misvik biology

Karolinska
Institutet

Bioinformatics Analysis of Nano-based Omics Data

Penny Nymark, Pekka Kohonen, Vesa Hongisto and Roland Grafström

Hands-on Workshop on Nano Safety Assessment, 10th February, 2016,
Technology Park, Basel

ENM
eNanoMapper

Overview of workshop session, 10th February, Basel

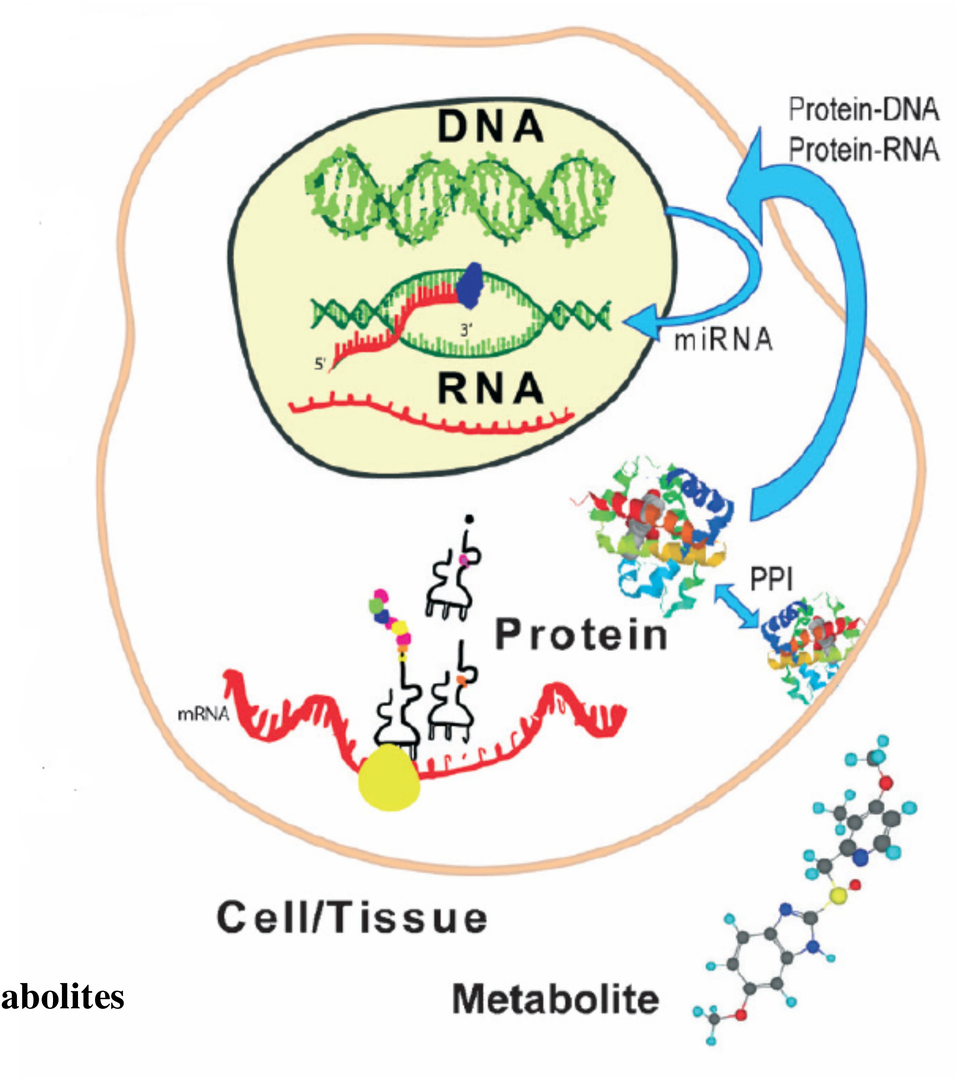
- Short background presentation (15-20 minutes)
  - → Intro to omics data
  - → Intro to nano-omics data
  - → Intro to Chipster

- Hands-on interpretation of the data directly in Chipster (follow tutorial – 20-25 minutes)
  - → ”Clicking” in a ready made session through a *guest* account at UPPMAX
  - → Go through the tutorial
  - → Discuss further interpretations

- Discussion (5-10 minutes)

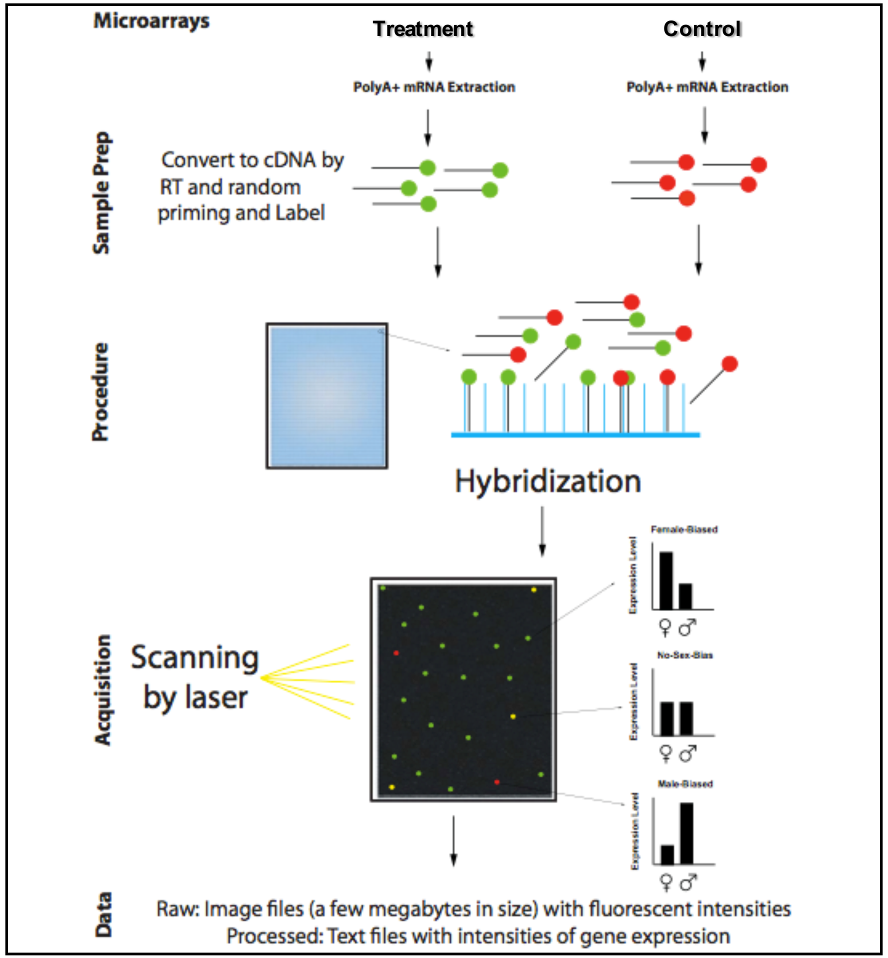
Different ‘omes and technologies to assess them

