Maastricht University Leading in Learning!

SaferNano Design bioHC

BiGCaT bioinformatics

Gene expression data analysis

SaferNanoDesign 29.05.2018

Dr. Friederike Ehrhart

Department for Bioinformatics – BiGCaT

Maastricht University

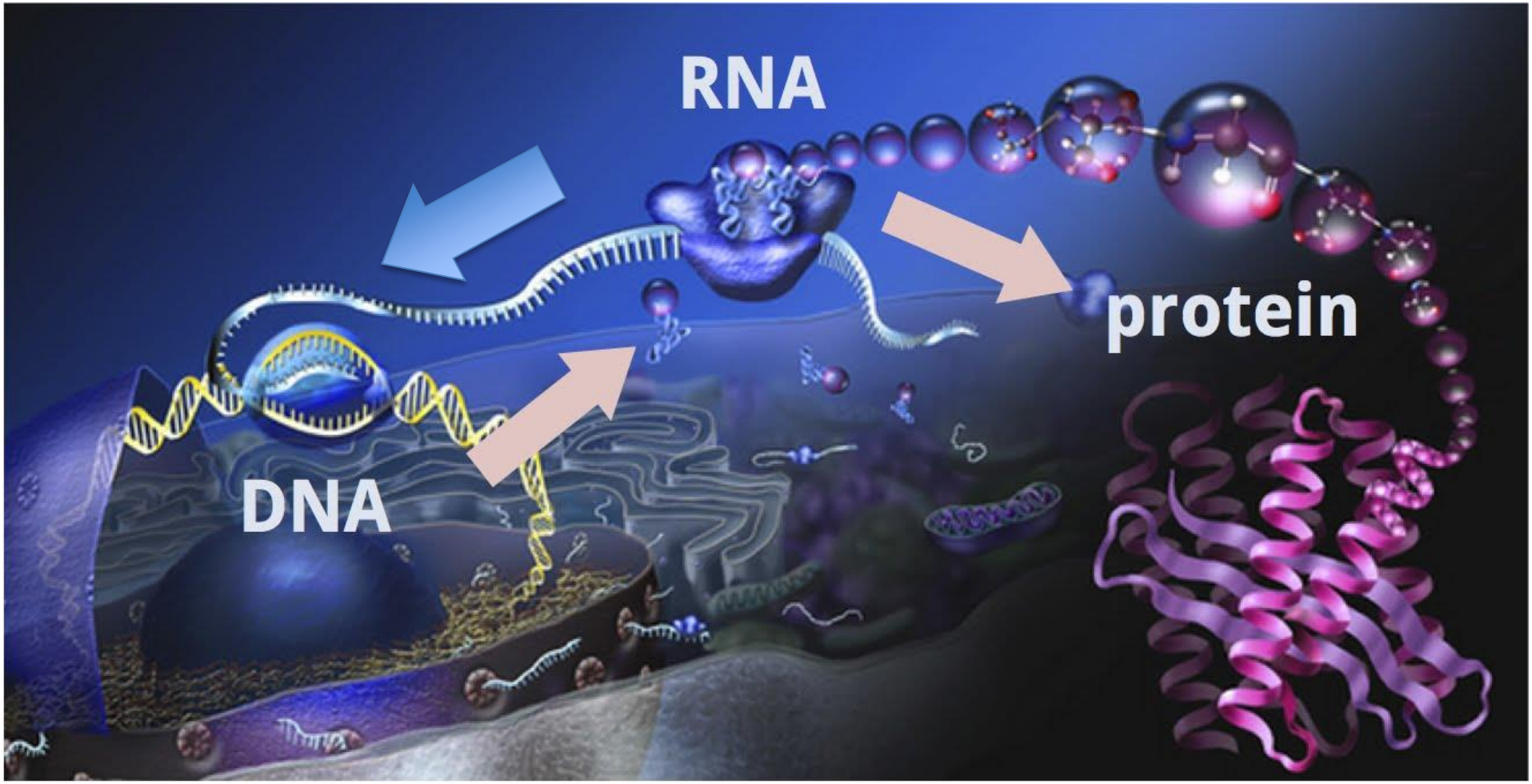
friederike.ehrhart@maastrichtuniversity.nl

ORCID: 0000-0002-7770-620X

Content

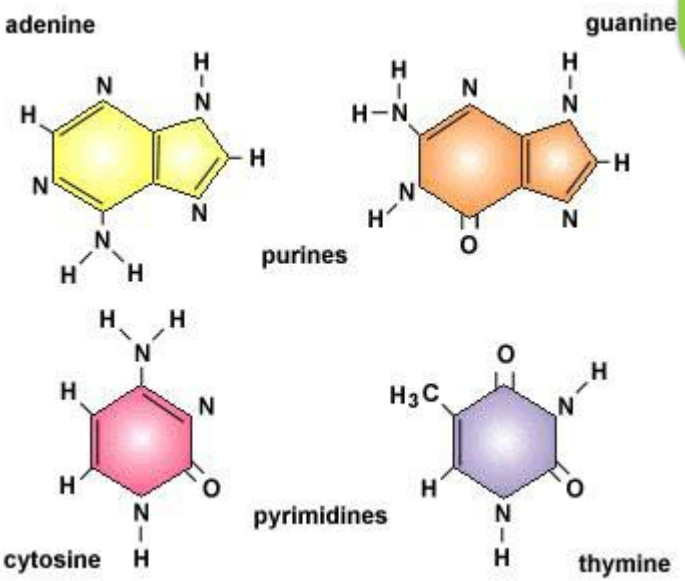
1. The dogma of molecular biology – DNA/RNA/protein relationship
2. Data and omics data – and their research strategies
3. Bioinformatics and databases
4. How to do gene expression data analysis – ArrayAnalysis.org and PathVisio
5. Limitations and pitfalls

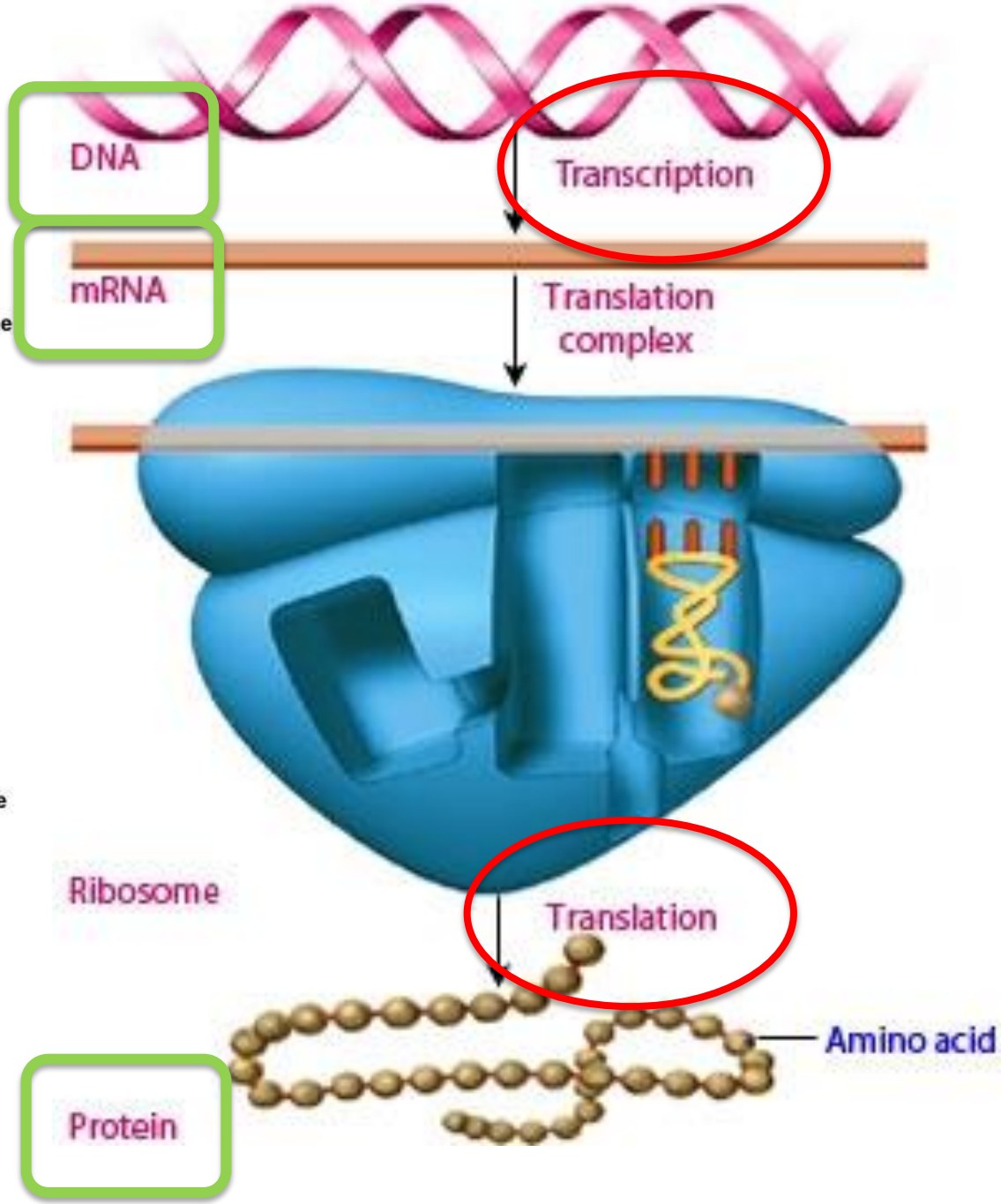
1. The dogma of molecular biology



http://www.youtube.com/watch?feature=playr_detailpage&v=9kOGOY7vthke

DNA

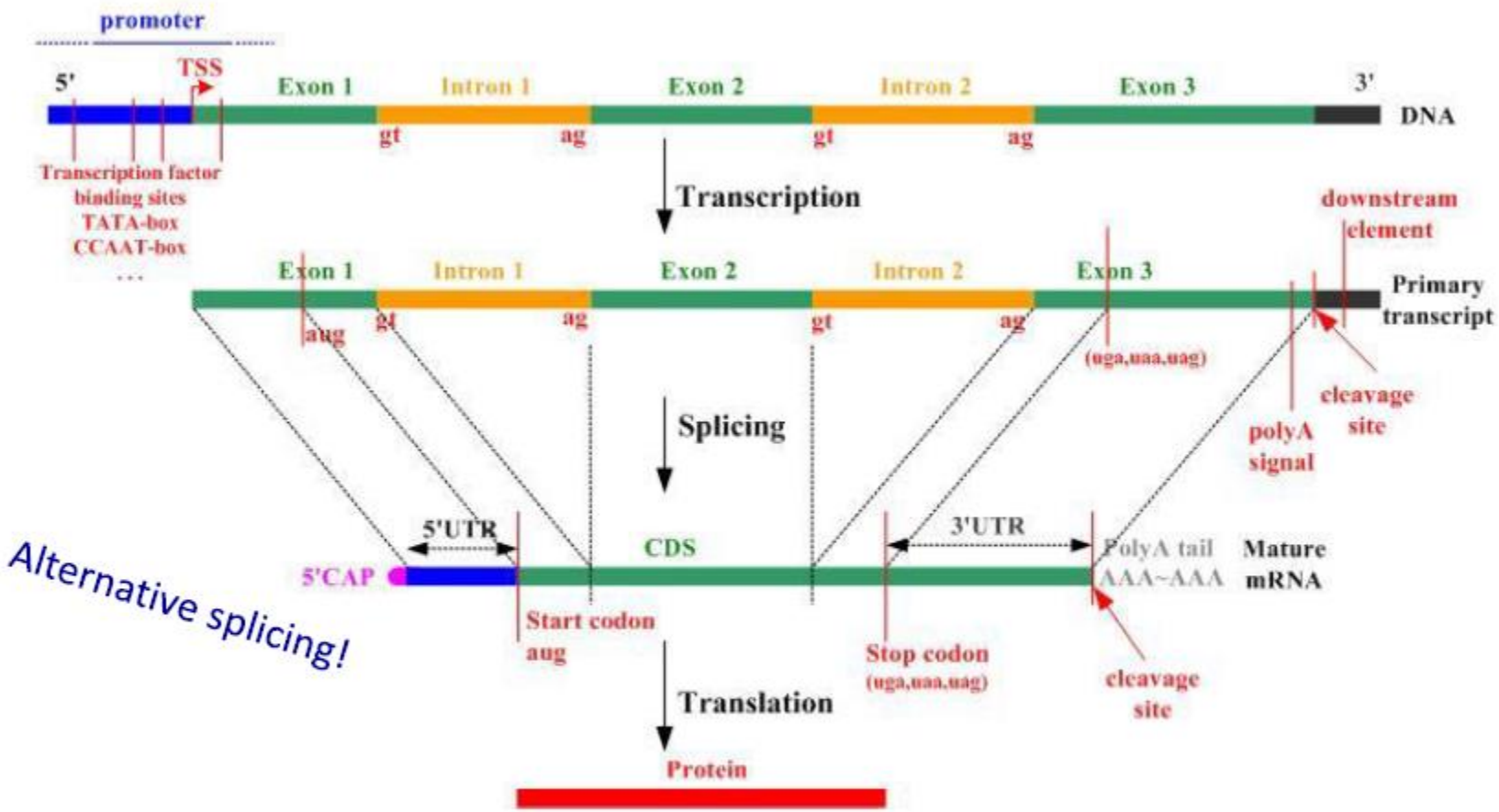




http://www.wonderwhizkids.com.sg/biology/wonders-of-the-inner-world

DNA and genes

A gene is a locus (or region) of the DNA that encodes a functional RNA.



Alternative splicing!

CDS = Coding DNA Sequence
UTR = UnTranslated region

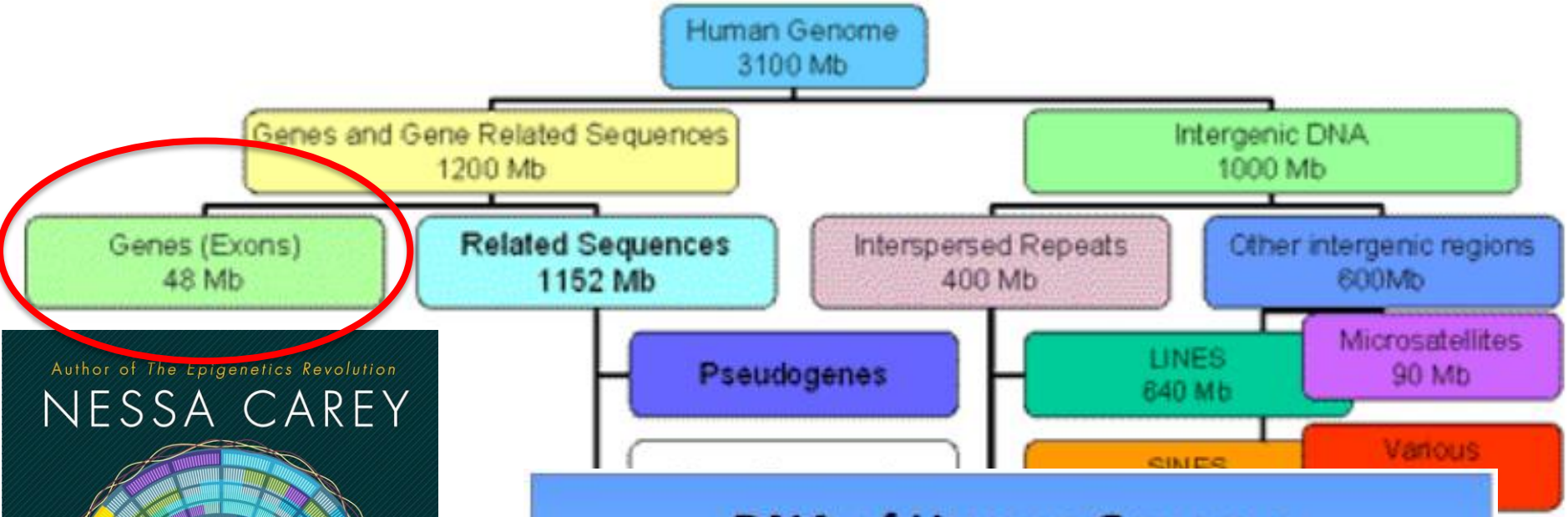
Number of genes per species

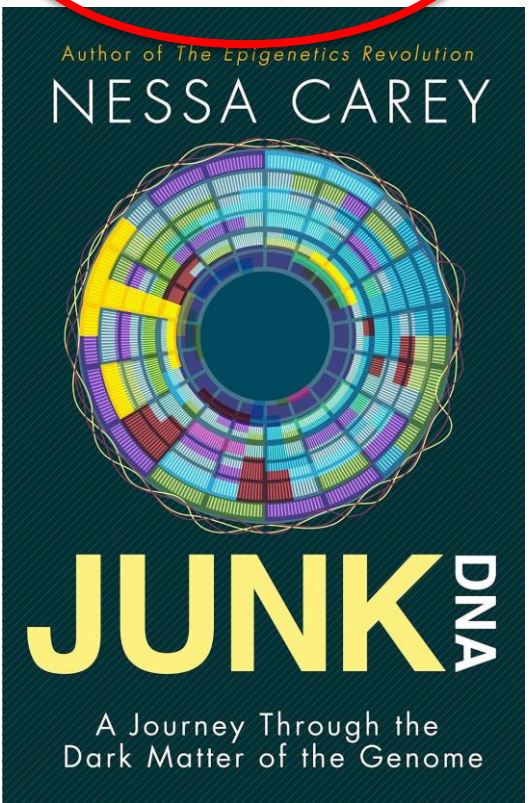
Species and Common Name	Estimated Total Size of Genome (bp)*	Estimated Number of Protein-Encoding Genes*
Saccharomyces cerevisiae (unicellular budding yeast)	12 million	**6,000**
Trichomonas vaginalis	160 million	**60,000**
Plasmodium falciparum (unicellular malaria parasite)	23 million	**5,000**
Caenorhabditis elegans (nematode)	95.5 million	**18,000**
Drosophila melanogaster (fruit fly)	170 million	**14,000**
Arabidopsis thaliana (mustard; thale cress)	125 million	**25,000**
Oryza sativa (rice)	470 million	**51,000**
Gallus gallus (chicken)	1 billion	**20,000-23,000**
Canis familiaris (domestic dog)	2.4 billion	**19,000**
Mus musculus (laboratory mouse)	2.5 billion	**30,000**
Homo sapiens (human)		

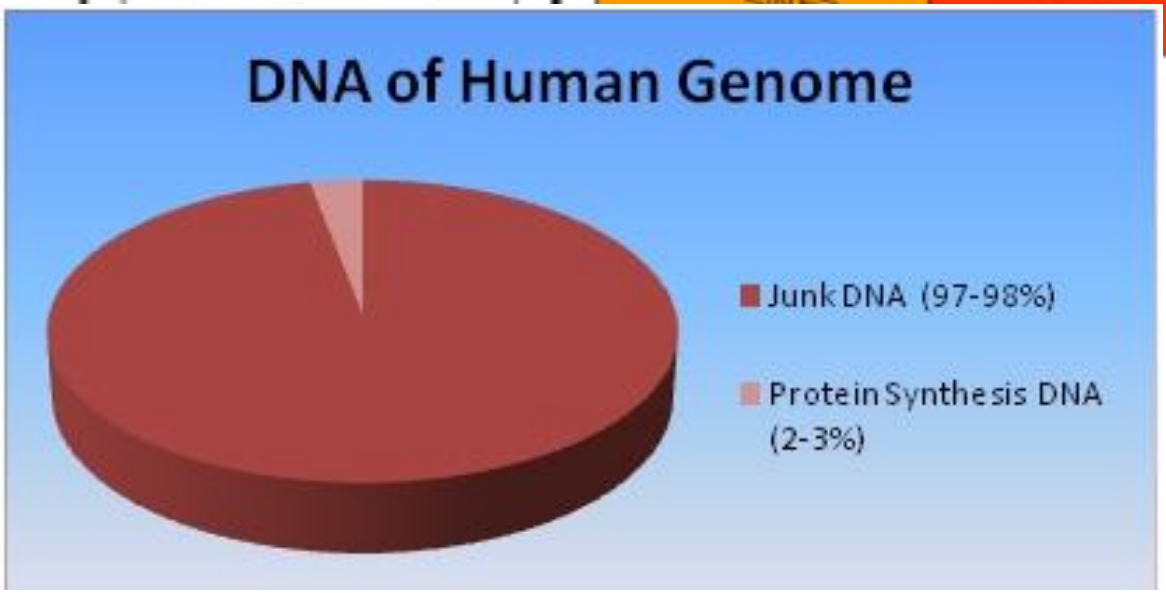
Plants and amphibians with huge genomes (not in table) do not have huge amounts of genes

Pray, L. (2008) Eukaryotic genome complexity. Nature Education 1(1)

Organization of the human genome







RNA

mRNA: messenger RNA – will be translated into protein

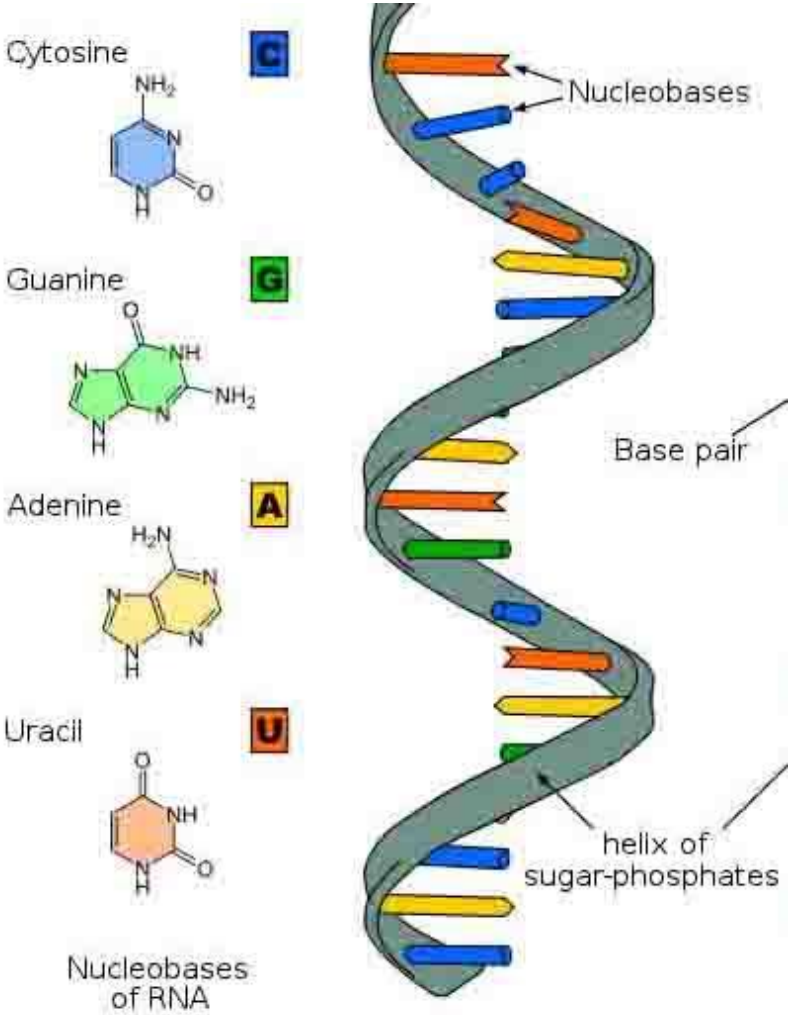
rRNA: ribosomal RNA – forms ribosomes

tRNA: transfer RNA – brings the right amino acids to the ribosomes

siRNA: silencer RNA – blocks specific mRNA

miRNA: micro RNA – regulative effect on specific translation of proteins

lncRNA: long non-coding RNA – regulative effect on specific DNA regions

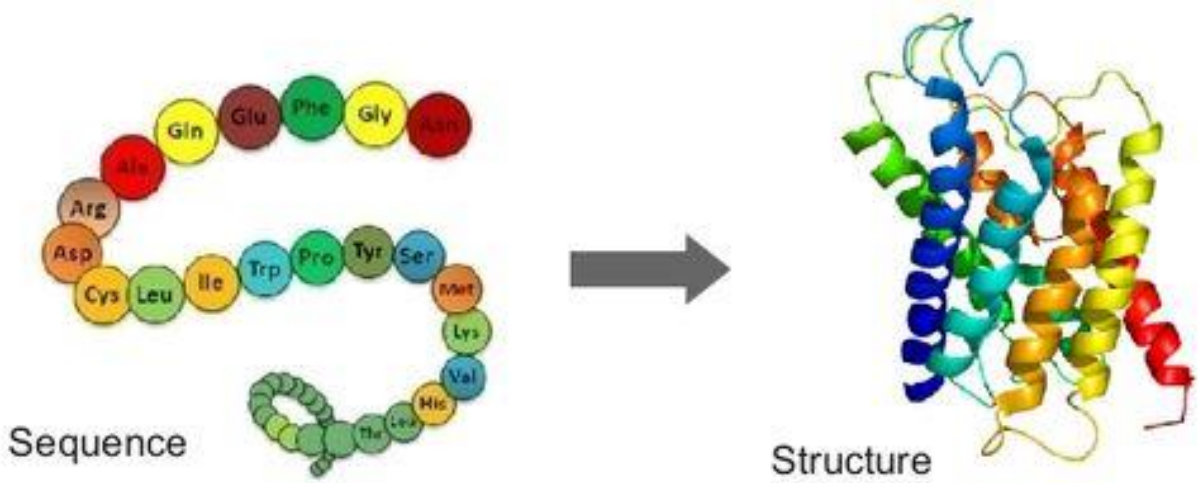


Protein

21 different amino acids

Protein function:

- Enzymes: catalysis of chemical reactions (alcohol dehydrogenase, lactase)
- Cell signaling: transmit signals within cells (antibodies, insulin)
- Structural proteins: form extracellular matrix (collagen) or generate mechanical force (myosin)
- https://www.youtube.com/watch?v=yKW4F0Nu-UY



2. Data and omics data

Single biomolecules

High throughput

DNA

⇩

RNA

⇩

PROTEIN

Sequencing and gene identification

Sequencing and gene expression

Identification and structure determination

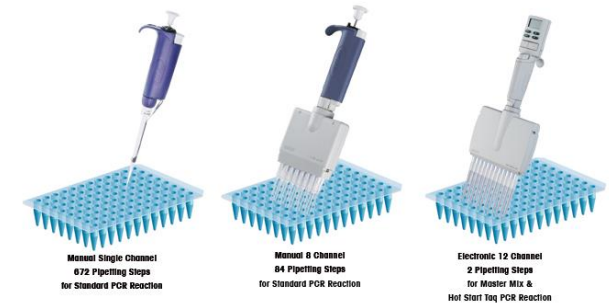
GENOME

⇩

TRANSCRIPTOME

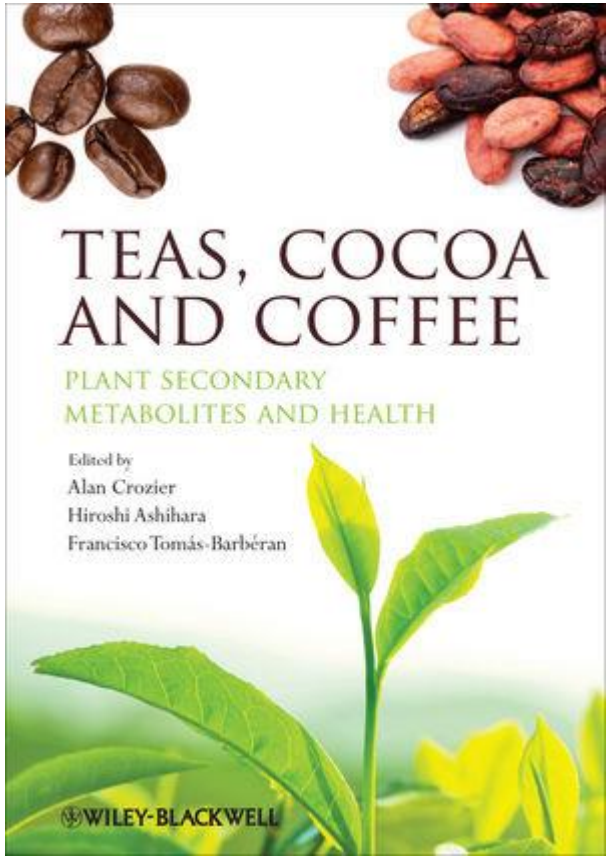
⇩

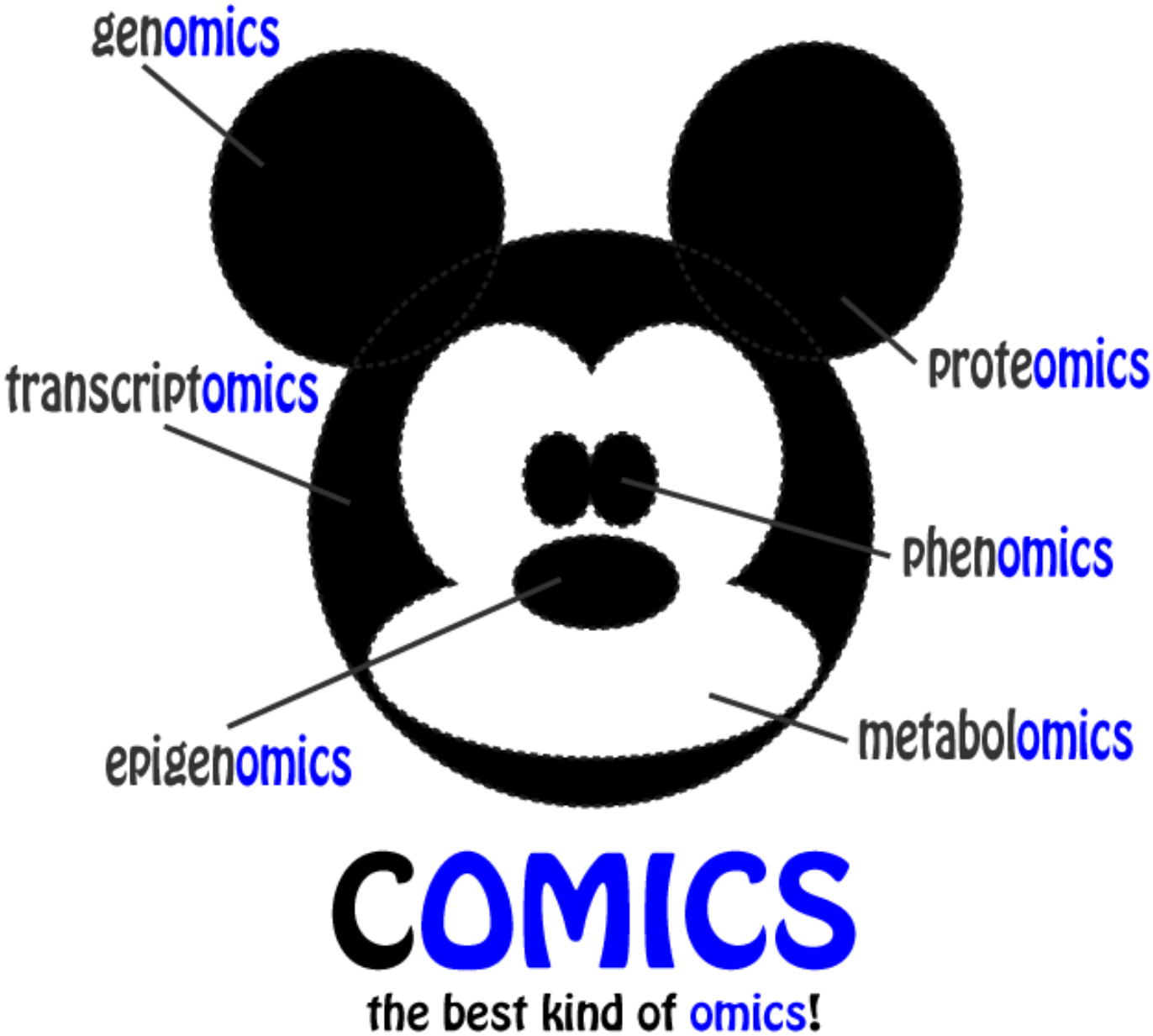
PROTEOME



The size of omics data

- 3,100,000,000 bp DNA per individual
- 22.000 (protein coding) genes
- 120.000 transcripts
- 100.000 proteins
- 40.000 metabolites





http://www.stathiskanterakis.com/?p=286

Research strategies

Hypothesis-driven research "reductionistic"

Hypothesis/Theory

Experiment

Falsification

Some hard thinking

New hypothesis

Data driven research "holistic"

Samples

Unbiased omics strategies

Data driven results

Data driven hypothesis/theory

Example: Nanomaterial toxicity assessment

Hypothesis – driven: Silver nanoparticles increases oxidative stress in Caco-2 cells

- In vitro assay – exposure scenario
- Oxidative stress assessment
  - Chance in SOD protein expression
  - Level of ROS
- Does the results confirm the hypothesis?

