# CancerHubs Data Explorer User Manual


This guide provides a walkthrough for using the **CancerHubs Data Explorer** online. It explains how to navigate each tab of the web interface and retrieve data and images generated by the app.

## 1. Overview

The **CancerHubs Data Explorer** is an interactive Shiny web application for visualising the output of the *CancerHubs* framework—a network-based approach to identify and rank cancer-relevant genes by integrating mutational burden, gene expression-prognosis correlations (PRECOG), and protein interactomics.

Cancer is increasingly understood as a systems-level disease, not merely the result of individual gene alterations. Rather, it emerges from the coordinated disruption of functional gene modules within the interactome, often driven by network-central "hub" genes that orchestrate oncogenic processes. This insight motivates the shift from single-gene analyses toward network-informed prioritisation strategies.

To exploit this complexity, CancerHubs introduces the **Network Score**, a novel metric that captures the extent to which a gene's encoded protein is embedded in a network of mutated interactors. This score is computed by quantifying the proportion and absolute number of mutated partners in a tumour-specific interactome, allowing identification of potential cancer-driving hubs that may be overlooked by traditional mutation frequency or expression-based analyses.

The Data Explorer builds upon this foundation, offering an accessible platform to explore these hubs across 11 tumour types. Designed for both computational and experimental researchers, it supports layered, intuitive exploration of gene-level and network-level features without requiring programming expertise.

This method assigns a **Network Score** to each gene, defined as:

$$\text{Network Score} = \frac{(\#\ \text{Mutated Interactors})^2}{\#\ \text{Total Interactors}}$$

This score captures the extent to which a gene's encoded protein is embedded within networks of cancer-mutated interactors. High scores highlight potential **mutated protein hubs** involved in cancer pathogenesis.

*Note that the Network Score does not take copy-number variations into account.*

### Key Features:

- Ranking genes across 11 tumour types using network-based prioritisation
- Visualising gene–gene interaction networks in 2D and 3D
- Filtering by mutation status, prognostic association, or both
- Exporting publication-ready plots and tables

