Supplementary Material

ProteomicsDB: A multi-omics and multi-organism resource for life science research

Patroklos Samaras[1], Tobias Schmidt[1], Martin Frejno[1], Siegfried Gessulat[1, 2], Maria Reinecke[1, 3, 4], Anna Jarzab[1], Jana Zecha[1], Julia Mergner[1], Piero Giansanti[1], Hans-Christian Ehrlich[2], Stephan Aiche[2], Johannes Rank[5, 6], Harald Kienegger[5, 6], Helmut Krcmar [5, 6], Bernhard Kuster[1, 7, *], Mathias Wilhelm[1, *]

[1] Chair of Proteomics and Bioanalytics, Technical University of Munich (TUM), Freising, 85354, Bavaria, Germany

[2] Innovation Center Network, SAP SE, Potsdam, 14469, Germany

[3] German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany

[4] German Cancer Research Center (DKFZ), Heidelberg, Germany

[5] Chair for Information Systems, Technical University of Munich (TUM), Garching, 85748, Germany

[6] SAP University Competence Center, Technical University of Munich (TUM), Garching, 85748, Germany

[7] Bavarian Biomolecular Mass Spectrometry Center (BayBioMS), Technical University of Munich (TUM), Freising, 85354, Bavaria, Germany

* To whom correspondence should be addressed. Tel: +49 8161 71 4202; Fax: +49 8161 71 5931; Email: mathias.wilhelm@tum.de, kuster@tum.de

