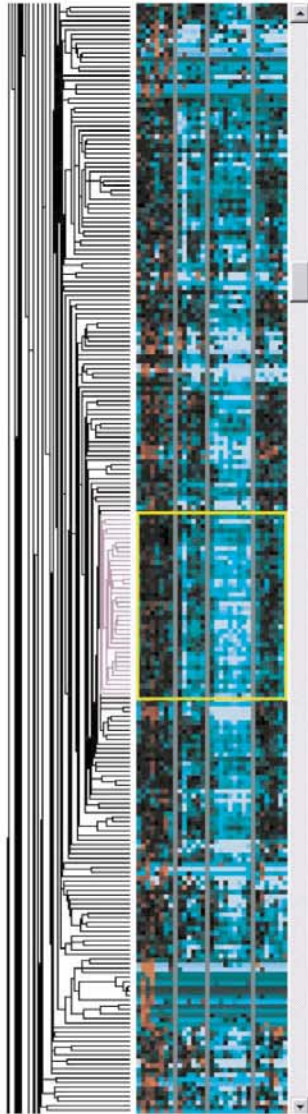


The Challenge:
Data for Thousands of Genes
and Multiple Replicates
at Several Time Points

Meeting the Challenge: Approaches to Microarray Data Analysis

- 1. Statistical methods**
- 2. Compare with known markers**
- 3. Pattern finding algorithms**
- 4. Map results on biological pathways**

Hierarchical Clustering Finds Patterns in Gene Expression



Genes with similar patterns of expression cluster next to each other in the hierarchical tree.

Clusters of genes may participate in the same biological pathway and are thought to be co-regulated by the same transcription factor.

 Fold Increase

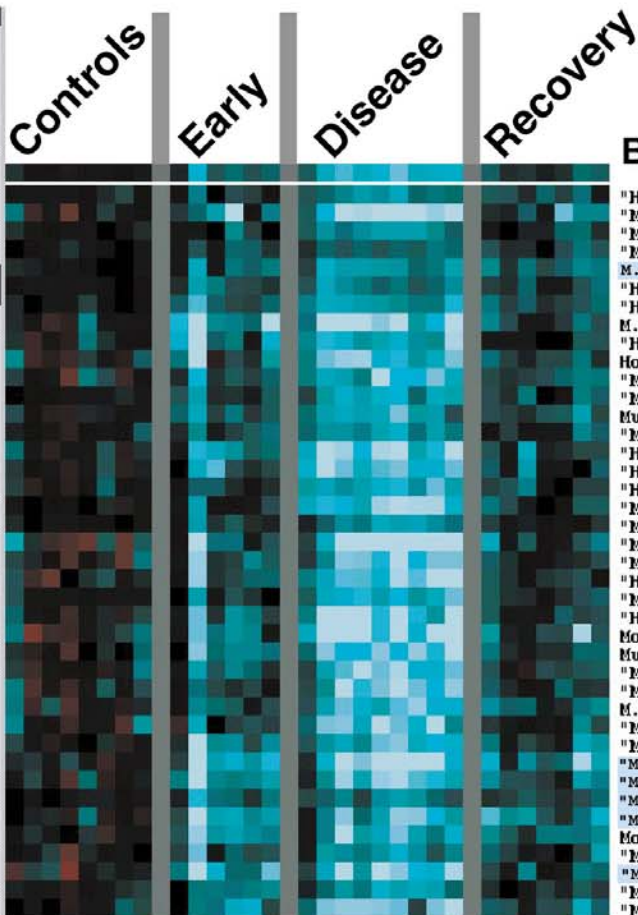
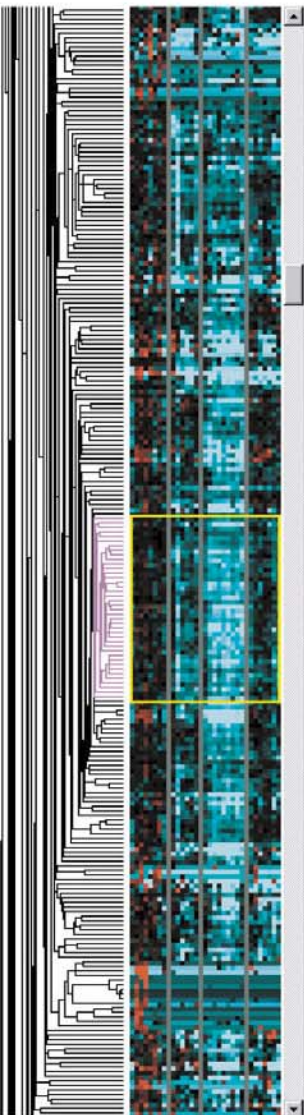
 Fold Decrease