Technology	Biological event
SNP arrays, WG-seq, aCGH	Copy number, LOH, Mutations, SNP
ChIP-Chip, ChIP-seq	DNA-protein interaction
Gene expression microarrays, RNA-seq, miR-seq	mRNA expression miRNA expression
Antibody or lysate arrays, 2D-gels, Mass spec	Protein expression
Yeast 2 hybrid, Lysate arrays	Protein-protein interaction (PPI)
Tissue microarrays	Protein expression and localization
Metabolomics, NMR	Quantification of metabolites

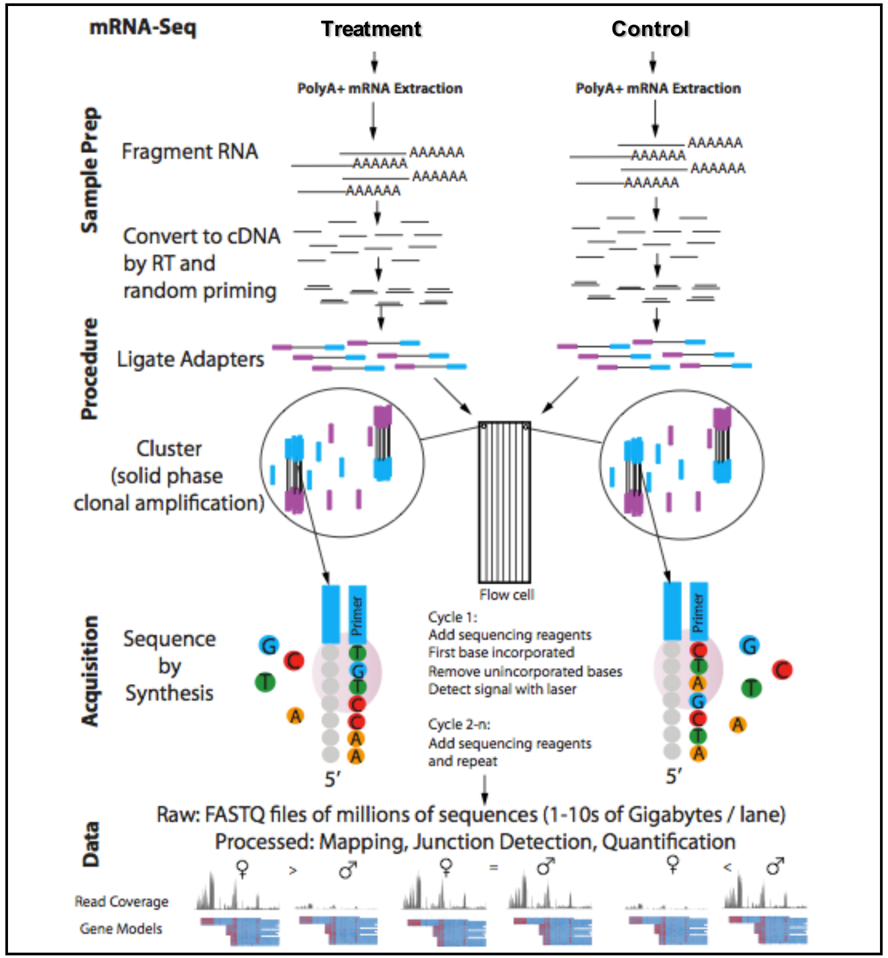


Adapted from Culhane et al, Cell Mol Life Sci, 64:3185-3200, 2007

Comparison of microarrays and RNA-sequencing technologies



Raw: Image files (a few megabytes in size) with fluorescent intensities
Processed: Text files with intensities of gene expression



Raw: FASTQ files of millions of sequences (1-10s of Gigabytes / lane)
Processed: Mapping, Junction Detection, Quantification

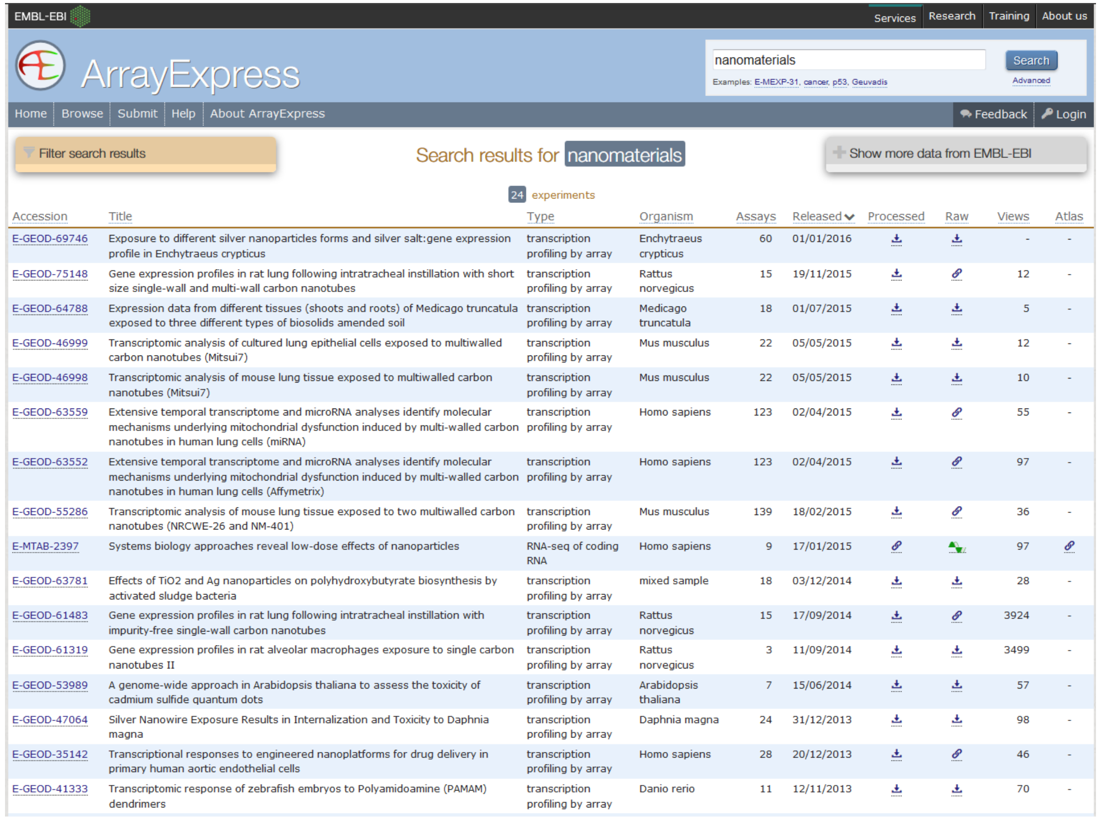
Microarrays are a legacy technology and many data sets are available for it.

Sequencing has greater sensitivity and provides more information.

Both technologies measure the transcriptome.
Hybridization and sequencing based methods produce similar results.

