

SBCNY

NIGMS funded Center

Experimental Methods in Systems Biology

Part of the Coursera Certificate in Systems Biology

Marc Birtwistle, PhD

Department of Pharmacology & Systems Therapeutics

Fall 2014, Week 7a, Interpreting Omics Experiments



Icahn School
of Medicine at
**Mount
Sinai**

Outline

- What can I do with unbiased omics data—mRNA sequencing and/or mass spec-based proteomics?
 - Lists of differentially expressed genes/transcripts/proteins
 - Gene set enrichment analysis
 - Ontologies
 - Transcription factor analysis
 - Protein-protein interaction analysis
 - Kinase-substrate analysis
 - Putting it all together with Enrichr
- This is a brief preview of a subsequent course in our series taught by Prof. Ma'ayan.
- Case Study: Identifying HIPK2 as a Critical Kinase Driving Kidney Fibrosis

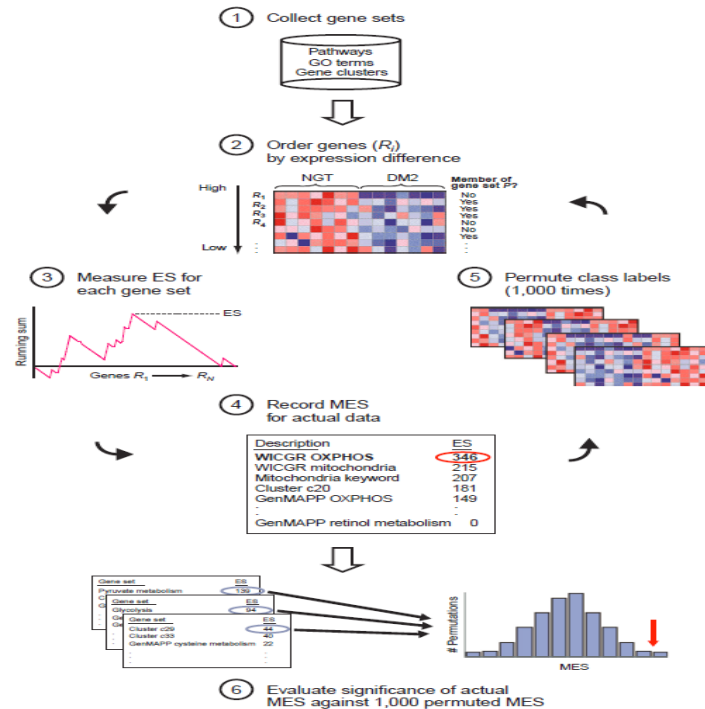
A Cornerstone of Omics Biology— Differential Expression

- Now I have quantified transcript levels (or protein levels) genome wide under two (or more) relevant conditions.
- What has changed?
- Differential expression analysis answers this question.
 - Implementation and theory is very heavy on computation and statistics—covered in much greater detail in Prof. Ma'ayan's course.
 - Here let us assume we can take such data and get a high confidence list
- Note: mRNA and protein data often give complementary information (although some obviously will be redundant)

A volcano plot titled "p-value versus fold-change". The y-axis is labeled "-log10 p-value" and ranges from 6.68 to 3.99. The x-axis is labeled "log2 fold change" and ranges from -6.0 to 6.0. A horizontal red dashed line is drawn at approximately y = 1.3. The plot contains numerous data points, with a dense cluster of grey points centered around (0, 1.3). Points with higher -log10 p-values (above the red line) are colored blue. Three red arrows point to specific blue data points in the upper left quadrant, at approximately (-6.2, 3.1), (-4.8, 3.2), and (-4.2, 3.0).

