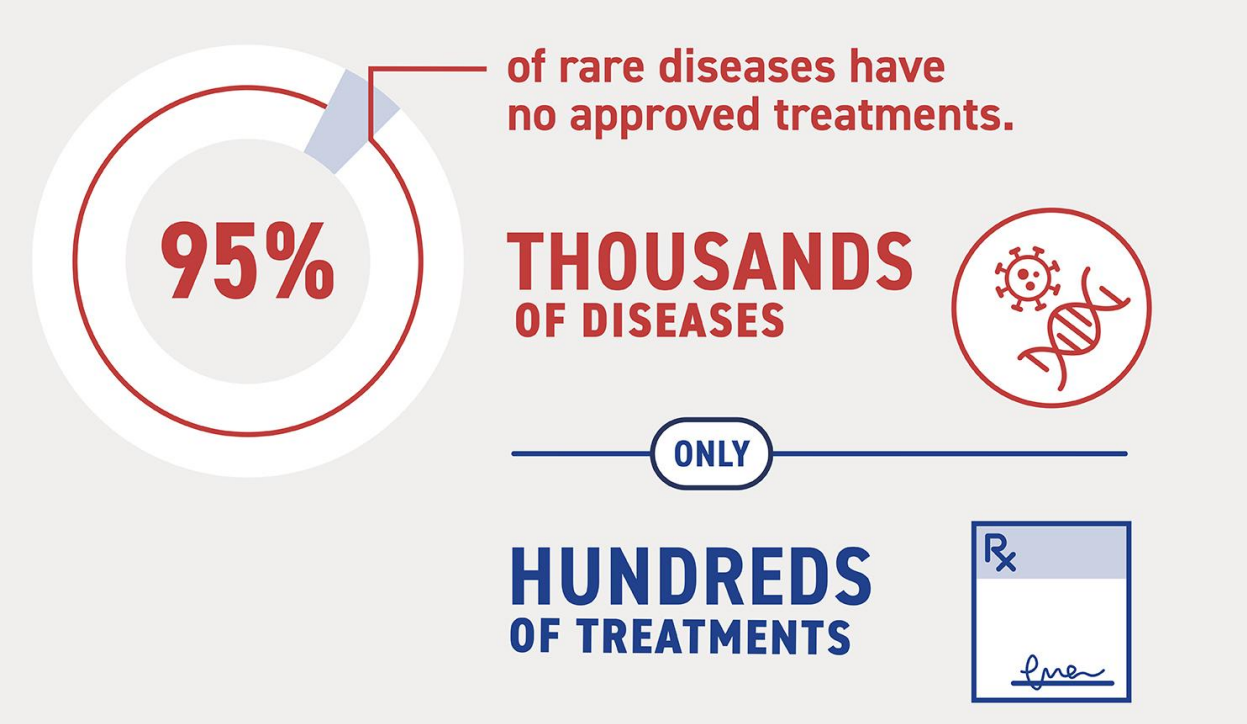
Transforming Data To Rare Disease Knowledge

Uncover insights into rare diseases to facilitate discoveries for new therapeutics

https://raresource.nih.gov rare-source@mail.nih.gov



Approximately 80% of over 10,000 rare diseases have a genetic association but most lack approved treatments.

There are more than 10,000 known rare diseases and only a few hundred have safe, effective treatments. One challenge for researchers, patients, and other community partners is the ability to access and coordinate data from disparate sources where they are siloed, and not organized or harmonized, making it difficult to integrate in a useful manner.

The ability to make connections between rare diseases, associated gene variants, and phenotypes is critical for accelerating translational research and therapeutic development.

Leveraging the power of commonalities

To harmonize and integrate relevant siloed data sources, the Therapeutic Development Branch in the intramural Division of Preclinical Innovation at NCATS conceptualized RARe-SOURCE®, an Integrated Bioinformatics Resource for Rare Diseases and in collaboration with the Advanced Biomedical and Computational Science developed and launched this platform for rare disease information.

The approach is to identify, extract, annotate, integrate, and enrich biomedical data from sources including the peer-reviewed scientific literature to help end-users uncover commonalities and share actionable knowledge that advances translational science and therapeutic discovery for rare diseases.

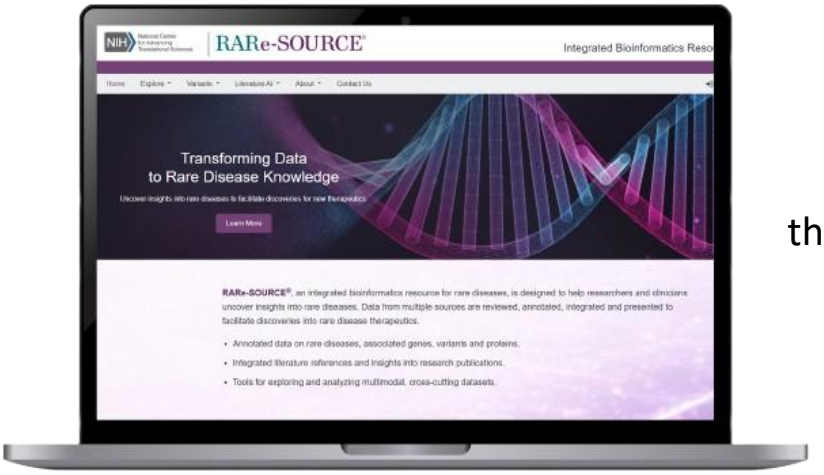
Integrate known public and private databases

Develop AI models for automating the mining of literature

Use manual curation to understand gaps in knowledge

Expand to thousands of rare diseases

RARe-SOURCE®, an Integrated Bioinformatics Resource for Rare Diseases, provides an innovative, data-driven foundation to support scientific insight and translational impact.

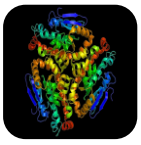


Explore various tools through interactive tables and visualizations

Diseases
Search for information on rare diseases.

Curation
Examine manually curated variants and their clinical associations.

Genes
Explore genes associated with rare disease and related details.



Proteins
Visualize proteins related to rare diseases.

Variants
Review genetic variants and their impacts.

Literature
Access literature related to rare diseases.

Section Overview: Explore Disease and Gene

The Explore Diseases and Genes section was built on a carefully developed foundation. Disease and gene information collected from NCATS' Rare Diseases Informatics Platform (RDIP) team was integrated and harmonized with additional features developed by the team, including publication access, aliases search, external cross references, and data export capabilities. Together, these elements form a unified structure with information strategically connected to enable meaningful insights.

WHY THIS MATTERS

Designed with all users in mind, the interface makes it simple to search, filter, and export results without any technical background required. Regular updates provided by NCATS' RDIP team, combined with automated updates to additional disease details, ensure the information remains comprehensive and reliable, with latest research and expanding disease relationships are always reflected.

By the end of the section, you'll learn how to navigate the disease and gene tables and explore linked information and cross-referenced databases.

Find Disease Information
Search and filter through rare diseases to view detailed disease information.

