Modeling of human pathways in animals: possibilities and limitations

Nadezhda T. Doncheva

Center for non-coding RNA in Technology and Health & Novo Nordisk Foundation Center for Protein Research, University of Copenhagen
Enable researchers to answer the following question:

Which animal models would be most useful to generate reliable hypotheses about human with respect to a given phenotype?
The big picture: AniGen project

▶ Enable researchers to answer the following question:

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

▶ Perform a comprehensive comparison between human and well-known animal models such as mouse, rat, and pig
The big picture: AniGen project

- Enable researchers to answer the following question:
  
  *Which animal models would be most useful to generate reliable hypotheses about human with respect to a given phenotype?*

- Perform a comprehensive comparison between human and well-known animal models such as mouse, rat, and pig

- *What can we do with the currently available data?*
What can we do with the currently available data?

- **Part I: State of the art**
  What kind of and how much data is available for animal models?

- **Part II: Pathways in animal models**
  How can we combine the available data to study pathways in animal models?
Part I: What kind of and how much data is available for animal models?
Part I: What kind of and how much data is available for animal models?

- Ensembl
- GENEONTOLOGY
- EggNOG 4.5
- RAIN
- STRING
- REACTOME
- TISSUES (Tissue expression database)
- GEO (Gene Expression Omnibus)
- DISEASES (Disease-gene associations mined from literature)
- PubMed
- KEGG
Part I: What kind of and how much data is available for animal models?

- Literature knowledge (PubMed)
- Tissue expression data (TISSUES)
- Protein interactions (STRING)
- Pathways (KEGG)
Text mining: PubMed abstracts

- ~30 mio abstracts on life sciences and biomedical topics
- Text-mining these abstracts by dictionary-based named entity recognition using *tagger* (Szklarczyk et al. (2015), *NAR*)
Text mining: PubMed abstracts

- ~30 mio abstracts on life sciences and biomedical topics
- Text-mining these abstracts by dictionary-based named entity recognition using *tagger* (Szklarczyk *et al.* (2015), *NAR*)

⇒ How many abstracts mention each organism of interest?
⇒ How many abstracts mention the genes of this organism?

<table>
<thead>
<tr>
<th>mentions</th>
<th>mouse</th>
<th>rat</th>
<th>pig</th>
</tr>
</thead>
<tbody>
<tr>
<td>organism</td>
<td>1 217 133</td>
<td>1 309 469</td>
<td>132 358</td>
</tr>
<tr>
<td>gene</td>
<td>652 808</td>
<td>405 278</td>
<td>21 917</td>
</tr>
</tbody>
</table>
Text mining: PubMed abstracts

- ~30 mio abstracts on life sciences and biomedical topics
- Text-mining these abstracts by dictionary-based named entity recognition using *tagger* (Szklarczyk et al. (2015), *NAR*)

⇒ How many abstracts mention each organism of interest?

⇒ How many abstracts mention the genes of this organism?

<table>
<thead>
<tr>
<th>organism</th>
<th>mouse (1 217 133)</th>
<th>rat (1 309 469)</th>
<th>pig (132 358)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>652 808</td>
<td>405 278</td>
<td>21 917</td>
</tr>
</tbody>
</table>

⇒ Challenge: pig is not studied as much as mouse and rat
Expression data: TISSUES database

- Covers human, mouse, rat, pig
- 14 transcriptomic datasets
- Text mining and manual curation
- Confidence scores comparable across datasets and organisms

https://tissues.jensenlab.org

Palasca et al. (2018), Database
Expression data: TISSUES database

- Covers human, mouse, rat, pig
- 14 transcriptomic datasets
- Text mining and manual curation
- Confidence scores comparable across datasets and organisms

⇒ However, comparable does not mean equal
⇒ Only few tissues covered by at least one/two datasets

https://tissues.jensenlab.org

Palasca et al. (2018), Database
Datasets (rows) and tissues covered (columns) in each organism; tissues supported by at least one (grey) or two (black) datasets are highlighted
Interaction databases: STRING

1380 mio interactions between 9.6 mio proteins in 2031 organisms

High-confidence interactions (score ≥ 0.8) for each evidence type
Interaction databases: STRING

1380 mio interactions between 9.6 mio proteins in 2031 organisms

High-confidence interactions (score $\geq 0.8$) for each evidence type

- Very few experimentally determined interactions for animals
- We need orthology transfer from human & data integration
Part II: Pathways in animal models
Part II: Pathways in animal models

How do we combine this data to study human pathways in animals?
Part II: Pathways in animal models

How do we combine this data to study human pathways in animals?