Existing omics data – including nano-specific

Two repositories recommended (often required) by <u>major scientific journals</u> to archive functional genomics data to support reproducible research



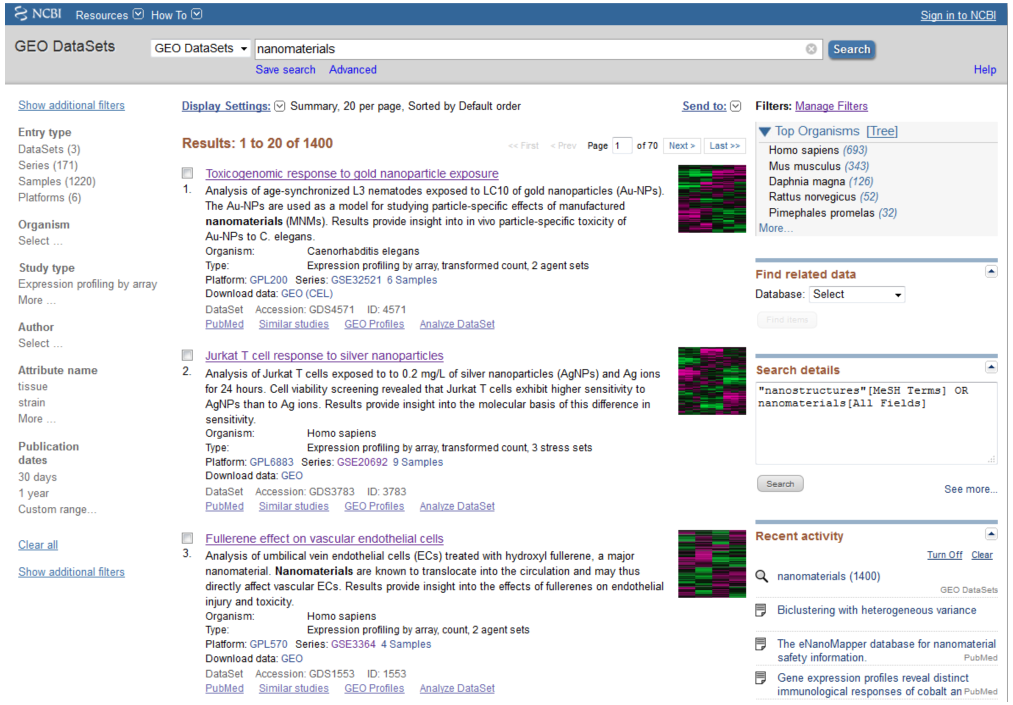
Search results for nanomaterials

24 experiments

Accession	Title	Type	Organism	Assays	Released	Processed	Raw	Views	Atlas
E-GEOD-69746	Exposure to different silver nanoparticles forms and silver salt:gene expression profile in Enchytraeus crypticus	transcription profiling by array	Enchytraeus crypticus	60	01/01/2016	⤓	⤓	-	-
E-GEOD-75148	Gene expression profiles in rat lung following intratracheal instillation with short size single-wall and multi-wall carbon nanotubes	transcription profiling by array	Rattus norvegicus	15	19/11/2015	⤓	🔗	12	-
E-GEOD-64788	Expression data from different tissues (shoots and roots) of Medicago truncatula exposed to three different types of biosolids amended soil	transcription profiling by array	Medicago truncatula	18	01/07/2015	⤓	⤓	5	-
E-GEOD-46999	Transcriptomic analysis of cultured lung epithelial cells exposed to multiwalled carbon nanotubes (Mitsui7)	transcription profiling by array	Mus musculus	22	05/05/2015	⤓	⤓	12	-
E-GEOD-46998	Transcriptomic analysis of mouse lung tissue exposed to multiwalled carbon nanotubes (Mitsui7)	transcription profiling by array	Mus musculus	22	05/05/2015	⤓	⤓	10	-
E-GEOD-63559	Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells (miRNA)	transcription profiling by array	Homo sapiens	123	02/04/2015	⤓	🔗	55	-
E-GEOD-63552	Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells (Affymetrix)	transcription profiling by array	Homo sapiens	123	02/04/2015	⤓	🔗	97	-
E-GEOD-55286	Transcriptomic analysis of mouse lung tissue exposed to two multiwalled carbon nanotubes (NRCWE-26 and NM-401)	transcription profiling by array	Mus musculus	139	18/02/2015	⤓	🔗	36	-
E-MTAB-2397	Systems biology approaches reveal low-dose effects of nanoparticles	RNA-seq of coding RNA	Homo sapiens	9	17/01/2015	🔗	●	97	🔗
E-GEOD-63781	Effects of TiO2 and Ag nanoparticles on polyhydroxybutyrate biosynthesis by activated sludge bacteria	transcription profiling by array	mixed sample	18	03/12/2014	⤓	⤓	28	-
E-GEOD-61483	Gene expression profiles in rat lung following intratracheal instillation with impurity-free single-wall carbon nanotubes	transcription profiling by array	Rattus norvegicus	15	17/09/2014	⤓	🔗	3924	-
E-GEOD-61319	Gene expression profiles in rat alveolar macrophages exposure to single carbon nanotubes II	transcription profiling by array	Rattus norvegicus	3	11/09/2014	⤓	⤓	3499	-
E-GEOD-53989	A genome-wide approach in Arabidopsis thaliana to assess the toxicity of cadmium sulfide quantum dots	transcription profiling by array	Arabidopsis thaliana	7	15/06/2014	⤓	⤓	57	-
E-GEOD-47064	Silver Nanowire Exposure Results in Internalization and Toxicity to Daphnia magna	transcription profiling by array	Daphnia magna	24	31/12/2013	⤓	⤓	98	-
E-GEOD-35142	Transcriptional responses to engineered nanoplatforms for drug delivery in primary human aortic endothelial cells	transcription profiling by array	Homo sapiens	28	20/12/2013	⤓	🔗	46	-
E-GEOD-41333	Transcriptomic response of zebrafish embryos to Polyamidoamine (PAMAM) dendrimers	transcription profiling by array	Danio rerio	11	12/11/2013	⤓	⤓	70	-

- **63 223 experiments**
- **1 900 034 assays**
- **40.13 TB of archived data**

Gene Expression Omnibus



Summary, 20 per page, Sorted by Default order

Results: 1 to 20 of 1400

☐ Toxicogenomic response to gold nanoparticle exposure

1. Analysis of age-synchronized L3 nematodes exposed to LC10 of gold nanoparticles (Au-NPs). The Au-NPs are used as a model for studying particle-specific effects of manufactured **nanomaterials** (MNMs). Results provide insight into in vivo particle-specific toxicity of Au-NPs to C. elegans.
Organism: Caenorhabditis elegans
Type: Expression profiling by array, transformed count, 2 agent sets
Platform: GPL200 Series: GSE32521 6 Samples
Download data: GEO (CEL)
DataSet Accession: GDS4571 ID: 4571
PubMed Similar studies GEO Profiles Analyze DataSet

☐ Jurkat T cell response to silver nanoparticles

2. Analysis of Jurkat T cells exposed to to 0.2 mg/L of silver nanoparticles (AgNPs) and Ag ions for 24 hours. Cell viability screening revealed that Jurkat T cells exhibit higher sensitivity to AgNPs than to Ag ions. Results provide insight into the molecular basis of this difference in sensitivity.
Organism: Homo sapiens
Type: Expression profiling by array, transformed count, 3 stress sets
Platform: GPL6883 Series: GSE20692 9 Samples
Download data: GEO
DataSet Accession: GDS3783 ID: 3783
PubMed Similar studies GEO Profiles Analyze DataSet