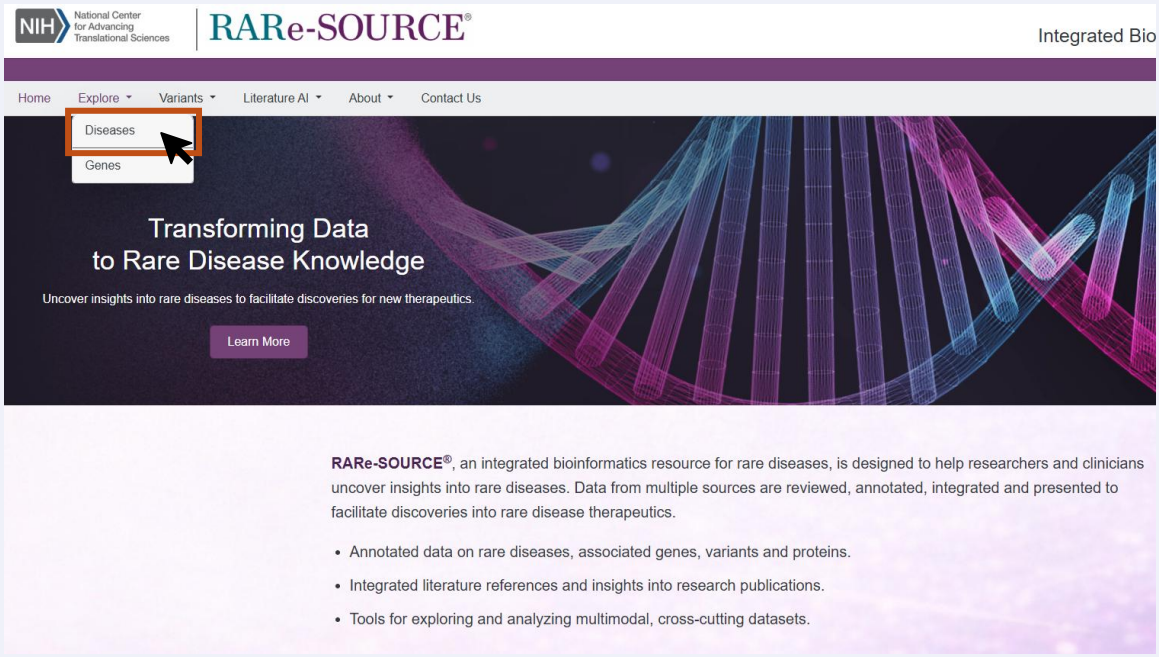
Find Gene Information
Search and filter through genes to view detailed gene information.

Explore Connections
Efficiently connect rare disease details and relationships.

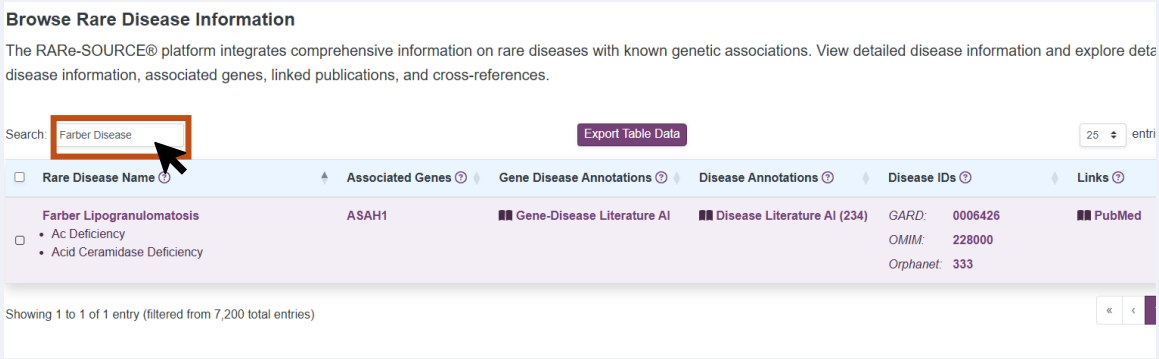
Export Findings
Download results to support your own research and analysis.

Example A: Exploring Diseases

1 Navigate to the “Explore” dropdown and select “Diseases” to launch the browse rare disease information page.

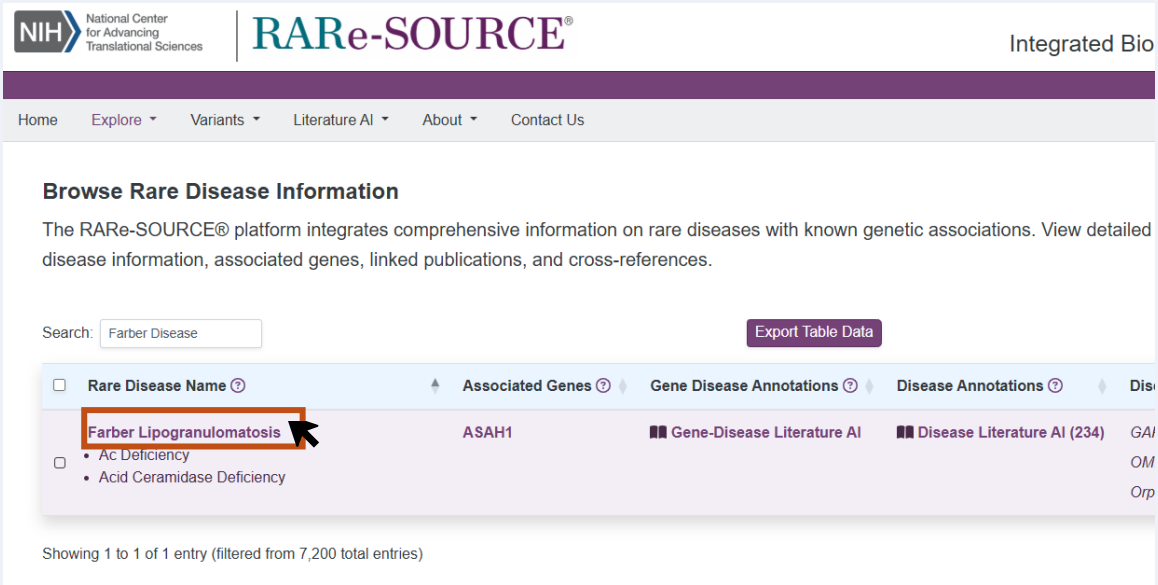


2 In the “Search” box, type in disease name. The results table shows disease information along with its synonyms, the associated gene(s), Literature AI links, cross-reference IDs, and additional links to other resources.

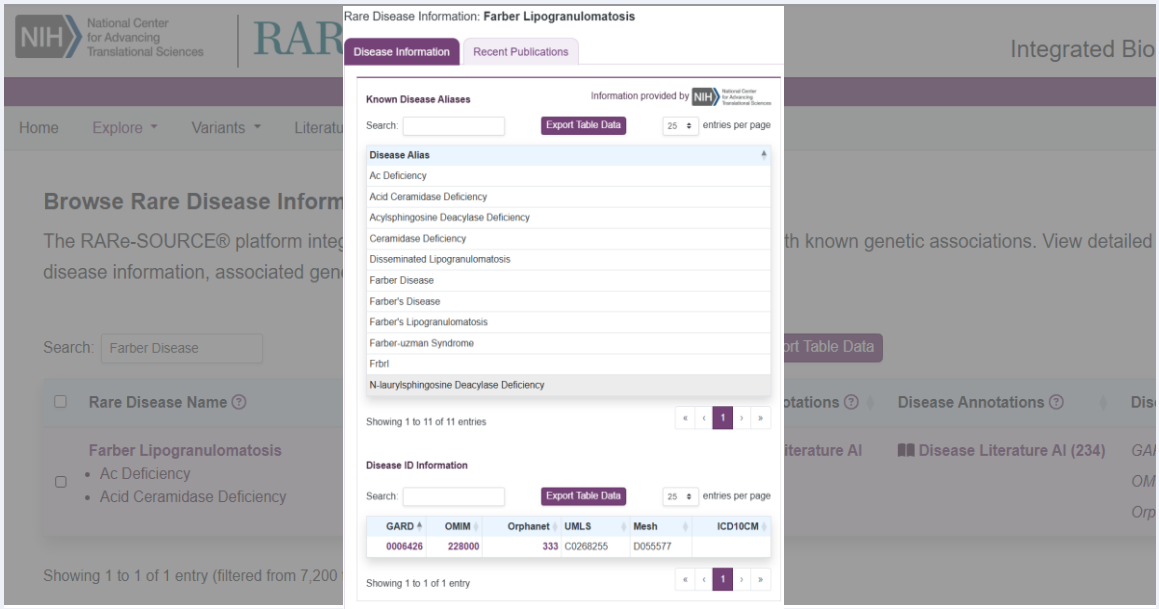


Example A: Exploring Diseases (Cont'd)

