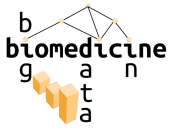
Documentation & Help

v 1.0

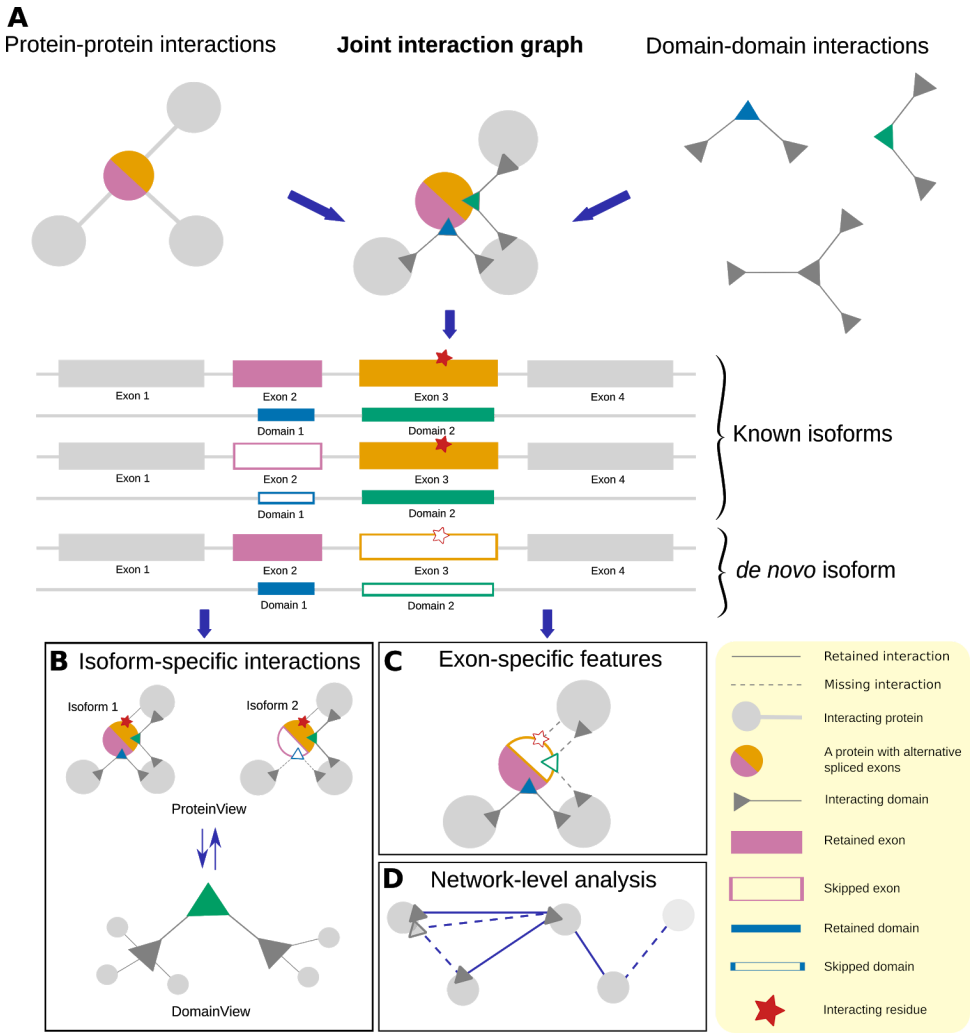


CompSysMed

Table of Contents

Overview of DIGGER workflow

PPI networks do not consider the influence of **alternative splicing**, even though experimental evidence suggests that the majority of protein isoforms have different interaction partners.



Domain Interaction Graph Guided ExploreR (DIGGER) integrates protein-protein interaction and domain-domain interactions into a joint graph and maps interacting residues to exons.

DIGGER provides a great way to visualize **PPI in structure context** using a **dynamic graph visualization** that can be toggled between a protein isoform and a domain-centric view.