 No Change

← Each row = one gene

↑ Each column = one mouse

Signal Dependent Genes - Downregulated



BLAST Definitions

"Homologous to sn 007021: PRE-MRNA SPLICING FACTOR SF2, P32 SUBUNIT PRECURSOR (GCIQ-R PROTEIN)"

"Mus musculus proteasome activator PA28 alpha subunit mRNA, complete cds"

"Mus musculus cdc37 homolog mRNA, complete cds"

"Mus musculus ornithine decarboxylase antisense gene, complete cds"

M.musculus mRNA for carnitine acetyltransferase

"Homologous to sn 000779: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.38)"

"Homologous to sn P11507: CALCIUM-TRANSPORTING ATPASE ENDOPLASMIC RETICULUM TYPE (EC 3.6.1.38)"

M.musculus ENO3 mRNA for enolase beta subunit

"Homologous to sn P47858: 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)"

Homologous to sn P23327: SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN PRECURSOR

"Mouse AE3 mRNA, complete cds"

"M.musculus glucose transporter 2 mRNA, complete cds"

Mus musculus aspartate aminotransferase gene 5'-flank and exon 1

"Mus musculus thioredoxin-dependent peroxide reductase (txr) mRNA, complete cds"

"Homologous to sn P47858: 6-PHOSPHOFRUCTOKINASE, MUSCLE TYPE (EC 2.7.1.11) (PHOSPHOFRUCTOKINASE)"

"Homologous to sn P11508: CALCIUM-TRANSPORTING ATPASE SARCOPLASMIC RETICULUM TYPE (EC 3.6.1.38)"

"Homologous to sn P35434: ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR (EC 3.6.1.34)."

"Mus musculus F1FOATP synthase complex E subunit (Atp5k) gene, complete cds"

"Mus musculus NAD(H)-specific isocitrate dehydrogenase gamma subunit precursor, mRNA, complete cds"

"M.musculus gene for dodecenoyl-CoA delta-isomerase, exons 1 and 2"

"Mus musculus cytochrome c oxidase subunit VIII-H precursor (COX8H) mRNA, complete cds"

"Homologous to sn P35745: ACYLPHOSPHATASE, MUSCLE TYPE ISOZYME (EC 3.6.1.7) (ACYLPHOSPHATE PHOSPHATASE)"

"Mus musculus CD-1 cardiac troponin I mRNA, complete cds"

"Homologous to sn P00566: CREATINE KINASE, M CHAIN (EC 2.7.3.2) (NU-2 PROTEIN)."

Mouse mRNA for protein with homology to transition protein 2 (TP2)

Mus musculus Selenium-binding liver protein mRNA

"Mus musculus (clone MAR1) aldose reductase mRNA, complete cds"

"Mus musculus vascular endothelial growth factor B 186 (VEGF-B) precursor, mRNA, complete cds"

M.musculus mRNA for NADP transhydrogenase

"Mus musculus aldehyde dehydrogenase (ALDH2) mRNA, nuclear gene encoding mitochondrial protein"

"Mouse cytosolic epoxide hydrolase mRNA, complete cds"

"Mus musculus 129SV carnitine palmitoyltransferase II mRNA, complete cds"

"Mus musculus medium-chain acyl-CoA dehydrogenase mRNA, complete cds"

"Mus musculus long-chain acyl-CoA dehydrogenase mRNA, complete cds"

"Mus musculus very-long chain acyl-CoA dehydrogenase, partial cds"

Mouse muscle creatine kinase mRNA (EC 2.7.3.2)

"Mus musculus isocitrate dehydrogenase mRNA, complete cds"