☐ Fullerene effect on vascular endothelial cells

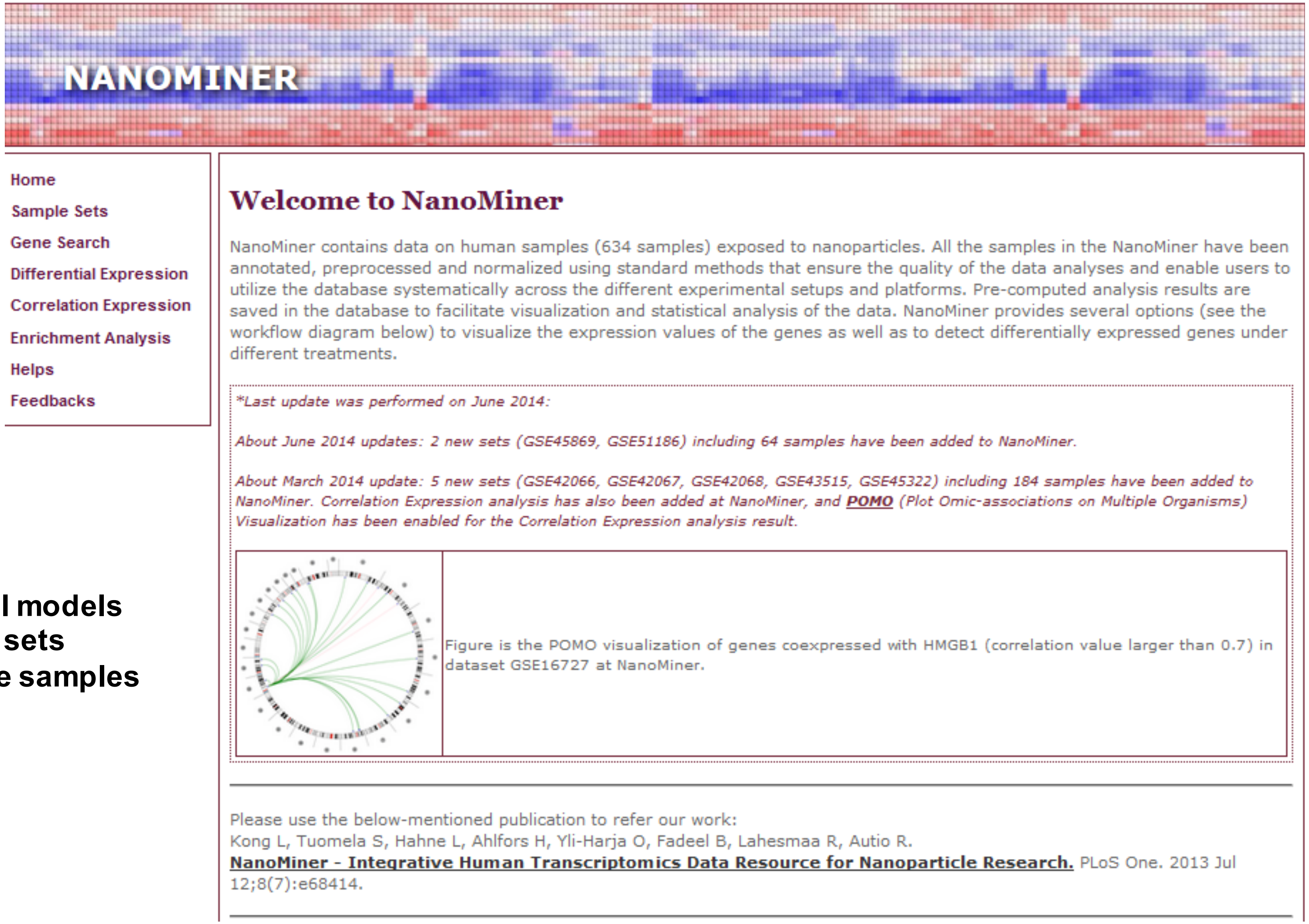
3. Analysis of umbilical vein endothelial cells (ECs) treated with hydroxyl fullerene, a major nanomaterial. **Nanomaterials** are known to translocate into the circulation and may thus directly affect vascular ECs. Results provide insight into the effects of fullerenes on endothelial injury and toxicity.
Organism: Homo sapiens
Type: Expression profiling by array, count, 2 agent sets
Platform: GPL570 Series: GSE3364 4 Samples
Download data: GEO
DataSet Accession: GDS1553 ID: 1553
PubMed Similar studies GEO Profiles Analyze DataSet

Search details:
```
"nanostructures"[MeSH Terms] OR
nanomaterials[All Fields]
```

Recent activity:
nanomaterials (1400)
Biclustering with heterogeneous variance
The eNanoMapper database for nanomaterial safety information.
Gene expression profiles reveal distinct immunological responses of cobalt an

- **DataSets: 3 848**
- **Series: 64 877**
- **Platforms: 15 365**
- **Samples: 1 671 418**

Nano-specific omics data gathered in one place



Welcome to NanoMiner

NanoMiner contains data on human samples (634 samples) exposed to nanoparticles. All the samples in the NanoMiner have been annotated, preprocessed and normalized using standard methods that ensure the quality of the data analyses and enable users to utilize the database systematically across the different experimental setups and platforms. Pre-computed analysis results are saved in the database to facilitate visualization and statistical analysis of the data. NanoMiner provides several options (see the workflow diagram below) to visualize the expression values of the genes as well as to detect differentially expressed genes under different treatments.

**Last update was performed on June 2014:*

About June 2014 updates: 2 new sets (GSE45869, GSE51186) including 64 samples have been added to NanoMiner.

*About March 2014 update: 5 new sets (GSE42066, GSE42067, GSE42068, GSE43515, GSE45322) including 184 samples have been added to NanoMiner. Correlation Expression analysis has also been added at NanoMiner, and **POMO** (Plot Omic-associations on Multiple Organisms) Visualization has been enabled for the Correlation Expression analysis result.*

Figure is the POMO visualization of genes coexpressed with HMGB1 (correlation value larger than 0.7) in dataset GSE16727 at NanoMiner.

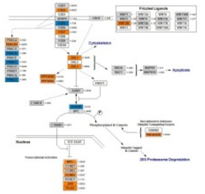
Please use the below-mentioned publication to refer our work:
Kong L, Tuomela S, Hahne L, Ahlfors H, Yli-Harja O, Fadeel B, Lahesmaa R, Autio R.
NanoMiner - Integrative Human Transcriptomics Data Resource for Nanoparticle Research. PLoS One. 2013 Jul 12;8(7):e68414.

