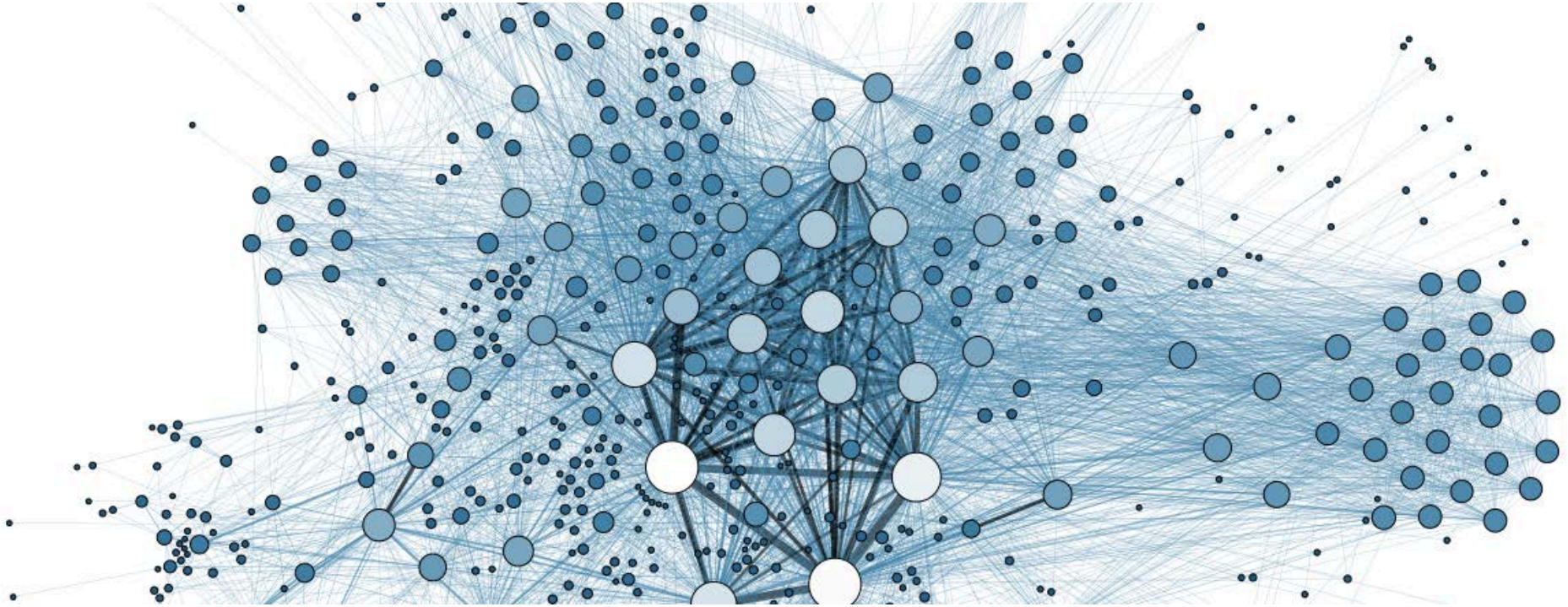
EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

