

**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, 29th September, 2016,
National Technical University of Athens, Athens, Greece

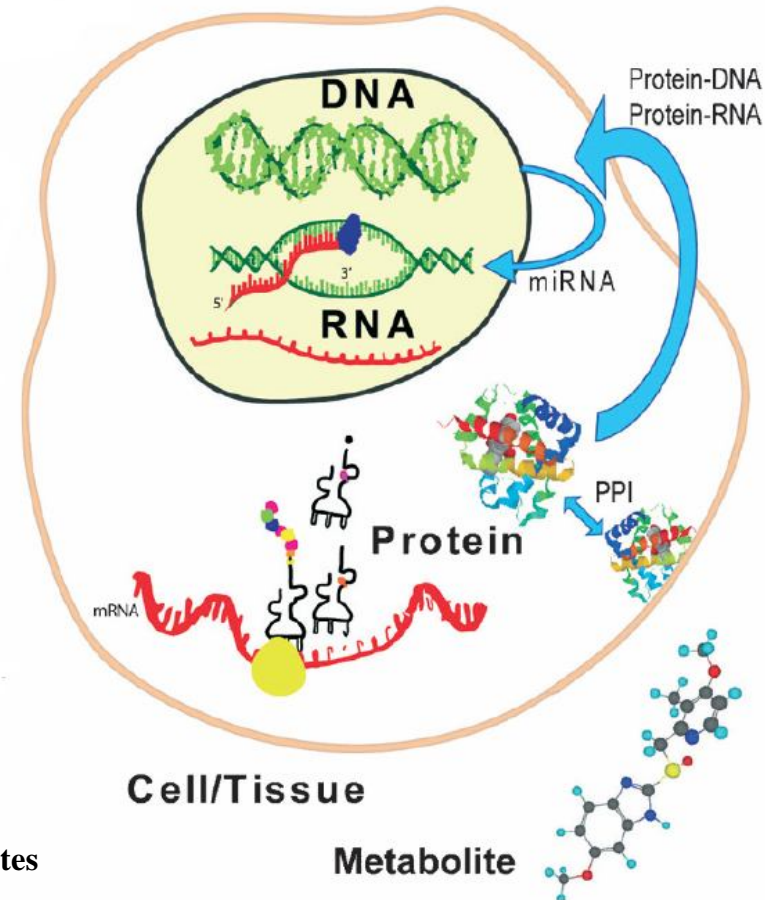


Overview of workshop session, 29th September, Athens

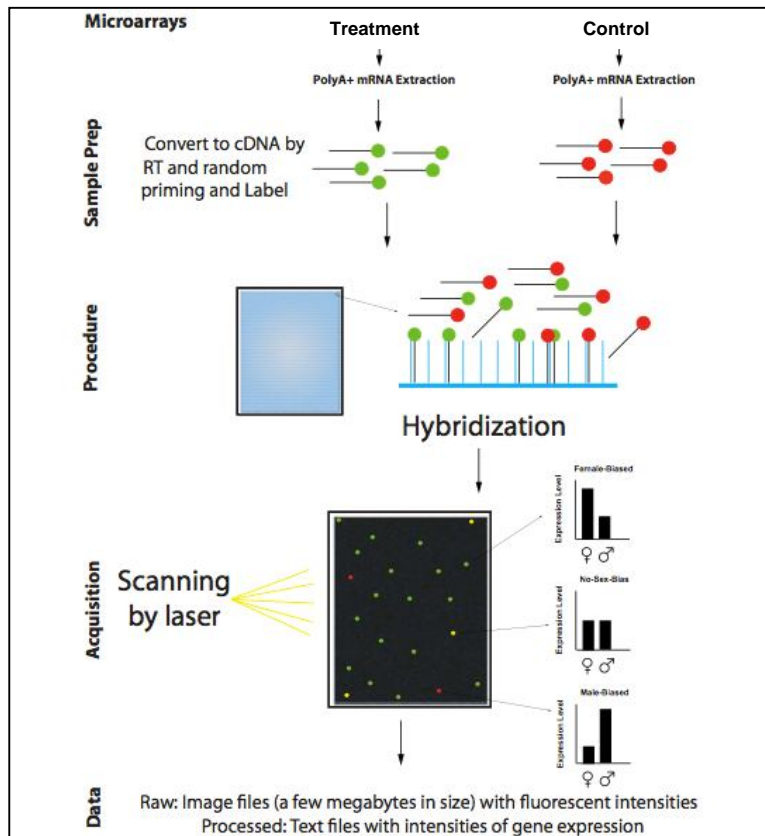
- Short background presentation (30 minutes including questions)
 - Intro to omics data
 - Intro to nano-omics data
 - Intro to Chipster
- Hands-on interpretation of the data directly in Chipster (follow tutorial – 60 minutes)
 - "Clicking" in a ready made session through a *guest* account at UPPMAX (20 min.)
 - Independent work according to the tutorial (30 min.)
 - Discuss further interpretations (10 min.)
- Further questions and discussion (30 minutes)