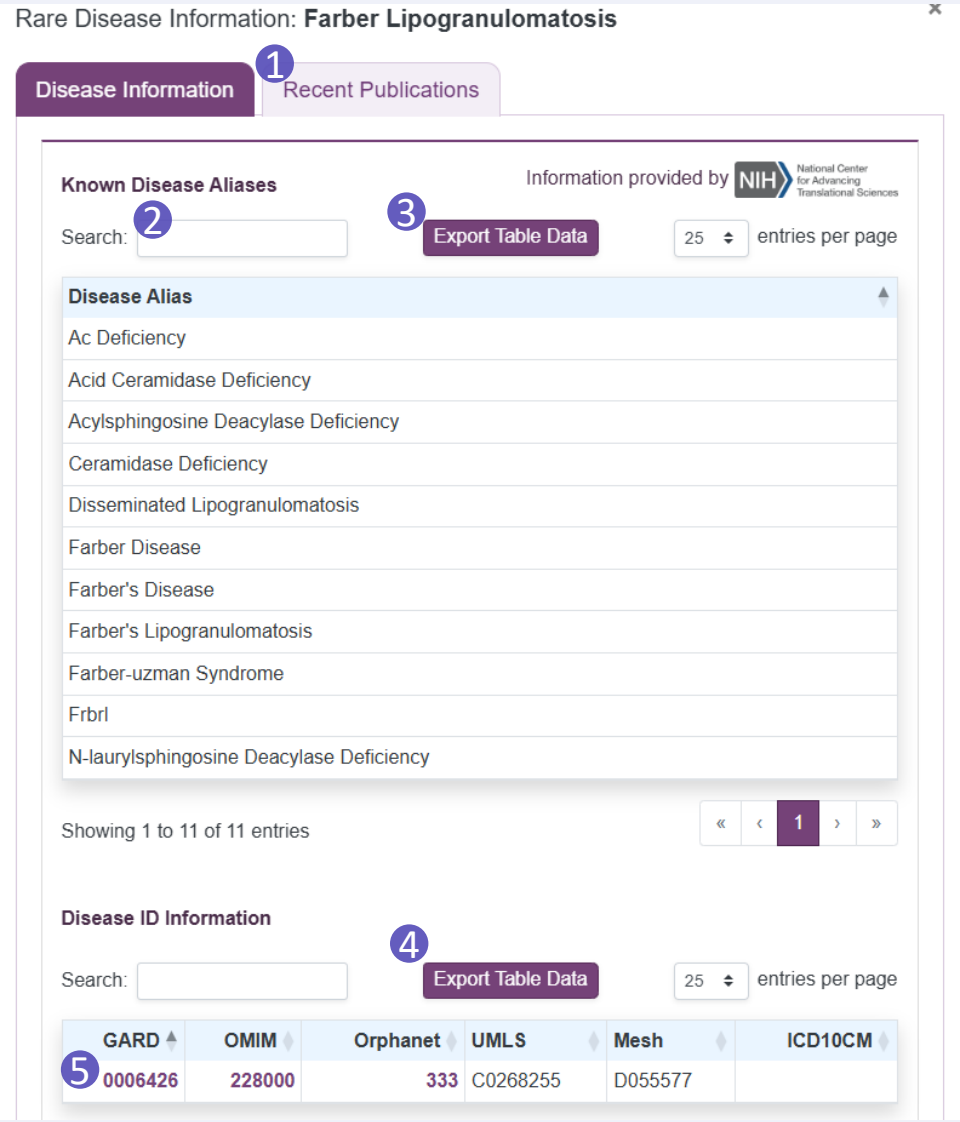
3 In the results table, select the disease name to open its information card in a pop-up window.



4 The information card will display a list of known aliases, most recent publications tab, and its IDs with links to other related resources. To view in detail, see next page.



Example A: Exploring Diseases (Cont'd)



1 **Recent Publications**
Access the three most recent publications related to the selected disease.

2 **Aliases Search**
Search through known aliases and synonyms to explore alternative names for the disease.

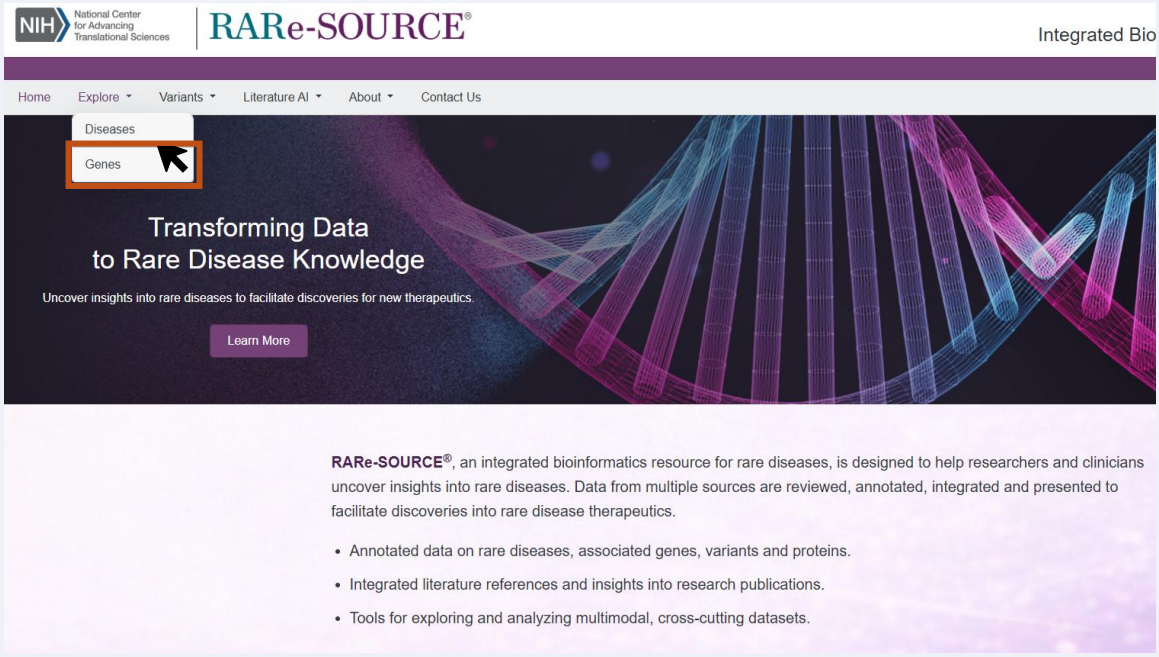
3 **Export Disease Aliases**
Download the full list of disease aliases as a CSV file.

4 **Export Disease Identifiers**
Download all linked resources IDs as a CSV file.

5 **Disease Identifier Links** Click any ID to be directed to its corresponding external resource.