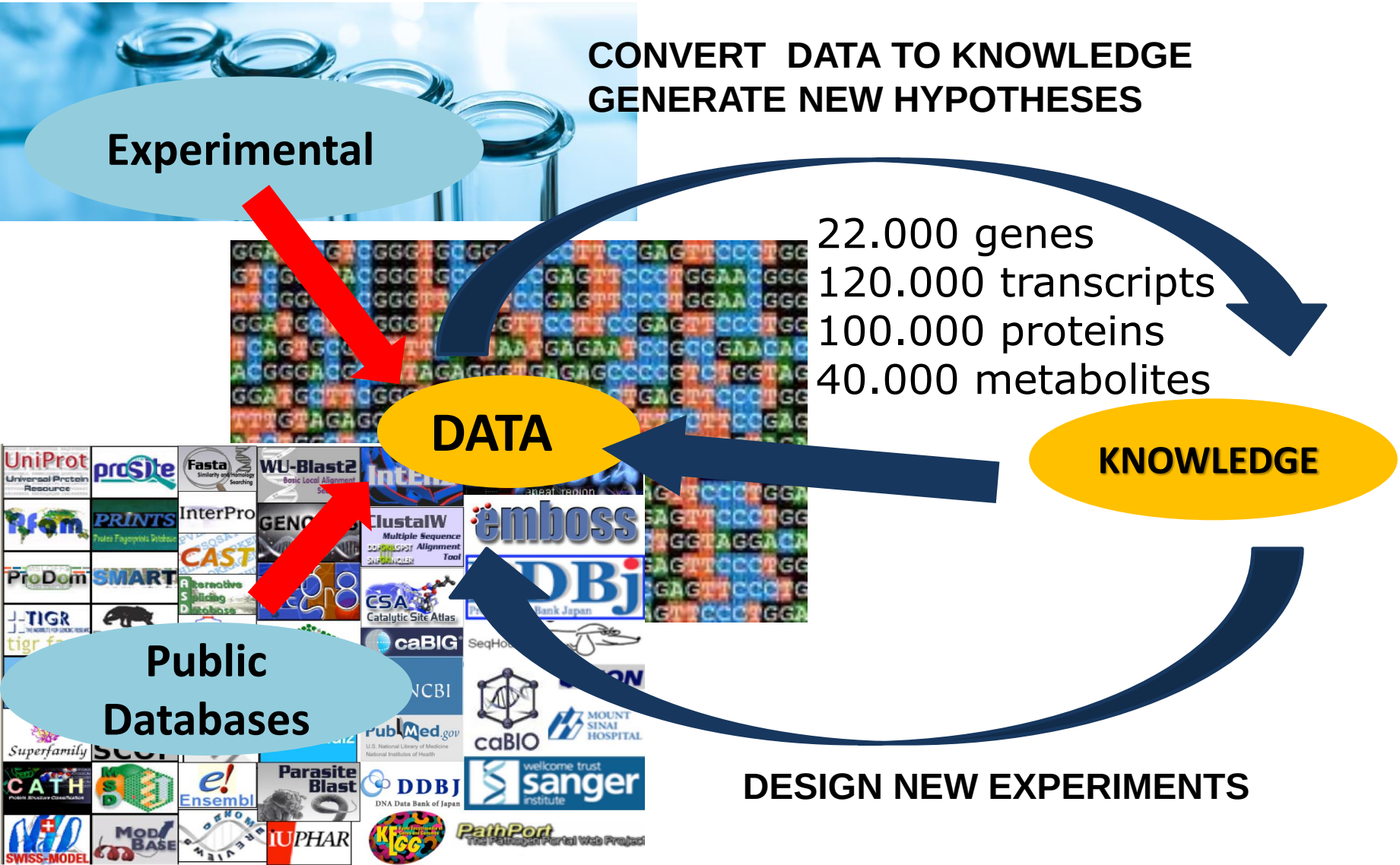
Data – driven: What happens to the transcriptome of Caco-2 cells after exposure to silver nanoparticles

- In vitro assay – exposure scenario
- Collect transcriptome (RNA) and run e.g. RNA-sequencing or microarray analysis
- Data analysis
- Result: list of changed gene expression
- Interpretation

3. What is Bioinformatics?

Bioinformatics uses **"informatics" techniques** (from applied math, computer science, statistics, etc.) to **understand** and **organize** biological information, like genes, proteins and molecules on a **large-scale.**

Why Bioinformatics?

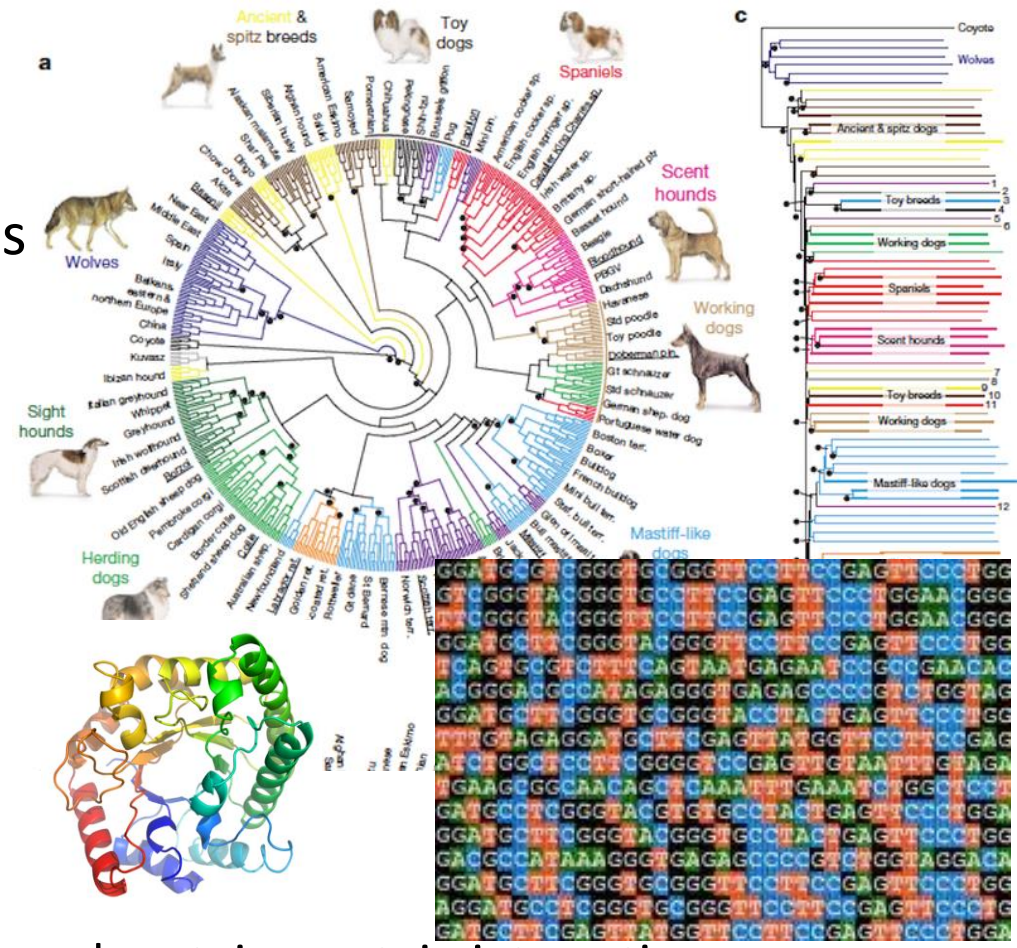


Bioinformatics tools and methods

- pattern recognition
- data mining
- machine learning algorithms
- visualization

Examples:

- sequence alignment (BLAST)
- gene finding
- genome assembly
- drug design, drug discovery
- protein structure alignment
- protein structure prediction
- prediction of gene expression and protein–protein interactions
- genome-wide association studies
- the modeling of evolution and cell division/mitosis.



Biological databases

1. Global nucleotide/protein sequence storage databases:

- GenBank of NCBI (National Center for Biotechnology Information)
- The European Molecular Biology Laboratory (EMBL) Ensembl database
- The DNA Data Bank of Japan (DDBJ)

2. Genome-centered databases

- NCBI genomes
- Ensembl Genome Browser
- UCSC Genome Bioinformatics Site

3. Protein Databases

- UniProt

4. Metabolite Databases

- HMDB, ChEBI

5. Interaction Databases

- Pathways: WikiPathways, KEGG, Reactome

6. Nanomaterial Databases

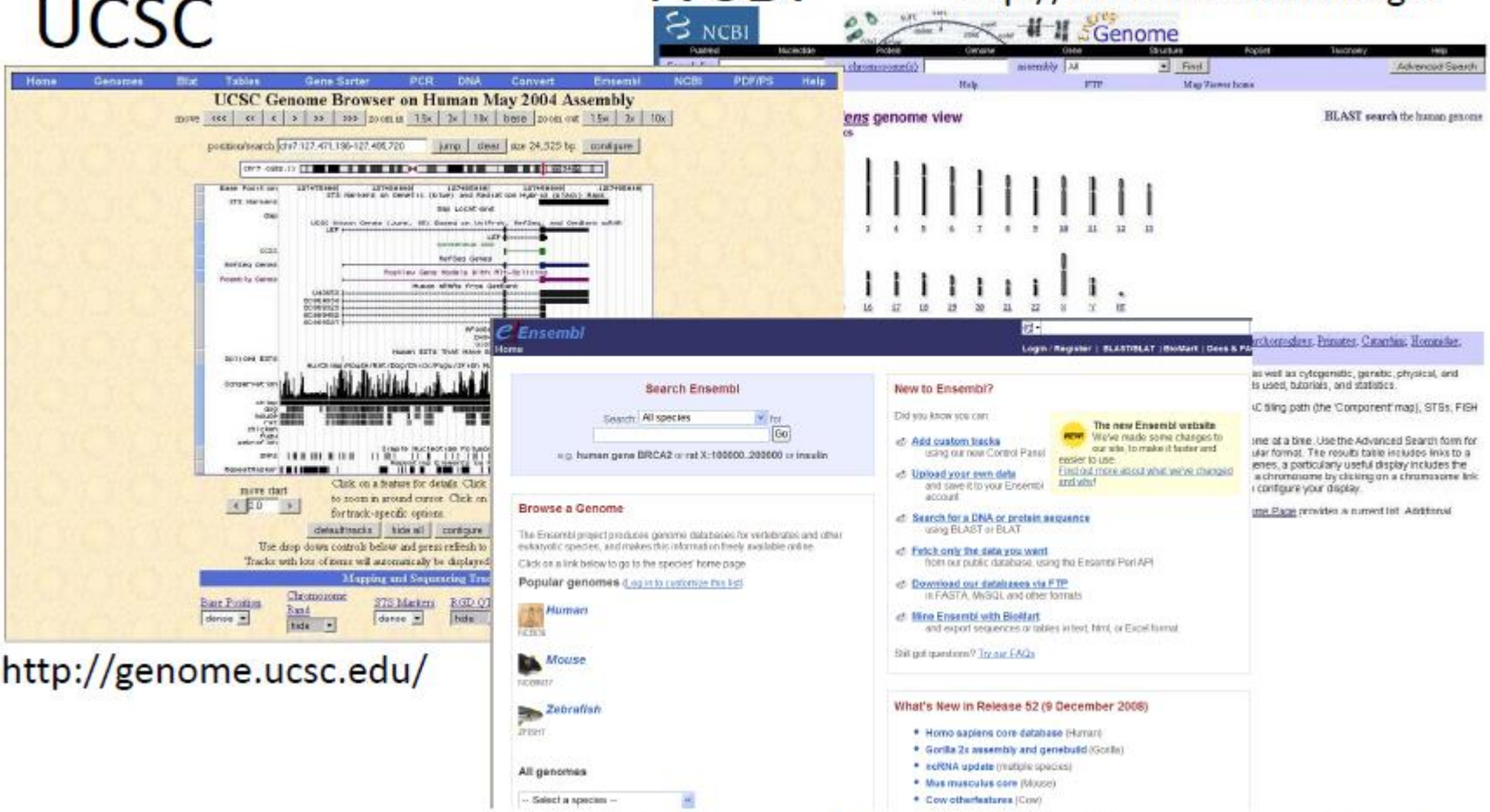
- eNanoMapper, Nanowerk, nature.nano

Genome centered databases

UCSC

NCBI

http://www.ncbi.nlm.nih.gov

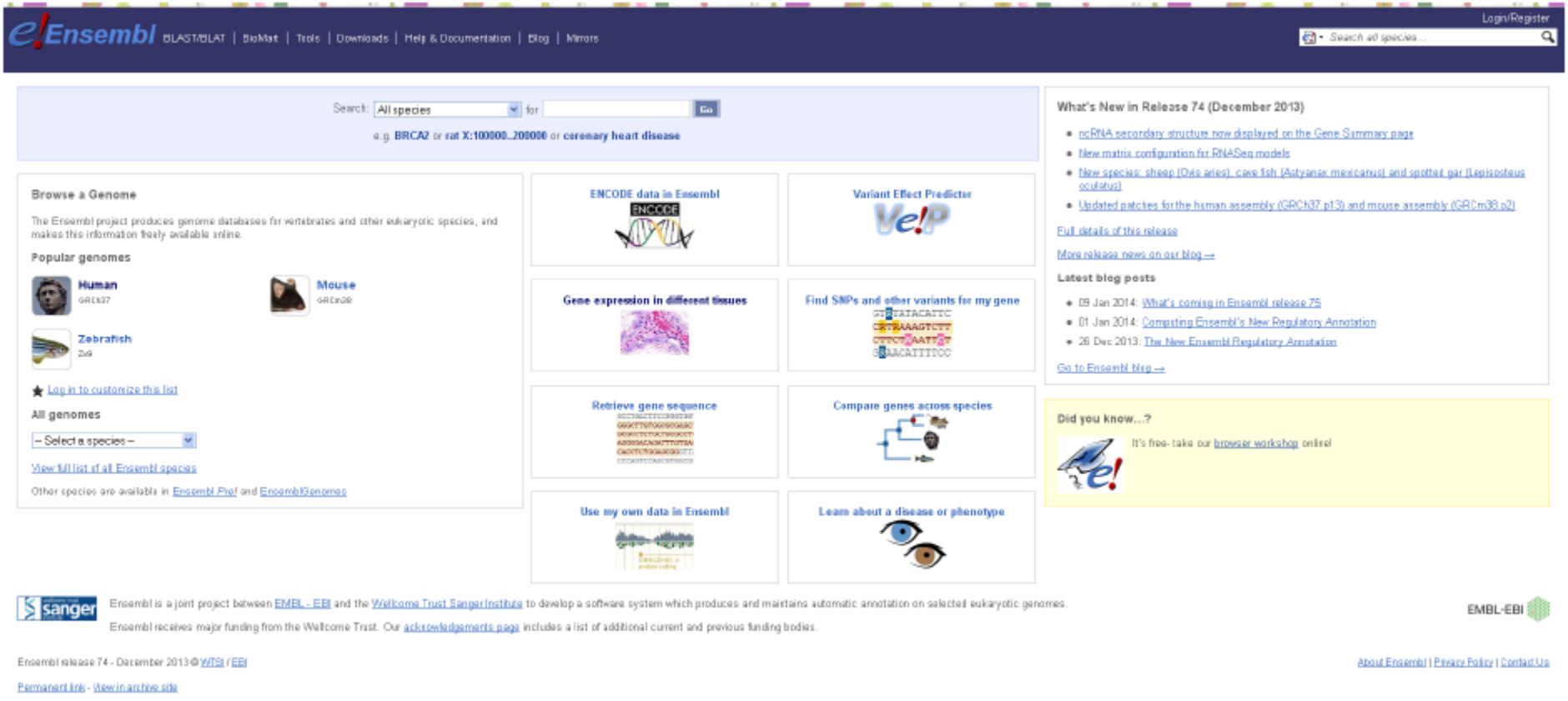


http://genome.ucsc.edu/

Ensembl

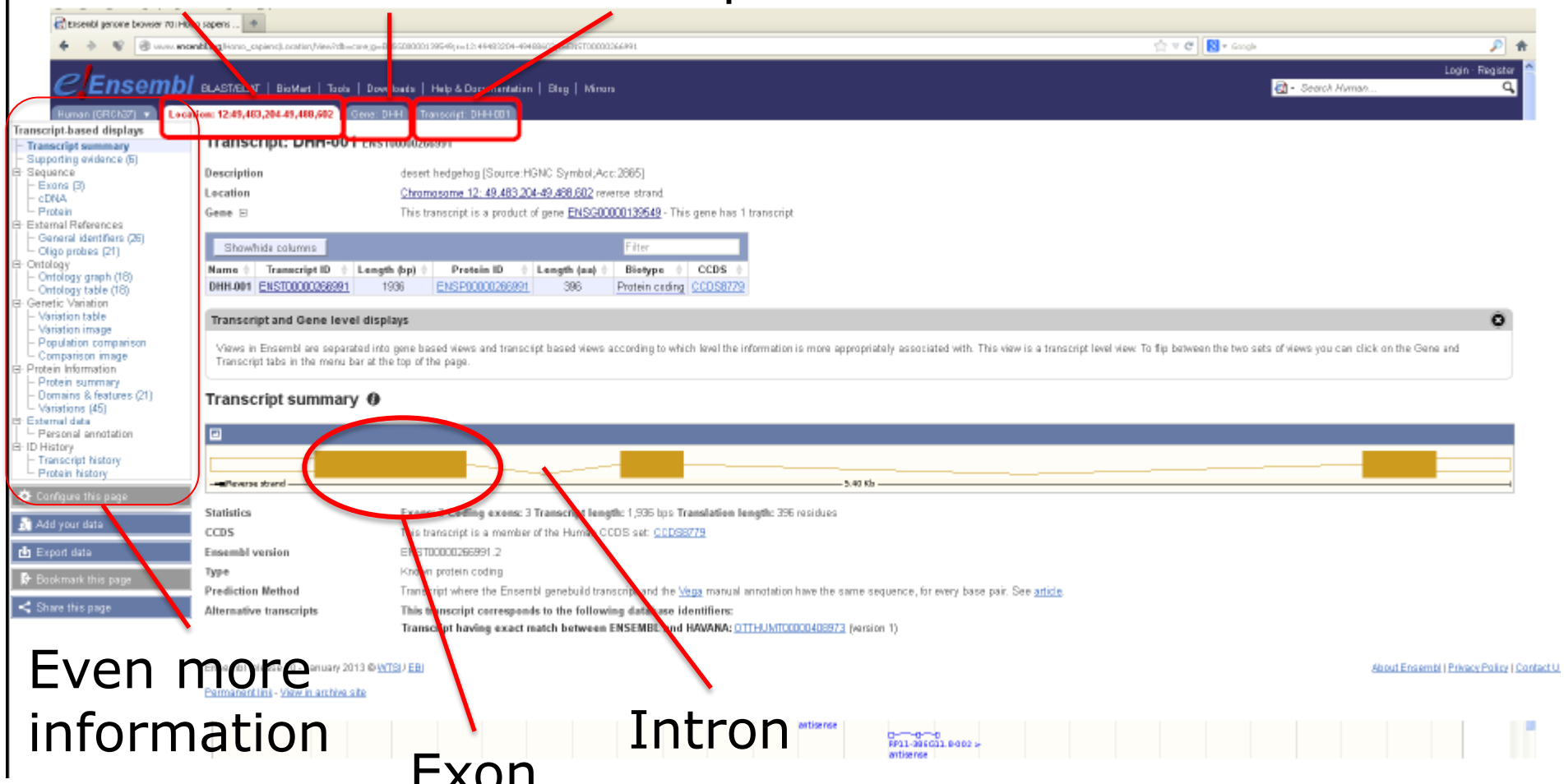
http://www.ensembl.org/

Ensembl


Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Search: All species for Go
e.g. BRCA2 or rat X:100000..200000 or coronary heart disease
Browse a Genome
The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.
Popular genomes
Human
GRCh37
Mouse
GRCm38
Zebrafish
Zv9
Log in to customize this list
All genomes
– Select a species –
View full list of all Ensembl species
Other species are available in Ensembl Pre! and EnsemblGenomes
ENCODE data in Ensembl
Variant Effect Predictor
Gene expression in different tissues
Find SNPs and other variants for my gene
Retrieve gene sequence
Compare genes across species
Use my own data in Ensembl
Learn about a disease or phenotype
What's New in Release 74 (December 2013)
ncRNA secondary structure now displayed on the Gene Summary page
New matrix configuration for RNASeq models
New species: sheep (Ovis aries), cave fish (Astyanax mexicanus) and spotted gar (Lepisosteus oculatus)
Updated patches for the human assembly (GRCh37.p13) and mouse assembly (GRCm38.p2)
Full details of this release
More release news on our blog →
Latest blog posts
09 Jan 2014: What's coming in Ensembl release 75
01 Jan 2014: Computing Ensembl's New Regulatory Annotation
26 Dec 2013: The New Ensembl Regulatory Annotation
Go to Ensembl blog →
Did you know...?
It's free- take our browser workshop online!
sanger
Ensembl is a joint project between EMBL - EBI and the Wellcome Trust Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes.
Ensembl receives major funding from the Wellcome Trust. Our acknowledgements page includes a list of additional current and previous funding bodies.
EMBL-EBI
Ensembl release 74 - December 2013 © WTSI / EBI
Permanent link - View in archive site
About Ensembl | Privacy Policy | Contact Us

Example: DHH

Location Gene Transcript



Even more information

Exon

Intron

Where does all this information come from?

- Submissions (e.g. Sequences)
- Literature
- Curators and contributors
- Automated generation by computer tools
- High-throughput lab screenings
- Individual contributions and large scale contributions

Example: DHH

Ensembl genome browser 70 | Homo sapiens ...

www.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000139549;r=12:49483204-49488602;t=ENST00000266991

Ensembl | BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors | Login · Register | Search Human...

Human (GRCh37) | Location: 12:49,483,204-49,488,602 | Gene: DHH | Transcript: DHH-001

Transcript-based displays
- Transcript summary
- Supporting evidence (6)
- Sequence
  - Exons (3)
  - cDNA
  - Protein
- External References
  - General identifiers (26)
  - Oligo probes (21)
- Ontology
  - Ontology graph (16)
  - Ontology table (16)
- Genetic Variation
  - Variation table
  - Variation image
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (21)
  - Variations (45)
- External data
  - Personal annotation
- ID History
  - Transcript history
  - Protein history

Configure this page | Add your data | Export data | Bookmark this page | Share this page

Transcript: DHH-001 ENST00000266991

Description desert hedgehog [Source:HGNC Symbol;Acc:2865]

Location Chromosome 12: 49,483,204-49,488,602 reverse strand

Gene This transcript is a product of gene ENSG00000139549 - This gene has 1 transcript

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
DHH-001	ENST00000266991	1936	ENSP00000266991	396	Protein coding	CCDS8779

← **Identifiers**

Transcript and Gene level displays

Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated with. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Transcript summary

Reverse strand — 5.40 Kb

Statistics **Exons:** 3 **Coding exons:** 3 **Transcript length:** 1,936 bps **Translation length:** 396 residues

CCDS This transcript is a member of the Human CCDS set: CCDS8779

Ensembl version ENST00000266991.2

Type Known protein coding

Prediction Method Transcript where the Ensembl genebuild transcript and the Vega manual annotation have the same sequence, for every base pair. See article

Alternative transcripts **This transcript corresponds to the following database identifiers:**
Transcript having exact match between ENSEMBL and HAVANA: OTTHUMT00000408973 (version 1)

Ensembl release 70 - January 2013 © WTSI / EBI | About Ensembl | Privacy Policy | Contact Us

Permanent link - View in archive site

antisense | RP11-395G11.8-002 > antisense

Learn more in the practical!

Unique identifiers – a game of names

- RefSeq:
  - Chromosome: NC_
  - mRNA: NM_
  - Protein: NP_
- Genbank:
  - Many types of IDs
- NCBI gene ID:
  - Number
- OMIM ID:
  - Number
- Pubmed ID:
  - Number
- No common identifier for nanoparticles yet

Ensembl identifiers

ENSG### Ensembl **Gene** ID

ENST### Ensembl **Transcript** ID

ENSP### Ensembl **Peptide** ID

ENSE### Ensembl **Exon** ID

For other species than human a suffix is added:

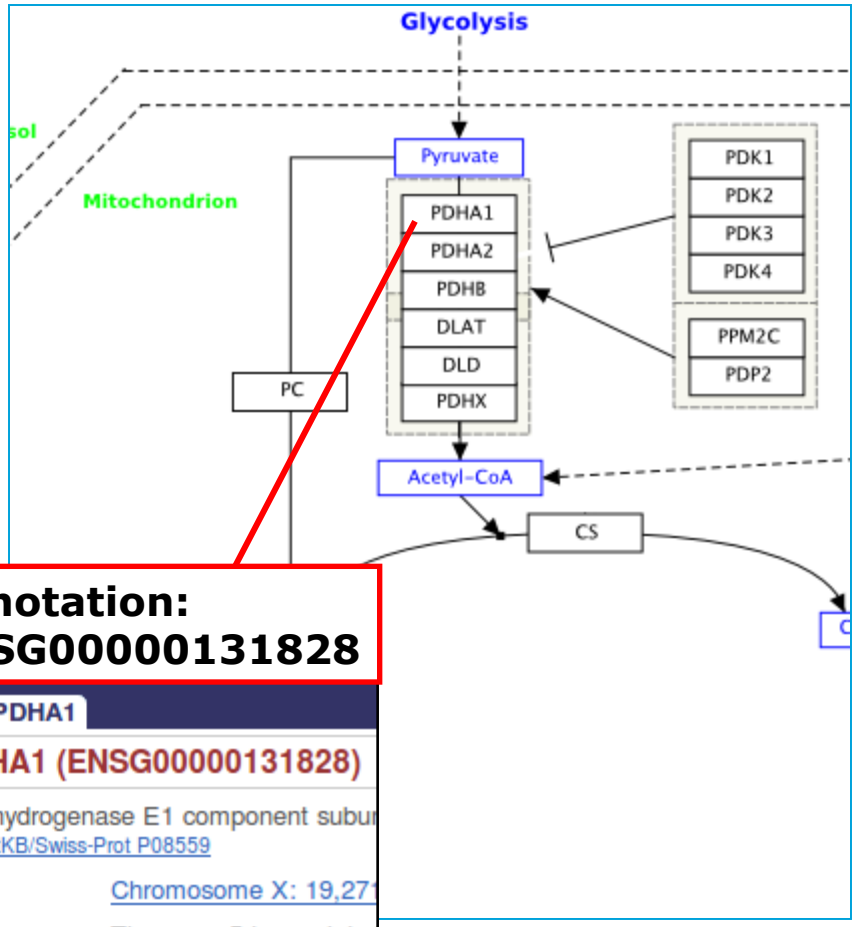
MUS (*Mus musculus*) for mouse: ENSMUSG###

DAR (*Danio rerio*) for zebrafish: ENSDARG###, etc.