Different 'omes and technologies to assess them

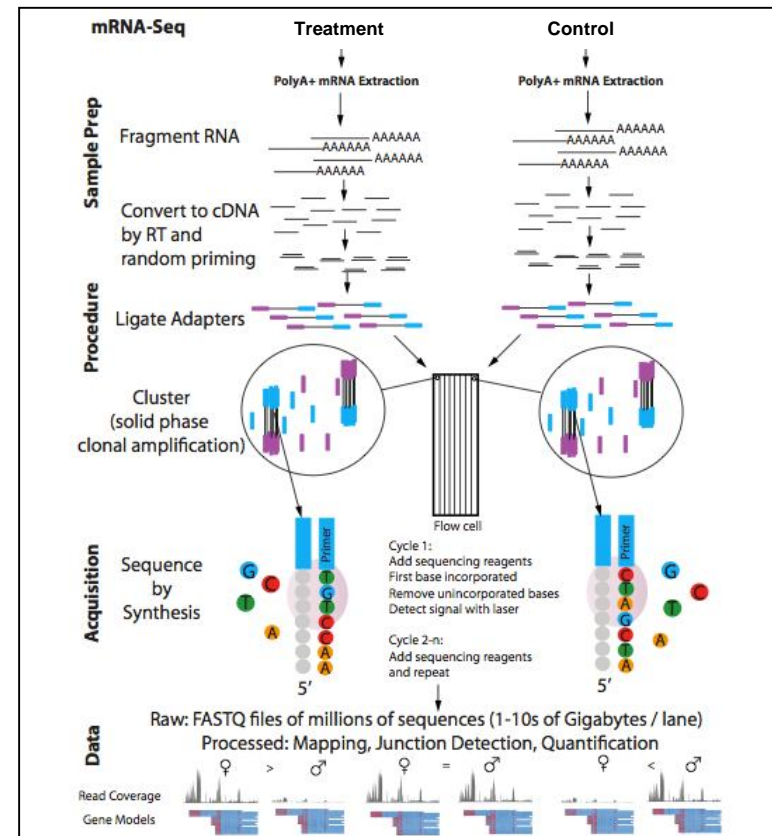
<u>Technology</u>	<u>Biological event</u>
↓	↓
SNP arrays, WG-seq, aCGH	Copy number, LOH, Mutations, SNP
ChIP-Chip, ChIP-seq	DNA-protein interaction
Gene expression microarrays, RNA-seq, miR-seq	mRNA expression miRNA expression
Antibody or lysate arrays, 2D-gels, Mass spec	Protein expression
Yeast 2 hybrid,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



Comparison of microarrays and RNA-sequencing technologies



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 major scientific journals to archive functional genomics data to support reproducible research

