Hands-on Workshop on Nano Safety Assessment

Workshop:	NMP projects eNanoMapper, NanoFASE, GUIDEnano and SUN joint event
DATE / PLACE:	10 February 2016 / Technology Park Basel, Switzerland
TIME:	11.00-11.45

TITLE:	**Bioinformatics Analysis of Nanotechnology-based Omics Data**

SPEAKER:	Penny Nymark
AUTHORS:	Penny Nymark[1, 2], Pekka Kohonen[1, 2], Vesa Hongisto[1] and Roland Grafström [1, 2] [1] *Misvik Biology Oy, Finland* [2] *Karolinska Intsitutet, Institute of Environmental Medicine, Sweden*

ABSTRACT

This demonstration will get you started on the basic use of Chipster. It will guide you through the import of microarray data, normalization, quality checks, basic analyses methods and visualization of your data. For additional and deeper guidance on how to use Chipster, see the webpage http://chipster.csc.fi/ or contact the developers at chipster@csc.fi. The website also includes guidance on how to analyze NGS data in Chipster.

The demo username and password to Chipster is *guest* (does not allow running analyses)

If you would like to use Chipster in the future, please contact:
Finnish users: the developers at CSC in Finland (http://chipster.csc.fi/)
Swedish users: UPPMAX in Sweden (http://www.uppmax.uu.se/installed-software)
Dutch users: the Dutch TechCentre for Life Sciences (http://www.dtls.nl/ctmm-trait-makes-chipster-available-biomedical-studies/)
German users: DKFZ (https://www.dkfz.de/gpcf/chipster0.html)
International users (coming soon – until then contact the developers in Finland):
https://www.egi.eu/news-and-media/newsletters/Inspired_Issue_19/chipster.html

Courses, manuals and tutorials are available through:
http://chipster.csc.fi/
https://www.youtube.com/channel/UCnL-Lx5gGlW01OkskZL7JEQ
http://www.elixir-finland.org/next-generation-sequencing-data-analysis-with-chipster/

HANDS-ON EXERCIZES

Data set

mRNA data (transcriptomics)

- Raw mRNA data – Affymetrix CEL files
- Array type: Affymetrix HT HG-U133a 2.0 (=hgu133av2 in Chipster)
- GEO accession number: GSE42067

Experiment set-up

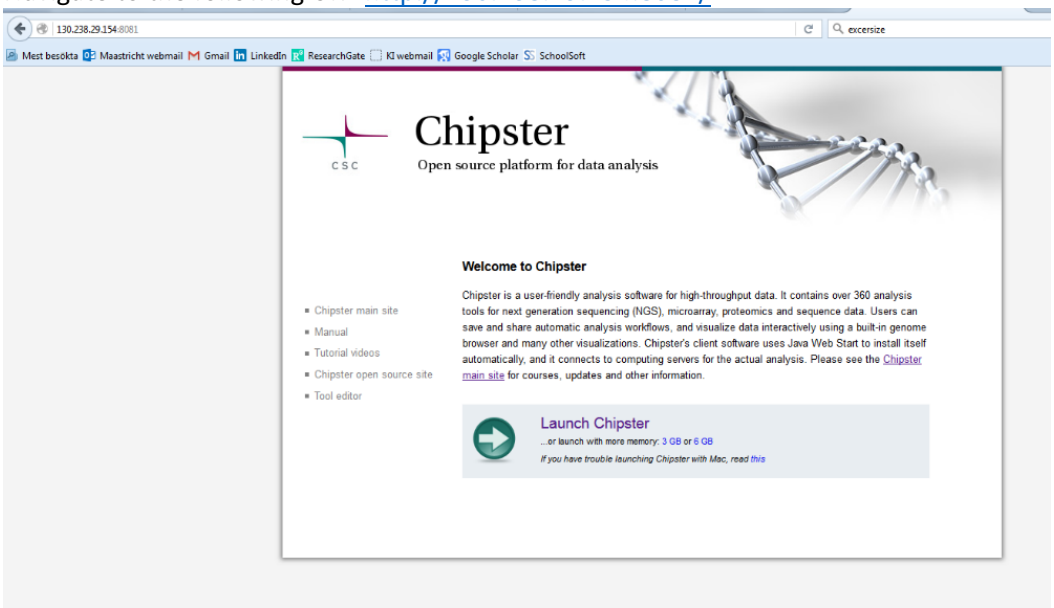
- Human small airway epithelial cells (SAE) exposed to 10 and 100 µg/ml titanium dioxide nanobelts (TiO_2) and Cheaptubes multi-walled carbon nanotubes (MWCNT) for 1h and 24h (3 replicates)
    - 24 samples
- Unexposed SAE cells (3 replicates)
    - 6 samples

Chipster analysis session available to course participants

- 'Chipster_session_Workshop_Basel_100216_hands_on.zip'

Open a pre-made analysis session in Chipster to explore the results

1. Open Chipster through the UPPMAX server.
    a. Open a browser *(Chrome* or *Firefox)*
    b. Navigate to the following URL http://130.238.29.154:8081/

