

# Tissue-specific pathway comparison in mammalian organisms

Nadezhda T. Doncheva<sup>1,2</sup>

Lars Juhl Jensen<sup>2</sup>, Jan Gorodkin<sup>1</sup>

<sup>1</sup>Center for non-coding RNA in Technology and Health &

<sup>2</sup>Novo Nordisk Foundation Center for Protein Research  
University of Copenhagen, Denmark



**DET FRIE  
FORSKNINGSRÅD**  
DANISH COUNCIL  
FOR INDEPENDENT  
RESEARCH

# AniGen project goal

- ▶ *Which animal models would be most useful to generate reliable hypotheses about human with respect to a given phenotype?*

# AniGen project goal

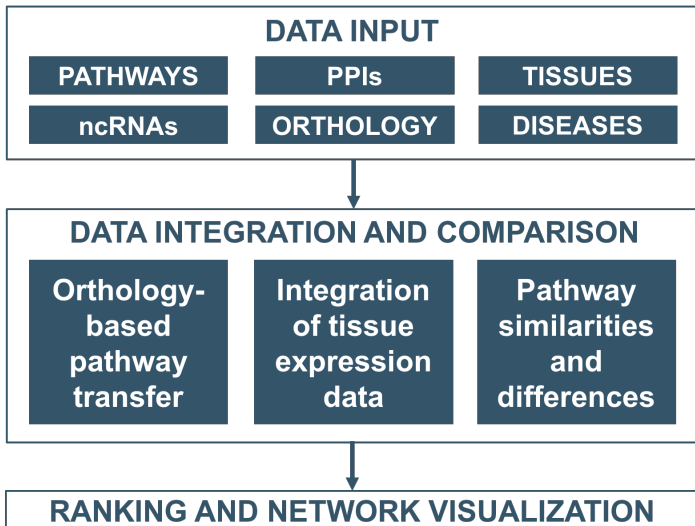
- ▶ *Which animal models would be most useful to generate reliable hypotheses about human with respect to a given phenotype?*
- ▶ Comprehensive comparison between mammalian organisms on different levels, e.g. tissue expression and pathways

# AniGen project goal

- ▶ *Which animal models would be most useful to generate reliable hypotheses about human with respect to a given phenotype?*
- ▶ Comprehensive comparison between mammalian organisms on different levels, e.g. tissue expression and pathways
- ▶ Experimental analysis and validation by our collaborators



# Pathway analysis workflow



How much is the pathway overlap between human and model organisms?

# Pathways in model organisms

- ▶ Only few experimentally determined interactions
- ⇒ Pathway resources use orthology transfer from human

# Pathways in model organisms

- ▶ Only few experimentally determined interactions
- ⇒ Pathway resources use orthology transfer from human
- ▶ However, no unified and consistent definition of orthology and no good consensus

# Pathways in model organisms

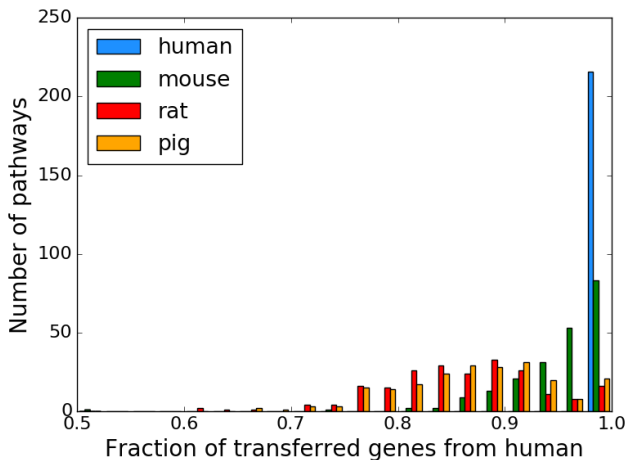
- ▶ Only few experimentally determined interactions
- ⇒ Pathway resources use orthology transfer from human
- ▶ However, no unified and consistent definition of orthology and no good consensus
- ⇒ Perform our own consistent orthology-based pathway transfer

# Pathways in model organisms

- ⇒ We use curated human pathways and do a consistent orthology transfer using eggNOG (Huerta-Cepas *et al.* 2016)