4

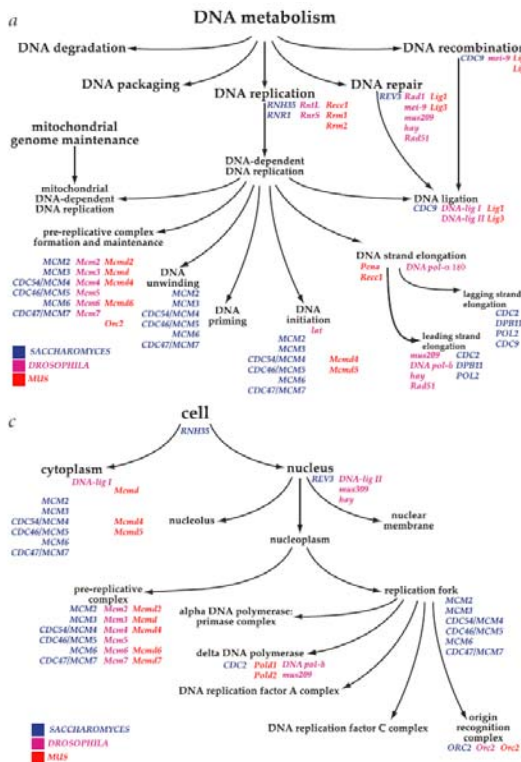
Gene Set Enrichment Analysis



- Rather than “gene-at-a-time”, look at coordinate regulation of many genes
- Rank genes in data
 - E.g. expression level
- “Walk” down ranked list and compute enrichment score as a running sum of overlap with pre-existing gene sets
 - Maximum enrichment score (MES) indicates how relevant a pre-existing gene set is for your data

Subramanian et al., PNAS, 2005; <http://www.broadinstitute.org/gsea/index.jsp>
 Mootha et al., Nat Genetics, 2003

Gene Ontology

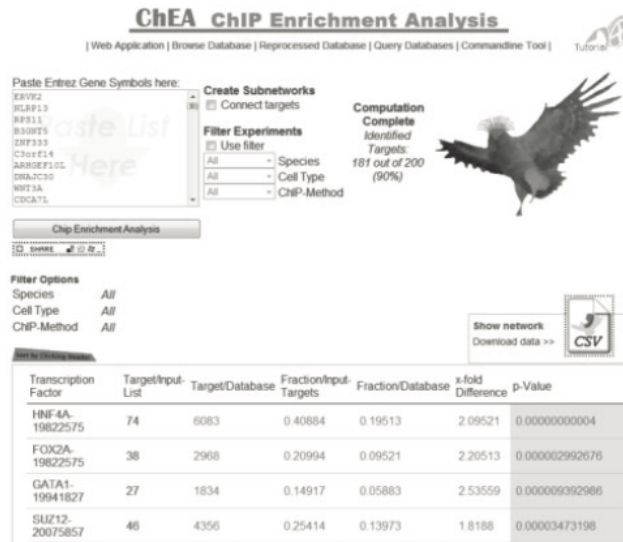


Ashburner et al., Nat Genetics, 2000

www.geneontology.org

- Gene sets organized by:
 - Biological Process
 - Molecular Function
 - Cellular Component

Transcription Factor Analysis

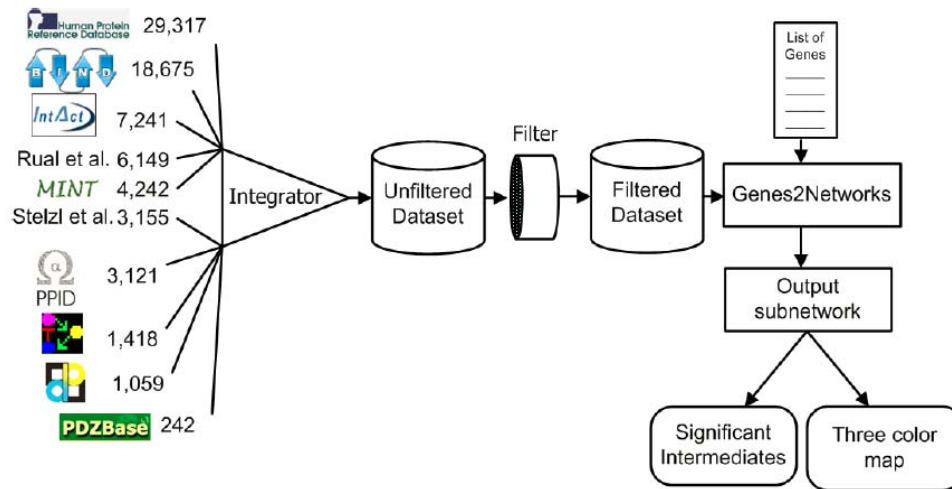


Lachmann et al. **ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments.** *Bioinformatics* 26(19):2438-44 (2010)