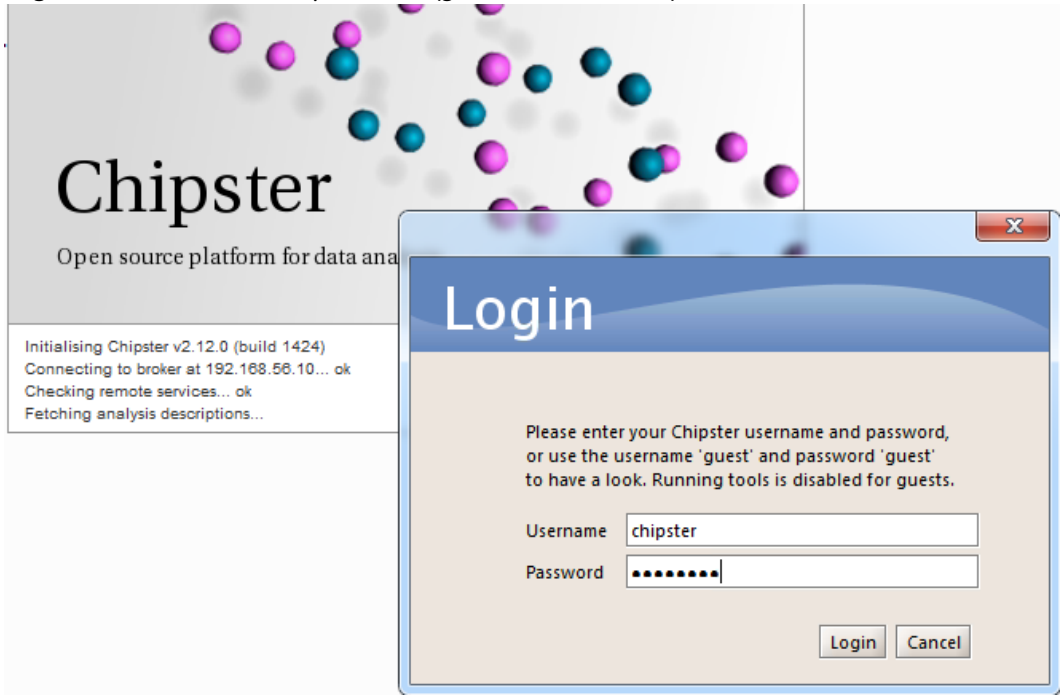


    c. Click on **Launch Chipster**

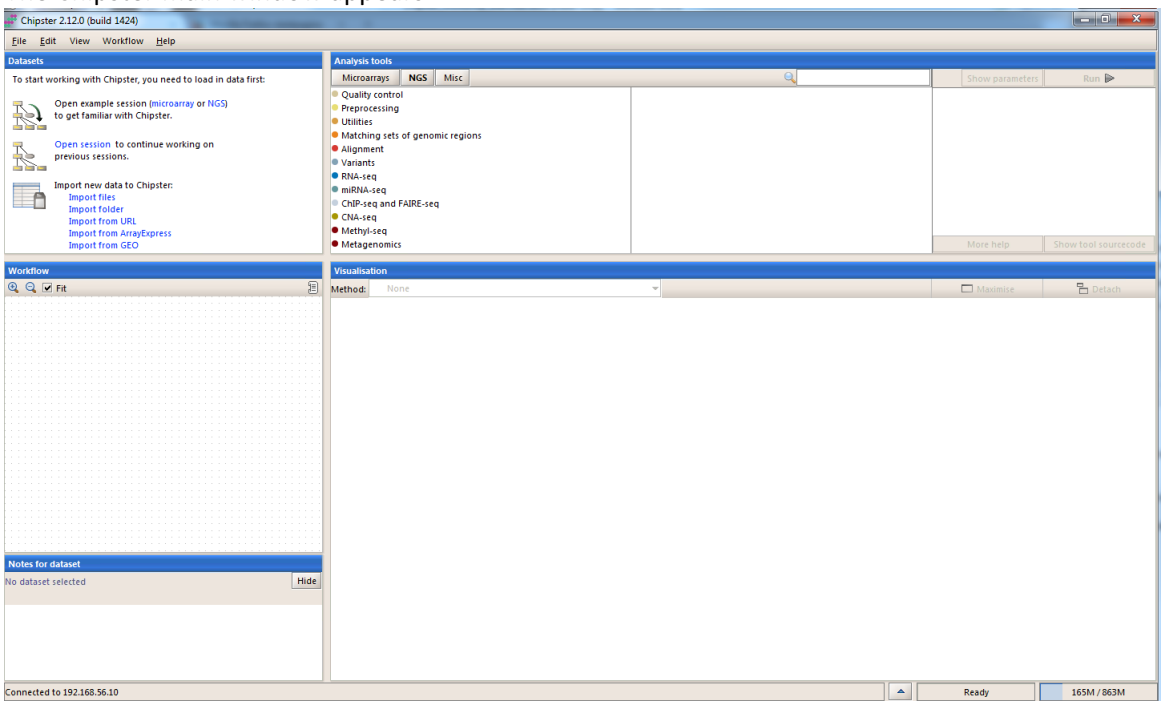
d. The file **chipster.jnlp** begins to download. Keep this file and open it (Chrome) or open it directly (Firefox)

e. A security message appears. Click **Run**

f. Login with username and password (*guest* in both cases)



g. The Chipster main window appears



2. Open the session that has been provided by the course organizers

3. Click once on the group of imported raw Affymetrix CEL files
   Q1: How many samples are there?

4. Double click on the normalized file
   Q2: What information can you find in that file?

5. Double click on the phenodata file
   Q3: What type of information does that file contain?
   Q4: What other information would you/could you add to that file?

