



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Advanced Computational Systems Biology Practical Sessions (Day 2)

2024-06-18

Bárbara Zita Peters Couto (barbara-zita.peters-couto@uni-heidelberg.de)

PhD student

Sebastian Lobentanzer (sebastian.lobentanzer@uni-heidelberg.de)

Postdoc Fellow



www.saezlab.org



@saezlab



Today's plan

- 1) Theoretical introduction to prior knowledge
- 2) Practical 1: OmniPath exploration
- 3) Practical 2: Expanding our insights from data with prior knowledge

Learning outcomes:

- **understand how biologists can leverage prior knowledge**
- **get exposed to data science basics: some python and the prior knowledge database OmniPath**



Some questions to start:

- How would you define prior knowledge?
- Where do you think/know prior knowledge is used?
- How do you think prior knowledge can be useful for biologists?
- Do you think prior knowledge can be misleading?



Prior Knowledge (PK) is important in many fields

prior knowledge

collocation in English

MEANINGS of **prior** and **knowledge**

These words are often used together. Click on the links below to explore the meanings. Or, [see other collocations with **knowledge**](#).

prior *adjective* [before noun]

UK /praɪə/ US /praɪr/

formal

existing or happening before something else, or before a ...

[See more at **prior**](#)

knowledge *noun*

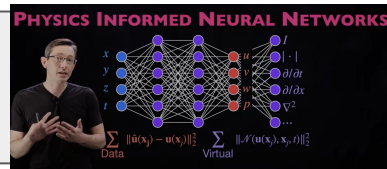
UK /ˈnɒl.ɪdʒ/ US /ˈnɑː.lɪdʒ/

understanding of or information about a subject that you get by experience or study, either known by one person or by ...

Psychology → effect of PK on learning outcomes?

Business → data-driven decisions based on prior customer knowledge

Physics →

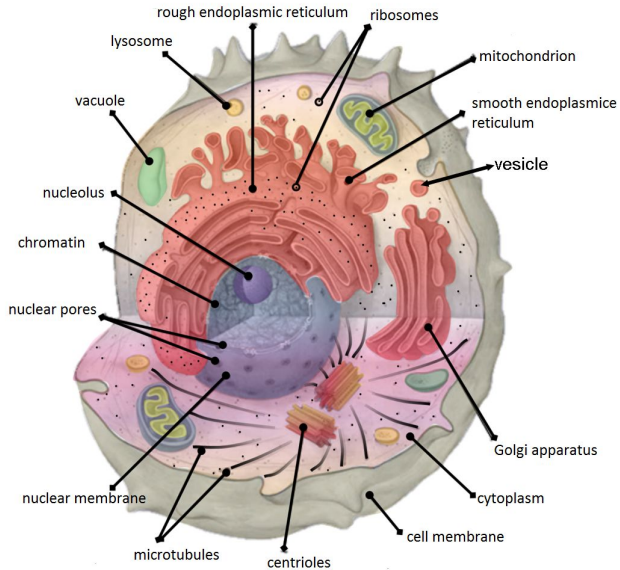


Medicine → clinical trials on drug repurposing



Prior Knowledge is also important in biology

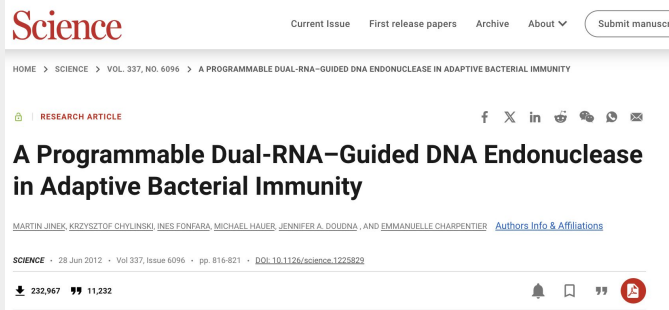
- We have been collecting biological data for centuries
- Recent high throughput technologies (e.g. NGS) enable us to understand the different omics layers
- Scientific literature contains this information
- Prior Knowledge = already existing understanding of biological process



