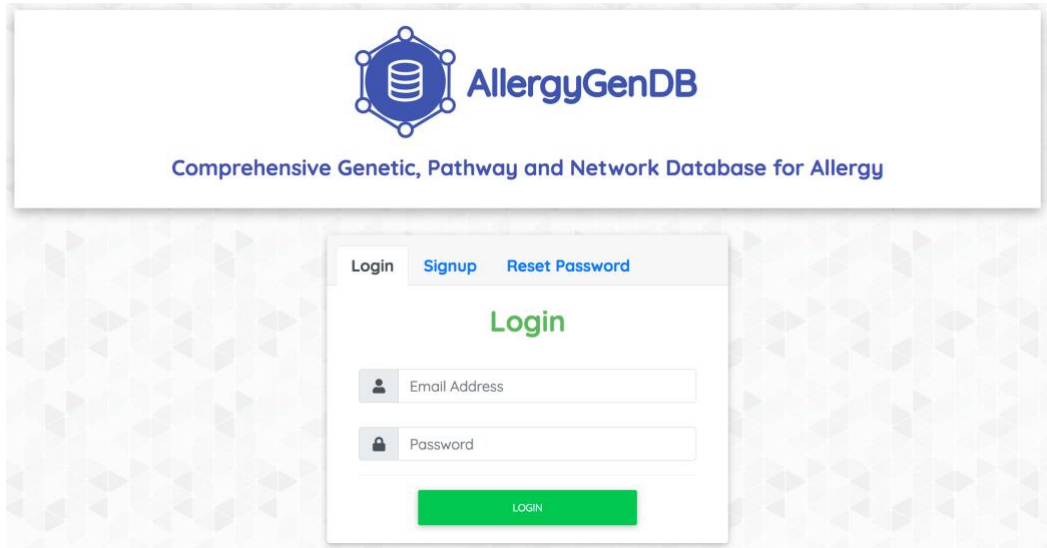




# AllergyGenDB User Manual

## 1. Login Page

User has to sign in with their email address in order to access AllergyGenDB. First time users will need to register by providing a valid email address and a password. Users can also change their account password by using the Reset Password form.

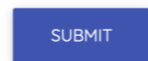


## 2. Input Forms

The AllergyGenDB homepage layout has two sections – **Literature Mining** and **Curated Databases**, as shown in the pictures below.



<b>Disease(s) (required):</b>	<input checked="" type="checkbox"/> Atopic Dermatitis <input type="checkbox"/> Food Allergy <input type="checkbox"/> Asthma <input type="checkbox"/> Allergic Rhinitis <input type="checkbox"/> Eosinophilic Esophagitis
<b>Query (required):</b>	<input checked="" type="radio"/> SNPs <input type="radio"/> Genes <input type="radio"/> Pathways
<b>Search Scope:</b>	Define Search Scope? No <input type="radio"/> Yes <input checked="" type="radio"/>
<b>Attributes:</b>	<input type="checkbox"/> Variant Annotation <input type="checkbox"/> Genome Variation
<b>Visualization:</b>	<input type="checkbox"/> Word Cloud <input type="checkbox"/> Venn Diagram <input type="checkbox"/> Jaccard Index Matrix <input type="checkbox"/> PPI Networks





PubMed Data Mining	Curated Databases
Database (required):	<input checked="" type="radio"/> GWAS Catalog <input type="radio"/> dbGaP <input type="radio"/> Monarch
Disease(s) (required):	<input type="text" value="Select Disease(s)"/>
Query (required):	<input checked="" type="radio"/> SNPs <input type="radio"/> Genes
Attributes:	<input type="checkbox"/> Variant Annotation <input type="checkbox"/> Genome Variation
Visualization:	<input type="checkbox"/> Word Cloud <input type="checkbox"/> Venn Diagram <input type="checkbox"/> Jaccard Index Matrix

**Literature Mining** section allows users to extract the disease-associated genes, SNPs and pathways from PubMed, while **Curated Databases** section supports curated association retrieval from GWAS Catalog, dbGaP and Monarch. In addition, both sections provide multiple omics annotations and data visualization for the genes and variants. Users can submit the input information and run the data analysis by clicking the **Submit** button. Once **Submit** button is clicked, a **Start Over** button will appear which can be used to clear all the previously entered input information.