Example B: Exploring Genes

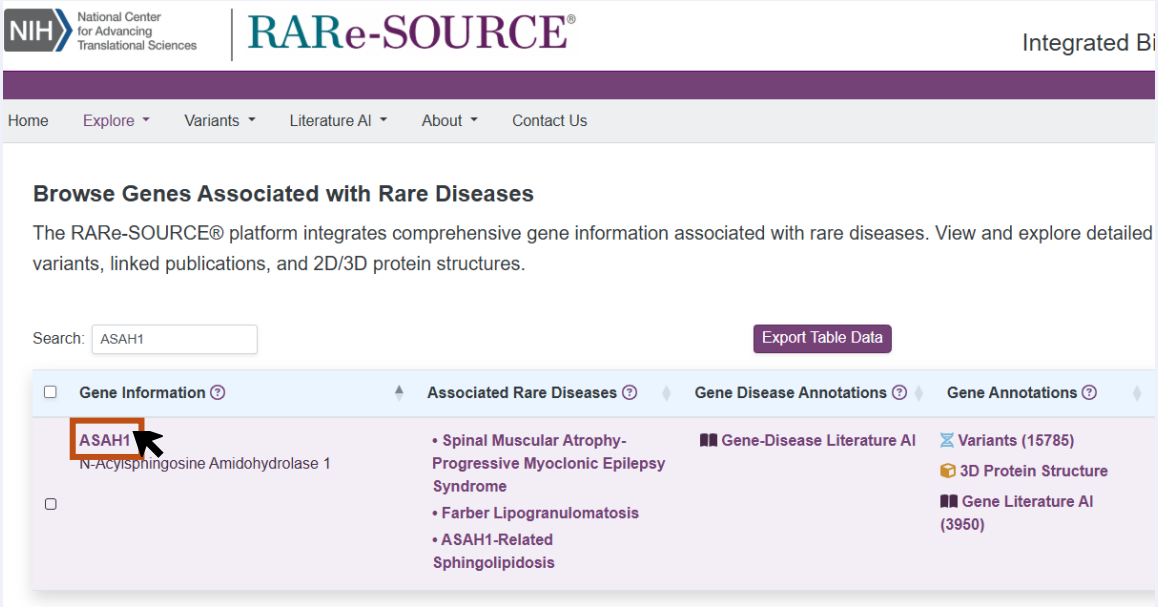
1 Navigate to the “Explore” dropdown and select “Genes” to launch the browse gene information page.



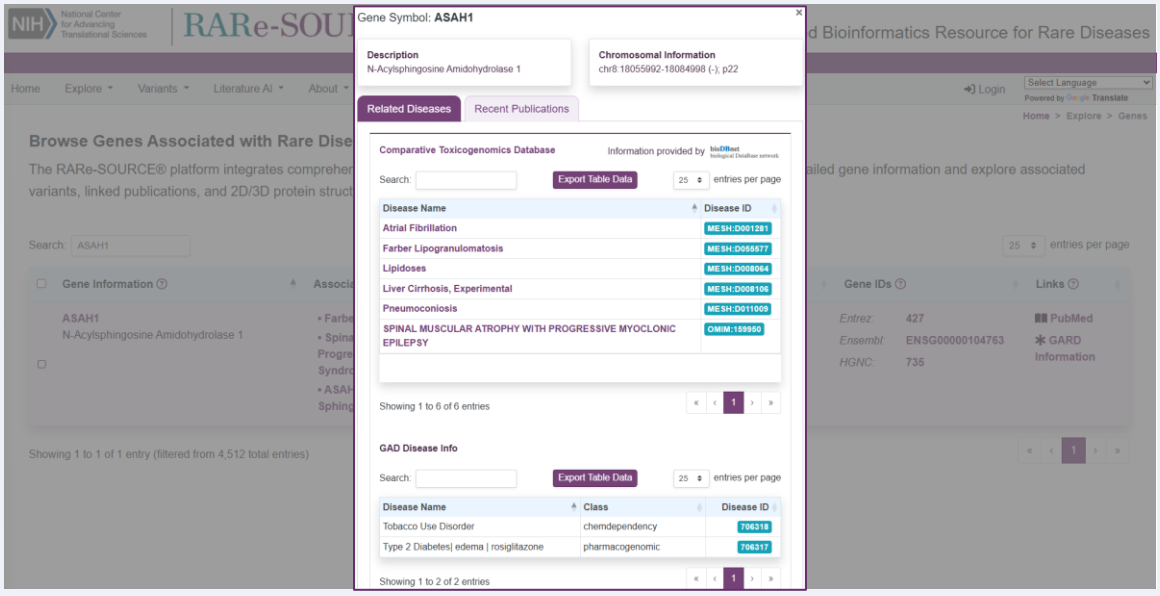
2 In the “Search” box, type in gene name. The results table shows gene information along with its synonyms, the associated disease(s), Literature AI links, cross-reference IDs, and additional links to other resources.

Example B: Exploring Genes (Cont'd)

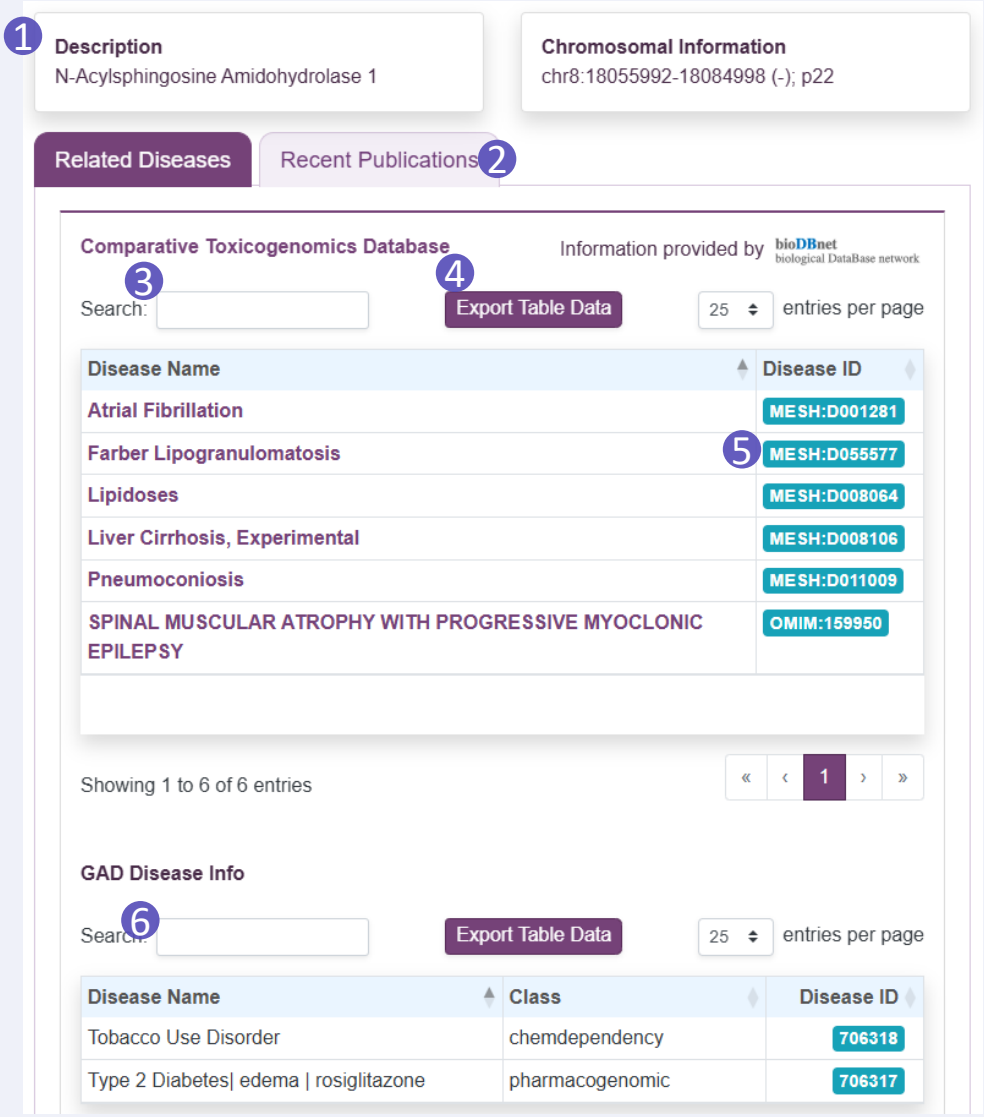
3 In the results table, select the gene name to open its information card in a pop-up window.



4 The information card will display the gene description, most recent publications tab, and associated diseases tab with IDs and links to other related resources. To view in detail, see next page.



Example B: Exploring Genes (Cont'd)



1 **Gene Description**
View the full scientific name and description of the selected gene.

2 **Recent Publications**
Access top three most recent publications related to the selected gene.

3 **CTD Search**
Search related diseases linked to the gene provided by bioDBnet.

4 **Export CTD**
Download the full list of associated diseases and their IDs as a CSV file.

