Maastricht University

BiGCaT department of bioinformatics

Pathway Analysis

BGK2007 - Blok 2.5

20 April 2016

Martina Summer-Kutmon, PhD

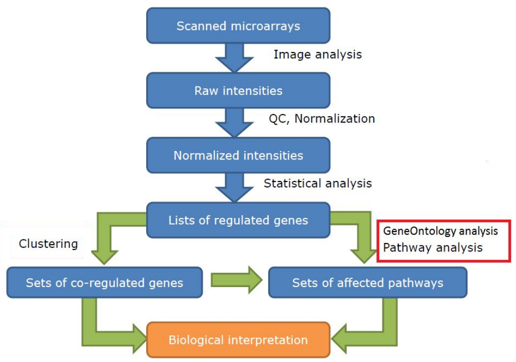
Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- Pathway analysis
- Software: PathVisio

Break

- Gene Ontology
- GO analysis
- Software: GOrilla

Outline



Scanned microarrays

Image analysis

Raw intensities

QC, Normalization

Normalized intensities

Statistical analysis

Lists of regulated genes

Clustering

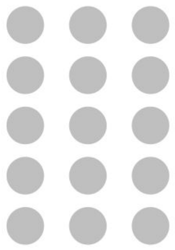
GeneOntology analysis
Pathway analysis

Sets of co-regulated genes

Sets of affected pathways

Biological interpretation

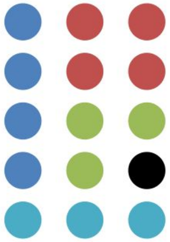
Introduction



Quantify

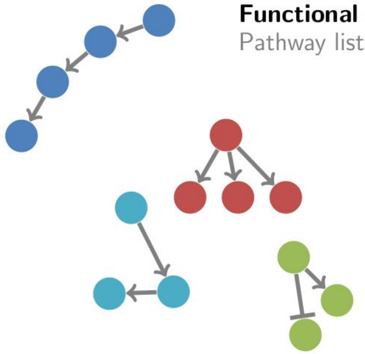
Isolated Data Points

Introduction



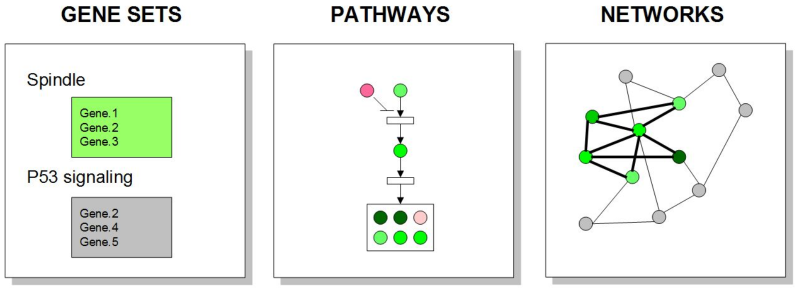
Comparative statistics
Isolated lists
Clustering
Isolated groups
Gene sets
Functional gene sets

Introduction



Functional organization
Pathway lists

Functional enrichment

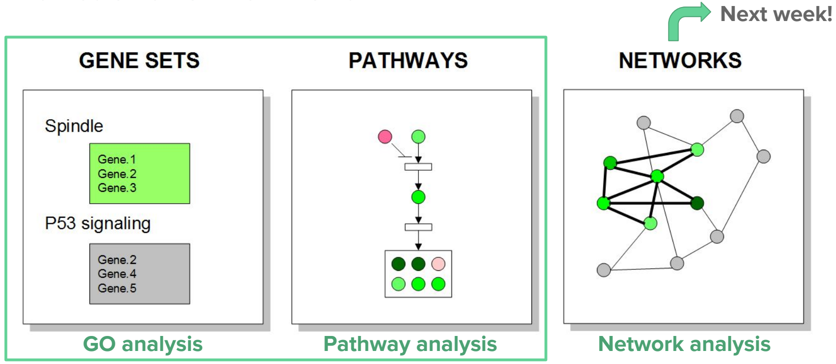


GENE SETS
Spindle
Gene.1
Gene.2
Gene.3
P53 signaling
Gene.2
Gene.4
Gene.5

PATHWAYS

NETWORKS

Functional enrichment



Next week!

GENE SETS
Spindle
Gene.1
Gene.2
Gene.3
P53 signaling
Gene.2
Gene.4
Gene.5
GO analysis

PATHWAYS
Pathway analysis

NETWORKS
Network analysis

Outline

- Introduction
- **Biological pathways**
- Pathway databases: WikiPathways
- Pathway analysis
- Software: PathVisio

Break

- Gene Ontology
- GO analysis
- Software: GOrilla

Biological pathways

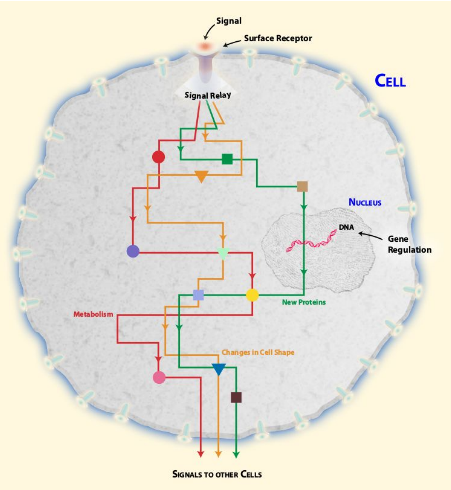
“A biological pathway is a series of **interactions among molecules** in a cell that leads to a certain **product or change** in a cell.”