### 2.1. Literature Mining

In the **Literature Mining** section, diseases-associated genes, SNPs and pathways are collected from PubMed by using the pre-annotated file collection from PubTator. For each gene and variant, we use the Log of the Product of Frequency (LPF) to measure how relevant a gene or variant is to the disease. The Log of the Product of Frequency (LPF) is a measure that explains the strength of association or co-occurrence of the gene or SNP to the disease under analysis. It is calculated as follows:

$$LPF(\text{gene/SNP, disease}) = \log_{10}\left(\frac{X}{G} * \frac{X}{T}\right)$$

Where X is the number of abstracts containing both gene and disease term, G is the number of abstracts containing the gene, and T is the number of abstracts containing the disease term.



The screenshot shows the 'PubMed Data Mining' section of the AllergyGenDB interface. It features a form with several sections: 'Disease(s) (required):' with checkboxes for Atopic Dermatitis (checked), Food Allergy, Asthma, Allergic Rhinitis, and Eosinophilic Esophagitis; 'Query (required):' with radio buttons for SNPs (selected), Genes, and Pathways; 'Search Scope:' with a toggle for 'Define Search Scope?' set to 'Yes'; 'Attributes:' with checkboxes for Variant Annotation and Genome Variation; and 'Visualization:' with checkboxes for Word Cloud, Venn Diagram, Jaccard Index Matrix, and PPI Networks. A 'SUBMIT' button is at the bottom right. Callout boxes on the right explain: 'Select one or multiple diseases' (pointing to the disease checkboxes), 'Select one query at a time' (pointing to the query radio buttons), 'Select if define custom search scope or not' (pointing to the search scope toggle), and 'Select one or more gene/SNP attributes (optional)' (pointing to the attribute checkboxes). A bottom callout box states: '"Cloud Tag Graph" is available when selecting a single disease; "Venn Diagram" and "Jaccard Index Matrix" are available when selecting two or more diseases (optional)'. A red arrow points from this box to the 'Visualization' section.

### 2.1.1. Select Disease(s)

Users are required to select one or more allergic disease(s) they are interested in. There are five available allergic diseases including asthma, atopic dermatitis, food allergy, allergic rhinitis and eosinophilic esophagitis.

### 2.1.2. Select Query Type

Users can retrieve the disease-associated SNPs, genes or pathways. In addition, protein-protein interactions (PPI) network is also available for each disease.

### 2.1.3. Select Search Scope

Instead of using the comprehensive PubMed data, users can define the search scope by entering a list of PMIDs. Once the PMID list is provided, the query result will be retrieved based on given PMID list.

The screenshot shows the 'Search Scope' section of the AllergyGenDB interface. It includes a toggle for 'Define Search Scope?' which is currently set to 'Yes'. Below this is a text input field labeled 'Enter a list of PMIDs:' with a placeholder text: 'Please enter PMIDs ... (one PMID per line, or multiple PMIDs per line, separated by commas or spaces)'. A 'SAMPLE' button is located below the input field.

### 2.1.4. Select Attributes

When SNPs or Gene is selected for a single disease, users are able to get the functional annotation information for SNPs or Genes.



### 2.1.5. Select Data Visualization Method

AllergyGenDB provides three types of visualization: when single disease is selected, the **Cloud Tag Graph** can be used to displays how frequently genes, SNPs or pathways appear in the literature of a given disease; when multiple diseases are selected, the **Venn Diagram** can be used to visualize the number of overlapping genes, SNPs or pathways among two or more allergic diseases, and the **Jaccard Index Matrix** will compute the Jaccard similarity between each disease pairs based on the queried genes, SNPs or pathways. The similarity between two sets using Jaccard index is given by:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

where  $0 \leq J(A, B) \leq 1$ .