Quality control and filtering

6. Check the Quality Control (QC) files (RLE, NUSE, RNA-degradation, spike-in and simpleaffy plots)
   Q5: What seems to be the quality of the samples?
   Q6: Can you figure out the reason for the two different overall levels of distributions of Relative Log Expression (RLE) values in the RLE plot? Does it indicate bad quality?

   TIP1: Look at the phenodata file to see any correlation between the different distributions and the the sample annotations. Such differences can be due to biological differences, but also to technical batch effects.

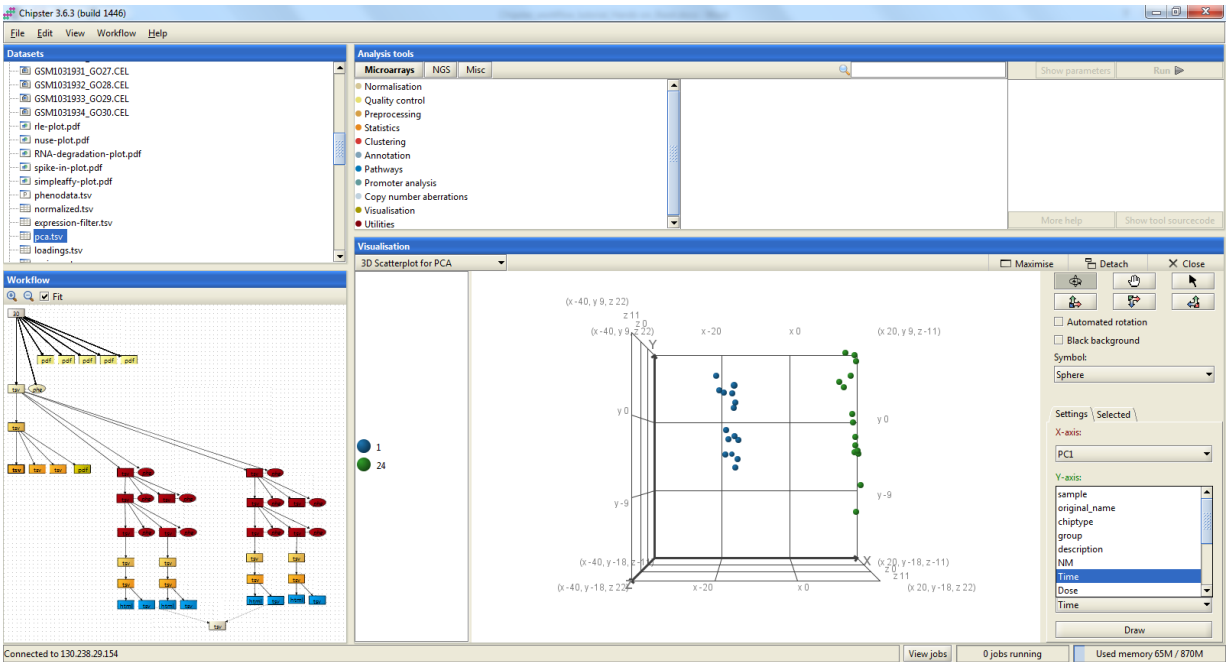
TIP2: Refer to the Chipster manual for descriptions of QC plots (http://chipster.csc.fi/manual/tools.html#quality)

7. Click once on one of the preprocessed files (expression-filter.tsv)
   Q7: How was the normalized data filtered?
   Q8: Why do you think it was filtered this way?

Clustering and Visualization

8. Click once on the PCA file and then open it with the "3D scatter plot for PCA" option in the "Visualization" window.
   Q9: What seems to be the main effect on the grouping of the samples?
   TIP: Color the data according to different phenodata parameters (e.g. 'NM' or 'time') in the scroll-down menu on the right (see example Figure below).

   Q10: What other initial conclusions can you draw from the PCA plot?



9. Click on the heatmap file
   Q11: What conclusions can you draw from it?

Statistical analysis

10. Check the four files with differentially expressed genes (two-sample.tsv)
    Q12: How many genes are significantly differentially expressed for each treatment?
    TIP1: Check how many rows the file has – upper left corner.

TIP2: To know which exposure the genes are associated with, go backwards (or up) in the workflow and see which samples have been extracted as basis for the analysis. The best way is to check the phenodata file.