"Mus musculus long chain fatty acyl CoA synthetase mRNA, complete cds"

"Mus musculus sterol carrier protein-2 (SCP-2) gene, complete cds"

"Mouse alpha-tubulin isotype M-alpha-4 mRNA, complete cds"

Components of Fatty Acid Degradation

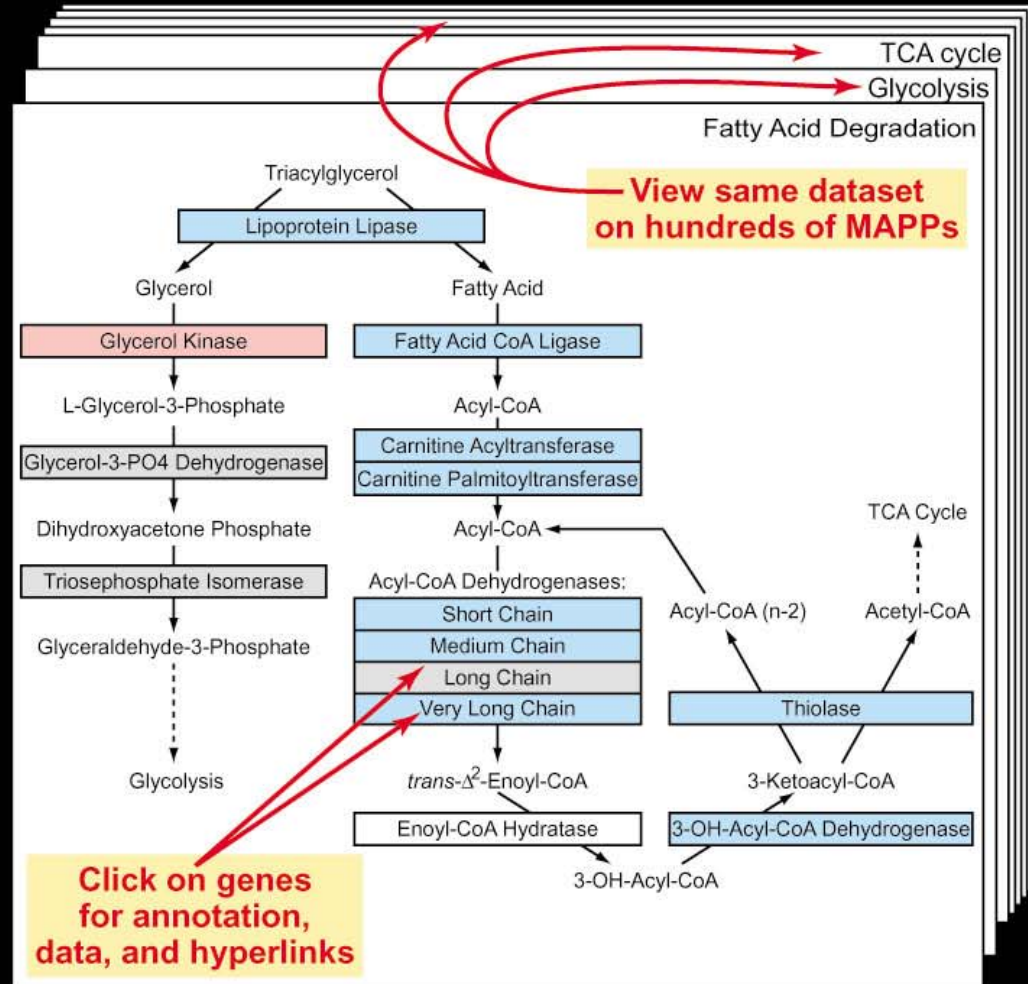
Meeting the Challenge: Approaches to Microarray Data Analysis

1. Statistical methods
2. Compare with known markers
3. Pattern finding algorithms
4. Map results on biological pathways

GenMAPP

Gene MicroArray Pathway Profiler

- ◆ MAPPs represent biological pathways and other functional groupings of genes
- ◆ Graphics tools for drawing MAPPs
- ◆ Underlying gene database
- ◆ Import expression data and set criteria to automatically color the MAPPs according to the data