## 2.2. Curated Databases

The screenshot shows the 'Curated Databases' section of the AllergyGenDB interface. It contains several input fields and radio buttons:

- Database (required):** Radio buttons for GWAS Catalog (selected), dbGaP, and Monarch. Callout: "Select one database at a time".
- Disease(s) (required):** A text input field labeled "Select Disease(s)". Callout: "Select one or more diseases".
- Query (required):** Radio buttons for SNPs (selected) and Genes. Callout: "Select one query at a time".
- Attributes:** Checkboxes for Variant Annotation and Genome Variation. Callout: "Select one or more gene/SNP attributes (optional)".
- Visualization:** Checkboxes for Cloud Tag Graph, Venn Diagram, and Jaccard Index Matrix. Callout: "Cloud Tag Graph" is available when choosing a single disease; "Venn Diagram" and "Jaccard Index Matrix" are available when choosing two or more diseases (optional).

A blue 'SUBMIT' button is located at the bottom right of the form.

### 2.2.1. Select Database

In the **Curated Databases** section, three databases, GWAS Catalog, dbGaP and Monarch are made available to users.

### 2.2.2. Select Disease(s)

For each curated database, the allergic diseases listed below are

Public Database	Allergic Disease
<b>GWAS</b>	Asthma, Recalcitrant Atopic Dermatitis, Atopic Eczema, Food Allergy Measurement, Allergic Rhinitis, Seasonal Allergic Rhinitis



<b>dbGaP</b>	Asthma, Atopic Dermatitis, Food Allergy, Allergic Rhinitis
<b>Monarch</b>	Asthma, Asthma (Nasal Polyps and Aspirin Intolerance), Asthma (Susceptibility to), Dermatitis, Allergic Contact Dermatitis, Atopic Eczema, Rhinitis, Seasonal Allergic Rhinitis, Allergic Rhinitis

Besides major allergic diseases, AllergyGenDB also provides over 1000 other diseases in GWAS Catalog and dbGaP for users to explore

### 2.2.3. Select Query

User can retrieve the disease-associated SNPs, genes or pathways.

### 2.2.4. Select Attributes

When SNPs or Gene is selected for a single disease, users are able to get the functional annotation information for SNPs or Genes.

### 2.2.5. Select Data Visualization Method

Similar to **Literature Mining** section, **Curated Databases** section also supports three types of data visualization methods, **Cloud Tag Graph**, **Venn Diagram** and **Jaccard Index Matrix**. Note that the **Cloud Tag Graph** displays how significant the genes or SNPs are based on the association p-values.

## 3. Output

After users submit the input form, a result area will appear below the input section.

### 3.1. Disease Association Tables

The disease associated table will display the related gene, SNP or pathway information. The output table that can be ordered by any columns by clicking the sorting arrow in the column header. On the top of the table five buttons are provided for users to access the output: the table can be copied using **Copy** button, downloaded using **Excel**, **CSV** or **PDF** buttons, and printed using **Print** button. Here a sample output of the Atopic Dermatitis associated gene information is shown in the table below.



### Gene Information

Copy CSV Excel PDF Print

Search:

Gene Symbol	Gene Id	Disease	LPF	Abstract Count
<a href="#">A2M</a>	2	Atopic Dermatitis	-7.28	2
<a href="#">ABCA1</a>	19	Atopic Dermatitis	-7.08	2
<a href="#">ABCB1</a>	5243	Atopic Dermatitis	-8.50	1
<a href="#">ABCB6</a>	10058	Atopic Dermatitis	-6.86	2
<a href="#">ABCG4</a>	64137	Atopic Dermatitis	-6.07	1
<a href="#">ABHD5</a>	51099	Atopic Dermatitis	-6.37	1
<a href="#">ABO</a>	28	Atopic Dermatitis	-7.16	3
<a href="#">ACAA1</a>	30	Atopic Dermatitis	-5.71	2
<a href="#">ACE</a>	1636	Atopic Dermatitis	-6.81	8
<a href="#">ACHE</a>	43	Atopic Dermatitis	-7.36	3

