Analysis of Microarray Data

Keith Baggerly and Kevin Coombes
Section of Bioinformatics
Department of Biostatistics and Applied Mathematics
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org
kcoombes@mdanderson.org

21 October 2004
Lecture 15: Regulatory networks

- PathArt
- Ingenuity
INTRODUCTION TO MICROARRAYS

PathArt

PathArt™ is a curated database of biomolecular interactions with tools for searching, analysis and visualization of data for use by microarray researchers and identification of potential drug targets. Pathway diagrams in PathArt™ are dynamically generated from data in the database. PathArt™ is accessible via any Java enabled browser and provides enterprise wide access to data stored in an Oracle database.

Features

* Coverage of about 800 regulatory and signaling pathways across species.
* Browse pathways by organism, disease and other classifications.
* Coverage of protein-protein interaction.
* Information on knockout and mutagenesis studies.
* Search for pathways by specific genes.
* Coverage of 17 high priority diseases and disease responsive genes.
* Generate customized reports on genes and interactions of interest.
* Allows use of microarray expression data to search relevant pathways based on expression level. Allows use of Agilent and Affymetrix data.
* Comprehensive information on all participating biomolecules, relevant scientific journals, and clinical data.
PathArt as pathway browser
PTEN mediated pathway
PTEN pathway diagram
PathArt symbol legend
Evidence supports inclusion of genes in diagrams
INTRODUCTION TO MICROARRAYS

Links go live to NCBI

PTEN protects p53 from Mdm2 and sensitizes cancer cells to chemotherapy.

Mayer LD, Dixon JE, Durden DL, Tenks NK, Donner DB.

Department of Microbiology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

The PTEN tumor suppressor gene inhibits phosphatidylinositol 3-kinase (PI3K)/Akt signaling that promotes translocation of Mdm2 into the nucleus. When repressed by the tumor protein, Mdm2 is degraded. The ability of PTEN to inhibit the nuclear entry of Mdm2 increases the cellular control and transcription of the p53 tumor suppressor protein. Restoration of PTEN into U87MG/PTEN-null glioblastoma cells increases p53 activity and expression of p53 target genes and induces cell cycle arrest. U87MG/PTEN glioblastoma cells are more sensitive than U87MG/PTEN-null cells to death induced by etoposide, a chemotherapeutic agent that induces DNA damage. Furthermore, tumor suppressor proteins have been suggested to act individually to suppress cancers. Our results establish a direct connection between the activities of two major tumor suppressors and show that they act together to stop and to suppress malignancies. PTEN protects p53 from survival signals, permitting p53 to function as a guardian of the genome. By virtue of its capacity to protect p53, PTEN can sensitize tumor cells to chemotherapy that relies on p53 activity. p53 induces PTEN gene expression, and here it is shown that PTEN protects p53, indicating that a positive feedback loop may amplify the cellular response to stress, damage, and cancer.

PMID: 11729185 [PubMed - indexed for MEDLINE]
**PathArt Menu to load microarray data**

![Image of PathArt Menu]

**INTRODUCTION TO MICROARRAYS**

PathArt Menu to load microarray data

---

**© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly**
PathArt with array info loaded
Microarray-selected pathways
Pathways are organized by disease
The FGF pathway with array data
Weaknesses of PathArt

1. Dreadful interface
   (a) File selection box often starts blank (TAB display icons).
   (b) No easy way to return to previous screens. (Often have to reload microarray data to get back to first page.)

2. Diagrams are rigid
   (a) Can shrink, but not move things around
   (b) Typical diagram is wider than the screen, but not very deep
   (c) Often impossible to see all the information at once

3. Cannot sort pathways by number or proportion of hits

4. No statistical way to decide if a pathway is “enriched” for differentially expressed genes
INTRODUCTION TO MICROARRAYS

Ingenuity

Helping Life Science Researchers Understand Biological Mechanisms

Ingenuity delivers systems biology expertise to biologists and bioinformaticians through pathway analysis software, genome-scale compatible network databases, and knowledge management services and infrastructure, resulting in increased R&D productivity and faster drugs to market.

Biologists
Ingenuity Pathways Analysis, a new way for scientists to compare and understand biological states. New scientists can concurrently analyze multiple datasets across different experimentation platforms and...

Bioinformaticians
Ingenuity Pathways Knowledge Base is the world's largest curated database of biological networks created from millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, drugs, and diseases.

