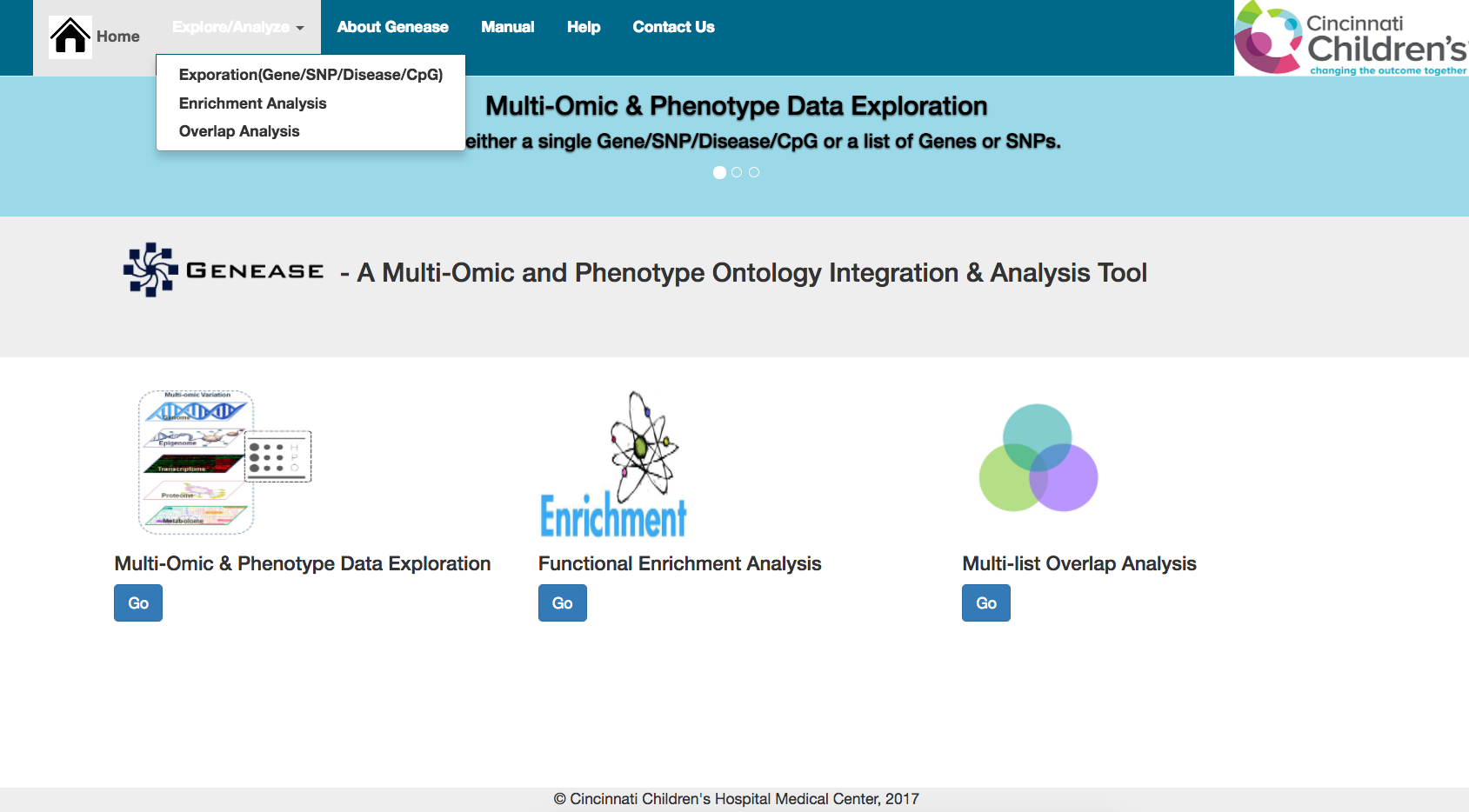
**GENEASE USER MANUAL**

GENEASE is a dynamic, real-time bioinformatics tool which can be used to explore a single or sets of genes, SNPs, or a methylation sites providing simultaneous access to multiple external databases. Additionally, GENEASE can be used to perform enrichment analysis of functional annotation terms in candidate sets of genes/SNPs and find overlaps among multiple candidates based on the enriched annotations.

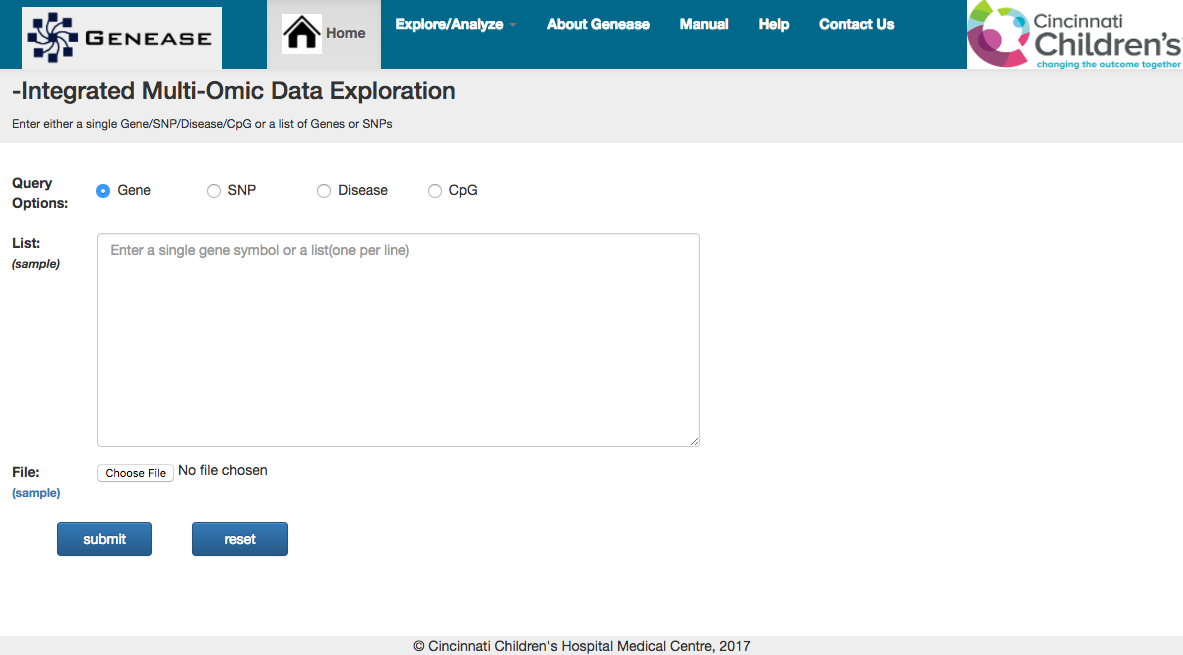


***GENEASE home screen***

GENEASE can be broadly divided into three modules. All modules are easily navigable from the main screen in the tool by either clicking the “Go” button underneath each module image or by selecting the module in the “Explore/Analyze” menu.

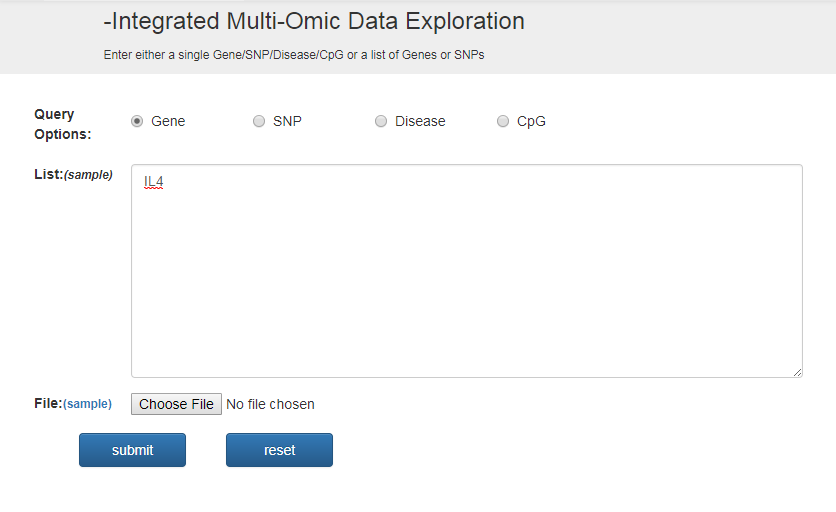
**Exploration Module:**

GENEASE exploration output utilizes a multi-tiered grid layout with a “sliding” window design. External databases are rendered in IFRAMEs, which are web browsers embedded within a main web browser. This means external applications like KEGG pathway, Ensembl gene, Gen2Phen, PubMed etc., are loaded within the GENEASE workspace. When user submits a query by Gene, SNP, disease, or methylation-site, related information, client applications are displayed in categorized layouts where user can navigate through the available information.



*GENEASE explore screen*

1. **Gene Query:**



*Gene explore query*

Gene query based on HGNC symbols result in an output having summarized information about the gene embedded from with links to different databases categorized into information categories namely Expression, Pathway, and Protein etc. A sample use-case of gene exploration could be:

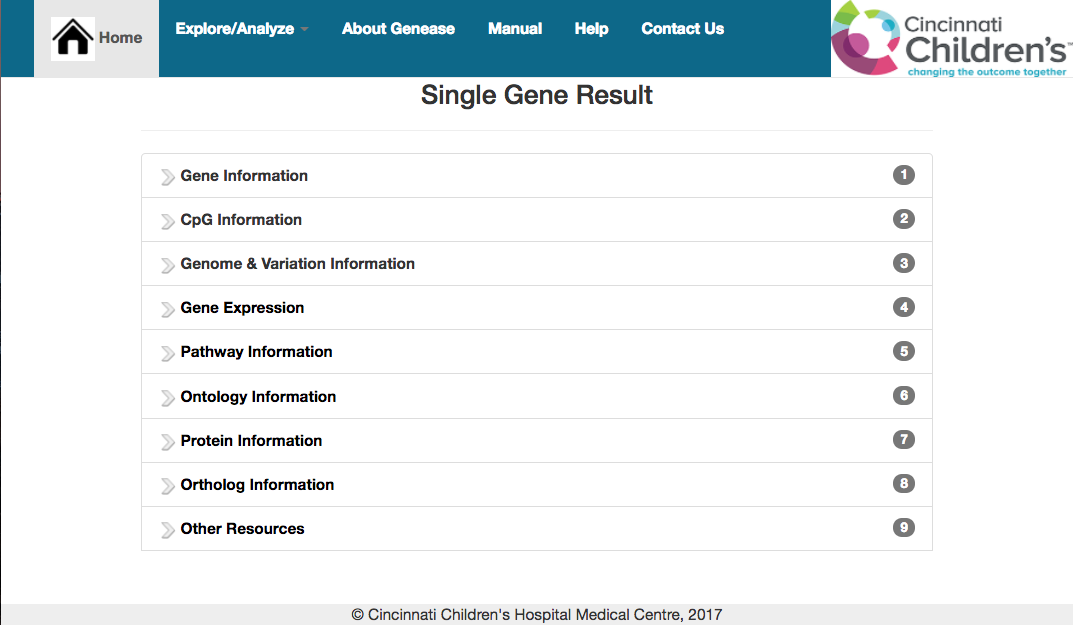
*What are the diseases associated with specific genes of interest. Example: IL13?*

*Where (or in which tissue) is IL13 expressed? What are the significant eQTLs mapped to IL13?*

*What are the GO terms, metabolic pathways related to IL13?*

*What are the methylation sites mapped to IL13?*

Supply the input gene symbol in the explore screen by selecting “Gene” (default) as the input type and click “Submit”.



*Single Gene explore result*

Underneath each “slider” either information about the gene or link to an external database can be found.

Summary of gene information can be explored by clicking on the “**Gene Information**“ slider.