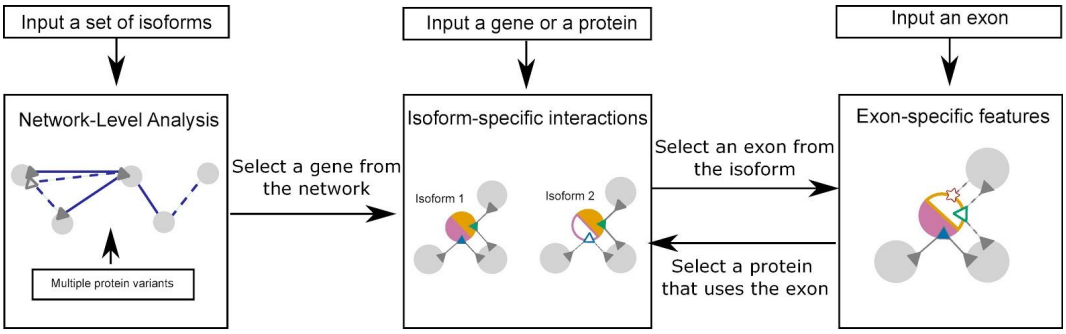
By comparing the structure of different isoforms, DIGGER identifies the surfaces specific to an isoform and missing in others. In this way, **isoform-specific interactions** can be precisely identified.

Furthermore, the tool maps the protein features encoded by a selected exon, to judge the functional role of individual exons in the PPI.

DIGGER is ideally suited to investigate the difference of interactions between the isoforms, analyze the effect of **isoform-switch,** To or explore how alternative splicing events such as **exon skipping** lead to altered interactions of protein isoforms.

DIGGER's joint PPI and domain-domain interaction network can also be used for subnetwork extraction, providing a basis for network analysis. From a list of proteins or transcripts, the tool **re-score the PPI edges** based on the structure evidence of the input proteins.

General workflow for alternative splicing analysis\



DIGGER allows the users to query exons or isoforms individually or as a set to visually explore their interactions. The three modes of DIGGER can be used interchangeably.

Isoform-Level Analysis

After selecting this mode, the user can query a gene (for example, “NCK2”, “ENSG00000058404”) or a protein variant (for example, “ENST00000066544”, “ENSP00000358396 ”).

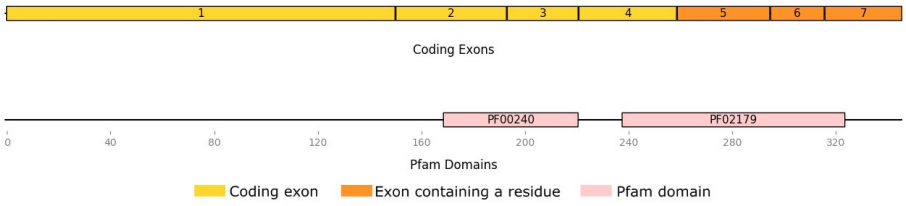
If the input is a gene, a table of annotated variants to explore will be shown. After choosing a protein variant (either from the table or as an input) the user can visualize its structure, which includes the exons and their corresponding domains.

Transcript name	Transcript ID	Pfam domains	Link
FSD2-201	ENST00000334574	PF00041 ; PF00622	(Visualize)
FSD2-202	ENST00000541889	PF00622	(Visualize)

From the table, the user can also switch to the exon-level analysis mode to further explore the functional role of individual exons (more details in the Exon-Level section).

Protein Features: Exons specific features

Domain and exon architecture: The transcript BAG1-210 has 7 exons and 2 unique protein domains.

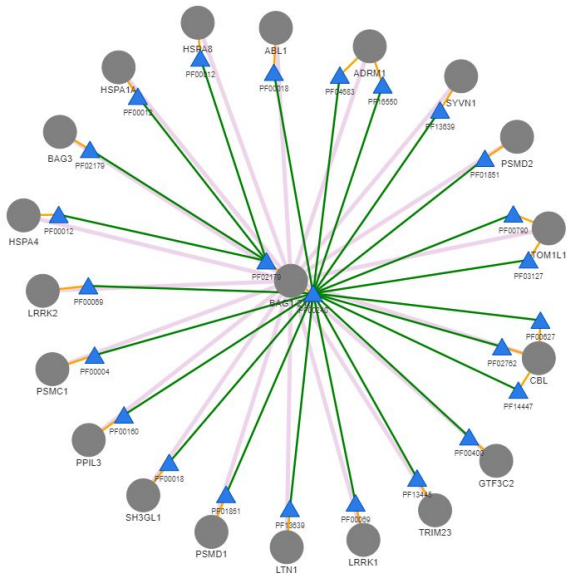


Select an exon from the table to assess the consequence of its skipping or choose different tabs to explore the isoform BAG1-210 interactions.