Definition Wikipedia 2016

Biological pathways

Types of pathways:

- Metabolic pathways
- Gene regulation pathways
- Signal transduction pathways



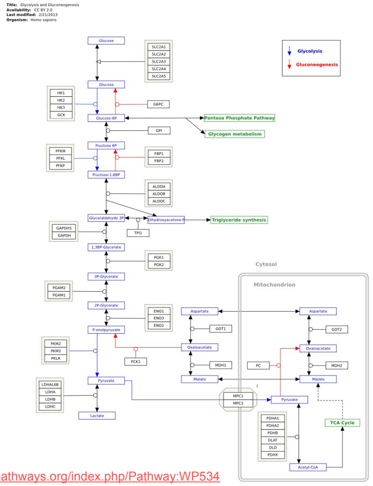
http://www.genome.gov/27530687

Metabolic pathways

- Series of chemical reactions occurring within a cell
- Enzymes are catalyzing these conversions
- Example: Glycolysis
    - breaks down glucose and forms pyruvate with the production of two molecules of ATP

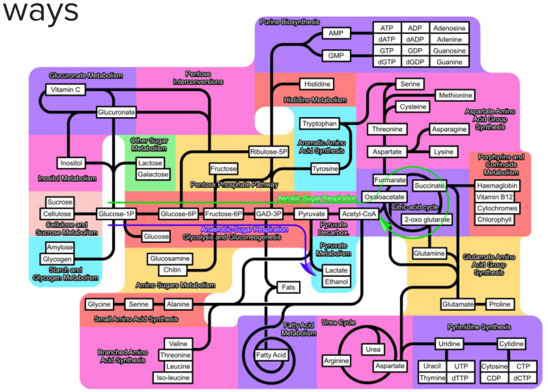
Metabolic pathways

- Series of chemical reactions occurring within a cell
- Enzymes are catalyzing these conversions
- Example: Glycolysis
    - breaks down glucose and forms pyruvate with the production of two molecules of ATP



http://wikipathways.org/index.php/Pathway:WP534

Metabolic pathways



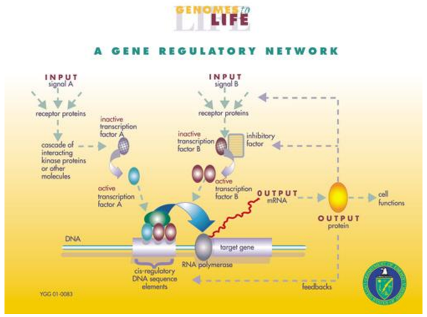
https://en.wikipedia.org/wiki/Metabolic_pathway

Gene regulation pathways

- Pathways that regulate the level of gene expression
- Often sub-pathways of larger pathways
- Transcription factors

Gene regulation pathways

- Pathways that regulate the level of gene expression
- Often sub-pathways of larger pathways
- Transcription factors



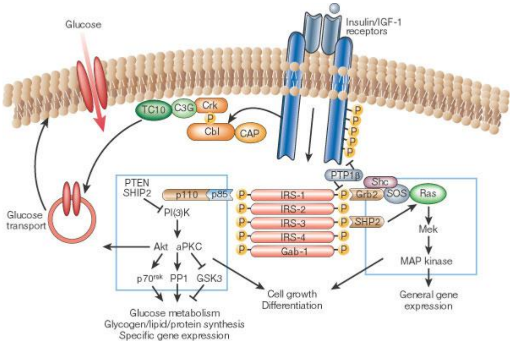
https://en.wikipedia.org/wiki/Gene_regulatory_network

Signal transduction pathways

- Extracellular signaling molecule
- Activation of a specific receptor
- Triggers a biochemical chain of events inside the cell
- Eventually eliciting a response

Signal transduction pathways

- Extracellular signaling molecule
- Activation of a specific receptor
- Triggers a biochemical chain of events inside the cell
- Eventually eliciting a response



http://www.nature.com/nature/journal/v414/n6865/images/414799a-f2.2.jpg

---

Outline

- Introduction
- Biological pathways
- **Pathway databases: WikiPathways**
- Pathway analysis
- Software: PathVisio

Break

- Gene Ontology
- GO analysis
- Software: GOrilla

Pathway databases

PathGuide (www.pathguide.org)
List of > 500 pathway and interaction related resources
109 resources containing pathway information

Pathway databases

PathGuide (www.pathguide.org)
List of > 500 pathway and interaction related resources
109 resources containing pathway information

Commonly used pathway databases:

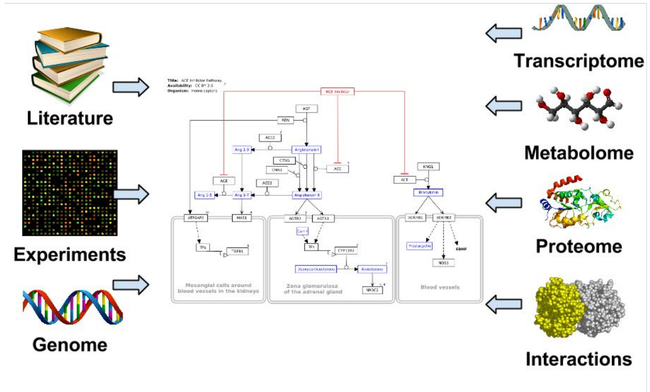
- WikiPathways (www.wikipathways.org)
- Reactome (www.reactome.org)
- KEGG (www.genome.jp/kegg)
- BioCyc (www.biocyc.org)
- ...

Pathway databases

Be aware!

- Only about 50% of the human protein coding genes are present in biological pathways.
- A lot of information is hidden in literature and researchers' minds.
- Detailed functions and mechanisms for many proteins still unknown.

Pathway databases



WikiPathways

- Pathway database developed in 2008 at Maastricht University (BiGCaT) and Gladstone Institutes in San Francisco

WikiPathways

- Pathway database developed in 2008 at Maastricht University (BiGCaT) and Gladstone Institutes in San Francisco

WikiPathways - What is a wiki?

“A wiki is an application, typically a web application, which allows collaborative modification, extension, or deletion of its content and structure.”

Definition Wikipedia 2016

WikiPathways

Wikis:

- Simple formatting rules
- Collection of knowledge
- Multiple contributors
- Curation of knowledge

WikiPathways

Wikis:

- Simple formatting rules
- Collection of knowledge
- Multiple contributors
- Curation of knowledge

Examples:

- General: Wikipedia, Wikiquote, Wikibooks, Wikiversity

WikiPathways

Wikis:

- Simple formatting rules
- Collection of knowledge
- Multiple contributors
- Curation of knowledge

Examples:

- General: Wikipedia, Wikiquote, Wikibooks, Wikiversity
- Biological: Wikispecies, Proteopedia, GeneWiki, WikiPathways

WikiPathways

Wikis:

- Simple formatting rules
- Collection of knowledge
- Multiple contributors
- Curation of knowledge

Examples:

- General: Wikipedia, Wikiquote, Wikibooks, Wikiversity
- Biological: Wikispecies, Proteopedia, GeneWiki, WikiPathways
- Specific topics: Jedipedia, Simpson wiki, Lostpedia

WikiPathways

A wikipedia for pathways - www.wikipathways.org

- Everybody can contribute and share pathways
- Everybody can edit and curate pathways
- Everybody can use the pathway collections

WikiPathways
Pathways for the People

WikiPathways

A wikipedia for pathways - www.wikipathways.org

- Everybody can contribute and share pathways
- Everybody can edit and curate pathways
- Everybody can use the pathway collections
- Not just images but fully annotated models
- Interactive pathway viewer and editor
- Full pathway analysis software available - PathVisio

WikiPathways
Pathways for the People

WikiPathways

A wikipedia for pathways - www.wikipathways.org

- Everybody can contribute and share pathways
- Everybody can edit and curate pathways
- Everybody can use the pathway collections
- Not just images but fully annotated models
- Interactive pathway viewer and editor
- Full pathway analysis software available - PathVisio

- Important: New findings can be added immediately - very fast!