<http://amp.pharm.mssm.edu/lib/chea.jsp>

- Genes that change expression together may share common transcription factors
- This might be inferred by combining omics data with prior knowledge
- One tool: ChEA
 - Original publication: 189,933 interactions (and growing) that describe binding of 92 transcription factors to >30,000 genes
- Other tools out there
 - E.g. TRANSFAC, PAINT (Vadigepalli et al., OMICS, 2003)

Protein-Protein Interaction Analysis

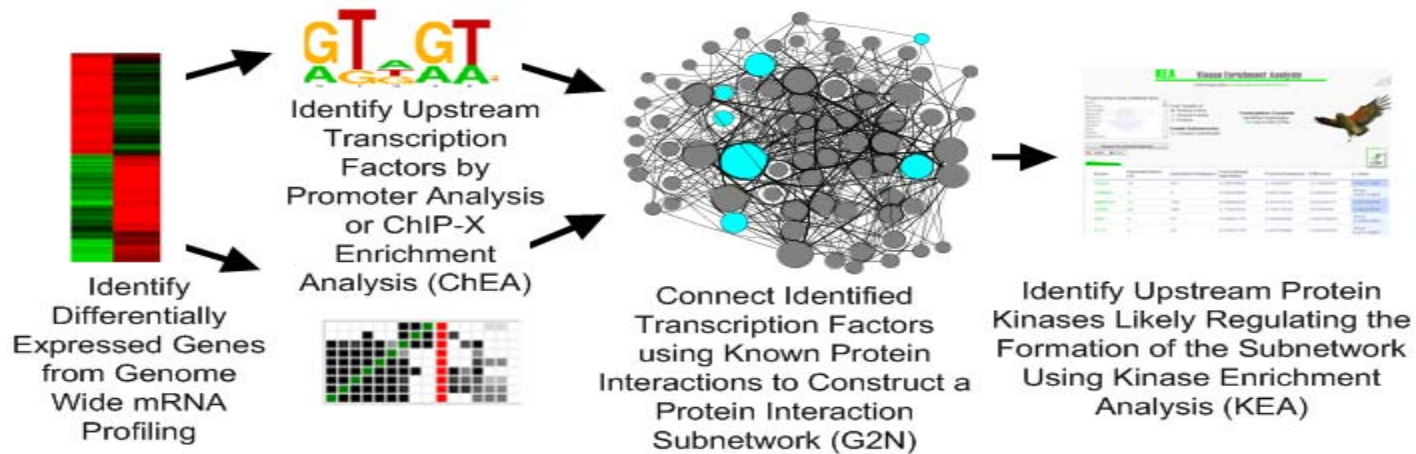


Berger et al. **Genes2Networks: connecting lists of gene symbols using mammalian protein interactions databases.**
BMC Bioinformatics 8:372 (2007)

<http://actin.pharm.mssm.edu/genes2networks>

- Many databases contain information on protein-protein interactions
 - E.g. Yeast two hybrid screens
- A given list of genes can be cross-referenced against these interactions
- Generates a putative network of interactions relevant to the gene list
- Can be biased:
 - Well-studied proteins
 - Stable interactions