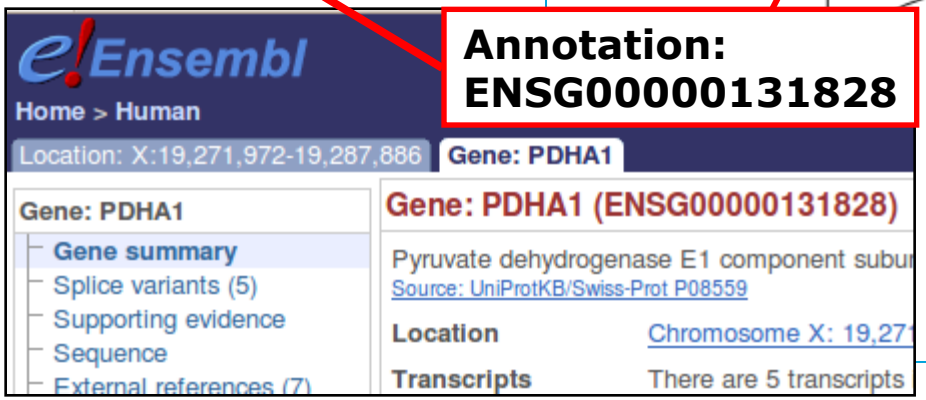
Identifier Mapping

	A	B	C	D	E	F	G
1	Probeset	GSM143409	GSM143411	GSM143412	GSM143413	GSM143414	GSM143415
2	1415670_at	208.9	171.3	186.1	179.7	226	2
3	1415671_at	330.7	281.9	301.1	355.4	300.9	354
4	1415672_at	488.5	453.5	474.8	477.8	477.5	460
5	1415673_at	90.1	103.2	64.3	90.8	96.1	52
6	1415674_a_at	167.5	187.9	168.3	174.8	148.5	210
7	1415675_at	70.1	96.6	81.1	66	70.6	75
8	1415676_a_at	1142.2				.2	1238

Identifier Mapping



Annotation: ENSG00000131828



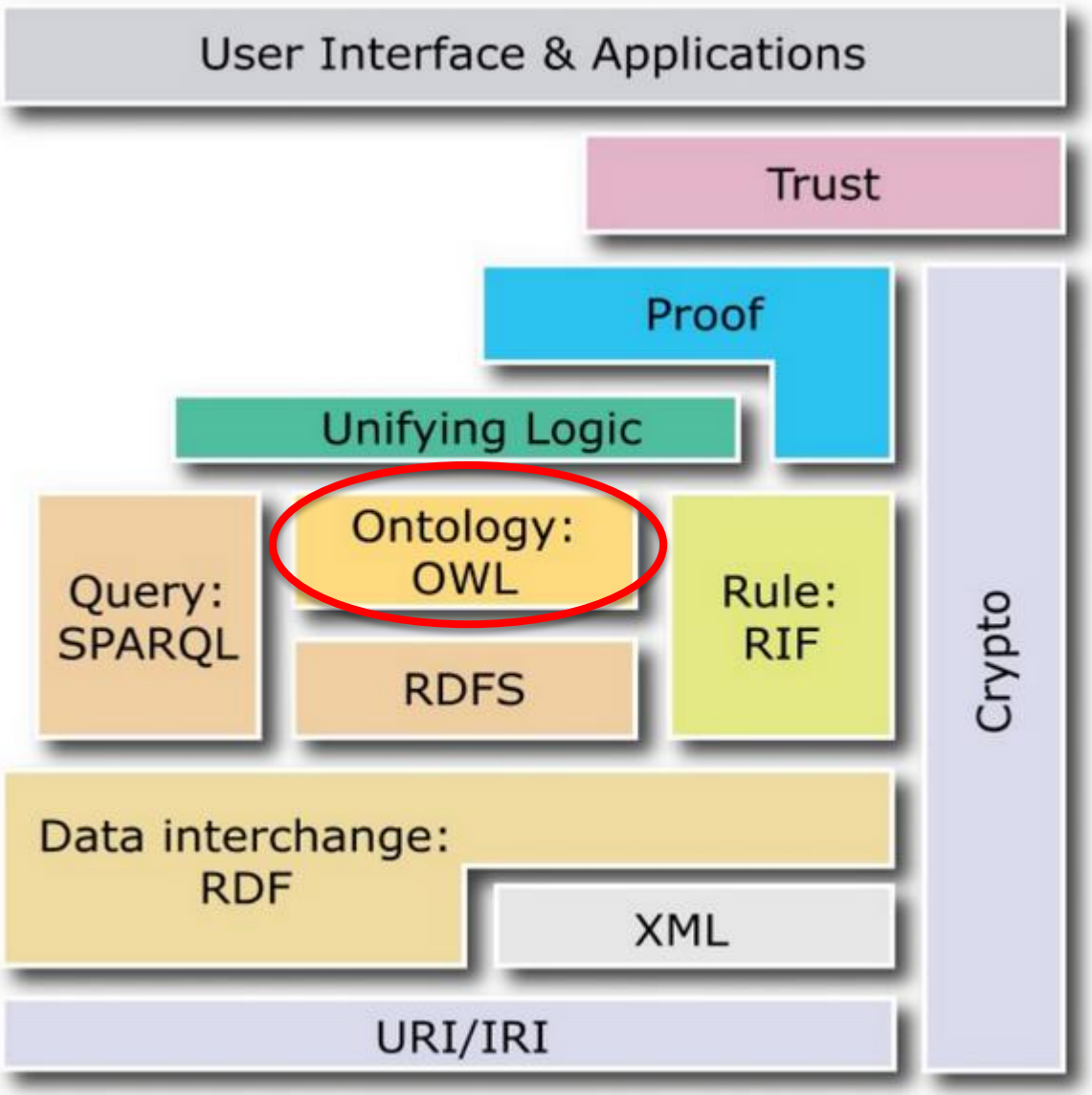
Mapping database: BridgeDb.org

Data integration by ontology

Ontology:

- Controlled language
- Relationship between terms
  - Hierarchy
  - Is_a/has_a
- Machine readable (OWL)

- Repositories
  - AberOwl
  - Bioportal
  - OLS (EBI ontology lookup service)



Anatomy of a GO term

id: GO:0006094	unique GO ID
name: gluconeogenesis	term name
namespace: process	ontology
def: The formation of glucose from noncarbohydrate precursors, such as pyruvate, amino acids and glycerol. [http://cancerweb.ncl.ac.uk/omd/index.html]	definition
exact_synonym: glucose biosynthesis	synonym
xref_analog: MetaCyc:GLUCONEO-PWY	database ref
is_a: GO:0006006 (glucose metabolic process)	parentage
is_a: GO:0006092 (main pathway of carbohydrate metabolism)	

The 3 Gene Ontologies

- **Molecular Function** = elemental activity/task
  - the tasks performed by individual gene products; examples are ***carbohydrate binding*** and ***ATPase activity***
- **Biological Process** = biological goal or objective
  - broad biological goals, such as ***mitosis*** or ***purine metabolism***, that are accomplished by ordered assemblies of molecular functions
- **Cellular Component** = location or complex
  - subcellular structures, locations, and macromolecular complexes; examples include ***nucleus***, ***telomere***, and ***RNA polymerase II holoenzyme***

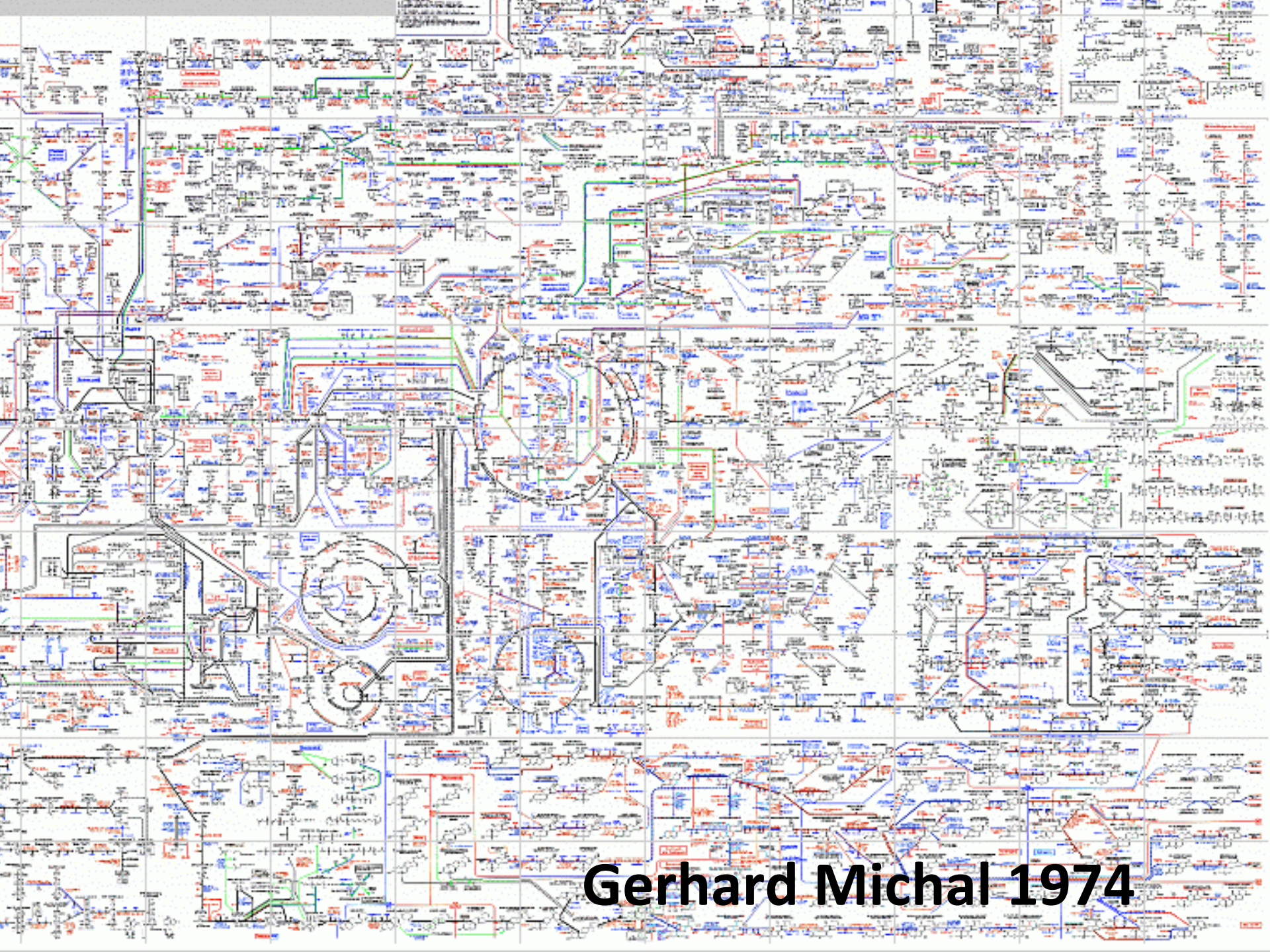
Searching and Browsing GO

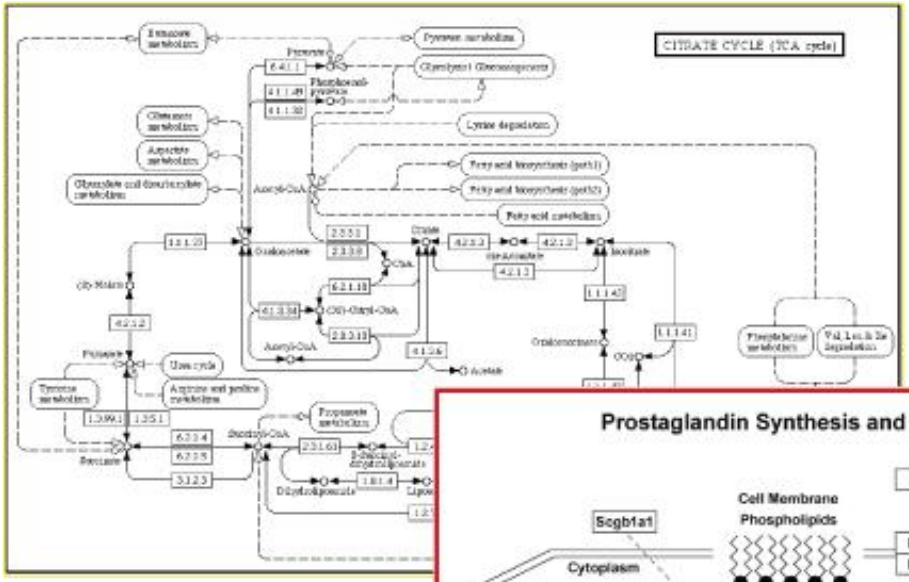
- AmiGO
  - http://www.godatabase.org
- Downloads
  - http://www.godatabase.org/dev/database/
  - XML or as a MySQL database dump
- http://www.geneontology.org/GO.tools.annotation.shtml
  - Annotate gene by sequence similarity.

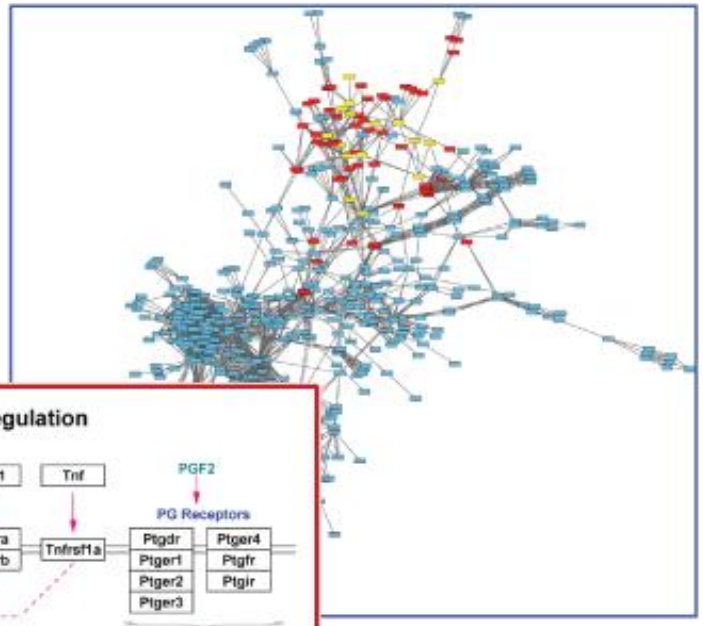
Back to databases…

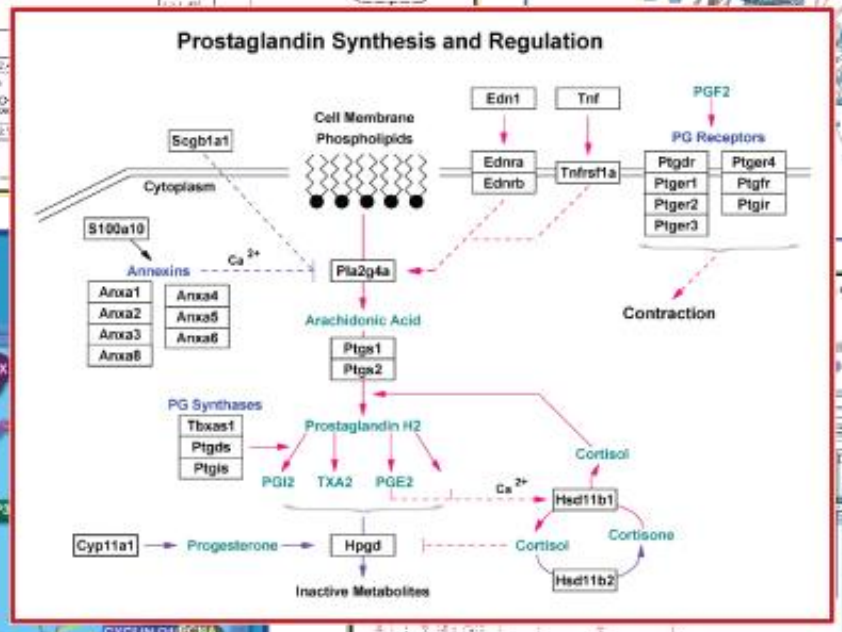
Single entity + interaction

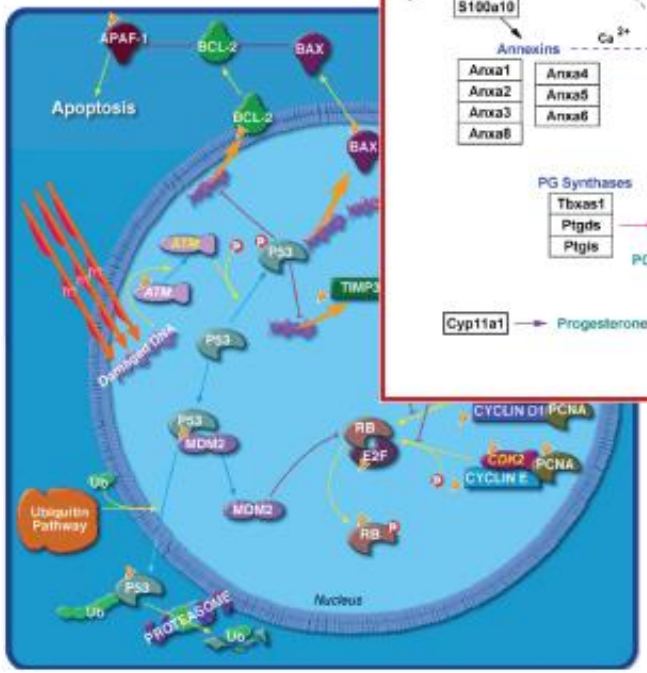
= biological pathways!

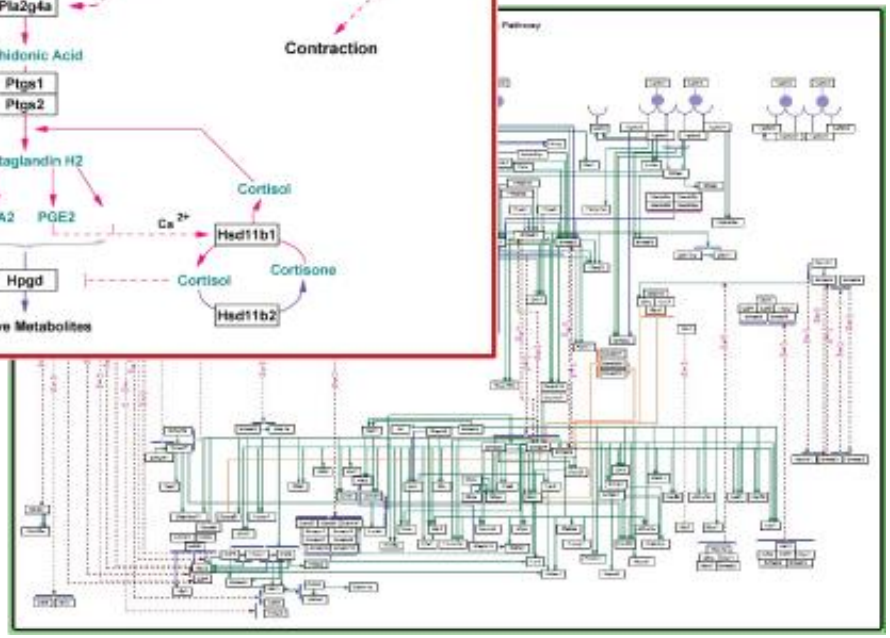

Gerhard Michal 1974


CITRATE CYCLE (TCA cycle)



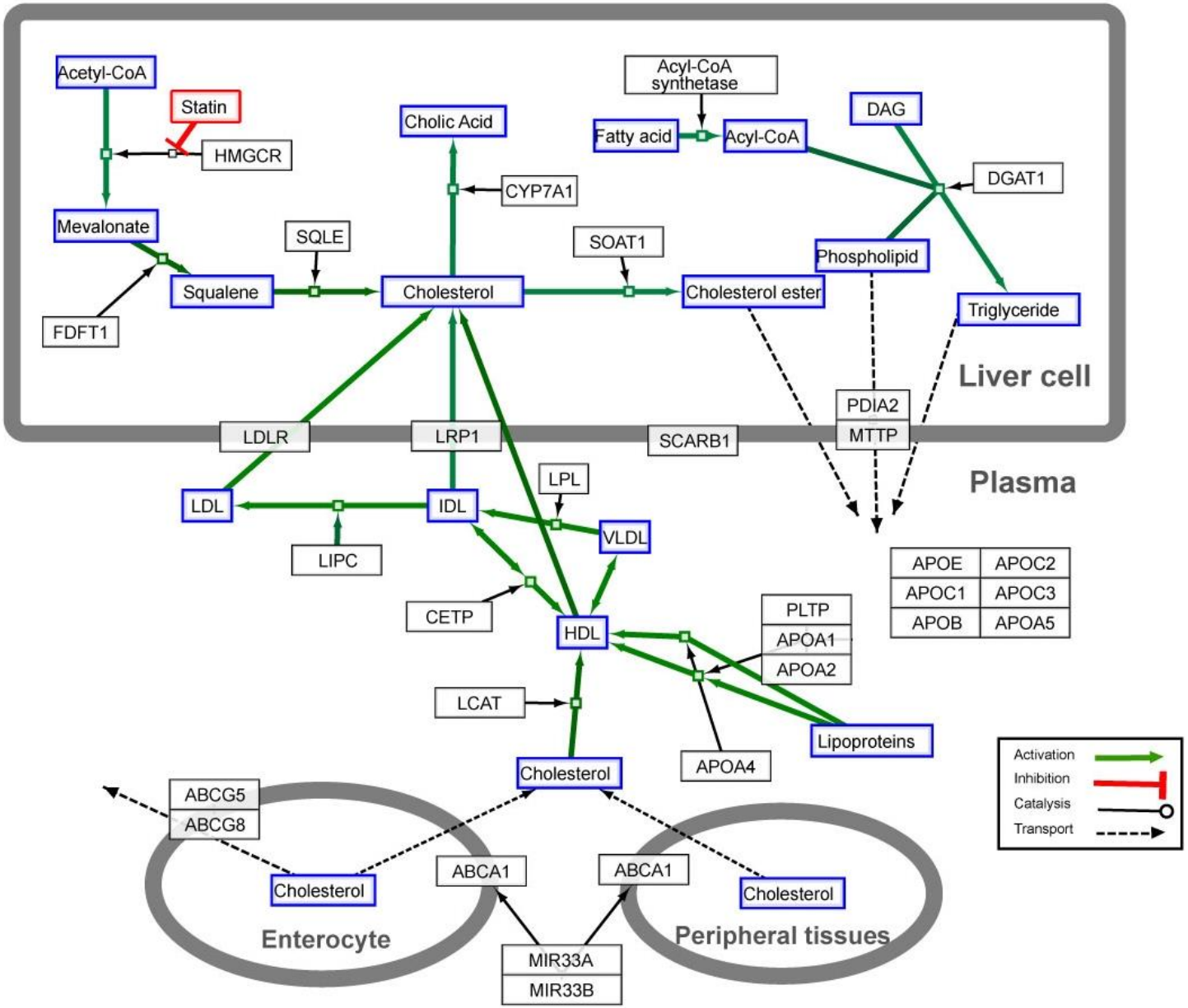

Prostaglandin Synthesis and Regulation
Cell Membrane
Phospholipids
Cytoplasm
Edn1
Tnf
PGF2
PG Receptors
Scgb1a1
Ednra
Ednrb
Tnfrsf1a
Ptgdr
Ptger1
Ptger2
Ptger3
Ptger4
Ptgfr
Ptgir
S100a10
Annexins
Anxa1
Anxa2
Anxa3
Anxa8
Anxa4
Anxa5
Anxa6
Pla2g4a
Arachidonic Acid
Contraction
Ptgs1
Ptgs2
PG Synthases
Tbxas1
Ptgds
Ptgis
Prostaglandin H2
PGI2
TXA2
PGE2
Cortisol
Hsd11b1
Cortisone
Hsd11b2
Cyp11a1
Progesterone
Hpgd
Inactive Metabolites

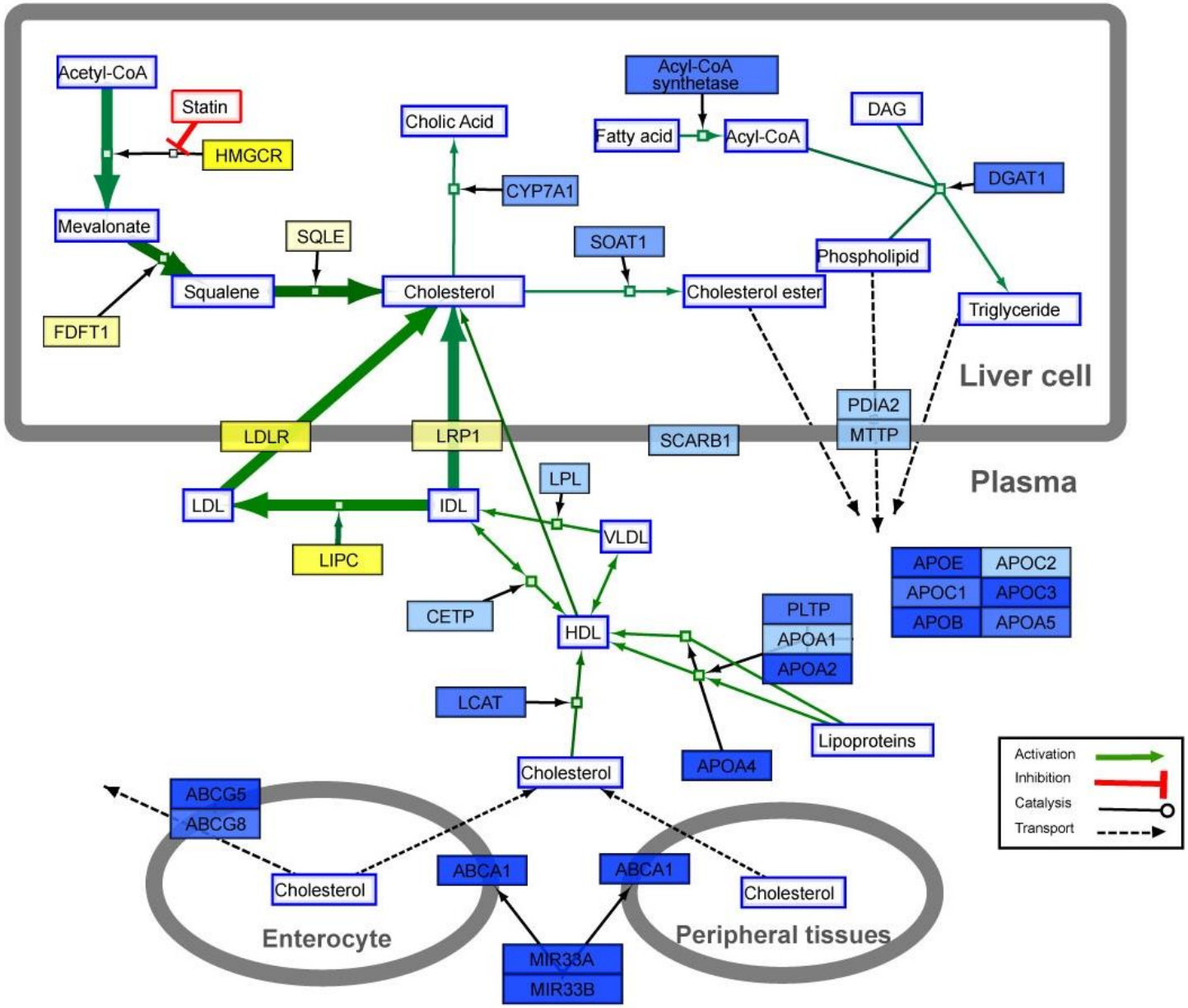

APAF-1
BCL-2
BAX
Apoptosis
ATM
P53
TIMP3
Damaged DNA
MDM2
RB
E2F
CYCLIN D1
PCNA
CDK2
CYCLIN E
Ubiquitin Pathway
Ub
PROTEASOME
Nucleus

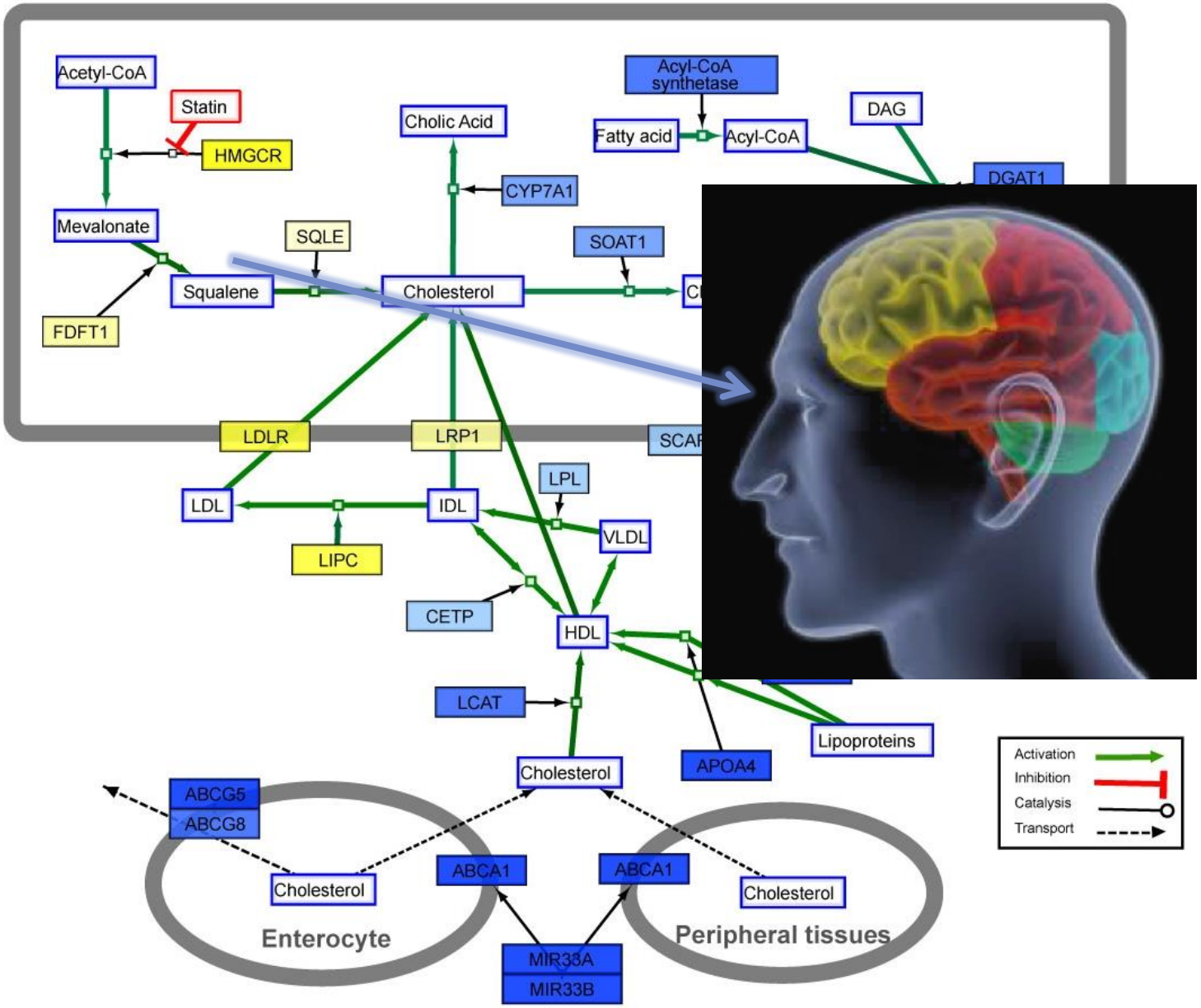


WikiPathways and How to Change the World

(or at least a small corner of the world)


Acetyl-CoA
Statin
HMGCR
Mevalonate
FDFT1
Squalene
SQLE
Cholesterol
Cholic Acid
CYP7A1
SOAT1
Cholesterol ester
Acyl-CoA synthetase
Fatty acid
Acyl-CoA
DAG
DGAT1
Phospholipid
Triglyceride
Liver cell
PDIA2
MTTP
SCARB1
LDLR
LRP1
Plasma
LPL
LDL
IDL
VLDL
LIPC
CETP
HDL
PLTP
APOA1
APOA2
APOA4
LCAT
Lipoproteins
APOE
APOC2
APOC1
APOC3
APOB
APOA5
Activation
Inhibition
Catalysis
Transport
Cholesterol
ABCG5
ABCG8
Cholesterol
Enterocyte
ABCA1
ABCA1
Cholesterol
Peripheral tissues
MIR33A
MIR33B


Acetyl-CoA
Statin
HMGCR
Mevalonate
FDFT1
SQLE
Squalene
Cholesterol
Cholic Acid
CYP7A1
SOAT1
Cholesterol ester
Acyl-CoA synthetase
Fatty acid
Acyl-CoA
DAG
DGAT1
Phospholipid
Triglyceride
Liver cell
PDIA2
MTTP
SCARB1
Plasma
LDLR
LRP1
LDL
IDL
LIPC
LPL
VLDL
CETP
HDL
PLTP
APOA1
APOA2
APOA4
LCAT
Lipoproteins
APOE
APOC2
APOC1
APOC3
APOB
APOA5
Cholesterol
ABCG5
ABCG8
Cholesterol
ABCA1
ABCA1
Cholesterol
Enterocyte
Peripheral tissues
MIR33A
MIR33B
Activation
Inhibition
Catalysis
Transport


Acetyl-CoA
Statin
HMGCR
Mevalonate
FDFT1
Squalene
SQLE
Cholesterol
Cholic Acid
CYP7A1
SOAT1
Acyl-CoA synthetase
Fatty acid
Acyl-CoA
DAG
DGAT1
LDLR
LRP1
LDL
IDL
LIPC
LPL
VLDL
CETP
HDL
LCAT
Lipoproteins
APOA4
Cholesterol
ABCG5
ABCG8
Cholesterol
ABCA1
ABCA1
Cholesterol
Enterocyte
Peripheral tissues
MIR33A
MIR33B
Activation
Inhibition
Catalysis
Transport

Nanomaterial database

How to represent nanomaterials in a database?