WikiPathways
Pathways for the People

Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- **Pathway analysis**
- Software: PathVisio

Break

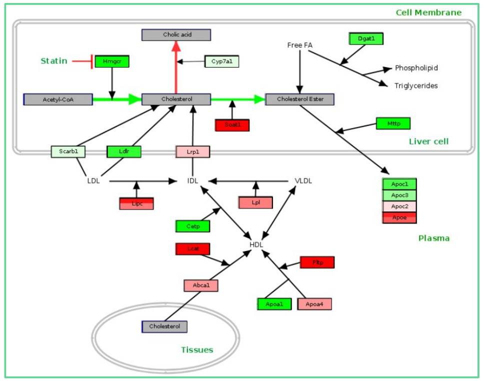
- Gene Ontology
- GO analysis
- Software: GOrilla

Pathway analysis

Hmgcr
Dgat1
Ldlr
Mttp
Soat1
Lipc
Pltp
Lcat

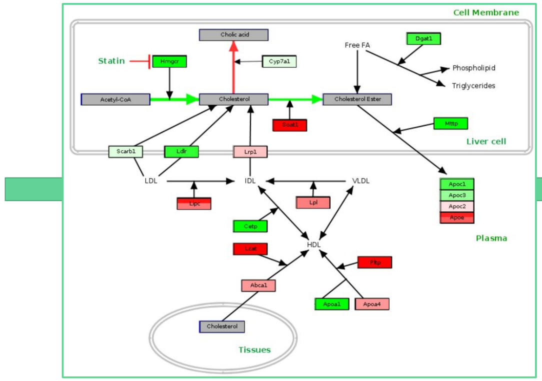
Pathway analysis

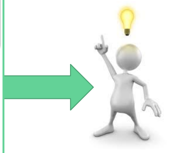
Hmgcr
Dgat1
Ldlr
Mttp
Soat1
Lipc
Pltp
Lcat



Pathway analysis

Hmgcr
Dgat1
Ldlr
Mttp
Soat1
Lipc
Pltp
Lcat





Biological Interpretation

Pathway analysis

“A picture is worth a thousand words.”

- Intuitive and simple
- Puts data into a biological context → analysis on functional level
- More efficient than looking up single gene information

Pathway analysis

“A picture is worth a thousand words.”

- Intuitive and simple
- Puts data into a biological context → analysis on functional level
- More efficient than looking up single gene information
- Reduce complexity by grouping genes, proteins and other molecules → several hundred pathways instead of thousands of genes
- Higher explanatory power than a simple gene list
- Visual representation

Pathway Analysis

In Research:

1. Pathway as a working hypothesis
    - Describe the mechanisms you are studying in a pathway model
    - Find out if data supports hypothesis
    - Fast, easy and with a graphical representation

Pathway Analysis

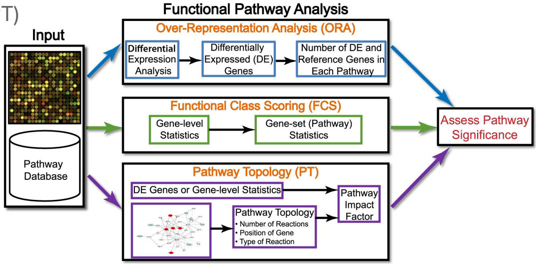
In Research:

1. Pathway as a working hypothesis
    - Describe the mechanisms you are studying in a pathway model
    - Find out if data supports hypothesis
    - Fast, easy and with a graphical representation

2. Find pathways altered in experimental data
    - Pathways that are different between experimental settings (healthy vs. disease / before vs. after treatment)
    - Pathways that change over time

Pathway analysis methods

1. Overrepresentation analysis (ORA)
2. Functional Class Scoring (FCS)
3. Pathway Topology Based (PT)



Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Computational Biology, 8(2), e1002375.

Pathway analysis methods

Overrepresentation analysis:

1. Input gene list ➔ e.g. significantly up- or down-regulated genes
2. Background gene list ➔ e.g. all measured genes
3. Perform statistical test ➔ e.g. Fisher's exact test (hypergeometric test)

$$\text{Z-score} = \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$$

$$\text{Z-score} = \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$$

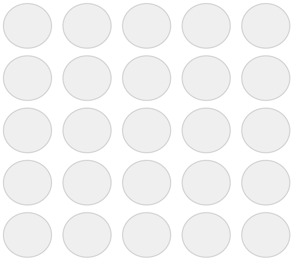
Z-Score ORA

Z-Score is calculated for each pathway ➔ result is ranked list of pathways

1. N = background list (total number of measured genes in experiment)
2. R = input list (number of changed genes in experiment)
3. n = total number of genes in pathway
4. r = number of changed genes in pathway

$$\text{Z-score} = \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$$

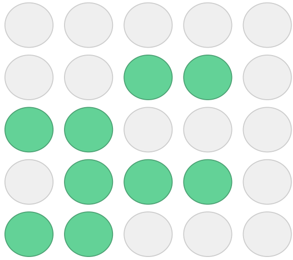
Z-Score ORA



1. **N = 25**
    - background list (total number of measured genes in experiment)
2. R =
    - input list (number of changed genes in experiment)
3. n =
    - total number of genes in pathway
4. r =
    - number of changed genes in pathway

Z-score $= \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$

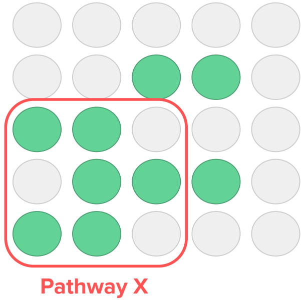
Z-Score ORA



1. N = 25
    - background list (total number of measured genes in experiment)
2. **R = 9**
    - input list (number of changed genes in experiment)
3. n =
    - total number of genes in pathway
4. r =
    - number of changed genes in pathway