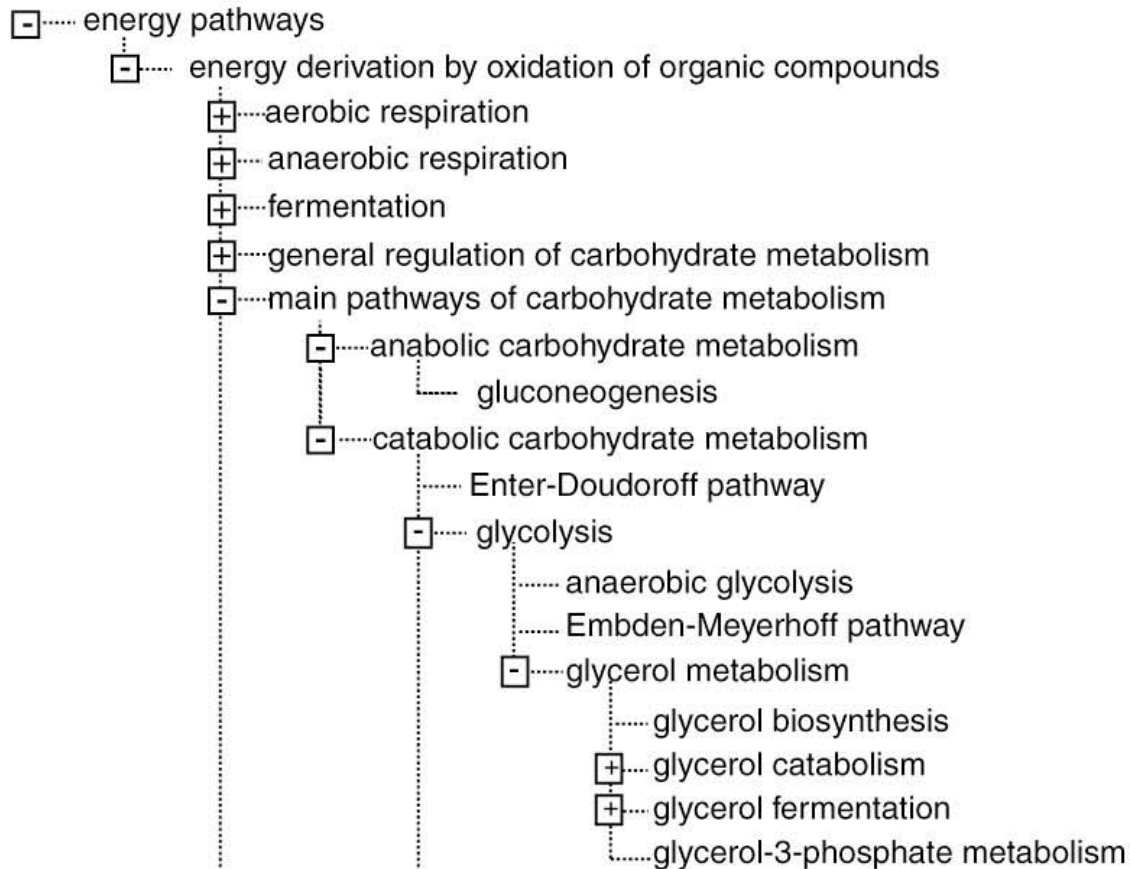


www.GenMAPP.org

- ◆ GenMAPP and accessory programs are distributed free-of-charge to all researchers.
- ◆ MAPPs contributed by the biological community are posted on the web site.
Currently, there are ~50 pathway MAPPs each for mouse, rat and human, and a few for yeast
- ◆ Contact information for the MAPP author is included on each MAPP, promoting the exchange of information.