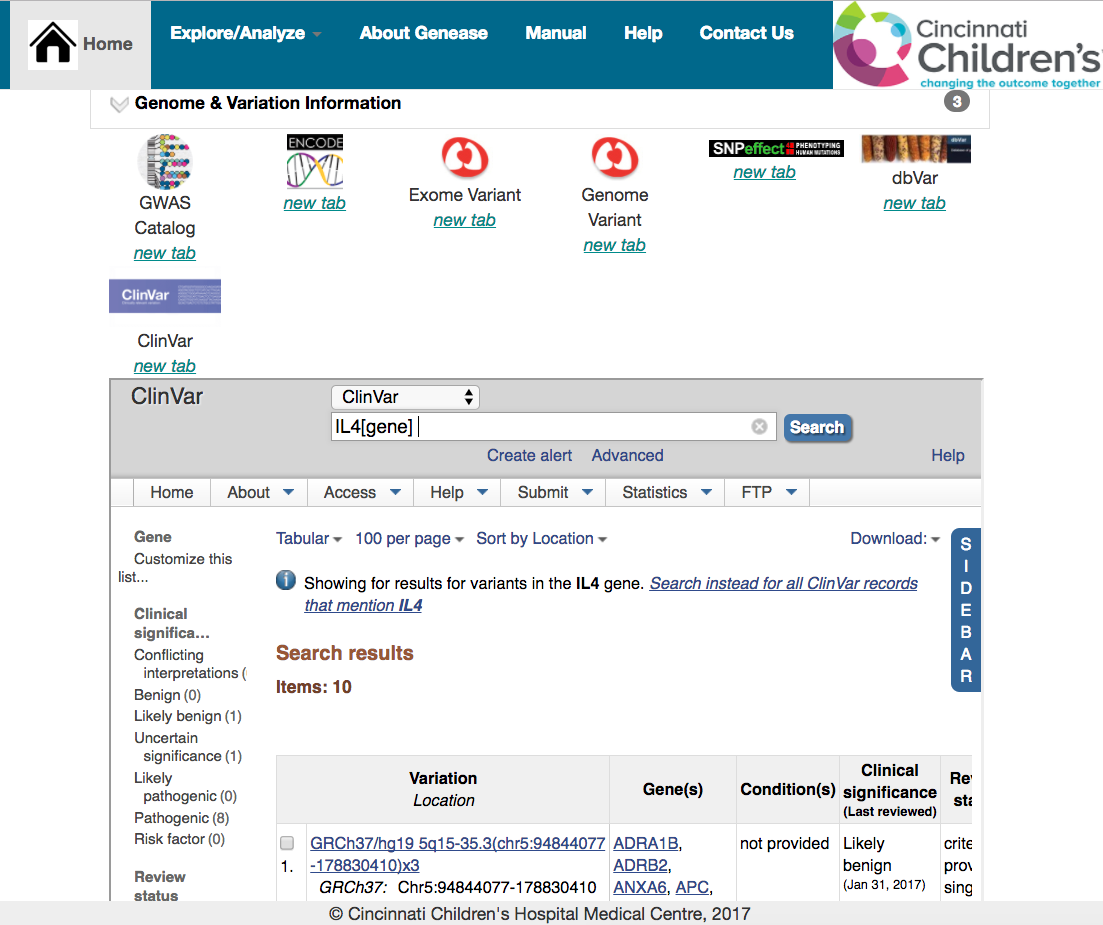


*Summarized gene information*

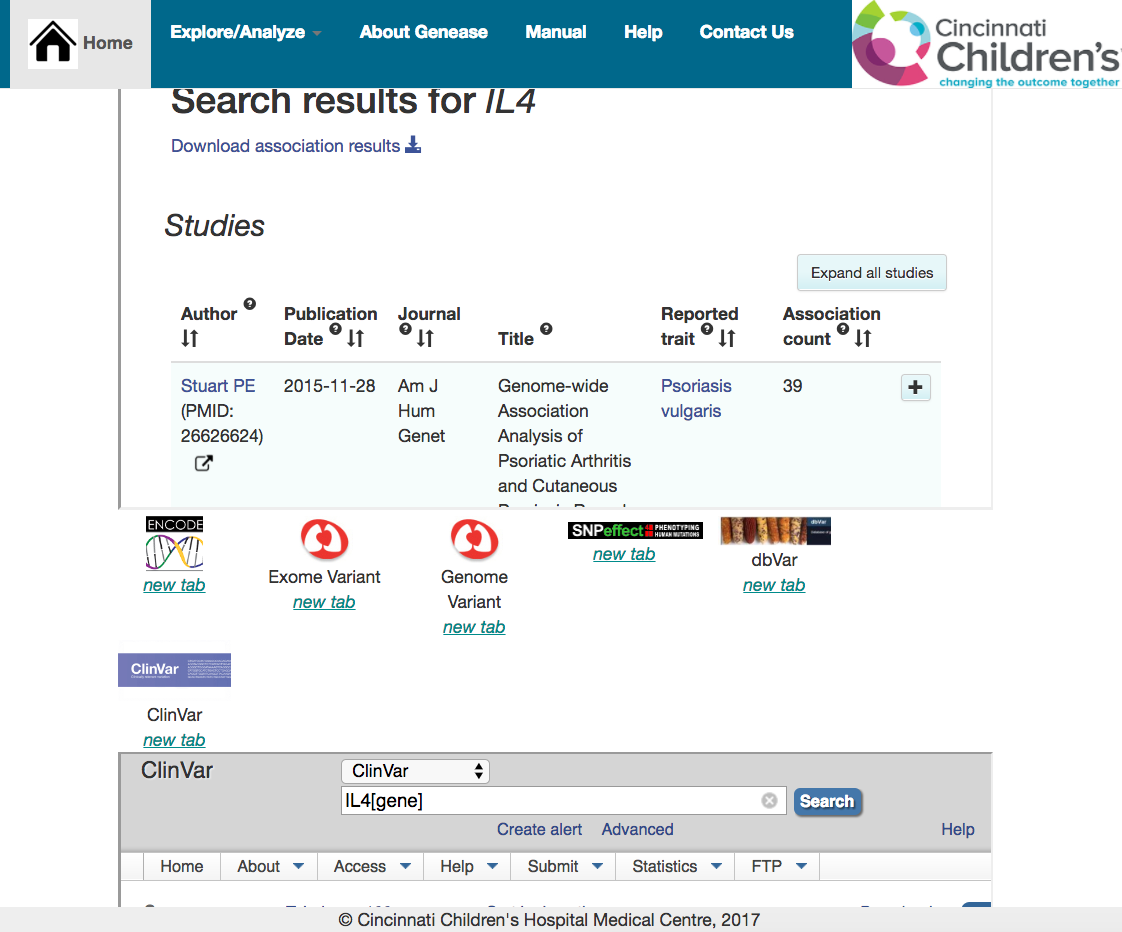
In case of external links, a single-click on the logo will open the database in the same screen as an HTML iframe with the input appended automatically to the request.

To open the link in a separate tab click on the “new tab” link underneath the logo.

Multiple external databases can be explored simultaneously by clicking their logos.



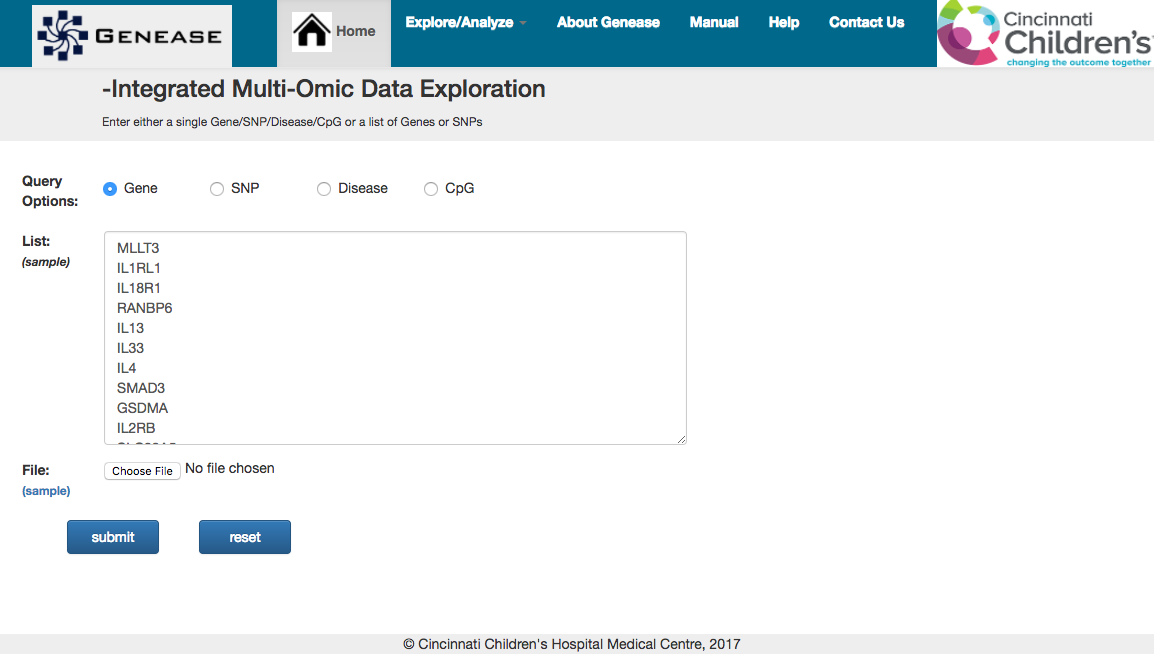
*Exploring “ClinVar” database for ‘IL4’ linked variants*



*Exploring “GWAS Catalog” and “ClinVar” databases simultaneously*

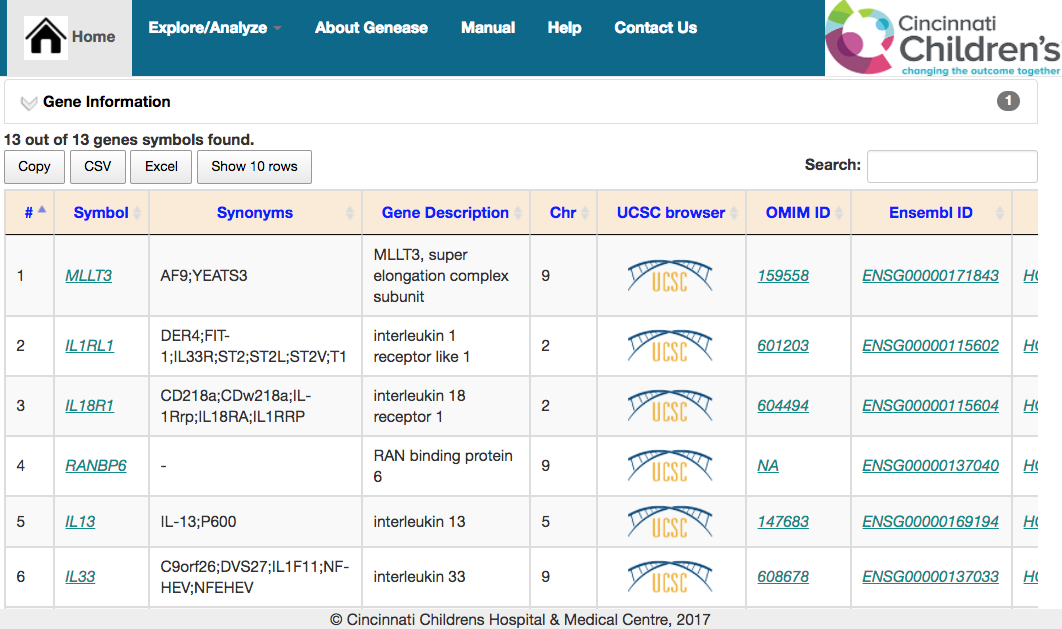
GENEASE can be used to simultaneously explore a list of genes. Select “Gene” (default) as the query option and enter the list of genes in the “List” text-area or upload a file.

To understand the working of the tool, sample genes can be used by clicking on “sample” link beside “List” or “File” input category.