Z-score $= \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$

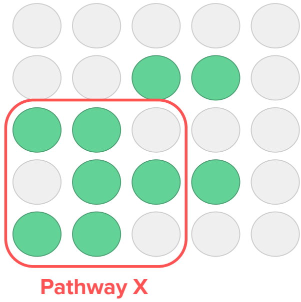
Z-Score ORA



1. N = 25
    - background list (total number of measured genes in experiment)
2. R = 9
    - input list (number of changed genes in experiment)
3. **n = 9**
    - total number of genes in pathway
4. r =
    - number of changed genes in pathway

Z-score $= \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$

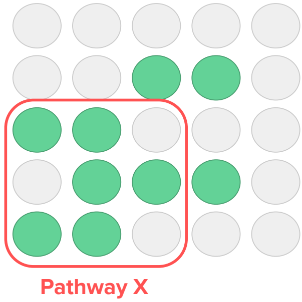
Z-Score ORA



1. N = 25
    - background list (total number of measured genes in experiment)
2. R = 9
    - input list (number of changed genes in experiment)
3. n = 9
    - total number of genes in pathway
4. **r = 6**
    - number of changed genes in pathway

Z-score $= \frac{(r - n\frac{R}{N})}{\sqrt{n\frac{R}{N}(1-\frac{R}{N})(1-\frac{n-1}{N-1})}}$

Z-Score ORA



1. N = 25
    - background list (total number of measured genes in experiment)
2. R = 9
    - input list (number of changed genes in experiment)
3. n = 9
    - total number of genes in pathway
4. r = 6
    - number of changed genes in pathway

Z-Score for pathway X = 2.347

Z-Score ORA

What does the Z-Score tell you?

- Z-Score > 1.96
  - Significantly more genes than expected are changed in the pathway → different between the experimental settings
- Z-Score = 0
  - Distribution of changed genes in the pathway is the same as in the complete dataset
- Z-Score = < -1.96
  - Significantly less genes than expected are changed in the pathway → very stable

---

Z-Score ORA

Be aware!!

What does the Z-Score tell you?

- Z-Score > 1.96
  - Significantly more genes than expected are changed in the pathway → different between the experimental settings
- Z-Score = 0
  - Distribution of changed genes in the pathway is the same as in the complete dataset
- Z-Score = < -1.96
  - Significantly less genes than expected are changed in the pathway → very stable

---

Z-Score ORA

Be aware!

- ORA and FCS **do not** take pathway topology into account!
- You don’t know yet where the changes occur in the pathway.
- Always look at the pathway diagrams and study the changes to make the right conclusions!

Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- Pathway analysis
- **Software: PathVisio**

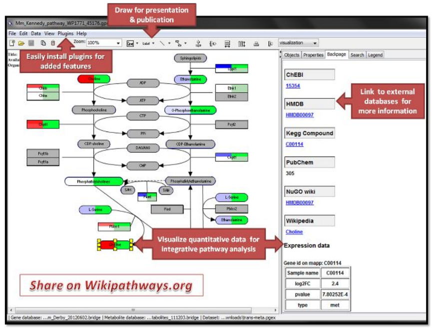
Break

- Gene Ontology
- GO analysis
- Software: GOrilla

Pathway analysis with PathVisio

PathVisio 3.2.2 (www.pathvisio.org)

- Draw and annotate biological pathways
- Visualize your experimental data on pathways
- Find pathways regulated in your dataset



Draw for presentation & publication

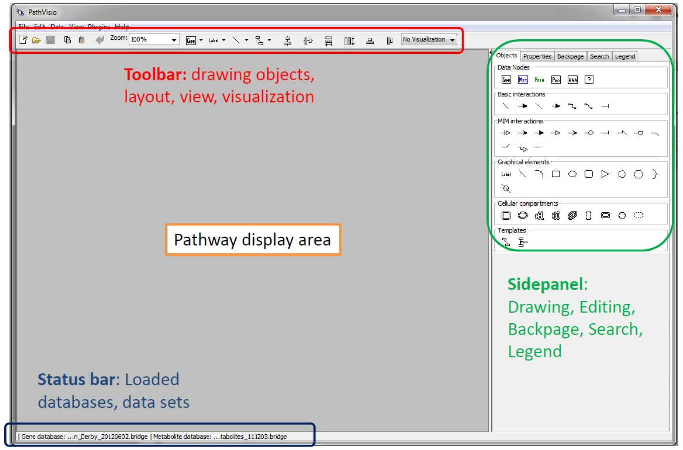
Easily install plugins for added features

Link to external databases for more information

Visualize quantitative data for integrative pathway analysis

Share on Wikipathways.org

PathVisio



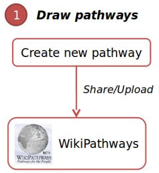
Toolbar: drawing objects, layout, view, visualization

Pathway display area

Sidepanel: Drawing, Editing, Backpage, Search, Legend

Status bar: Loaded databases, data sets

Pathway analysis with PathVisio



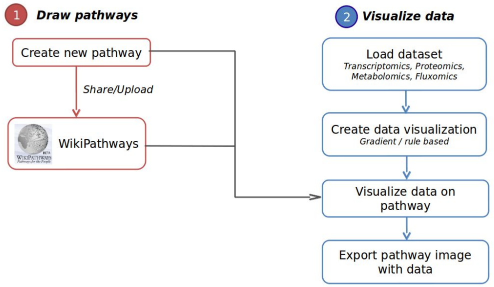
1 ***Draw pathways***

Create new pathway

Share/Upload

WikiPathways

Pathway analysis with PathVisio



1 ***Draw pathways***

Create new pathway

Share/Upload

WikiPathways

2 ***Visualize data***

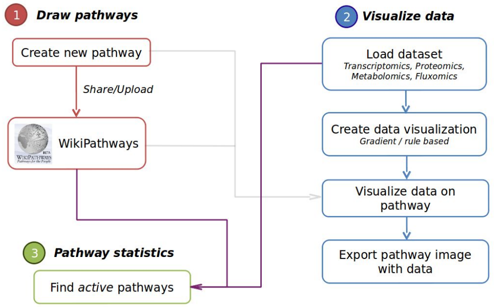
Load dataset
Transcriptomics, Proteomics, Metabolomics, Fluxomics