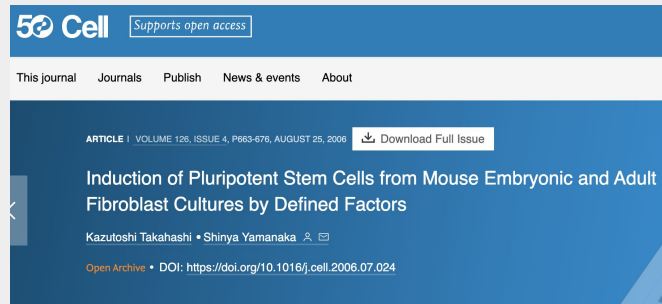


Prior knowledge has played a crucial part in biological discoveries

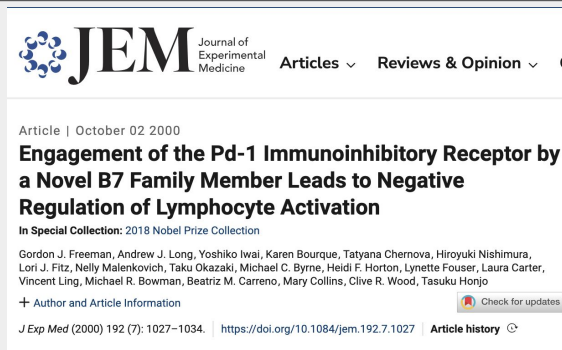
Prior understanding of bacterial immunity enabled CRISPR/Cas9 for genome editing



Prior understanding of cellular reprogramming and embryonic stem cell biology enabled identification of factors needed to make iPS cells

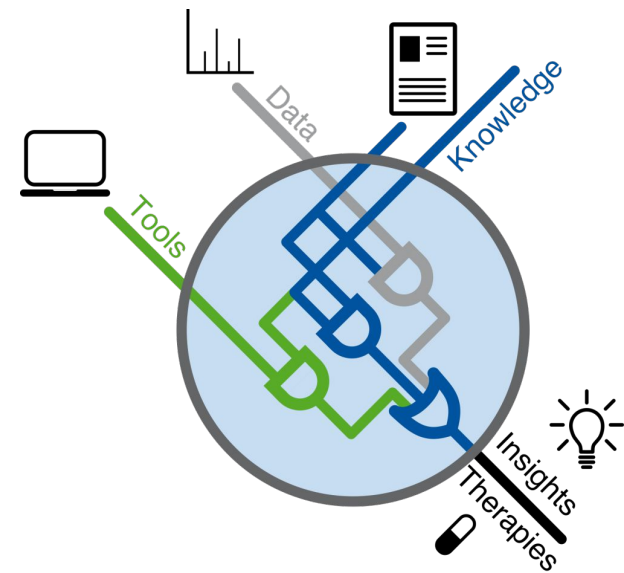


Prior understanding of surface proteins in immune cells enabled identification of PD1-blockade for cancer immunotherapy





In omics analyses, PK is beneficial for:



- Generating hypotheses
- Moving towards causal inference
- Avoid reinventing the wheel
- Filtering out data / Dimensionality reduction
 - Increasing statistical robustness
 - Prioritising signal over noise
- Data integration