Showing 1 to 10 of 1,470 entries

Previous 1 2 3 4 5 ... 147 Next

*Gene information result of the Atopic Dermatitis related genes from literature mining.*

### 3.2. Attribute Tables

Attribute tables for a disease-related genes and SNPs will be displayed after the association table. Major SNP related annotations include functional annotations of non-coding variants from RegulomeDB and GWAVA, regulatory variant annotations from RBP-Var and SNP2TFBS databases and links to VEP and Open Targets Genetcis. Apart from multiple annotations, links to various multi-omics databases such as Ensembl, ENCODE and ClinVar are included. Similarly, all the genes are annotated with eQTL counts from GTEx and gene orthologs in other model organisms. Additional links are also included to databases from multiple functional categories such as KEGG, GTEx, Roadmap Epigenomics, and ENCODE. Here we retrieve gene attribute tables using five Eosinophilic Esophagitis-associated genes as an example.



### eQTL(s)

Search:

Gene Symbol	Tissue Count	eQTLs(GTEX)
ABCB1	6	1eQTLs(Adipose_Subcutaneous); 1eQTLs(Artery_Tibial); 2eQTLs(Brain_Hypothalamus); 25eQTLs(Heart_Atrial_Appendage); 74eQTLs(Testis); 1eQTLs(Whole_Blood)
ABCE1	8	2eQTLs(Adipose_Subcutaneous); 2eQTLs(Brain_Cerebellar_Hemisphere); 1eQTLs(Cells_Transformed_fibroblasts); 38eQTLs(Esophagus_Gastroesophageal_Junction); 12eQTLs(Esophagus_Muscularis); 6eQTLs(Lung); 1eQTLs(Skin_Sun_Exposed_Lower_leg); 1eQTLs(Thyroid)
ALOX15	17	27eQTLs(Artery_Tibial); 176eQTLs(Brain_Anterior_cingulate_cortex_BA24); 88eQTLs(Brain_Caudate_basal_ganglia); 22eQTLs(Brain_Cerebellar_Hemisphere); 182eQTLs(Brain_Cerebellum); 174eQTLs(Brain_Cortex); 41eQTLs(Brain_Hippocampus); 174eQTLs(Brain_Hypothalamus); 49eQTLs(Brain_Nucleus_accumbens_basal_ganglia); 109eQTLs(Colon_Sigmoid); 95eQTLs(Colon_Transverse); 41eQTLs(Esophagus_Gastroesophageal_Junction); 186eQTLs(Esophagus_Muscularis); 10eQTLs(Heart_Atrial_Appendage); 166eQTLs(Heart_Left_Ventricle); 13eQTLs(Skin_Sun_Exposed_Lower_leg); 172eQTLs(Whole_Blood)
AR	13	4eQTLs(Adipose_Visceral_Omentum); 1eQTLs(Artery_Tibial); 3eQTLs(Brain_Cortex); 537eQTLs(Brain_Frontal_Cortex_BA9); 2eQTLs(Brain_Hypothalamus); 245eQTLs(Brain_Nucleus_accumbens_basal_ganglia); 368eQTLs(Cells_Transformed_fibroblasts); 1eQTLs(Colon_Sigmoid); 118eQTLs(Colon_Transverse); 4eQTLs(Esophagus_Muscularis); 1eQTLs(Lung); 1eQTLs(Muscle_Skeletal); 499eQTLs(Thyroid)
ATP4A	3	2eQTLs(Adrenal_Gland); 1eQTLs(Brain_Spinal_cord_cervical_c-1); 44eQTLs(Thyroid)