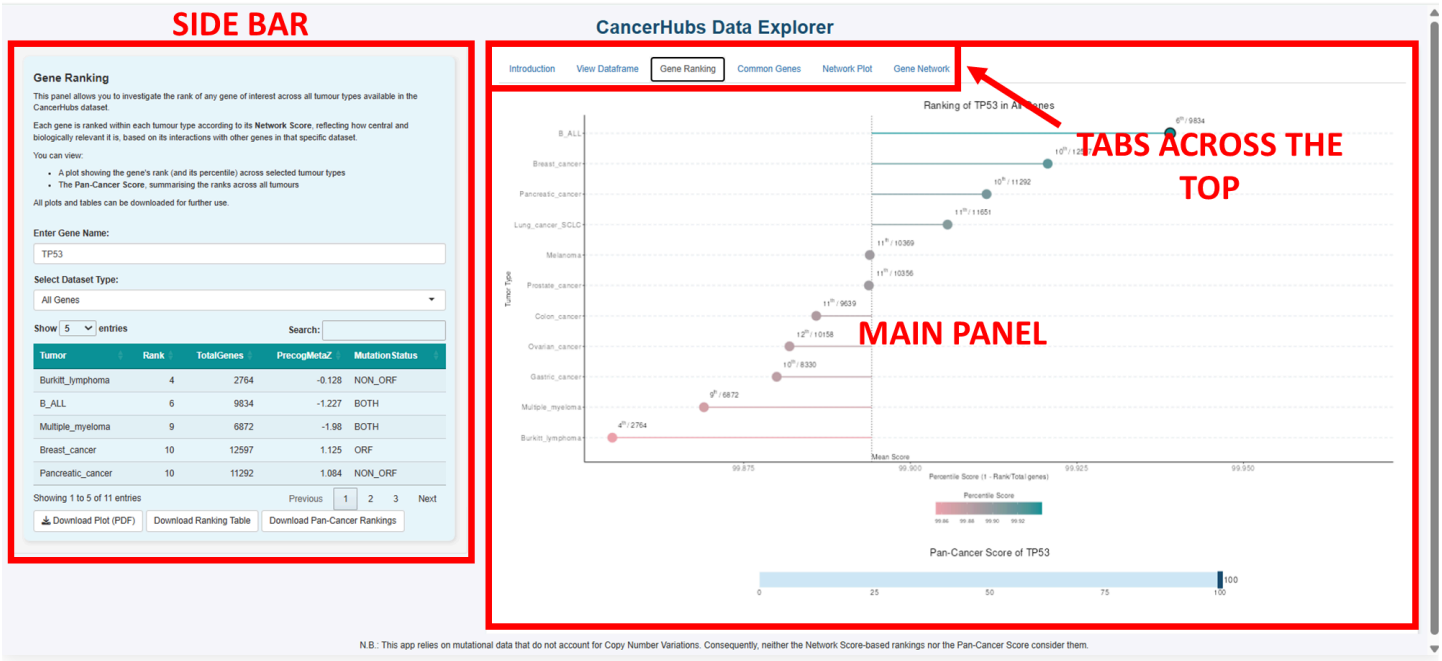
## 2. Accessing the App

The CancerHubs Data Explorer is available online—no installation is necessary.All necessary data and packages are loaded automatically upon opening the website. Simply visit: https://cancerhubs.app/

## 3. Application Layout

The interface is divided into a sidebar (left) and a main panel (right).

- **Sidebar** – Provides navigation and filtering options for each feature.
- **Main Panel** – Displays tables, plots, and network visualisations.
- **Tabs across the top** – Allow you to switch between the main functions: View Dataframe, Gene Ranking, Common Genes, Network Plot (3D), and Gene Network (2D).



# 4. Gene Subsets

Users can focus on different evidence-based gene categories:

- **All Genes** – All scored genes that are either mutated, show prognostic relevance, or both. Genes with neither mutations nor significant PRECOG scores are excluded during preprocessing.
- **Only Mutated** – Genes that harbour mutations in the selected tumour type but do not exhibit significant prognostic correlation (i.e., PRECOG meta-Z < |1.96|).
- **PRECOG** – Genes with prognostic relevance, defined by a meta-Z score ≥ |1.96| from the PRECOG dataset, regardless of their mutation status.
- **Only PRECOG** – Genes with significant meta-Z scores (≥ |1.96|) but that are not mutated in the selected tumour dataset.

