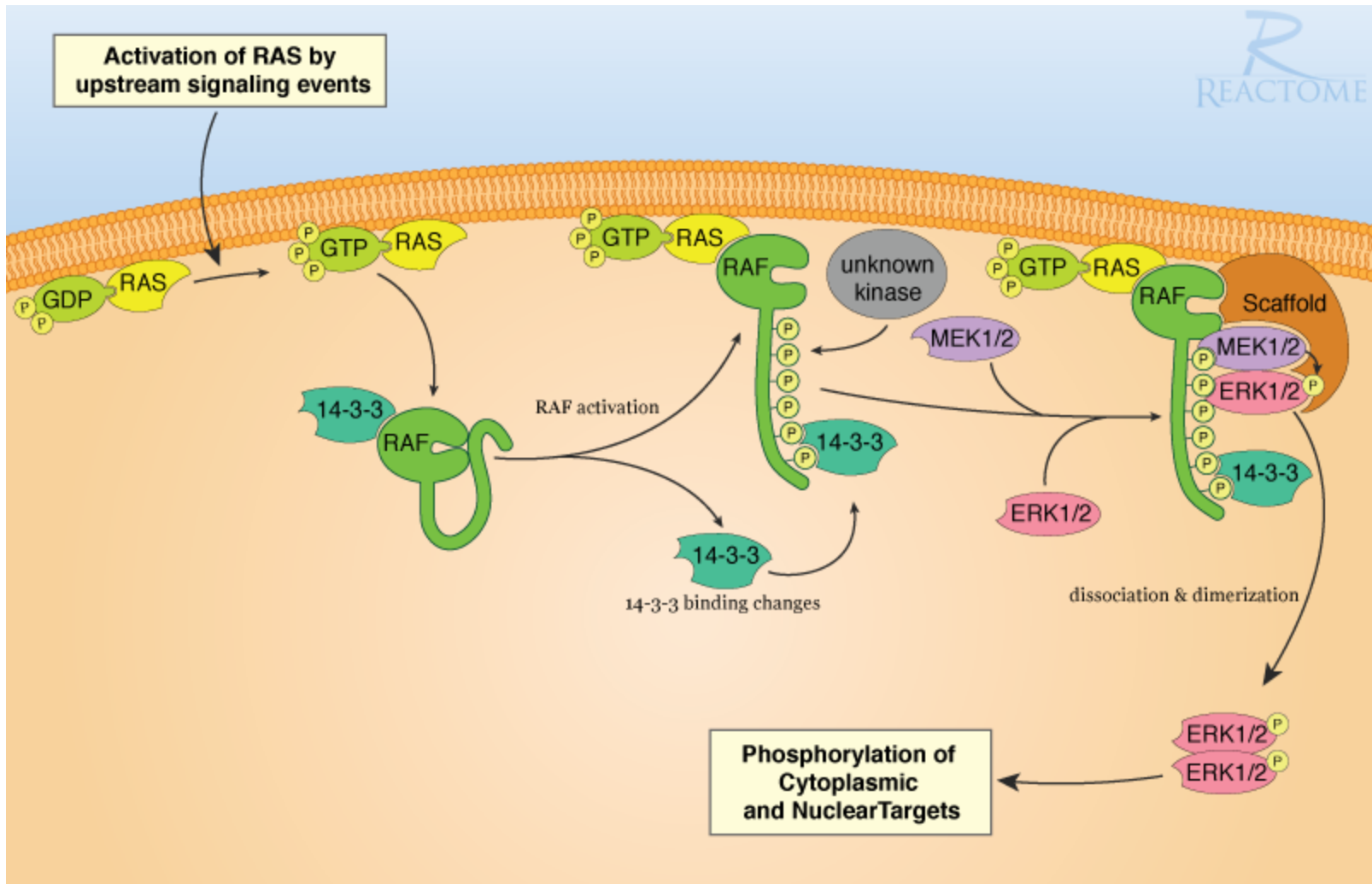


## **PIBI: Pathways**

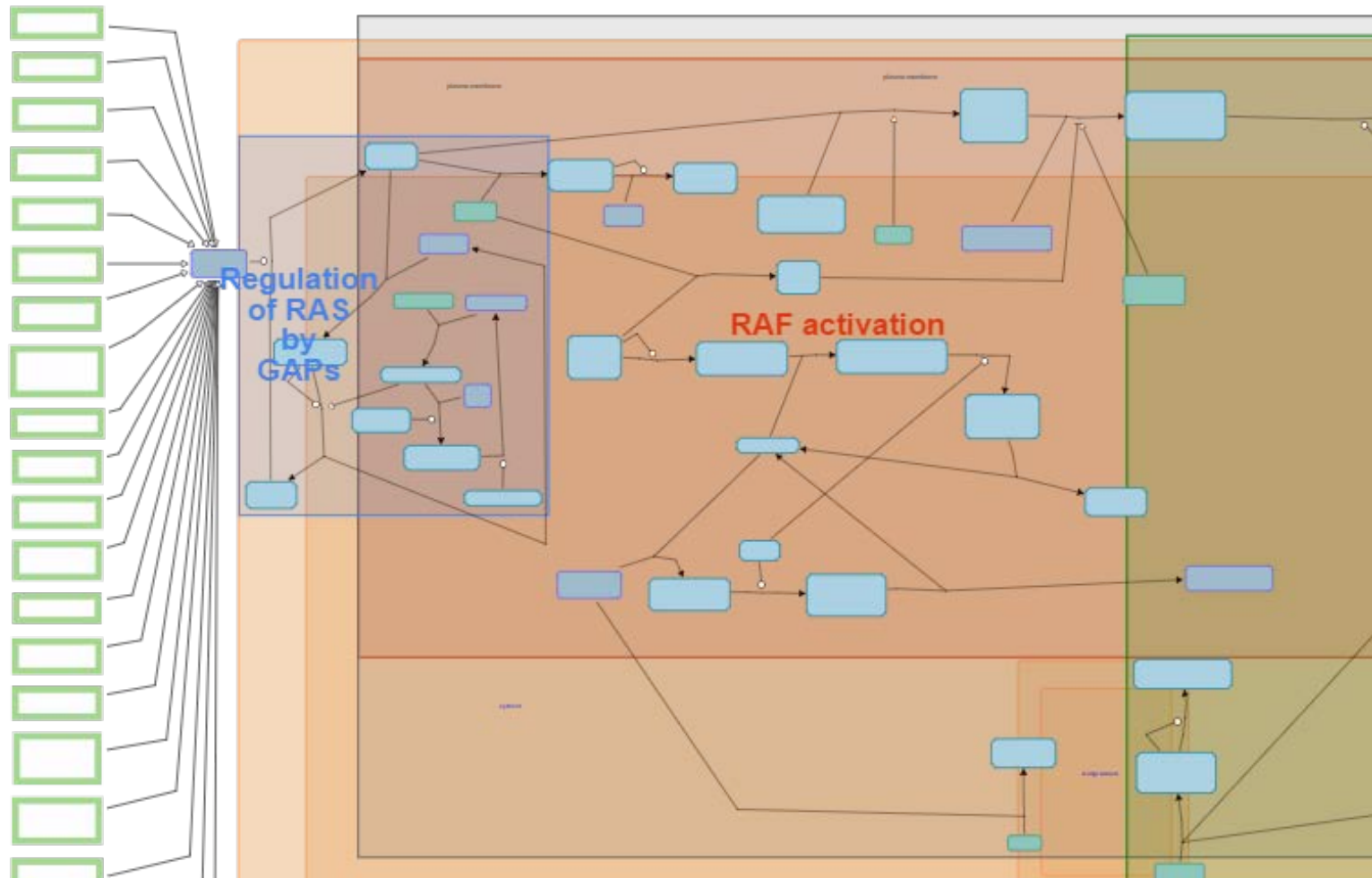
Summer Term 2017

August 21 – September 1





## Reactome



The diagram illustrates the MAPK signaling pathways, showing the cascades from various receptors to transcription factors and cellular outcomes. The pathways are organized into columns representing the hierarchy: MAPKKKK, MAPKKK, MAPKK, MAPK, and Transcription factors.