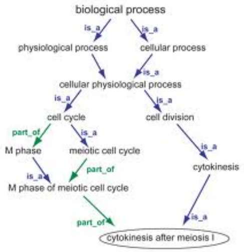
human cell models
28 sample sets
634 unique samples

“Omics”-Analysis Workflow

→

→





I. Experimental design

Frame a biological question

↓

Choose an omics platform

↓

Decide on biological replicates

↓

Samples for profiling

II. Pre-processing and initial statistical analyses

Quality control of raw data

↓

Assessment of test cases

↓

Normalization

↓

Evaluation of the data

↓

Differential expression

↓

Significant genes/proteins

III. Computational interpretation

Similar gene patterns

↓

Over-represented categories

↓

Networks of highly connected genes

↓

Biological interpretation

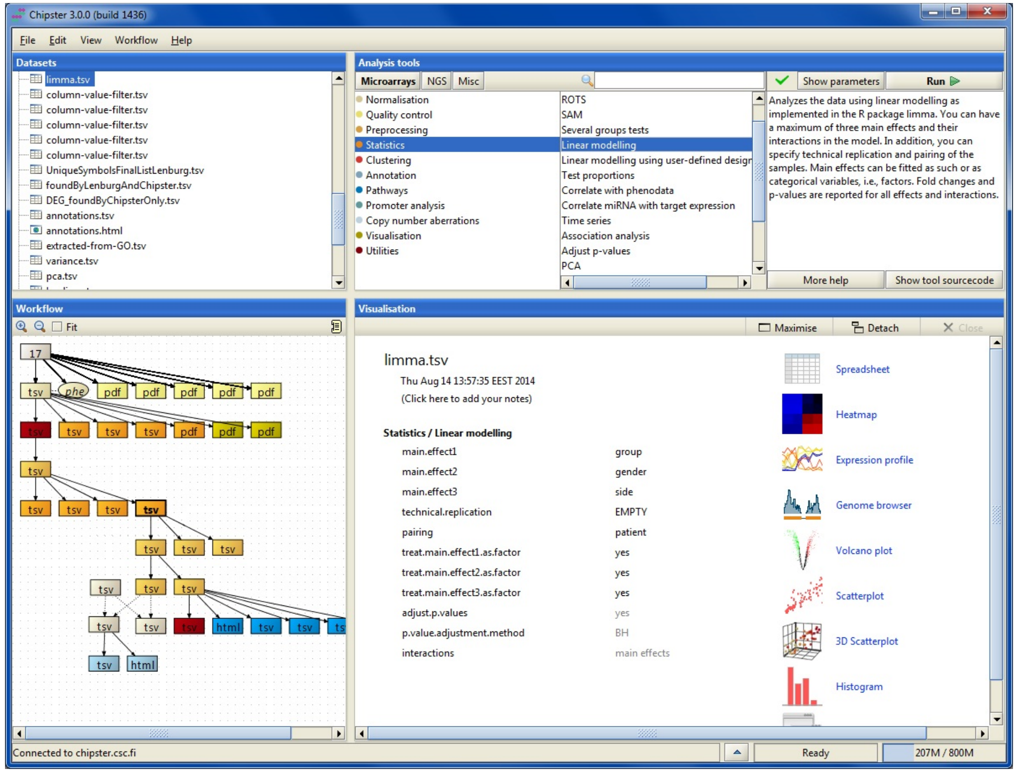
Chipster

Open source platform for data analysis

http://chipster.csc.fi

- Developed in 2008 by the Finnish IT Centre for Science
  - → Updated frequently
  - → Strong user support

- Increased use lately among researchers around the world
  - → Finland – Finnish IT Centre for Science
  - → Sweden - Uppsala University
  - → The Netherlands - Dutch TechCentre for Life Sciences
  - → Germany - DKFZ German Cancer Research Centre
  - → ELIXIR-project provides courses
  - → EGI-project will soon provide cloud services for all European researchers



Analyzes the data using linear modelling as implemented in the R package limma. You can have a maximum of three main effects and their interactions in the model. In addition, you can specify technical replication and pairing of the samples. Main effects can be fitted as such or as categorical variables, i.e., factors. Fold changes and p-values are reported for all effects and interactions.

limma.tsv

Thu Aug 14 13:57:35 EEST 2014

(Click here to add your notes)

Statistics / Linear modelling

main.effect1	group
main.effect2	gender
main.effect3	side
technical.replication	EMPTY
pairing	patient
treat.main.effect1.as.factor	yes
treat.main.effect2.as.factor	yes
treat.main.effect3.as.factor	yes
adjust.p.values	yes
p.value.adjustment.method	BH
interactions	main effects

Open source, server installation packages available, http://chipster.sourceforge.net/

Chipster What is it?

User-friendly analysis software and workflow tool

- → Intuitive graphical user interface (GUI)
- → Provides easy access to over 350 analysis tools (R/Bioconductor)
- → No programming or command line experience required
- → Analysis steps taken can be saved as an automatic workflow, which can be shared

Free, open source software

Compatible with

- → Largely all types of microarray data
- → All types of NGS data (ChIP-seq, RNA-seq and miRNA-seq, CNA-seq)

Chipster Goals

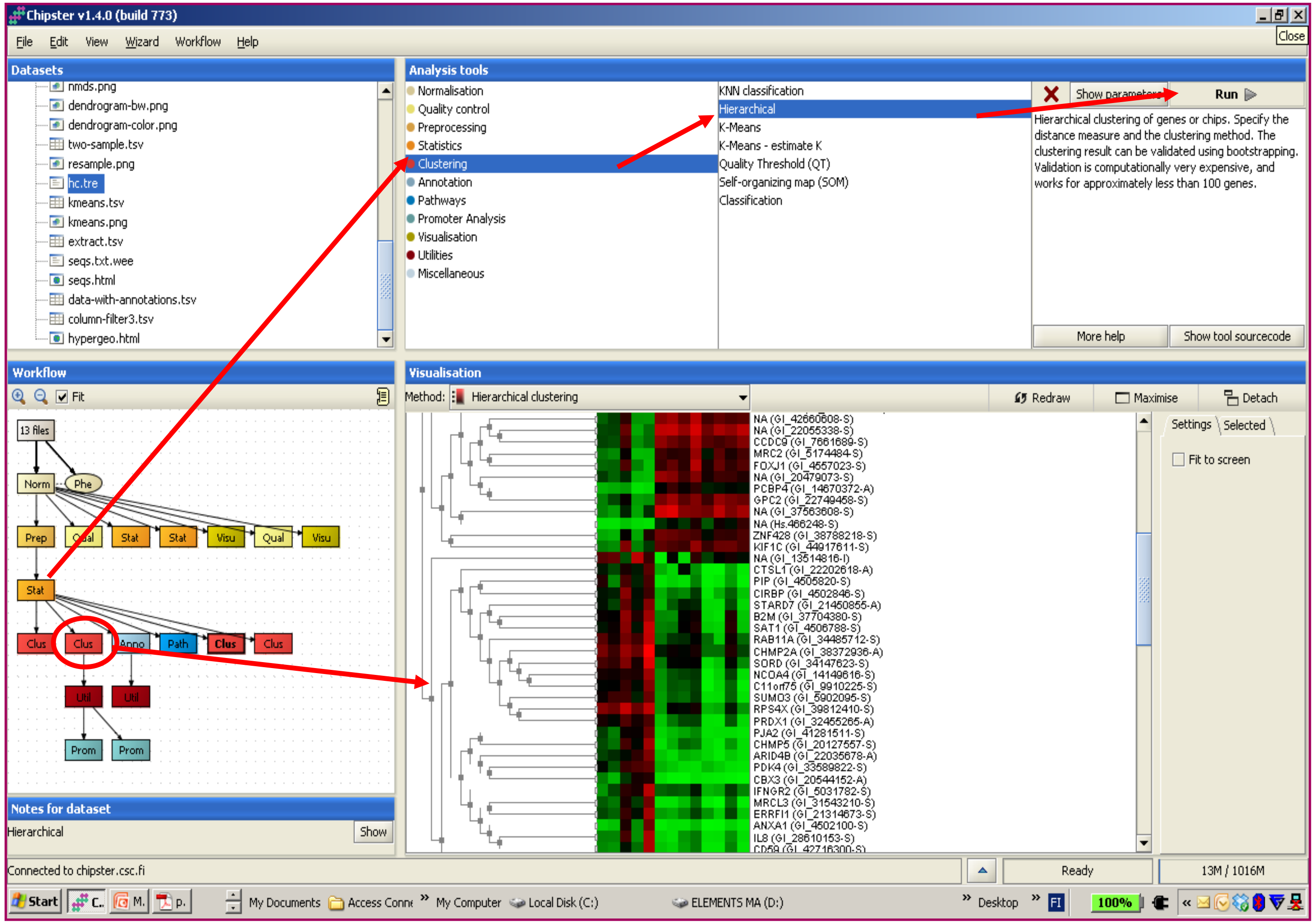
Enable researchers without programming skills or extensive bioinformatics knowledge to:

- access an extensive selection of up-to-date tools for high-throughput data analysis
- work with the data through a graphical and intuitive user interface
- combine tools into automatic workflows that can be shared
- integrate different types of data and analysis workflows
- interpret results in meaningful and efficient visualizations

Chipster How does it look?

- **Select data**
- **Select tool category**
- **Select tool**
- **Click run**

- **View the results by double-clicking**



Hierarchical clustering of genes or chips. Specify the distance measure and the clustering method. The clustering result can be validated using bootstrapping. Validation is computationally very expensive, and works for approximately less than 100 genes.

Method: Hierarchical clustering

Notes for dataset: Hierarchical

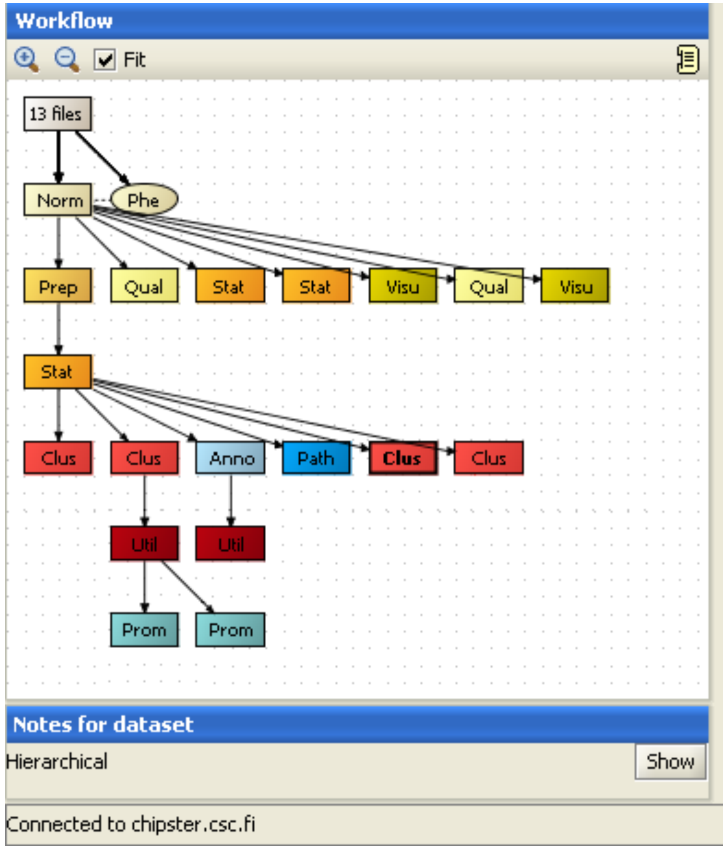
Connected to chipster.csc.fi

Workflow view – analysis session

- Shows the relationship of the data sets
- In order to continue working later on, you can save the analysis session

  → The session file is saved on your own computer, but you can also take it with you and continue on another computer by simple copying. Sessions can also be shared with other colleagues

- A workflow can also be automated
- You can of course also save multiple sessions separately