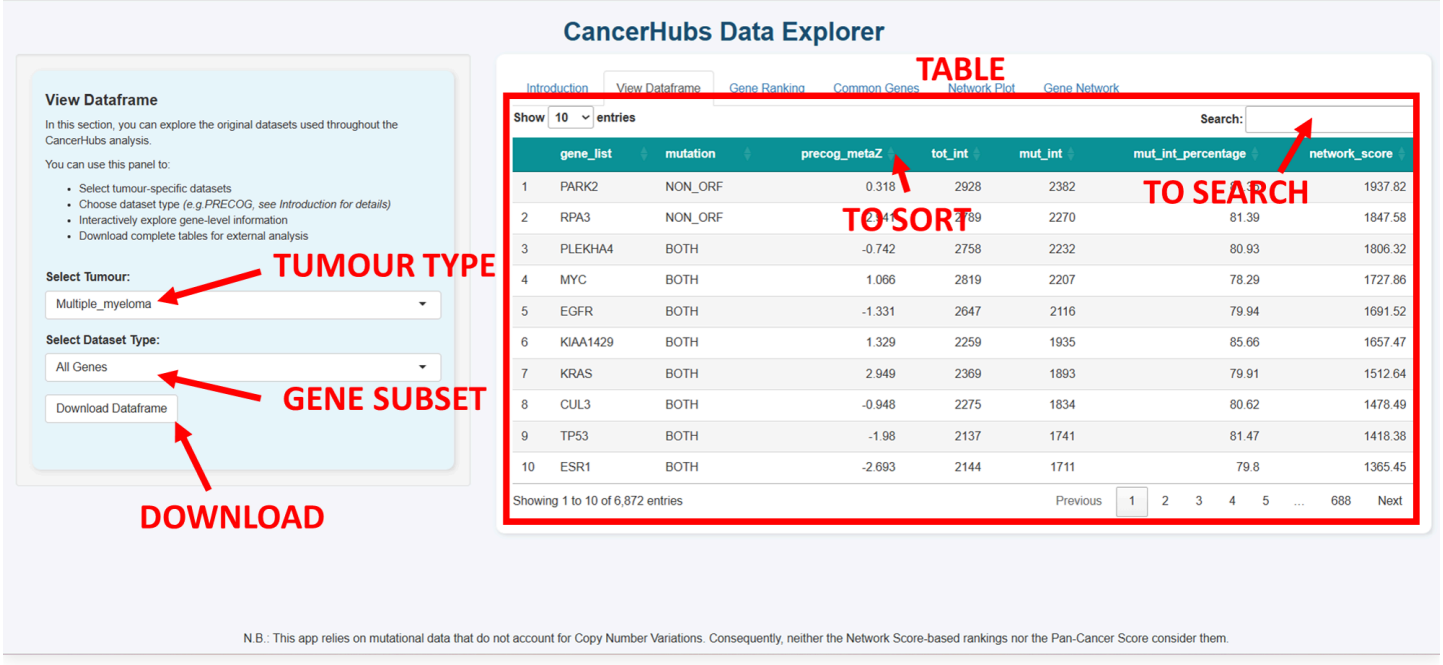
**How mutation status is defined:** Any type of nucleotide substitution (including synonymous and non-coding changes) is considered. Genes are labelled based on where mutations fall: *ORF* (in coding regions), *NON_ORF* (non-coding regions), or *BOTH*. Thresholds for retaining mutated genes are adjusted based on mutation frequency and correlation with prognosis, as detailed in the CancerHubs pipeline.

**How prognostic status is defined:** Genes with meta-Z scores ≥ 1.96 or ≤ −1.96 are considered statistically significant (95% confidence). Scores ≥ 2.58 or ≤ −2.58 are considered highly significant (99% confidence), and are retained even if the gene is not mutated.

# 5. Features

## 5.1 View Dataframe

This panel allows users to explore tumour-specific gene tables used in the CancerHubs analysis. The screenshot above illustrates the key components of the interface:

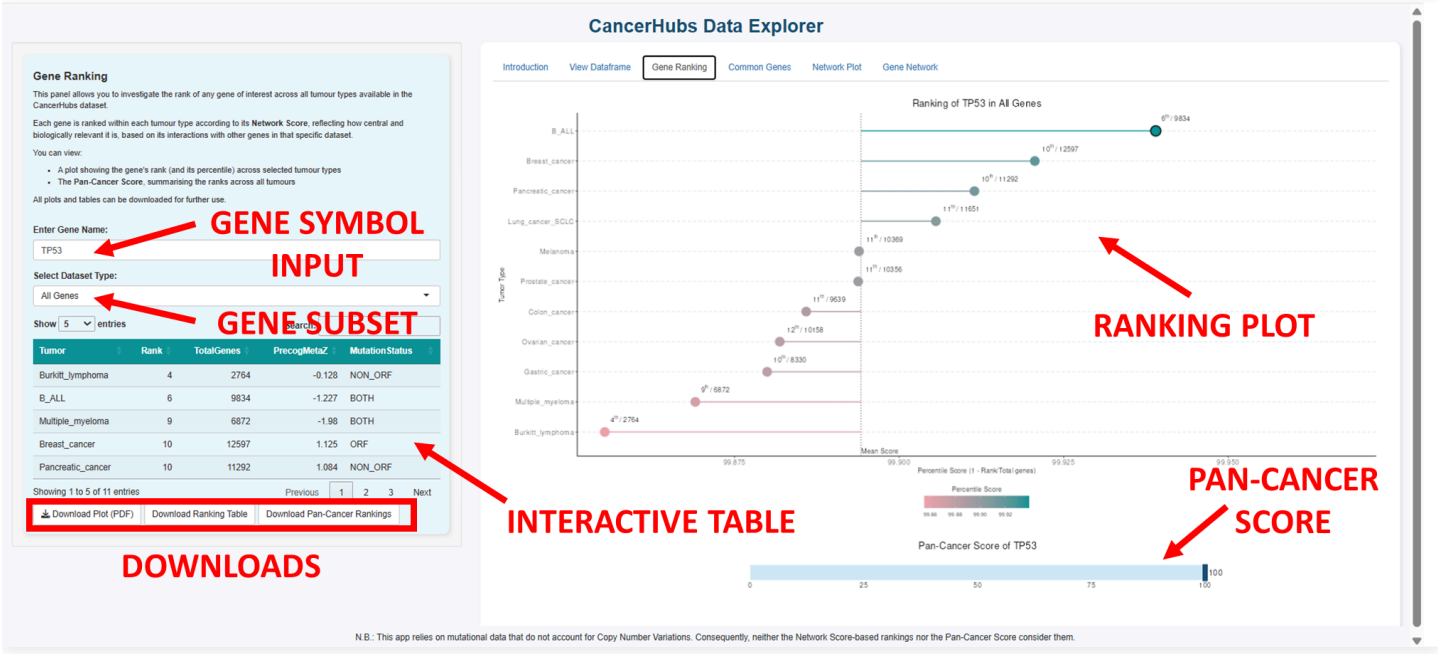


- **Tumour Type**: dropdown to select among 11 available tumour datasets.
- **Gene Subset**: filter based on mutation status and prognostic annotation (e.g. All Genes, PRECOG, Only Mutated).
- **Table**: dynamically rendered data including columns for mutation type, PRECOG meta-Z score, number and percentage of mutated interactors, and final Network Score.
- **Search**: type a gene symbol to instantly locate it in the table.
- **Sort**: click any column header to sort genes by that metric.
- **Download**: export the visible table as CSV or Excel by clicking the button below the dropdowns.

*Note:* The application uses mutation data that does not account for copy number variations. Network Score rankings and pan-cancer comparisons do not consider CNVs.

## 5.2 Gene Ranking

This panel allows you to investigate the ranking of any gene of interest across tumour types, based on its **Network Score**. The figure above highlights the main features:

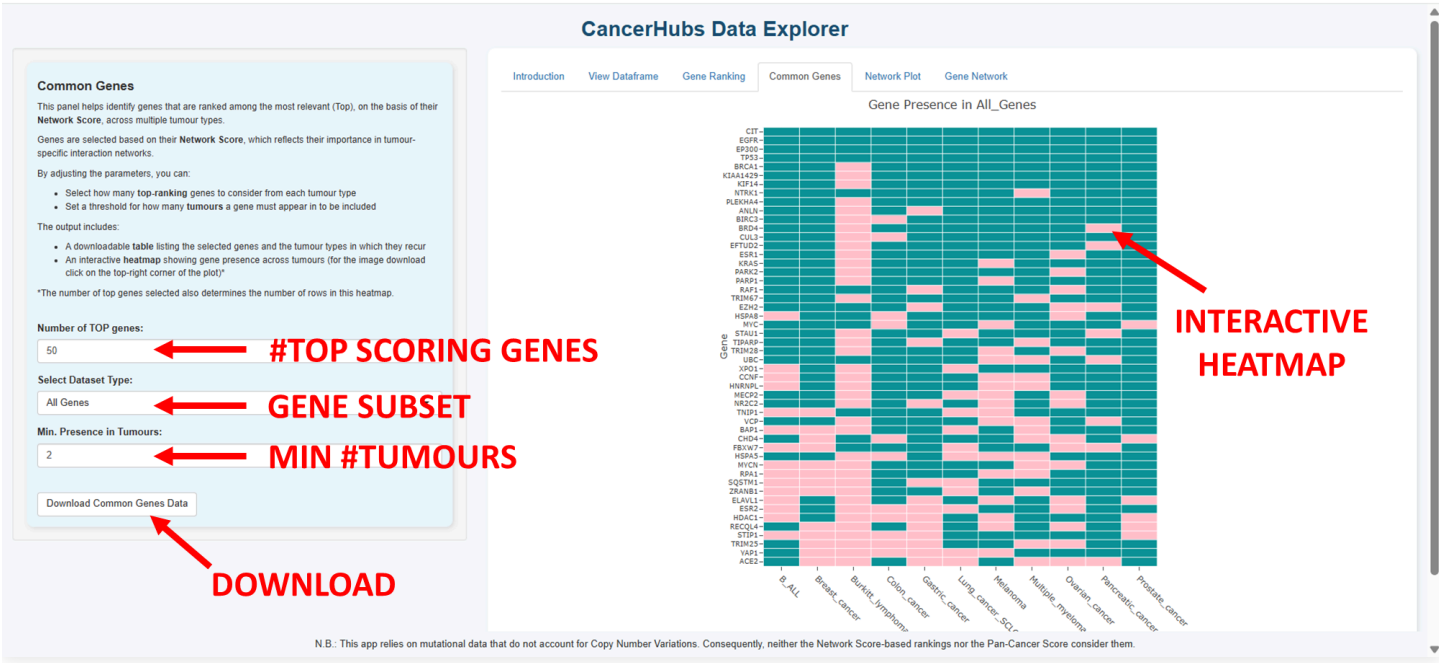



- **Gene Symbol Input**: enter any gene to query its network rank across cancers.
- **Gene Subset**: select the context for ranking (e.g. All Genes, PRECOG).
- **Ranking Plot**: a horizontal lollipop chart showing the percentile position of the gene in each tumour.
- **Interactive Table**: tabular view of Network Scores, mutation type, and ranks.
- **Pan-Cancer Score**: summarises a gene's centrality across tumours in a unified score scaled 0–100.