FACULTY OF
SCIENCE

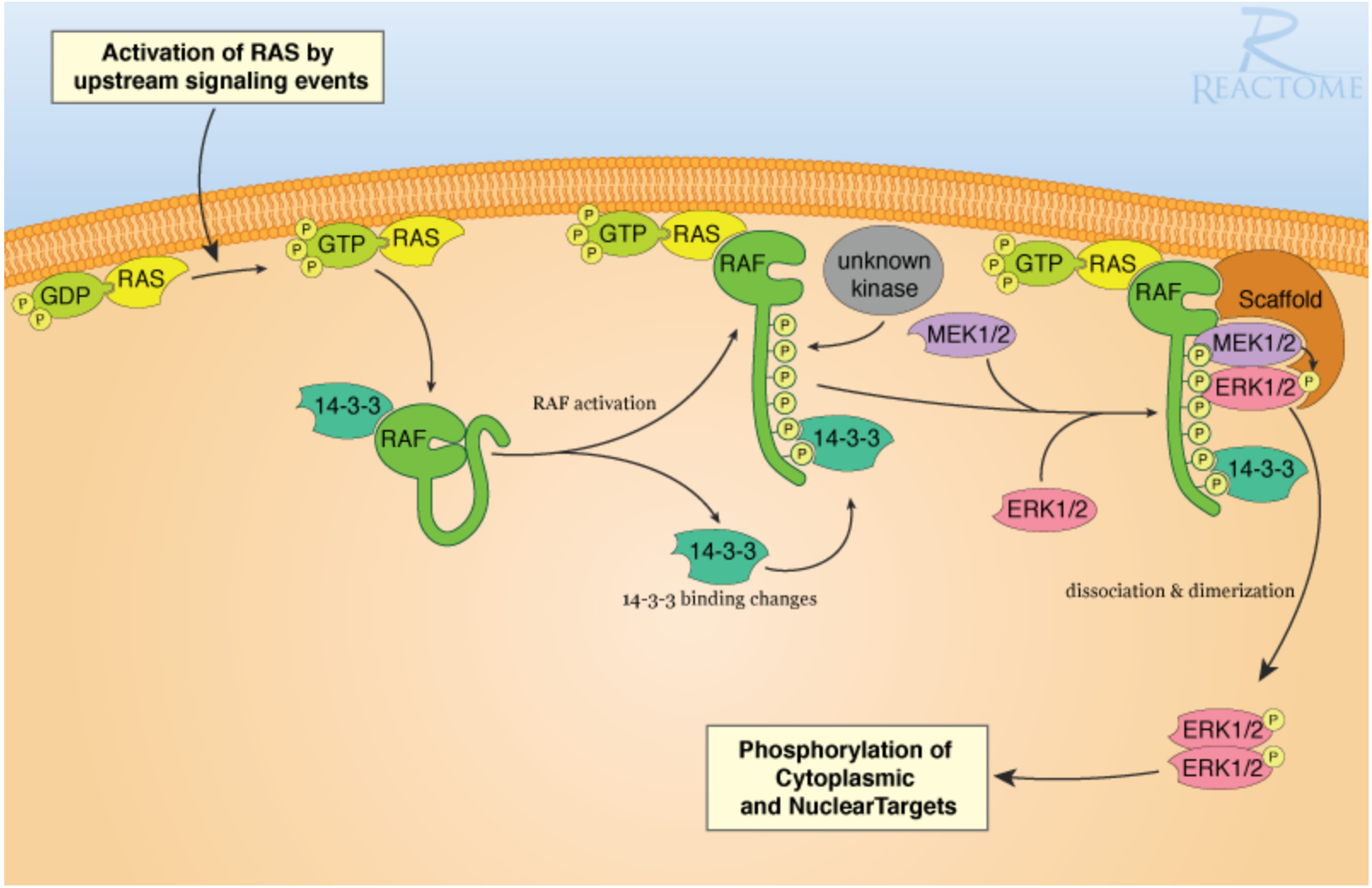


PIBI: Pathways

Summer Term 2017

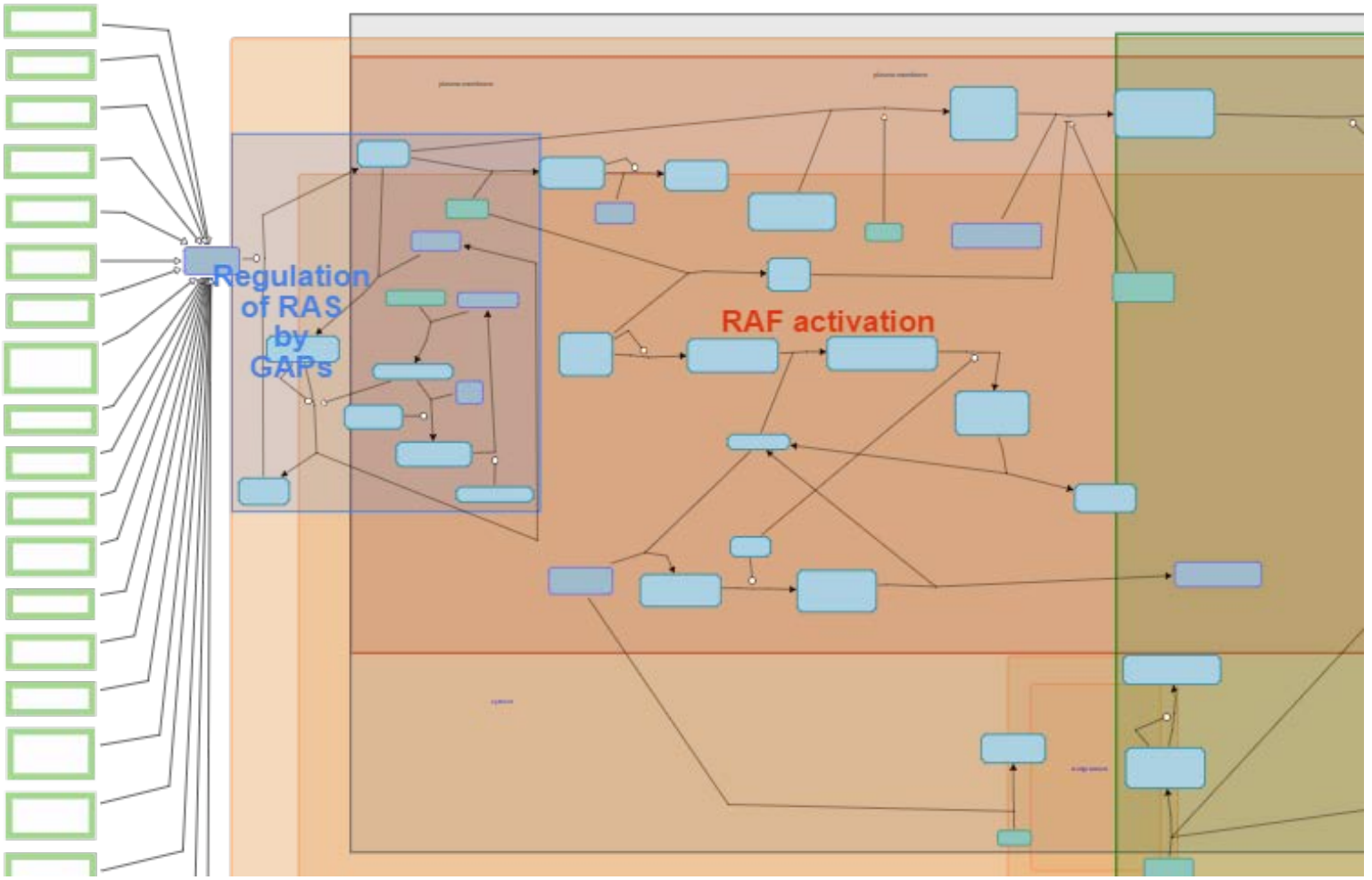
August 21 – September 1

Pathway databases



Pathway databases

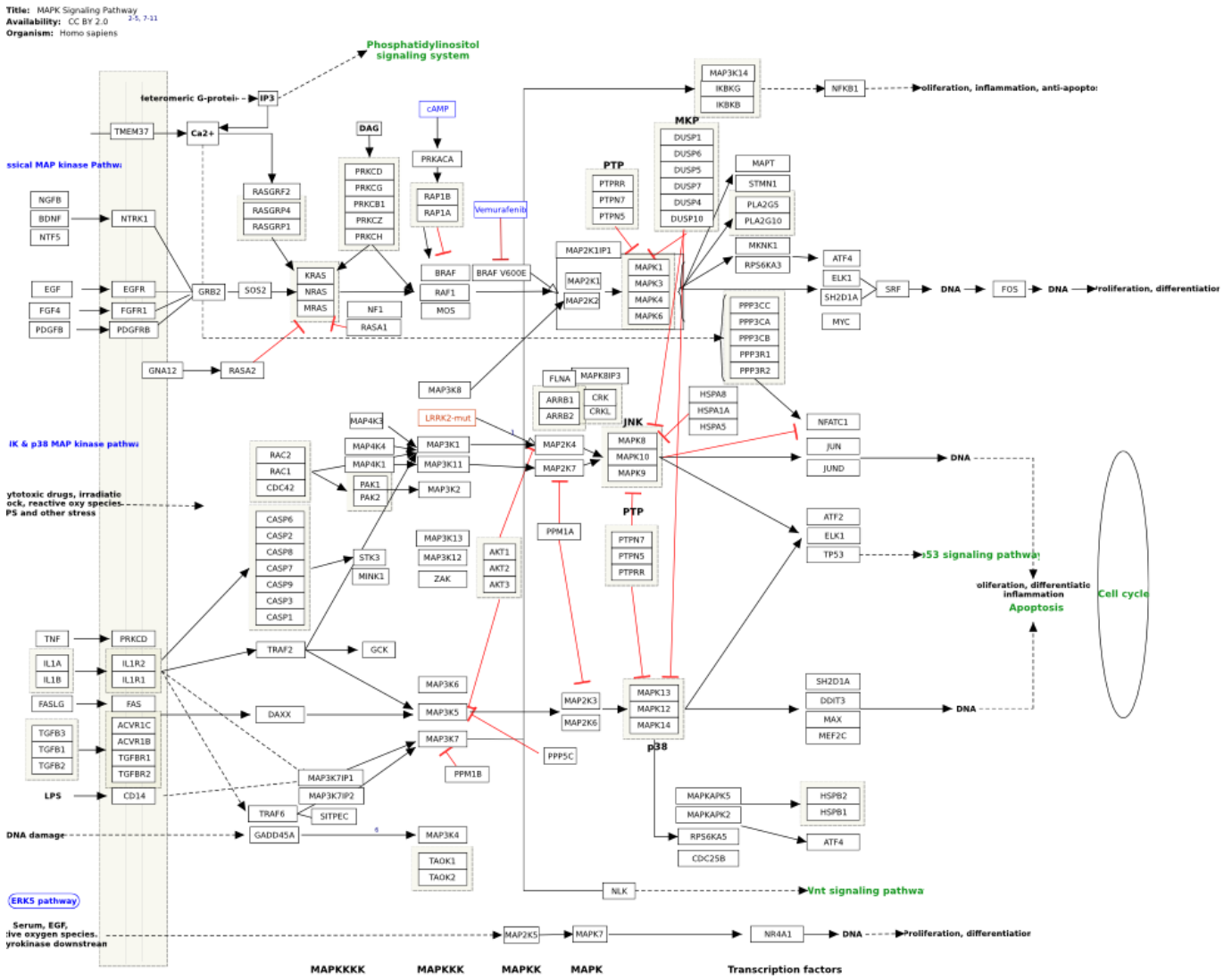
Reactome



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

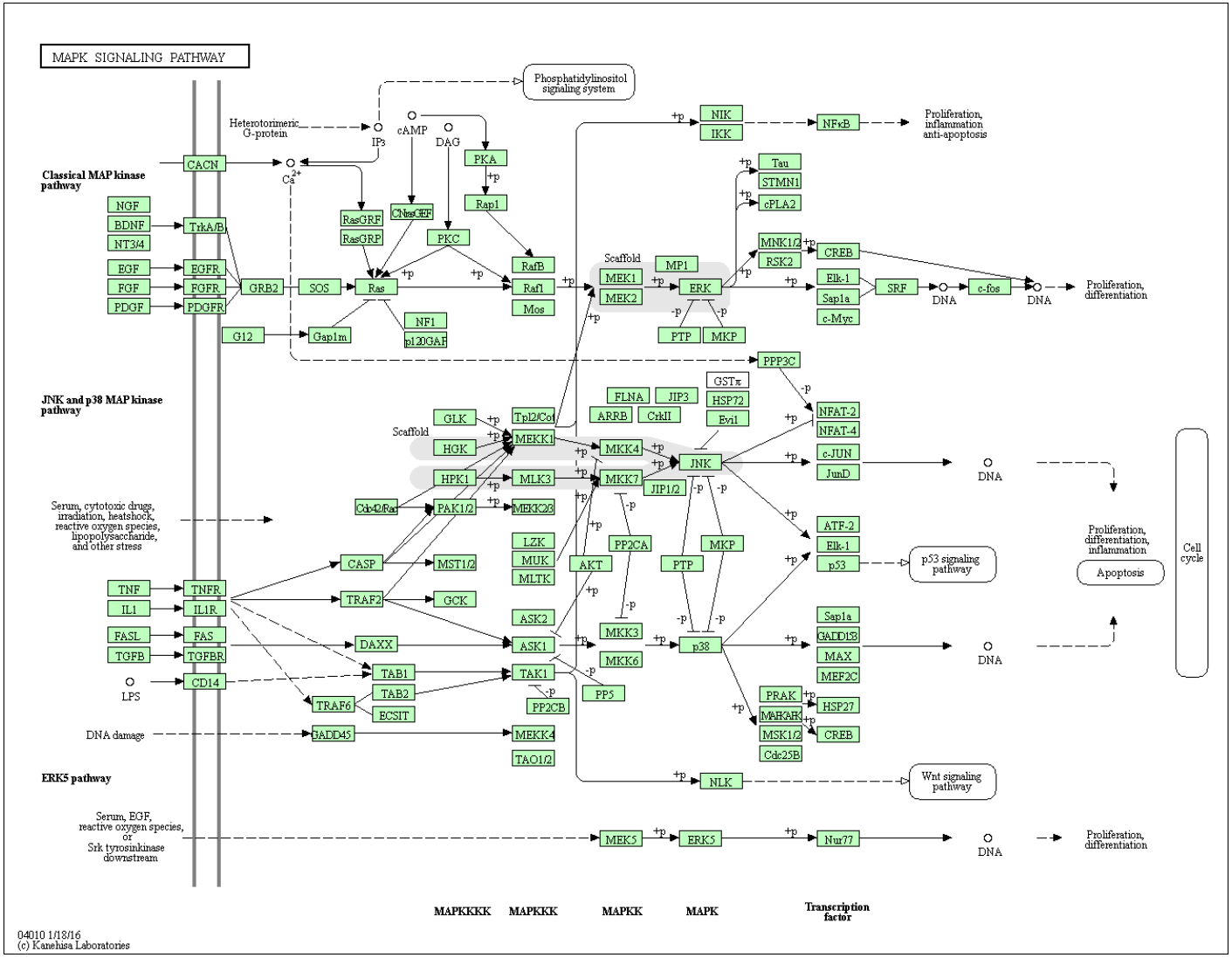
Pathway databases

WikiPathways

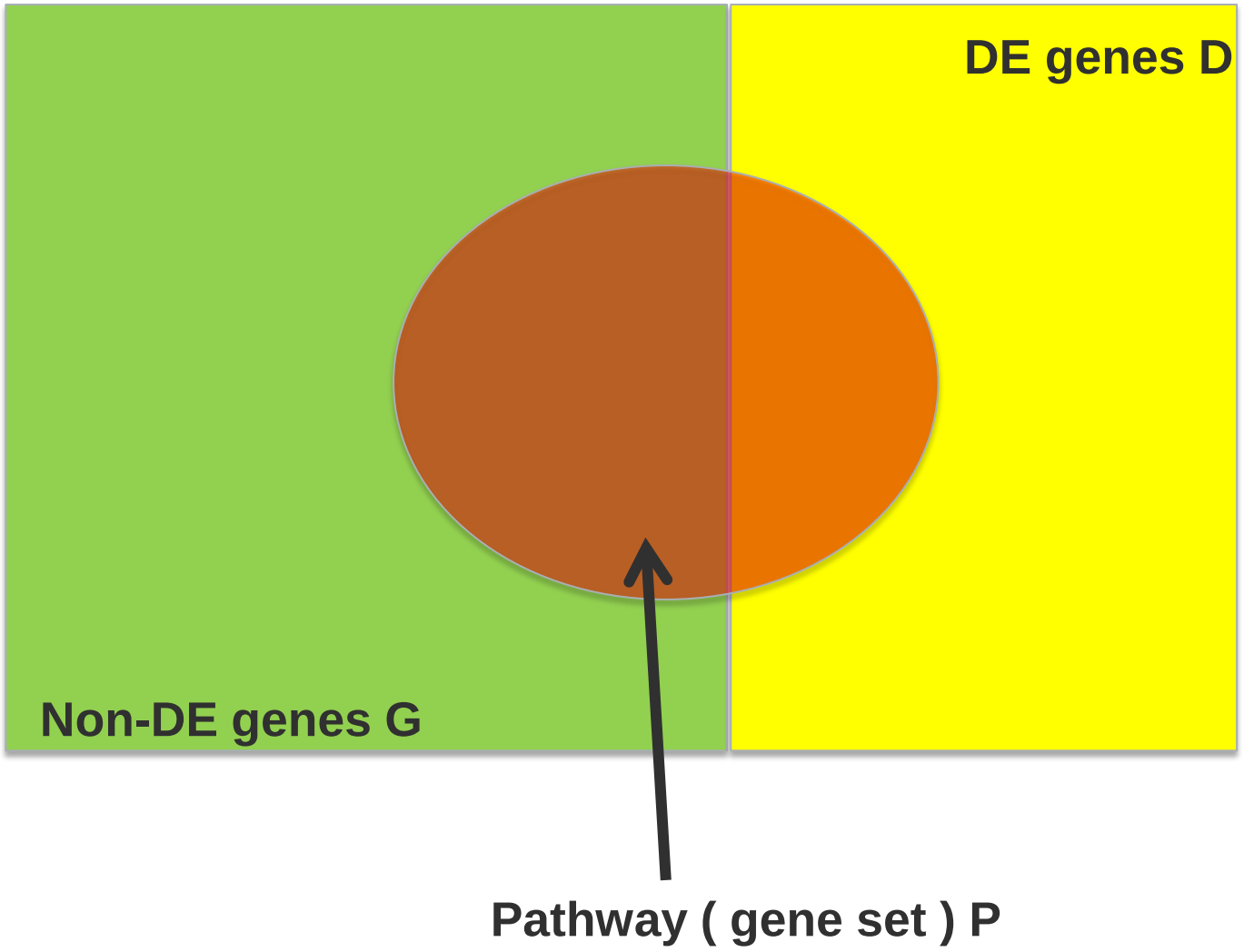


Pathway databases

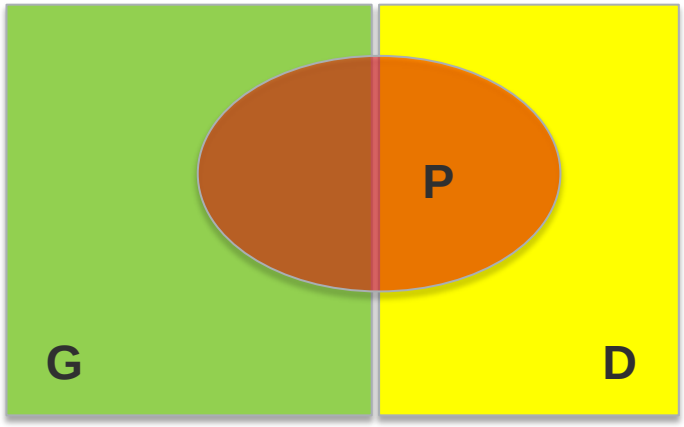
KEGG