- **eggNOG**: orthology relationships between mammals
- **KEGG**: curated & high-quality human pathways
- **TISSUES**: healthy tissue expression data in mammals
Orthology-based transfer of 216 human pathways

35% of the KEGG pathways overlap completely between human & mouse, while only 10% between human & pig or human & rat.

We assess the pathway gene/interaction overlap between human and other organisms to highlight their similarities.
Orthology-based transfer of 216 human pathways

- 35% of the KEGG pathways overlap completely between human & mouse, while only 10% between human & pig or human & rat.
Orthology-based transfer of 216 human pathways

35% of the KEGG pathways overlap completely between human & mouse, while only 10% between human & pig or human & rat

We assess the pathway gene/interaction overlap between human and other organisms to highlight their similarities
Are there pathway differences?

Overlay the orthology-transferred KEGG pathways with tissue expression data from the TISSUES database.

For each pathway, organism & tissue, the respective gene is expressed if it has a score above a given confidence cutoff.

A pathway is considered expressed in a tissue if 90% of the pathway genes are expressed above a given confidence cutoff.
Are there pathway differences?

human pathway
Are there pathway differences?

- **Human pathway**
- **Transferred to mouse**

Overlay the orthology-transferred KEGG pathways with tissue expression data from the TISSUES database.

▶ For each pathway, organism & tissue, the respective gene is expressed if it has a score above a given confidence cutoff.

▶ A pathway is considered expressed in a tissue, if 90% of the pathway genes are expressed above a given confidence cutoff.
Are there pathway differences?

⇒ Overlay the orthology-transferred KEGG pathways with tissue expression data from the TISSUES database
Are there pathway differences?

- Overlay the orthology-transferred KEGG pathways with tissue expression data from the TISSUES database

- For each pathway, organism & tissue, the respective gene is expressed if it has a score above a given confidence cutoff
Are there pathway differences?

⇒ Overlay the orthology-transferred KEGG pathways with tissue expression data from the TISSUES database

► For each pathway, organism & tissue, the respective gene is expressed if it has a score above a given confidence cutoff

► A pathway is considered expressed in a tissue, if 90 % of the pathway genes are expressed above a given confidence cutoff
Tissue distribution of the 216 transferred pathways

Given tissue & organism of interest, which are the expressed pathways?

Are these pathways tissue-specific or broadly expressed?

Are there organism- and/or tissue-specific pathways?
Are there pathways or tissues, for which pig is better represented than mouse?
Organism- & tissue-specific pathways

beta-Alanine metabolism
Glycosaminoglycan degradation
Circadian rhythm
N-Glycan biosynthesis
Renal cell carcinoma
Amino sugar and nucleotide sugar metabolism
GABAergic synapse
Glycine, serine and threonine metabolism
Phenylalanine metabolism
Autoimmune thyroid metabolism
Porphyrin and chlorophyll metabolism
Tyrosine metabolism
mTOR signaling pathway
Acute myeloid leukemia
DNA replication
Adipocytokine signaling pathway
Drug metabolism - cytochrome P450
GnRH signaling pathway
Complement and coagulation cascades
Dorso-ventral axis formation
Ascorbate and aldarate metabolism
Fructose and mannose metabolism
Pentose phosphate pathway
Thyroid cancer
Spliceosome
Renin-angiotensin system
Hedgehog signaling pathway
Butanoate metabolism
Type I diabetes mellitus
Starch and sucrose metabolism
Cholinergic synapse
Basal transcription factors
Phototransduction
Vitamin B6 metabolism
Alanine, aspartate and glutamate metabolism
One carbon pool by folate
Inositol phosphate metabolism
Histidine metabolism
Dopaminergic synapse
Organism- & tissue-specific pathways

Is pig a good model for diseases related to spleen?
Conclusions and future work

- Transferred a set of mammalian pathways from human pathways
- Integrated them with tissue expression data
- Identified a set of tissue-and organism-specific pathways
Conclusions and future work

- Transferred a set of mammalian pathways from human pathways
- Integrated them with tissue expression data
- Identified a set of tissue- and organism-specific pathways
- Include non-coding RNAs in the pathways (RAIN)
- Take into account the pathway / network structure
- Include expression data from our mouse and pig disease models
- Perform the same analysis using gene-disease associations
Acknowledgements

▶ Oana Palasca
▶ Jan Gorodkin
▶ Lars Juhl Jensen
▶ Reza Yarani
▶ Flemming Pociot
▶ Thomas Litman
▶ RTH group
▶ Jensen group
THANK YOU!