*Gene list explore query*

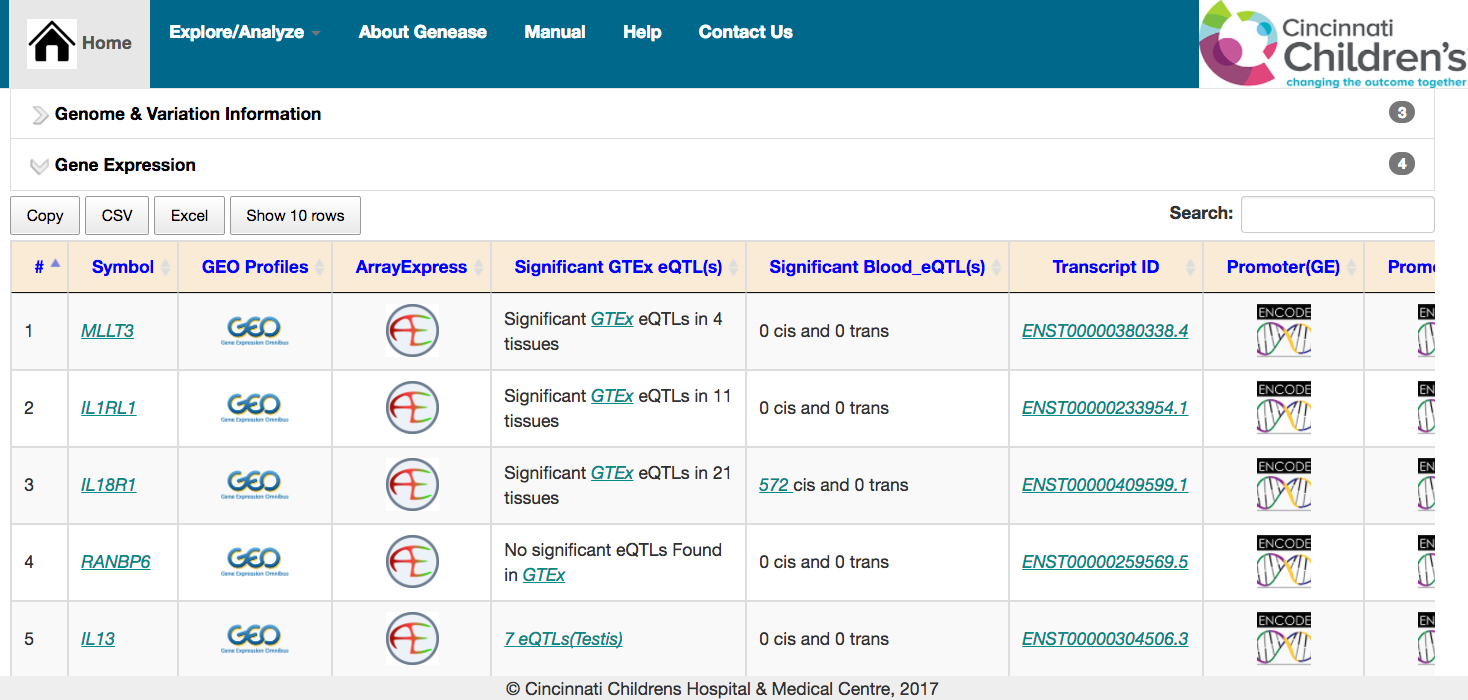
Result layout consists of “sliders” like in the single gene input case with each slider containing information in tables.



*Summarized information – gene list*

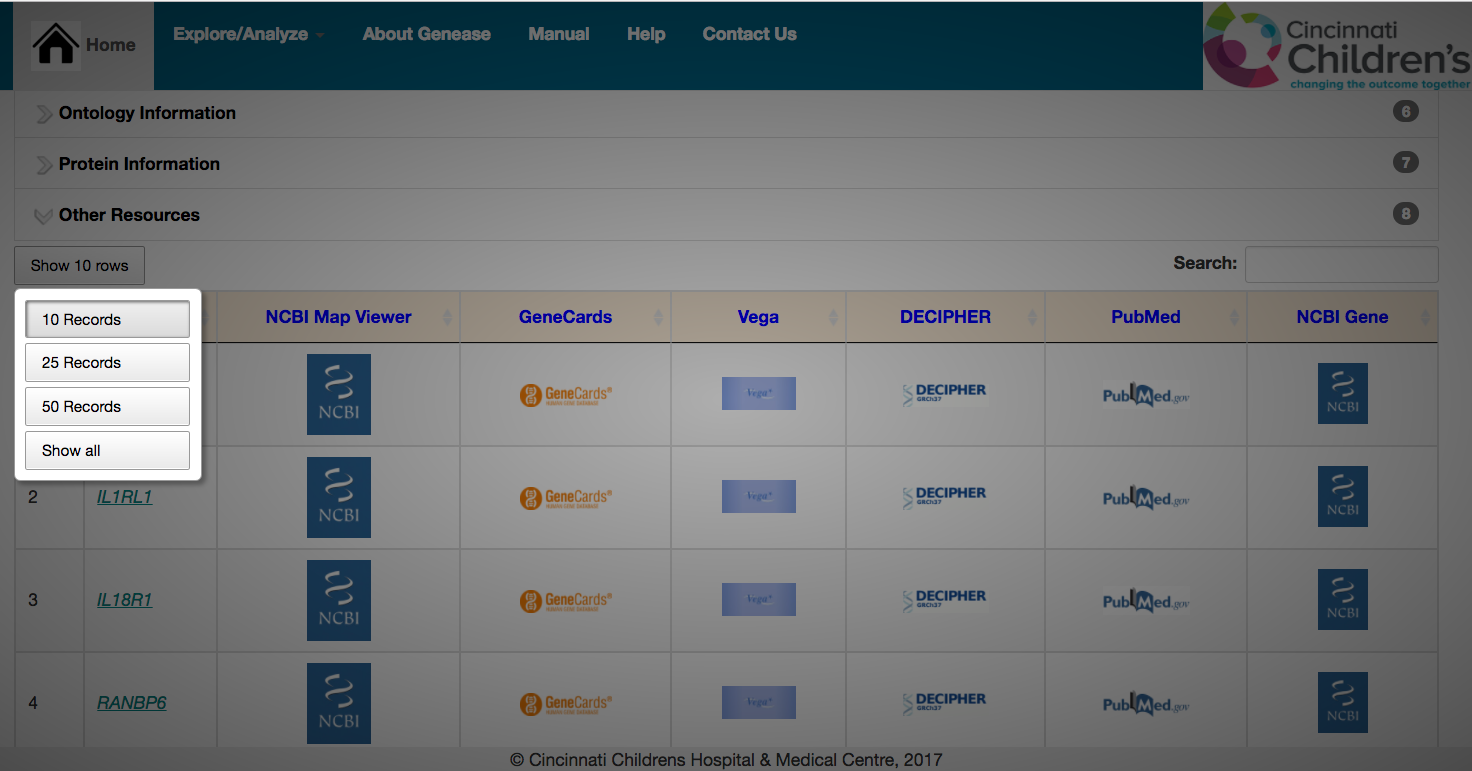
Any table containing “downloadable” (text) data can be downloaded using either “CSV” or “Excel” buttons just below the sliders.

For sliders having NO text data, these buttons would NOT be available.



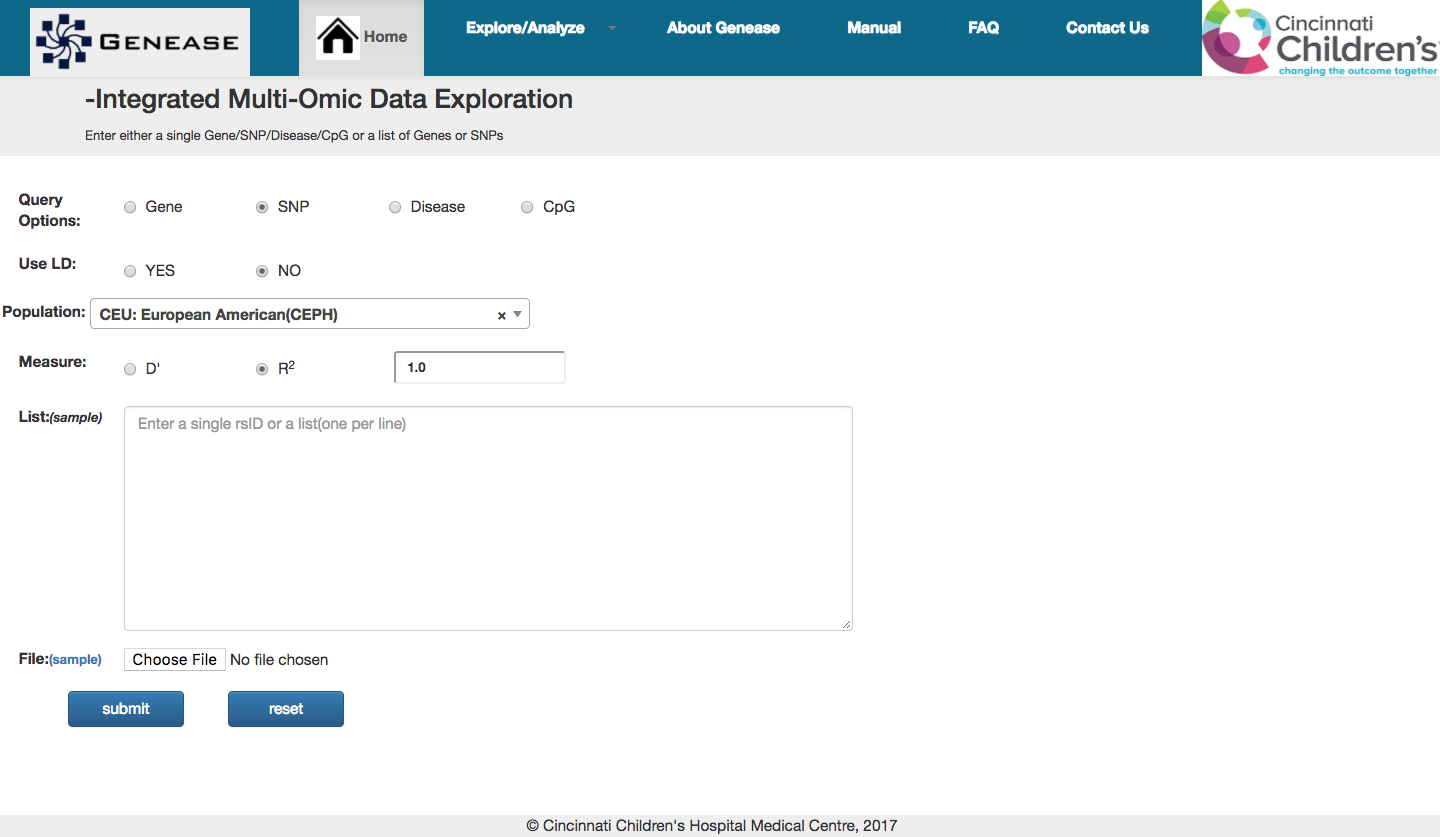
*Gene Expression Information – gene list*

Number of rows displayed in the table can be adjusted using “Show 10 rows” button.



*Gene Exploration slider – adjusting number of rows*

1. **SNP Query:**



*SNP explore query*

SNP query based on rsIDs result in gene information like genes associated with the SNP, description, chromosome, position, locus, consequence allele and iframes linked to various SNP related data resources like RegulomeDB, HaploReg, GWAVA, dbSNP, GTEx etc. A sample use-case of SNP exploration query could be:

*What is the functional class of the SNP - coding, non-coding? Example: rs1800234*

*Where is locus of rs1805017? And what gene(s) is it mapped to?*

*What are the significant eGenes associated with rs1800234?*

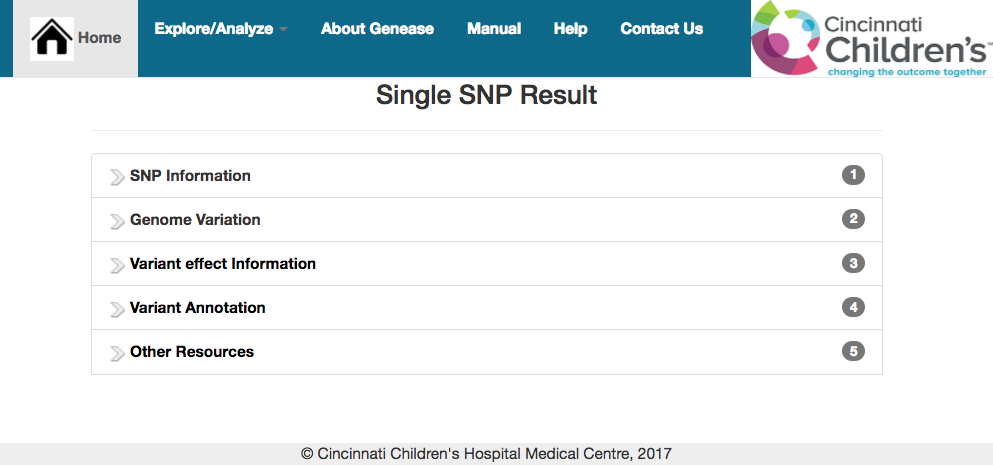
*What is the HaploReg score of rs1800234? Is it a functional SNP?*

*What are the proxy variants of rs1800234 in European ancestry?*

To explore any particular variant, select “SNP” as the query option and enter the rsID in the input text area. Click “Submit” button.

