

Instructions for Filling the Tables

Below are the guidelines for filling in the table. Please follow these instructions carefully to ensure consistency and accuracy.

Column Labels and Descriptions SC

- **Gene/Protein**

- Note the name of the gene/Protein for which connections to a disease or a treatment/chemical/medicine are described. Label it additionally with a (G) if it is stated to be the gene and (P) if it is a Protein. Otherwise, add none. Example: Protein(P)
- add a (mut) label if the Gene or Protein is described to have a mutation. Example: Gene(G)(mut)
- Add (fam) if for gene families instead of single genes. Example: Gene(G)(fam)
- Add (path) if the pathway of the gene interaction is mentioned. Example: Gene(G)(path)

- **Disease**

- If a disease is connected to a gene, specify the disease here. If the disease is a cancer use the following format disease(cancer name)(cancer type). Enter (NA) if a category is not given. Use the information in the text and your understanding of cancer to fill in this information.

For cancer types use the labels: Carcinoma, Sarcoma, Lymphoma, Leukemia, Myeloma.

Example: Cancer(lung cancer)(carcinoma)

Treatment/Exposure

- If a treatment (drug/chemical/radiation/knock-out, etc.) is mentioned that was used and a connection to the gene is mentioned, write it here. Example: DAS, DADS
- **Connection Disease**
- Use "connected" if no specific interaction is described; there is just some relationship between the gene and the disease.
- Use "key gene" if the gene is directly described as a key gene or it is very clear that this particular gene plays a key role in the interaction.
- Use "marker" if something is used as a marker gene.

- Use “driver” if the gene/protein makes the disease worse
- Use “suppressor” if the gene/protein counteracts (aspects of) the disease
- **Connection Treatment**
 - Use "connected" if no specific interaction is described; there is just some relationship between the gene/protein and the treatment.
 - Use "target" if a treatment specifically targets this gene/protein.
 - Use “activator/suppressor” if the treatment activates/suppresses the gene/protein. Add (ind) if the connection is indirect. Example: activator(ind)
- **Disease Mechanism**
 - Describe the mechanism to which the connection between the gene and the disease is tied. Use keywords, not full sentences. Examples include risk factor, better outcomes, proliferation, cell cycle progression, apoptosis, pro inflammation, anti inflammation
- **Treatment Mechanism**
 - Describe the mechanism to which the connection between the gene and the treatment is tied. Use keywords, not full sentences. Examples include risk factor, susceptibility, resistance, growth inhibition, autophagy
- **Associated Factors**
 - Enter information about other contexts or more specific info that don't fit into the other categories. Examples include age, mechanisms of action, outcomes, etc.
- **Cellline**
 - Specify the cell line mentioned in which the connection between the gene and other entities is described. If the exact cell line is not mentioned, add more general info like patient, human cancer cells, mouse cancer cells etc.
- **Species**
 - Specify the species being studied (e.g., human, mouse, rat, etc.).
- **Tissue**
 - Specify the tissue in which the connection was observed. If only the organ is stated, note the organ. If not specified, write NA.
 - If the paper talks about metastases, write metastases. If the organ of the metastases is known, note only the organ.
- **Confidence**
 - Use the labels: low, medium, high.
 - **Low:** No source/no experimental data, low certainty language.
 - **Medium:** Source/experimental data but uncertain language or certain language but no source.
 - **High:** Source/experimental data and high certainty language.
- **Type of Evidence**

Use the following labels (direct, indirect, citation). Explanations:

 - **Direct**

- **Examples:**
 - In Vivo: Observed directly in living organisms.
 - In Vitro: Demonstrated in isolated biological systems.
- **Indirect**
 - **Examples:**
 - Correlational: Inferred from statistical correlations or associations.
 - Comparative: Based on comparisons across species, conditions, or datasets.
 - Predictive: Based on computational models or theoretical predictions.
 - Observational: Suggests but does not prove a relationship.
 - Anecdotal: Based on personal accounts or individual cases.
 - Speculative: Based on conjecture or extrapolation from limited data.
 - Analogical: Based on analogies or similarities with other systems.
 - **Citation:** Given a citation in a text that does not specify the type of evidence. Formulations like: studies have shown or just a statement and a citation number behind it.
- **Sentence**
 - Each sentence has a number given like this: Sentence[1]:. Provide the sentence numbers that support the statements made in the rest of the row. This often includes multiple sentences.