GenMAPP demonstration

Gene Ontology (GO) Provides a Framework for Biology



Hierarchy of biological process, cellular component, and molecular function terms

Genes assigned to terms by model organism databases

MAPPFinder Rapidly Identifies Global Trends in Expression Data

Hundreds of genes
meeting a user-defined
expression criterion

GO process, component,
function terms and
local MAPPs

```
graph TD; A[Hundreds of genes meeting a user-defined expression criterion] --> C[Calculate % genes changed per term or MAPP]; B[GO process, component, function terms and local MAPPs] --> C; C --> D[List of GO terms and MAPPs ranked by z score];
```

Calculate
% genes
changed
per term
or MAPP

List of GO terms and MAPPs ranked by z score

Processes Downregulated with Ro1-induced Cardiomyopathy

GO term	% Changed		Z score
fatty acid metabolism	41	(10/17)	7.2
peroxisome organization	67	(4/6)	4.9
mitochondrion	22	(24/109)	4.8
muscle contraction regulation	57	(4/7)	4.4
translation elongation factor	40	(4/10)	3.4
oxidoreductase	17	(20/120)	3.0
connexon channel	33	(4/12)	2.9

Processes Upregulated with Ro1-induced Cardiomyopathy

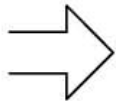
GO term	% Changed		Z score
serine protease inhibitor	47	(14/30)	4.7
N-linked glycosylation	100	(4/4)	4.7
Golgi membrane	75	(6/8)	4.7
cell-matrix adhesion	45	(9/20)	3.7
intracellular protein transport	28	(28/99)	3.6
integrin-mediated signaling	43	(9/21)	3.5
small GTPase-mediated signaling	33	(17/52)	3.4
complement activation	46	(6/13)	3.1
endoplasmic reticulum	28	(21/76)	3.0
cytoskeleton organization	26	(24/92)	2.8

MAPPFinder demonstration

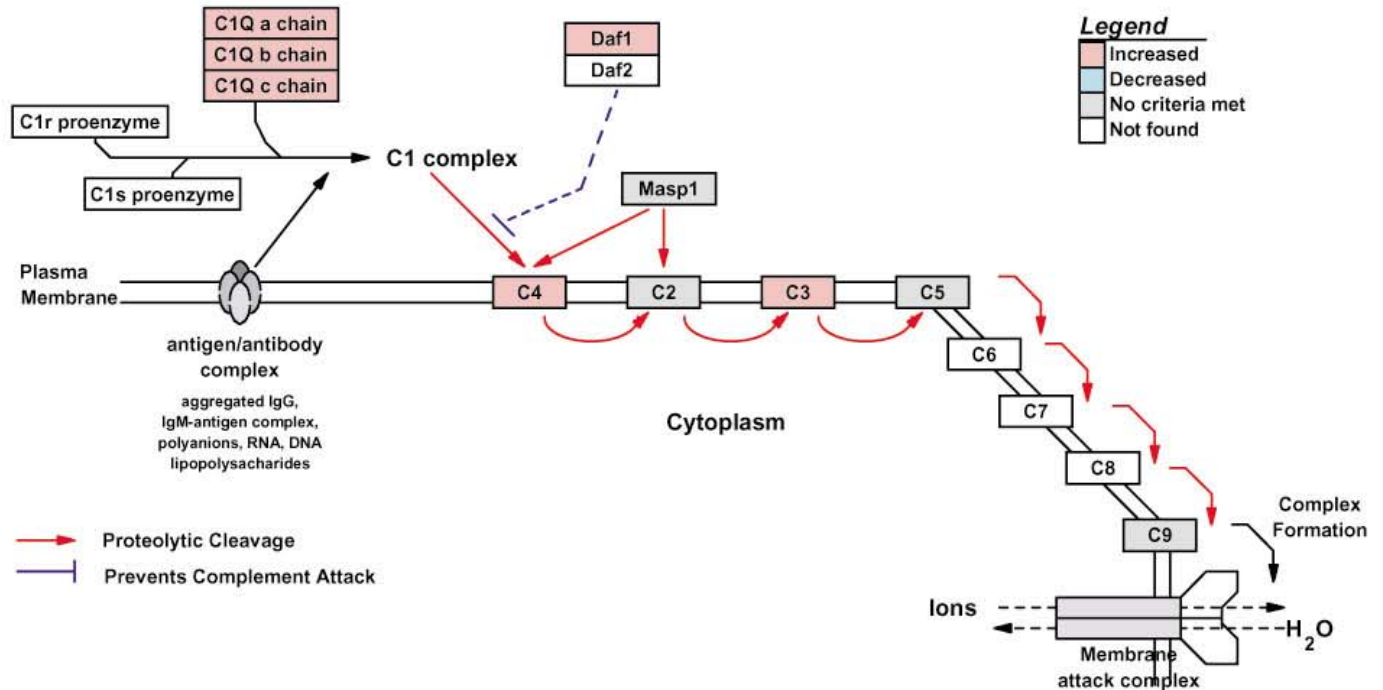
The Gene Ontology Classifications are a Great Starting Point for Building Detailed Pathway MAPPs