Create data visualization
Gradient / rule based

Visualize data on pathway

Export pathway image with data

Pathway analysis with PathVisio



1 **Draw pathways**

Create new pathway

Share/Upload

WikiPathways

2 **Visualize data**

Load dataset
Transcriptomics, Proteomics, Metabolomics, Fluxomics

Create data visualization
Gradient / rule based

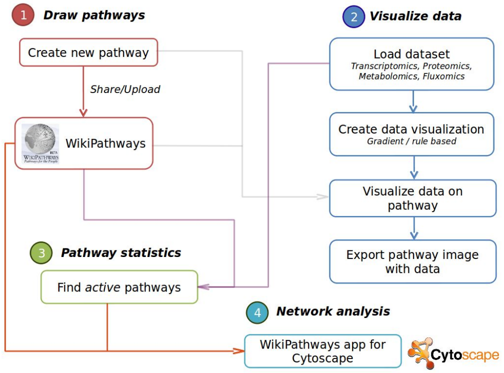
Visualize data on pathway

Export pathway image with data

3 **Pathway statistics**

Find *active* pathways

Pathway analysis with PathVisio



1 **Draw pathways**

Create new pathway

Share/Upload

WikiPathways

2 **Visualize data**

Load dataset
Transcriptomics, Proteomics, Metabolomics, Fluxomics

Create data visualization
Gradient / rule based

Visualize data on pathway

Export pathway image with data

3 **Pathway statistics**

Find *active* pathways

4 **Network analysis**

WikiPathways app for Cytoscape

Cytoscape

Pathway analysis with PathVisio

Pathways are more than just images - **fully annotated pathway models**

- Elements (data nodes and interactions) are annotated with biological database identifiers, e.g. Ensembl, UniProt, HMDB, …
- Elements are linked to publication references if available
- Users can add additional comments to elements

Identifier mapping in PathVisio

Data nodes and interactions can be annotated with biological database identifiers

- **Problem:** Many different databases ➔ different identifiers for the same biological element
- Manual mapping of identifiers is very time-consuming and error-prone

Identifier mapping in PathVisio

Data nodes and interactions can be annotated with biological database identifiers

- **Problem:** Many different databases ➔ different identifiers for the same biological element
- Manual mapping of identifiers is very time-consuming and error-prone

- **Solution**: BridgeDb framework ➔ identifier mapping databases for gene products, metabolites and interactions
  - Gene products based on Ensembl
  - Metabolites based on HMDB
  - Interactions based on Rhea

Identifier mapping in PathVisio

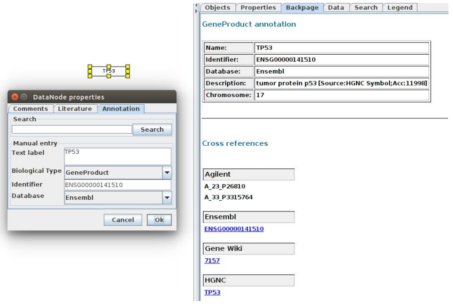
- Databases can be downloaded from
  - http://bridgedb.org/data/gene_database/

- **Important**: Load mapping database in PathVisio
  - Menu: Data ➔ Import Gene/Metabolite/Interaction Database

- Double click on element to add annotation
- Check “Backpage” side-tab to check annotation and mapped identifiers (cross references)

Identifier mapping in PathVisio

1. Draw new data node
2. Annotate with Xref (Identifier + Database)
3. Find linkouts and cross references in “Backpage” side-tab



Manual entry
Text label: TP53
Biological Type: GeneProduct
Identifier: ENSG00000141510
Database: Ensembl

GeneProduct annotation

Name:	TP53
Identifier:	ENSG00000141510
Database:	Ensembl
Description:	tumor protein p53 [Source:HGNC Symbol;Acc:11998]
Chromosome:	17

Cross references

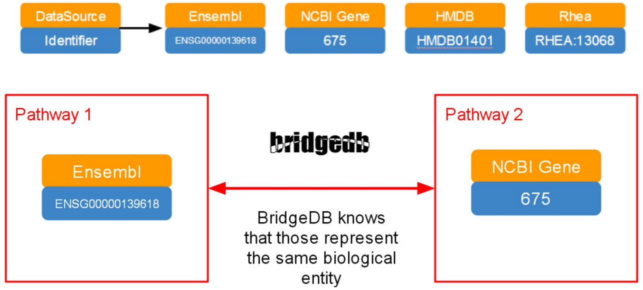
Agilent
A_23_P26810
A_33_P3315764

Ensembl
ENSG00000141510

Gene Wiki
7157

HGNC
TP53

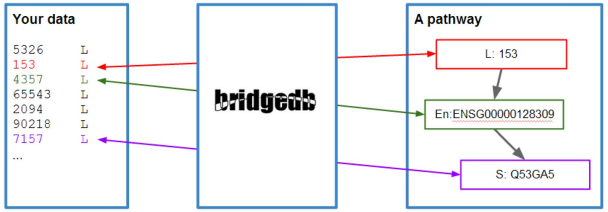
Identifier mapping in PathVisio



DataSource	Ensembl	NCBI Gene	HMDB	Rhea
Identifier	ENSG00000139618	675	HMDB01401	RHEA:13068

BridgeDB knows that those represent the same biological entity

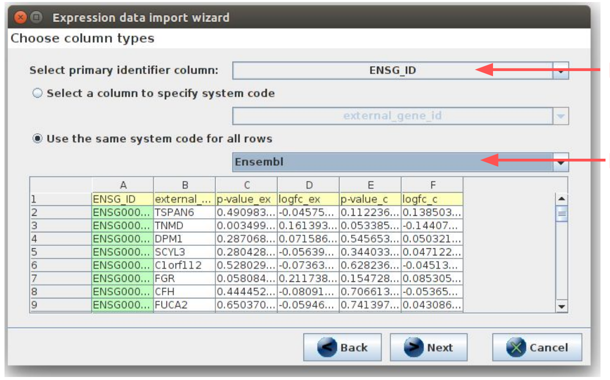
Identifier mapping in PathVisio



Importing experimental data in PathVisio

- Use only comma or tab-separated files (.csv / .txt)
    - In Excel → Save As ...
- One column must contain the identifier of one of the supported databases
- Menu: Data → Import expression data
- **Important!** Select the correct database (system code) when importing the data!

