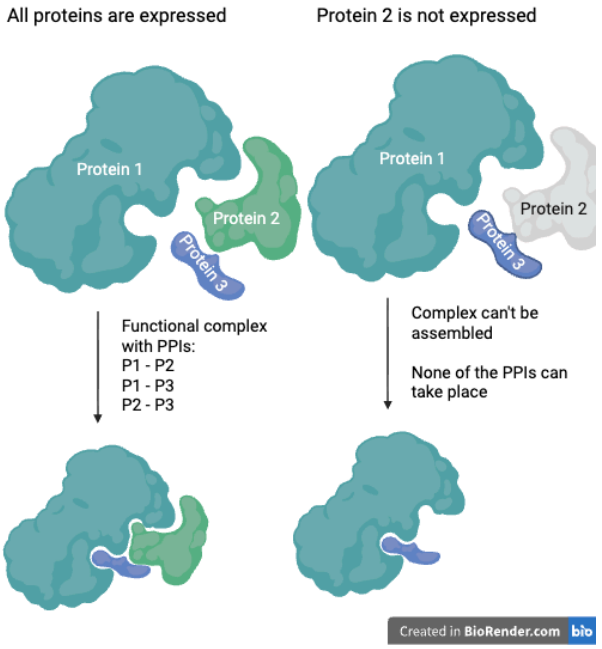
Attention: This project can have some collaboration with the project with *'Investigate inter- vs. intra-protein complex interaction biases in existing PPI databases'* towards the end. If you are applying with a friend (with whom you work well together) and who wants to take the other project, please indicate this in your application.

Project Proposal: DIGGER for intra-protein-complex interactions

Supervisor: Judith Bernett
Occasional co-supervision by Konstantin Pelz (TUM)

Background and Motivation



Existing protein-protein interaction (PPI) databases (e.g., STRING, BioGRID, HIPPIE) annotate pairwise interactions. However, most proteins function as part of multi-subunit complexes.

Which proteins are expressed is highly context-dependent. Tissue type, disease state, or subcellular localization, driven by differential gene expression or alternative splicing, critically determine which PPIs can occur. If a single protein is not expressed or is expressed as a different isoform, this disrupts the protein complex and, consequently, all associated interactions (depicted here). We would like to filter generic PPI networks based on experimental data (e.g., gene, transcript, or protein expression profiles) to retain only interactions stemming from expressed complexes.

This work should be implemented as a variation of DIGGER[1], a web tool that models context-specific PPIs at the domain level using alternative splicing data. By assessing whether splicing events remove interaction-mediating domains, DIGGER predicts which PPIs are likely disrupted. The integrated tool NEASE enables functional enrichment analysis to identify pathways significantly affected by these splicing events.

This project extends the idea of DIGGER to a higher level, focusing on PPIs disrupted within protein complexes. We want to subset the generic PPI network to a PPI network containing only interactions (A,B), for which we have an annotation that A and B are part of the same protein complex. We then query this subgraph with a gene, transcript, or protein expression vector to identify protein complexes that cannot be assembled. Like this, we want to make context-specific subgraphs.

Concrete steps:

1. Familiarize yourself with DIGGER: Read the publication[1] and the documentation, and understand the code structure. DIGGER was created with Django and deployed via Docker. Most of the code is written in Python. The data is stored in a PostgreSQL database.
2. Check out resources that annotate protein complexes, such as CORUM[2] or ComplexPortal, and read related work, such as COMPLEAT[3] or CPredictor 3.0[4].

1 Albrecht, Pelz, et al. DIGGER 2.0: digging into the functional impact of differential splicing on human and mouse disorders, Nucleic Acids Research, https://doi.org/10.1093/nar/gkaf384, Tool website: https://daisybio.ls.tum.de/digger/, Github: https://github.com/daisybio/DIGGER
2 Steinkamp et al. CORUM in 2024: protein complexes as drug targets, Nucleic Acids Research, Volume 53, Issue D1, 6 January 2025, Pages D651–D657, https://doi.org/10.1093/nar/gkae1033, Website: https://mips.helmholtz-muenchen.de/corum/
3 Balu et al. Complex portal 2025: predicted human complexes and enhanced visualisation tools for the comparison of orthologous and paralogous complexes. Nucleic Acids Research 53, D644–D650. https://doi.org/10.1093/nar/gkae1085, Website: https://www.ebi.ac.uk/complexportal/home
4 Xu et al. CPredictor3.0: detecting protein complexes from PPI networks with expression data and functional annotations. BMC Syst Biol 11 (Suppl 7), 135 (2017). https://doi.org/10.1186/s12918-017-0504-3