Complement Activation, Classical GO List

C1qa
C1qb
C1qc
C2
C3
C4
C4bp
C9
Cr2
Daf1
Daf2
Hc
Masp1



Complement Activation, Classical Pathway



Other Uses for GenMAPP & MAPPFinder

MAPPs

Any biologically relevant grouping of genes

e.g., clusters of co-expressed genes

genes trapped in ES cells by BayGenomics

Data

Any type of large-scale data

e.g., protein levels

MAPPFinder

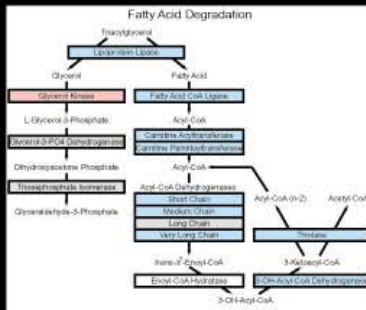
Analyze results from clustering algorithms

GenMAPP is a Powerful Tool for Analysis



New Insights

Energy pathways down-regulated in a mouse model of dilated cardiomyopathy

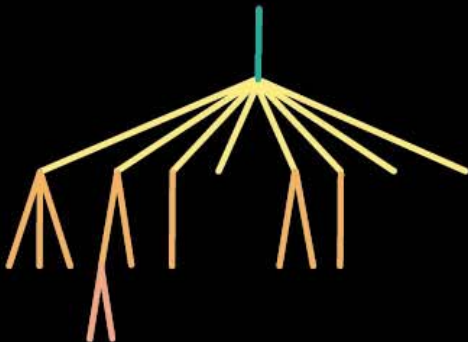


GenMAPP

View and share pathways and data

MAPPFinder

Combine with other types of analysis





<http://www.GenMAPP.org>

Bruce Conklin

Scott Doniger

Steven Lawlor

Nathan Salomonis

Karen Vranizan

**NIH - NHLBI
Programs for
Genomics Applications**

<http://baygenomics.ucsf.edu>

GenMAPP 1.0: Design & Implementation

- ◆ **MAPP files store gene IDs and vector coordinates for all graphical objects**
- ◆ **Separate Expression Dataset files store data and color-coding instructions**
- ◆ **Gene database provides IDs, annotation, and hyperlinks to GenBank, SWISS-PROT, and SGD for human, mouse, rat, and yeast**
- ◆ **Stand-alone, prototype program implemented in Visual Basic, accessory files are Access databases**

GenMAPP Version 2.0 (in Development)

◆ Expansion and integration of gene database

SWISS-PROT

GenBank

LocusLink

RefSeq

Unigene

InterPro

Gene Ontology

Genome-specific databases

H. sapiens

M. musculus

R. norvegicus

S. cerevisiae

C. elegans

D. melanogaster

D. rerio

A. thaliana

E. coli

◆ Users choose which catalogs and species to download

◆ Add custom gene catalogs and relationships

Resources and URLs

Download GenMAPP, MAPPFinder, and MAPPs

<http://www.genmapp.org/download.asp>

Interactive tutorial

<http://www.genmapp.org/tutorial.html>

Help files

<http://www.genmapp.org/GenMAPPHelp/HelpFiles/GenMAPP.htm>

http://www.genmapp.org/MAPPFinder_help.html

Help e-mail address

genmapp@gladstone.ucsf.edu

Publications

Dahlquist *et al.* (2002) *Nature Genetics* 31: 19-20

Doniger *et al.* (2003) *Genome Biology* 4: R7.1-R7.12

<http://www.genmapp.org/Publications.html>