5 **Disease ID Links**
Click any ID to be directed to its corresponding external resource.

6 **GAD Disease Search**
Search disease associations and classifications sourced from GAD database.

Section Overview: Explore LiteratureAI

One approach to extracting data is developing Natural Language Processing models to automate the mining of scientific literature and other databases. The LiteratureAI module does exactly this – scanning titles and abstracts of papers from Medline, identifying mentions of rare diseases and their associated genes, and extracting articles for users to review.

WHY THIS MATTERS

Searching for specific rare disease literature can be overwhelming and time consuming. LiteratureAI is built to be accessible to anyone, automatically accounting for as many known disease aliases and gene synonyms to ensure relevant articles are not missed. Disease or Gene search options organize results by focusing on publications most relevant to a specific disease or gene, helping users identify key articles without manually sorting through large volumes of literature. Additionally, the combined Gene-Disease Co-occurrence feature adds another layer by connecting disease and gene mentions together, giving users a more concise and connected view of the literature in one place.

By the end of the section, you'll learn how to navigate the literature mining tool, discover relevant scientific publications on rare diseases and their associated genes.

Search Literature
Find publication by rare disease, gene, or disease-gene co-occurrence.

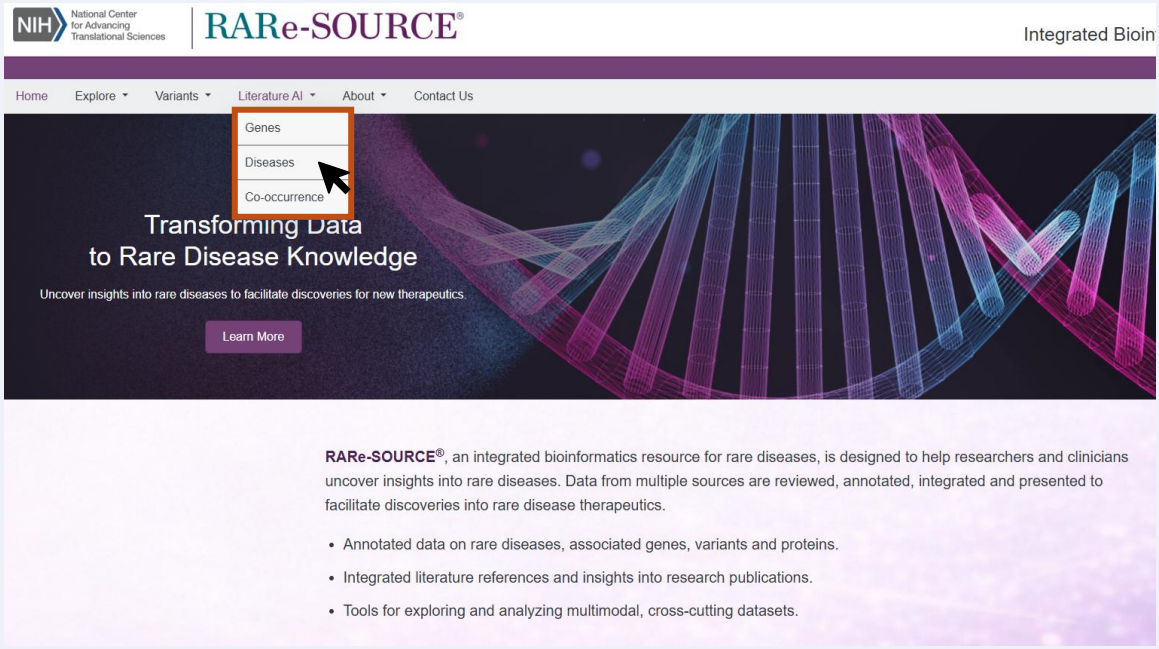
Identify Publication Trends
Use charts to track publication volume over time and across journals.

Refine Your Results
Sort and filter results easily without having to scroll through hundreds of results.

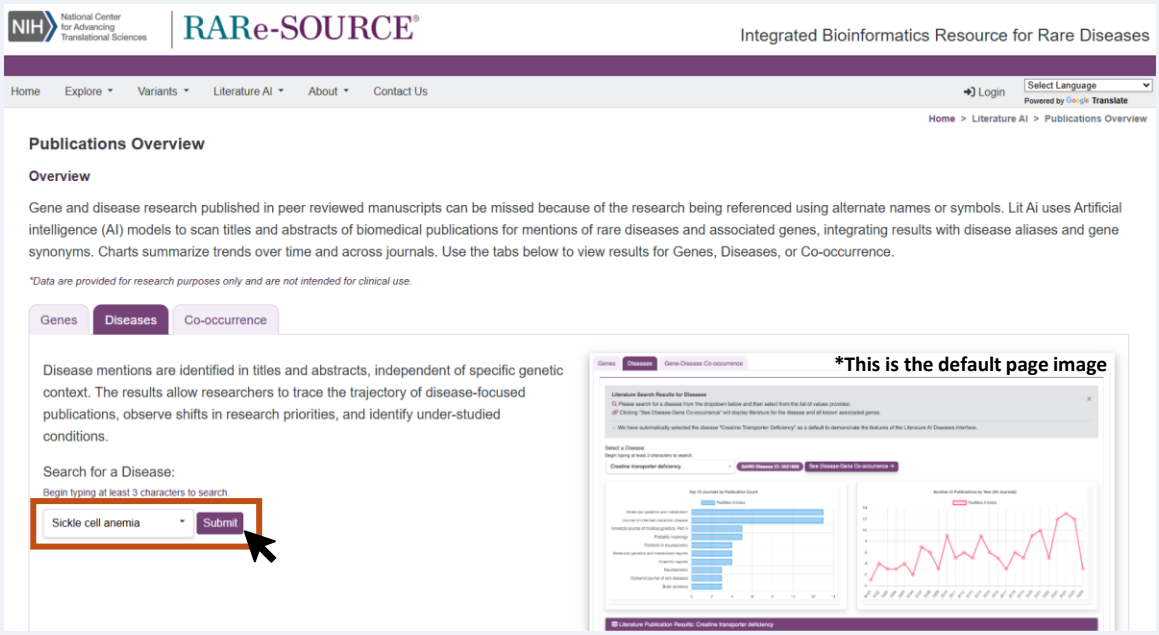
Export Findings
Download results to support your own research and analysis.

Example C: Explore Literature AI

1 Navigate to the “LiteratureAI” dropdown and select “Diseases,” “Genes,” or “Co-occurrence” to launch the publications overview page.

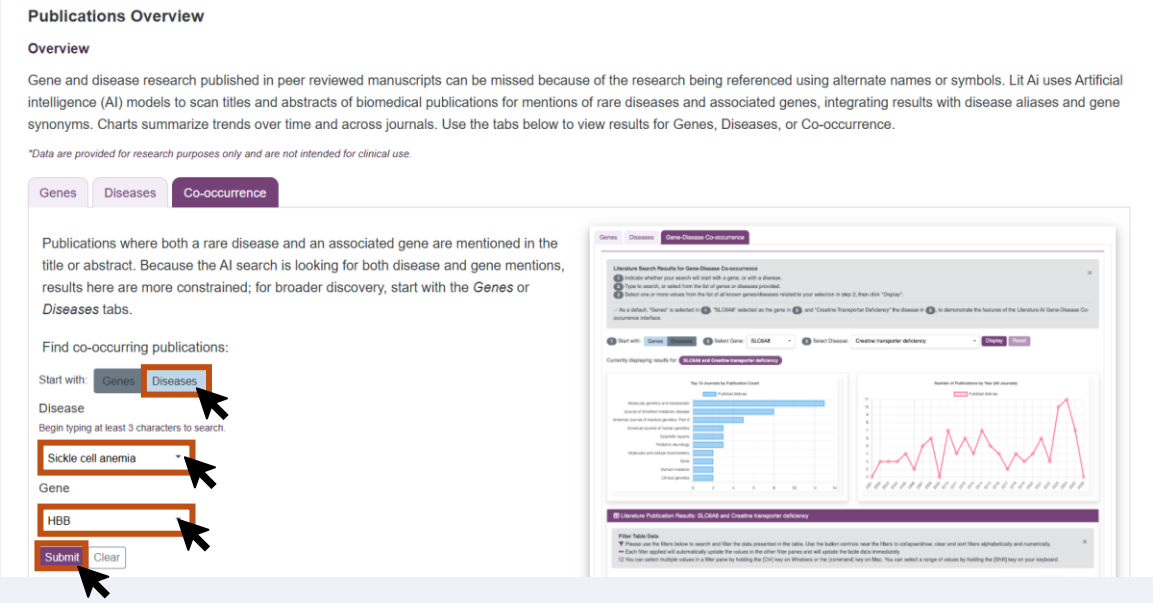


2a For disease or gene searches, type the disease or gene name in the “Search Box.” Click the “Submit” button to view the results page.

