Annotations in microarray data analysis

Jean Yee Hwa Yang
University of California, San Francisco
http://www.biostat.ucsf.edu/jean/

Statistics for Microarray Data
IBC Short Course, 11 July 2004

Examples of some available databases

- Literature databases
  - PubMed, Medline, OMIM
- Factual databases
  - Nucleic acid sequence: GenBank, EMBL, DDBJ, RefSeq
  - Amino acid sequence: SwissProt
  - 3D molecular structures: PDB
- Knowledge and other databases
  - Gene classification: Unigene, LocusLink
  - Gene Ontology: GO
  - Motif libraries: Prosite
  - Pathways: KEGG, WIT
  - Transcription factor: Transfac

Some public databases URL

- EBI http://www.ebi.ac.uk/Databases/
  - LocusLink http://www.ncbi.nlm.nih.gov/LocusLink
  - KEGG http://www.genome.jp/
  - WIT http://wit.mcs.anl.gov/WIT2/
  - TRANSFAC http://transfac.gbf.de/TRANSFAC/
- SwissProt http://www.ebi.ac.uk/swissprot/

Example

- Oligo ID M200000488
- GenBank accession NM_008355
- Nucleotide Sequence

GenBank

Search for NM_008355
Example

<table>
<thead>
<tr>
<th>Oligo ID</th>
<th>M200000488</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenBank accession</td>
<td>NM_008355</td>
</tr>
<tr>
<td>Name</td>
<td>Interleukin 13</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>IL13</td>
</tr>
<tr>
<td>UniGene</td>
<td>Mm.1284</td>
</tr>
<tr>
<td>Map Position</td>
<td>Chromosome 11: 29.0 Mb</td>
</tr>
</tbody>
</table>

**GO**
GO:0005144

- GO Browsers
  - AmiGO from BDGAP
  - MGI GO Browser
  - QuickGO at EBI and many others

---

**Gene Ontology**

- Cellular component
- Biologic process
- Molecular function

Directed acyclic graph structure

- Swiss-Prot
  - P20109
  - GO:0005144

- SwissProt
  - NiceProt View of Swiss-Prot:
    - P20109

- MGI: 96541

- Map Position: Chromosome 11: 29.0 Mb

- Map Position: Chromosome 11: 29.0 Mb

- Accession: MGI: 96541

- Accession: MGI: 96541

- Accession: MGI: 96541
**Example**

- **Oligo ID**: M200000488
- **GenBank accession**: NM_006036
- **Map Position**: Chromosome 11

**Swiss-Prot**: P0109
**GO**: GO:0005144
**Gene Symbol**: 813
**Name**: Interleukin 13

**KEGG: database of biochemical pathways**

- **Pathway**:
  - 00010: Glycolysis / Gluconeogenesis
  - 00020: Citrate cycle (TCA cycle)
  - 00030: Pentose phosphate

**MedLine**: 22738719
**PubMed**: 14897262

**Different database**

- A database assesses a unique identifier to each element: a database key.
- Each database rely on its own key (GenBank accession number, UniGene ID, GO ID...). This key is used to query the database and to retrieve the information.
- Different web sites for a summary:
  - Additional sites for human Genes:
    - GeneLynx: [http://www.genelinx.org](http://www.genelinx.org)
Information in microarray data

- Different levels / types of information:
  - Genes annotation
  - Samples annotations
  - Genes expression levels

Sample annotation

- Store microarray parameters, such as sample name, mouse strain, amplification protocols used.
MIAME

- An increasing number of journals are making it a requirement that publications involving microarray experiments conform to a set of guidelines and standards known collectively as MIAME.

What is MIAME?

- “[MIAME is a] set of guidelines [that] will then assist with the development of microarray repositories and data analysis tools.” -- Microarray Gene Expression Data Group (MGED)

- “[MIAME] aims to outline the minimum information required to unambiguously interpret microarray data and to subsequently allow independent verification of the data at a later stage if required. -- MGED

Software and database

- MIAMEXpress (a MIAME compliant microarray data submission tool) is currently available at EBI. http://www.ebi.ac.uk/miameexpress/

- MIAME compliant databases
  - GeneTraffic
  - Ipsogen Cancer Profiler
  - GeneSpring and GeNet

- See http://www.mged.org/Workgroups/MIAME/miame.html for more information.

Information in microarray data

- Different levels / types of information:
  - Genes annotation
  - Samples annotations
  - Genes expression levels

Expression databases

- Public repository
  - NCBI, Gene Expression Omnibus (GEO).
  - EBI, ArrayExpress.
    http://www.ebi.ac.uk/arrayexpress/