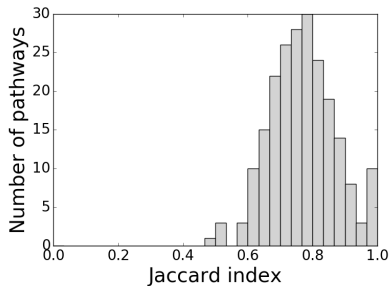
# Pathways in model organisms

- ⇒ We use curated human pathways and do a consistent orthology transfer using eggNOG (Huerta-Cepas *et al.* 2016)
- ▶ Example for the transfer of  $\sim 200$  KEGG pathways

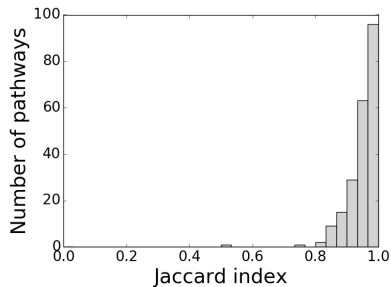


# KEGG pathways comparison

Human - mouse - pig - rat



Human - mouse

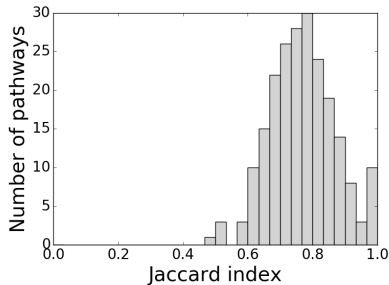


- ▶ Overlap is measured by the Jaccard index of the gene overlap

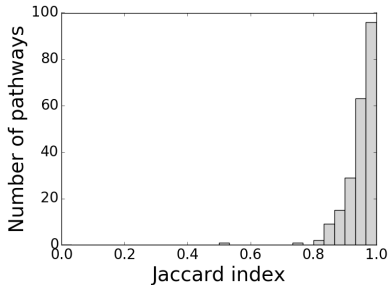


# KEGG pathways comparison

Human - mouse - pig - rat



Human - mouse



- ▶ Overlap is measured by the Jaccard index of the gene overlap
- ▶ Between human, mouse, rat and pig, all KEGG pathways overlap more than 50% and half of them - at least 75%
- ▶ Between human and mouse, the gene overlap is 100% for 1/3 of the pathways and above 80% for all pathways (except for 2)

## Next step

- ▶ Orthology-based comparison highlights the similarities between human and other organisms
- ▶ But we are also interested in the differences

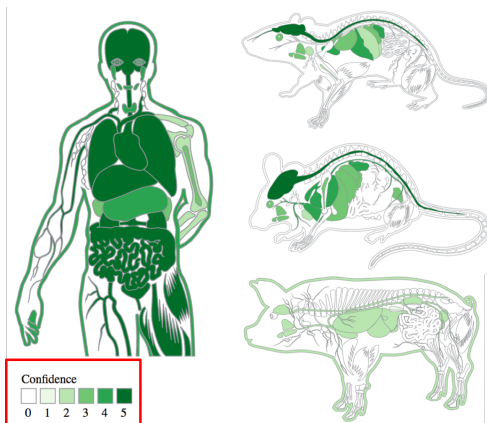
# Next step

- ▶ Orthology-based comparison highlights the similarities between human and other organisms
  - ▶ But we are also interested in the differences
- ⇒ Create tissue-specific pathways for each organism by integrating the expression data

Which tissue expression data do we use?

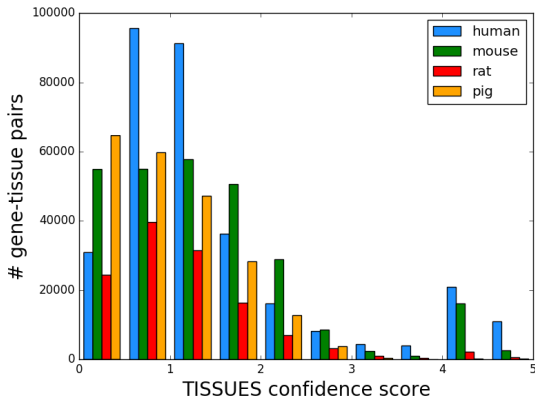
# TISSUES web resource

- ▶ Tissue expression for human, mouse, rat and pig
- ▶ Integrates data from transcriptomics, text mining and manual curation
- ▶ Provides confidence scores based on a gold standard
- ▶ That are comparable across datasets and organisms
- ▶ Palasca *et al.*, 2018.