- Nanomaterials
  - Core
  - Coating(s)
  - Linkage
  - Impurities
  - Components, internal structure, etc.
- Typical assay description
  - Property – value (range of values) – units *(Excel templates)*
- More complex description:
  - Experimental graph *(ISA-TAB / ISA-TAB-nano)*

- Existing data models
  - *BioAssay Ontology*
  - *OECD Harmonized Templates*
  - *CoDATA UDS*
  - *ISA-TAB- Nano*
- Commonalities:
  - Materials sample
  - Protocols, protocol parameters
  - Experimental conditions
  - Readouts
    - Measurements,
    - Measurement groups,
    - Raw data, derived data

https://data.enanomapper.net/

https://apps.ideaconsult.net/enanomapper/

Search

Home | Search | Data collections | Data upload | For developers | Help

eNanoMapper prototype database

A substance database for nanomaterial safety information

free text search

silver

Search

Search by identifier | by physchem parameters or biological effects | by composition | by citation | **Browse** | **Upload**

Search ▾ Nanomaterials ▾ Help ▾ About ▾

Integrated view of eNanoMapper database [contributors] and caNanoLab

Search

Current Selection

(x) silver

(x) substanceType:NPO_1892

- Data sources
- Nanomaterial type
  - silver 27
- P-CHEM (27)
- TOX (82)
- Cell
- Species
- Results
- References
- Protocols
- Instruments

< 1 2 > displaying 1 to 20 of 27

Ag (Harper2011 9)
P-CHEM.Nanomaterial surface chemistry ATOMIC COMPOSITION = more

material composition study

Ag (Harper2011 8)
P-CHEM.Nanomaterial surface chemistry ATOMIC COMPOSITION = more

material composition study

Ag (Harper2011 7)
P-CHEM.Nanomaterial surface chemistry ATOMIC COMPOSITION = more

material composition study

Ag (Harper2011 6)
P-CHEM.Nanomaterial surface chemistry ATOMIC COMPOSITION = more

material composition study

Ag (Harper2011 5)
P-CHEM.Nanomaterial surface chemistry ATOMIC COMPOSITION = more

material composition study

Ag (Harper2011 4)

Search by phys-chem parameter or biol. effect

Search substances by endpoint data

Update results

P-Chem

- [ ] 4.1. Appearance (S) [1]
- [ ] 4.2. Melting point / freezing point (S) [5]
- [ ] 4.26. Nanomaterial crystallite and grain size (S) [105]
- [ ] 4.27. Nanomaterial aspect ratio/shape (S) [70]
- [ ] 4.28. Nanomaterial specific surface area (S) [81]
- [ ] 4.29. Nanomaterial zeta potential (S) [289]
- [ ] 4.3. Boiling point (S) [5]
- [ ] 4.30. Nanomaterial surface chemistry (S) [368]
- [ ] 4.31. Nanomaterial dustiness (S) [1]
- [x] 4.5. Particle size distribution (Granulometry) (S) [600]
- [ ] 4.99. Physico chemical properties (other) (S) [227]

Tox

Update results

Search substances by endpoint data

Update results

P-Chem

Tox

- [ ] 7.2.1. Acute toxicity - oral (S) [1]
- [ ] 7.6.1. Genetic toxicity in vitro (S) [1]
- [ ] 7.7. Carcinogenicity (S) [1]
- [ ] 7.99. Toxicity (other) (S) [244]
- [ ] 8.100. Proteomics (S) [121]
- [ ] BAO_0002167. Genotoxicity Assay (S) [12]
- [ ] BAO_0002993. Cytotoxicity Assay (S) [20]
- [ ] BAO_0003009. Cell Viability Assay (S) [290]
- [ ] BAO_0010001. ATP Assay (S) [56]
- [ ] NPO_1709. LDH Release Assay (S) [35]
- [ ] NPO_1911. MTT Assay (S) [16]

Update results

ENM - ontology

BioPortal Browse Search Mappings Recommender Annotator Resource Index Projects Sign In | Help | Fee

eNanoMapper

Summary Classes Properties Notes Mappings Widgets

Details

ACRONYM	ENM
VISIBILITY	Public
BIOPORTAL PURL	http://purl.bioontology.org/ontology/ENM
DESCRIPTION	The eNanoMapper ontology covers the full scope of terminology needed to support research into nanomaterial safety. It builds on multiple pre-existing external ontologies such as the NanoParticle Ontology.
STATUS	Alpha
FORMAT	OWL
CONTACT	Egon Willighagen, egon.willighagen@maastrichtuniversity.nl Friederike Ehrhart, friederike.ehrhart@maastrichtuniversity.nl Gareth Owen, gowen@ebi.ac.uk Linda Rieswijk, linda.rieswijk@maastrichtuniversity.nl Jiakang Chang, jkchang@ebi.ac.uk Janna Hastings, hastings@ebi.ac.uk
HOME PAGE	https://github.com/enanomapper/ontologies
PUBLICATIONS PAGE	http://enanomapper.net/library
DOCUMENTATION PAGE	http://enanomapper.net/ontology
CATEGORIES	Chemical, Health
GROUPS	

Reviews

Add your review

No reviews available.

Submissions

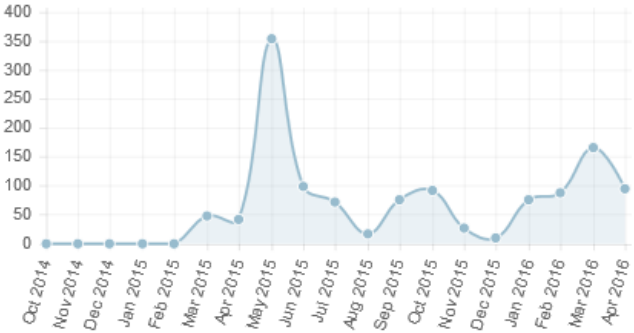
SUBMISSION	RELEASE DATE	UPLOAD DATE	DOWNLOADS
3.0 (Parsed, Indexed, Metrics, Annotator)	05/22/2016	05/22/2016	OWL | CSV | RDF/XML | Diff
3.0 (Archived)	03/01/2016	03/01/2016	OWL | Diff
2 (Archived)	01/29/2016	01/29/2016	OWL | Diff
2 (Archived)	01/28/2016	01/28/2016	OWL | Diff

Metrics

NUMBER OF CLASSES:	7937
NUMBER OF INDIVIDUALS:	196
NUMBER OF PROPERTIES:	2
MAXIMUM DEPTH:	10
MAXIMUM NUMBER OF CHILDREN:	1419
AVERAGE NUMBER OF CHILDREN:	6
CLASSES WITH A SINGLE CHILD:	441
CLASSES WITH MORE THAN 25 CHILDREN:	56
CLASSES WITH NO DEFINITION:	2756

Visits

Download as CSV



Nanoparticle

AberOWL Repository Browse

Back to Browse

AberOWL Repository Browse Ontologies » Add New Ontology Statistics Help Username Password

Back to Browse

nanoparticle

- entity
  - material entity
    - molecular entity
      - particle
        - primary particle
          - nanoparticle
        - secondary particle
    - biome
    - cell line
    - chemical substance
    - environmental material
    - fiat material part
    - instrument
    - organism
    - tissue
  - disposition
  - information content entity
  - process
  - quality

Object properties

ENM - eNanoMapper

Overview | Browse | DL Query | Visualise | PubMed | Data | SPARQL | Download

label	nanoparticle
oboid	NPO:707
SubClassOf:	primary particle
class	http://purl.bioontology.org/ontology/npo#NPO_70
http://purl.bioontology.org/ontology/npo#code	npo_707
http://purl.bioontology.org/ontology /npo#preferred_Name	nanoparticle
ontology	ENM
rdfs:comment	definition is partly based on astm e 2456-06 (termi nanotechnology).
type	class

Object properties

How to use this interactive machine readable biological data?

- Databases
  - Information (API – manually)
- Tools
  - Re-analysis
  - Modelling
  - Reference materials

4. How to **DO** data analysis?

Microarray experiments workflow

SETUP

↓

WET-LAB

↓

ANALYSIS

↓

PUBLICATION

E.g. Caco-2 cells exposed to silver nanoparticles

- Caco-2 small intestine cell line (human)
- Exposed to
  - 2.5 µg/ml Ag nanoparticles
  - 25 µg/ml Ag nanoparticles
  - 0.5 µg/ml $AgNO_3$ (soluble)
  - Control without exposure
- Data from GEO: **GSE62253**
- Publication:

http://informahealthcare.com/nan
ISSN: 1743-5390 (print), 1743-5404 (electronic)

Nanotoxicology

Nanotoxicology, 2015; 9(7): 852–860
© 2014 Informa UK Ltd. DOI: 10.3109/17435390.2014.980760

informa healthcare

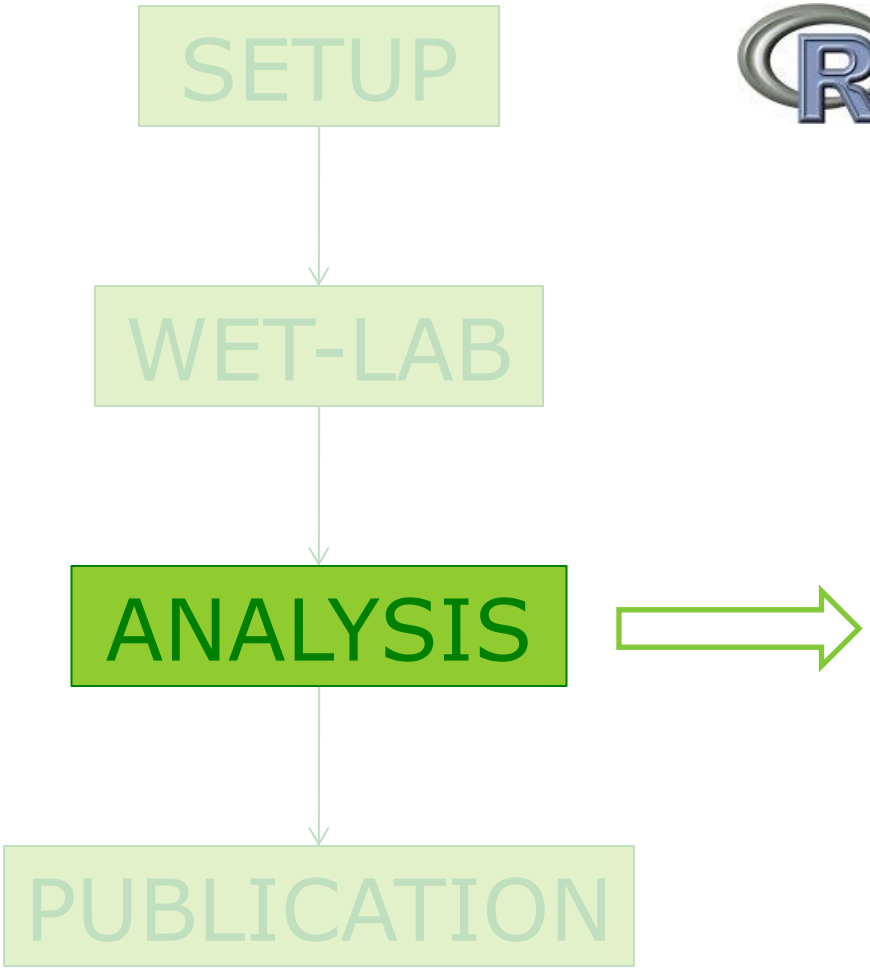
ORIGINAL ARTICLE

Molecular mechanism of silver nanoparticles in human intestinal cells

Linda Böhmert, Birgit Niemann, Dajana Lichtenstein, Sabine Juling, and Alfonso Lampen

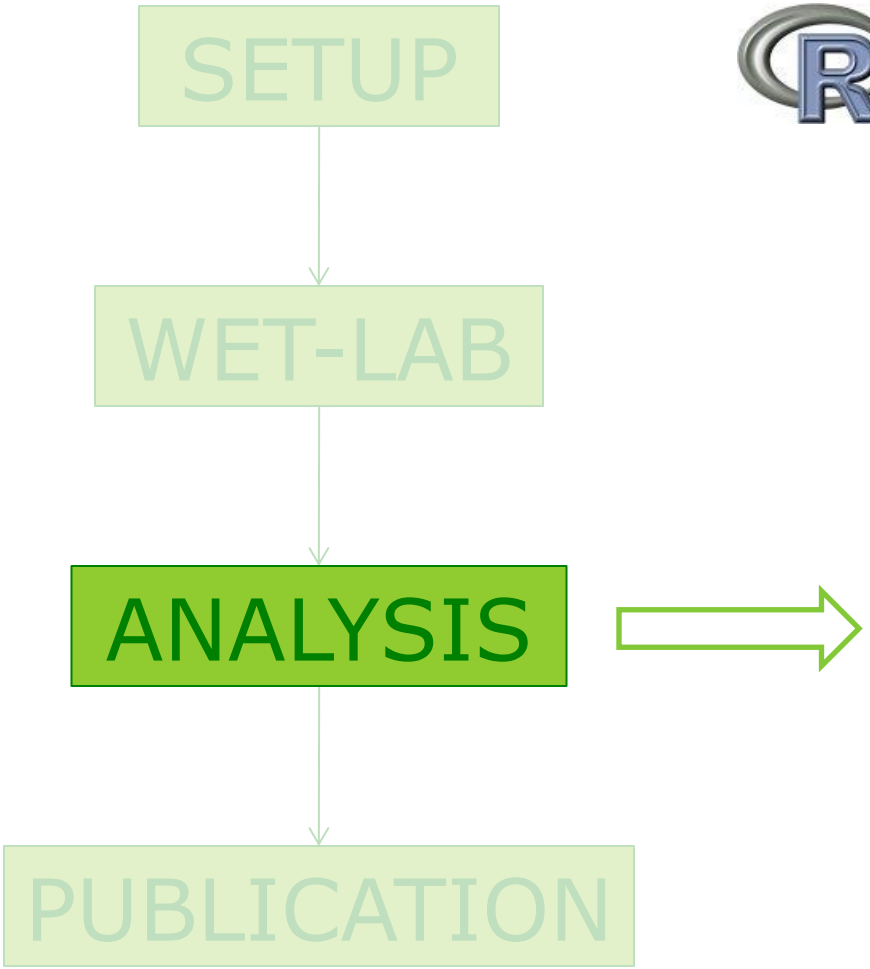
Department Food Safety, Federal Institute for Risk Assessment, Berlin, Germany

Microarray data analysis



- Image analysis
- Quality Control
- Pre-processing
  - Background correction
  - Normalisation
  - Filtering
  - Annotation

Microarray data analysis



- Statistical evaluation
  - T-test
  - ANOVA / modelling
- Further analysis
  - Significantly changed genes
  - LogFC (log2)

Microarray data analysis

SETUP

WET-LAB

ANALYSIS

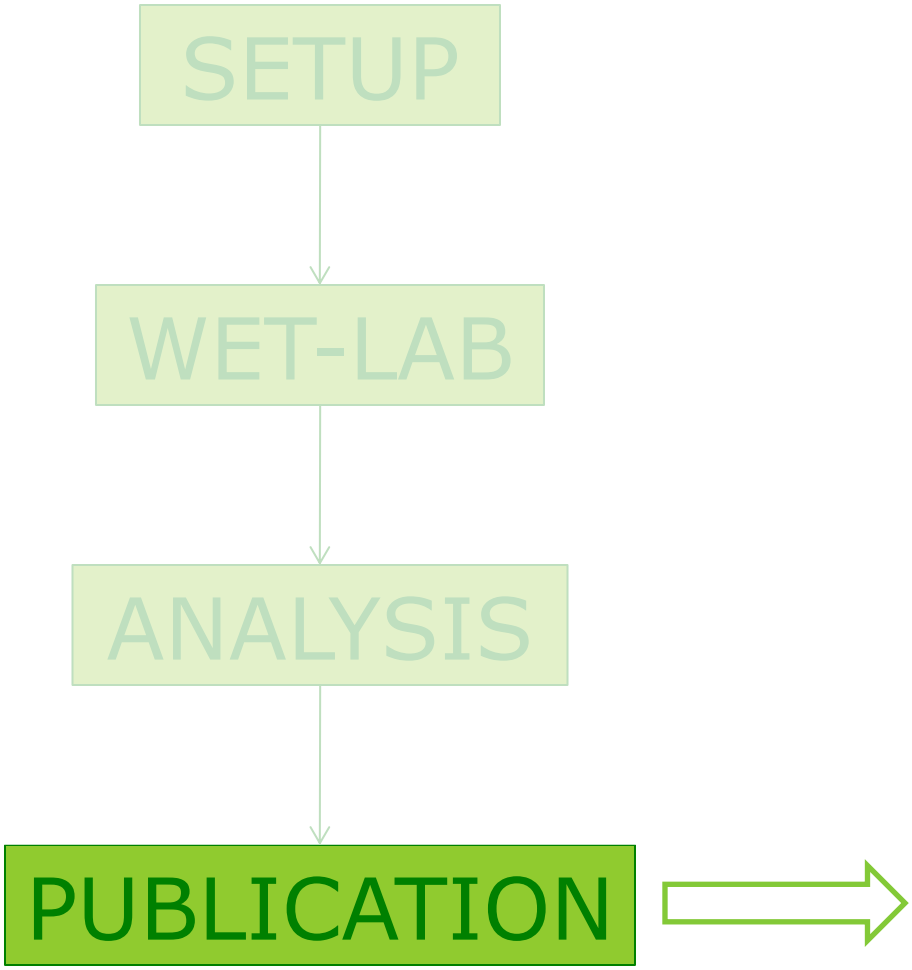
PUBLICATION

the Gene Ontology

GO-Elite

- Biological evaluation
  - Pathway analysis
  - Gene Ontology analysis
  - Network analysis
  - Etc…
- *Validation*
  - Technology
  - Biology / Literature

Microarray data analysis



- With publication of the paper, also the data has to be published: obligatory!
  - ArrayExpress at EBI
  - Gene Expression Omnibus (GEO) at NCBI
- Standard for publication: MIAME

Dual channel ('two colour') gene expression microarrays – 'spotted arrays'

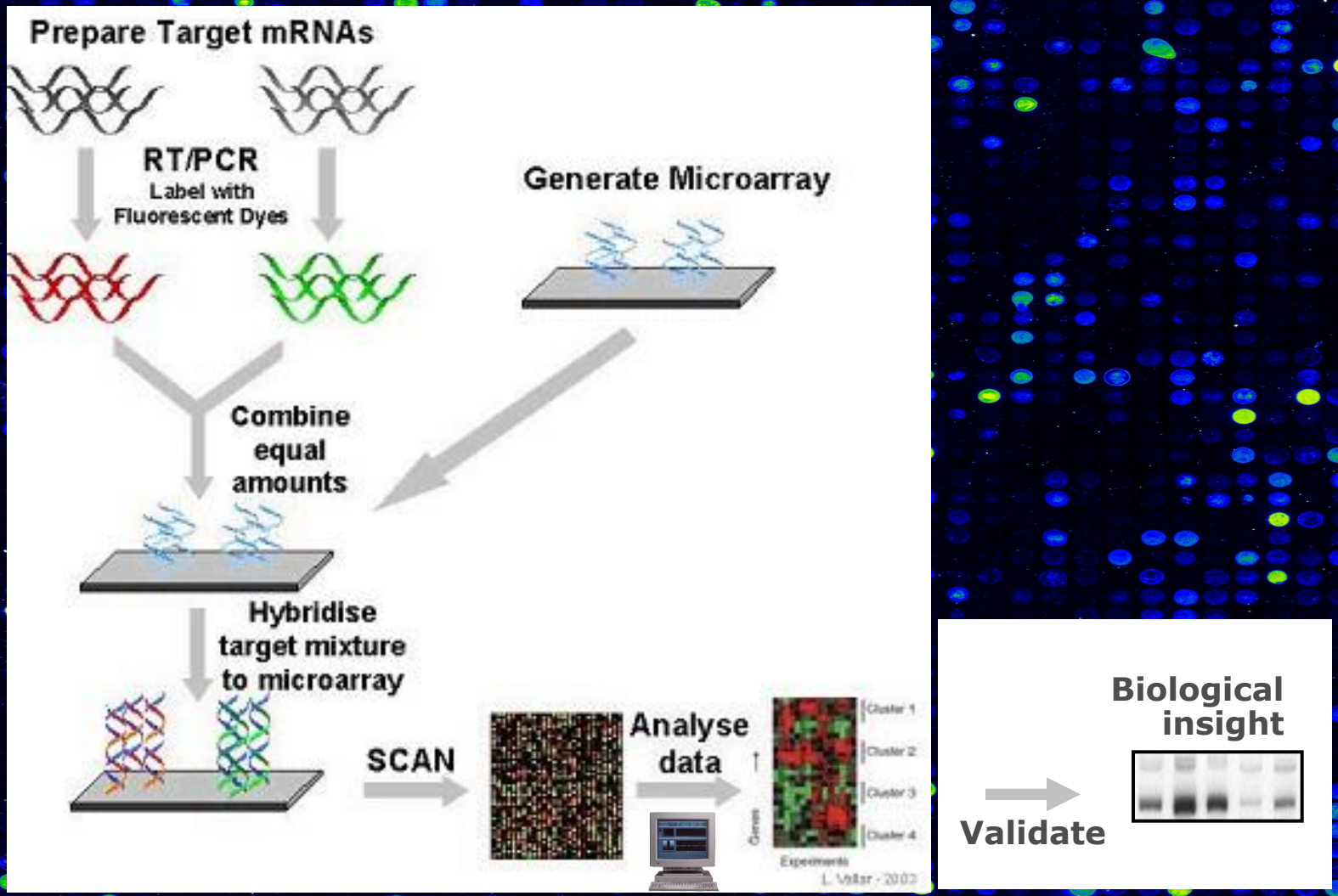
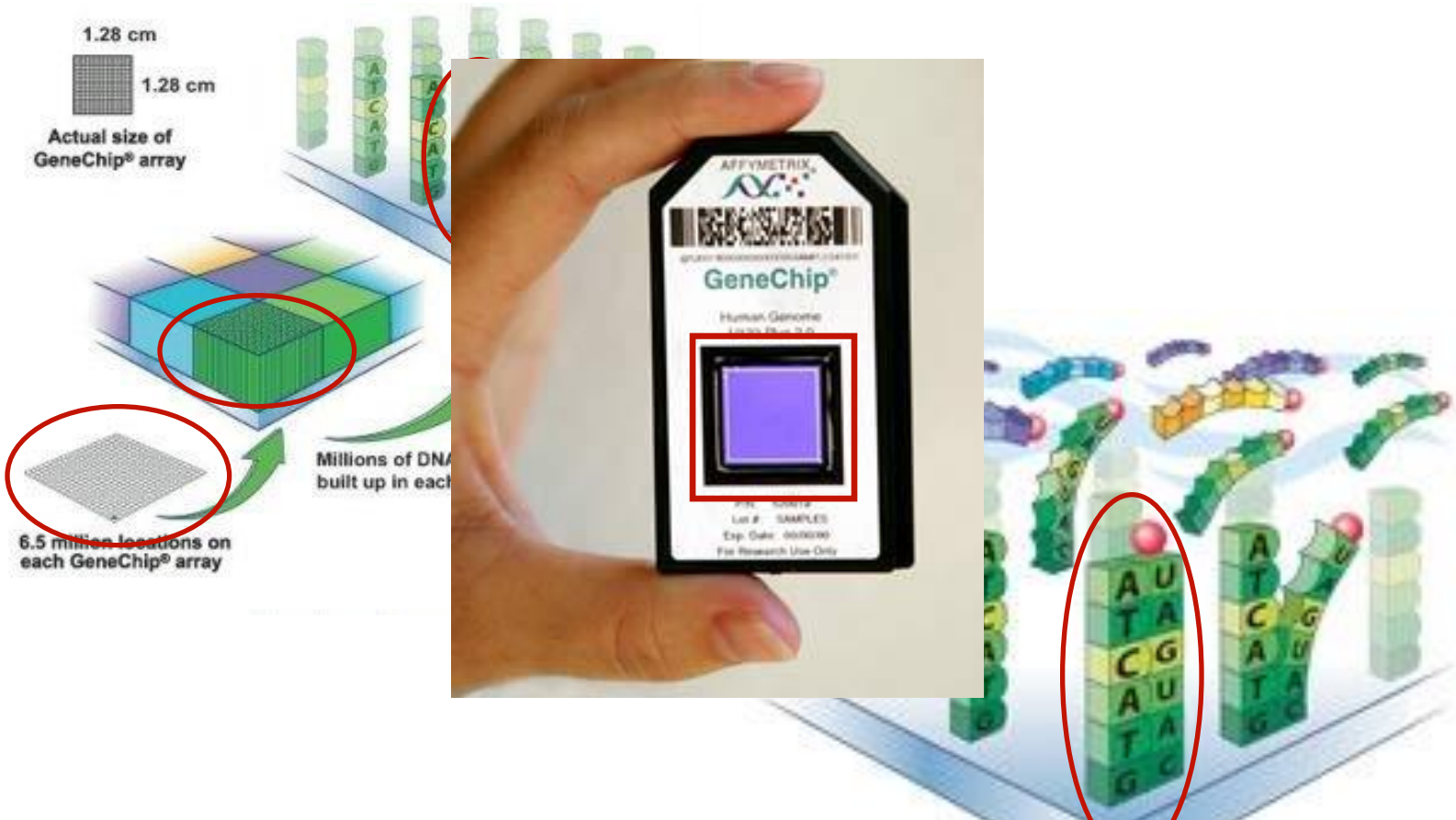


image from http://www.microarray.lu/en_print/MICROARRAY_Overview.shtml

Affymetrix chips: one sample per array



For Affymetrix chips each gene is measured by <u>dozens of probes</u> that are randomly distributed across the chip; these probes together form a <u>probeset</u>