Supplementary Figures

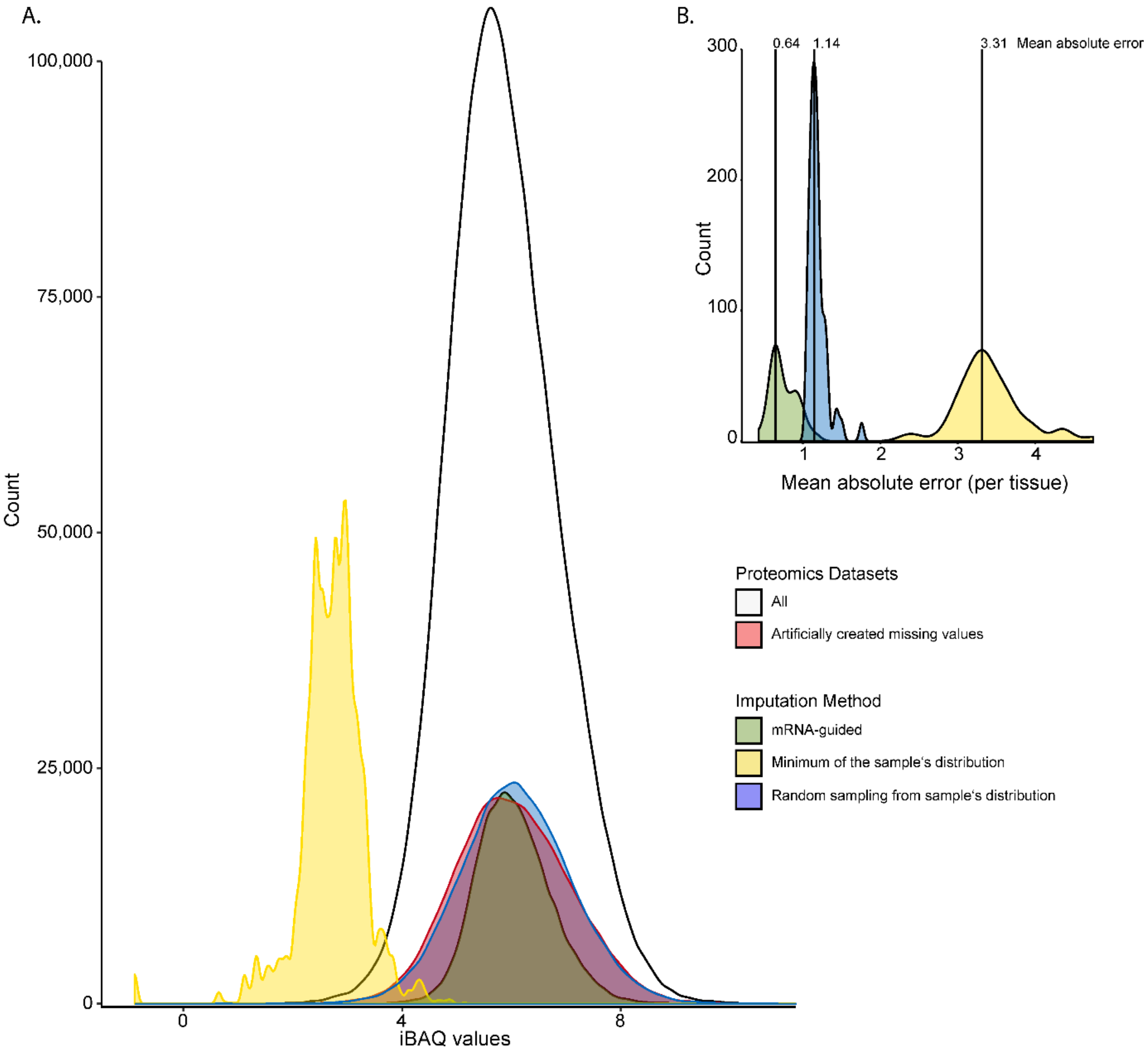
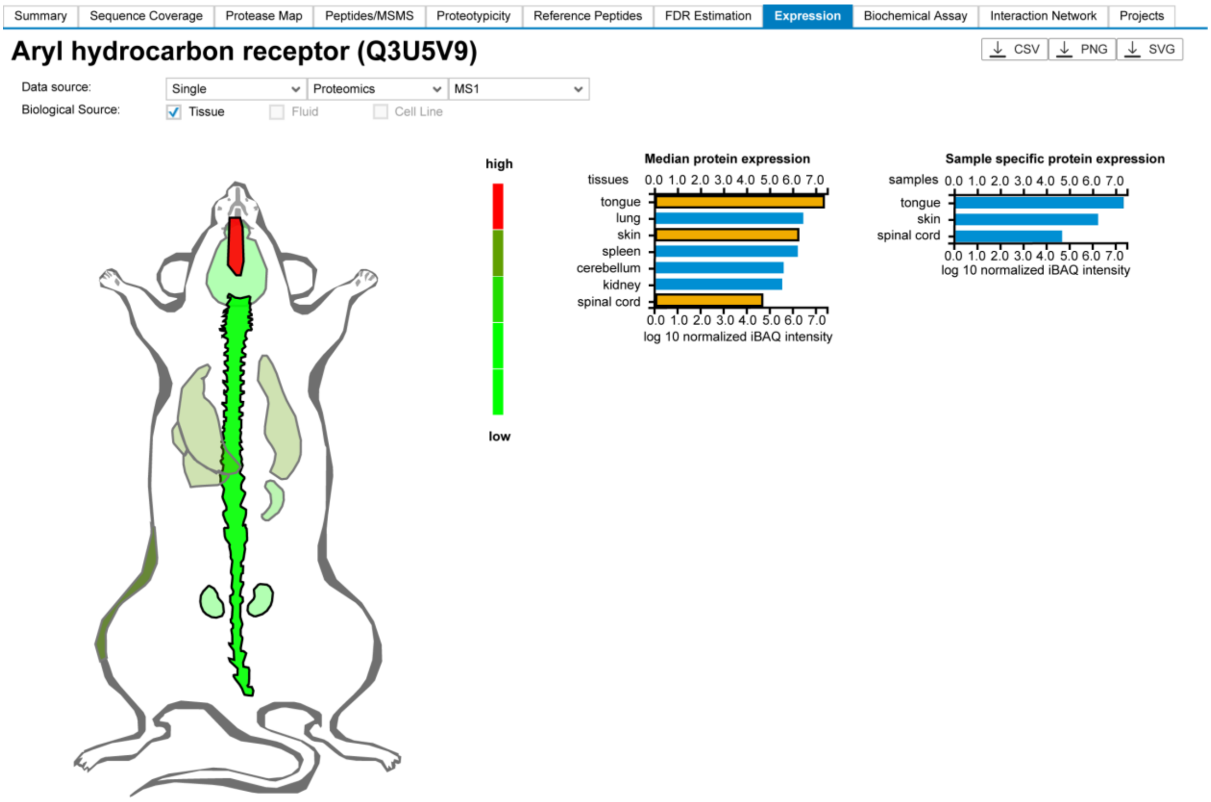