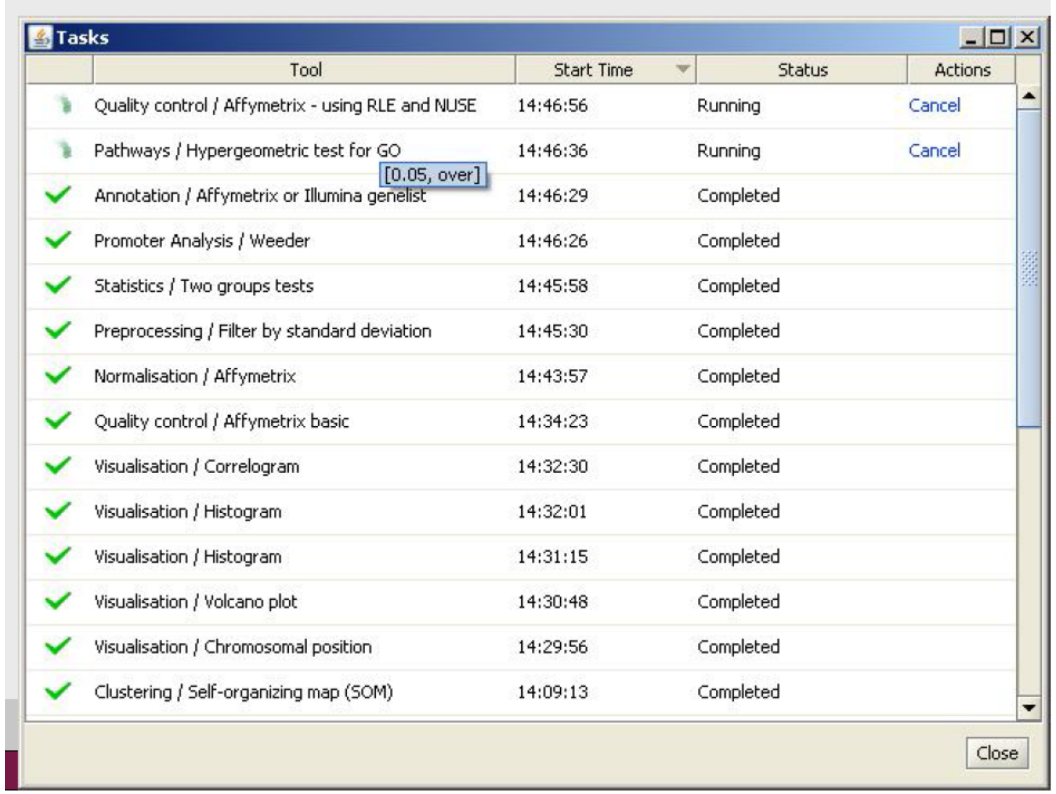


Hierarchical

Connected to chipster.csc.fi

Running many analyses simultaneously

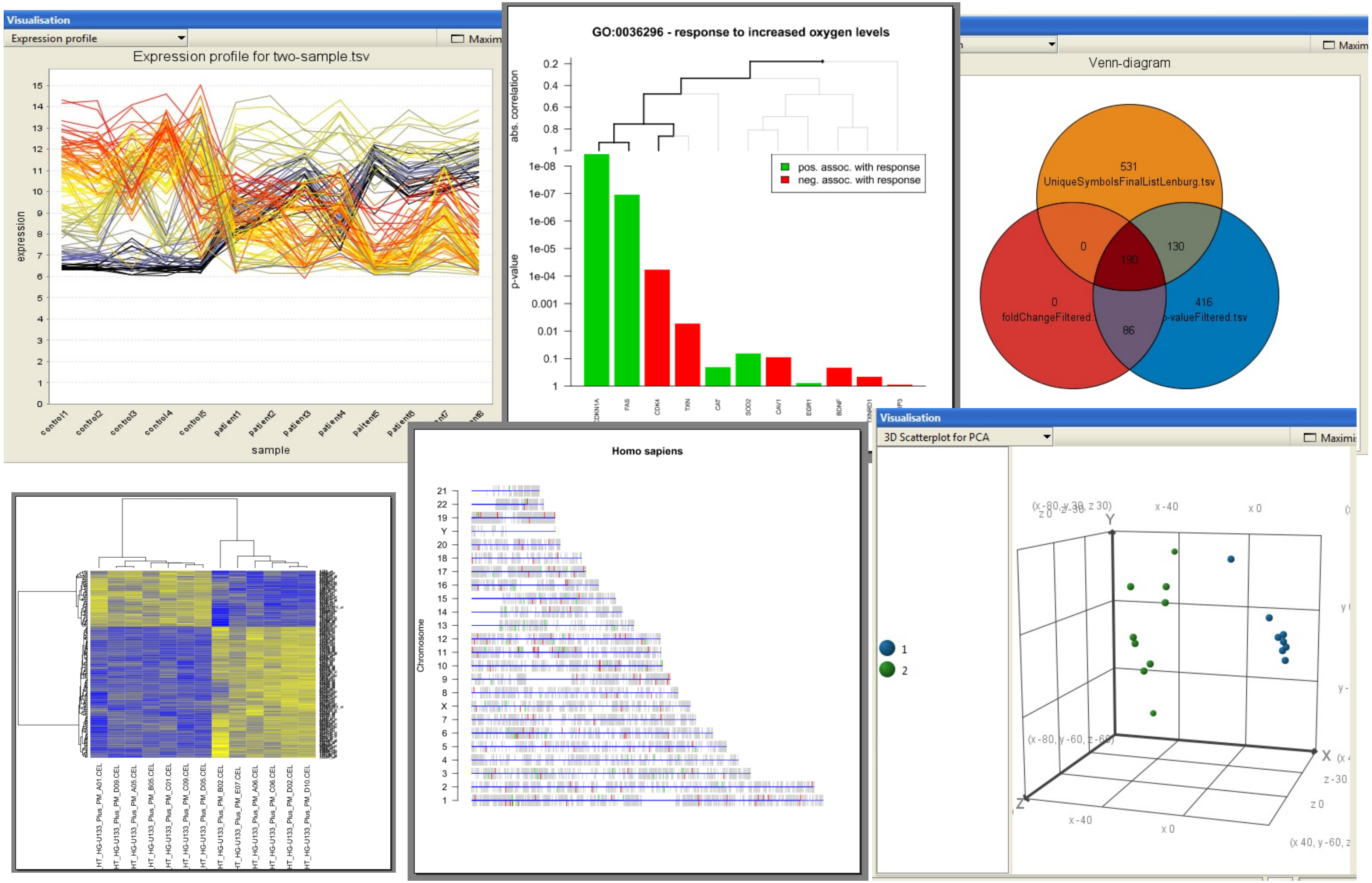
- You can have 10 analysis jobs running at the same time
- With the Task manager you can
  - → View the status
  - → Cancel jobs
  - → View time, parameters



	Tool	Start Time	Status	Actions
	Quality control / Affymetrix - using RLE and NUSE	14:46:56	Running	Cancel
	Pathways / Hypergeometric test for GO	14:46:36	Running	Cancel
✓	Annotation / Affymetrix or Illumina genelist	14:46:29	Completed	
✓	Promoter Analysis / Weeder	14:46:26	Completed	
✓	Statistics / Two groups tests	14:45:58	Completed	
✓	Preprocessing / Filter by standard deviation	14:45:30	Completed	
✓	Normalisation / Affymetrix	14:43:57	Completed	
✓	Quality control / Affymetrix basic	14:34:23	Completed	
✓	Visualisation / Correlogram	14:32:30	Completed	
✓	Visualisation / Histogram	14:32:01	Completed	
✓	Visualisation / Histogram	14:31:15	Completed	
✓	Visualisation / Volcano plot	14:30:48	Completed	
✓	Visualisation / Chromosomal position	14:29:56	Completed	
✓	Clustering / Self-organizing map (SOM)	14:09:13	Completed	

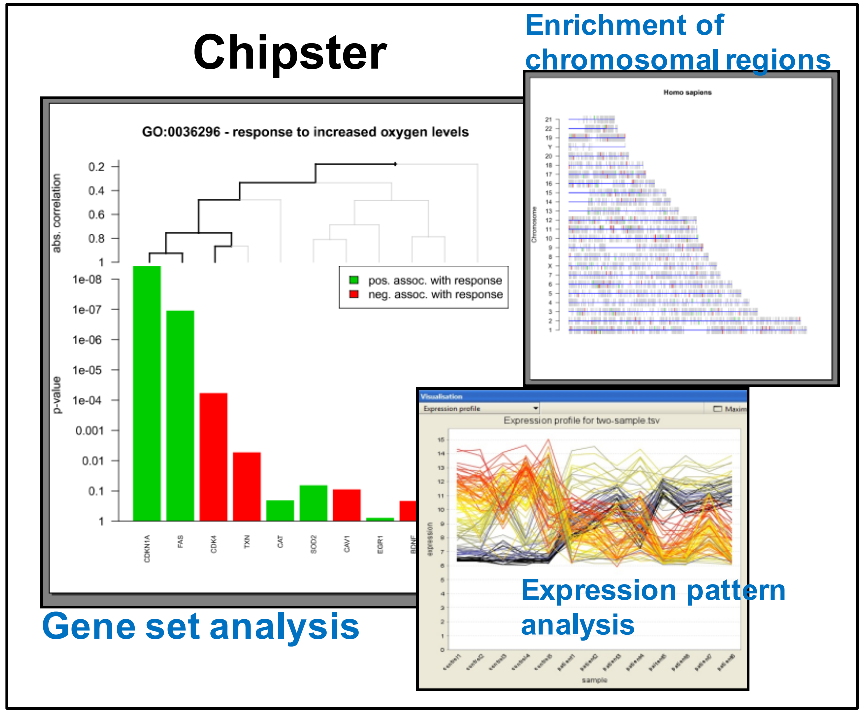
[0.05, over]