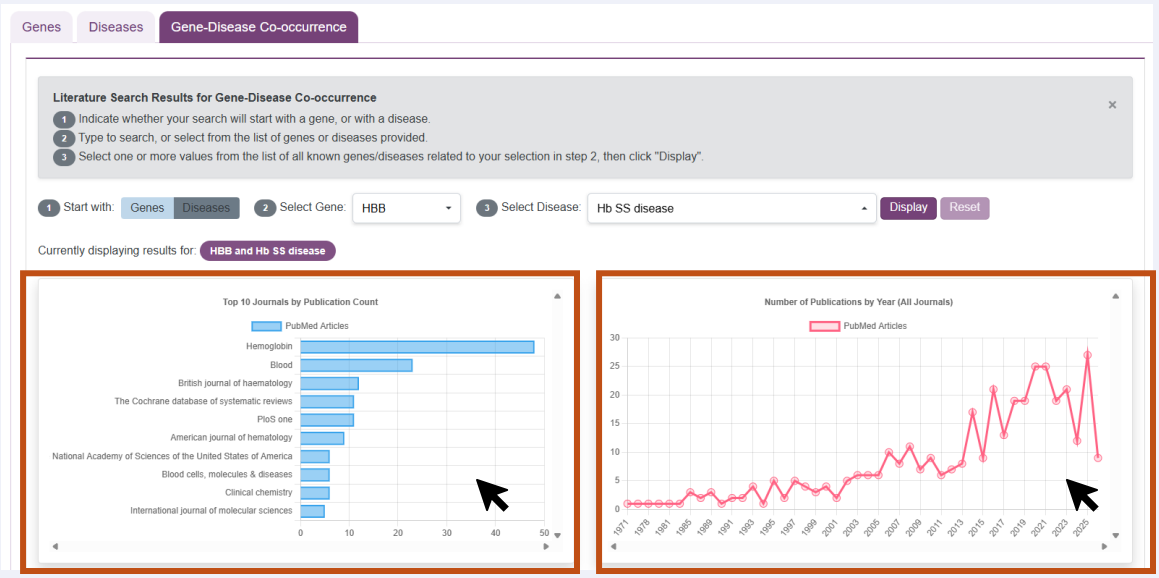


Example C: Explore LiteratureAI (Cont'd)

2b For Co-occurrence searches (Disease/Gene), select the "Gene" or "Disease" button. Then, type in your selections in the fields below and click the "Submit" button to view the results page.

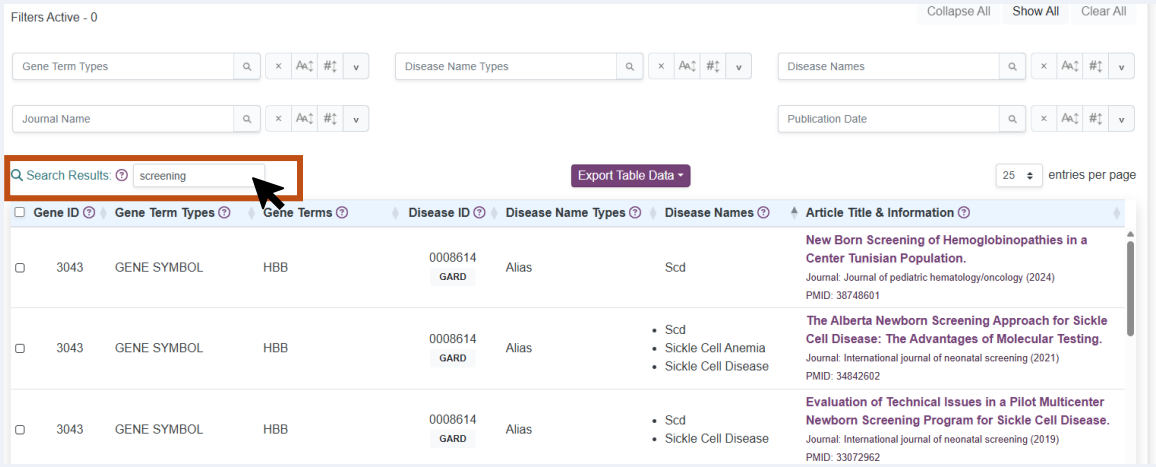


3 At the top of the results page, you can optionally view the two charts to see publication trends of the top journals by article count and the number of publications by year.

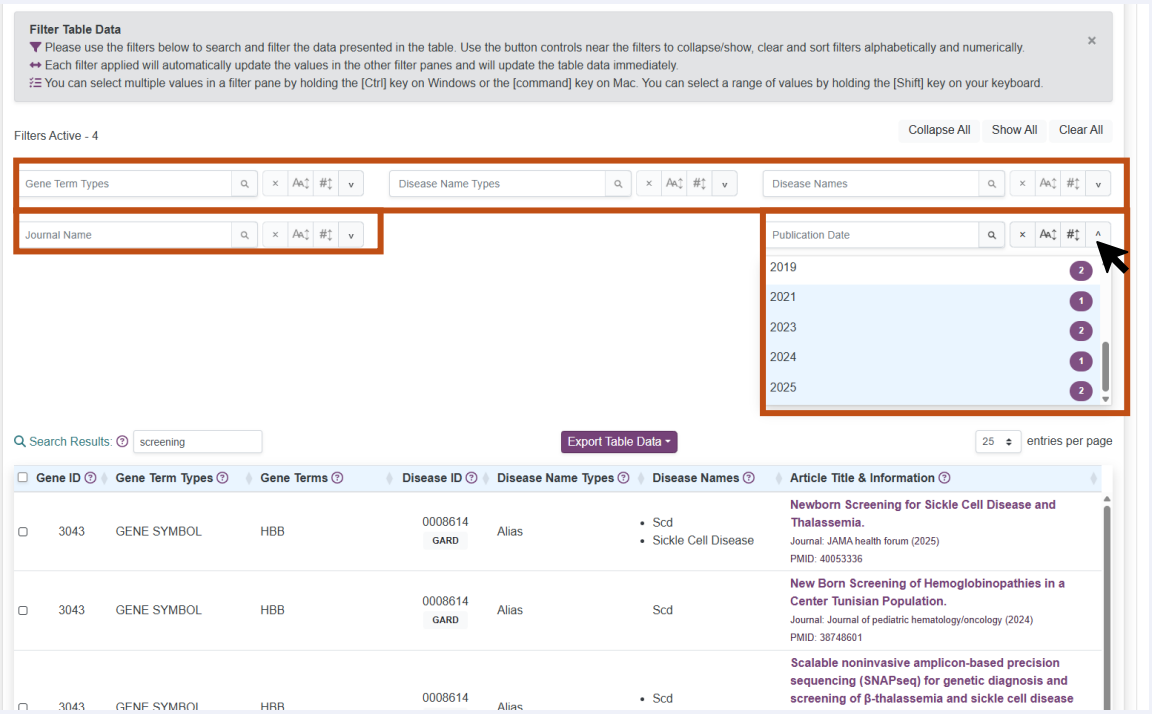


Example C: Explore LiteratureAI (Cont'd)

4 Publications are shown in the results table. To filter the articles, scroll down to the "Search Results" box and type in a keyword.