Affymetrix Chips

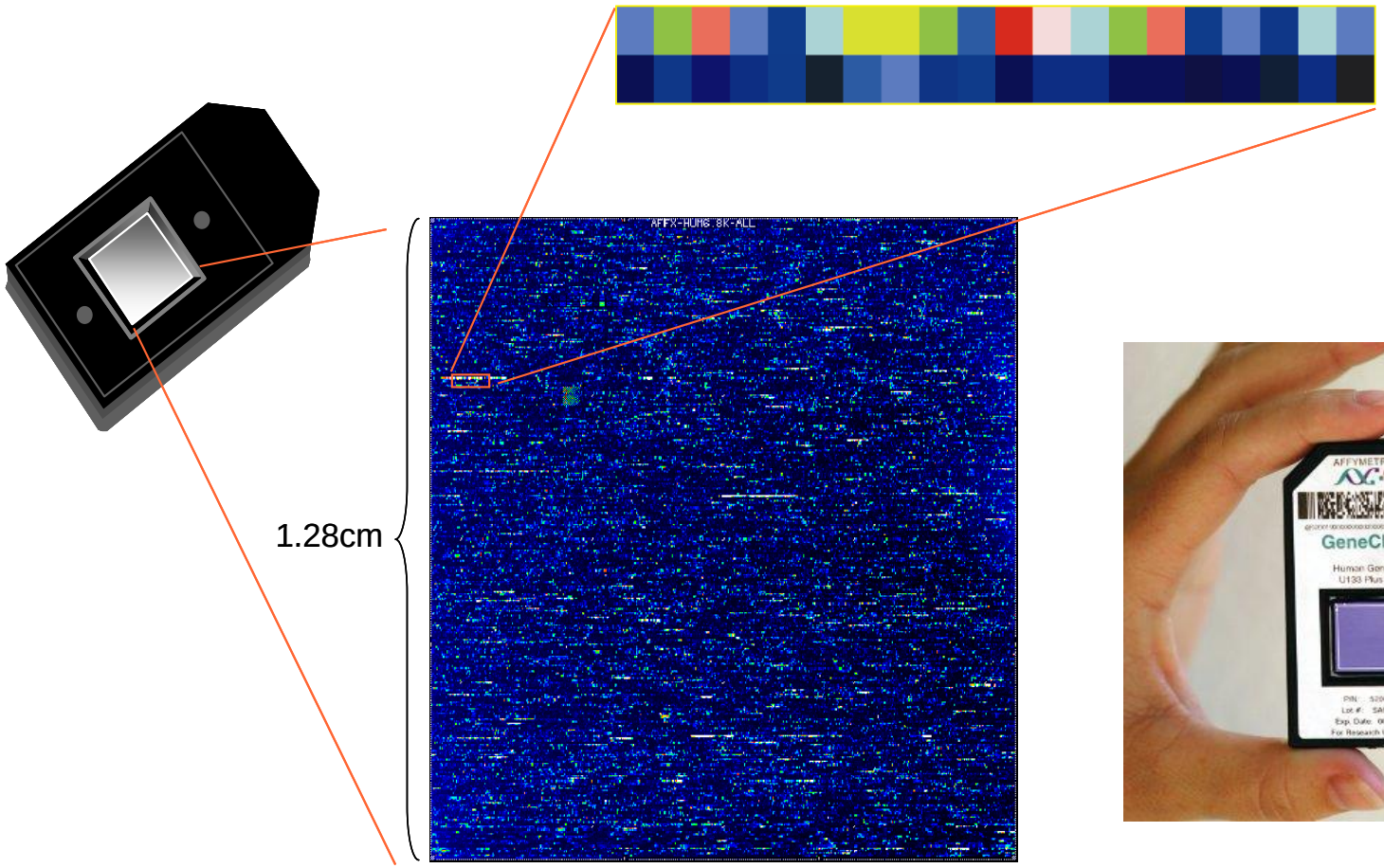


Image of Hybridized Probe Array

Image analysis

1	ORF	t0 green	t0 green bl	t0 red	t0 red bkg	t0.5 green	t0.5 green	t0.5 red	t0.5 red bk	t2 green	t2 green bl	t2 red	t2 red bkg
2	YHR007C	3570	1132	3643	692	3858	1213	5102	1052	2477	1351	3850	785
3	YOL109W	7534	1159	12218	622	7016	1386	5418	576	6119	1470	8272	872
4	YAL056W	1441	996	1043	569	2873	1062	2465	384	1984	1361	1537	858
5	YAL058W	2145	1168	1740	631	2623	1291	1768	670	2122	1535	1486	926
6	YAL059W	1894	1109	1578	575	2145	1052	801	442	1784	1385	1069	789
7	YAL060W	7927	1143	8770	694	9361	1484	5820	772	6740	1586	4029	978
8	YAL061W	5208	1171	5664	756	5914	1108	6008	494	3492	1376	3517	759
9	YAL062W	8258	1224	9527	664	5637	1836	22504	2094	4015	1474	21303	873
10	YAR002W	2374	1308	1838	752	3632	1156	2451	511	2675	1168	1881	643
11	YAR003W	2131	1230	1397	636	2668	1368	2265	580	1848	1184	1652	632
12	YAR007C	2183	1373	1553	794	3170	1179	6450	508	2191	1209	5920	650
13	YAR008W	1702	1214	964	603	2106	1397	1160	590	1635	1250	1743	662
14	YAR009C	4848	1356	4079	748	6508	1277	5457	493	4770	1191	3480	619
15	YAR010C	10550	1361	9306	748	11736	1503	10471	687	9254	1363	7756	742
16	YAL001C	1530	1118	1018	607	2221	1151	1233	421	1818	1407	1171	798
17	YAL002W	2302	1104	1881	614	2705	1493	2307	746	2102	1460	1603	892
18	YAL003W	6897	1160	7621	705	12021	1244	3263	479	6281	1450	2750	762
19	YAL004W	10306	1187	13176	718	12818	1568	8520	804	13036	1506	7086	811
20	YAL005C	9570	1305	13796	857	11039	1308	8848	594	9246	1470	4087	855
21	YAL007C	3041	1142	2768	665	4013	1530	2306	800	2629	1404	2471	834

- Start with a scanned microarray image
- Use software packages to recognise spots and compute (raw) intensities

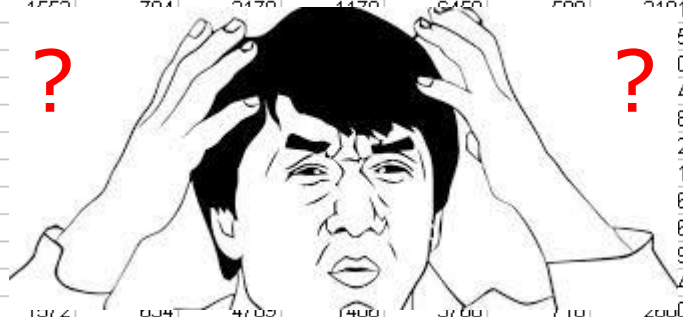
Example raw data file of a dual channel array

…are these values of good quality?

Microsoft Excel - spospread.tab

P1 | t5 green bkg

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ORF	t0 green	t0 green bk	t0 red	t0 red bkg	t0.5 green	t0.5 green	t0.5 red	t0.5 red bk	t2 green	t2 green bk	t2 red	t2 red bkg
2	YHR007C	3570	1132	3643	692	3858	1213	5102	1052	2477	1351	3850	785
3	YOL109W	7534	1159	12218	622	7016	1386	5418	576	6119	1470	8272	872
4	YAL056W	1441	996	1043	569	2873	1062	2465	384	1984	1361	1537	858
5	YAL058W	2145	1168	1740	631	2623	1291	1768	670	2122	1535	1486	926
6	YAL059W	1894	1109	1578	575	2145	1052	801	442	1784	1385	1069	789
7	YAL060W	7927	1143	8770	694	9361	1484	5820	772	6740	1586	4029	978
8	YAL061W	5208	1171	5664	756	5914	1108	6008	494	3492	1376	3517	759
9	YAL062W	8258	1224	9527	664	5637	[illegible]	22504	2094	4015	1474	21303	873
10	YAR002W	2374	1308	1838	752	3632	[illegible]	2451	511	2675	1168	1881	643
11	YAR003W	2131	1230	1397	636	2668	1368	2265	580	1848	1184	1652	632
12	YAR007C	2183	1373	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1209	5920	650
13	YAR008W	1702	1214	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1250	1743	662
14	YAR009C	4848	1356	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1191	3480	619
15	YAR010C	10550	1361	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1363	7756	742
16	YAL001C	1530	1118	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1407	1171	798
17	YAL002W	2302	1104	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1460	1603	892
18	YAL003W	6897	1160	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1450	2750	762
19	YAL004W	10306	1187	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1506	7086	811
20	YAL005C	9570	1305	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1470	4087	855
21	YAL007C	3041	1142	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1404	2471	834
22	YAL008W	3649	1274	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1675	5655	899
23	YAL009W	2067	1179	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1445	2019	826
24	YAL010C	2596	1144	2396	724	2807	1229	2026	756	2203	1498	1226	808
25	YAL011W	3971	1166	3777	668	5128	1360	3203	670	3017	1373	2448	778
26	YAL012W	3394	1239	2964	712	2653	1108	4221	611	3068	1430	1695	773
27	YAL013W	2812	1032	2763	568	2766	1320	2216	644	2085	1370	1347	808
28	YAL014C	2500	1324	1954	728	3683	1314	3212	536	2610	1121	1941	578
29	YAL015C	3010	1374	2236	753	3838	1120	2546	409	2646	1238	1570	644
30	YAL016W	4777	1260	4243	667	6863	1147	5379	449	5054	1183	2807	560
31	YAL017W	2534	1362	1828	735	3102	1214	1933	460	2659	1318	1758	706



background intensity

foreground intensity

Quality control

- Check for <u>technical failures</u> or <u>biological outliers</u>
- Check abnormalities on the array or dissimilarities between the arrays
- Decide to repeat / reject arrays if needed
- Think careful before repeating: introducing possible bias

example: QC for Affymetrix arrays using

http://www.arrayanalysis.org

ArrayAnalysis.org - Mozilla Firefox

Bestand Bewerken Beeld Geschiedenis Bladwijzers Extra Help

ArrayAnalysis.org

www.arrayanalysis.org

Meest bezocht Aan de slag Laatste nieuws

ARRAYANALYSIS.ORG welcome page

- Quick & easy on-line Affymetrix arrays Quality Control -

Get started | Download sources | Pipeline description | Documentation | Bug tracker

[Pipeline Description]

Sample quality
- Sample prep controls
- 3'/5' for control genes
- RNA degradation plot

Hybridization quality
- Spike-in controls
- Background intensity
- Percent present
- PMA calls table
- Pos/Neg controls
- Affx control profiles

Signal comparability
- Scale factor
- Boxplot
- Density histogram
- MA plots
- Reference Layout
- Pos/Neg COI plot
- 2D spatial images
- NUSE plot
- RLE plot

Array correlation
- Correlation plot
- PCA analysis
- Clustering

BiGCaT bioinformatics

Welcome to ArrayAnalysis.org !

Discover the integrated Affymetrix array Quality Control pipeline

Affymetrix microarrays of different technology versions are very often used in transcriptomics analysis.
Quality control and normalization approaches do exist, especially as packages in Bioconductor/R. However:
- Procedures are often different between teams.
- They are not always easy to access as they run through command lines.
- It is often not clear what the meaning of the specific settings and results are
- They are not always usable for the newer technology types of arrays.

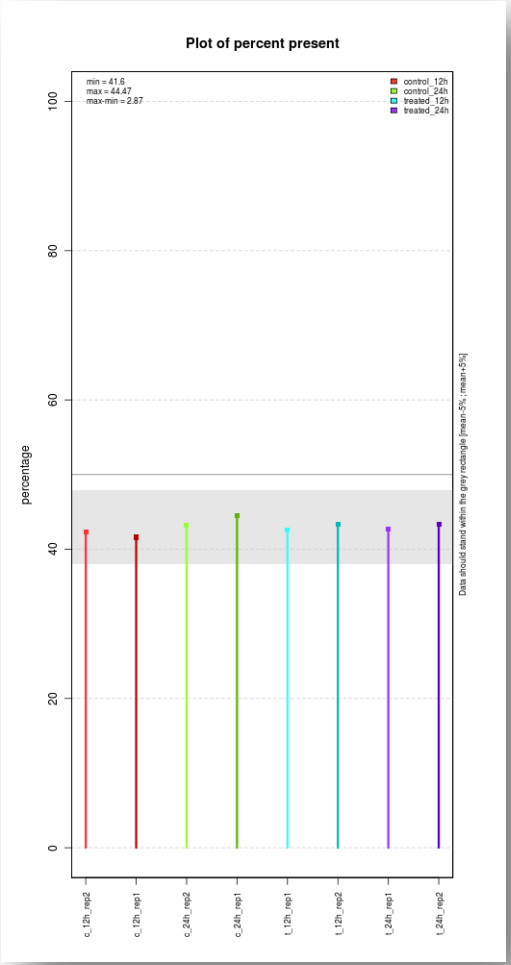
To tackle this, we proposed an automated well-documented and user-friendly pipeline for Affymetrix microarray quality control and normalization.

We gratefully acknowledge all authors of R/BioConductor packages used by affyAnalysisQC: affy, affycomp, affypdnn, affyPLM, affyQCReport, ArrayTools, bioDistm biomaRt, simpleaffy, yaqcaffy. (more...)

Tables and images of QC criteria

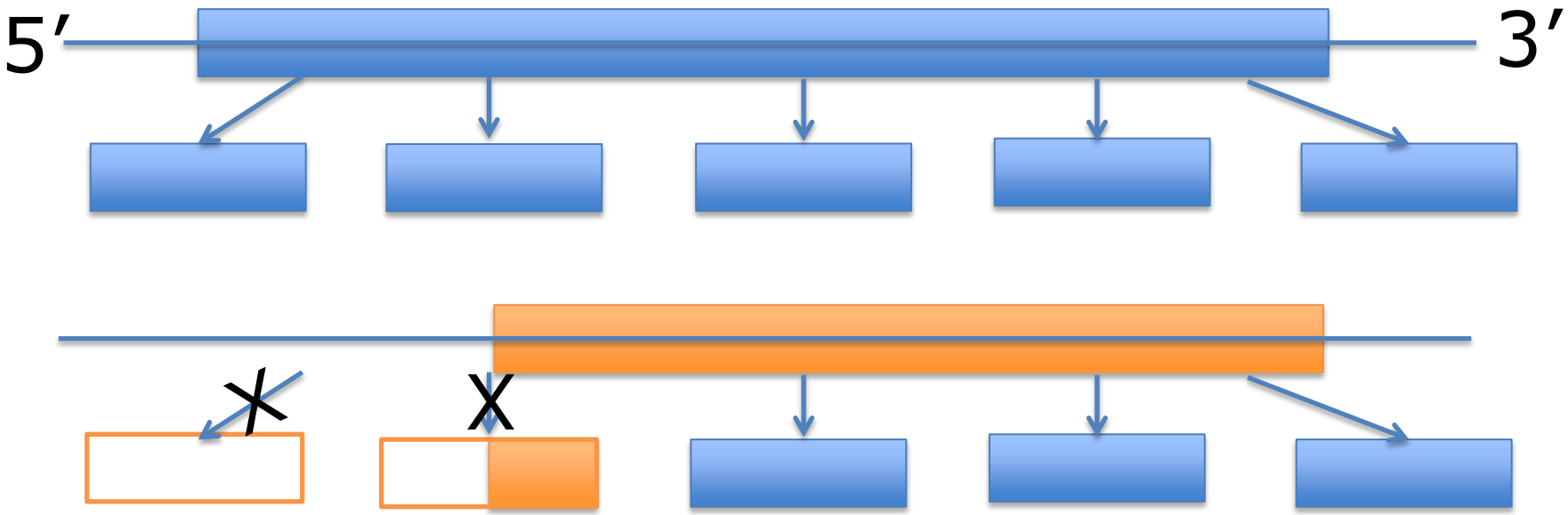
Affymetrix criteria

- ✓ Sample prep controls Lys < Phe < Thr < Dap
- ✓ Lys present
- ✓ Beta Actin 3’/5’ ≤ 3
- ✓ GAPDH 3’/5’ ≤ 1.25
- ✓ Hybridisation controls BioB < BioC < BioD < Crex
- ✓ BioB present
- ✓ Percentage present within 10%
- ✓ Background within 20 units
- ✓ Scaling factors within 3-fold from the average

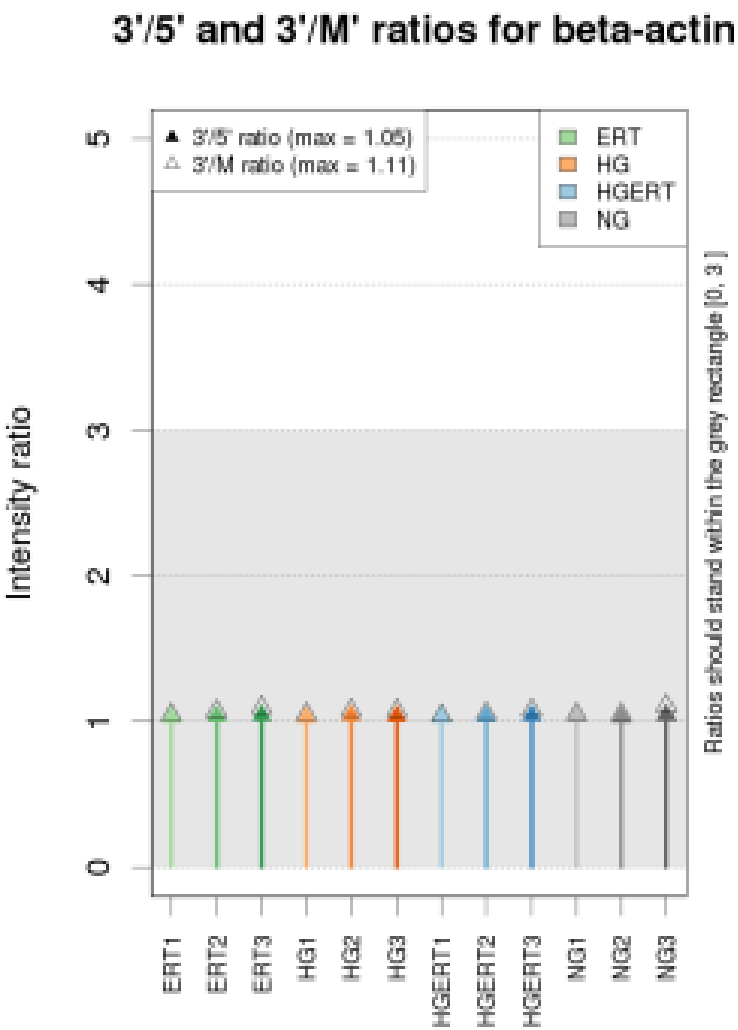


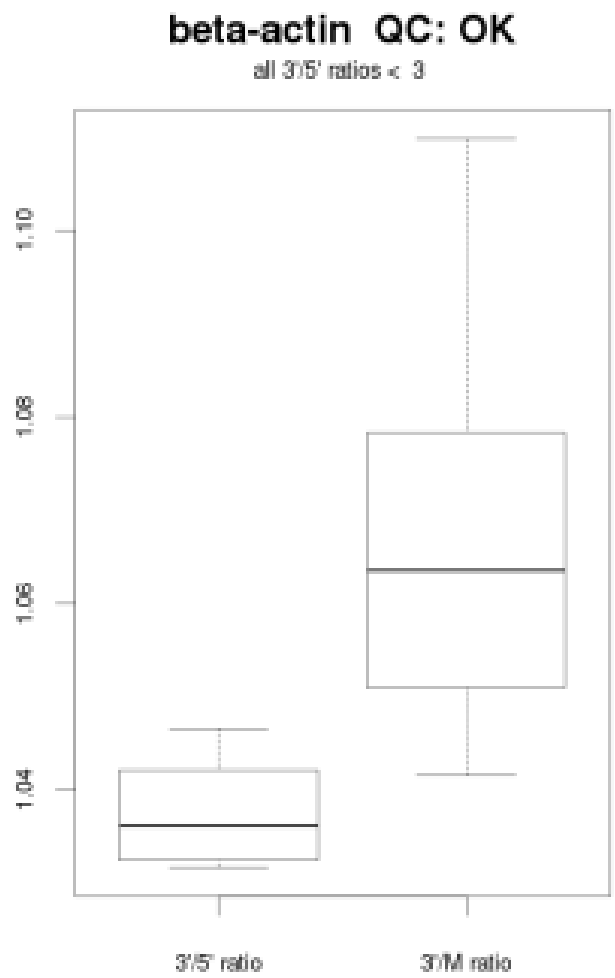
RNA degradation

- RNA degradation starts 5’ -> 3’
- Less fragments of 5’ end than 3’ indicates degradation

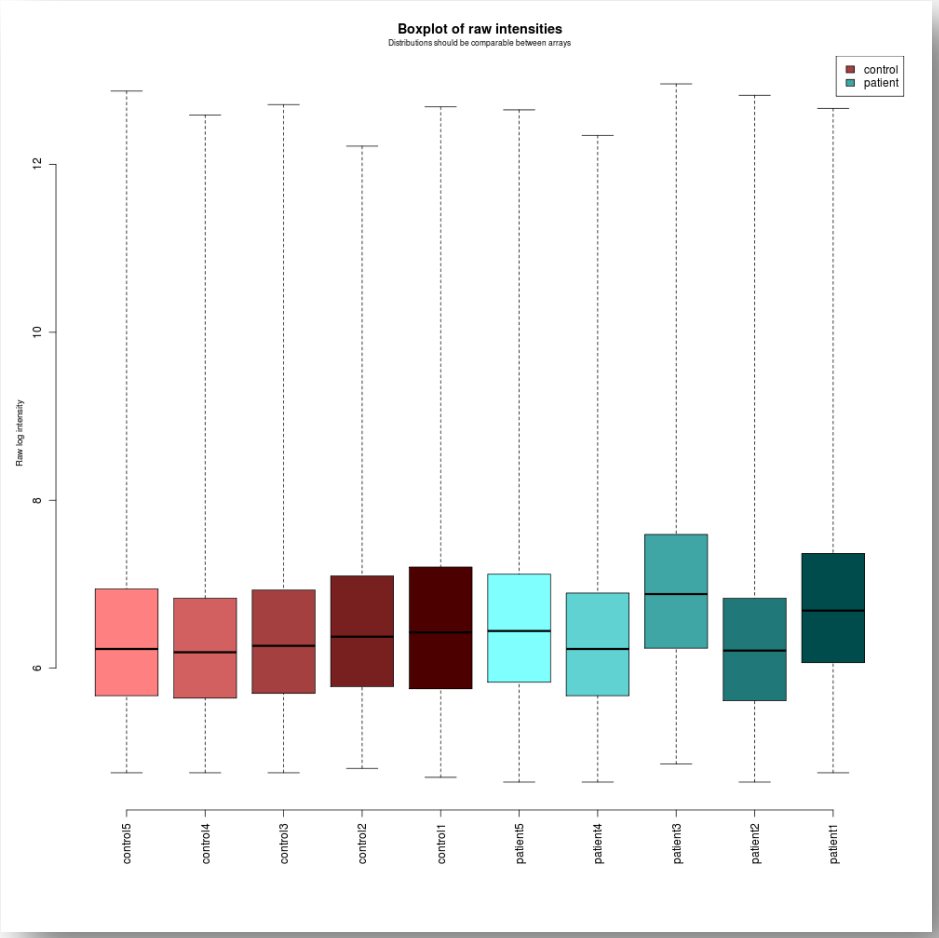


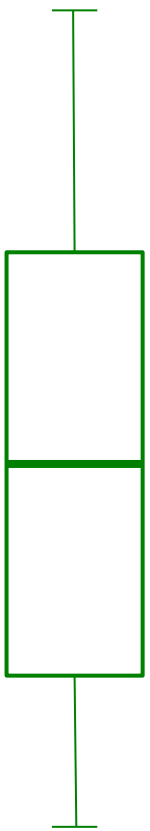
RNA degradation analysis





Average intensity boxplot





maximum

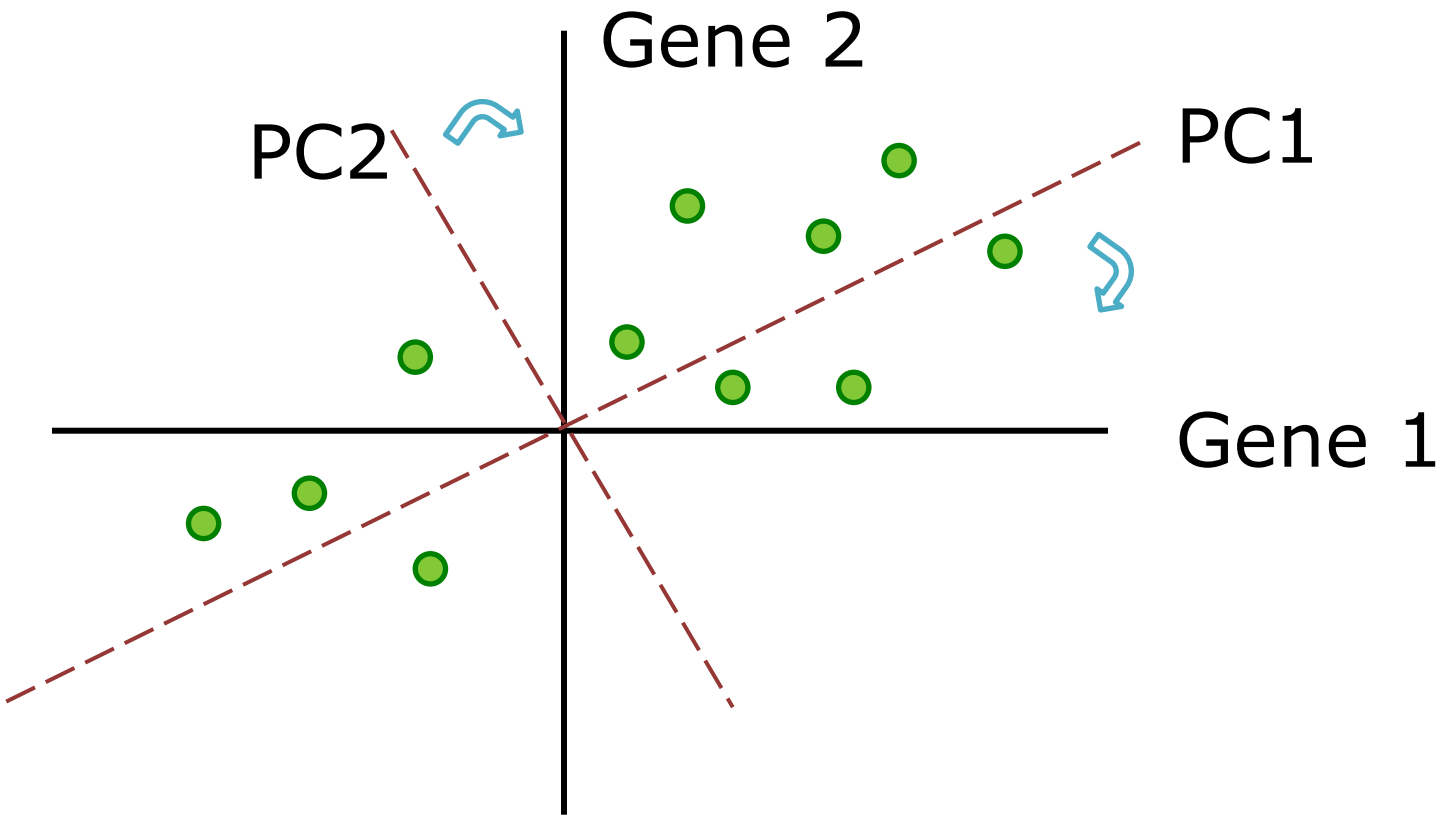
3rd quartile

median

1st quartile

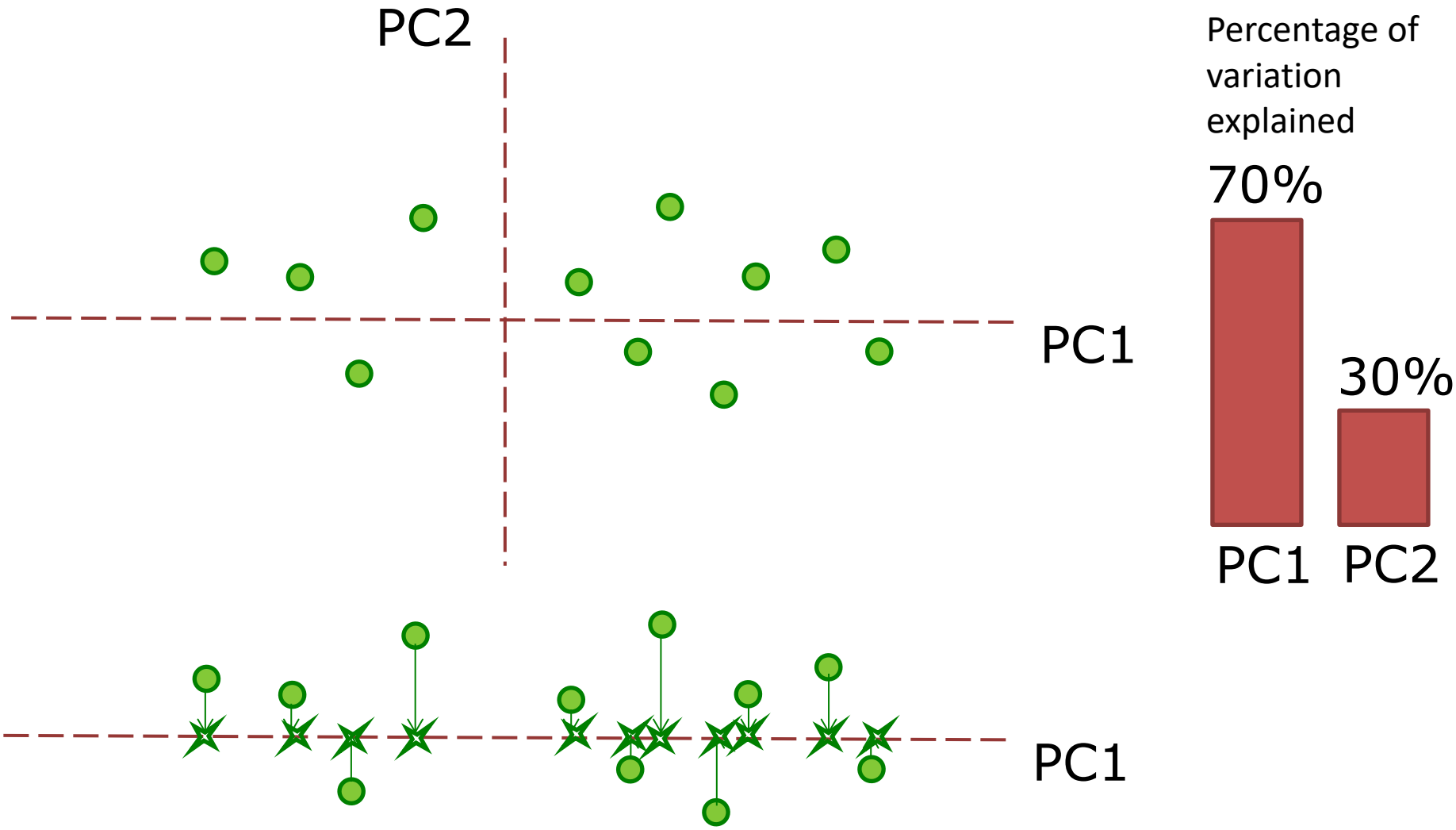
minimum

PCA = Principle Components Analysis



This is a simplified example: in reality PCA is used to reduce the dimensions of a multi-dimensional data set to 2 or 3 dimensions