Pathway and Gene Ontology Enrichment analysis

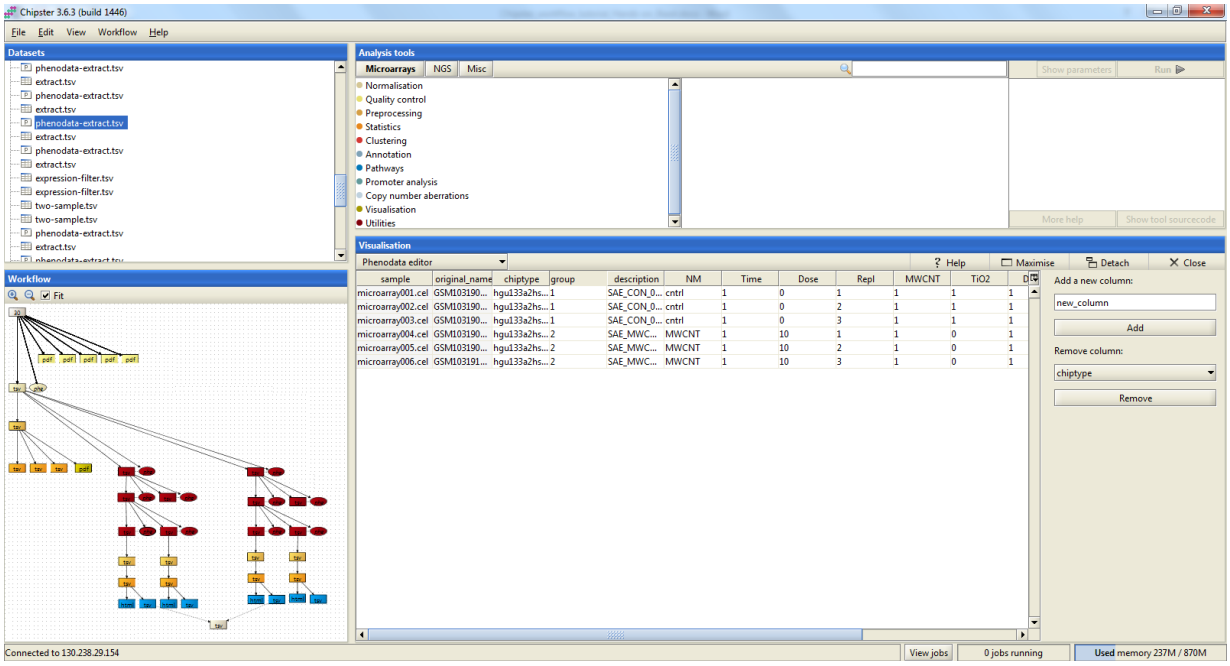
11. Click on the four GO files (hypergeo.tsv)
    Q13: How many enriched Gene Ontologies are there for each condition?
    Q14: Can you draw some conclusions based on the ontologies?

12. Choose the GO file (hypergeo.tsv) for condition MWCNT, Dose 10 µg/ml, 1h and the GO file for the condition TiO_2, Dose 10 µg/ml, 1h by holding down the Ctrl button. Open a Venn diagram by clicking on the icon in the 'Visualization' window.
    Q15: How many common GO terms are there between the MWCNT- and TiO_2-exposures at the lower dose at 1h?
    TIP: You can check which ones they are in the file that is linked to both the GO files (user-edited.tsv).

    Do the same for the 24h exposures (the other two hypergeo.tsv files).
    Q16: How many common GO terms are there?
    Q17: What conclusions can you draw from the answers to Q15 and 16?

Other useful functions

13. Click on one of the extracted files (extract.tsv)
    Q18: Which samples have been extracted from the original set of samples?
    TIP: Check the samples in the phenodata files (see example Figure below)



Export data from Chipster

14. Click on the right mouse button on any .tsv file that you would like to export for further analysis (e.g. a list of differentially expressed genes). Choose 'Export…' and save the file on the Desktop and open it in Excel.
    You can now use the list for analysis in online tools.
    Q19: What tool would you test the list(s) of genes in to interpret the data biologically?

Answers

A1: 30 samples
A2: Probeset identifiers, gene names, gene description, log2 transformed expression of each gene in each sample
A3: Sample annotations, such as exposure, dose, time and replicate pairs
A4: Anything, e.g. information from an endpoint assay, e.g. level of cyto-, geno- or immunotoxicity
A5: Good, since e.g. the RLE plot shows that the samples are all centered around 0.
A6: No it does not indicate bad quality, it is due to the large differences in gene expression between the two different time points tested – 1h and 24h.
A7: Genes with below log2 expression 6 and above log2 expression 100 in at least one sample were filtered out.
A8: Genes with low expression in all samples do not need to be analyzed and can be filtered out before statistical analysis to make it more robust. The upper limit (log2 expression 100) is in general never reached and can safely be used without losing genes with high expression.
A9: The time effect seems to be the strongest.
A10: It is also possible to notice that TiO2 seems to elicit a stronger effect on the cells at 24h, than MWCNTs (color the samples in the PCA according to NM).
A11: Similar conclusions can be drawn as from the PCA plot, but in addition stronger NM-specific clusters can be seen at 24h, indicating more NM-specific changes at this time point. Also the genes can be seen to cluster into two large differentially expressed clusters of genes, which are either up-regulated at 1h and down-regulated at 24h or vice versa.
A12:
MWCNT Dose 10 1h – 226 genes
MWCNT Dose10 24h – 22 genes
TiO_2 Dose10 1h – 319 genes
TiO_2 Dose10 24h - 316 genes
A13:
MWCNT Dose10 1h – 57 GOs
MWCNT Dose10 24h – 9 GOs
TiO_2 Dose10 1h – 112 GOs
TiO_2 Dose10 24h – 21 GOs
A14: The lists of ontologies are long and it is difficult to draw overall conclusions, however, the top GOs indicate:
MWCNT Dose10 1h - Changes in RNA metabolic processes
MWCNT Dose10 24h - Changes in MAPK cascades
TiO_2 Dose10 1h - Changes in primary metabolic processes
TiO_2 Dose10 24h - Changes in immune response