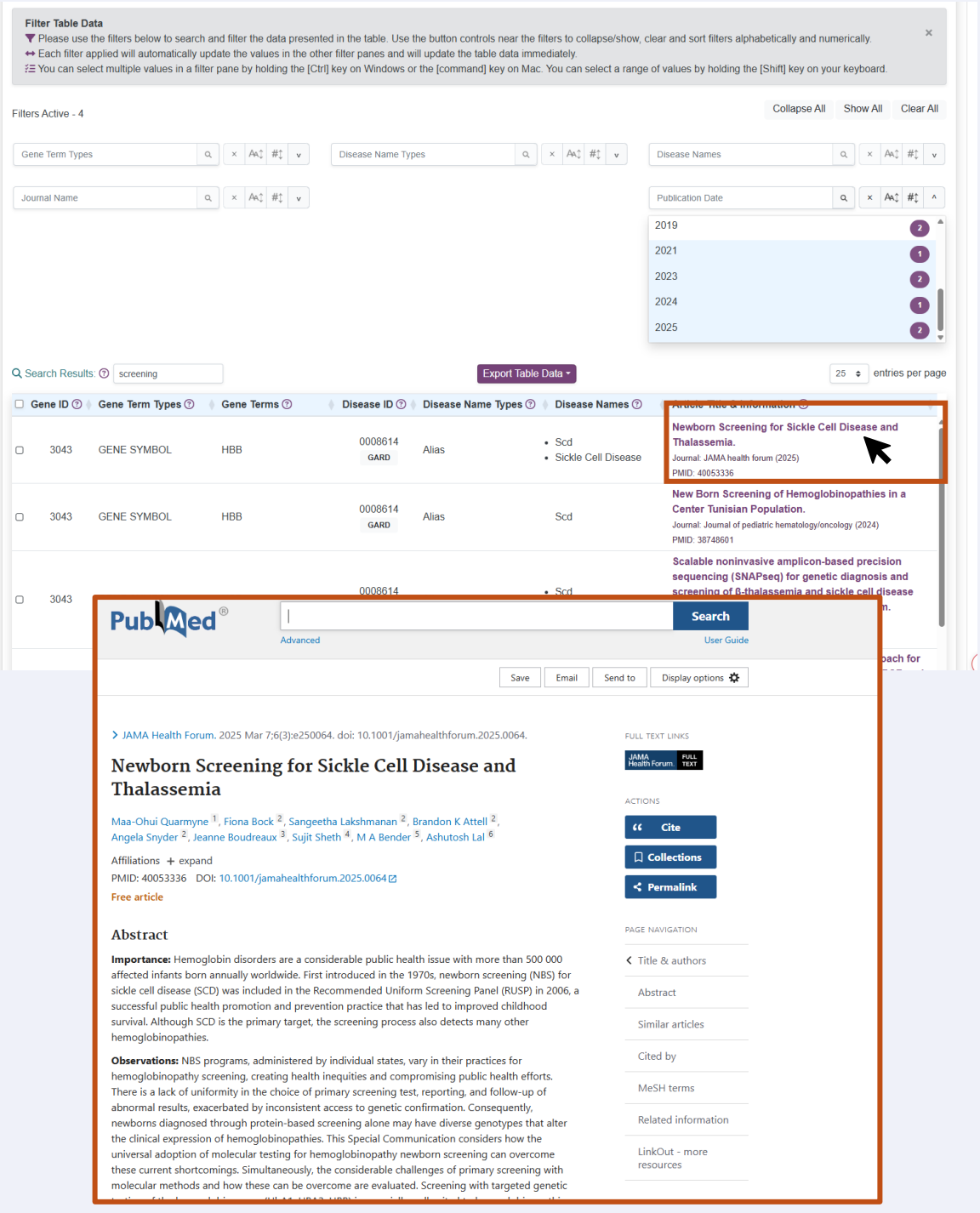


5 In the "Filter Table Data" section, you can refine results by gene or disease name, gene or disease type, journal name, or publication date.



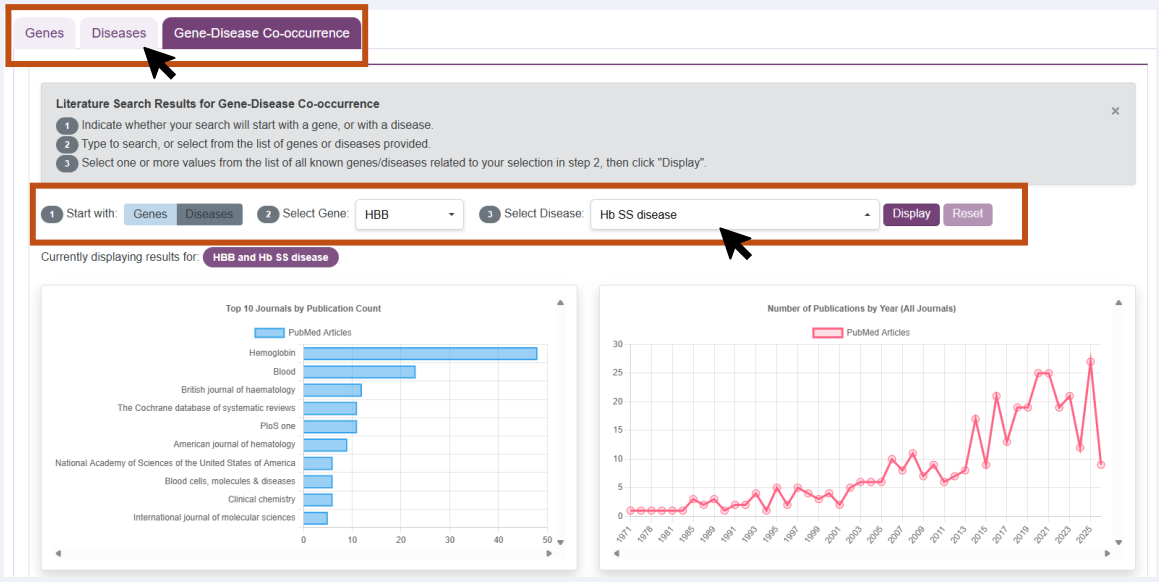
Example C: Explore LiteratureAI (Cont'd)

6 After using the search, sort, and filter features, click on the article title in the results table to access the publication in PubMed. Note that some articles are freely available, while others may require subscription access.

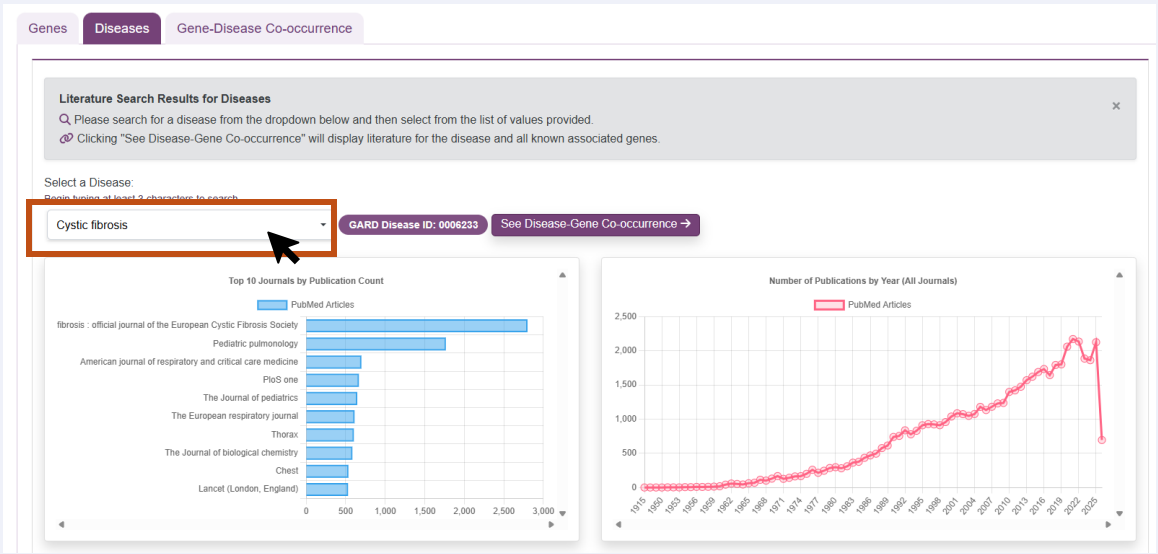


Example C: Explore LiteratureAI (Cont'd)