Importing experimental data in PathVisio



Choose column types

Select primary identifier column: ENSG_ID

○ Select a column to specify system code

external_gene_id

● Use the same system code for all rows

Ensembl

	A	B	C	D	E	F
1	ENSG_ID	external_...	p-value_ex	logfc_ex	p-value_c	logfc_c
2	ENSG000...	TSPAN6	0.490983...	-0.04575...	0.112236...	0.138503...
3	ENSG000...	TNMD	0.003499...	0.161393...	0.053385...	-0.14407...
4	ENSG000...	DPM1	0.287068...	0.071586...	0.545653...	0.050321...
5	ENSG000...	SCYL3	0.280428...	-0.05639...	0.344033...	0.047122...
6	ENSG000...	C1orf112	0.528029...	-0.07363...	0.628236...	-0.04513...
7	ENSG000...	FGR	0.058084...	0.211738...	0.154728...	0.085305...
8	ENSG000...	CFH	0.444452...	-0.08091...	0.706613...	-0.05365...
9	ENSG000...	FUCA2	0.650370...	-0.05946...	0.741397...	0.043086...

Identifier!

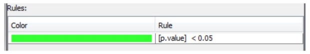
Database/system code!

Data visualization in PathVisio

Two different visualization styles:

- Gradient based visualization
    - Continuous values (e.g. log fold change)

- Rule based visualization
    - Distinct values like categories or criteria (e.g. p-value)



Data visualization in PathVisio

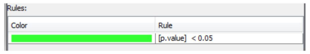
Two different visualization styles:

- Gradient based visualization
  - Continuous values (e.g. log fold change)

☑ Gradient:

-1.0 0.0 1.0

- Rule based visualization
  - Distinct values like categories or criteria (e.g. p-value)

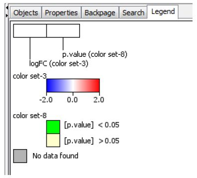


Rules:

Color	Rule
🟩	[p.value] < 0.05

- Advanced visualization - multiple visualizations shown at the same time (in columns)

Data visualization



p.value (color set-8)

logFC (color set-3)

color set-3

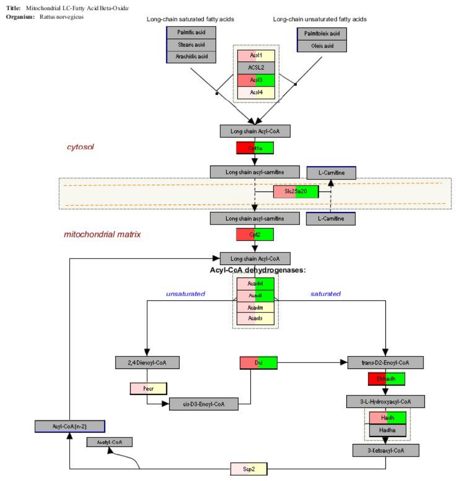
-2.0 0.0 2.0

color set-8

🟩 [p.value] < 0.05

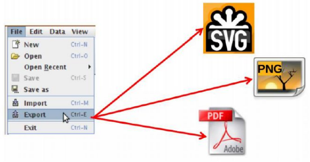
🟨 [p.value] > 0.05

⬜ No data found



Exporting figures with PathVisio

- Pathway diagrams with and without experimental data visualization can be exported as high quality images for publication



Pathway statistics

- Default statistics method: Z-Score ORA
- Menu: Data → Statistics

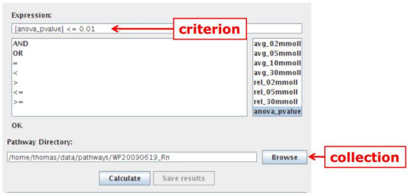
Pathway statistics

- Default statistics method: Z-Score ORA
- Menu: Data → Statistics

1. Define criterion for input list (changed genes)
    - e.g. logFC > 1 AND p.value < 0.05 (significant up-regulated genes)
2. Select pathway collection for background list (from WikiPathways)
3. PathVisio calculates Z-Score for each pathway and ranks them in the result table

---

Pathway statistics



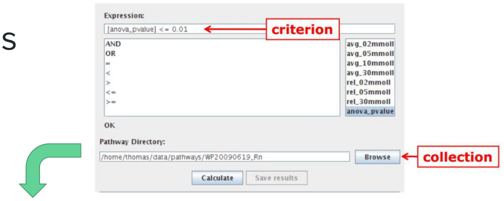
Expression:

[anova_pvalue] <= 0.01

Pathway Directory:

/home/thomas/data/pathways/WP20090619_Rn

Pathway statistics

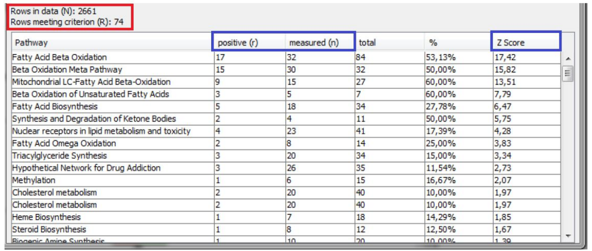


Expression:

[anova_pvalue] <= 0.01

Pathway Directory:

/home/thomas/data/pathways/WP20090619_Rn



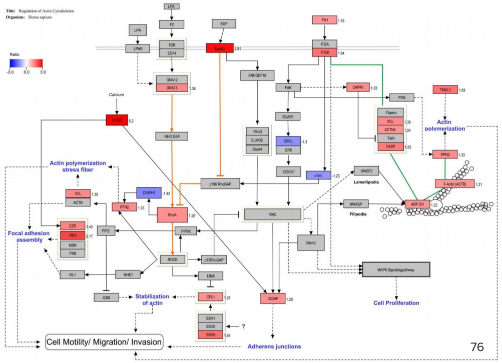
Rows in data (N): 2661
Rows meeting criterion (R): 74