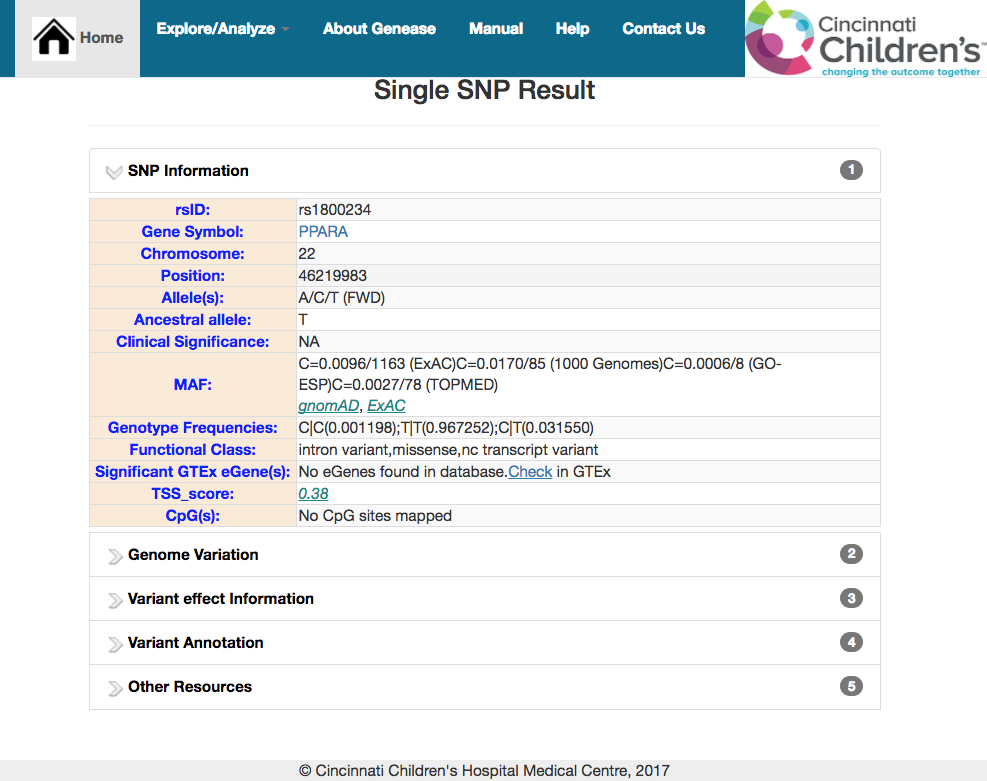
Additionally, “Use LD” option can be used to explore SNPs in linkage disequilibrium (proxy variants) for a given list of query SNPs. Source population, LD measures and strength of LD pattern can be chosen by the user to fetch the proxy variants.

The result screen of the non-LD query scenario, similar to gene exploration result, has sliders containing information about the SNP or link to another external database.



*Single SNP explore result*

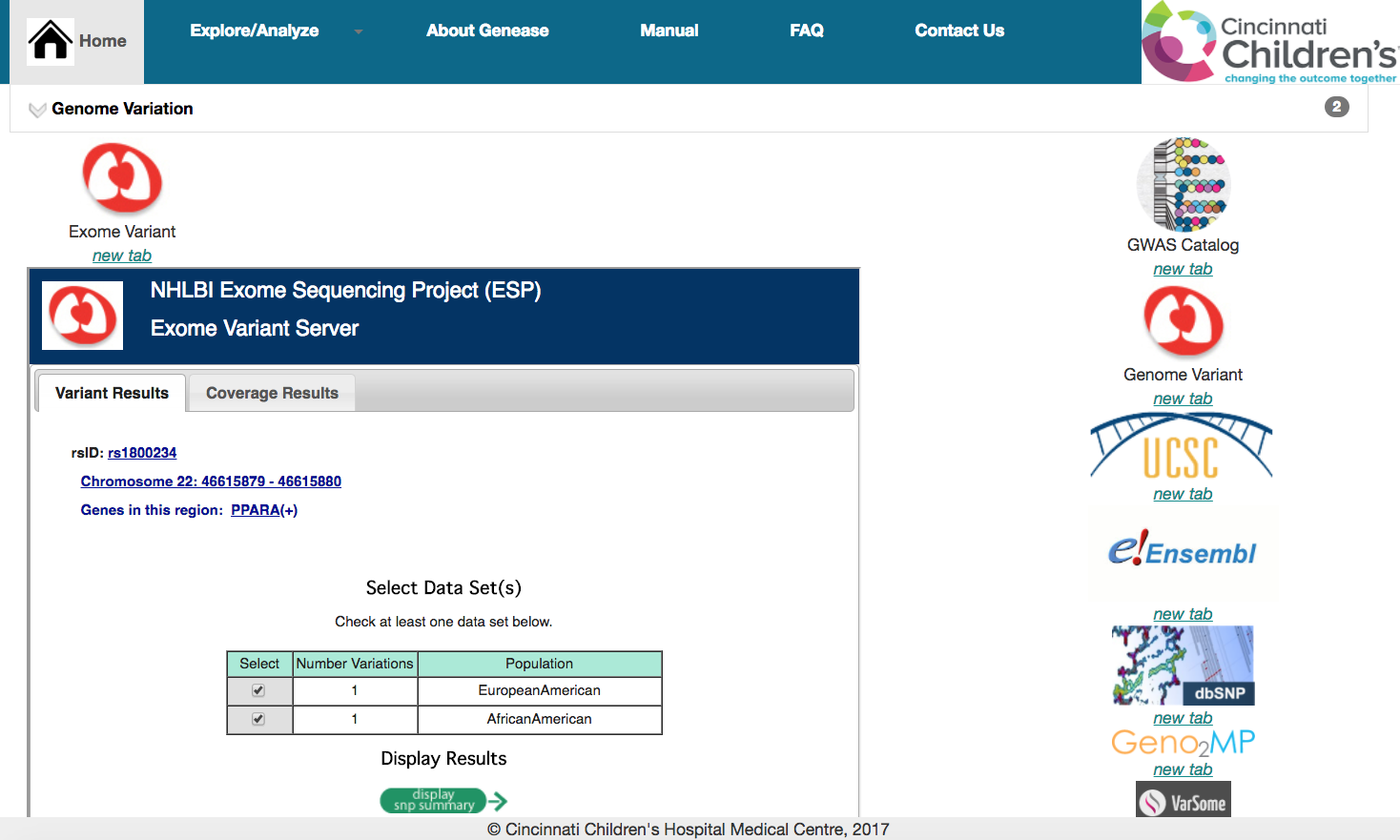
The “SNP Information” slider contains summarized information about the variant which includes position, genes mapped to, alleles and the functional class of the variant.



*Summarized SNP information*

In case of external links, a single-click on the logo will open the database in the same screen as an HTML iframe with the input appended automatically to the request.

To open the link in a separate tab click on the “new tab” link underneath the logo.

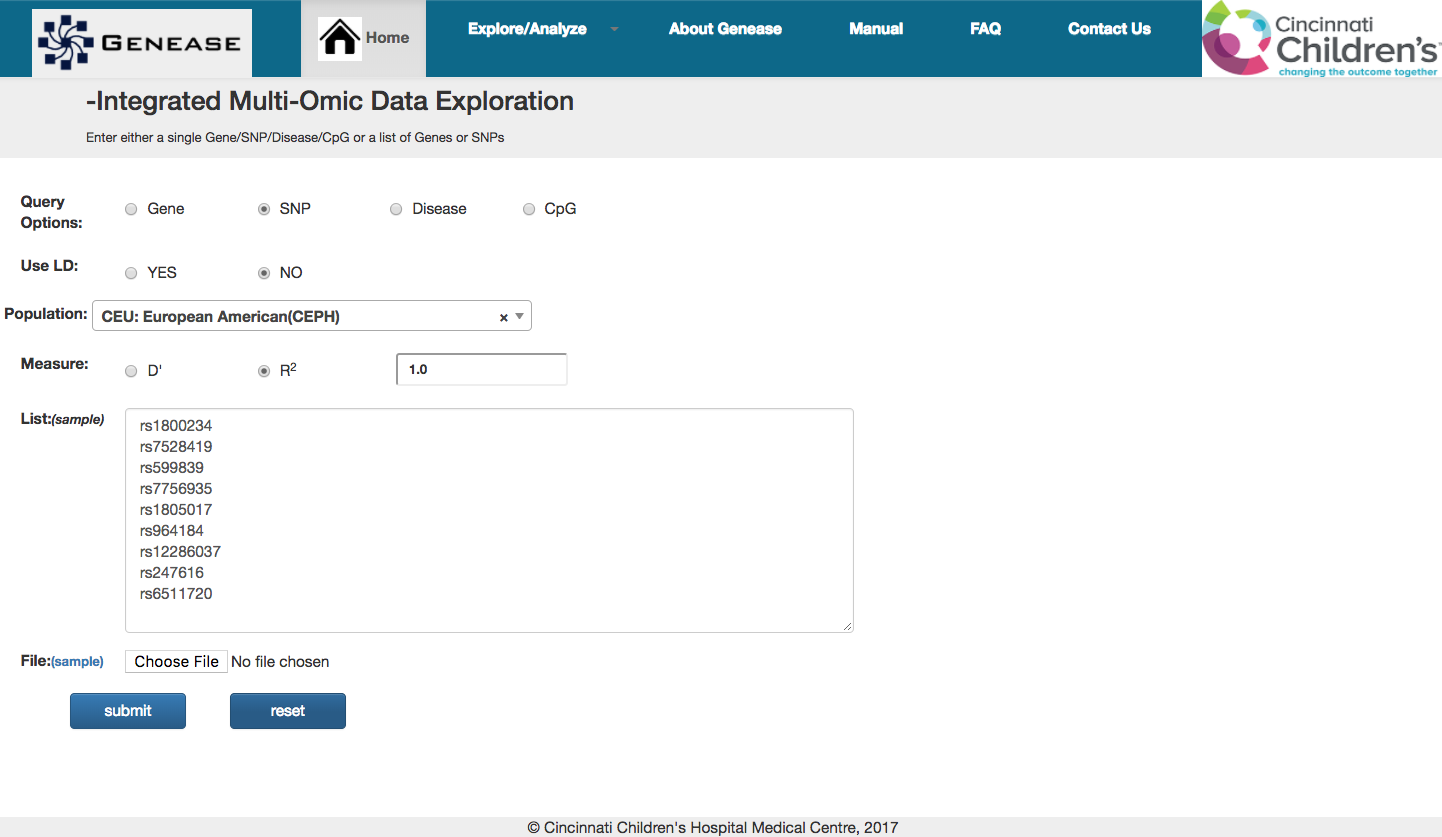


*Exploring “Exome Variant Server” database for rsID rs1800234*

Similar to the gene exploration case multiple external databases can be explored simultaneously by clicking on their respective logos.