In addition, you may want to look at the relation between number of genes and number of GOs for each exposure. E.g. for TiO2, both time points seem to affect similar numbers of genes, but the GOs are much fewer at 24h, potentially indicating a more focused response.
A15: 26 common GOs
A16: None
A17: The larger overlap of GO terms between the two exposures at 1h indicate that the cells initially respond similarly to both MWCNTs and TiO_2, but the lack of overlapping GO terms at 24h point towards different responses developing over time. This is in agreement with the observations from the PCA and heatmap. And also in agreement with the original study.
A18: The extracted files contain either only MWCNT-exposed samples (+ controls) or TiO_2-exposed samples (+ controls). Lower down in the workflow only one dose or one time point has been extracted.
A19: If you extract a list of differentially expressed genes, you can now analyze it in e.g.
PathVisio (free) http://www.pathvisio.org/
Ingenuity Pathway Analysis (commercial) http://www.ingenuity.com/products/ipa
ConsensusPathDB (free) http://consensuspathdb.org/
Reactome (free) http://www.reactome.org/
Enrichr (free) http://amp.pharm.mssm.edu/Enrichr/
or any other of the many bioinformatics pathway analysis tool.

Further interpretations based on an advanced session shown by the organizers

Dose response
Limma is used to identify dose response related differentially expressed genes and PCA plots are generated based on those genes (for MWCNT and TiO_2 separately). The figure shows a more clear dose response for TiO2 than for MWCNT, especially at 24h – hypothesis: MWCNTs are known to agglomerate with time…?

Figure below: TiO_2-related dose-response.

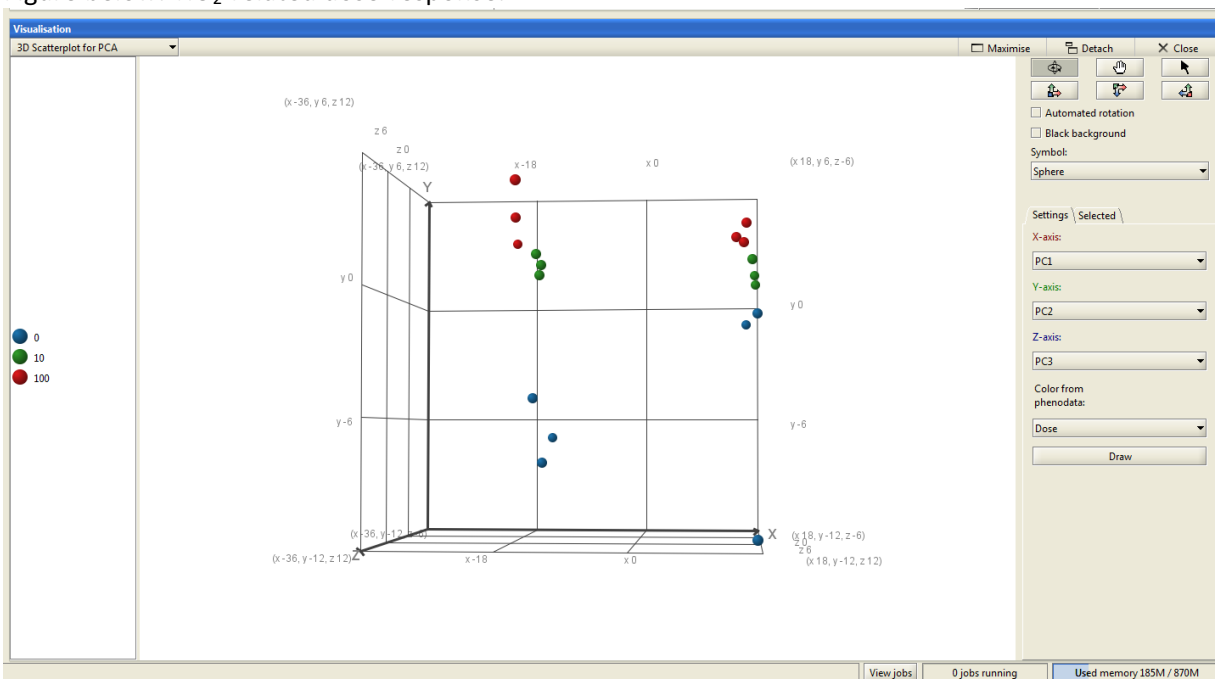
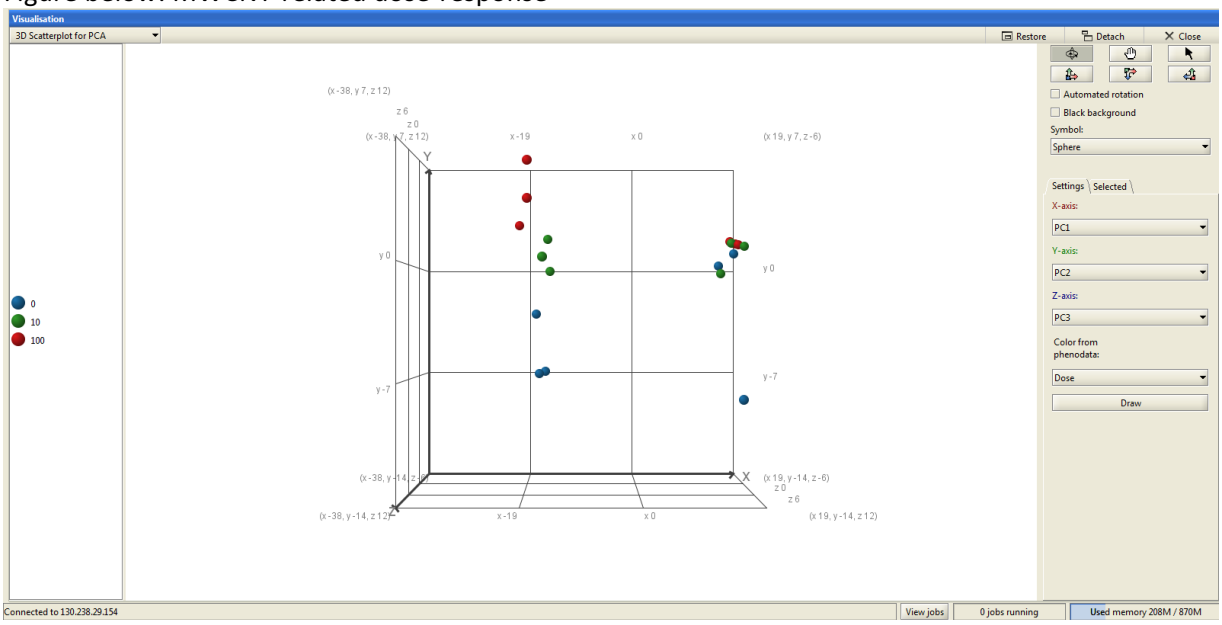