EMBL-EBI ArrayExpress 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 <i>Enchytraeus crypticus</i>	transcription profiling by array	<i>Enchytraeus crypticus</i>	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	<i>Rattus norvegicus</i>	15	19/11/2015	↓	↓	12	-
E-GEOD-64788	Expression data from different tissues (shoots and roots) of <i>Medicago truncatula</i> exposed to three different types of biosolids amended soil	transcription profiling by array	<i>Medicago truncatula</i>	18	01/07/2015	↓	↓	5	-
E-GEOD-46999	Transcriptomic analysis of cultured lung epithelial cells exposed to multiwalled carbon nanotubes (Mtsu7)	transcription profiling by array	<i>Mus musculus</i>	22	05/05/2015	↓	↓	12	-
E-GEOD-46998	Transcriptomic analysis of mouse lung tissue exposed to multiwalled carbon nanotubes (Mtsu7)	transcription profiling by array	<i>Mus musculus</i>	22	05/05/2015	↓	↓	10	-
E-GEOD-63359	Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells (mH4C)	transcription profiling by array	<i>Homo sapiens</i>	123	02/04/2015	↓	↓	35	-
E-GEOD-63352	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	<i>Homo sapiens</i>	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	<i>Mus musculus</i>	139	18/02/2015	↓	↓	36	-
E-MTAB-2397	Systems biology approaches reveal low-dose effects of nanoparticles	RNA-seq of coding RNA	<i>Homo sapiens</i>	9	17/01/2015	↓	↓	97	↓
E-GEOD-63781	Effects of TiO ₂ 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	<i>Rattus norvegicus</i>	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	<i>Rattus norvegicus</i>	3	11/09/2014	↓	↓	3499	-
E-GEOD-53989	A genome-wide approach in <i>Arabidopsis thaliana</i> to assess the toxicity of cadmium sulfide quantum dots	transcription profiling by array	<i>Arabidopsis thaliana</i>	7	15/06/2014	↓	↓	57	-
E-GEOD-47064	Silver Nanowire Exposure Results in Internalization and Toxicity to <i>Daphnia magna</i>	transcription profiling by array	<i>Daphnia magna</i>	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	<i>Homo sapiens</i>	28	20/12/2013	↓	↓	46	-
E-GEOD-41333	Transcriptomic response of zebrafish embryos to Polyamidoamine (PAMAM) dendrimers	transcription profiling by array	<i>Danio rerio</i>	11	12/11/2013	↓	↓	70	-

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



NCBI GEO DataSets Search results for **nanomaterials** (Results: 1 to 20 of 1400)

1. **Toxicogenomic response to gold nanoparticle exposure**
 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 (GCL)
 DataSet Accession: GDS4571 ID: 4571
 Published Similar studies GEO Profiles Analyze DataSet

2. **Jurkat T cell response to silver nanoparticles**
 Analysis of Jurkat T cells exposed 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
 Published Similar studies GEO Profiles Analyze DataSet

3. **Fullerene effect on vascular endothelial cells**
 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: GDS1563 ID: 1563
 Published Similar studies GEO Profiles Analyze DataSet

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

Nano-specific omics data gathered in one place

NANOMINER

- Home
- Sample Sets
- Gene Search
- Differential Expression
- Correlation Expression
- Enrichment Analysis
- Helps
- Feedbacks

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.

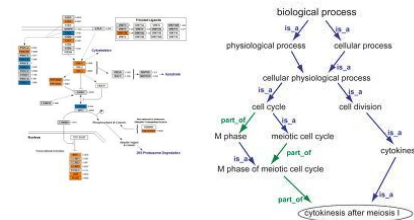
*About March 2014 update: 5 new sets (GSE42066, GSE42067, GSE42068, GSE43515, GSE45322) including 184 samples have been added to NanoMiner. Correlation Expression analysis has also been added at NanoMiner, and **POMO** (Plot Omic-associations on Multiple Organisms) Visualization has been enabled for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