Exon ID	Exon rank in transcript	Number of interaction interface mapped to the exon	Corresponding domain ID	Protein features encoded by the exon
ENSE00003847903	1	0	-	Exon Page
ENSE00002346915	2	0	PF00240	Exon Page
ENSE00003574940	3	0	PF00240	Exon Page
ENSE00003684099	4	0	PF02179	Exon Page
ENSE00003569638	5	13	PF02179	Exon Page
ENSE00003506454	6	6	PF02179	Exon Page
ENSE00003785100	7	6	PF02179	Exon Page

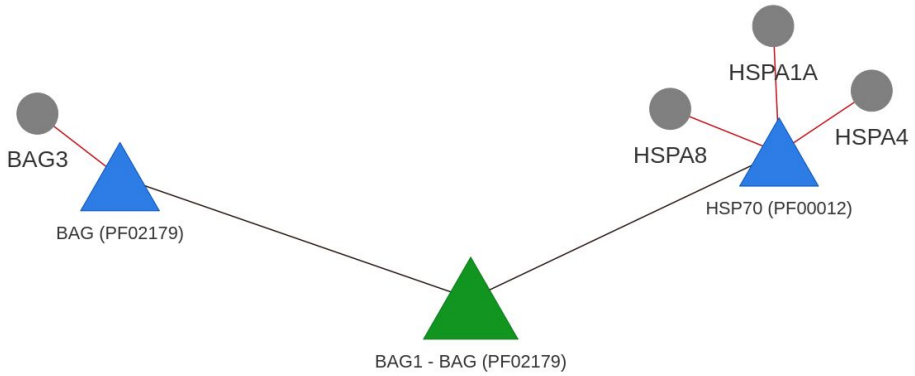
The exons in orange have residue interfaces mapped to them. In addition, every exon is shown with its corresponding Pfam domain with a link for more details and the residue and the domain encoded by the exon.

ProteinView: A structurally annotated PPI



The first feature of this mode is to allow the user to query genes or proteins and visualize their **structurally annotated interactions**: proteins interactions together with specific interacting domains (triangle nodes).

DomainView: Domain-specific interactions

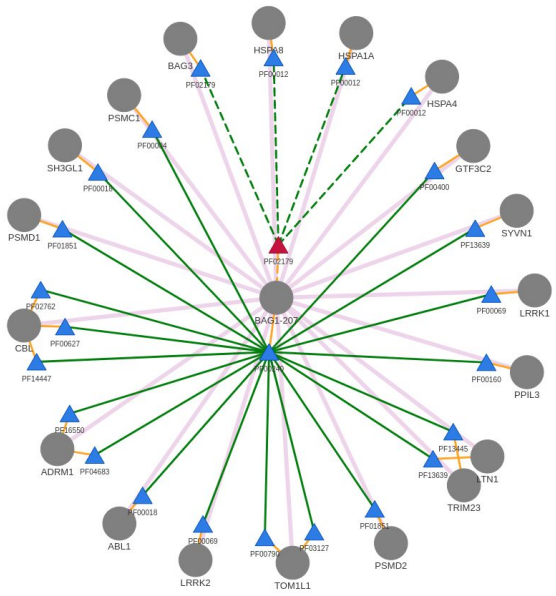


Protein domains are often shared between different isoforms. The DomainView focuses on the domains as independent units and highlights their roles on the PPI. This view is not only useful to study spliced domains but can also be extended for other applications such as studying coding disease variants affecting a protein domain or analyzing specific drugs targeting a domain unit. For instance, DomainView for the BAG domain from the last example is shown. This domain is also known to be spliced in a few isoforms.