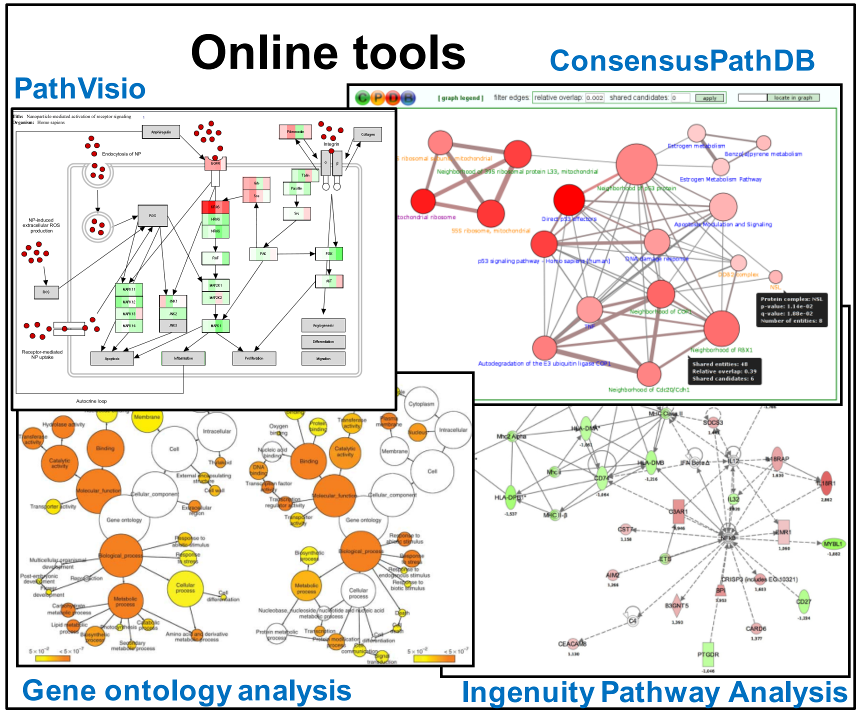
Chipster Visualizations



Clustering, Pathway analysis, Visualization

- Interpreting the results is the most complex and the most laborious part of bioinformatics and usually includes testing several tools (including online ones)



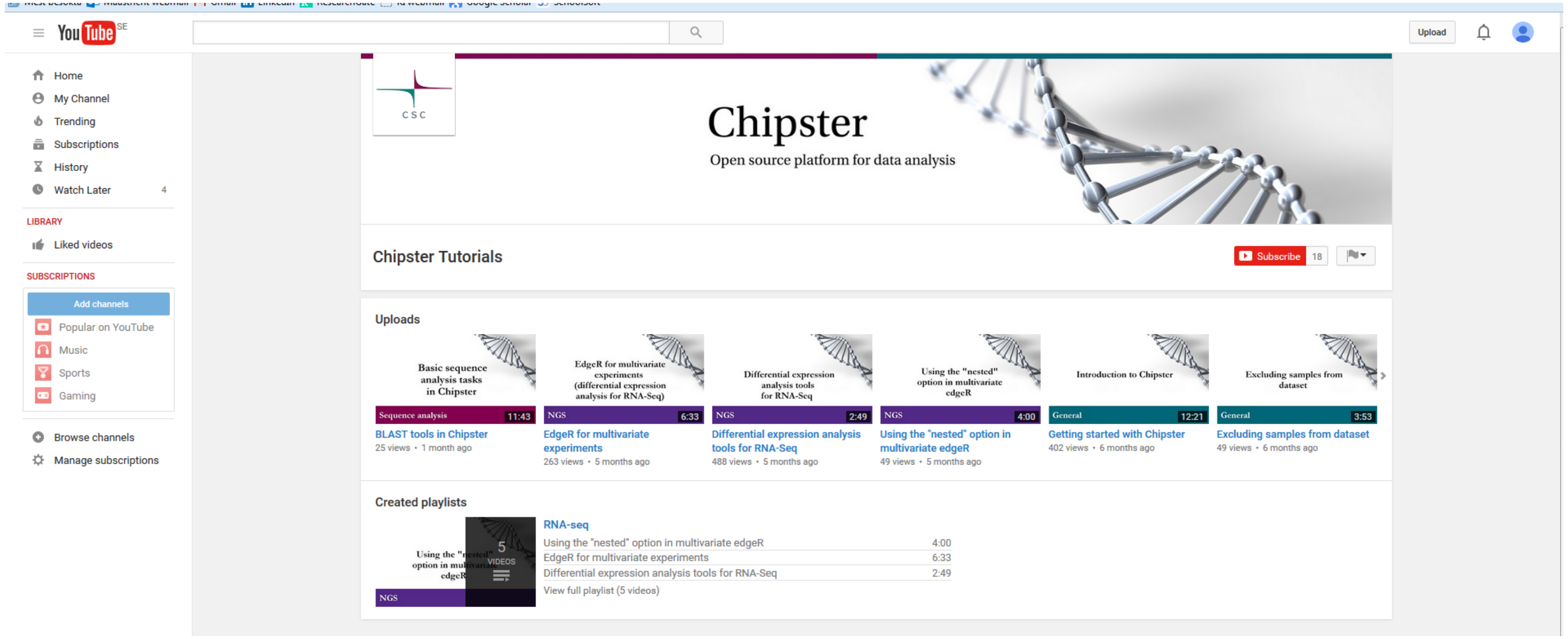


Examples of online tools useful for Pathway/Gene Ontology Analysis

- **Ingenuity Pathway Analysis (IPA)**
- **ConsensusPathDB (CPDB)**
- **Enrichr**
- **Reactome**
- **PathVisio**

Chipster tutorials

- http://chipster.csc.fi/
- https://www.youtube.com/channel/UCnL-Lx5gGIW01OkskZL7JEQ



Chipster Tutorials

Uploads

BLAST tools in Chipster
25 views • 1 month ago

EdgeR for multivariate experiments
263 views • 5 months ago

Differential expression analysis tools for RNA-Seq
488 views • 5 months ago

Using the "nested" option in multivariate edgeR
49 views • 5 months ago

Getting started with Chipster
402 views • 6 months ago

Excluding samples from dataset
49 views • 6 months ago

Created playlists

RNA-seq

Using the "nested" option in multivariate edgeR	4:00
EdgeR for multivariate experiments	6:33
Differential expression analysis tools for RNA-Seq	2:49

View full playlist (5 videos)

Hands-on session
Looking at a nano-related data set - GSE42067

Experimental set up

Model

- Human small airway epithelial cells (SAE)

Nanomaterials

- MWCNTs (Cheaptubes)
- TiO_2 nanobelts

Doses

- 10 µg/ml
- 100 µg/ml

Time

- 1 h
- 24 h

Workflow

- Import into Chipster
- Quality control
- Normalize
- Annotate samples
- Preprocess/Filter data
- Visualize
- Identify differentially expressed genes
- Cluster and visualize
- Pathway analysis

Hands-on session conclusions

Original publication conclusions

Tilton et al, Nanotoxicology. 2014 Aug;8(5):533-48.
doi: 10.3109/17435390.2013.803624.

- Low toxicity by MWCNTs
- High toxicity by TiO_2
- Early (1h) response - common to both NP types
- Late (24 h) response - NP-specific
- NP-specific responses:
  - → MWCNT uniquely up-regulated cell proliferation, anti-apoptotic and DNA repair mechanisms associated with cell survival
  - → TiO_2-NB differentially regulated inflammatory responses associated with cellular stress

Participant conclusions?