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
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

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

Adapted from Culhane et al, Cell Mol Life Sci, 64:3185-3200, 2007
Comparison of microarrays and RNA-sequencing technologies

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

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

Sequencing has greater sensitivity and provides more information.
Existing omics data – including nano-specific

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

- 63,223 experiments
- 1,900,034 assays
- 40.13 TB of archived data
Nano-specific omics data gathered in one place

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

Penny Nymark eNanoMapper Workshop, 10th February, 2016, Basel
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
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
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

- [https://www.youtube.com/channel/UCnL-Lx5gGIW01OkskZL7JEQ](https://www.youtube.com/channel/UCnL-Lx5gGIW01OkskZL7JEQ)
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

- 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?