3. Obtain a subset of the generic PPI network that contains only interactions within complexes.
4. Adapt the DIGGER functionalities by replacing the domain-domain interaction information and isoform information with complex information:

Isoform-Level Analysis → Complex level analysis

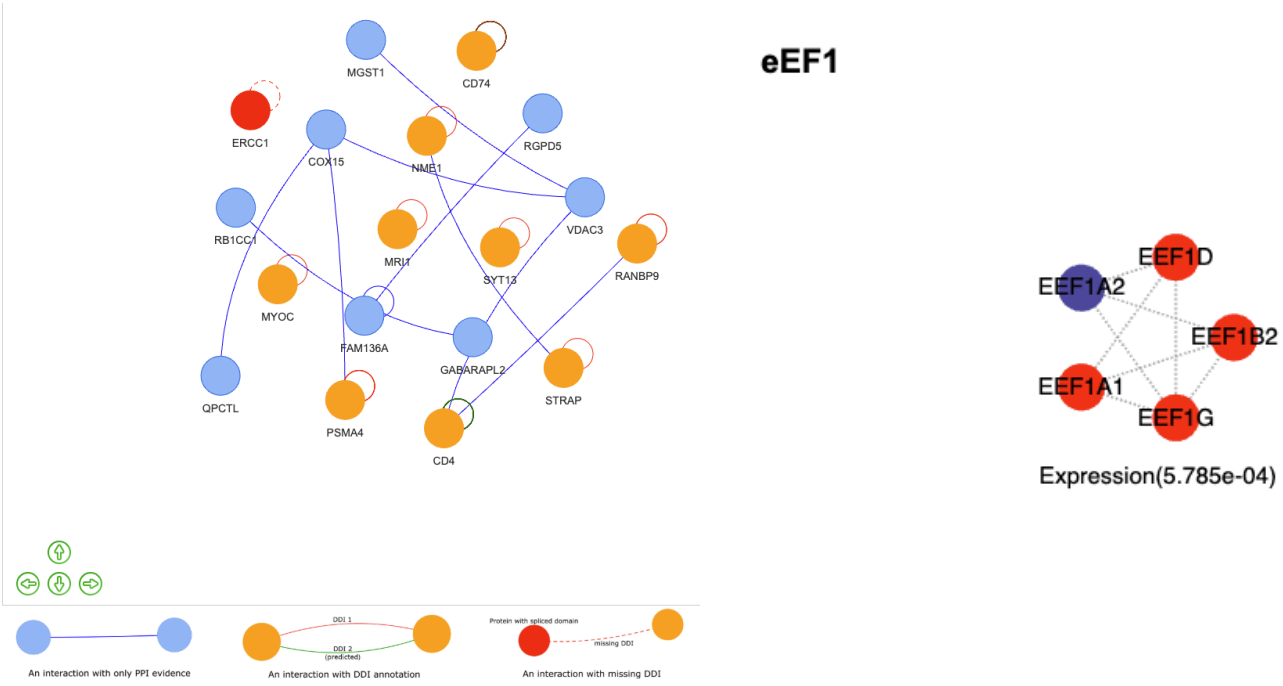
For example, TP53: we now don't want to show the transcripts and annotated domains, but the complexes and interaction partners.

List of the gene **TP53** transcripts with known Pfam domains.

Transcript name	Transcript ID	Pfam domains	Present / Missing interacting domains in the isoform	Link	Effect on Interactions
TP53-201	ENST00000269305	PF00870, PF07710, PF08563, PF18521		Visualize	Show Interactions
TP53-205	ENST00000445888	PF00870, PF07710, PF08563, PF18521		Visualize	Show Interactions
TP53-222	ENST00000615910	PF00870, PF07710, PF08563, PF18521		Visualize	Show Interactions
TP53-219	ENST00000610292	PF00870, PF07710, PF18521		Visualize	Show Interactions
TP53-203	ENST00000413465	PF00870, PF08563, PF18521		Visualize	Show Interactions

Network-level analysis of isoforms → Network-level analysis of complexes

Instead of inputting a list of transcripts and obtaining the PPI enriched with DDI information, we want to input a gene/transcript/protein expression vector and obtain the complex PPI (right screenshot from CORUM).



NEASE: functional enrichment analysis of alternative splicing events in the context of protein-protein interaction networks → adapt for complexes: which pathways are significantly disrupted by the given expression?