ASSIGNMENT

Allocated specific protein encoding gene

protein contains specific functional domain(s) of interest
Proteins contain functional domains (often many)
Functional domains initially identified experimentally > invitro
determine consensus domain signature (many experiments)
specific amino acid/nucleotide sequences present

eg: Phosphoglycerate kinase signature  (PATTERN)

Consensus aa pattern:  11 amino acids

Once determined protein sequences can be searched computationally for the presence of known domains

>blast search
Starting point

identify domains present in protein

The Arabidopsis Information Resource (TAIR)
http://www.arabidopsis.org/

General information about genes/proteins
domain composition
links to relevant publications etc

Example from paper

Pathogenesis-related protein-2 (PR-2) AT3G57260
Protein domain composition provides some functional information in isolated context.

Many proteins contain similar individual domains or domain combinations.

eg: kinase = 1700 in Arabidopsis
Wall associated kinase family (around 30)

Based on structure > Similar function:

Bind extracellular ligands (molecules) > signal to cytoplasm via kinase

Little functional specificity within family ?? Biological processes???
How can proteins be associated with a specific biological process?

Assumption in eukaryotes:
Most cellular responses (eg: development/stress/disease)

- not mediated by the action of a single gene/protein
  rather,

- network of genes/proteins function coherently to achieve a common cellular response

cell coordinates expression of functionally related genes

expressed at same time
In order to make end product, proteins must be present together.

Functionally related genes are co-expressed.
ANALYSIS OVERVIEW

GOI: GENE OF INTEREST

EXPRESSION CORRELATION ANALYSIS

ECGG: EXPRESSION CORRELATED GENE GROUP

GO ANALYSIS

CONDITION SPECIFIC EXPRESSION ANALYSIS

PROMOTER CONTENT ANALYSIS

ASSOCIATION OF GOI ECGG NETWORK WITH SPECIFIC BIOLOGICAL PROCESSES

Unknown biological function

Performed over many diverse conditions (over 130)

~ top 50 genes

Putative cis-elements - co-regulation

Functional enrichment
Arabidopsis co-expression tool

http://www.arabidopsis.leeds.ac.uk/act/

Expression correlation analysis (many experiments) identify co-expressed genes

Read : Frequently asked questions

Only use probes with single genes

Select top 50 genes
Gene Ontology

http://babelomics.bioinfo.cipf.es/EntryPoint?loadForm=fatigo

Identify functional enrichment

Select significant

If none select closest to significance to report
Genevestigator

www.genevestigator.ethz.ch

Use to identify experiments where gene/s are differentially expressed

Select all available experiments

Screen
  stimulus
  mutants
  development

Scroll over slide, lists repository – link to experiment
MICROARRAY REPOSITORIES

Nottingham Arabidopsis Stock Centre's microarray database (NASCArrays)

http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl

Array Express

http://www.ebi.ac.uk/microarray-as/ae/

Gene Expression Omnibus (GEO)

Promoter analysis: binding site enrichment

DNA (template)

Transcription factors

Promoter analysis: binding site enrichment

Promoter analysis: binding site enrichment

Proteins synthesised that have functional roles in the cell

Measuring mRNA levels tells us what proteins cells are making
Promoter analysis

Athena

visualization tool

http://www.bioinformatics2.wsu.edu/cgi-bin/Athena/cgi/home.pl

Search for enriched transcription factor binding sites (TFBS)

Highly co-expressed genes likely to be co-regulated

contain common TFBS in promoters (enriched)
bind common TFs
## Promoter enrichment

Motif = TTGAC

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of prom. in a dataset</th>
<th>Number of promoters containing the pattern</th>
<th>Total number of patterns in dataset</th>
<th>Promoter mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>BG model</td>
<td>31353</td>
<td>27523 (87.8%)</td>
<td>70245</td>
<td>2.24</td>
</tr>
<tr>
<td>Cluster1</td>
<td>26</td>
<td>25 (96.2%)</td>
<td>78</td>
<td>2.99</td>
</tr>
</tbody>
</table>
POBO
http://ekhidna.biocenter.helsinki.fi/poxo/pobo/
Additional notes

Link to Arabidopsis metabolic pathways

AraCyc Pathways (TAIR)
http://www.arabidopsis.org/biocyc/index.jsp

Can input gene ID or pathway name

http://www.gramene.org/pathway/

http://www.genome.jp/kegg/kegg2.html
http://www.genome.jp/kegg/pathway.html
For presentation of expression results can use MEV

Download at
http://www.tm4.org/mev/

Install, use to make heatmaps

See instructions