Column Labels and Descriptions GG

- **Gene/Protein**
 - Name the Gene/Protein for which connections to (another) Gene/Protein are described. Label it additionally with a (G) if it is stated to be the gene and (P) if it is a Protein.
 - add a (mut) label if the Gene or Protein is described to have a mutation. Example: Gene(G)(mut)
- **Connected Gene/protein**
 - The gene/Protein for which connections to the aforementioned Gene/Protein are described. Label it additionally with a (G) if it is stated to be the gene and (P) if it is a Protein. Example: Protein(P)
 - add a (mut) label if the Gene or Protein is described to have a mutation. Example: Gene(G)(mut)
- **Connection Type**
 - activator / inhibitor
 - Use “activator” if some positive regulation between the Genes/Proteins is described.

- Use “inhibitor” if some negative regulation between the Genes/Proteins is described.
- Add (co) if the gene/inhibitor is co-regulating the Gene with another Gene/Protein.
- Add (ind) if the gene is involved in the regulation process but unclear if itself or one of its downstream targets is involved in the regulation.
- Only if activation/inhibition is not mentioned, use: connected/ binding / target
 - Use "connected" if no specific interaction is described; there is just some relationship between the genes/Proteins. Add (ind) if the gene is involved in the regulation process but unclear if itself or one of its downstream targets is involved in the regulation.
 - Use “binding” if the Gene/Protein binds to the connected Gene/Protein.
 - Use “target” if the Gene/Protein targets the connected Gene/Protein.
- Example: activator(ind)
- **Regulation level**
 - At what level does the regulation take place? Choose the most fitting label based on the information provided in the text. The (L) label is to be added if localisation is mentioned in the context of regulation.
label definitions:
 - epigenetic: regulation happens through methylation and other changes to the DNA and histones that change DNA folding.
 - transcription: regulation happens in the process of producing mRNA
 - splicing: different splicing patterns change the mRNA
 - mrna(L): regulation happens through longevity and state of mRNA
 - translation(L): regulation happens in the process of producing a protein from mRNA
 - Protein(L): regulation happens through stability and state of the protein
 - other: if another method is mentioned
 - **Disease**
 - If a disease is connected to the mentioned genes, specify the disease here. If the disease is a cancer, use the following format: disease(cancer name)(cancer type) enter NA if a category is not given. Use the information in the text and your understanding of cancer to fill in this information.
 - For cancer types use the labels: Carcinoma, Sarcoma, Lymphoma, Leukemia, Myeloma.
 - Example: Cancer(lung cancer)(carcinoma) or parkinson
- **Treatment/Exposure**
 - If a treatment (drug/chemical/radiation, etc.) is mentioned that was used and a connection to the gene is mentioned, write it here.
- **Associated Factors**

- Enter information about other contexts that don't fit into the other categories. Examples include age, inflammation, etc.
- **Cellline**
 - Specify the cell line mentioned in which the connection between the gene and other entities is described. If the exact cell line is not mentioned, add more general info like patient, human cancer cells, mouse cancer cells etc.
- **Species**
 - Specify the species being studied (e.g., human, mouse, etc.).
- **Tissue**
 - If the connection was observed within a specific tissue rather than a cell culture, specify the tissue type. If only the organ is stated, note the organ. If not specified, write NA.
 - If the paper talks about metastases, write metastases. If the organ of the metastases is known, note only the organ.
- **Confidence**
 - Use the labels: low, medium, high.
 - **Low:** No source/no experimental data, low certainty language.
 - **Medium:** Source/experimental data but uncertain language or certain language but no source.
 - **High:** Source/experimental data and high certainty language.
- **Type of Evidence**
 - Use the following labels (direct, indirect, citation). Explanations:
 - **Direct**
 - In Vivo: Observed directly in living organisms.
 - In Vitro: Demonstrated in isolated biological systems.
 - **Indirect**
 - Correlational: Inferred from statistical correlations or associations.
 - Comparative: Based on comparisons across species, conditions, or datasets.
 - Predictive: Based on computational models or theoretical predictions.
 - Observational: Suggests but does not prove a relationship.
 - Anecdotal: Based on personal accounts or individual cases.
 - Speculative: Based on conjecture or extrapolation from limited data.
 - Analogical: Based on analogies or similarities with other systems.
 - **Citation:** Given a citation in a text that does not specify the type of evidence. Formulations like: studies have shown or just a statement and a citation number behind it.
- **Sentence**

- Each sentence has a number given like this: Sentence[1]:. Provide the sentence numbers that support the statements made in the rest of the row. This often includes multiple sentences.