PCA = Principle Components Analysis



Now we reduced the two dimensional data set to one dimension, thereby explaining (keeping) 70% of the original variation

Making all your data comparable:

PRE-PROCESSING

QC and pre-processing

- Ensure signal comparability <u>within</u> each array
  - Stains on the array
  - Gradient over the array

- Ensure comparable signals <u>between</u> all arrays
  - Degraded / low quality sample
  - Failed hybridisation
  - Too low or high overall intensity

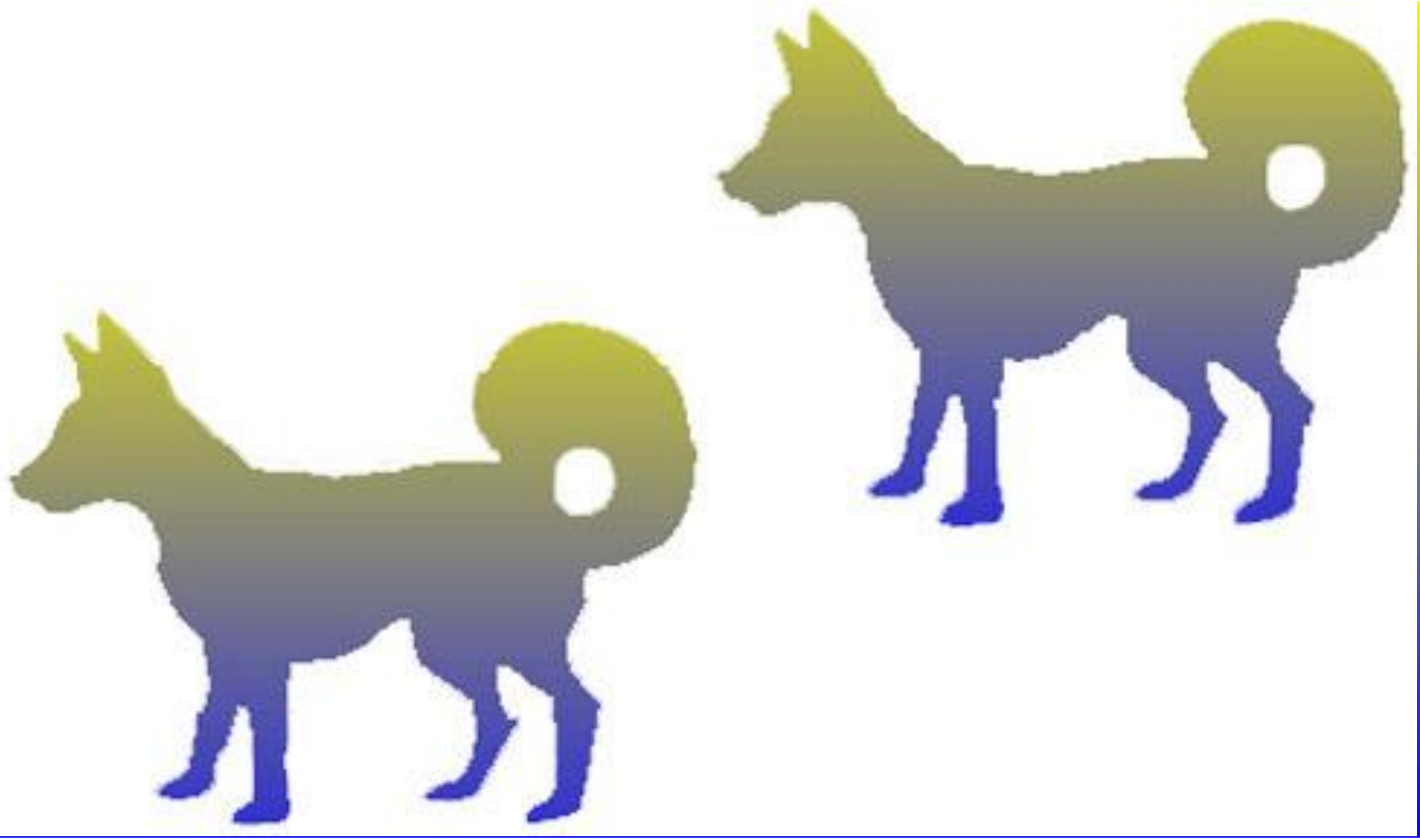
- Some effects can be <u>corrected</u> for, others require <u>removal</u> of data from the set

Background correction

- Background signal needs to be corrected for
  - for example signal of remaining non-hybridized mRNA

- Three types of background
  - Overall slide background
    - Can be corrected for by subtracting mean background, or by subtracting mean of empty spots
  - Local slide background
    - Same as previous, but per slide region
  - Specific background
    - For example cross-hybridization, can be corrected for by mismatch probes (in case of Affymetrix arrays)

The importance of background



Uneven background

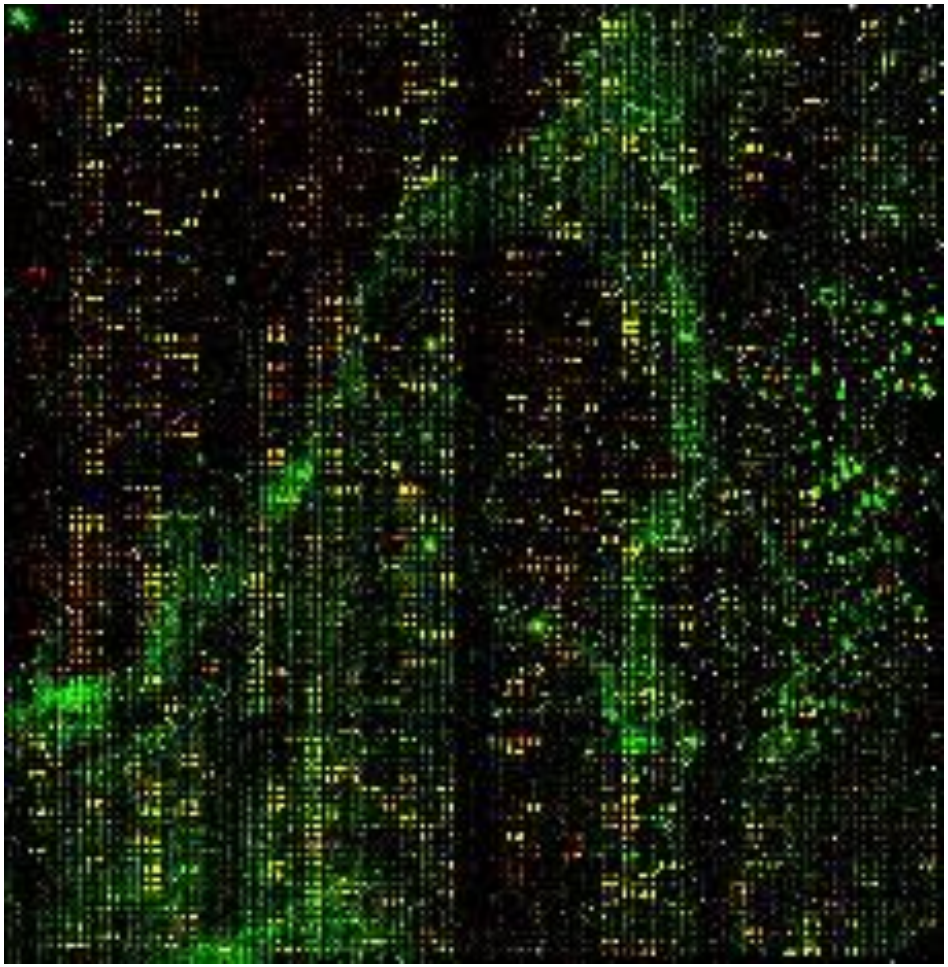
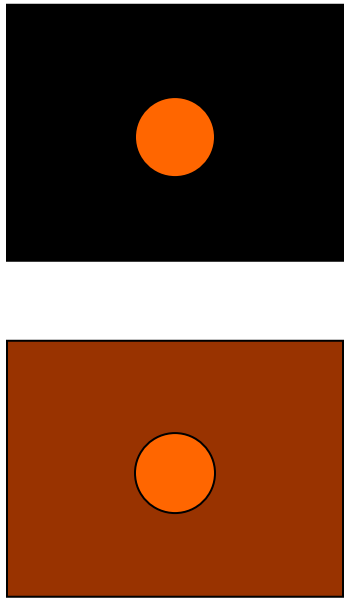


image from www.systemsbiology.nl

Background correction



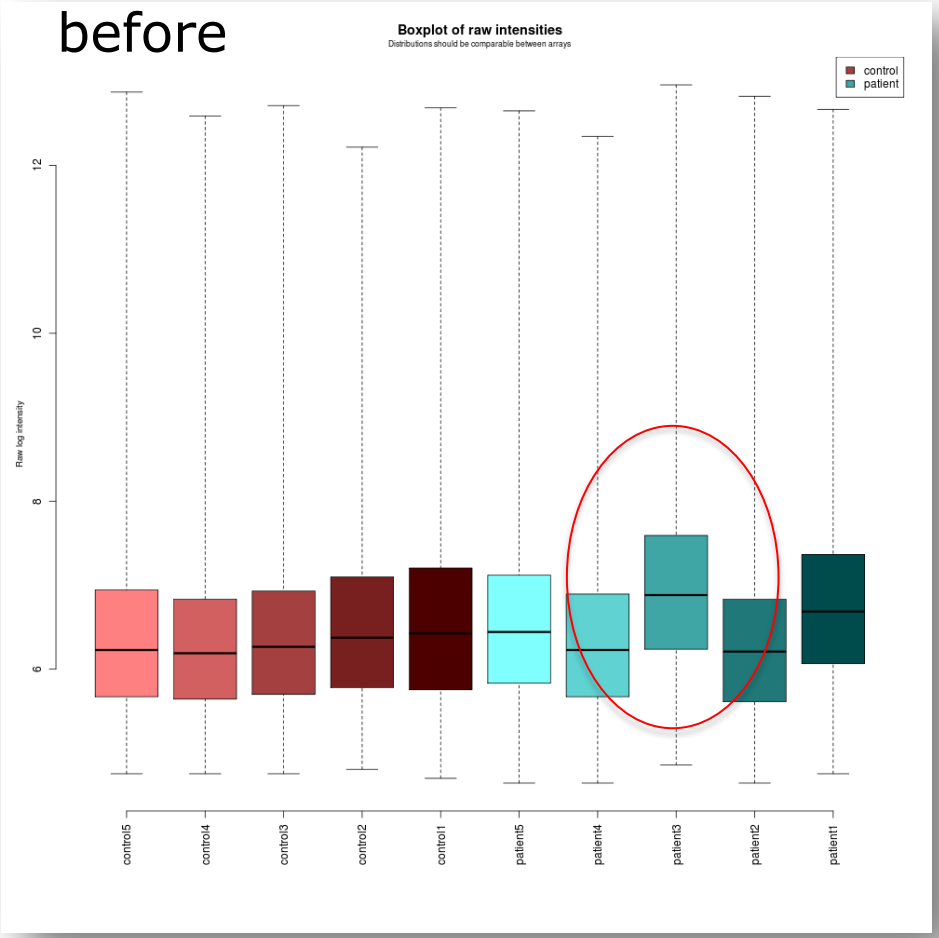
Measure the intensity of the background around the spot as well as the intensity of the spot itself

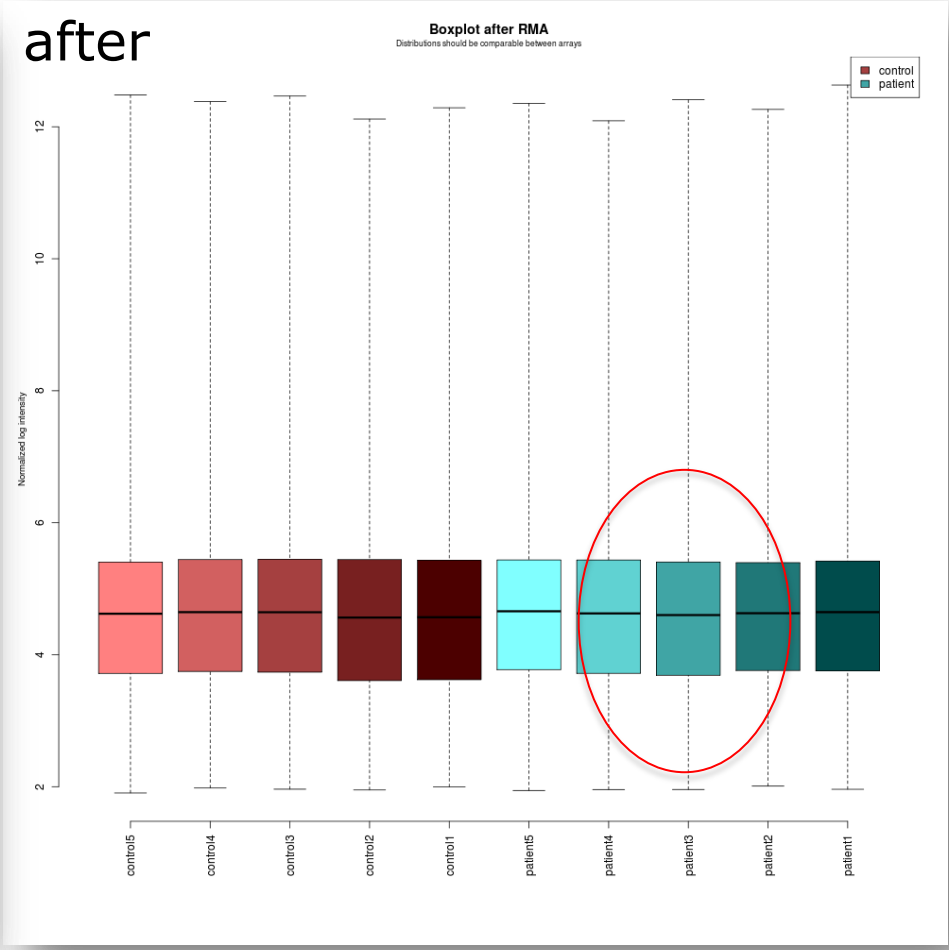
- Reported intensity = spot intensity – background intensity
- More advanced methods are generally used
  - prevent negative values

Normalisation

- Adjusting values
- Between-slide normalisation: correct for experimental differences between slides
  - e.g. one may have an overall higher signal due to differences in hybridisation
- Within-slide normalisation: correct for within slide variations
  - by applying normalisation per region, per spot group etc.
- For dual channel arrays: between-channel normalisation

Boxplot before and after normalization



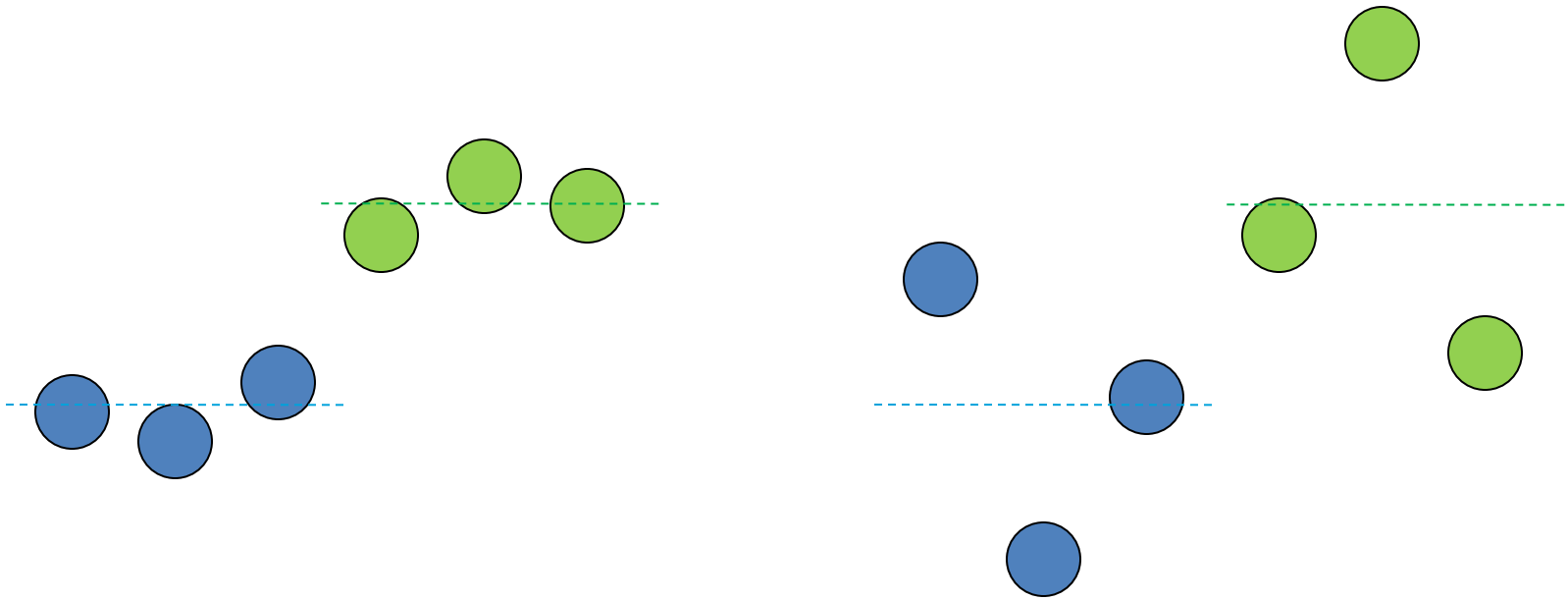


Finding differentially expressed genes:

STATISTICAL ANALYSIS

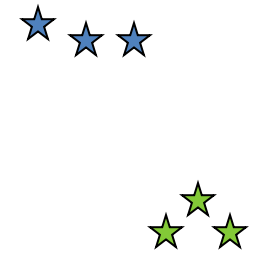
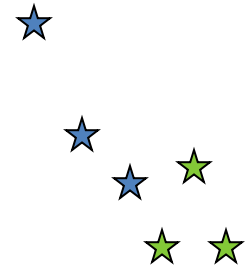
Which genes have changed? (I)

- “Every gene that has changed two-fold is relevant”
- Doesn’t take variation into account



Which genes have changed? (II)

Often people use both the difference and statistical significance between two groups to determine the list of differentially expressed genes

	Large difference	Small difference
Significant		
Non-Significant		

Comparing experimental groups

- The ratio between the (average) expression in two experimental groups is generally called the fold change
- Generally we do no not work directly with the fold, but with the logged fold change, which is called the log ratio or log fold change
- Why?
  - Intuitive understanding log ratio -> + is up, - down
  - Visualization

Considerations for the t-test

- Requirements
  - Adequate sample size (n)
  - Knowledge on the experimental grouping
  - Normally distributed data
    - “Gaussian curve”
  - P value

Example results table

	A	B	C	D	E	F	G	H
1	logFC	Fold Chan	AveExpr	t	P.Value	adj.P.Val	B	ID
2	-7.23489	-150.633	7.863147	-30.6765	2.37E-11	1.02E-07	15.14212	170496
3	-2.67637	-6.39244	11.12548	-20.2298	1.53E-09	3.30E-06	12.15308	24614
4	-4.92565	-30.3926	7.786763	-19.26	2.49E-09	3.58E-06	11.75189	24296
5	-2.7171	-6.5755	7.613864	-17.415	6.73E-09	7.27E-06	10.90447	29569
6	-4.97479	-31.4458	9.115975	-16.4341	1.19E-08	1.03E-05	10.40304	24300
7	-2.14602	-4.42604	9.631401	-15.1037	2.72E-08	1.96E-05	9.657943	266602
8	-1.80955	-3.50532	6.405987	-14.5628	3.88E-08	2.40E-05	9.331303	678701
9	2.696893	6.484038	6.947894	13.93027	5.98E-08	3.23E-05	8.930261	25256
10	2.0373	4.104765	7.227198	13.19942	1.01E-07	4.84E-05	8.439403	29301
11	1.848977	3.602446	7.935563	12.39145	1.85E-07	7.99E-05	7.859381	192268
12	-4.19194	-18.2768	7.498633	-12.1536	2.23E-07	8.60E-05	7.680627	25355
13	-2.54811	-5.84867	6.233037	-12.0639	2.39E-07	8.60E-05	7.612213	308100
14	-1.14175	-2.20649	5.738167	-11.8145	2.92E-07	9.69E-05	7.419026	29242
15	2.512379	5					078	64313
16	-2.05217	-					524	29230
17	1.684944	3					374	25231
18	-1.41899	-					342	25427
19	-2.07639	-					545	29469
20	1.421941	2					988	24252
21	3.428312	1					281	79243
22	-1.73615	-					991	89784
23	2.274691						542	84029
24	1.58322						904	24538
25	-2.93092	-7.62596	6.194612	-10.2059	1.16E-06	0.000209	6.060767	24299

Voila!
List of differentially expressed genes!

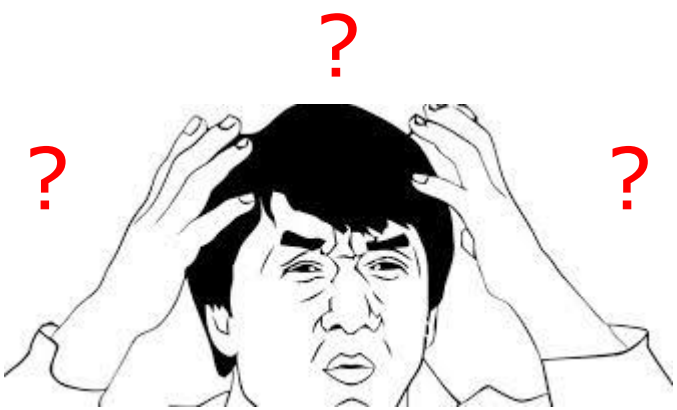
Possible filtering on:

- ✓ P value
  - ✓ Only significant changes
- ✓ logFC or FC
  - – Is a significant change with a FC of 1.001 relevant?
- ✓ Average expression
  - – Recall that lowly expressed genes are less reliably measured

www.moillusions.com

Gene ontology enrichment analysis (overrepresentation analysis (ORA))

- NNT NAD(P) transhydrogenase
- DHRS2 Dehydrogenase/reductase SDR family member 2
- ME3 NADP-dependent malic enzyme
- SDHC Succinate dehydrogenase cytochrome b560 subunit
- BCO2 Beta,beta-carotene 9',10'-oxygenase
- SURF1 Surfeit locus protein 1



GO: mitochondria oxidation-reduction process !

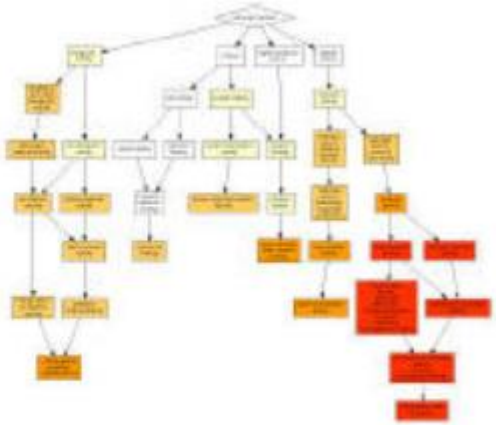
BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 16 2012, pages 2209–2210
doi:10.1093/bioinformatics/bts366

Databases and Ontologies Advance Access publication June 27, 2012

GO-Elite: a flexible solution for pathway and ontology over-representation

Alexander C. Zambon[1], Stan Gaj[2], Isaac Ho[3], Kristina Hanspers[3], Karen Vranizan[3], Chris T. Evelo[2], Bruce R. Conklin[3,4], Alexander R. Pico[3] and Nathan Salomonis[3,*]

GORILLA



***G**ene **O**ntology en**RI**chment ana**L**ysis and visua**L**iz**A**tion tool*

Running example Usage instructions GOrilla Ne

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: Browse... No file selected.

Background set:

Or upload a file: Browse... No file selected.

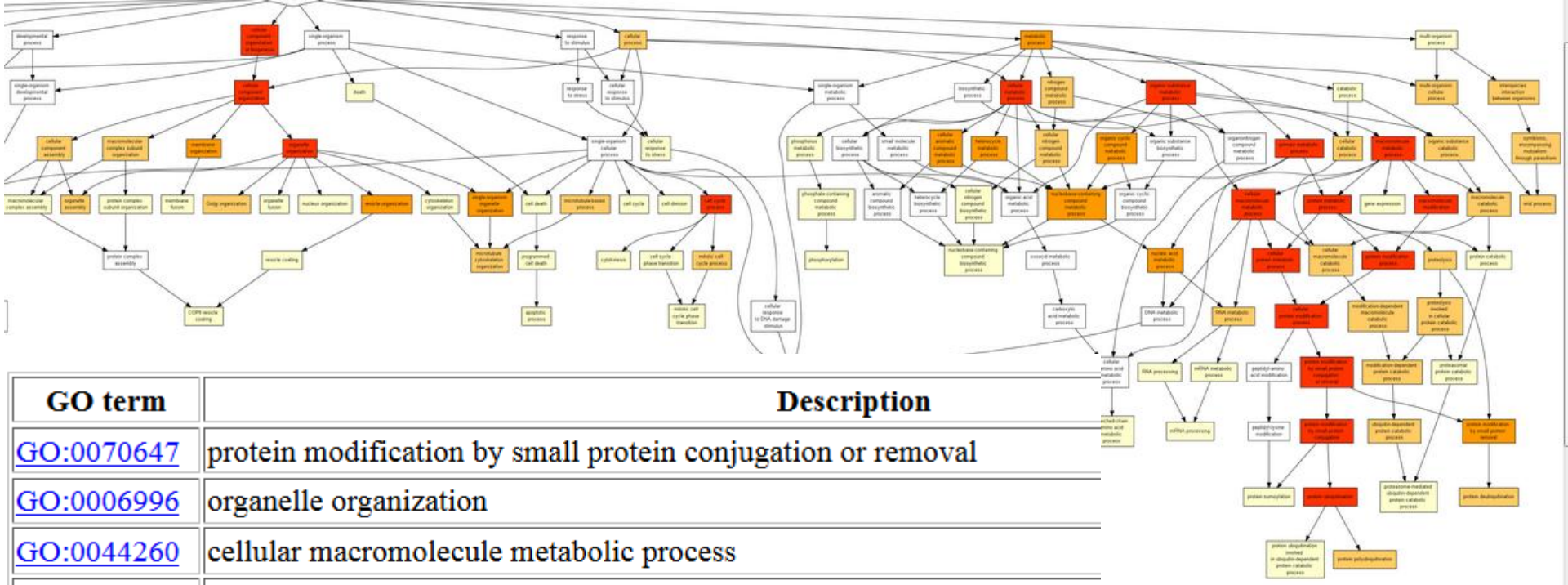
Step 4: Choose an ontology

Process Function Component All

Search Enriched GO terms

2 list of genes (identifiers):

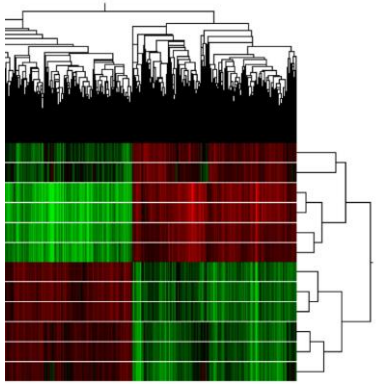
- **Target set:** list of changed genes
- **Background set:** list of all investigated genes



GO term	Description
GO:0070647	protein modification by small protein conjugation or removal
GO:0006996	organelle organization
GO:0044260	cellular macromolecule metabolic process
GO:0044238	primary metabolic process
GO:0008104	protein localization
GO:0044267	cellular protein metabolic process
GO:0033036	macromolecule localization
GO:0006464	cellular protein modification process
GO:0036211	protein modification process
GO:0044237	cellular metabolic process
GO:0043170	macromolecule metabolic process
GO:0043412	macromolecule modification
GO:0048193	Golgi vesicle transport
GO:0032446	protein modification by small protein conjugation
GO:0071704	organic substance metabolic process

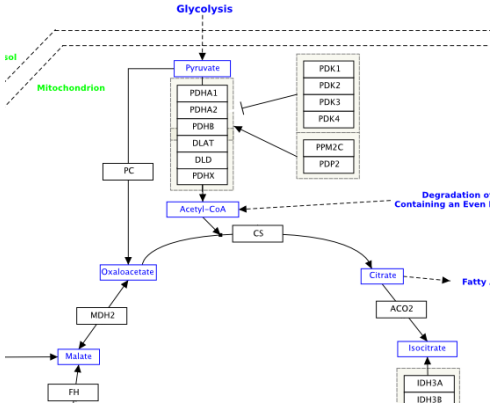
Pathway analysis

Why Pathway Analysis?