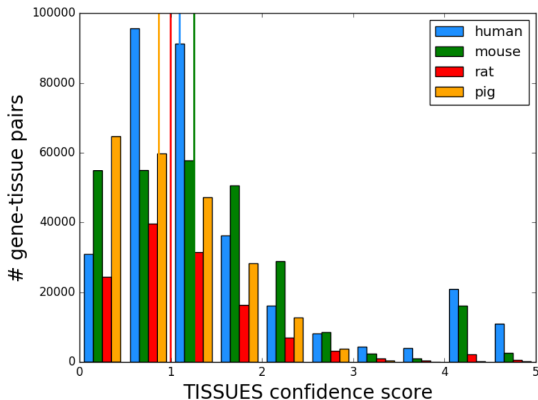
# When is a gene expressed?

- ▶ We need a TISSUES confidence cutoff
- ▶ However, the same cutoff for all organisms is not suitable, since amount and quality of data varies a lot between organisms



# When is a gene expressed?

- ▶ We need a TISSUES confidence cutoff
- ⇒ Use organism-specific cutoffs based on the 50 percentile of confidence scores for each organism (or even tissue?)



Organism-specific  
confidence cutoff  
Human: 1.09  
Mouse: 1.25  
Rat: 0.99  
Pig: 0.86

How do we integrate the pathway and tissue expression data?

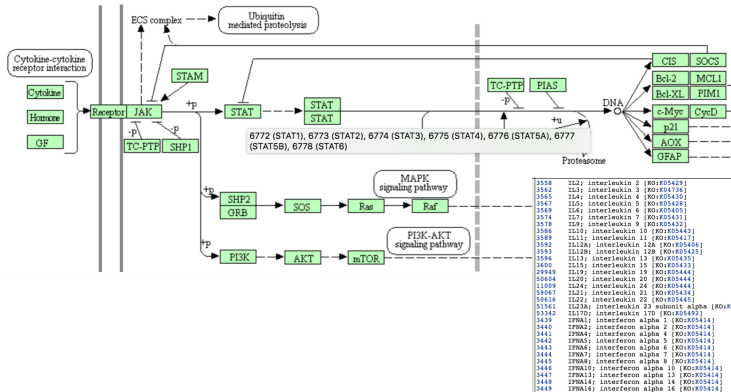


How do we integrate the pathway and tissue expression data?

*Exploratory analysis on a KEGG pathway*

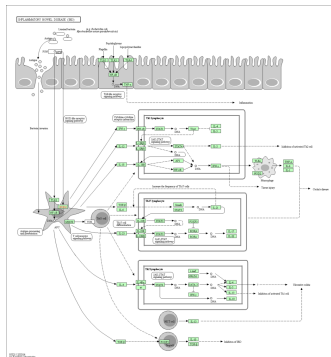
# Case study: JAK-STAT signaling pathway

## JAK-STAT SIGNALING PATHWAY

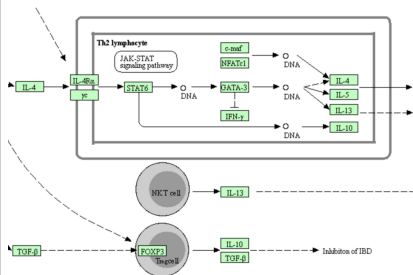
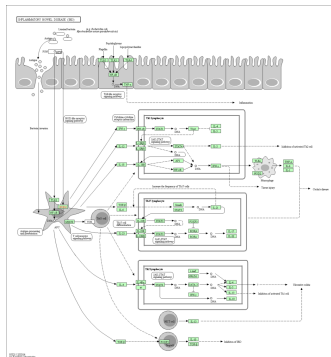


- ▶ KEGG pathway (hsa04630) is very general, each protein box combines several genes, the pathway entry contains more than 100 genes