**Classical MAPK Pathway:** Receptors (NGF, BDNF, NTF5, EGF, FGF4, PDGFRB) activate NTRK1, EGFR, FGFR1, and PDGFRB, which activate GRB2, leading to SOS2, RASGRF2, RASGRF4, and RASGRP1. These activate KRAS, NRAS, and RASGRP1, which then activate BRAF, RAF1, and MOS. BRAF V600E is a common mutation. These pathways lead to MAP2K1IP1, MAP2K1, and MAP2K2, which activate MAPK1, MAPK3, MAPK4, and MAPK6. These MAPKs then activate MAP3K14, IKKKG, and IKKB, leading to NFKB1, which promotes proliferation, inflammation, and anti-apoptosis.

**JNK/SAPK Pathway:** Receptors (TNF, IL1A, IL1B, FASLG, TGFBR3, TGFBR1, TGFBR2, LPS) activate PRKCD, IL1R2, IL1R1, FAS, ACVR1C, ACVR1B, TGFBR1, TGFBR2, and CD14. These activate TRAF2, TRAF6, and TRAF1, which lead to GSK3, IKKKG, and IKKB. These pathways lead to MAP2K3, MAP2K6, and MAP2K7, which activate MAPK8, MAPK10, and MAPK9. These MAPKs then activate MAP3K13, MAPK12, and MAPK14, leading to NFATC1, JUN, JUND, ATF2, ELK1, TP53, SH2D1A, DDIT3, MAX, and MEF2C. These transcription factors promote proliferation, differentiation, inflammation, and apoptosis.

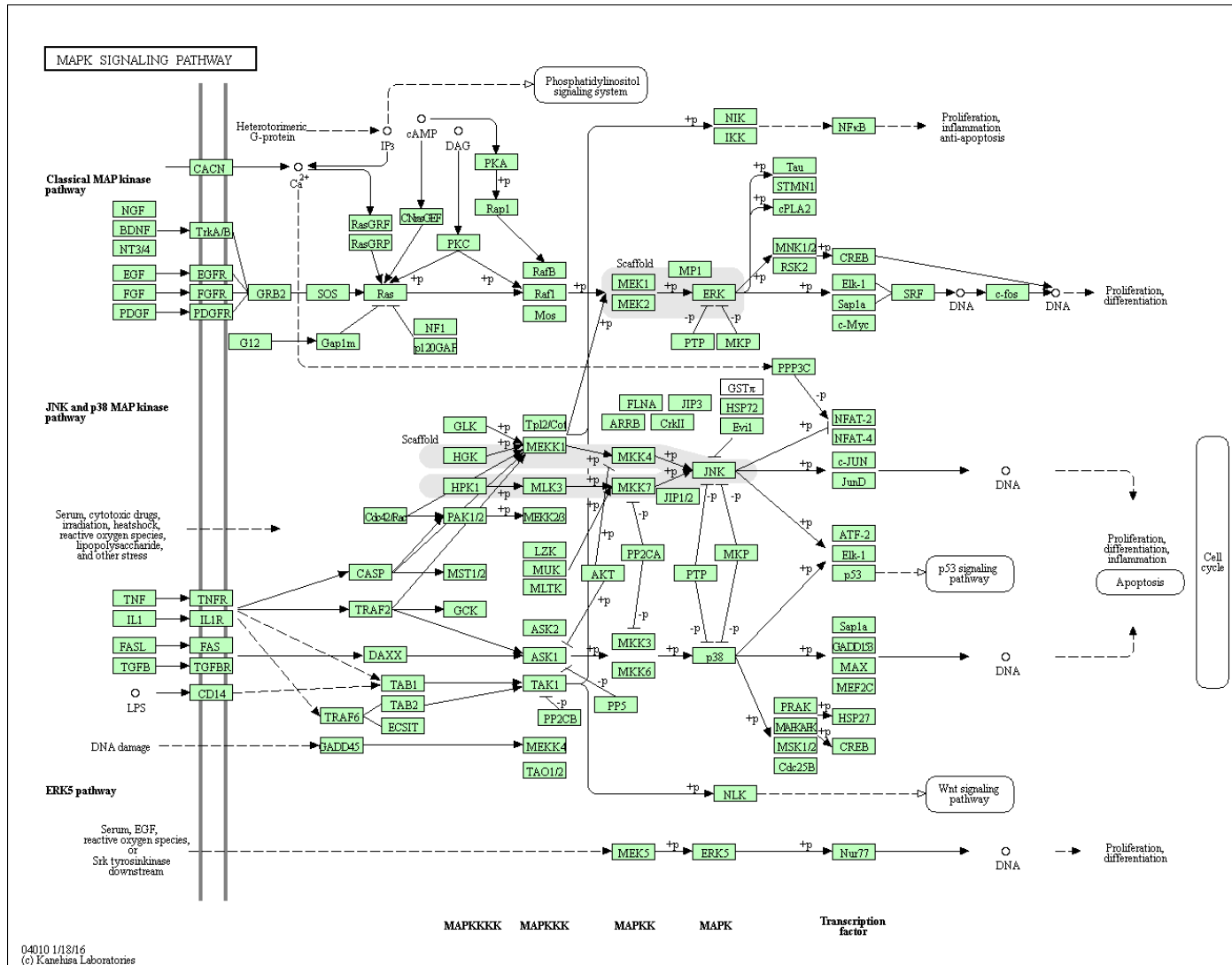
**p38 Pathway:** Receptors (TNF, IL1A, IL1B, FASLG, TGFBR3, TGFBR1, TGFBR2, LPS) activate PRKCD, IL1R2, IL1R1, FAS, ACVR1C, ACVR1B, TGFBR1, TGFBR2, and CD14. These activate TRAF2, TRAF6, and TRAF1, which lead to GSK3, IKKKG, and IKKB. These pathways lead to MAP2K3, MAP2K6, and MAP2K7, which activate MAPK8, MAPK10, and MAPK9. These MAPKs then activate MAP3K13, MAPK12, and MAPK14, leading to NFATC1, JUN, JUND, ATF2, ELK1, TP53, SH2D1A, DDIT3, MAX, and MEF2C. These transcription factors promote proliferation, differentiation, inflammation, and apoptosis.