Expression2Kinases: Identifying Upstream Pathways from Gene Expression



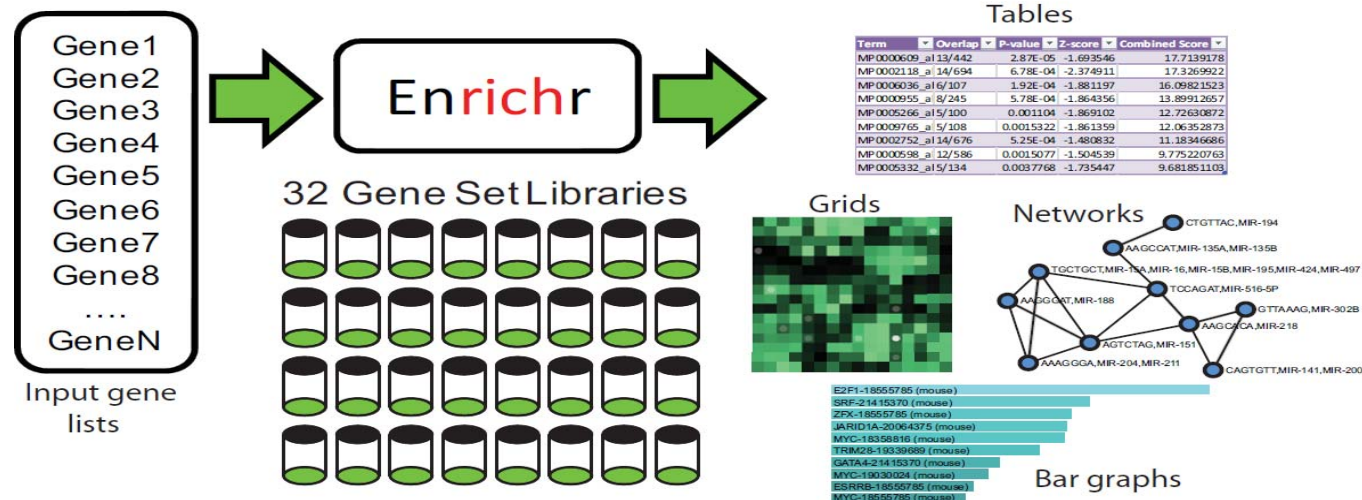
Chen et al. **Expression2Kinases: mRNA Profiling Linked to Multiple Upstream Regulatory Layers.** *Bioinformatics*. 2011 Nov 10.

www.maayanlab.net/X2K

Lachmann et al. **KEA: kinase enrichment analysis.** *Bioinformatics* 25(5):684-6 (2009)

<http://amp.pharm.mssm.edu/lib/kea.jsp>

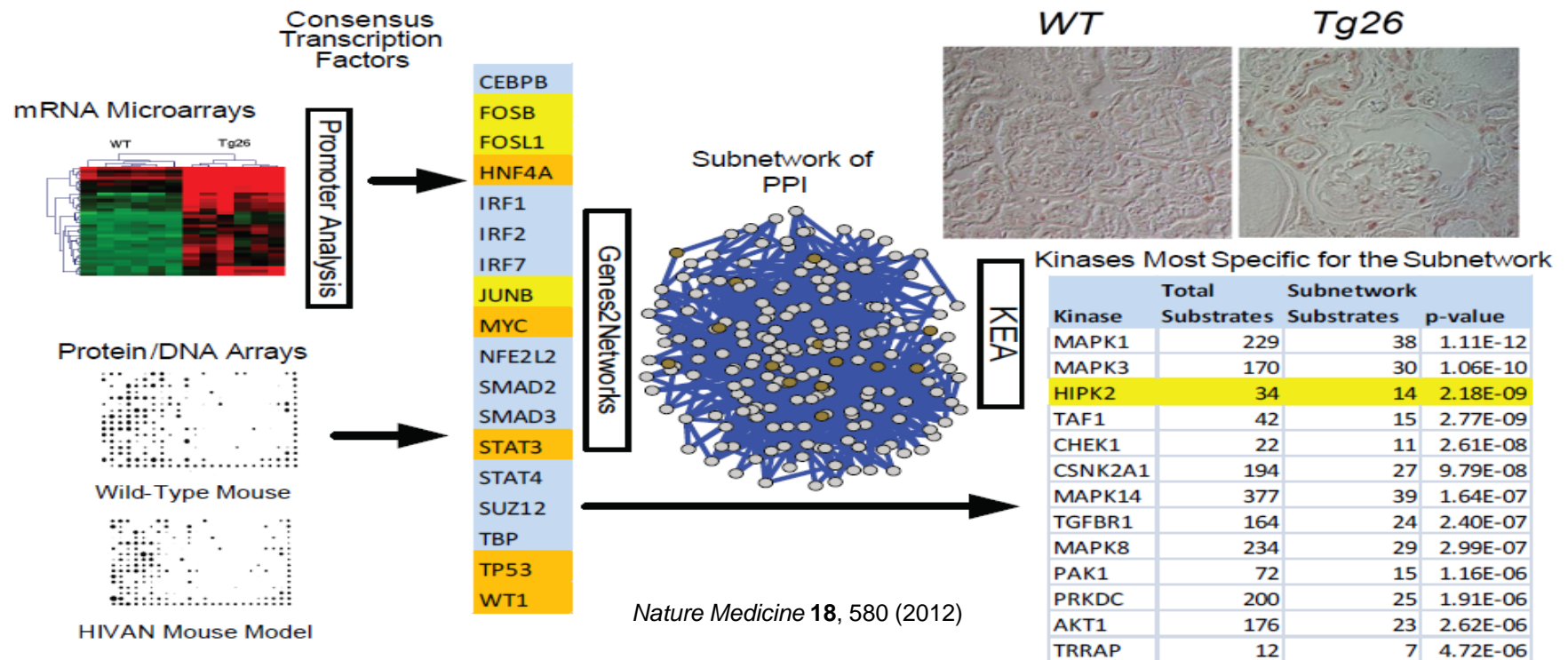
Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool



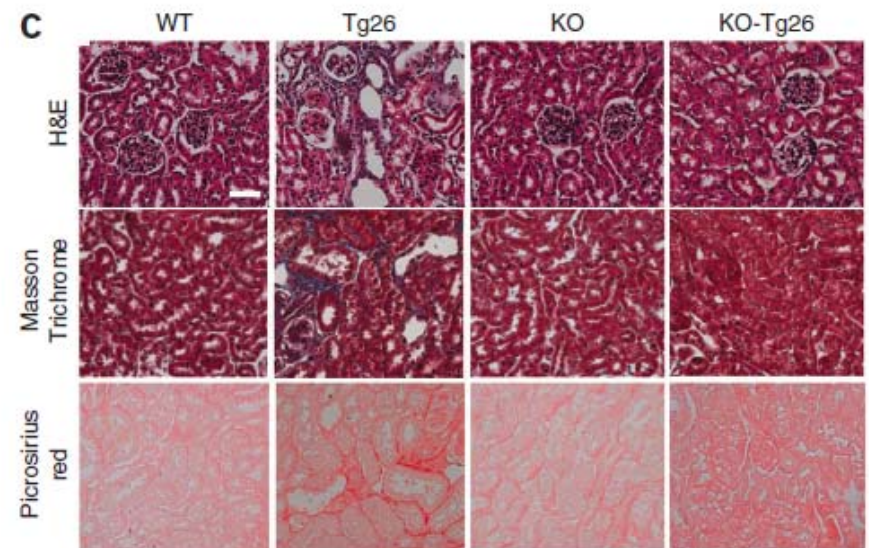
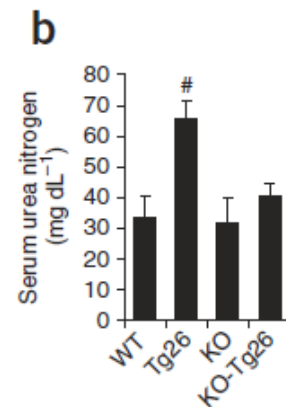
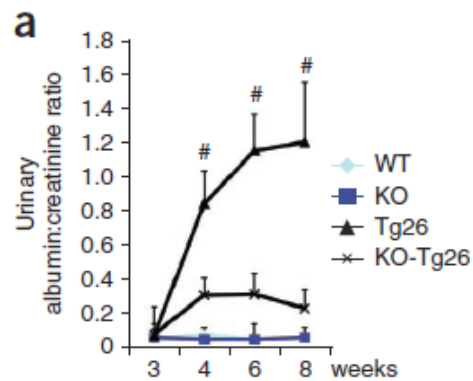
Edward Y. Chen et al. BMC Bioinformatics (2013)

<http://amp.pharm.mssm.edu/Enrichr/index.html>

Case Study: Identifying HIPK2 as a Critical Kinase Driving Kidney Fibrosis



Case Study: Identifying HIPK2 as a Critical Kinase Driving Kidney Fibrosis



Nature Medicine **18**, 580 (2012)