# Example: KEGG IBD pathway

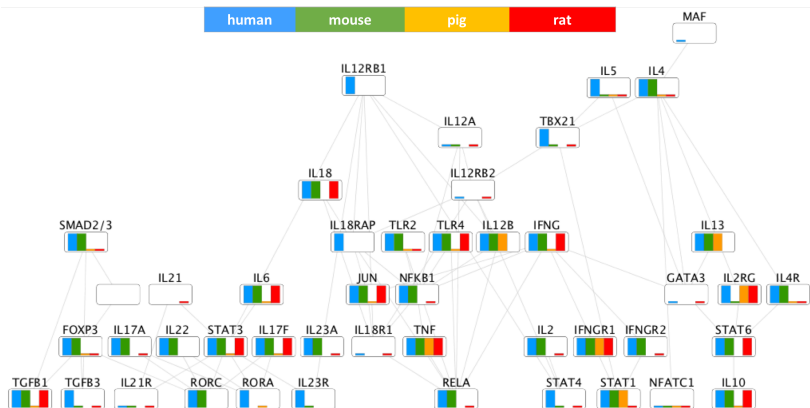


# Example: KEGG IBD pathway



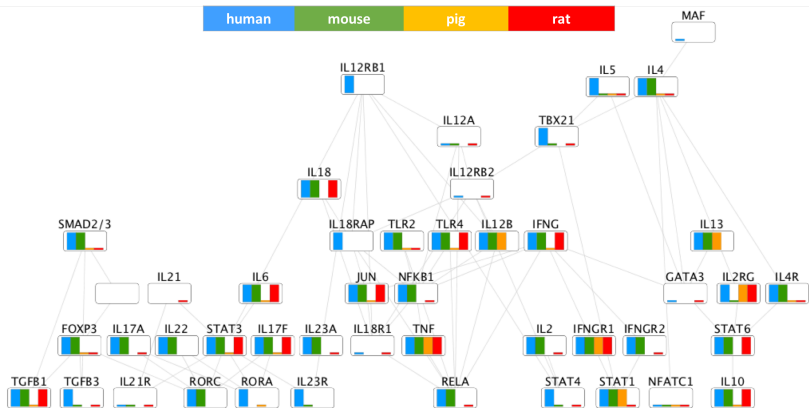
- ▶ The IBD pathway (hsa05321) is much better annotated in KEGG

# KEGG IBD pathway in Cytoscape



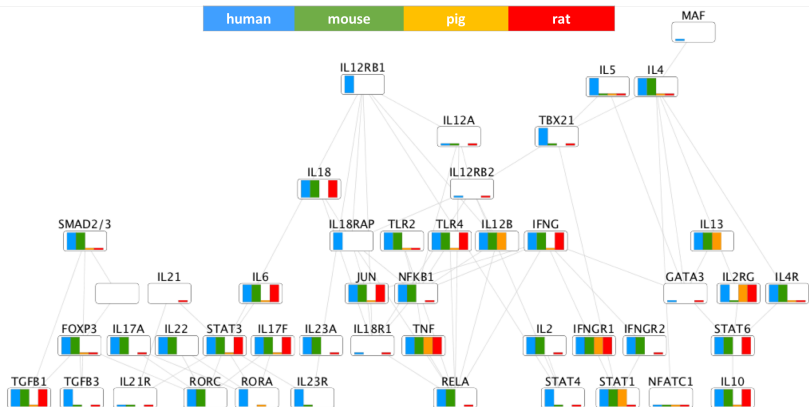
- Annotated with TISSUES expression evidence for **colon**: full bar for confidence above 1.0 and a stripe for lower confidence.

# KEGG IBD pathway in Cytoscape



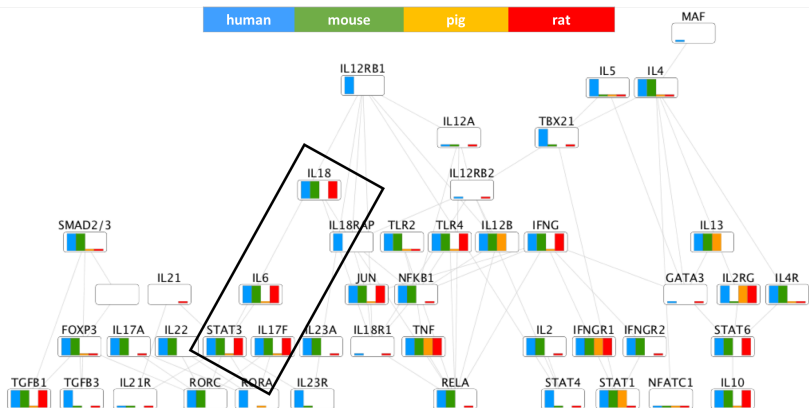
- ▶ ~ 60 % of the genes are expressed in mouse and human colon tissue, while only 26 % for rat and 15 % for pig.