### Genome and Variation Information

Search:

Gene Symbol	ENCODE	SNPeffect	dbVar	ClinVar	Biogps
ABCB1					
ABCE1					
ALOX15					
AR					
ATP4A					

### Gene Expression Information

Search:

Gene Symbol	GEO Profiles	GTEX eQTL(s)	Promoter(GE)	Promoter(cis)
ABCB1				
ABCE1				
ALOX15				
AR				
ATP4A				





### Pathway Information

Search:

Gene Symbol	KEGG	Reactome	BioCarta
ABCB1			
ABCE1			
ALOX15			
AR			
ATP4A			

### Ontology Information

Search:

Gene Symbol	Cellular Component	Biological Process	Molecular Function
ABCB1	AmiGO	AmiGO	AmiGO
ABCE1	AmiGO	AmiGO	AmiGO
ALOX15	AmiGO	AmiGO	AmiGO
AR	AmiGO	AmiGO	AmiGO
ATP4A	AmiGO	AmiGO	AmiGO

### Protien Information

Search:

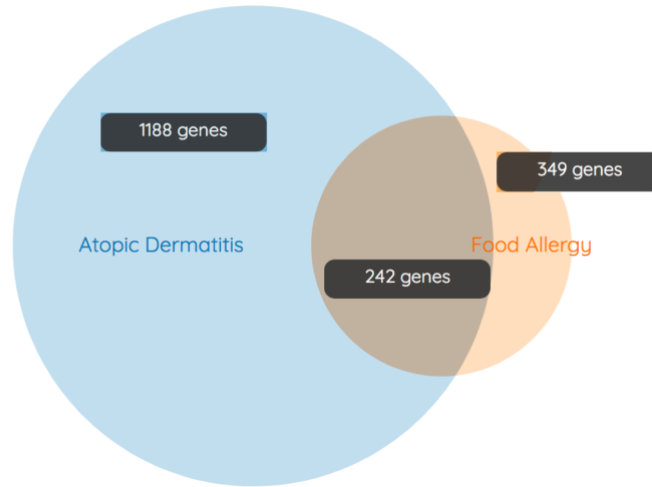
Gene Symbol	Protien Atlas	Uniprot	PFam	SMART
ABCB1				
ABCE1				
ALOX15				
AR				
ATP4A				





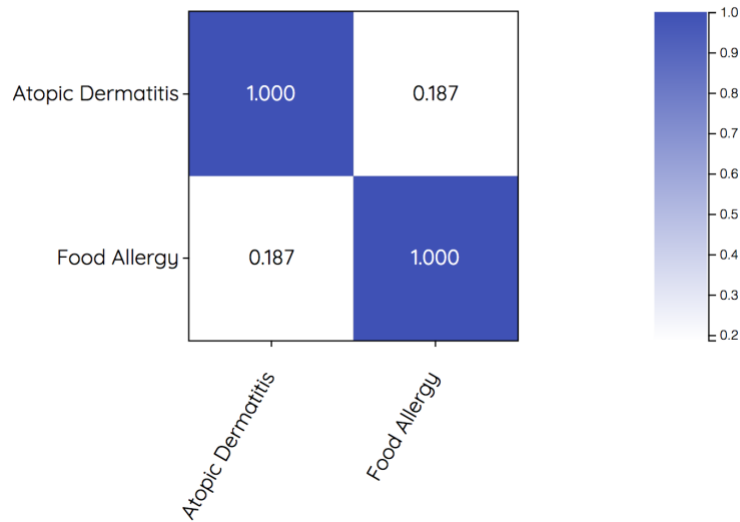


### Overlap Between Diseases



An example of overlapping Venn diagram for Atopic Dermatitis related and Food Allergy related genes.

### Jaccard Index Similarity Matrix



Jaccard index matrix of Atopic Dermatitis and Food Allergy based on the related genes.