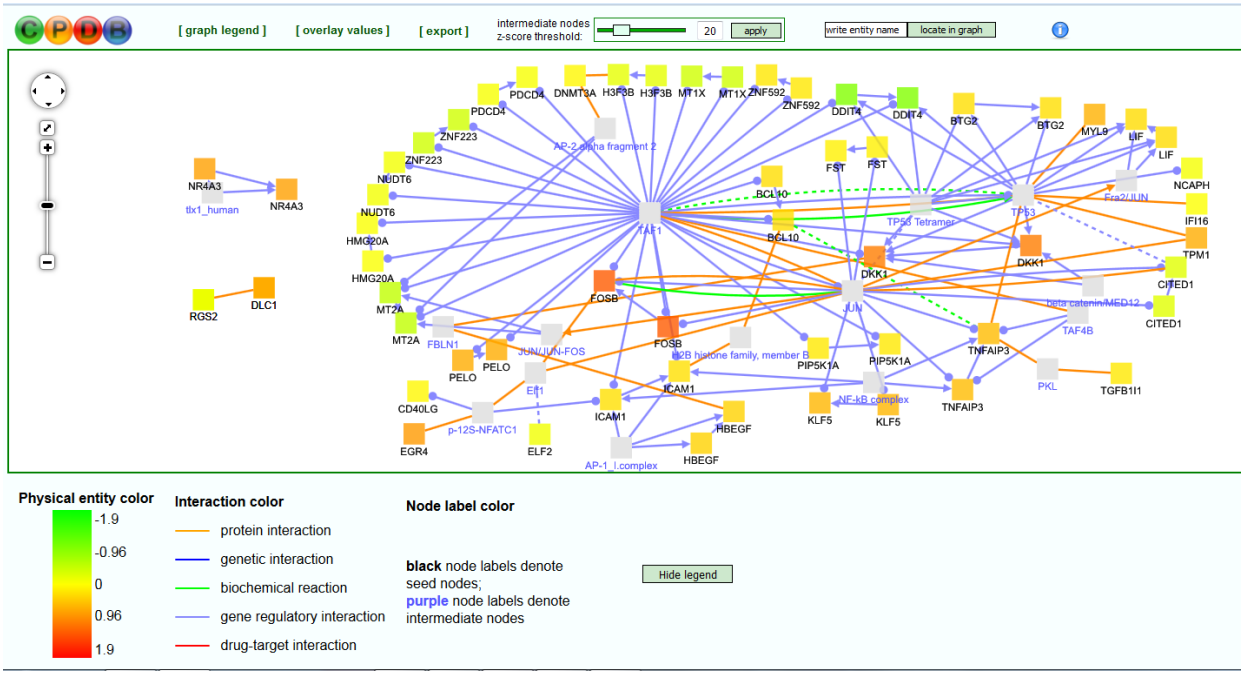
Figure below: MWCNT-related dose-response