Compare between Isoforms

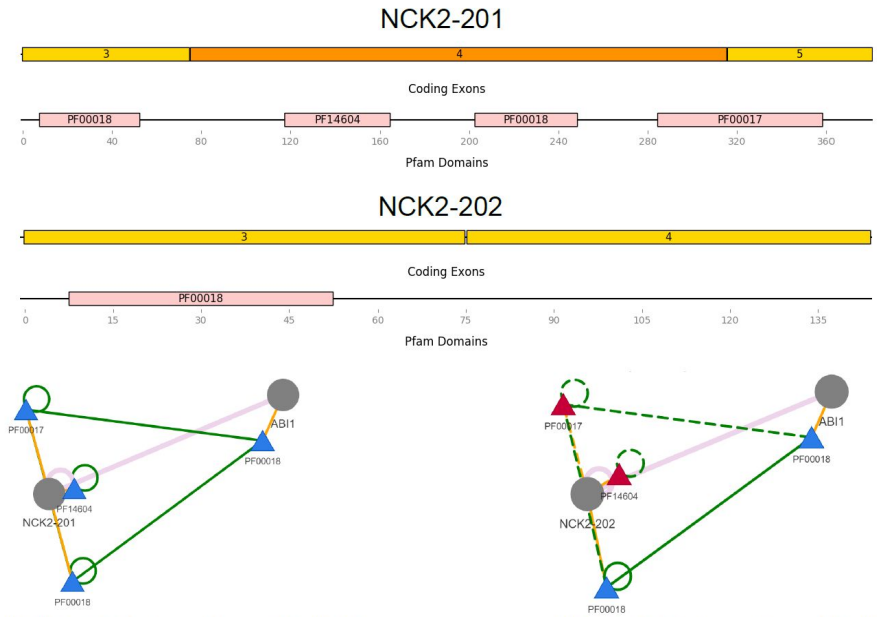
DIGGER provides **isoforms-specific interactions** and analyzes the effect of the **spliced domains** on the PPI. For example, an isoform of BAG1 is known to be missing the domain BAG (PF02179). Since the interaction with the proteins: BAG3, HSPA1A, HSPA8, and HSPA4 are mediated by this domain, these interactions are most probably specific to the main isoform and missing for the isoforms lacking the domain.

The ProteinView highlights spliced domains



Considering that the PPI databases usually focus on the main isoforms and neglect these splicing variants, this mode can be useful to explore **the effect of alternative splicing in the PPI**. The user can open multiple tabs and compare the exon/domain composition and the interactions of these protein variants.

InteractionView: Visualize a single interaction



NCK2-201 interaction with ABI1 NCK2-202 interaction with ABI1

DIGGER provides a **score** for every structurally annotated PPI for the specific isoform. Based on the percentage of **spliced interacting domains** in that isoform. For example, there are two domains known to mediate the interaction between two proteins NCK2 and ABI2. One of them is spliced out for isoform NCK2-202 (the triangle node in **red**), the interaction is scored 0.5 for this isoform and 1.0 for NCK2-201. The user can visualize each **interaction individually using InteractionView**.