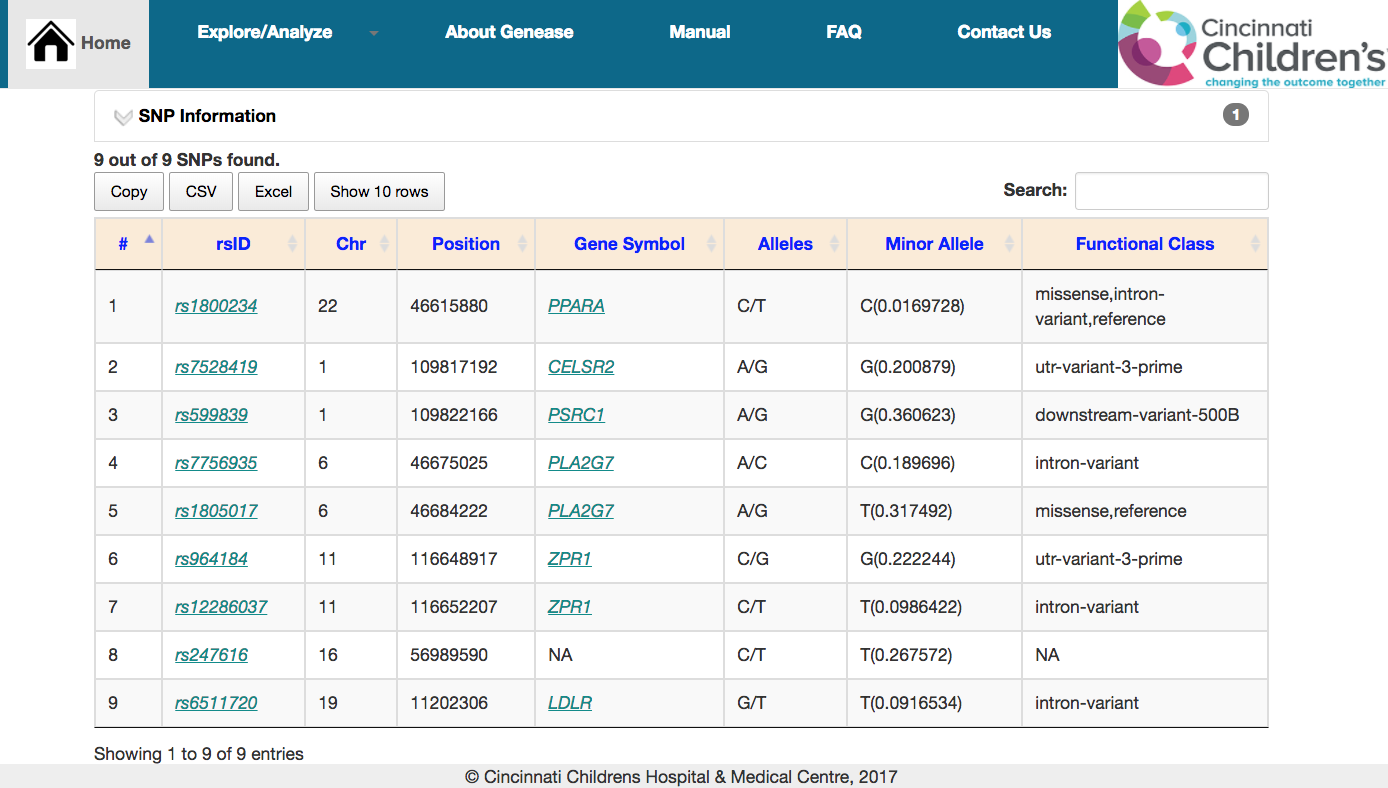
GENEASE can also be used to simultaneously explore a list of SNPs. Select “SNP” as the query option and enter the list of rsIDs in the “List” text-area or upload a file containing the same.

To understand the working of the tool, sample rsIDs can be used by clicking on “sample” link beside “List” or “File” input category.



*SNP list explore query*

The same information sliders as in “Single SNP exploration” procedure can be seen. Each slider contains information in a tabular format

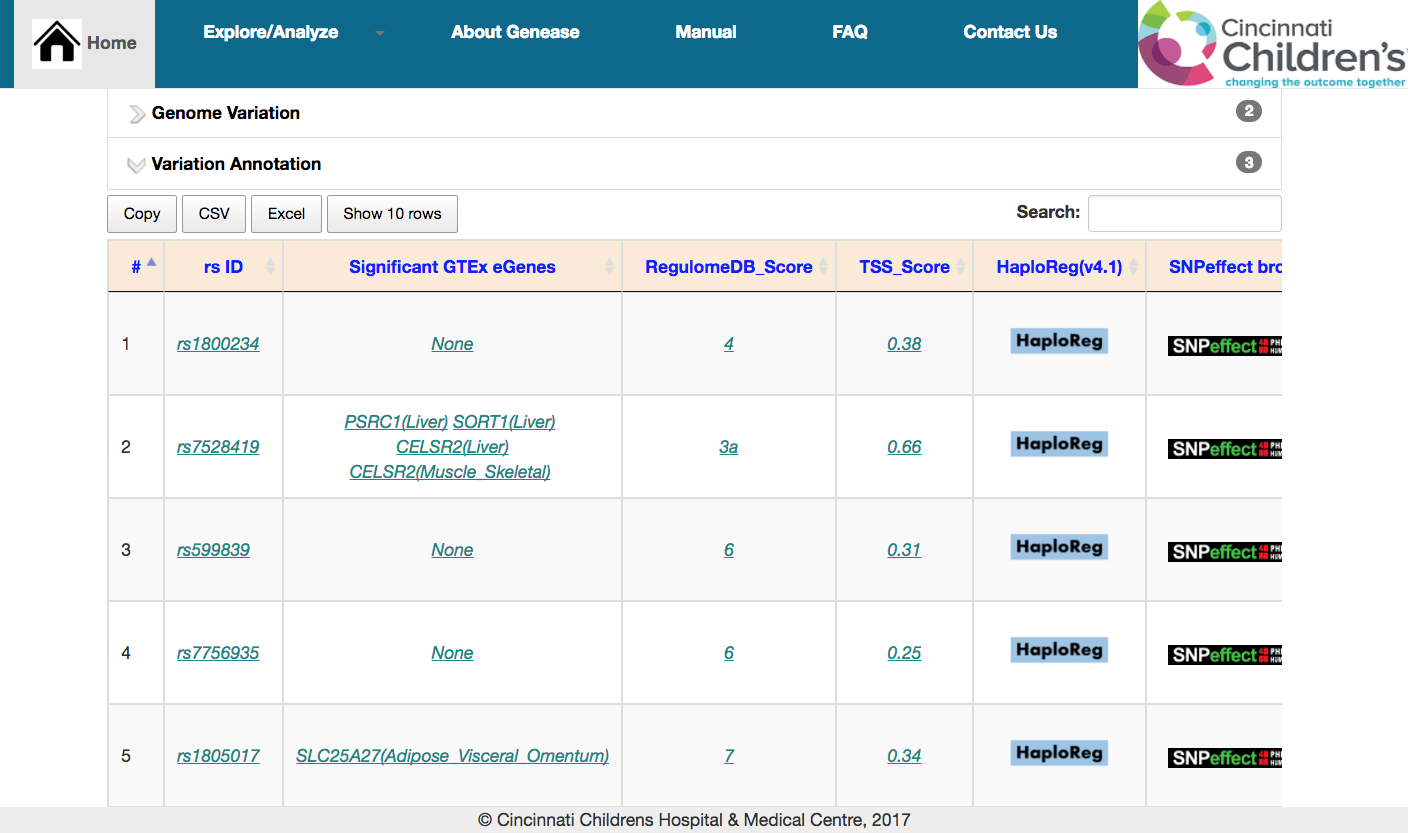


*Summarized SNP information – snp list*

Any table containing “downloadable” (text) data can be downloaded using either “CSV” or “Excel” buttons just below the sliders.

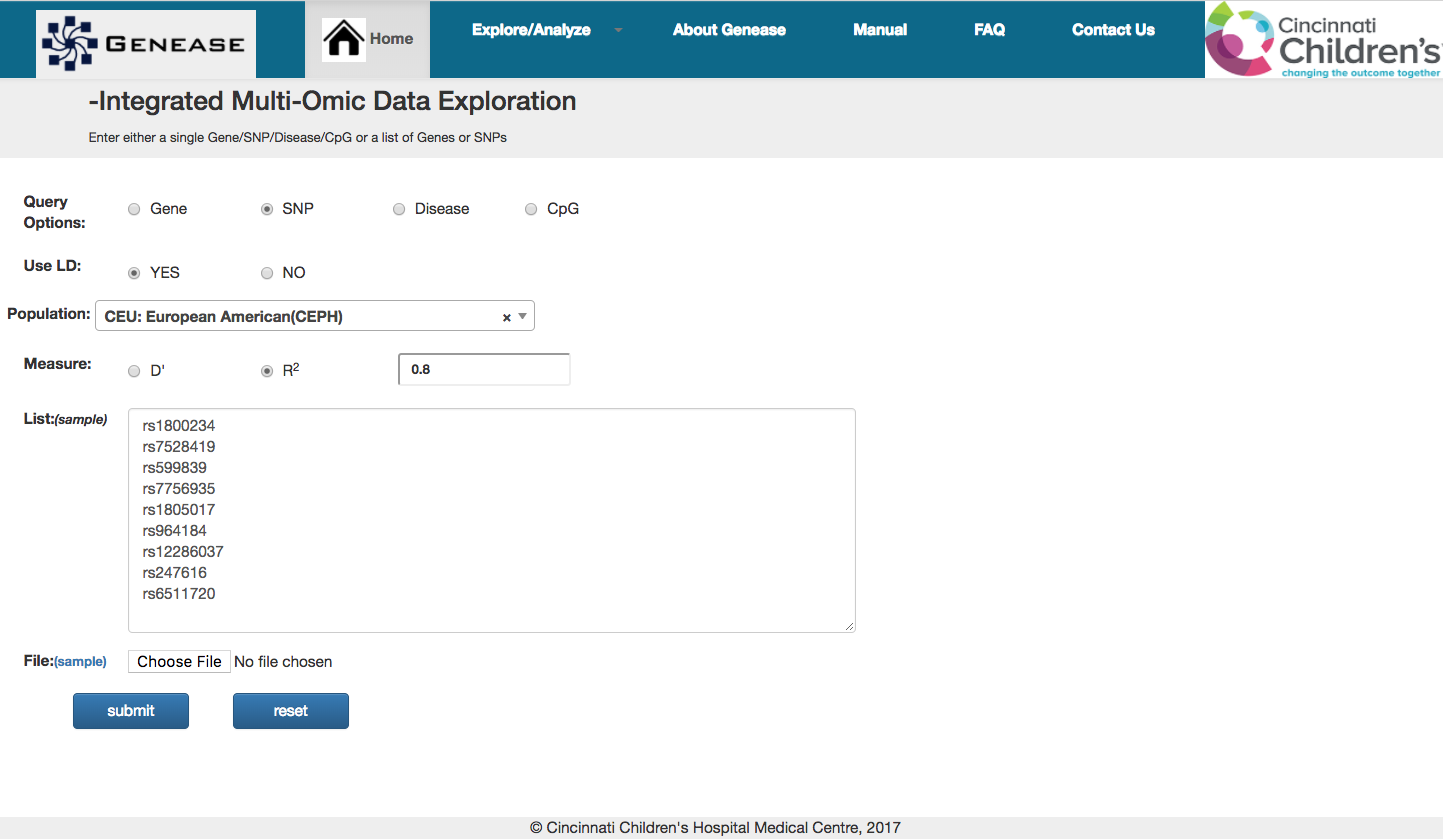
For sliders having NO text data, these buttons would NOT be available.

Number of rows displayed in the table can be adjusted using “Show 10 rows” button.