Gene network Analysis

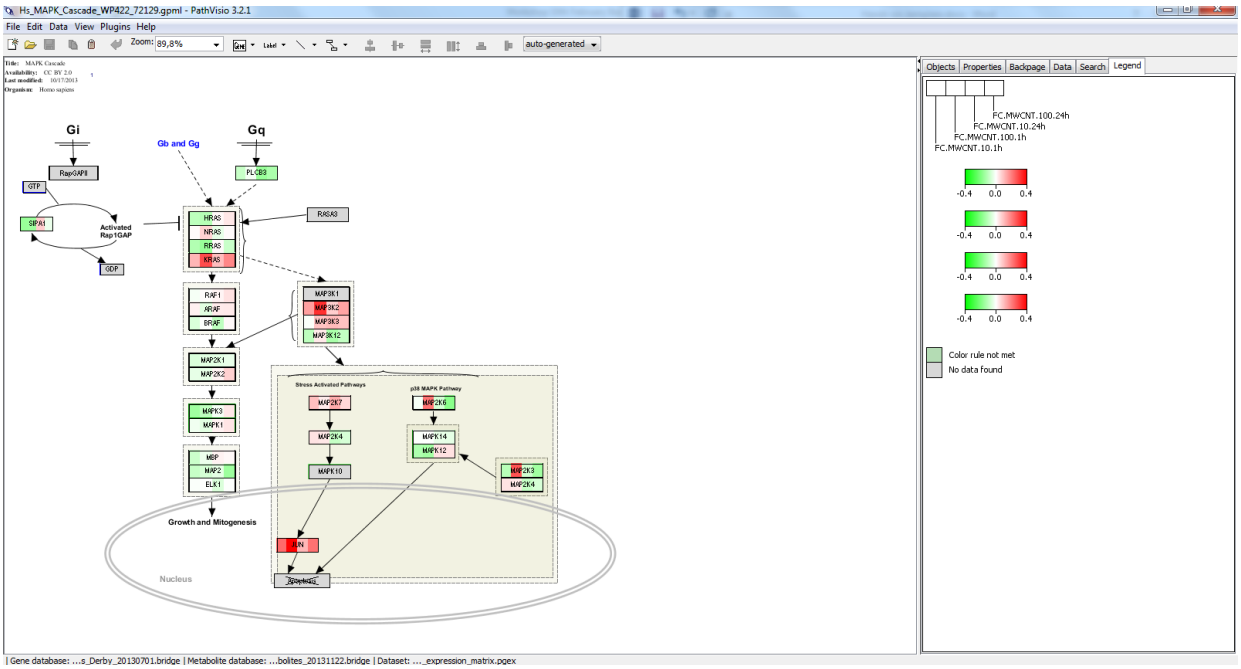
ConsensusPathDB was used to perform an induced network module analysis (http://consensuspathdb.org/) of the genes related to the dose-response following MWCNT exposure. The resulting gene (and protein) interaction network shows that a large part of the genes are interconnected with (and thereby potentially regulated by) the transcription factor TAF1, indicating a role for this molecule in the dose-response to MWCNTs.

See figure below for the identified gene interaction network showing the expression level of the genes for exposure MWCNT, 100 µg/cm^2, 1h.

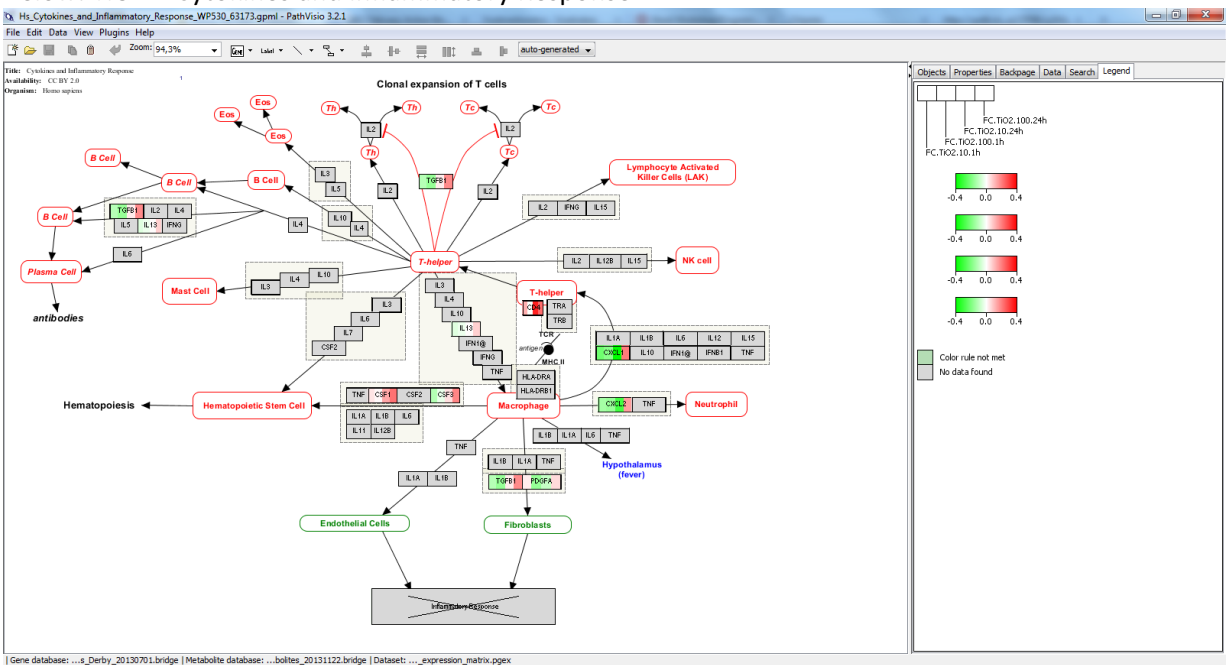


PathVisio (http://www.pathvisio.org/) was also used to visualize pathways related to MAPK cascades (for MWCNT) and immune response (for TiO2), using a matrix with the expression of all differentially expressed genes for all conditions.