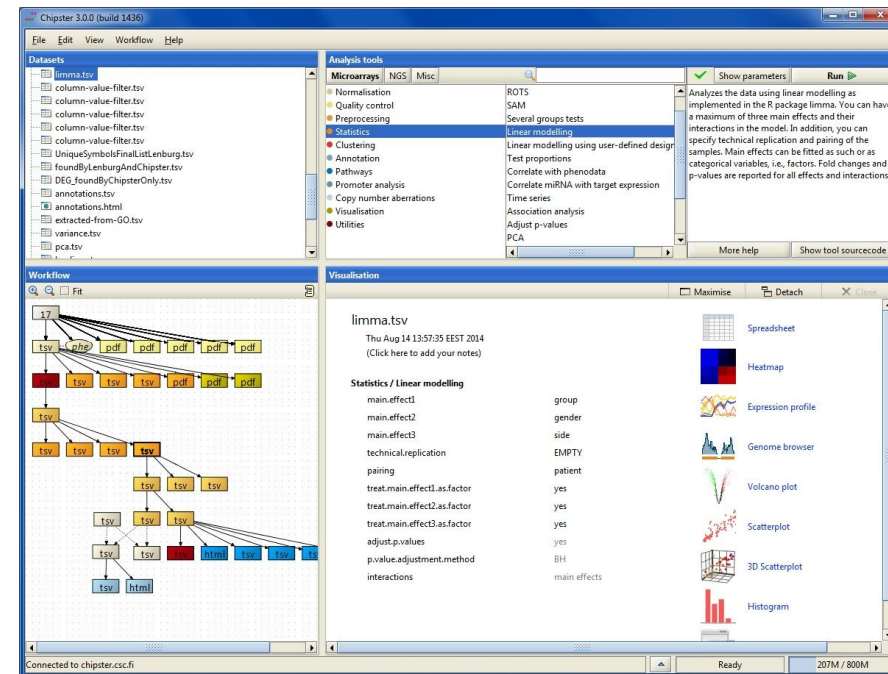
Chipster

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



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

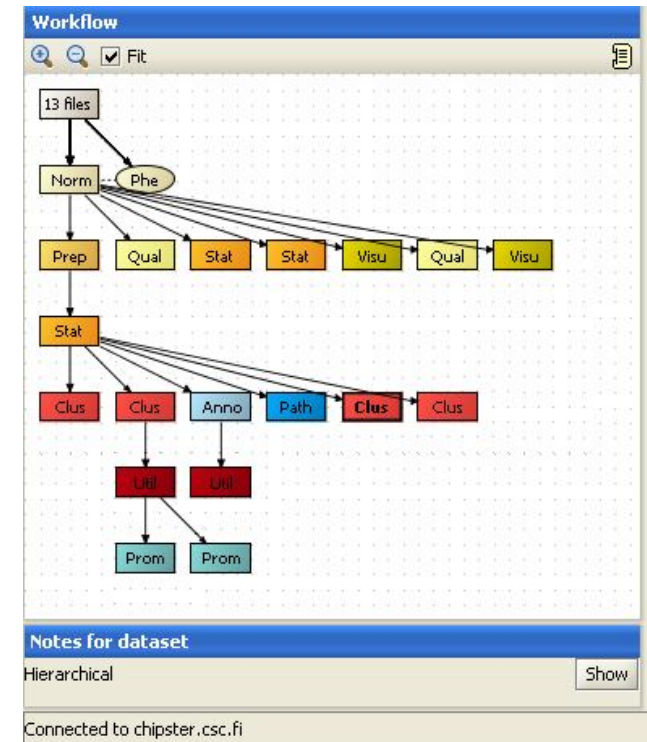
How does it look?

- Select data
- Select tool category
- Select tool
- Click run
- View the results by double-clicking