FREE TRIAL
Ingenuity Pathways Analysis

ROI of Pathways Software
New White Paper - FREE
INTRODUCTION TO MICROARRAYS


Pathways Analysis

Ingenuity Pathways Analysis - A new way for scientists to compare and understand biological states. Now scientists can concurrently analyze multiple datasets across different experimentation platforms and identify key functions and pathways that distinguish biological states. This innovative technology is especially useful for dose response assessment, patient stratification for clinical trials, time course experiments, and simultaneous analysis of gene lists and protein lists. The web-delivered application makes use of the Ingenuity Pathways Knowledge Base, the world’s largest curated database consisting of millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, drugs, and diseases. These new features help life science researchers in early discovery to clinical trials gain novel insights in a shorter period of time than previously possible.

Biological researchers have traditionally examined functional genetic information to elucidate fundamental cellular processes and unravel the etiology of human disease. In today’s post-genome

© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly
INTRODUCTION TO MICROARRAYS

Ingenuity three week free trial

The Ingenuity Pathways Analysis application is focused on providing a powerful computational solution to this problem. A therapeutic researcher inputs into the Pathways Analysis application a set of genes and/or a set of proteins, (e.g. a set of differentially regulated genes on the mRNA transcript level). Utilizing this data in conjunction with Ingenuity’s vast, computable, biological knowledge base, the software dynamically computes a set of relevant pathways and presents this information to the scientist. The researcher analyzes these novel gene and protein networks, and associated biological research, to verify that the networks are differentially modulated in disease and may serve as a target-rich environment for therapeutic manipulation.
INTRODUCTION TO MICROARRAYS

Ingenuity start page

© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly

GS01 0163: ANALYSIS OF MICROARRAY DATA
Creating a new project
Ingenuity project page
Uploading microarray data
Uploading microarray data
Uploading microarray data
Finding the networks (analyzing)
Finding the networks (analyzing)
Waiting for the analysis to complete
INTRODUCTION TO MICROARRAYS

Ingenuity networks

© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly

GS01 0163: ANALYSIS OF MICROARRAY DATA
The first network: MYC
INGENUITY FIGURE LEGEND

Network Explorer & Canonical Pathways Node Shapes

Network Explorer Node Shapes

Canonical Pathways Node Shapes

Network Explorer & Canonical Pathways Edges

Network Explorer Edge Types

Canonical Pathways Edge Types

Edge Labels

A. Activation
B. Binding
C. Expression
D. Inhibition
E. Gene
F. Gene Expression
G. Gene Regulation
H. Gene Silencing
I. Gene Silencing
J. Inhibition
K. Gene Silencing
L. Translation
M. RNA Stability
N. Protein Stability
O. Post-translational Modification
P. Transcription
Q. Translation
R. Protein-Protein Interaction
S. Protein-DNA Interaction
T. Protein-DNA Interaction
U. Protein-DNA Interaction
V. Protein-DNA Interaction
W. Protein-DNA Interaction
X. Protein-DNA Interaction
Y. Protein-DNA Interaction
Z. Protein-DNA Interaction

Legend

This legend provides an overview of the main features of Network Explorer and Canonical Pathways, including node shapes and colors as well as edge labels and types. For details, see Network Explorer and Canonical Pathways.
Evidence for interactions
Second network
Third network
Fourth network
Fifth network
Sixth network
INTRODUCTION TO MICROARRAYS

CAV1 node details

© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly

GS01 0163: ANALYSIS OF MICROARRAY DATA
Links to the literature are harder to find.
Links to literature are live

Caveolin-1 null mice are viable but show evidence of hyperproliferative and vascular abnormalities.


Department of Molecular Pharmacology and The Albert Einstein Cancer Center, The Albert Einstein College of Medicine, Bronx, New York 10467, USA.

Caveolin-1 is the principal structural protein of caveolae membranes in fibroblasts and endothelia. Recently, we have shown that the human CAV-1 gene is localized to a suspected tumor suppressor locus, and mutations in Cav-1 have been implicated in human cancer. Here, we created a caveolin-1 null (Cav-1−/−) mouse model, using standard homologous recombination techniques, to assess the role of caveolin-1 in caveole biogenesis, endocytosis, cell proliferation, and endothelial nitric-oxide synthase (eNOS) signaling. Surprisingly, Cav-1−/− null mice are viable. We show that these mice lack caveolin-1 protein expression and plasmaendothelial caveoles. In addition, analysis of cultured fibroblasts from Cav-1−/− null embryos reveals the following: (i) a loss of caveolin-2 protein expression; (ii) defects in the endocytosis of a known caveolar ligand, i.e., fluorescent isothiocyanate-albumin; and (iii) a hyperproliferative phenotype. Importantly, these phenotypic changes are reversed by recombinant expression of the caveolin-1 (C-1)-DNA. Furthermore, examination of the lung parenchyma (an endothelial-rich tissue) shows hypercellularity with thickened alveolar septa and an increase in the number of vascular endothelial growth factor receptor (VEGF-1) positive endothelial cells. As predicted, endothelial cells from Cav-1−/− null mice lack caveolae membranes. Finally, we examined eNOS signaling by measuring the physiological response of aortic rings to various stimuli. Our results indicate that eNOS activity is up-regulated in Cav-1−/− null mice, and to mimic these findings a specific NOS inhibitor, nitro-L-arginine methyl ester. These findings are in accordance with previous in vitro studies showing that caveolin-1 is an endogenous inhibitor of eNOS. Thus, caveolin-1 expression is required to stabilize the caveolin-2 protein product, to mediate the caveolar endocytosis of specific ligands, to negatively regulate the proliferation of certain cell types, and to provide tonic inhibition of eNOS activity in endothelial cells.