Gene set enrichment



Gene set enrichment: hypergeometric test



$$P(\,|P \cap D| = k) = \frac{\binom{|D|}{k}\binom{|G|}{|P|-k}}{\binom{|G \cup D|}{|P|}}$$

$$E(\,|P \cap D|\,) = \frac{|D|}{|D \cup G|} \cdot |P|$$

If $E(\,|P \cap D|\,) < |P \cap D|$ then $pval = \sum_{k=|P \cap D|}^{|P|} P(\,|P \cap D| = k\,)$

If $E(\,|P \cap D|\,) \geq |P \cap D|$ then $pval = \sum_{k=0}^{|P \cap D|} P(\,|P \cap D| = k\,)$

Gene set enrichment: issues

- Topology is ignored
- Pathways are historically evolved categorization and not necessarily the best way to functionally interpret omics data
- Statistical significance of differential expression may not be the most appropriate way to measure “activity” in a network context

- Biology is organized across different layers of organization
  (→ multi-omics aka ultra-omics aka trans-omics aka super-omics aka mega-omics aka poly-omics aka …)

Tasks for today

- Access KEGG over its REST API, extract gene sets for all pathways and store them in an appropriate data format
- Perform a (simple) gene set enrichment analysis with the genes you found in your DE notebook
- Visualize gene expression in the context of a pathway

Representational state transfer webservices

- Stateless: the same query returns the same result every time
- No login or user info required
- For the tasks in this course, the returned values are mostly plain text tables

RESTful webservices

Example:

General info about the Homo Sapiens genome in KEGG

http://rest.kegg.jp/info/hsa

```
T01001          Homo sapiens (human) KEGG Genes Database
hsa             Release 83.0+/08-10, Aug 17
                Kanehisa Laboratories
                39,524 entries

linked db       pathway
                brite
                module
                ko
                genome
                enzyme
                disease
                drug
                dgroup
                ncbi-geneid
                ncbi-proteinid
                uniprot
```