The screenshot displays the Chipster v1.4.0 (build 773) interface. The top-left panel, 'Datasets', lists various files including 'nmds.png', 'dendrogram-bw.png', 'dendrogram-color.png', 'two-sample.tsv', 'resample.png', 'hc.tre', 'kmeans.tsv', 'kmeans.png', 'extract.tsv', 'seqs.txt.wee', 'seqs.html', 'data-with-annotations.tsv', 'column-filter3.tsv', and 'hypergeo.html'. The top-right panel, 'Analysis tools', shows a tree view with categories like 'Normalisation', 'Quality control', 'Preprocessing', 'Statistics', 'Clustering', 'Annotation', 'Pathways', 'Promoter Analysis', 'Visualisation', 'Utilities', and 'Miscellaneous'. The 'Clustering' category is selected, and the 'Hierarchical' tool is chosen. A 'Run' button is visible next to it. The bottom-left panel, 'Workflow', shows a flowchart of the analysis process. A red circle highlights the 'Clus' step, and a red arrow points from it to the 'Visualisation' panel. The bottom-right panel, 'Visualisation', displays a heatmap with a dendrogram on the left. The heatmap shows a color scale from green to red, representing gene expression levels. The y-axis lists gene identifiers such as 'NA (GI_42660808-S)', 'NA (GI_22055338-S)', 'CCDC9 (GI_7661689-S)', 'MR22 (GI_5174494-S)', 'FDX1 (GI_4557023-S)', 'NA (GI_20479073-S)', 'PCBP4 (GI_14670372-A)', 'SPD2 (GI_22740496-S)', 'NA (GI_37563808-S)', 'NA (hs_486248-S)', 'ZNF428 (GI_38788218-S)', 'KIF1C (GI_44617811-S)', 'NA (GI_13814816-I)', 'CTSL1 (GI_22202818-A)', 'PIP (GI_4605520-S)', 'CIRBP (GI_4502846-S)', 'STAR7 (GI_21460855-A)', 'S2M (GI_37704830-S)', 'SAT1 (GI_4606788-S)', 'RAB11A (GI_34485712-S)', 'CHMP2A (GI_38372836-A)', 'SORB (GI_34747623-S)', 'NCOA4 (GI_14149816-S)', 'C11orf5 (GI_9910225-S)', 'SUN3 (GI_593206-S)', 'RPS4X (GI_36812410-S)', 'PRDX1 (GI_32465265-A)', 'PUJ2 (GI_41281511-S)', 'CHMP2 (GI_20127557-S)', 'ARID4B (GI_22036678-A)', 'PDHA4 (GI_3569822-S)', 'CBX3 (GI_205444162-A)', 'IFNGR2 (GI_5031782-S)', 'MR13 (GI_31543210-S)', 'ERRF1 (GI_121314873-S)', 'ANXA1 (GI_4602100-S)', 'IL8 (GI_28810153-S)', and 'CD59 (GI_47216300-S)'. The status bar at the bottom indicates 'Ready' and '13M / 1016M'.