Pathway	positive (r)	measured (n)	total	%	Z Score
Fatty Acid Beta Oxidation	17	32	84	53,13%	17,42
Beta Oxidation Meta Pathway	15	30	32	50,00%	15,82
Mitochondrial LC-Fatty Acid Beta-Oxidation	9	15	27	60,00%	13,51
Beta Oxidation of Unsaturated Fatty Acids	3	5	7	60,00%	7,79
Fatty Acid Biosynthesis	5	18	34	27,78%	6,47
Synthesis and Degradation of Ketone Bodies	2	4	11	50,00%	5,75
Nuclear receptors in lipid metabolism and toxicity	4	23	41	17,39%	4,28
Fatty Acid Omega Oxidation	2	8	14	25,00%	3,83
Triacylglyceride Synthesis	3	20	34	15,00%	3,34
Hypothetical Network for Drug Addiction	3	26	35	11,54%	2,73
Methylation	1	6	15	16,67%	2,07
Cholesterol metabolism	2	20	40	10,00%	1,97
Cholesterol metabolism	2	20	40	10,00%	1,97
Heme Biosynthesis	1	7	18	14,29%	1,85
Steroid Biosynthesis	1	8	12	12,50%	1,67
Biogenic Amine Synthesis	1	10	20	10,00%	1,39

Research examples

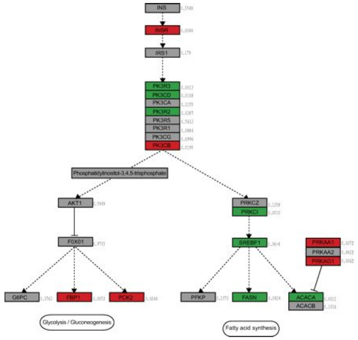
- Actin cytoskeleton regulation
- Tamoxifen resistant cell line, MCF-7-TamR vs. control
- Red = upregulated
- Blue = downregulated

Zhou, Changhua, et al. "**Proteomic analysis of acquired tamoxifen resistance in MCF-7 cells reveals expression signatures associated with enhanced migration.**" Breast Cancer Res 14.2 (2012): R45.



Research examples

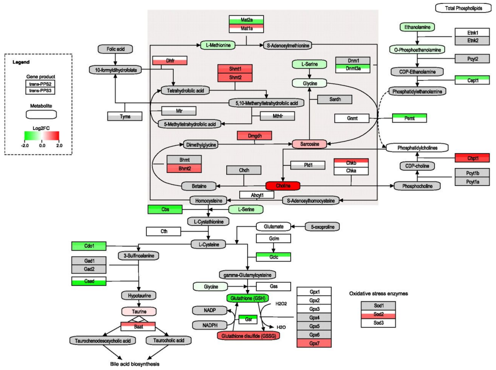
- Insulin signaling pathway
- Comparison high vs. low ratio of visceral to visceral + subcutaneous fat
- Red = genes are upregulated
- Green = genes are downregulated



Kursawe, Romy, et al. **"Cellularity and adipogenic profile of the abdominal subcutaneous adipose tissue from obese adolescents: association with insulin resistance and hepatic steatosis."** Diabetes 59.9 (2010): 2288-2296.

Research examples

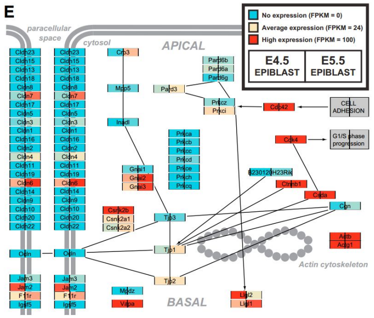
- Multi-omics - transcriptomics + metabolomics
- High fat diet vs. control



Rubio-Aliaga, Isabel, et al. "**Alterations in hepatic one-carbon metabolism and related pathways following a high-fat dietary intervention.**" Physiological genomics 43.8 (2011): 408-416.

Research examples

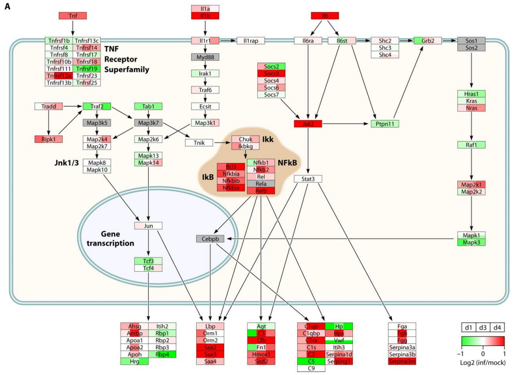
- Time-series data
- Expression in the preimplantation (left) and postimplantation epiblast (right)



Boroviak, Thorsten, et al. "**Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis.**" Developmental cell 35.3 (2015): 366-382.

Research examples

- Time-series data
- Early response after virus infection



Tisoncik, Jennifer R., et al. **"Into the eye of the cytokine storm."** Microbiology and Molecular Biology Reviews 76.1 (2012): 16-32.

Questions Pathway Analysis?

Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- Pathway analysis
- Software: PathVisio

Break

- **Gene Ontology**
- GO analysis
- Software: GOrilla

Gene Ontology

www.geneontology.org

- Founded in 1998
- Ontology = formal naming and definition of types, properties and interrelationships of entities for a particular domain
- Gene Ontology = description of gene functions (controlled vocabulary!)

Gene Ontology

www.geneontology.org

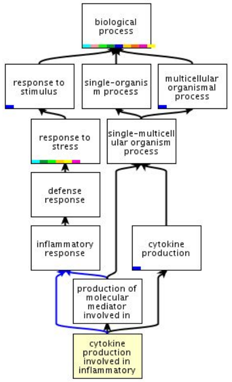
- Founded in 1998
- Ontology = formal naming and definition of types, properties and interrelationships of entities for a particular domain
- Gene Ontology = description of gene functions (controlled vocabulary!)

- Hierarchical structure
  - **Molecular function** (molecular activity of gene products)
  - **Cellular component** (where gene products are active)
  - **Biological process** (pathways made up of the activities of multiple gene products)

Gene Ontology structure

- All annotations in the ontology are called terms
- Terms are organized hierarchically
  - Terms on top are more general
  - Terms on bottom are more specific

Gene Ontology structure



- All annotations in the ontology are called terms
- Terms are organized hierarchically
  - Terms on top are more general
  - Terms on bottom are more specific
- If a gene product plays a role in the “*cytokine production involved in inflammatory response*”, then it will automatically be annotated with all parent terms (up-propagation)

Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- Pathway analysis
- Software: PathVisio

Break

- Gene Ontology
- **GO analysis**
- Software: GOrilla

GO analysis

- Pathway collections are still incomplete
- It is much harder to define the exact mechanisms and interactions than associating a protein with a biological process

GO analysis

- Pathway collections are still incomplete
- It is much harder to define the exact mechanisms and interactions than associating a protein with a biological process

- GO provides nearly 600.000 annotations for human (about 100.000 experimentally validated)
- Genes associated with a term can be considered a gene set for ORA or Functional Class Scoring methods (e.g. Gene set enrichment analysis, GSEA)