Note

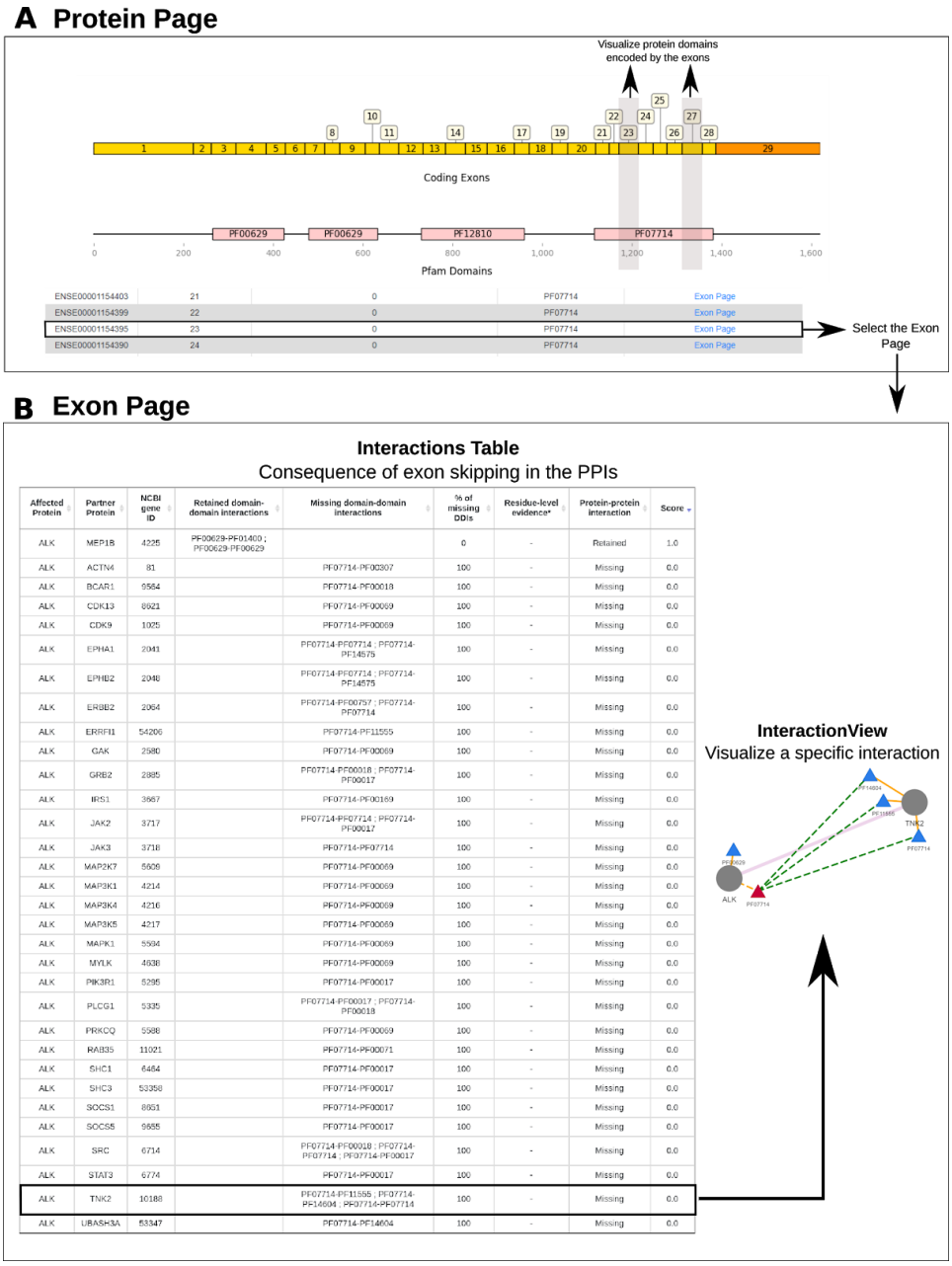
In this mode, DIGGER compares between different isoforms but does not consider splicing of the partner protein. In order to analyze the interactions of two specific isoforms or more check the **Network-Level Analysis.**

This mode is based on the annotated transcripts from the Ensembl database. In order to analyze non-annotated proteins (or splicing events) check the **Exon-level analysis**.

Exon-Level Analysis

In the exon-level analysis mode, we focus on the **functional role of individual exons** in the PPI. This is particularly handy to explore the consequence of **a splicing event that results in a non-annotated protein**. From the protein page, the user can choose an exon or input the Ensembl id/coordinate directly from the exon-level page (for example, "ENSE00002224211 157135471 157135708").

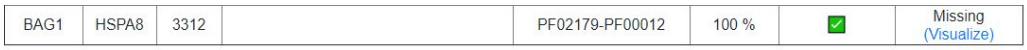
The consequence of exon skipping on PPI



The example below shows how to navigate from the **Isoform-level** analysis to the **Exon Level** analysis. In this mode, all domains and residues (if any) encoded by a specific exon are shown. In addition, ProteinView and DomainView simulate the consequence of the skipping of the exon in the PPI. In the figure example, the exons 21-28 from the ALK gene encode for the domain PF00714. This domain mediates the interaction with 32 partners. Consequently, DIGGER identifies these PPIs as affected in case of skipping any of these exons. The percentage of missing mediated domains is shown in the table in addition to a visualization of all interactions together (ProteinView) or individually (InteractionView). In the graphical visualization, the spliced domains are shown in red and their edges are dashed (missing edge).