**ERK5 Pathway:** Receptors (TNF, IL1A, IL1B, FASLG, TGFBR3, TGFBR1, TGFBR2, LPS) activate PRKCD, IL1R2, IL1R1, FAS, ACVR1C, ACVR1B, TGFBR1, TGFBR2, and CD14. These activate TRAF2, TRAF6, and TRAF1, which lead to GSK3, IKKKG, and IKKB. These pathways lead to MAP2K3, MAP2K6, and MAP2K7, which activate MAPK8, MAPK10, and MAPK9. These MAPKs then activate MAP3K13, MAPK12, and MAPK14, leading to NFATC1, JUN, JUND, ATF2, ELK1, TP53, SH2D1A, DDIT3, MAX, and MEF2C. These transcription factors promote proliferation, differentiation, inflammation, and apoptosis.

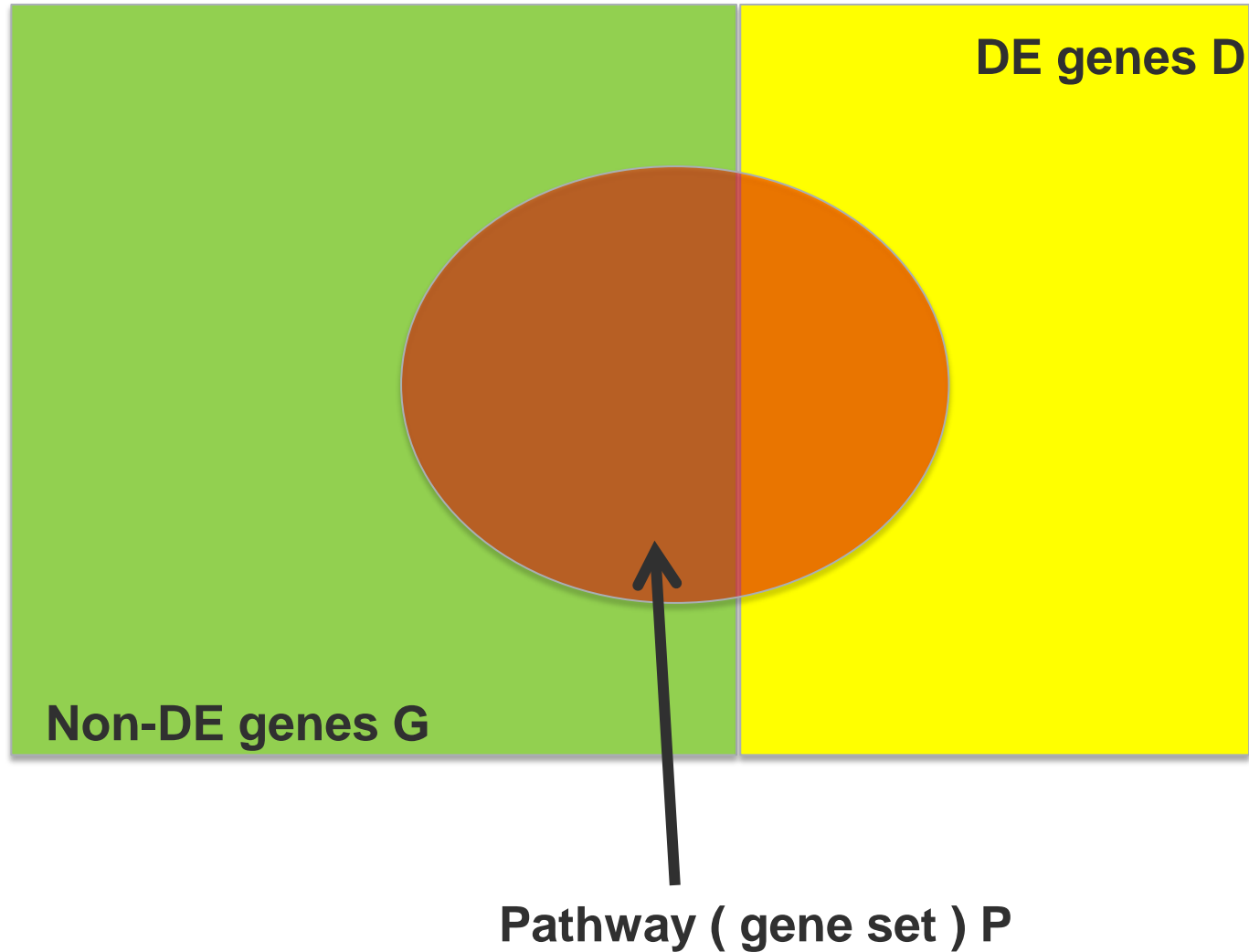
**Cell Cycle:** The pathways lead to DNA, which promotes proliferation, differentiation, inflammation, and apoptosis.



KEGG

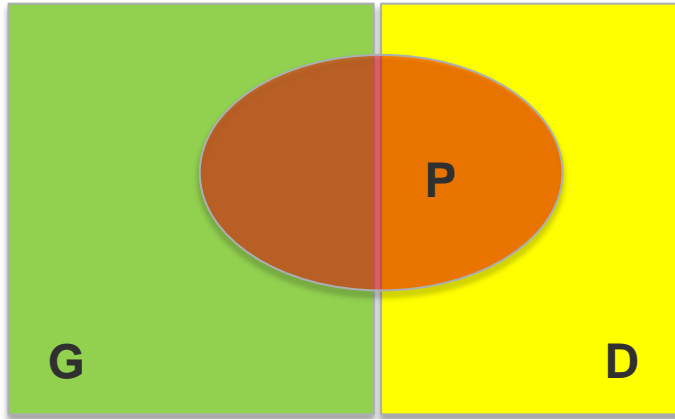








# Gene set enrichment: hypergeometric test



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

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

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  $pval = \sum_{k=0}^{|P \cap D|} P(|P \cap D| = k)$



- 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 ... )
-





- 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 webservice

- 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
-



## 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
```