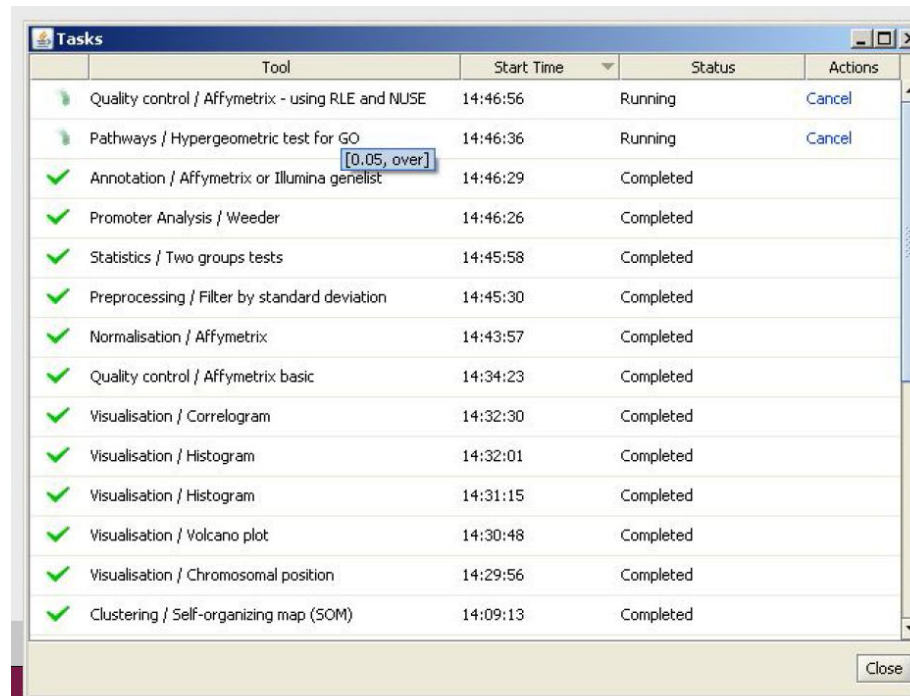
Workflow view – analysis session

- 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



Running many analyses simultaneously

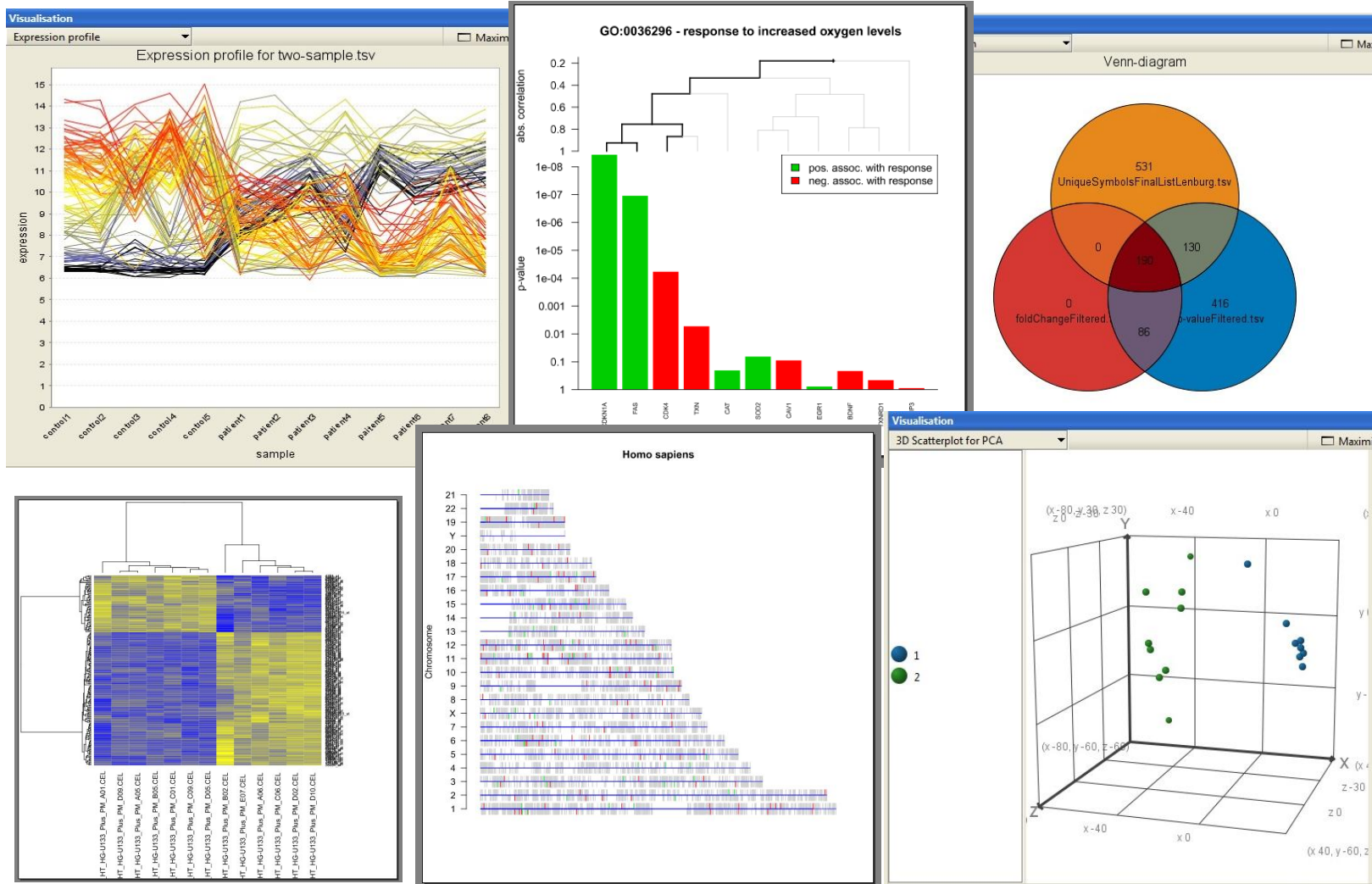
- You can have 10 analysis jobs running at the same time
- With the Task manager you can
 - View the status
 - Cancel jobs
 - View time, parameters