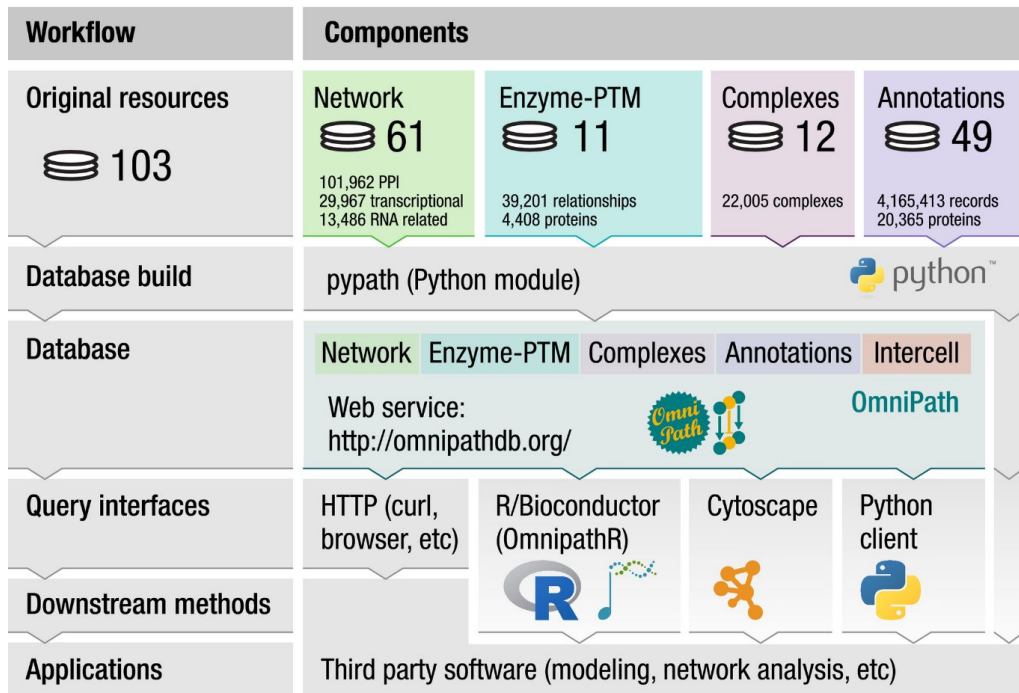
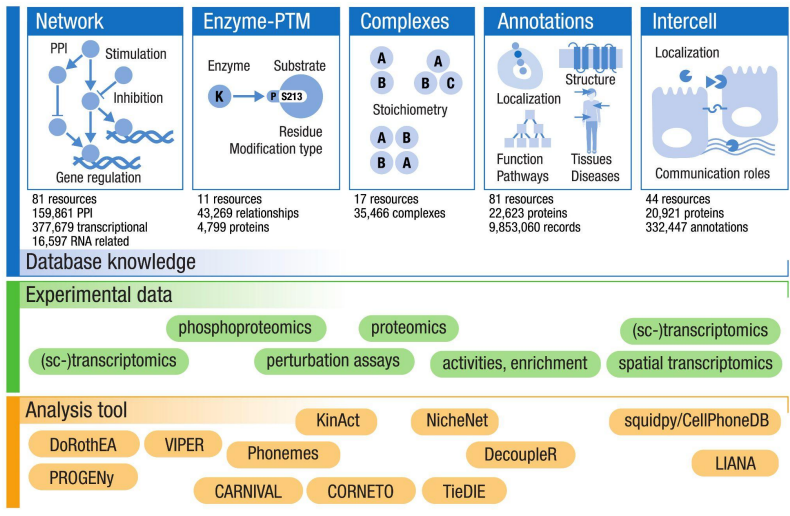


There are many different sources with omics prior knowledge

- Databases
 - Networks (e.g. intra- & inter- cellular PPIs, TF target interactions, miRNA-mRNA interactions) → e.g. DoRothEA, CellPhoneDB, ENCODE, Pathway Commons, KEGG, and many others
 - Protein complexes → e.g. hu.MAP, ICELLNET
 - Annotations of proteins (e.g. function, structure, localization) → e.g. MSigDB, UniProt
- Sometimes also single publications
- Labs usually construct their own PK resource based on a combination of others, with custom filtering based on their specific task



OmniPath is a comprehensive PK database for omics data

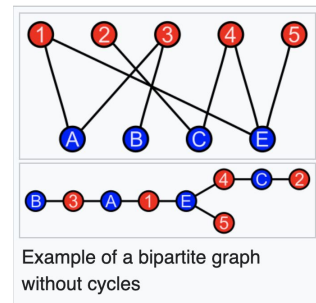




Prior knowledge is usually represented as:

- 1) Tabular format → Data Frames → invented in 1990s at Bell Labs
 - Basis for many stats & ML methods
- 2) Graphs → stores nodes and relationships
 - ideal for highly connected data
 - [BioCypher](#) as a recent tool to generate PK graphs
- Example: Bipartite networks can be represented in both tabular format, or graph
- Knowledge representation is an active area of research

Gene	Function
A	1,3
B	3
C	2,4
E	1,4,5





Prior knowledge can have limitations

PK can be:



- Inaccurate
- Outdated
- **Incomplete**
- Biased
- Unstructured
- Noisy
- Unoriginal

There is a need to continuously improve and update prior knowledge as new evidence becomes available.

Prior Knowledge (e.g. protein structure) can be predicted:

Article | [Open access](#) | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

[John Jumper](#) , [Richard Evans](#), [Alexander Pritzel](#), [Tim Green](#), [Michael Figurnov](#), [Olaf Ronneberger](#), [Kathryn Tunyasuvunakool](#), [Russ Bates](#), [Augustin Židek](#), [Anna Potapenko](#), [Alex Bridgland](#), [Clemens Meyer](#), [Simon A. A. Kohl](#), [Andrew J. Ballard](#), [Andrew Cowie](#), [Bernardino Romera-Paredes](#), [Stanislav Nikolov](#), [Rishub Jain](#), [Jonas Adler](#), [Trevor Back](#), [Stig Petersen](#), [David Reiman](#), [Ellen Clancy](#), [Michal Zielinski](#), ... [Demis Hassabis](#)  [+ Show authors](#)

[Nature](#) **596**, 583–589 (2021) | [Cite this article](#)

1.60m Accesses | **13k** Citations | **3591** Altmetric | [Metrics](#)



Can we use data-driven approaches instead of PK approaches?

[nature](#) > [articles](#) > [article](#)

Article | Published: 19 October 2017

Mastering the game of Go without human knowledge

[David Silver](#) , [Julian Schrittwieser](#), [Karen Simonyan](#), [Ioannis Antonoglou](#), [Aja Huang](#), [Arthur Guez](#),
[Thomas Hubert](#), [Lucas Baker](#), [Matthew Lai](#), [Adrian Bolton](#), [Yutian Chen](#), [Timothy Lillicrap](#), [Fan Hui](#),
[Laurent Sifre](#), [George van den Driessche](#), [Thore Graepel](#) & [Demis Hassabis](#)

[Nature](#) **550**, 354–359 (2017) | [Cite this article](#)

364k Accesses | **4963** Citations | **2511** Altmetric | [Metrics](#)

Article | Published: 31 May 2023

Transfer learning enables predictions in network biology

[Christina V. Theodoris](#) , [Ling Xiao](#), [Anant Chopra](#), [Mark D. Chaffin](#), [Zeina R. Al Sayed](#), [Matthew C. Hill](#),
[Helene Mantineo](#), [Elizabeth M. Brydon](#), [Zexian Zeng](#), [X. Shirley Liu](#) & [Patrick T. Ellinor](#) 

[Nature](#) **618**, 616–624 (2023) | [Cite this article](#)

109k Accesses | **52** Citations | **567** Altmetric | [Metrics](#)

- But data-driven approaches need data
- Biological data usually suffers from curse of dimensionality → data-driven methods might not be applicable
- In biology, PK can be very useful



Conclusion

Prior knowledge is integral for:

- Contextualizing findings for more confident & broader hypothesis generation
- Reducing data dimensionality for more interpretable and less noisy findings
- Integrating data from other types to make analysis comprehensive

But we need to be aware of limitations!



Now onto the practicals!

You will find the jupyter notebook here:

<https://shorturl.at/EZD3L>

Go to File > Save a copy in Drive → work on that copy



Sources

To prepare this presentation I used:

- perplexity.ai, check it out!
- Wikipedia for figures
- Ideker, T., Dutkowski, J. & Hood, L. Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell* 144, 860–863 (2011)
- Türei, D., et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol Syst Biol* (2021)