Below: MWCNT – MAPK Cascade



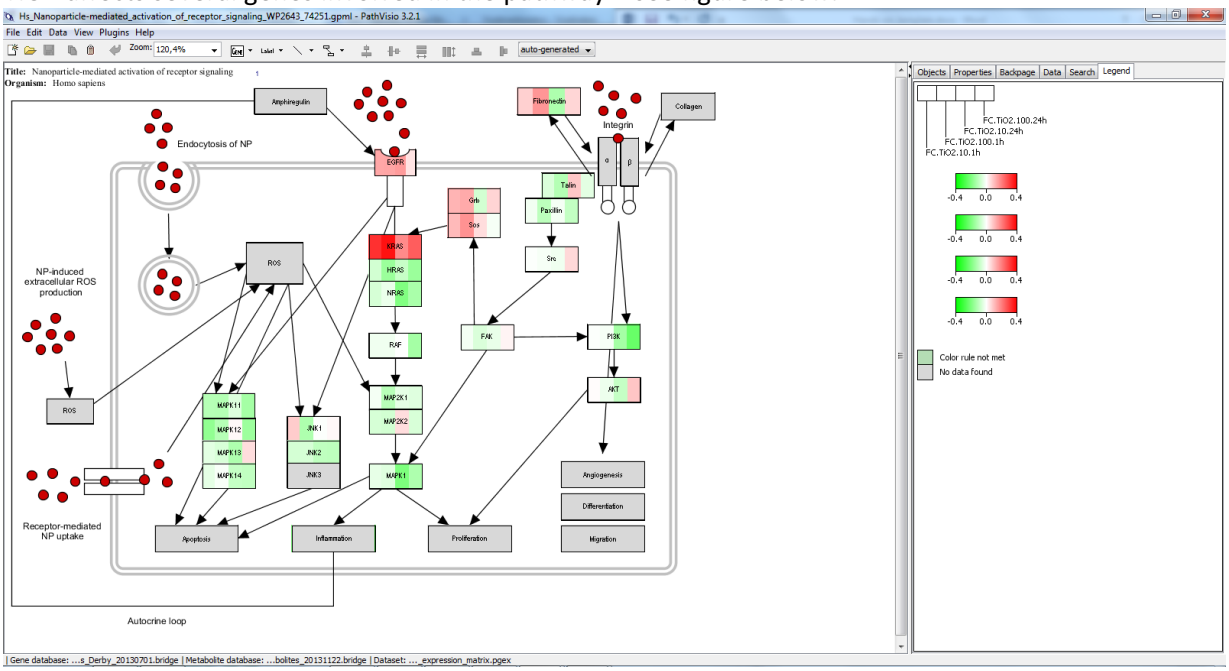
Below: TiO2 – Cytokines and inflammatory Response



Curiosa related to PathVisio and WikiPathways

PathVisio also contains the WikiPathway "Nanoparticle mediated activation of receptor signaling" which may be an interesting starting point for nanoparticle-related bioinformatics gene network analysis.

TiO2 affects several genes involved in the pathway – see figure below.



REFERENCES

A selection of databases and tools relevant for nanomaterial-related bioinformatics

ArrayExpress https://www.ebi.ac.uk/arrayexpress/
GEO http://www.ncbi.nlm.nih.gov/geo/
NanoMiner http://compbio.uta.fi/estools/nanommune/index.php/
Chipster http://chipster.csc.fi/
Chipster UPPMAX (Sweden) http://130.238.29.154:8081/
PathVisio http://www.pathvisio.org/
Ingenuity Pathway Analysis (commercial) http://www.ingenuity.com/products/ipa
ConsensusPathDB http://consensuspathdb.org/
Reactome http://www.reactome.org/
Enrichr http://amp.pharm.mssm.edu/Enrichr/

Publications

Grafström RC, Nymark P, Hongisto V, Spjuth O, Ceder R, Willighagen E, Hardy B, Kaski S, Kohonen P. *Toward the Replacement of Animal Experiments through the Bioinformatics-driven Analysis of 'Omics' Data from Human Cell Cultures.* Altern Lab Anim. 2015 Nov;43(5):325-32

Nymark P, Wijshoff P, Cavill R, van Herwijnen M, Coonen ML, Claessen S, Catalán J, Norppa H, Kleinjans JC, Briedé JJ. *Extensive temporal transcriptome and microRNA analyses identify molecular mechanisms underlying mitochondrial dysfunction induced by multi-walled carbon nanotubes in human lung cells.* Nanotoxicology. 2015;9(5):624-35. doi: 10.3109/17435390.2015.1017022

Kohonen P, Ceder R, Smit I, Hongisto V, Myatt G, Hardy B, Spjuth O, Grafström R. *Cancer biology, toxicology and alternative methods development go hand-in-hand.* Basic Clin Pharmacol Toxicol. 2014 Jul;115(1):50-8. doi: 10.1111/bcpt.12257. Review.

Kohonen P, Benfenati E, Bower D, Ceder R, Crump M, Cross K, Grafström RC, Healy L, Helma C, Jeliazkova N, Jeliazkov V, Maggioni S, Miller S, Myatt G, Rautenberg M, Stacey G, Willighagen E, Wiseman J and Hardy B. *The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing.* Mol. Inf., 32: 47–63. doi: 10.1002/minf.201200114

Tilton SC, Karin NJ, Tolic A, Xie Y, Lai X, Hamilton RF Jr, Waters KM, Holian A, Witzmann FA, Orr G. *Three human cell types respond to multi-walled carbon nanotubes and titanium dioxide nanobelts with cell-specific transcriptomic and proteomic expression patterns.* Nanotoxicology. 2014 Aug;8(5):533-48. doi: 10.3109/17435390.2013.803624

Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, Koski M, Käki J, Korpelainen EI. *Chipster: user-friendly analysis software for microarray and other high-throughput data.* BMC Genomics. 2011 Oct 14;12:507. doi: 10.1186/1471-2164-12-507