# KEGG IBD pathway in Cytoscape



- ▶ ~ 60 % of the genes are expressed in mouse and human colon tissue, while only 26 % for rat and 15 % for pig.
- ⇒ Use organism- or tissue-specific TISSUES confidence cutoffs instead of an universal one

# KEGG IBD pathway in Cytoscape



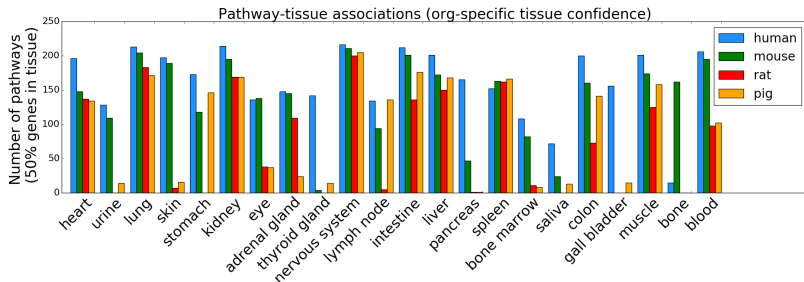
- Find groups of connected genes with similar "expression pattern" across organisms/tissues



How to integrate the pathway and tissue  
expression data?

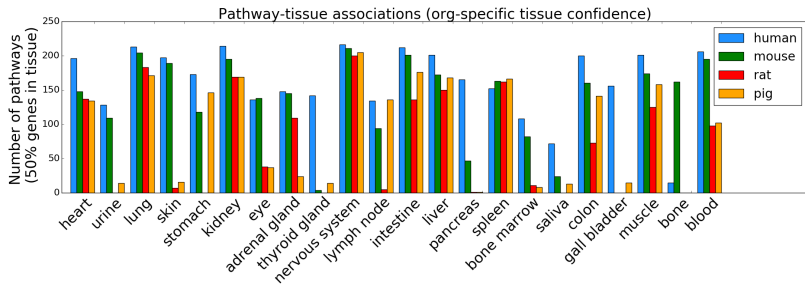
*Pathway-tissue associations*

# Pathway-tissue associations



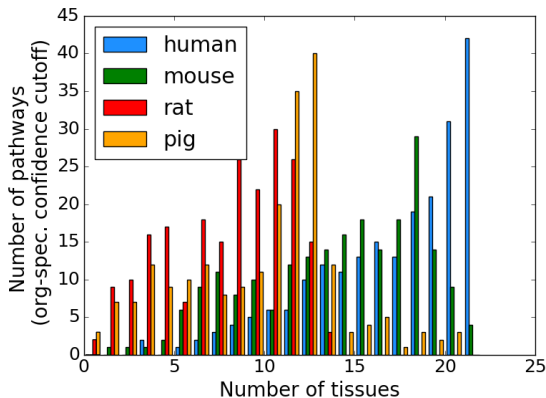
- ▶ Number of pathways "expressed" in a given tissue
- ▶ Measure "pathway expression" by a proportion (e.g. 50%) of confidently expressed pathway genes
- ▶ Use fixed confidence cutoff (universal, org-specific, tissue-specific)

# Pathway-tissue associations



- ▶ Number of pathways "expressed" in a given tissue
  - ▶ Measure "pathway expression" by a proportion (e.g. 50%) of confidently expressed pathway genes
  - ▶ Use fixed confidence cutoff (universal, org-specific, tissue-specific)
- ⇒ Are there tissue-conserved pathways among the organisms?

# Distribution of pathway-tissue associations



- ▶ Are there tissue-specific pathways?
- ▶ How many pathways are "expressed" in how many tissues (X% of the pathway genes confidently expressed)?

# Next steps

- ▶ Given our comprehensive pathway-tissue-organism framework, what interesting questions can we answer?
- ▶ Perform a disease-related comparison for a disease studied in several model organisms
- ▶ Provide pathway transfer and tissue annotation in Cytoscape

# Acknowledgements

▶ Jan Gorodkin

▶ Lars Juhl Jensen

▶ Oana Palasca

▶ Helen Cook

▶ Reza Yarani

▶ Flemming Pociot

▶ Thomas Litman



# THANK YOU!