- Intuitive to biologists
  - Puts data in biological context
  - More intuitive way of looking at your data
  - More efficient than looking up gene-by-gene

- Computational analysis
  - Overrepresentation analysis
  - Network analysis



PathVisio

- http://www.pathvisio.org/downloads/
- PathVisio is a free open-source biological pathway analysis software that allows you to **draw, edit and analyze biological pathways**.
- Direct down- and upload to WikiPathways.org via WikiPathways App

Biological Context

- Statistical results:
  - 1,300 genes are significantly regulated after treatment with X
- Biological Meaning:
  - Is a certain biological pathway activated or deactivated?
  - Which genes in these pathway are significantly changed?

How to use PathVisio

1. Prepare your data
2. Import your data
3. Find enriched pathways
4. Create a visualization
5. Export pathway

1. Prepare your data!

File Format

- PathVisio accepts **Tab delimited text files**
- Prepare and export from Excel

A1 | probeset

	A	R	S	T	U	V	W
1	probeset	system	avg_02mmoll	avg_05mmoll	avg_10mmoll	avg_30mmoll	rel_02mm
2	1367452_at	X	11.29566606	11.41559784	11.50689421	11.51252231	-0.211228
3	1367453_at	X	9.571538603	9.64925023	9.64421156	9.611808199	-0.072672
4	1367454_at	X	9.999264981	9.808902886	10.4537435	10.8402463	-0.454478
5	1367455_at	X	11.11130052	11.01312681	11.48128423	11.63689531	-0.369983
6	1367456_at	X	11.55919016	11.72946117	11.57407478	11.67939786	-0.014884
7	1367457_at	X	9.333530241	9.645627303	9.879507778	9.903407488	-0.545977
8	1367458_at	X	7.445770323	7.545258507	7.737163677	7.907681903	-0.291393
9	1367459_at	X	11.75766074				
10	1367460_at	X	11.31811043				
11	1367461_at	X	10.73966188				
12	1367462_at	X	9.685481596				
13	1367463_at	X	10.3369898				
14	1367464_at	X	9.419124099				
15	1367465_at	X	10.5962418				

File name: GSE12817_anova.txt

Save as type: Text (Tab delimited) (*.txt)

- Excel Template (*.xltx)
- Excel Macro-Enabled Template (*.xltm)
- Excel 97-2003 Template (*.xlt)
- Text (Tab delimited) (*.txt)
- Unicode Text (*.txt)
- XML Spreadsheet 2003 (*.xml)

File Format

- Export from R

```
write.table(myTable, file = txtFile,
            col.names = NA, sep = "\t", quote = FALSE, na = "NaN")
```

Identifier Systems

PathVisio accepts many identifier systems:

- Probes
  - Affymetrix, Illumina, Agilent,...
- Genes and Proteins
  - Entrez Gene, Ensembl, UniProt, HUGO,...
- Metabolites
  - ChEBI, HMDB, PubChem,...

Nicotine Activity on Chromaffin Cells (Homo sapiens) – WikiPathways

www.wikipathways.org/index.php/Pathway:WP1603

Nicotine Activity on Chromaffin Cells (Homo sapiens) – WikiPathways | Pathguide: the pathway resource list

Log in / create account

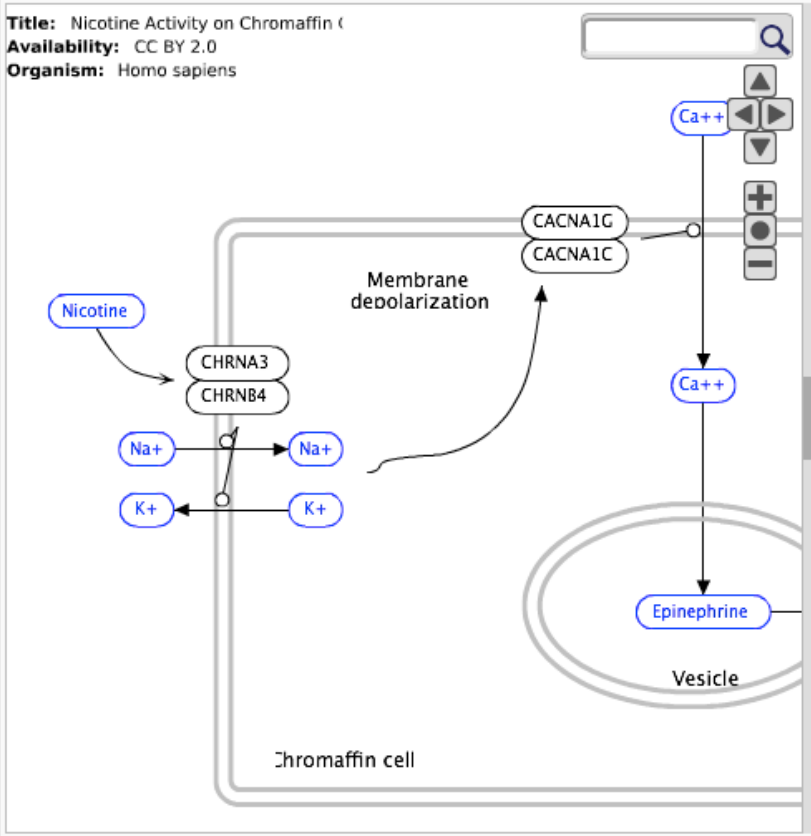
pathway | discussion | view source

Nicotine Activity on Chromaffin Cells (Homo sapiens)

Identifier Mapping

Kristina Hanspers, Alexander Pico

Title: Nicotine Activity on Chromaffin (
Availability: CC BY 2.0
Organism: Homo sapiens



CACNA1C

Annotated with: 775 (Entrez Gene)

Find pathways with CACNA1C...

External references:

- Affy
  - M92269_f_at
  - 7953040
  - Z34822_f_at
  - 33623_g_at
  - 38002_s_at
- Agilent
- Ensembl
- Entrez Gene
- Gene Wiki
- GeneOntology
- HGNC
- IPI
- Illumina
- OMIM

1 Curation Tags
2 Description
3 Ontology Tags
4 Bibliography
5 Categories
6 History
7 External references

Log in to edit pathway | not working? | Download

WIKIPATHWAYS BETA
Pathways for the People

search

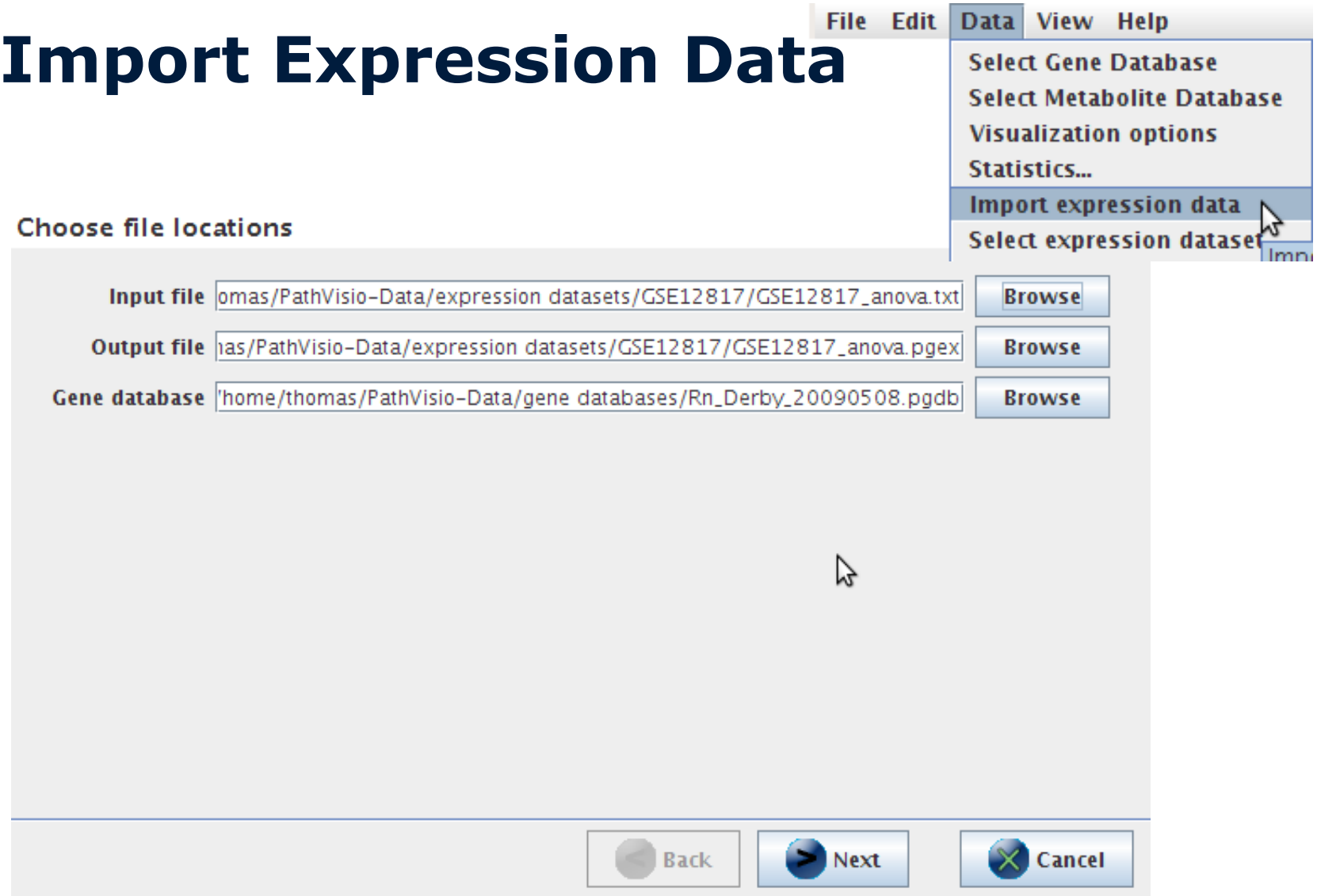
navigation
- Home
- Help

pathway
- Create
- Browse
- Wish List
- Download
- Web service API

overview
- Recent Changes
- Most Viewed
- Most Edited
- New Pathways
- Statistics

community
- About us
- Contact us
- How to cite
- Curation events
- BiGCaT portal
- CIRM portal
- GenMAPP portal
- Micronutrient portal

2. Import Expression Data



Identifier mapping database

Your data

Entrez Gene

5326
153
4357
65543
2094
90218
...

A pathway

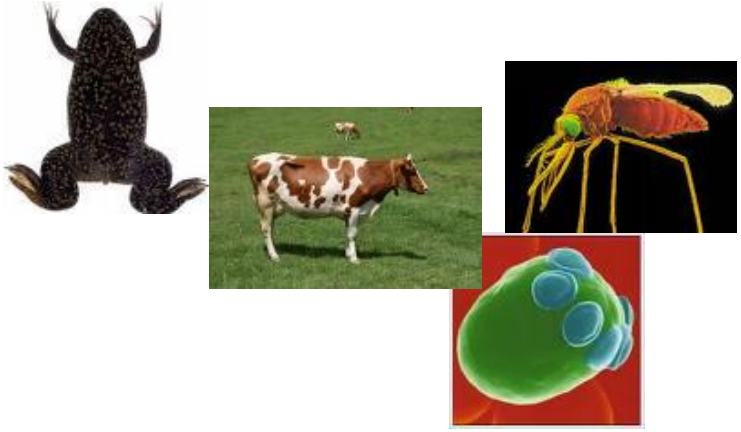
4357

ENS0002114

P4235

Load BridgeDB files

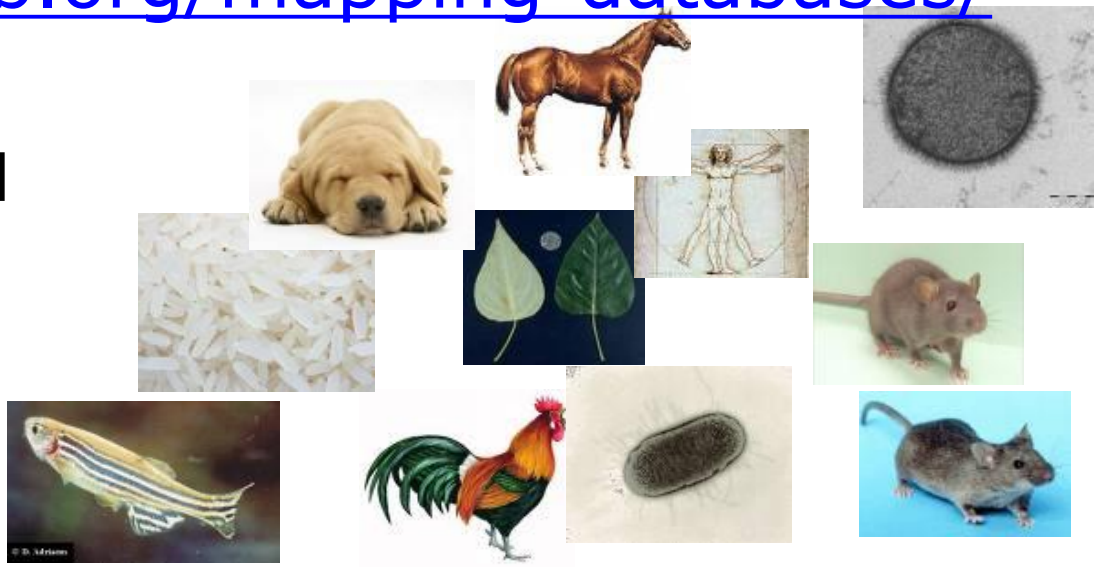
ID mapping database



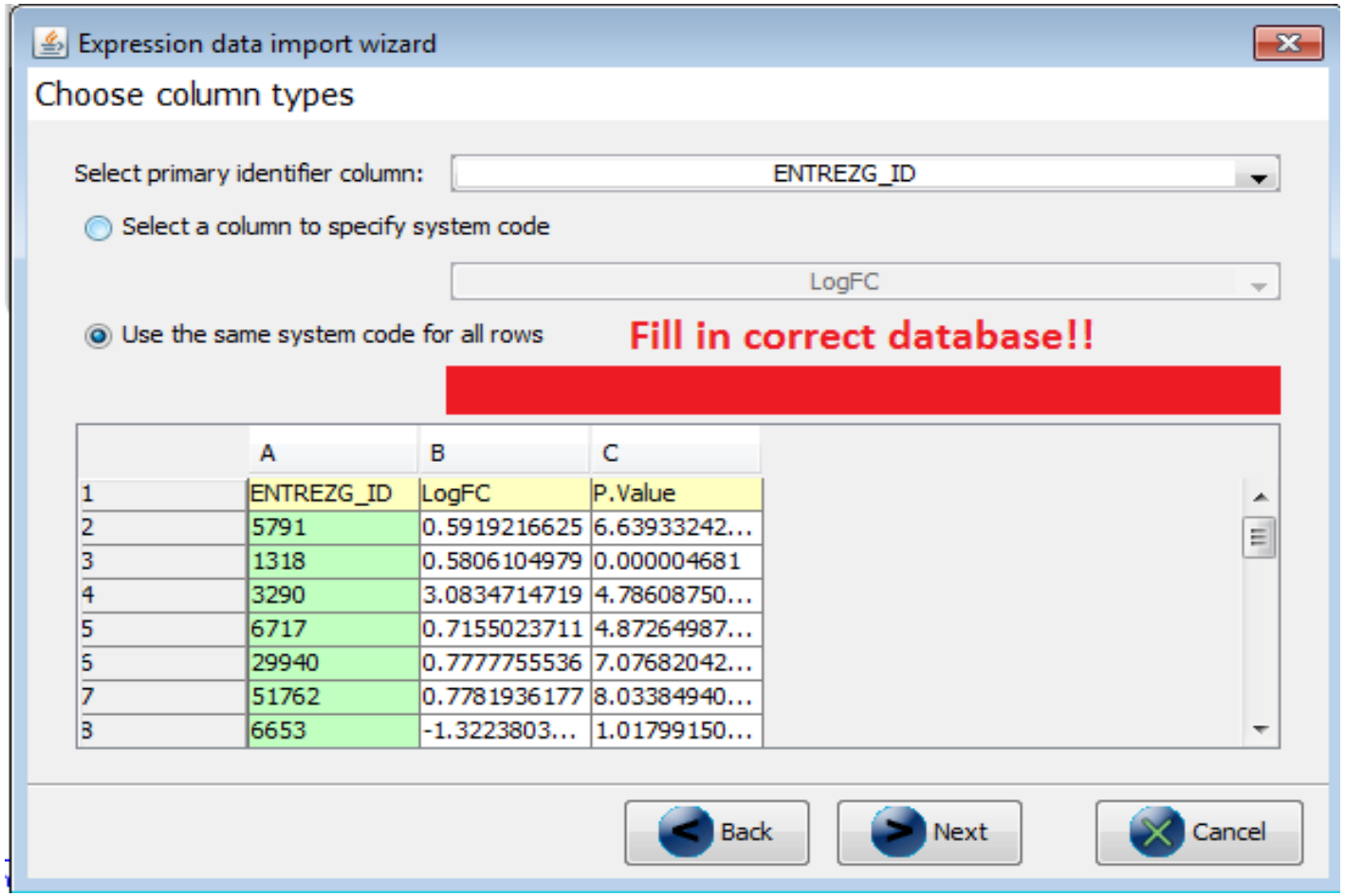
- Download from http://www.pathvisio.org/downloads/download-bridgedbs/
- OR

http://www.bridgedb.org/mapping-databases/

32 species supported



Identifier and System Code



	A	B	C
1	ENTREZG_ID	LogFC	P.Value
2	5791	0.5919216625	6.63933242...
3	1318	0.5806104979	0.000004681
4	3290	3.0834714719	4.78608750...
5	6717	0.7155023711	4.87264987...
6	29940	0.7777755536	7.07682042...
7	51762	0.7781936177	8.03384940...
8	6653	-1.3223803...	1.01799150...

Exception File

Perform import

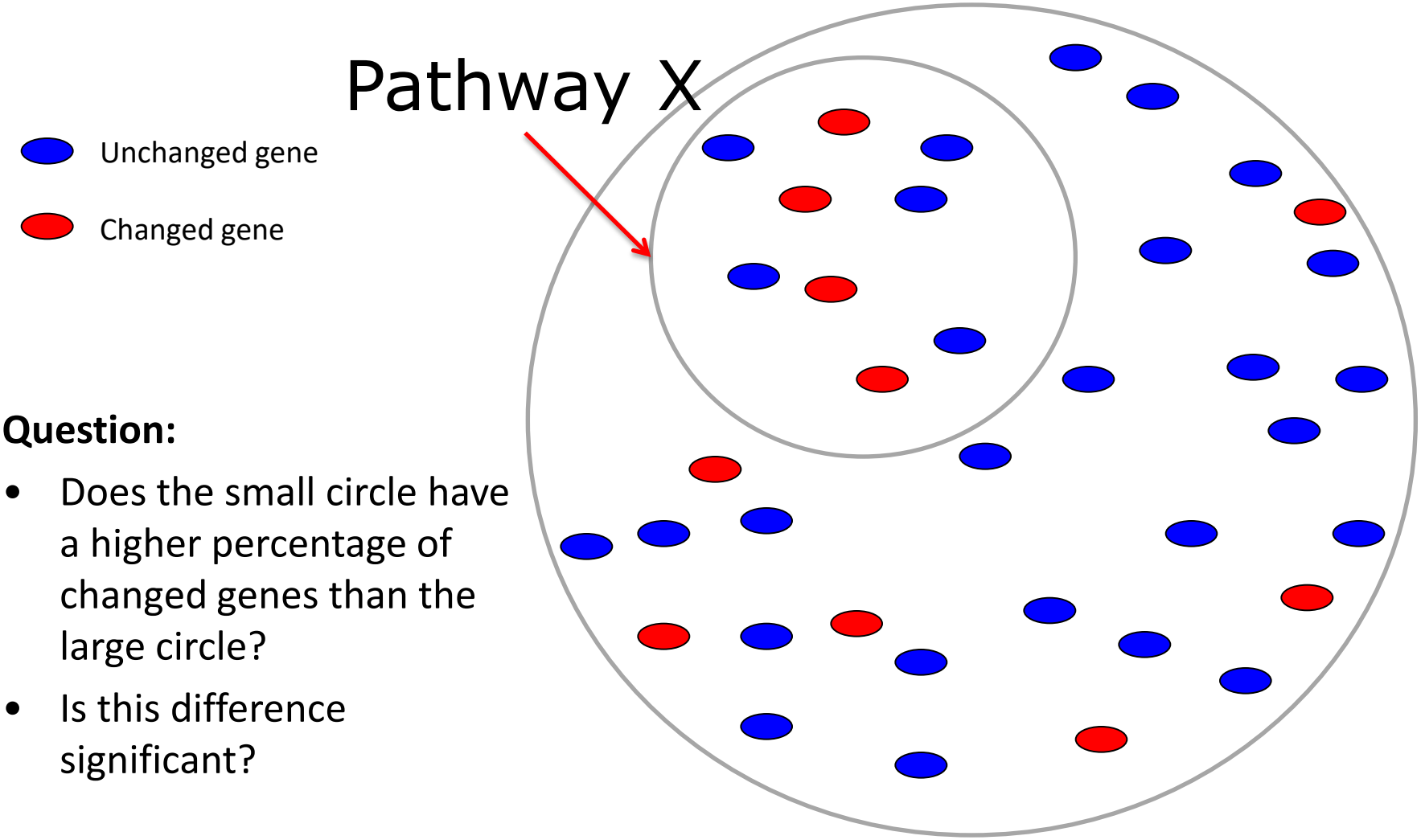
Finalizing database (this may take some time)

```
Creating expression dataset
> Processing headers
> Processing headers
> Processing lines
31099 rows of data were imported succesfully
Finalizing database
Finalizing database
```

Exceptions file

```
Line 4: X:1367454_at    Could not look up this identifier in the synonym database
Line 12:    X:1367462_at    Could not look up this identifier in the synonym database
Line 21:    X:1367471_at    Could not look up this identifier in the synonym database
Line 24:    X:1367474_at    Could not look up this identifier in the synonym database
Line 25:    X:1367475_at    Could not look up this identifier in the synonym database
Line 30:    X:1367480_at    Could not look up this identifier in the synonym database
Line 34:    X:1367484_at    Could not look up this identifier in the synonym database
Line 35:    X:1367485_at    Could not look up this identifier in the synonym database
Line 39:    X:1367489_at    Could not look up this identifier in the synonym database
Line 40:    X:1367490_at    Could not look up this identifier in the synonym database
```

3. Find „enriched" pathways by applying pathway statistics



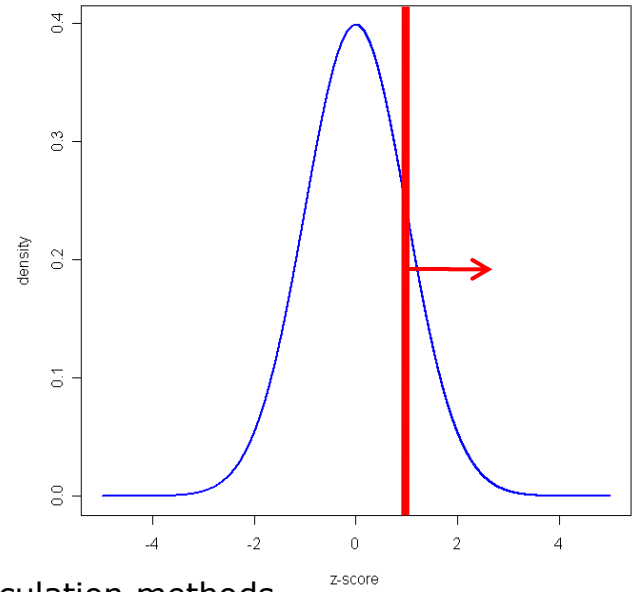
Question:

- Does the small circle have a higher percentage of changed genes than the large circle?
- Is this difference significant?

Calculate Z-scores

- The Z-score can be used as a measure for how much a subset of genes is different from the rest

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



- r = changed genes in Pathway
- n = total genes in Pathway
- R = changed genes
- N = total genes

<u>Other enrichment calculation methods</u>

Ackermann M et al., A general modular framework for gene set enrichment analysis, BMC bioinformatics, 2009

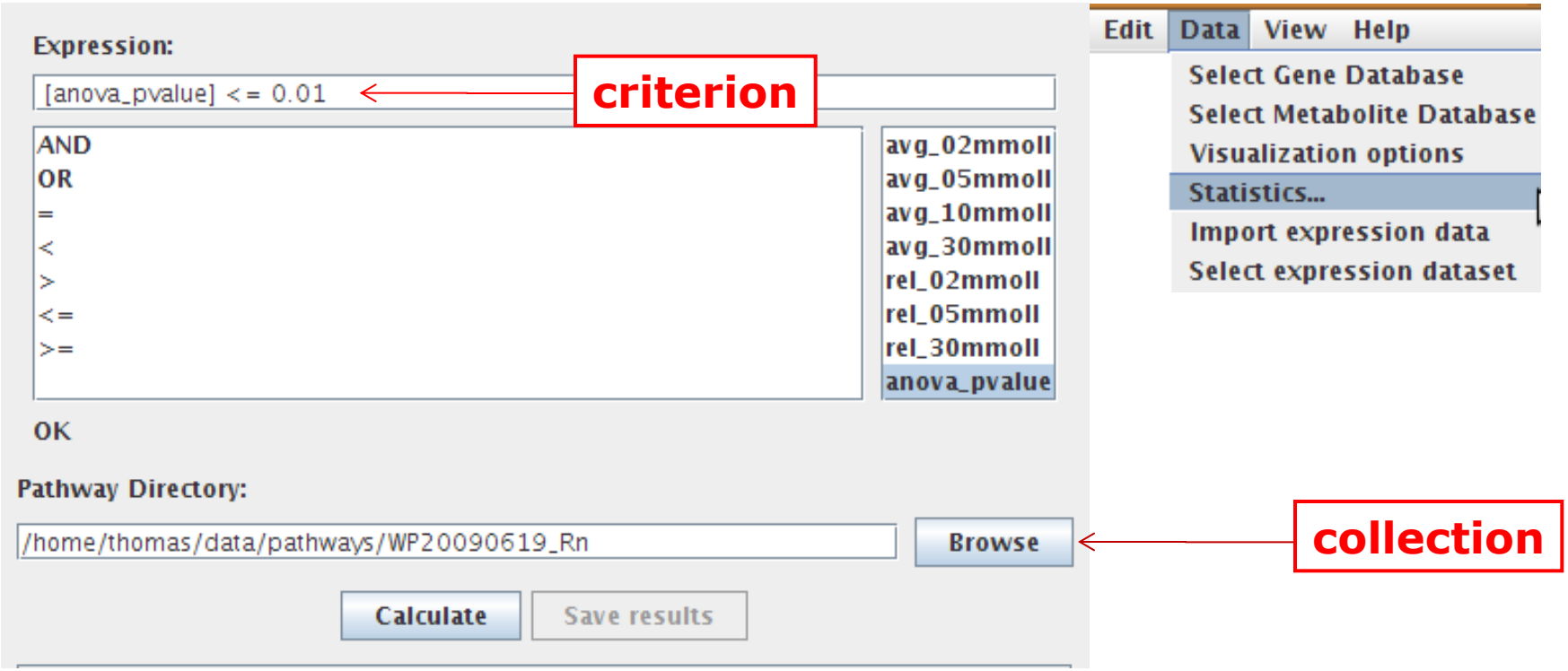
Z-score

- The Z-score is a ranking method.
  - High Z-score → selection is very different from the rest of the dataset
  - Z-score = 0 → selection is not different at all

Criteria

Define criterion and select pathway collection

([LogFC] < -1 OR [LogFC] > 1) AND [P.Value] < 0.05



Z-score Calculation

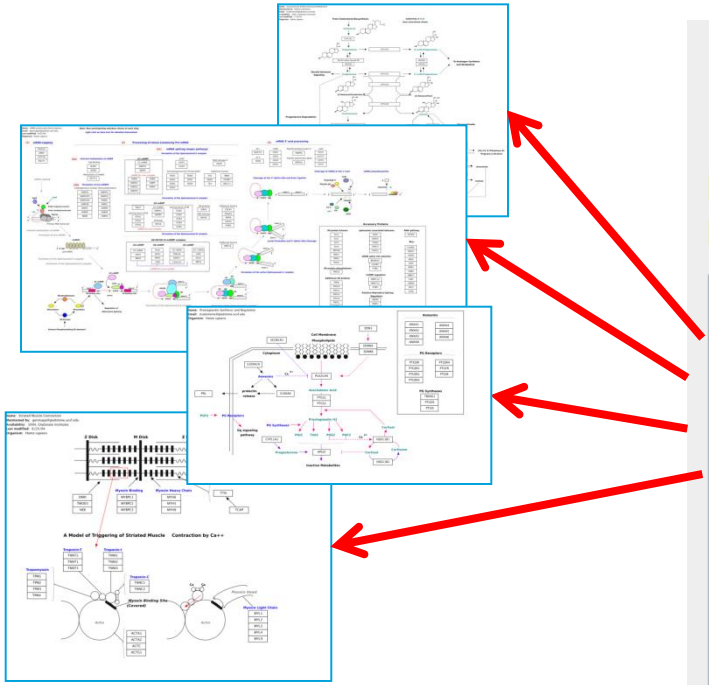
r = changed genes in Pathway
n = total genes in Pathway