If an exon does not encode any known domains or contain any interacting residues, Exon Page will contain the information about the corresponding protein interactions and a list of protein variants with this exon.

Residue-level evidence further specify the individual exons role.



In addition to the domain-domain interactions, DIGGER integrates residue-level information for structurally resolved interaction interfaces. These interacting residues are mapped to their corresponding exons.

In Isoform-level mode, the exons with residue mapped to them are shown in **orange**. Furthermore, in Exon-level mode, the specific interfaces are shown in a separate table. If an interaction has both DDI and residue level evidence, it is highlighted in the table with a **checkmark .** This shows that the selected exon (in this example an exon of gene BAG1) encodes for a domain that interacts with the protein (HSPA8) and for a residue interface for the same partner protein. This interaction is then considered with a **high level of confidence** to be specific to the selected exon.

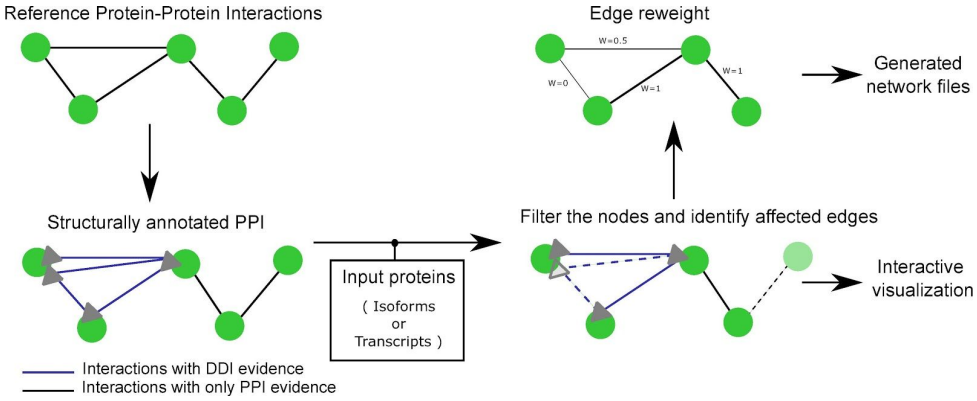
In addition, this feature can **narrow** down the function of the coding region. For instance, when multiple exons are known to encode one domain or when multiple domains mediate one interaction.

Note

The structures used for this annotation are typically derived from the full-length transcript and thus do not reflect the influence of the exon itself on protein folding.

Network-Level Analysis

This mode extends the isoform level analysis to the network level to visualize interactions between two isoforms or more (**up to 2000**). To run this mode, the user can input a list **of genes, transcripts, or isoforms using Ensembl identifiers**. DIGGER also supports the count matrix as it is typically output by tools like **Cufflinks and Kallisto**. The first column of your file should correspond to transcript ids and optionally a column named **'tpm'** or **'counts'** or **'FPKM'**.
In the case of gene IDs, DIGGER will generate a subnetwork from PPI of the interactions between the input list and highlight the edges with a structural annotation.



DIGGER constructs a specific PPI from a list of transcripts or isoforms

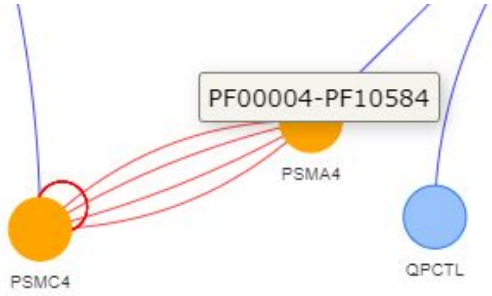
More interestingly, the user can input a list of transcripts or isoforms. DIGGER then identifies all missing domains in the subnetwork and detects affected interactions and reweighs them (Be aware, the run time could take some time).