	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	

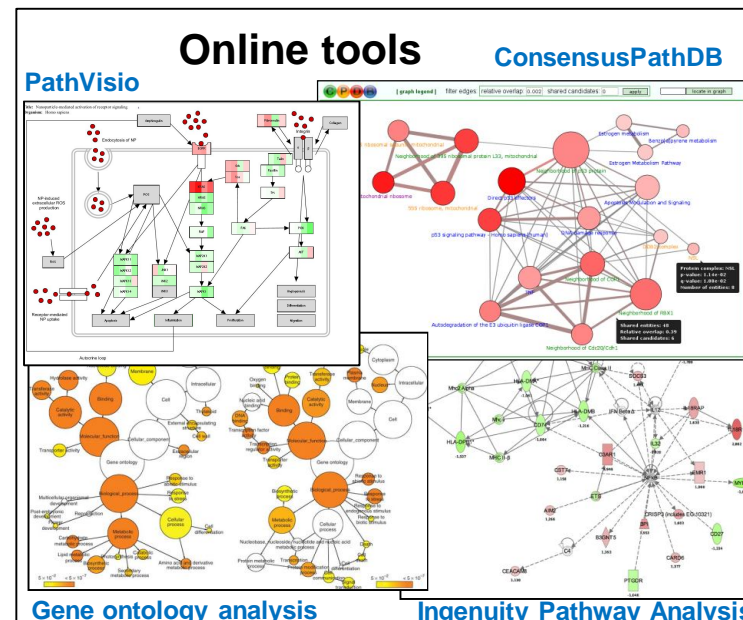
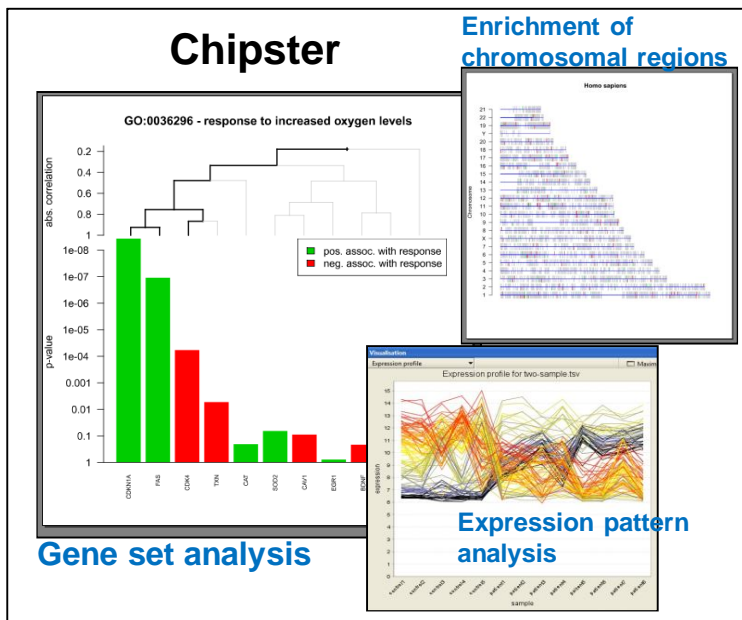
Chipster

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

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

Chipster tutorials

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

The screenshot displays the YouTube channel for 'Chipster', an open-source platform for data analysis. The channel banner features the CSC logo and a DNA double helix. Below the banner, the 'Chipster Tutorials' section lists several video uploads:

Video Title	Category	Duration	Views	Time Ago
Basic sequence analysis tasks in Chipster	Sequence analysis	11:43	25	1 month ago
EdgeR for multivariate experiments (differential expression analysis for RNA-Seq)	NGS	6:33	263	5 months ago
Differential expression analysis tools for RNA-Seq	NGS	2:49	488	5 months ago
Using the "nested" option in multivariate edgeR	NGS	4:00	49	5 months ago
Introduction to Chipster	General	12:21	402	6 months ago
Excluding samples from dataset	General	3:51	49	6 months ago

Below the uploads, a 'Created playlists' section shows a playlist titled 'RNA-seq' containing 5 videos:

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

Hands-on session

Looking at a nano-related data set - [GSE42067](#)

Experimental set up

Model

- Human small airway epithelial cells (SAE)

Nanomaterials

- MWCNTs (Cheaptubes)
- TiO₂ 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 TiO₂
- 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
 - TiO₂-NB differentially regulated inflammatory responses associated with cellular stress

Participant conclusions?