Figure S1. Histograms for comparing different imputation methods. A. The white distribution represents all iBAQ protein expression values in ProteomicsDB. The red distribution represents the missing values we created by random sampling 10% of the white distribution. The green, yellow and blue distributions represent the imputed missing values based on the 3 different methods respectively: mRNA-guided, minimum of a sample’s distribution and random sampling from each sample’s distribution. B. Histograms of the mean absolute error per tissue for each imputation method.



Aryl hydrocarbon receptor (Q3U5V9)

Data source: Single | Proteomics | MS1

Biological Source: ☑ Tissue ☐ Fluid ☐ Cell Line

Figure S2. Interactive expression body map for *Mus musculus.* As previously available for *Homo sapiens,* the same interactive body map idea is used for visualizing quantitative expression data for every organism hosted by ProteomicsDB.

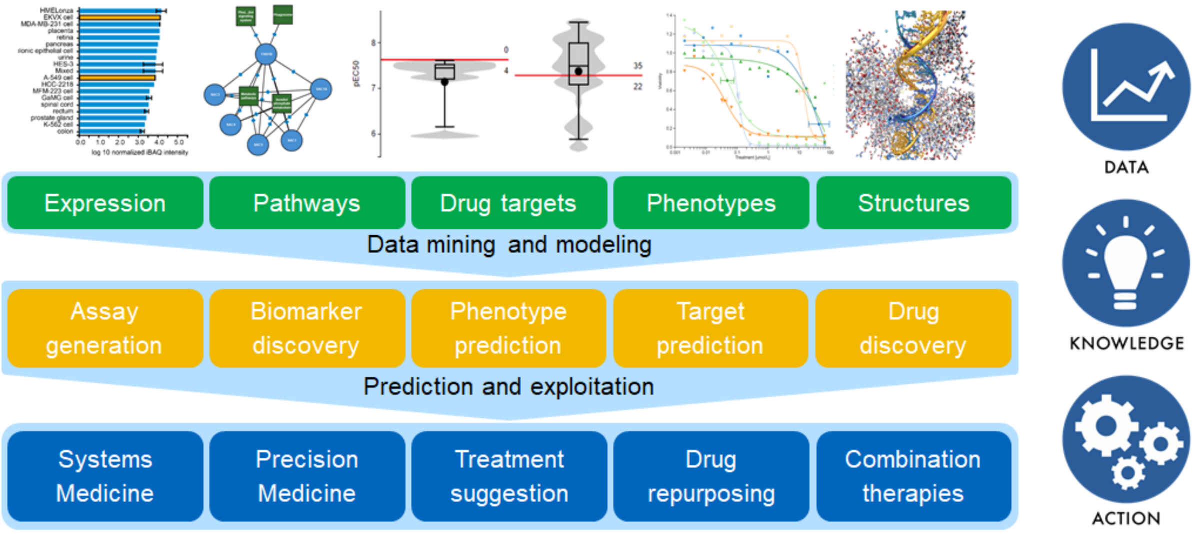


Figure S3. The future of ProteomicsDB: From data to knowledge to action. Future releases of ProteomicsDB will contain protein structure data. Applying data mining and knowledge discovery methods, new information will be generated which will be used to build new tools that will deliver the extracted knowledge to the user. The final goal of ProteomicsDB is to provide tools that will combine all data coming from the previous layers and take part in the decision making in modern research.