- **Download Buttons**: export the plot, the ranking table, or the pan-cancer summary.

*Note:* The Pan-Cancer Score is obtained by converting each tumour-specific rank into `1 - (rank / total genes)`, summing these values across all tumours, and rescaling so the top-scoring gene reaches 100.

## 5.3 Common Genes

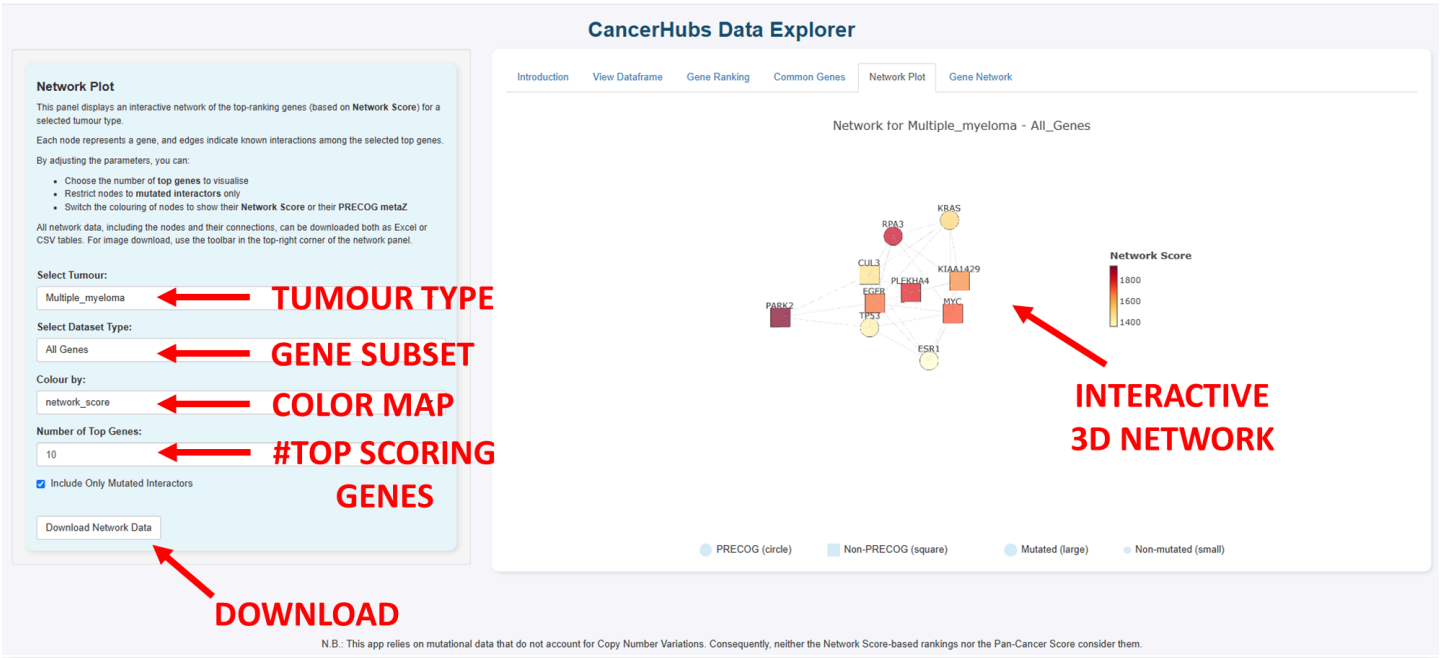
This panel helps identify genes that consistently rank among the top-scoring candidates across multiple tumour types. The interface includes:




- **Top Scoring Genes**: number of top-ranked genes to consider from each tumour (e.g. top 50).
- **Gene Subset**: choose the gene category of interest (e.g. All Genes, PRECOG).
- **Min. # Tumours**: specify the minimum number of tumour types in which a gene must appear to be considered recurrent.
- **Download Button**: export the underlying data table listing which genes appear in which tumours.
- **Interactive Heatmap**: a binary matrix where teal indicates a gene is present in the top list of a given tumour and pink indicates absence.

## 5.4 Network Plot (3D)

This panel visualises the top-ranking genes in a selected tumour type as a 3D network based on known BioGRID interactions. Nodes represent genes, and edges represent direct protein–protein interactions. The interface elements are:

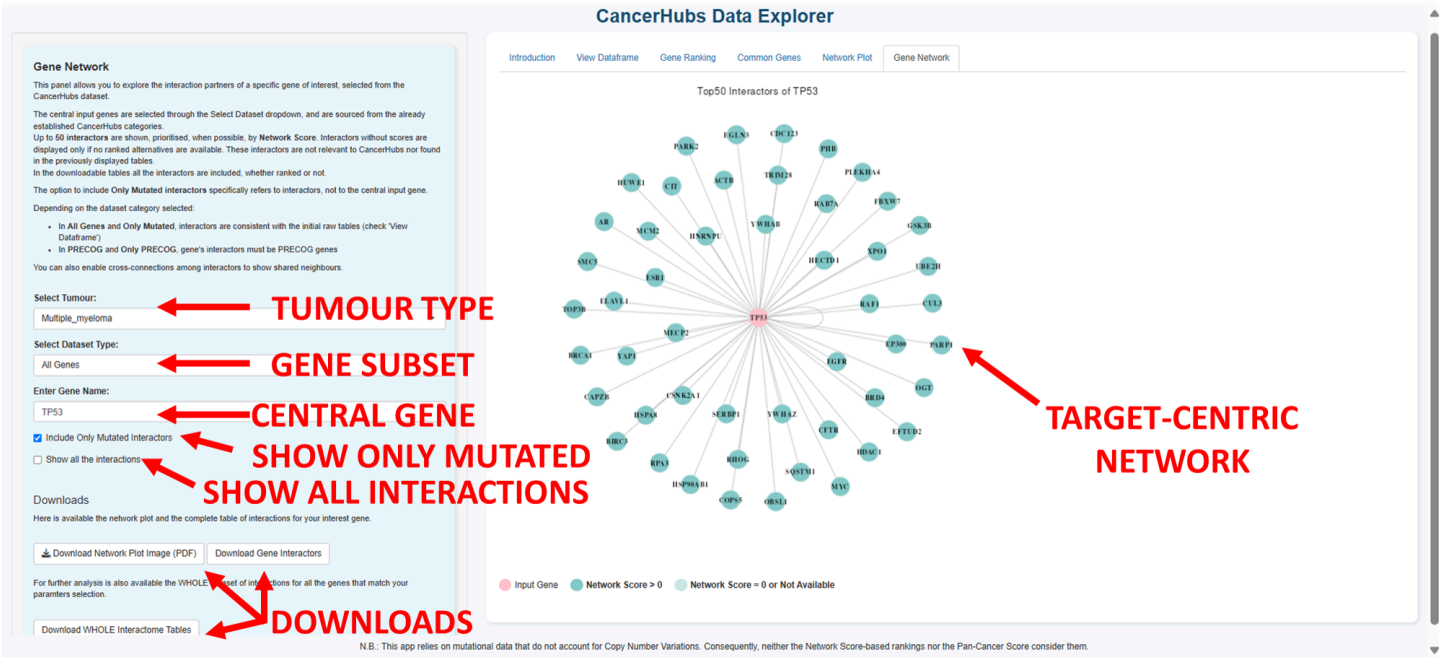



- **Tumour Type**: select the dataset to visualise (e.g. Multiple Myeloma, Breast Cancer).
- **Gene Subset**: filter for All Genes, Only Mutated, PRECOG, etc.

- **Colour Map**: choose to colour nodes by Network Score or PRECOG meta-Z score.
- **# Top Scoring Genes**: set the number of genes shown in the network (e.g. 10, 20, 50).
- **Include Only Mutated Interactors**: toggle to filter out non-mutated interaction partners.
- **Interactive 3D Network**: rotate, zoom, and inspect nodes dynamically.
- **Download**: export edge and node attribute tables for external use.

## 5.5 Gene Network (2D)

This panel allows users to explore the interactome neighbourhood of a specific gene of interest in a selected tumour context. The diagram presents a 2D radial layout of the selected gene (highlighted in pink) with up to 50 interactors arranged around it.




- **Tumour Type**: choose from the available tumour datasets.
- **Gene Subset**: filter the search space (All Genes, PRECOG, etc.).
- **Central Gene**: input the gene symbol you wish to explore.
- **Show Only Mutated Interactors**: limit the network to mutated interactors.
- **Show All Interactions**: toggle visibility of interactor–interactor links.
- **Target-Centric Network**: visualises direct partners of the central node, prioritised by Network Score.
- **Downloads**: export the network image, selected interactors, or entire interactome as tables.

*Note:* The "Only Mutated Interactors" checkbox filters for genes with any type or number of mutations, matching the set used in the Network Score calculation. In contrast, selecting the PRECOG or Only PRECOG subsets restricts the displayed interactors to those identified as significant in the PRECOG database.

## 6. Data Sources

The app retrieves data directly from the main CancerHubs repository:

- All Results Dataset : The analysis output, provides information and scores for each gene across cancer types.
- Gene Interactors List : The curated set of genes and all their corresponding interactors.
- Formatted Datasets : Preprocessed datasets used to run the analysis, obtained by formatting data from literature.
- BioGRID Interactors Data : The complete interaction records derived from the BioGRID database.
- Mutational Data Summary : A PDF document summarizing the literature sources used to extract mutational data.

## 7. Troubleshooting

If you encounter any issues while using the app, refer to the table below:

| **Problem** | **Solution** |
|---|---|
| No results found | Ensure the gene symbol is valid and present in the dataset. |
| Slow network rendering | Reduce the number of genes or interactors displayed. |
| Page not loading correctly | Check your internet connection and refresh your browser. |

## 8. Citing CancerHubs

If you use this application in your work, please cite:

> *Ivan Ferrari, Federica De Grossi, Giancarlo Lai, Stefania Oliveto, Giorgia Deroma, Stefano Biffo, Nicola Manfrini.* ***CancerHubs: a systematic data mining and elaboration approach for identifying novel cancer-related protein interaction hubs.*** *Briefings in Bioinformatics, 2025. https://doi.org/10.1093/bib/bbae635*

## 9. Support

For assistance, contact the maintainers listed in the README file. Contributions are welcome via GitHub pull requests.

*Last updated: 10/07/2025*