*Variant Information – SNP list*

1. **SNP****Query****(LD):**

**

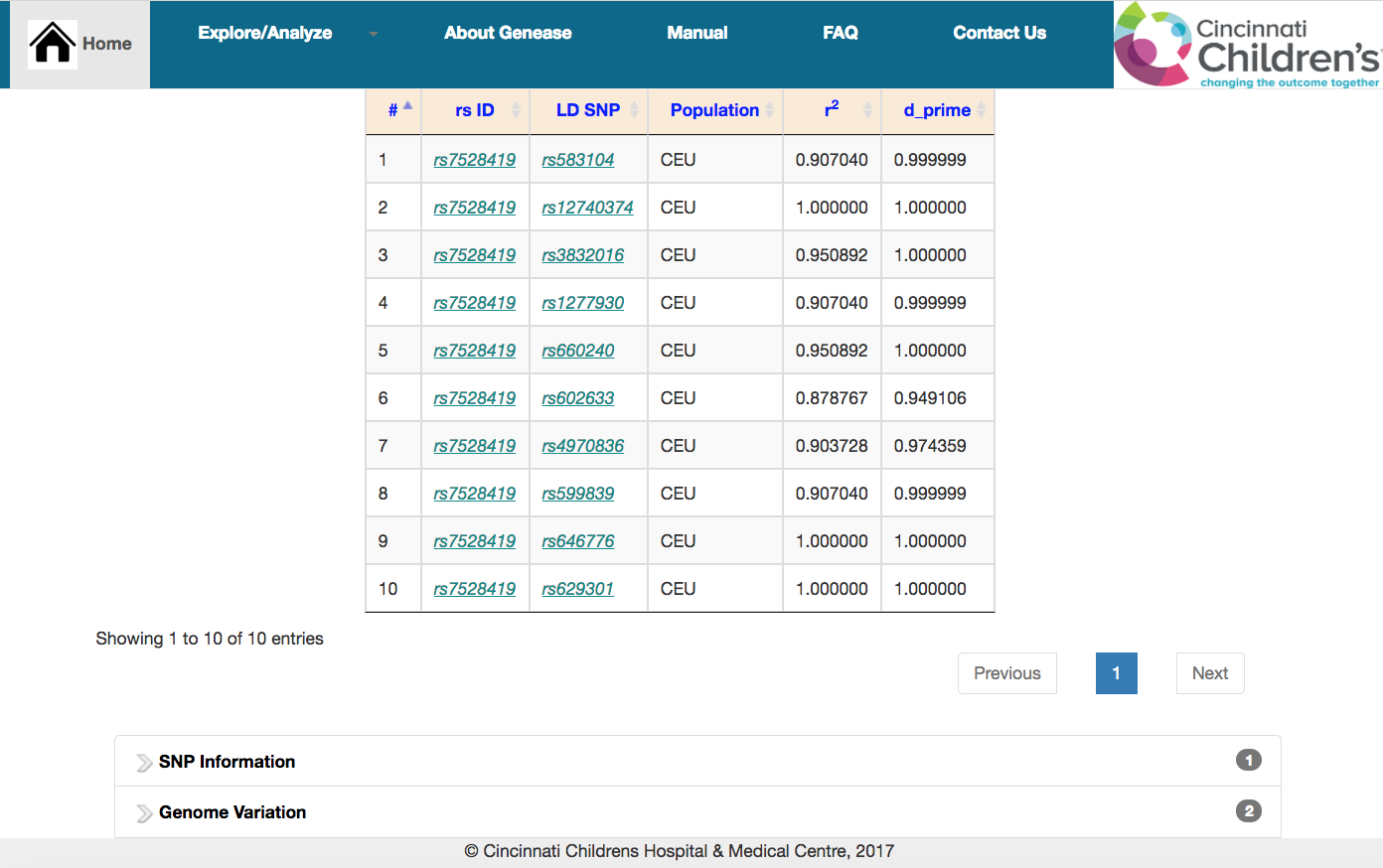
*SNP explore query for proxy variants*

“**Use LD**” option can be used to explore proxy SNPs (LD) for a given list of query SNPs.

“**Population**” dropdown can be used to select the ancestry of choice. All the populations in 1000 genomes phase 3 are supported.

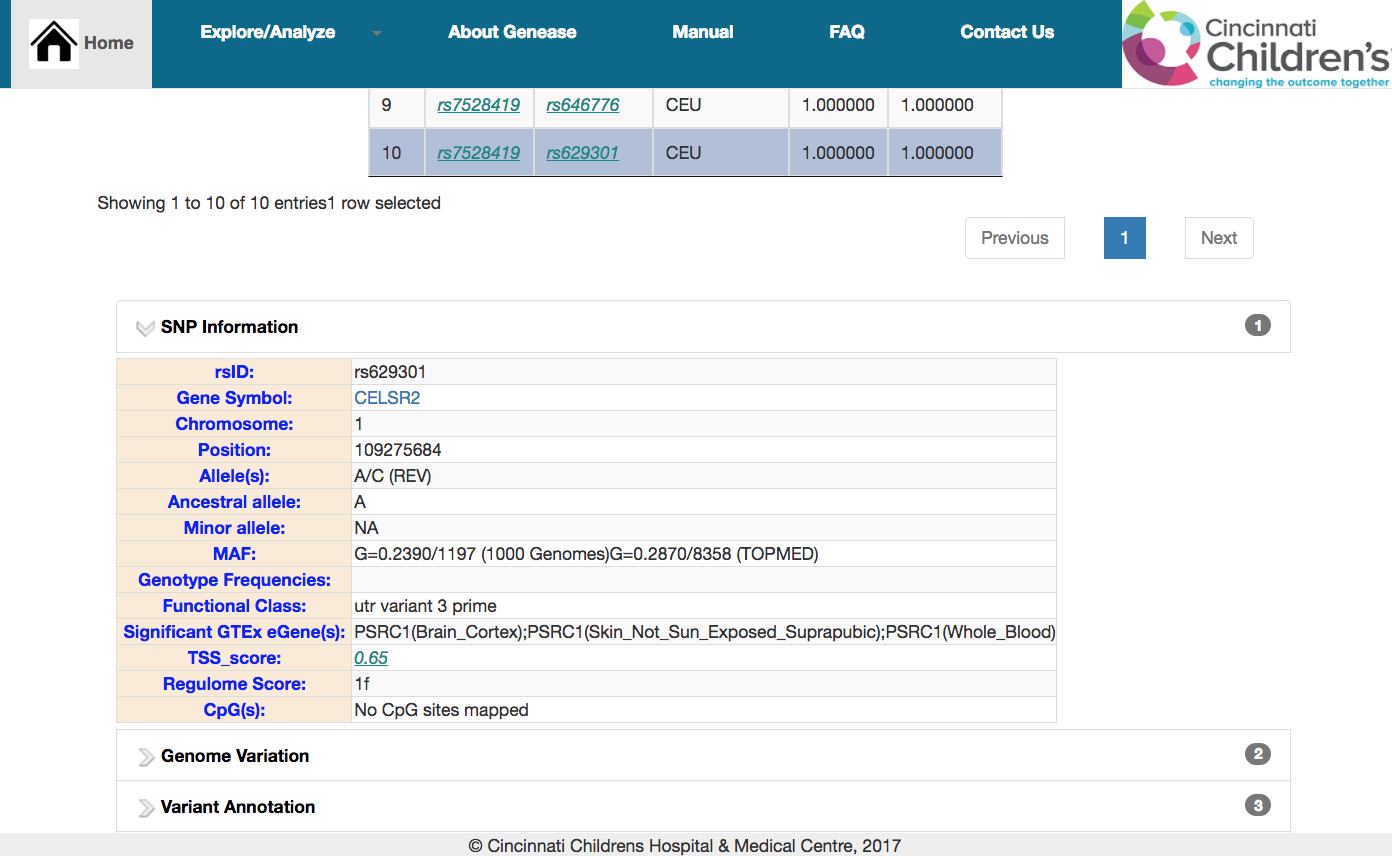
Additionally, LD measures used to filter the SNPs can also selected and the strength of LD pattern can be entered as a free text in the text field.

The result screen of the LD query scenario, firstly has an HTML downloadable table listing the query SNPs and the proxy variants for each of the query SNP



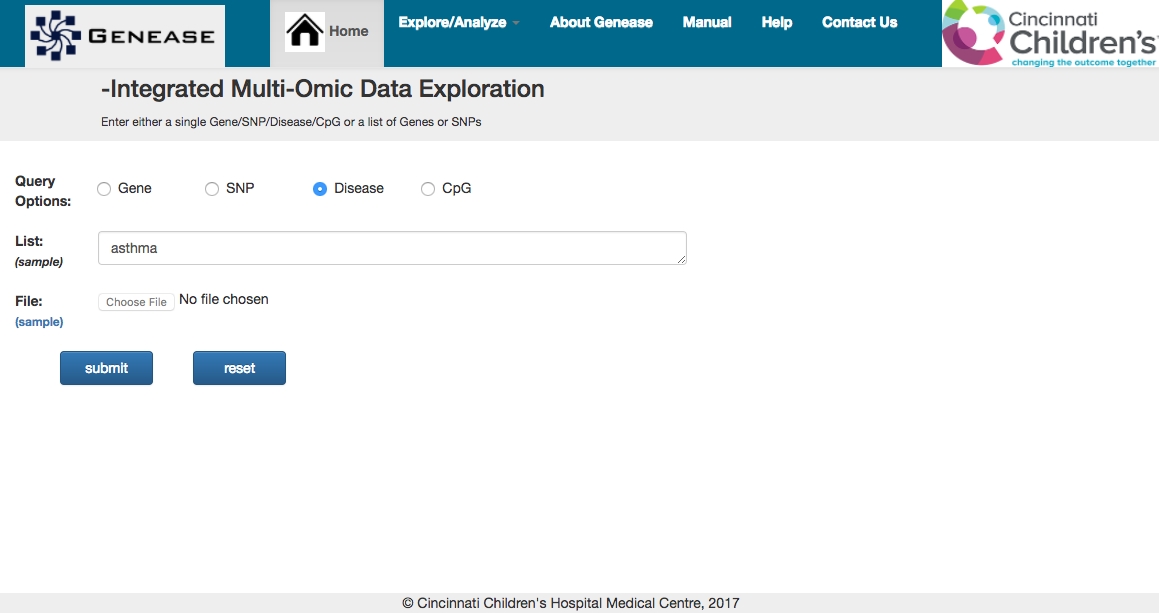
*LD SNP explore result with list of proxy variants for rs7528419*

Any proxy variant can be further explored by either clicking on the rsID (in a new tab) or by selecting the corresponding row (in the same screen).



*In page exploration of proxy variant in LD SNP explore result.*

1. **Disease Query**



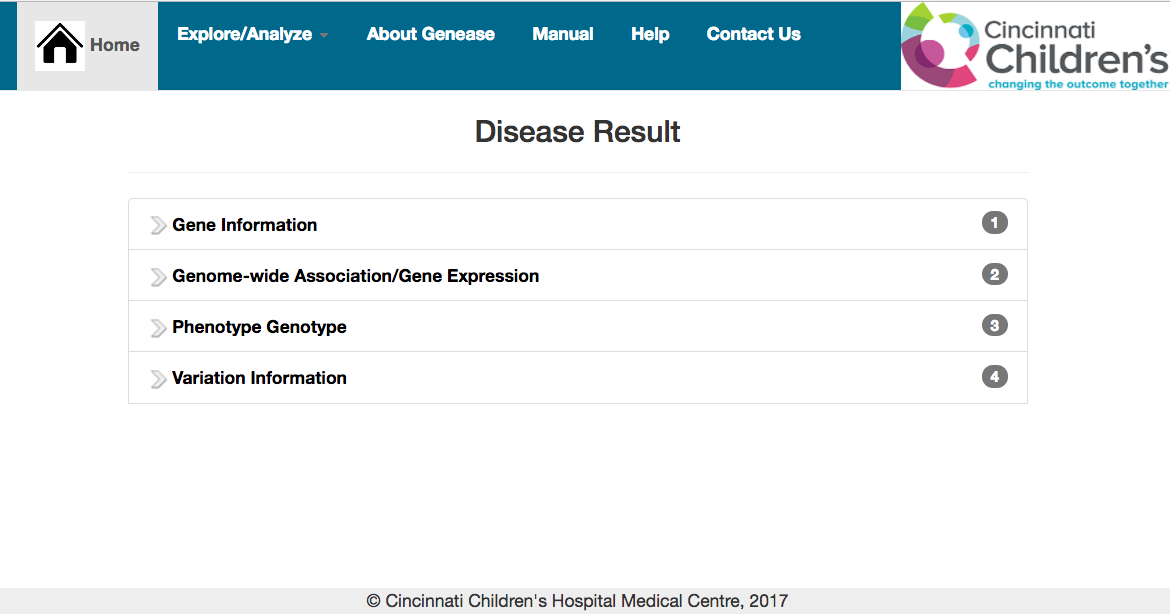
*Disease explore query*

Disease query output provides gene entries associated with a disease. External mutation-phenotype databases like SwissVar, NCBI OMIM, GWAS Catalog, dbGaP etc. are embedded into the output including the literature for the specific disease. A sample use-case could be:

*What are the known genes associated with the disease (example: asthma) and the gene status?*

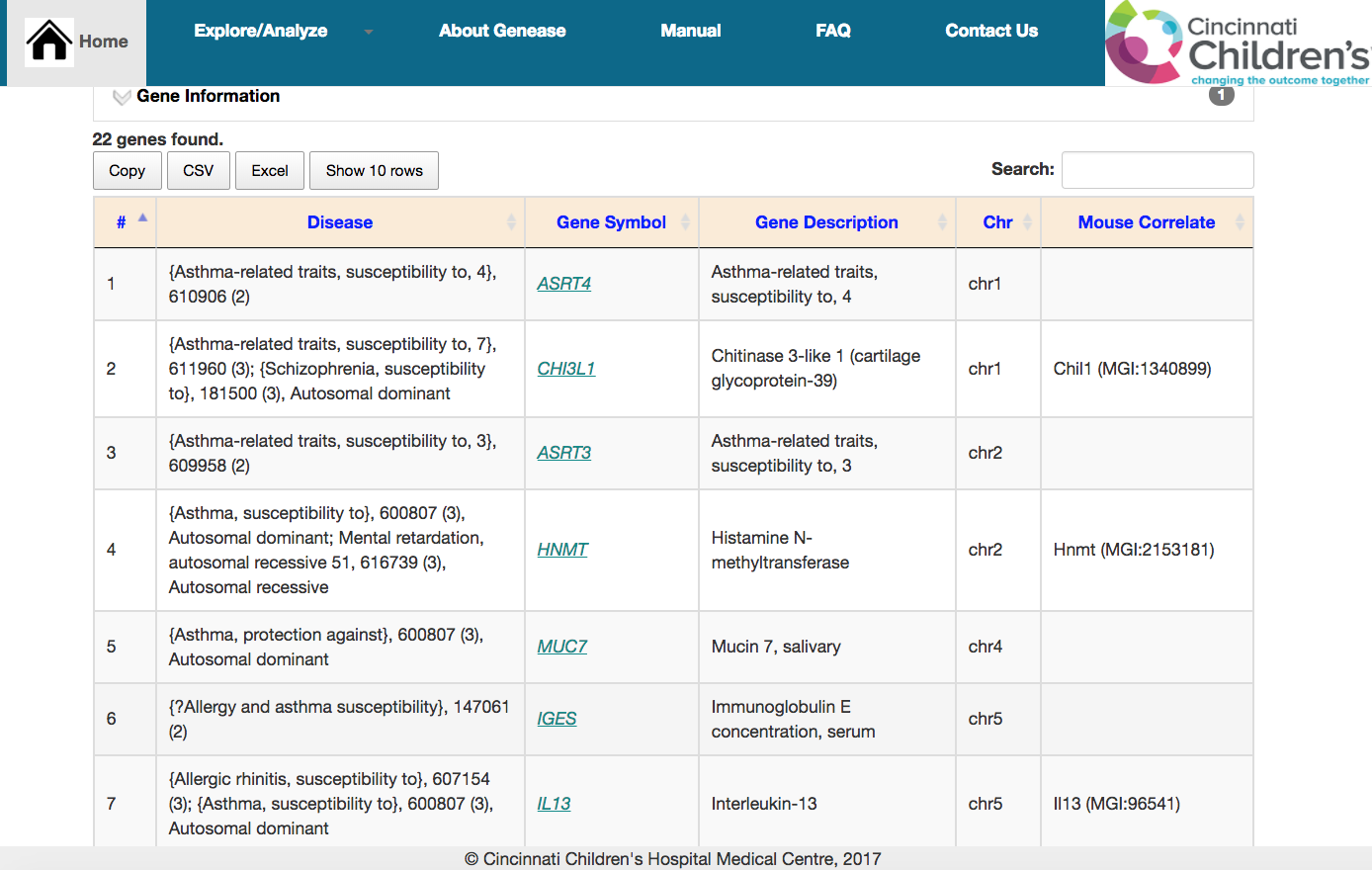
*What variants associated with ‘asthma’ through GWA studies? What gene expression datasets are available related to disease?*

The output format contains sliders similarity to Gene and SNP exploration case.



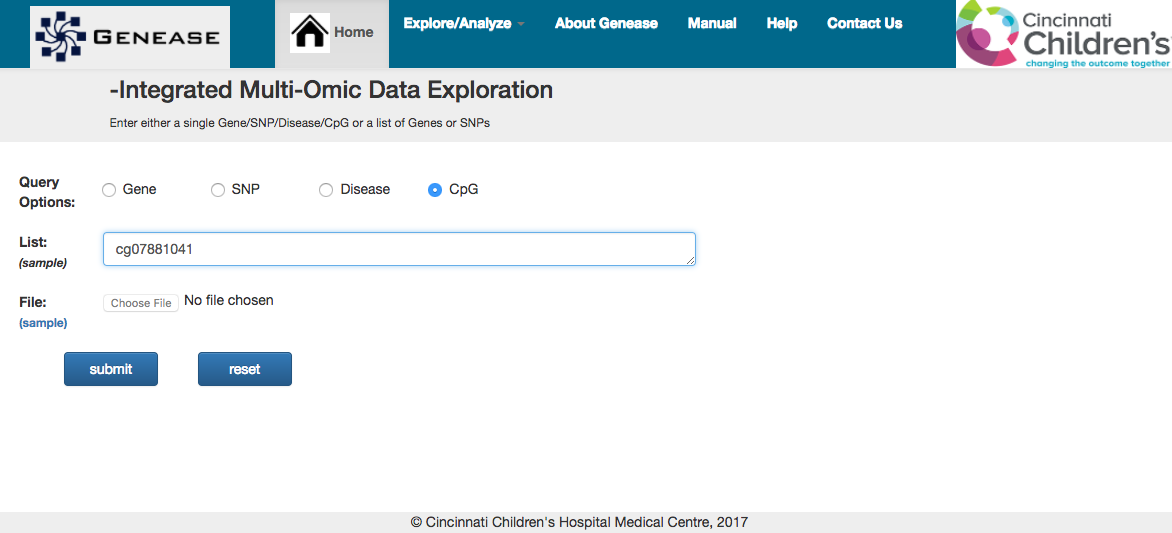
*Disease explore result*

The “Gene Information” table consists of summarized gene information mapped to the input disease retrieved from OMIM.



*Summarized information of genes mapped to asthma*

Other sliders contain links to external links having information about disease input.

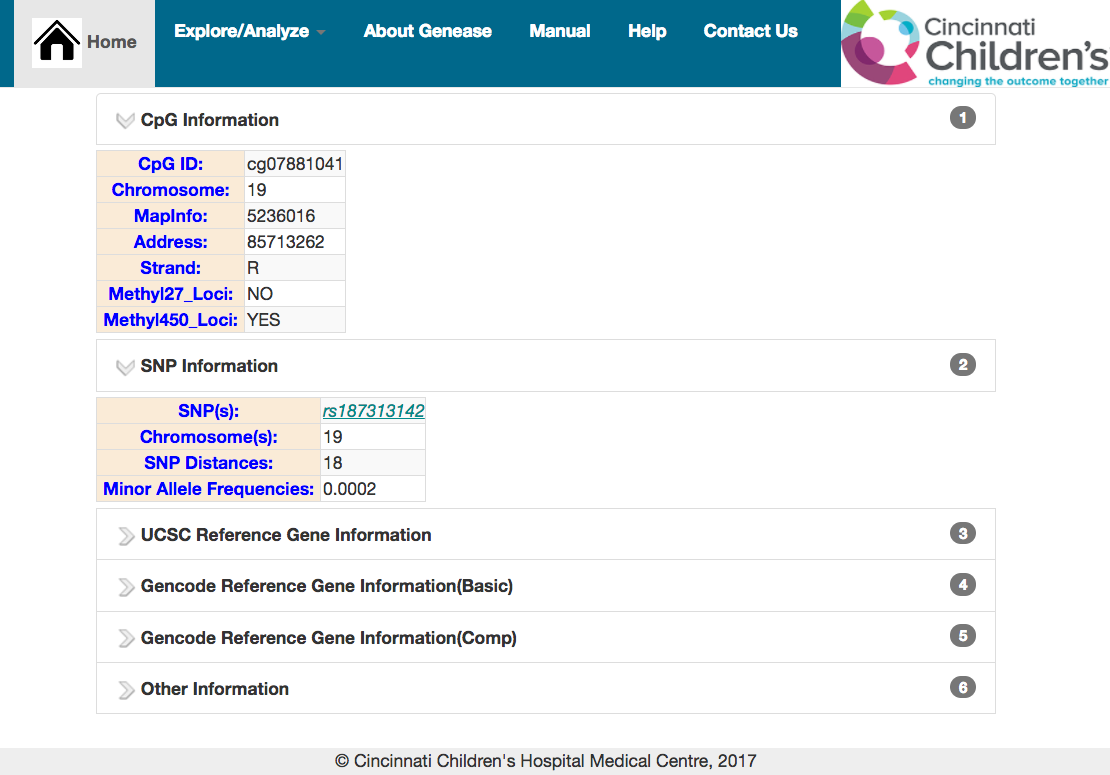
1. **CpG site query:**

*CpG site query*

CpG site query displays basic locus information about methylation site. The result also contains information about SNPs, genes mapped to them and evidence counts in tabular format.

“CpG Information” slider contains basic information about the methylation site including the position, chromosome, strand etc.

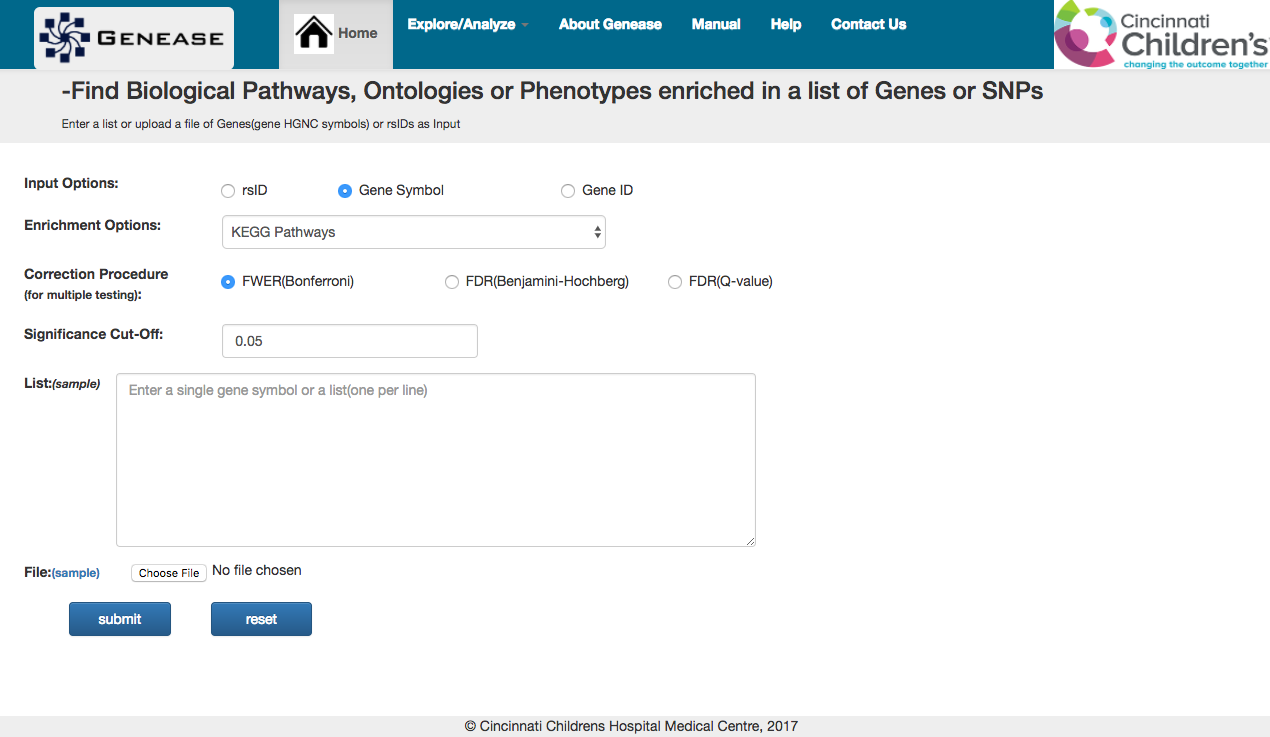
“SNP Information” slider contains information about variants mapped to the methylation site.