PMID: 11457855 [PubMed - indexed for MEDLINE]
I N T R O D U C T I O N T O M I C R O A R R A Y S

Ingenuity statistical information

FAQs about Statistical Calculations

To assist you with understanding how Ingenuity Pathways Analysis calculates the statistical values displayed in Functional Analysis and Global Functions and Pathways, here are answers to some frequently asked questions.

- **How are the significances/p-values in Ingenuity Pathways Analysis calculated?** These calculations are based on the hypergeometric distribution calculated via the computationally efficient Fisher's Exact Test for 2x2 contingency tables. More precisely, it is the right-tailed Fischer Exact test we are employing. Right-tailed here refers to the fact that we only show over-represented functional/pathway annotations, that is annotations which have more focus genes than expected by chance. We currently do not show left-tailed, that is under-represented annotations, which have significantly fewer focus genes than expected by chance.

- **Why are the significance/p-value calculations in the application not based on a binomial distribution?** The difference between the hypergeometric and binomial distributions is that the hypergeometric one calculates probabilities without replacement and the binomial assumes replacement. Since each gene can only be used once in each p-value calculation, no replacement should be considered, so the binomial distribution cannot be used.

- **Are Bonferroni corrections used for significance/p-value calculations in Ingenuity Pathways Analysis?** No, we do not apply a Bonferroni correction, which is one of several ways to attempt to correct for testing the same data against multiple hypotheses (annotations). That correction increases the false negative rate by overcorrection (causing more annotations to fall below 0.05 and thus not be shown). For that reason we currently do not attempt to correct for multiple testing. Applying the correction would not change the order of the results (annotations in descending order by p-value would remain in same order before and after correction) so users should have the highest confidence in the annotations with the lowest p-values, and can discount annotations with higher p-values as they see fit.

- **If I got a 50% hit to a small canonical pathway (10 focus genes/ 20 in the pathway) does that score differently than a 50% hit to a big pathway (200 focus genes/ 400 in the pathway)?** It is quite likely the statistic is dependent upon a number of factors. At least two factors would affect the calculation in this example, a more significant p-value will result if the relative proportion of focus genes from a dataset implicated in a specific pathway is greater. For example if there are 200 focus genes in the data set, the 50/200 focus gene pathway would have a greater proportion than the 10/20 focus gene pathway (50/200 vs. 10/200) and this would contribute to a lower p-value for the 50/100 pathway. However, a more significant p-value will also result if the proportion of total genes in the pathway relative to the reference gene set is smaller. The 10/20 gene pathway would have a smaller proportion than the 50/100 pathway, therefore this would contribute to a lower p-value for the 10/20 pathway. Since we would not necessarily expect these two counter-acting proportions to exactly cancel each other out, the resulting p-value scores would likely be different.

- **If I am using a customized chip that focuses on a particular pathway or disease, for example angiogenic genes, how would that impact the p-values in Global Functional Analysis?** The p-values that are generated depend on the exact question you are asking.

Question 1: How do the functions for these angiogenic genes compare to the functions for all the mammalian genes (for which there is information)?

To answer this question you can load the identifiers and their expression values for the entire chip into Ingenuity Pathways Analysis and select a cutoff value, but you must choose to use the IPKB genes as your reference gene set. Alternatively you can load just the significant angiogenic genes into Ingenuity Pathways Analysis and choose to use the IPKB genes as your reference gene set.

Question 2: If only a subset of the angiogenic genes are significant, how do the functions for these significant angiogenic genes compare to the functions for all the angiogenic genes on my chip (for which there is information)?

© Copyright 2004, Kevin R. Coombes and Keith A. Baggerly

GS01 0163: ANALYSIS OF MICROARRAY DATA