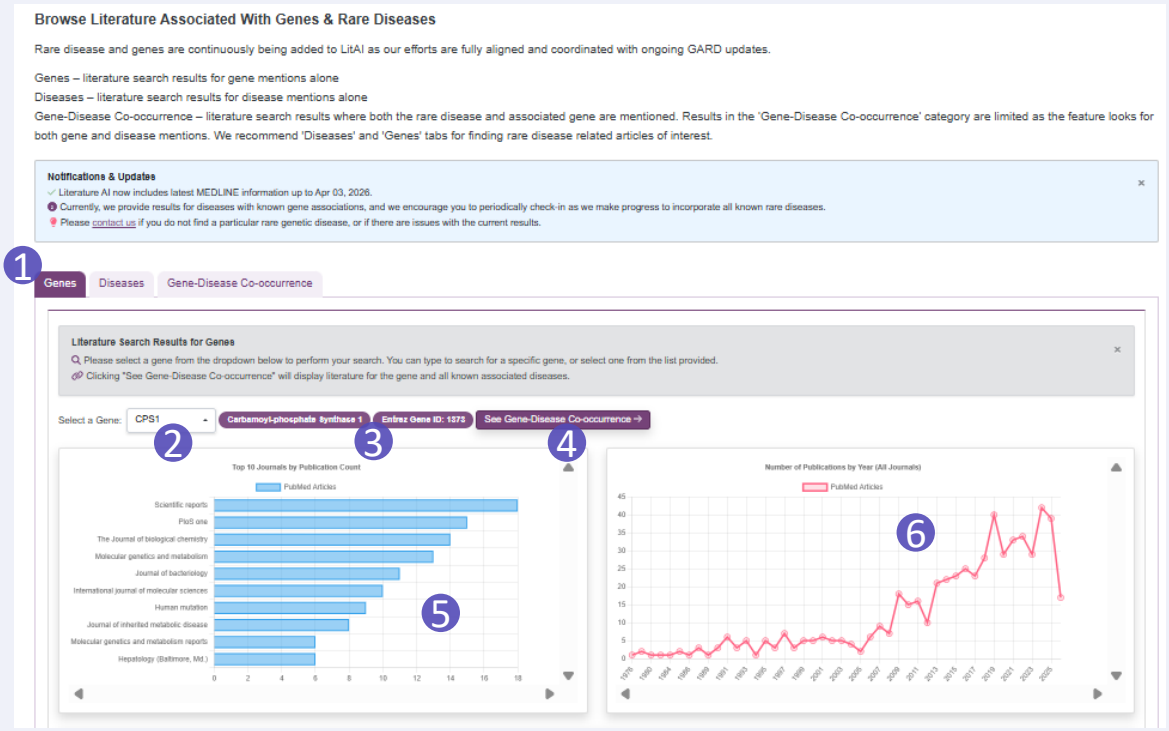
7 To start a new search or make modifications to an existing search, utilize the different tabs at the top of the page to switch between the different types of literature searches.



8 Repeat steps 3 through 6 to continue exploring articles and diving further into the publications. To view the comprehensive LiteratureAI features, see the next page.



Example C: Explore LiteratureAI (Cont'd)



1 **Category Tabs**
Switch between LiteratureAI disease, gene, or gene-disease co-occurrence.

2 **Search Box**
Search for a disease, gene, or gene-disease co-occurrence.

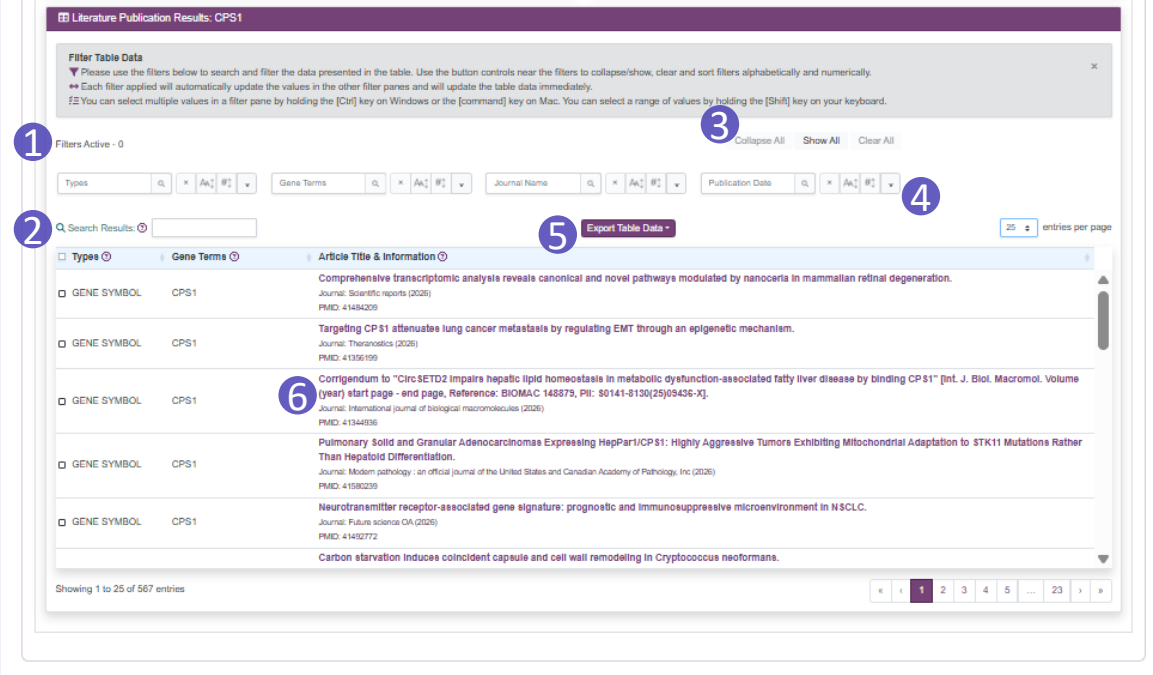
3 **Disease Information**
Displays disease, gene, or disease-gene details.

4 **Gene-Disease Co-occurrence Button**
In disease or gene mode, quickly swap to co-occurrence search.

5 **Top Publication Journals**
Bar chart showing the top 10 journals by publication count.

6 **Publication Trends by Year**
Line graph tracking research output over time.

Example C: Explore LiteratureAI (Cont'd)



1 Filtering
Narrow results to main terms, aliases, journals, and publication dates.

2 Keyword Search
Use keywords to match terms from article titles, journal names, and PMIDs.

3 View Settings
Use button control to collapse/ show, and clear filters.

4 Advance Filtering
Filter alphabetically or numerically, hold ctrl key for multiple selections.

5 Export Result Table
Download detailed comma separated values file of all results and corresponding annotations.

6 PubMed Access
Click article title to be directed to full PubMed publication.

Conclusion

RARe-SOURCE®, an integrated bioinformatics resource for rare diseases, is designed to help researchers and clinicians uncover insights into rare diseases. Data from multiple sources are reviewed, annotated, integrated and presented to facilitate discoveries into rare disease therapeutics.

Visit https://raresource.nih.gov/ to learn more