Outline

- Introduction
- Biological pathways
- Pathway databases: WikiPathways
- Pathway analysis
- Software: PathVisio

Break

- Gene Ontology
- GO analysis
- **Software: GOrilla**

GO analysis with GOrilla

http://cbl-gorilla.cs.technion.ac.il/

- **Again**: input and background gene list needed
    - Input / Target gene set = list of changed genes
    - Background gene set = list of measured genes

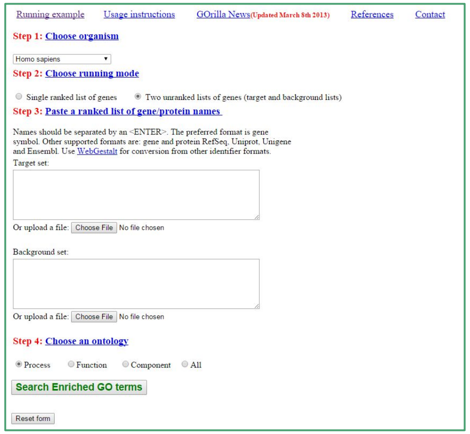
GO analysis with GOrilla

http://cbl-gorilla.cs.technion.ac.il/

- **Again**: input and background gene list needed
    - Input / Target gene set = list of changed genes
    - Background gene set = list of measured genes

- Tool performs Functional Class Scoring method (GSEA)
- Identifies GO terms that are overrepresented in the input/target gene set
- Provides a GO tree (visual, can be very large) and a list of significant GO terms

GO analysis with GOrilla



Step 1: Choose organism

Homo sapiens ▾

Step 2: Choose running mode

○ Single ranked list of genes ● Two unranked lists of genes (target and background lists)

Step 3: Paste a ranked list of gene/protein names

Names should be separated by an <ENTER>. The preferred format is gene symbol. Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and Ensembl. Use WebGestalt for conversion from other identifier formats.

Target set:

Or upload a file: Choose File No file chosen

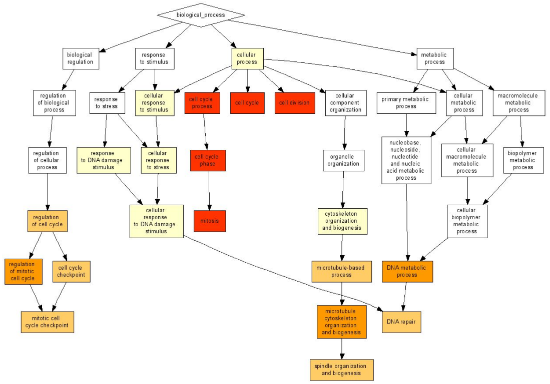
Background set:

Or upload a file: Choose File No file chosen

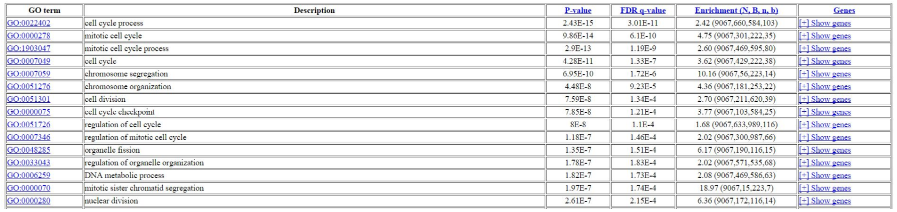
Step 4: Choose an ontology

● Process ○ Function ○ Component ○ All

GO analysis with GOrilla



GO analysis with GOrilla



GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0022402	cell cycle process	2.43E-15	3.01E-11	2.42 (9067,660,584,103)	[+] Show genes
GO:0000278	mitotic cell cycle	9.86E-14	6.1E-10	4.75 (9067,301,222,35)	[+] Show genes
GO:1903047	mitotic cell cycle process	2.9E-13	1.19E-9	2.60 (9067,469,595,80)	[+] Show genes
GO:0007049	cell cycle	4.28E-11	1.33E-7	3.62 (9067,429,222,38)	[+] Show genes
GO:0007059	chromosome segregation	6.95E-10	1.72E-6	10.16 (9067,56,223,14)	[+] Show genes
GO:0051276	chromosome organization	4.48E-8	9.23E-5	4.36 (9067,181,253,22)	[+] Show genes
GO:0051301	cell division	7.59E-8	1.34E-4	2.70 (9067,211,620,39)	[+] Show genes
GO:0000075	cell cycle checkpoint	7.85E-8	1.21E-4	3.77 (9067,103,584,25)	[+] Show genes
GO:0051726	regulation of cell cycle	8E-8	1.1E-4	1.68 (9067,633,989,116)	[+] Show genes
GO:0007346	regulation of mitotic cell cycle	1.18E-7	1.46E-4	2.02 (9067,300,987,66)	[+] Show genes
GO:0048285	organelle fission	1.35E-7	1.51E-4	6.17 (9067,190,116,15)	[+] Show genes
GO:0033043	regulation of organelle organization	1.78E-7	1.83E-4	2.02 (9067,571,535,68)	[+] Show genes
GO:0006259	DNA metabolic process	1.82E-7	1.73E-4	2.08 (9067,469,586,63)	[+] Show genes
GO:0000070	mitotic sister chromatid segregation	1.97E-7	1.74E-4	18.97 (9067,15,223,7)	[+] Show genes
GO:0000280	nuclear division	2.61E-7	2.15E-4	6.36 (9067,172,116,14)	[+] Show genes

---

Questions GO Analysis?

Practical session

1. Further investigating the real experimental dataset from the *Gene Expression* practical
2. Perform pathway analysis in PathVisio
3. Perform GO analysis with GOrilla

Practical session

1. Further investigating the real experimental dataset from the *Gene Expression* practical
2. Perform pathway analysis in PathVisio
3. Perform GO analysis with GOrilla

Goal:
Biological interpretation of the experimental dataset to better understand which processes are affected by the prophylactic human papillomavirus (HPV) vaccines.

Learning goals

1. What are biological pathways and which types are there?
2. Where can you find biological pathway data?
3. How does pathway analysis work?
4. How can you use PathVisio to perform pathway analysis?
5. What is the Gene Ontology?
6. How can you use GOrilla to perform GO analysis?
7. What are the advantages/disadvantages of GO/pathway analysis?

Questions?