Note

In contrast to Isoform-level mode, the spliced domains can be on both sides of the PPI edge.

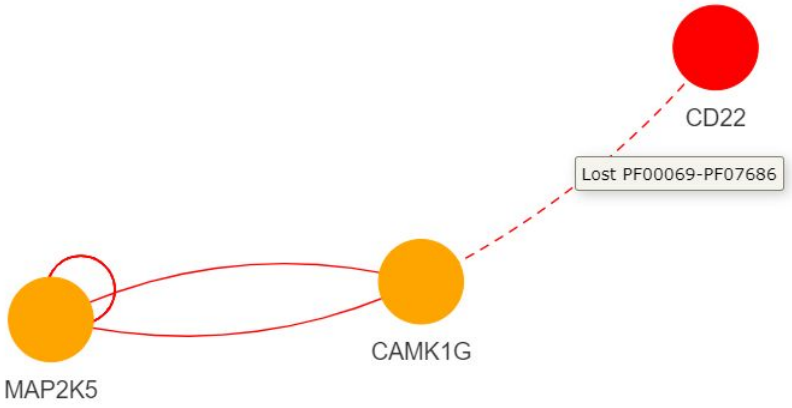
The user **should not** mix between the gene and transcript ids in a single query!

Visualize structurally annotated subnetwork



The dots, representing transcripts or isoforms with known DDI, marked orange, and the interactions are shown in red. Otherwise shown in blue if there is only PPI evidence. To highlight the interacting domains, the user could point to the interaction of interest.

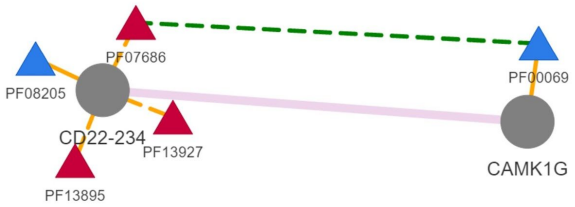
DIGGER highlight lost interfaces in PPI network level



The edges are shown in red dashed if the domain mediating the interaction is **missing in the user input of transcripts.** In this example, The protein CD22 is shown also in a red node since it is the one missing the interacting domain. This signifies that even if the proteins CD22 and CAMK1G are known to interact in the PPI database, this interaction is specific to another isoform of CD22 that was not included in the input. The user can click on any node to get details about the gene and its annotated variants. In this case, the isoform CD22-234 (ENST00000613136) missing the domain PF07686 is shown with a link to its page.

DIG deeper using DIGGER for a full exploration

CD22-225	ENST00000600424	PF07686	(Visualize)
CD22-234	ENST00000613136	PF08205	(Visualize)



DIGGER flexibility allows the user to **navigate through the database** for a **deeper understanding and exploring of AS impact in PPI**. From the Network-level analysis, the user can go to the specific gene and visualize the interactions of all the variants. For instance, CD22-2346 interaction with CAMK1G can be further visualized using InteractionView, which is the same edge that was dashed in our subnetwork. Furthermore, the user can **dig** deeper to analyze the **specific exons contribution** by choosing any of them from the protein page.

A re-weighted condition-specific network

Show 10 entries Search:

Protein 1 name	Protein 2 name	Retained domain-domain interactions	Missing domain-domain interactions	Score*	Source of the interaction
ERCC1	ERCC1	-	PF14520 -PF14520 ; PF00633 -PF00633	0.0	PPI-DDI
CD22	CAMK1G	-	PF07686 -PF00069	0.0	PPI-DDI
MBD3	MBD3	-	PF01429 -PF01429	0.0	PPI-DDI
BCL3	NFKB2	PF12796 -PF16179 ; PF12796 -PF12796	PF00023 -PF16179 ; PF00023 -PF12796 ; PF00023 -PF00554 ; PF00023 -PF00531	0.33	PPI-DDI

The new weight represents the confidence of the PPI. A score of 0 means that all known structurally resolved interfaces between the two proteins are missing in your list of isoforms.

Note

All the data and networks generated from DIGGER can be downloaded for further analysis.