Calculate | Save results

Rows in data (N): 3607
Rows meeting criterion (R): 1046

Pathway	positi... (r)	measu... (n)	total	%	Z Score
Translation Factors	29	50	50	58.00%	4.55
mRNA processing	58	125	131	46.40%	4.36
Cell cycle	43	89	94	48.31%	4.07
DNA Replication	23	42	49	54.76%	3.70
G1 to S cell cycle control	33	67	69	49.25%	3.69
DNA damage response	32	67	70	47.76%	3.42
TNF-alpha/NF-kB Signaling Pathway	71	186	188	38.17%	2.83
Tamoxifen metabolism	3	3	31	100.00%	2.71
FAS pathway and Stress induction of HSP...	18	38	43	47.37%	2.51
TGF-beta Receptor Signaling Pathway	56	151	152	37.09%	2.24
Wnt Signaling Pathway NetPath	42	109	110	38.53%	2.23
Androgen Receptor Signaling Pathway	41	112	114	36.61%	1.80
Delta-Notch Signaling Pathway	32	85	85	37.65%	1.78
IL-7 Signaling Pathway	18	44	45	40.91%	1.75
Keap1-Nrf2	7	14	16	50.00%	1.73
Glucuronidation	10	22	41	45.45%	1.71
Estrogen metabolism	8	17	33	47.06%	1.64

Z-score Calculation

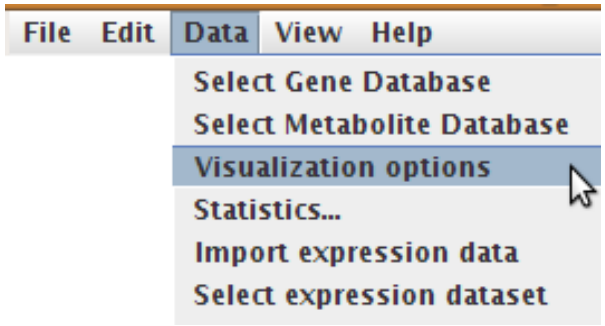


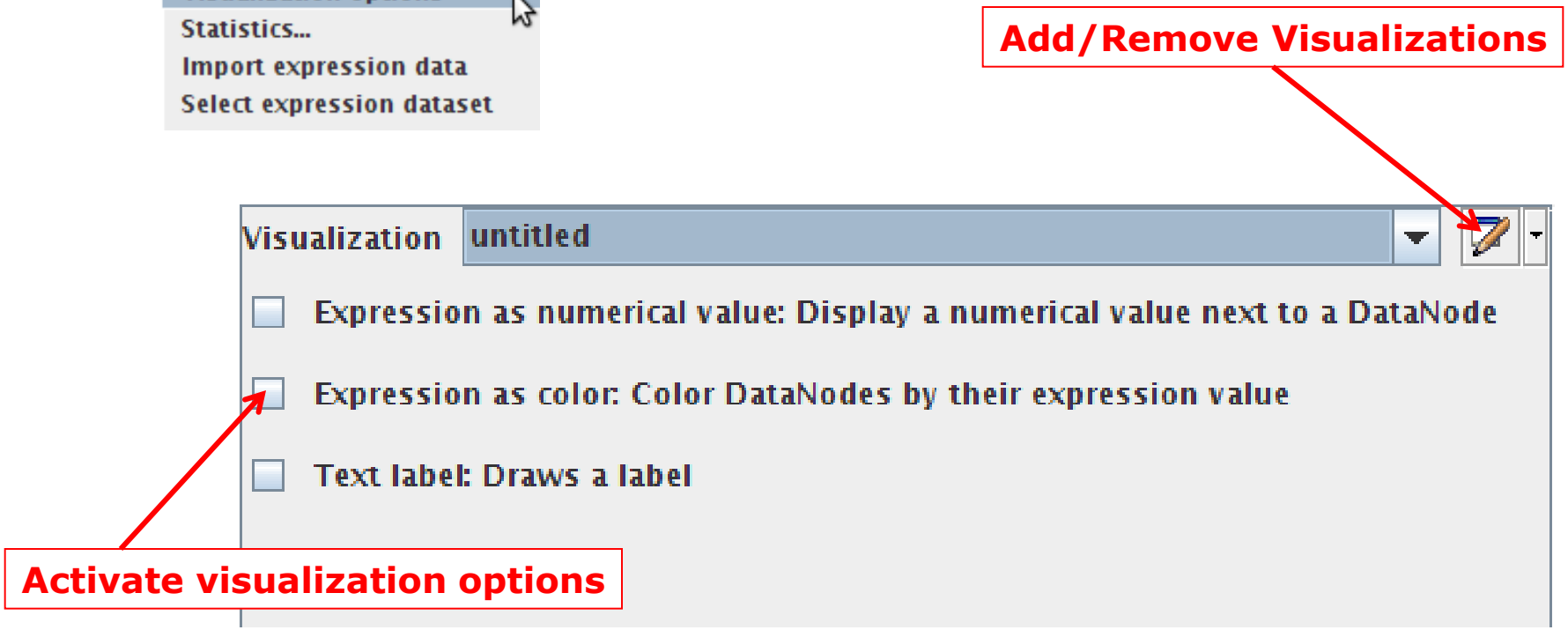
Calculate | Save results

Rows in data (N): 3607
Rows meeting criterion (R): 1046

Pathway	positi...	measu...	total
Translation Factors	29	50	50
mRNA processing	58	125	131
Cell cycle	43	89	94
DNA Replication	23	42	49
G1 to S cell cycle control	33	67	69
DNA damage response	32	67	70
TNF-alpha/NF-kB Signaling Pathway	71	186	188
Tamoxifen metabolism	3	3	31
FAS pathway and Stress induction of HSP...	18	38	43
TGF-beta Receptor Signaling Pathway	56	151	152

4. Create a Visualization

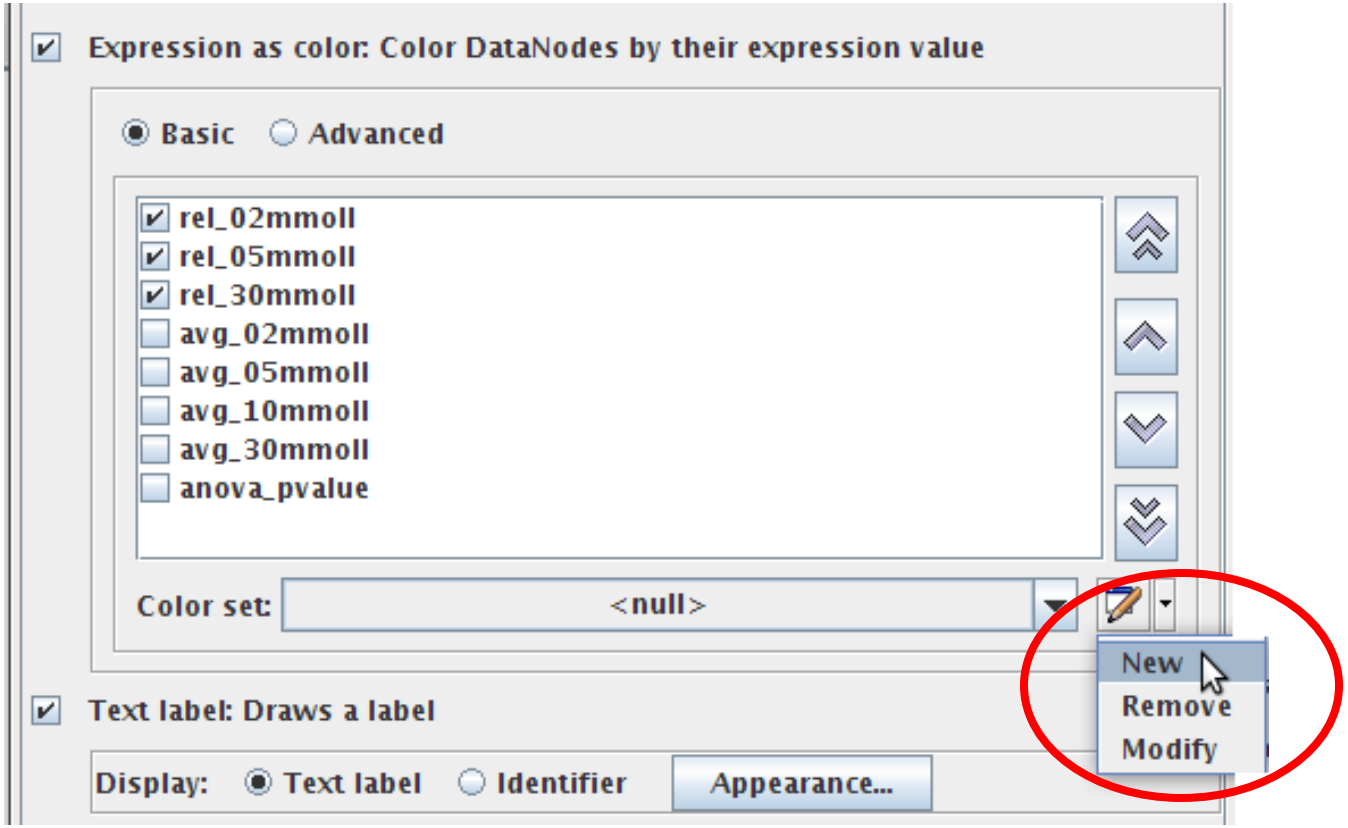




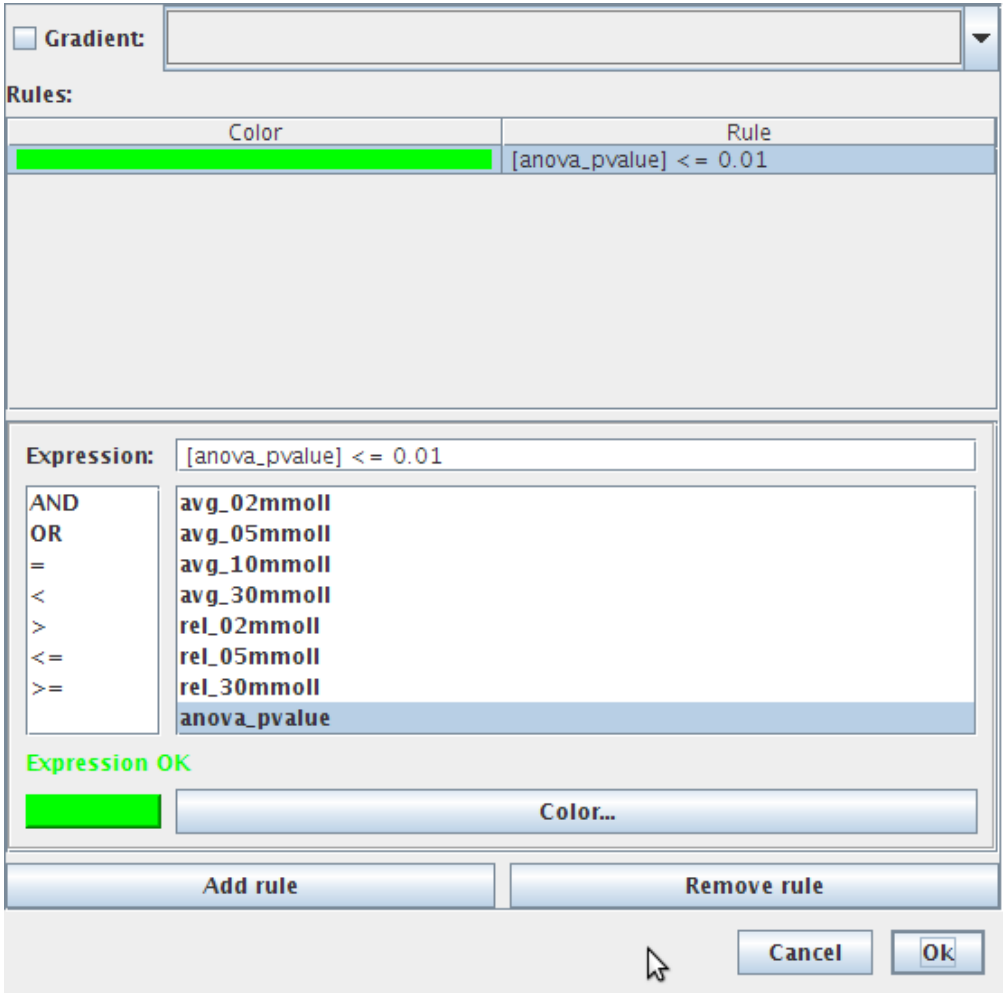
Visualizations

- Gradient based
  - Fold-change
- Rule based
  - Significant genes

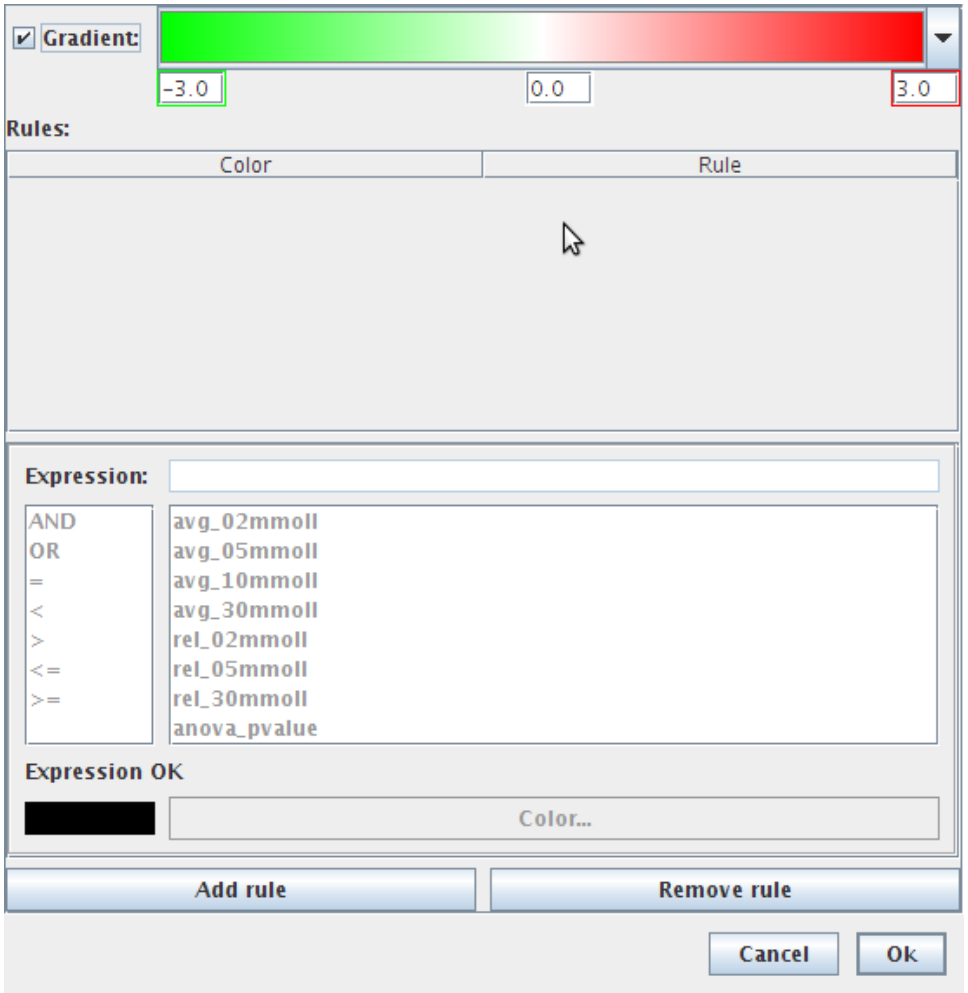
Color by Data Values



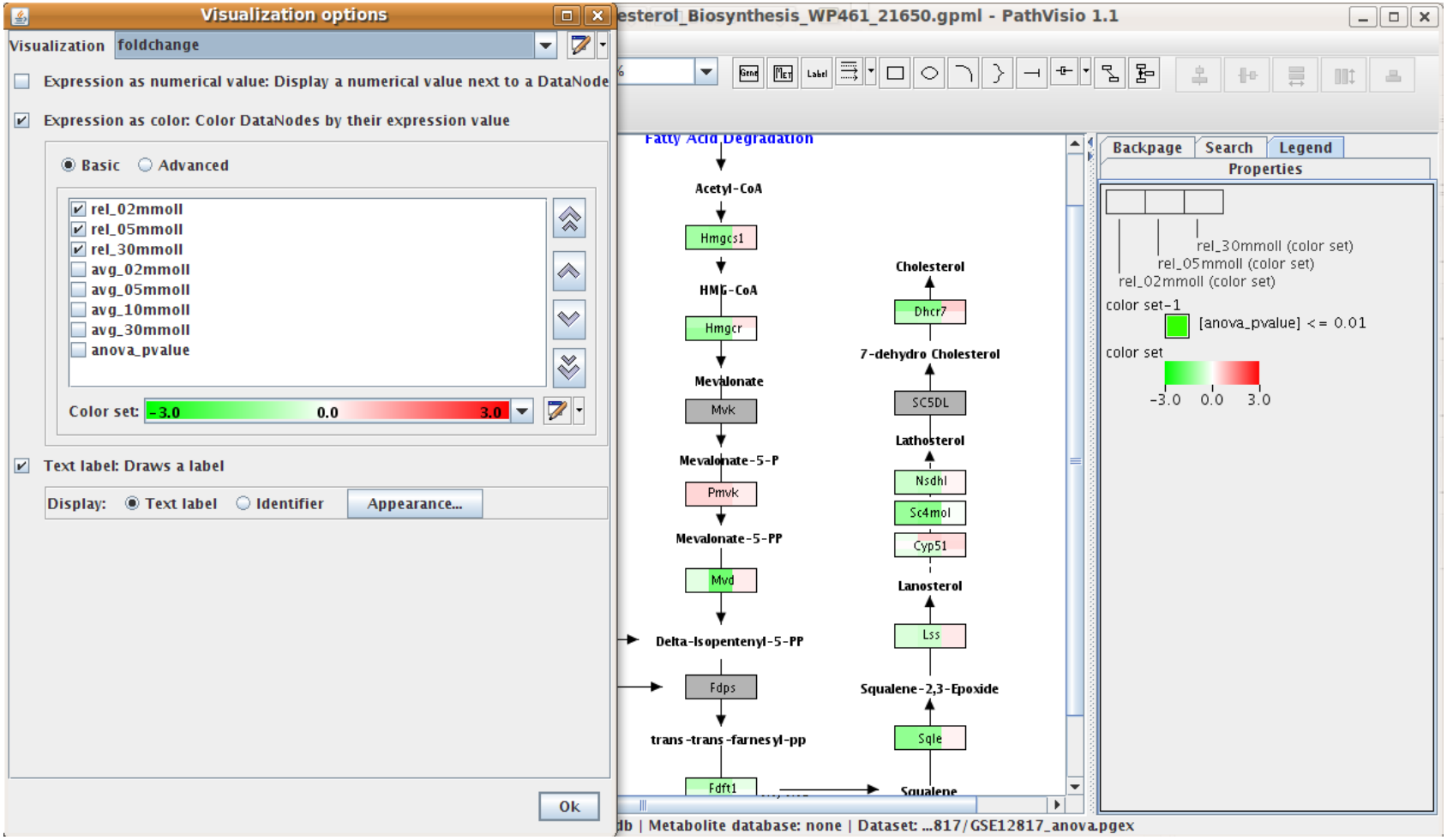
Color Set based on Criterion



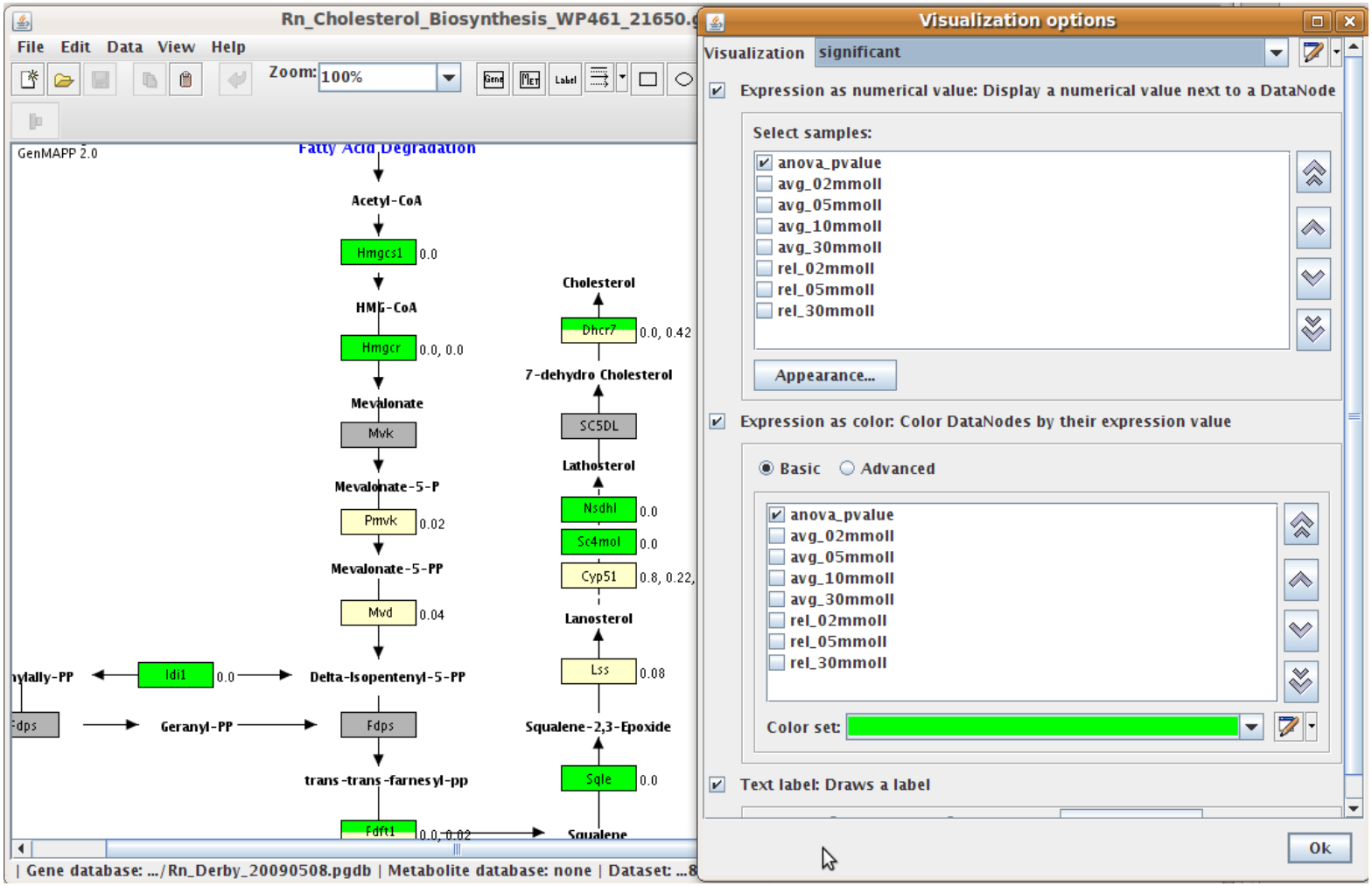
Color Set based on Gradient



Gradient based

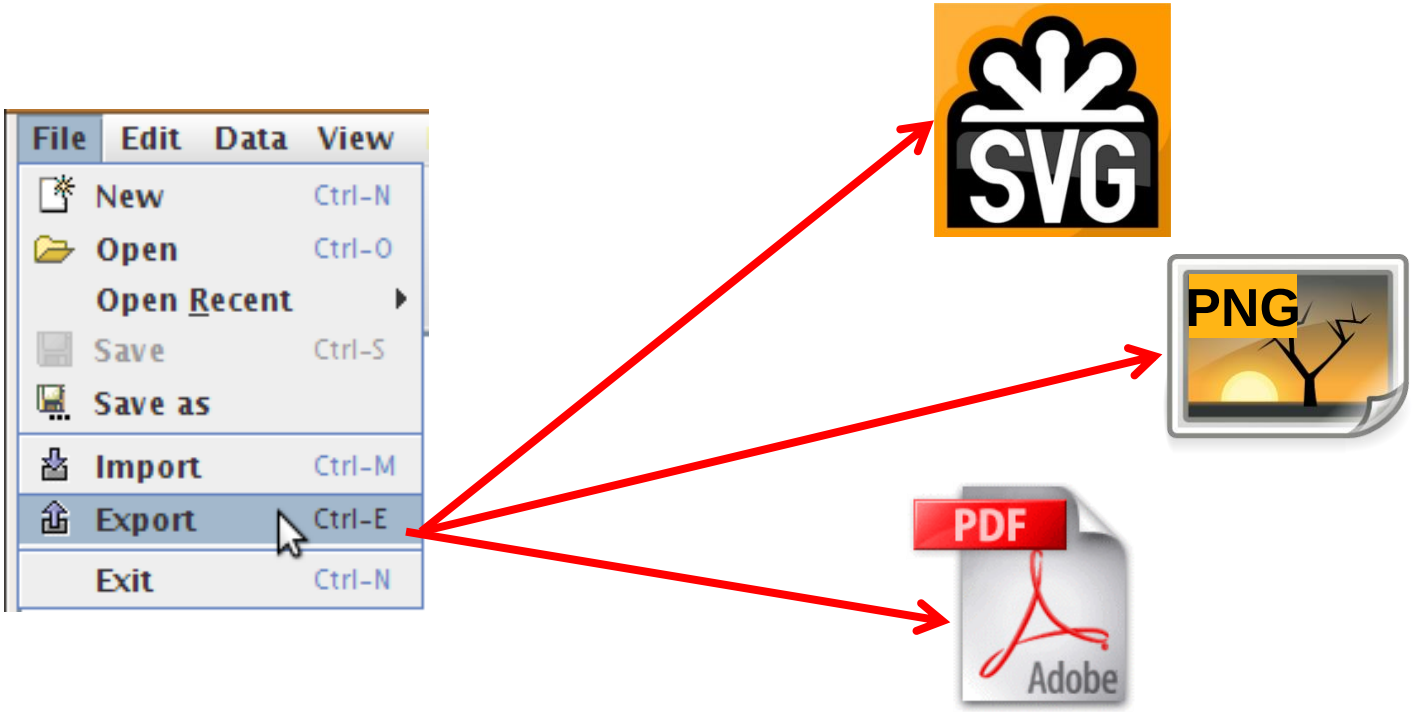


Rule based


Rn_Cholesterol_Biosynthesis_WP461_21650.
File Edit Data View Help
Zoom: 100%
GenMAPP 2.0
Fatty Acid Degradation
Acetyl-CoA
Hmgcs1 0.0
HMG-CoA
Hmgcr 0.0, 0.0
Mevalonate
Mvk
Mevalonate-5-P
Pmvk 0.02
Mevalonate-5-PP
Mvd 0.04
Idi1 0.0
Delta-Isopentenyl-5-PP
Geranyl-PP
Fdps
trans-trans-farnesyl-pp
Fdft1 0.0, 0.02
Squalene
Sqle 0.0
Squalene-2,3-Epoxide
Lss 0.08
Lanosterol
Cyp51 0.8, 0.22,
Sc4mol 0.0
Nsdhl 0.0
Lathosterol
SC5DL
7-dehydro Cholesterol
Dhcr7 0.0, 0.42
Cholesterol
| Gene database: .../Rn_Derby_20090508.pgdb | Metabolite database: none | Dataset: ...8
Visualization options
Visualization significant
Expression as numerical value: Display a numerical value next to a DataNode
Select samples:
anova_pvalue
avg_02mmoll
avg_05mmoll
avg_10mmoll
avg_30mmoll
rel_02mmoll
rel_05mmoll
rel_30mmoll
Appearance...
Expression as color: Color DataNodes by their expression value
Basic Advanced
anova_pvalue
avg_02mmoll
avg_05mmoll
avg_10mmoll
avg_30mmoll
rel_02mmoll
rel_05mmoll
rel_30mmoll
Color set:
Text label: Draws a label
Ok

5. Export Pathway

- Export to image formats



Which pathways do you expect to change when cells/tissues are exposed to nanoparticles?

- Oxidative stress
- DNA reparation
- Apoptosis
- Necrosis
- Fibrosis
- Inflammation
- Cytoskeleton
- Metal homeostasis
- Cholesterol metabolism

6. Limitations and pitfalls

- Tissue – cells specific gene expression
  - PathVisio tissue analyzer
- Snap shot
  - RNA half-life
  - Transcription onset
  - Sample preparation

- Dependent on database (+) or (-)

We want you to know:

- Know the basic biological molecules DNA-RNA-Protein and how they interact
- Have an idea about basic protein reactions (conversion of metabolites, signaling)
- Know about the most commonly investigated effects of nanoparticles *in vitro* and *in vivo*
- Recognize these effects in biological pathways
  - Know that typical affected pathways are oxidative stress, apoptosis, metal ion response
- Know about the variety of omics data and how to use it
- Have heard about a variety of tools and methods to assess the effects on pathway/system level: omics data, especially microarray and RNA-seq
- Know the limitations and pitfalls of omics data/systems biology analysis
- Know why databases are useful for biologic research
  - Remembers some of the databases for single entities and pathways
- Know about the basics of semantic web and data integration using ontologies
- Have heard about data repositories like ArrayExpress, GEO, eNanoMapper

Acknowledgements

Thanks for slides and support:

Chris T. Evelo

Egon Willighagen

Susan Coort

Lars Eijssen

Martina Summer-Kutmon

Andra Waagmeester