*Methylation site – summarized and SNP information*

The next 3 sliders contains information about genes mapped to the methylation sites belonging to two different genome sources.

**Enrichment Analysis:**

GENEASE Enrichment analysis query based on HGNC symbols, Entrez IDs or rsIDs result in a table displaying the mapped function and the p-value of the statistical test for enrichment/depletion. Optional bar plots indicating the top enriched and depleted terms are also included.

*GENEASE enrichment analysis screen*

A sample use-case for enrichment analysis scenario could be:

*What are the biological pathways mapped in the list of genes?*

*Which pathway functions are enriched? Which is the most significant one?*

*What are the genes from the list that contributed to the enrichment?*

Input Options” radio box indicates the type of inputs used. “Enrichment Options” dropdown menu can be used to select the annotation types to be tested for enrichment.

Users can also select the multiple correction procedure using the “Correction Procedure” checkboxes and the significant cut-off to be used can be entered in the text-box provided.

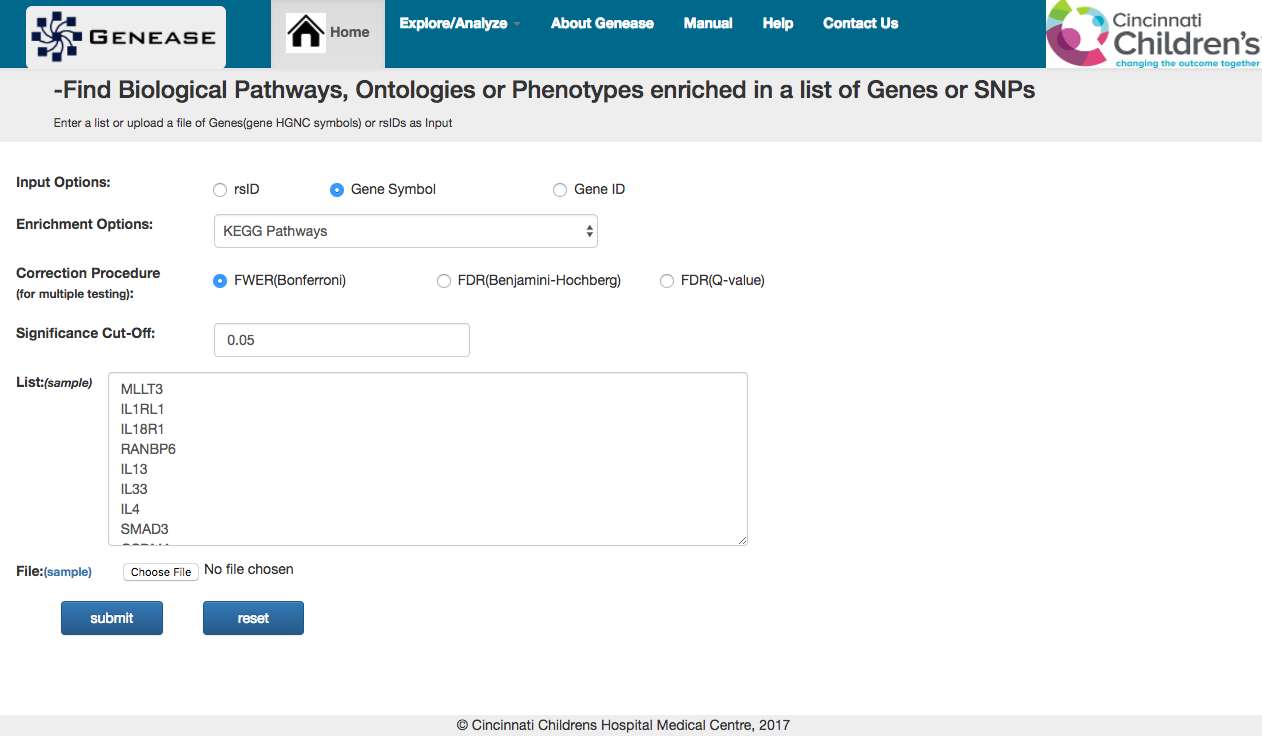
Input gene/SNP can be entered using “List” text-area or upload a file containing them.

“Sample” links provided both input types allows the users to use a sample list of inputs.

1. **Pathway enrichment analysis (genes)**

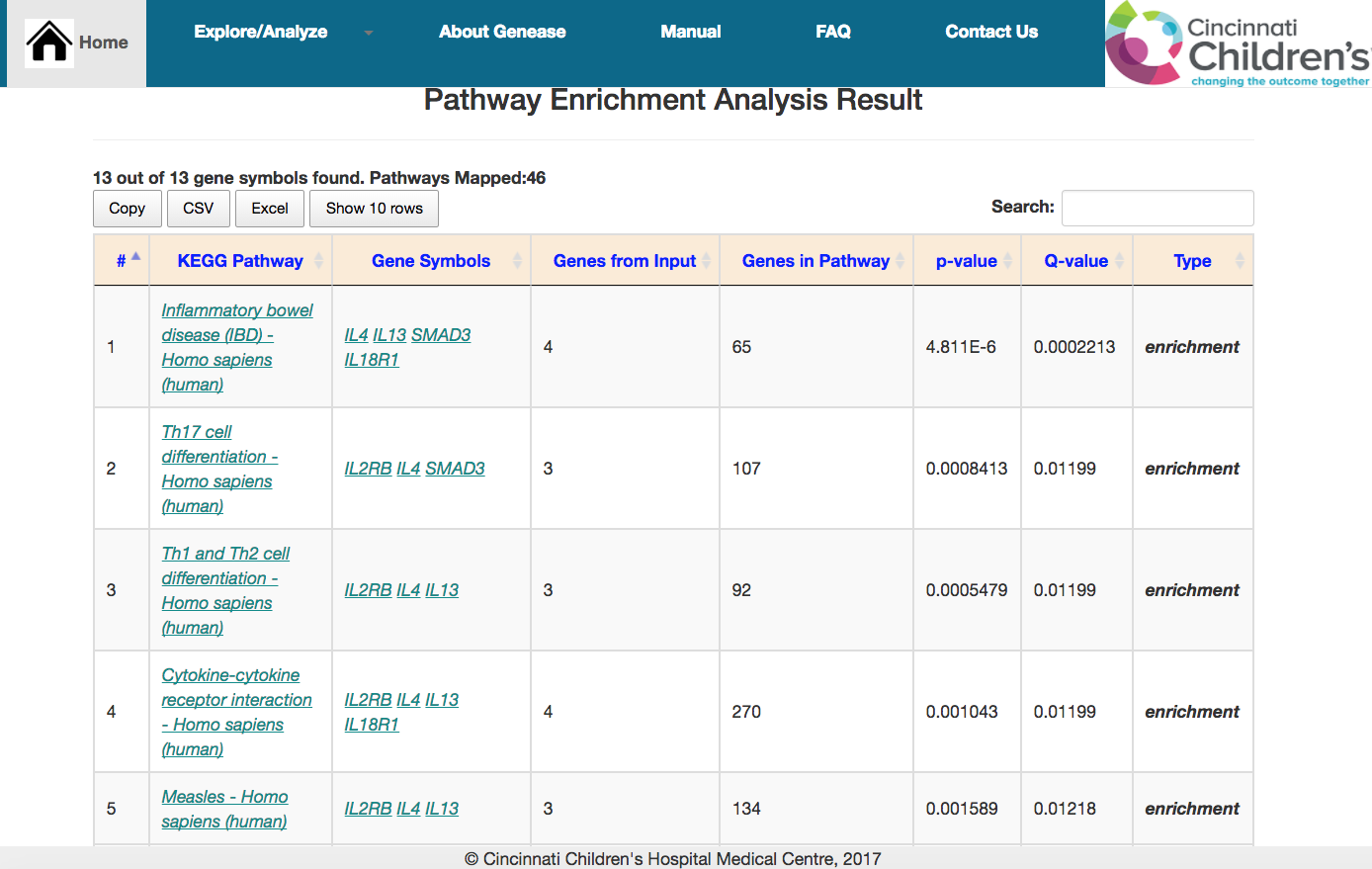
To perform KEGG pathway enrichment of a candidate gene list, select “Gene Symbol” (default) as the input option and “KEGG Pathways” as enrichment option.

Input gene symbols can be copied into the text-area or uploaded in a file.



*Pathway enrichment analysis query – gene list*

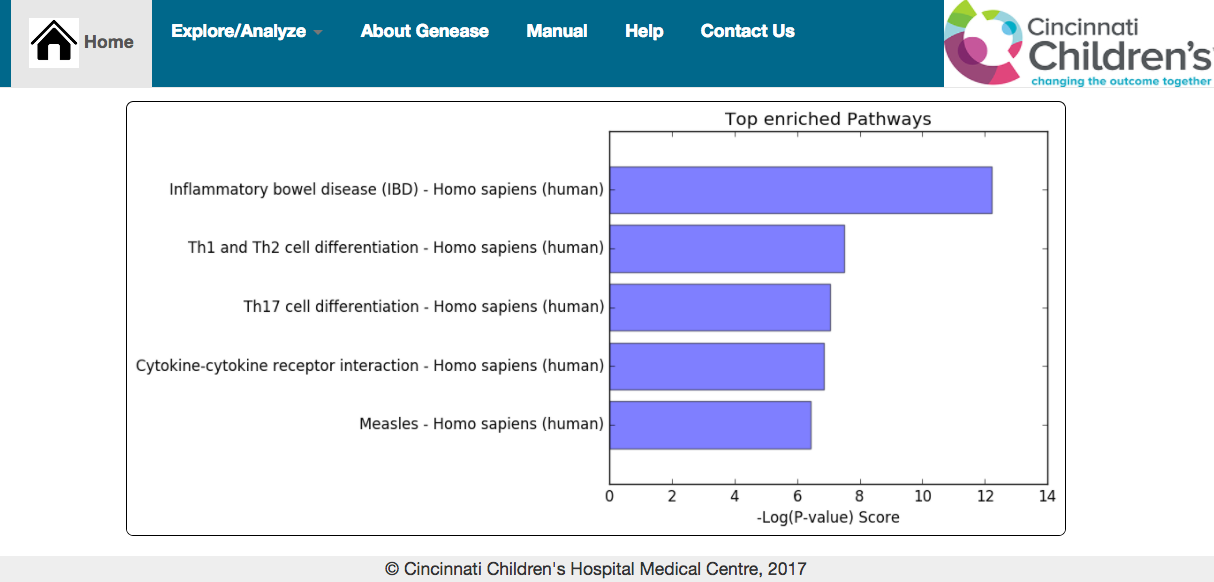
The enrichment result displays two optional bar plots which displays the top 5 enriched and depleted terms. Plots are not displayed if there are no enriched or depleted terms and a message is displayed in the result screen.

Result also contains a HTML sortable table which lists the pathways, genes from the sample that belong to the pathway and the corresponding counts.

*Pathway enrichment analysis result – gene list*

This table can be downloaded by clicking on the “CSV” or “Excel” buttons just above the table. Also the number of rows being displayed can be changed using the “Show X rows” button.

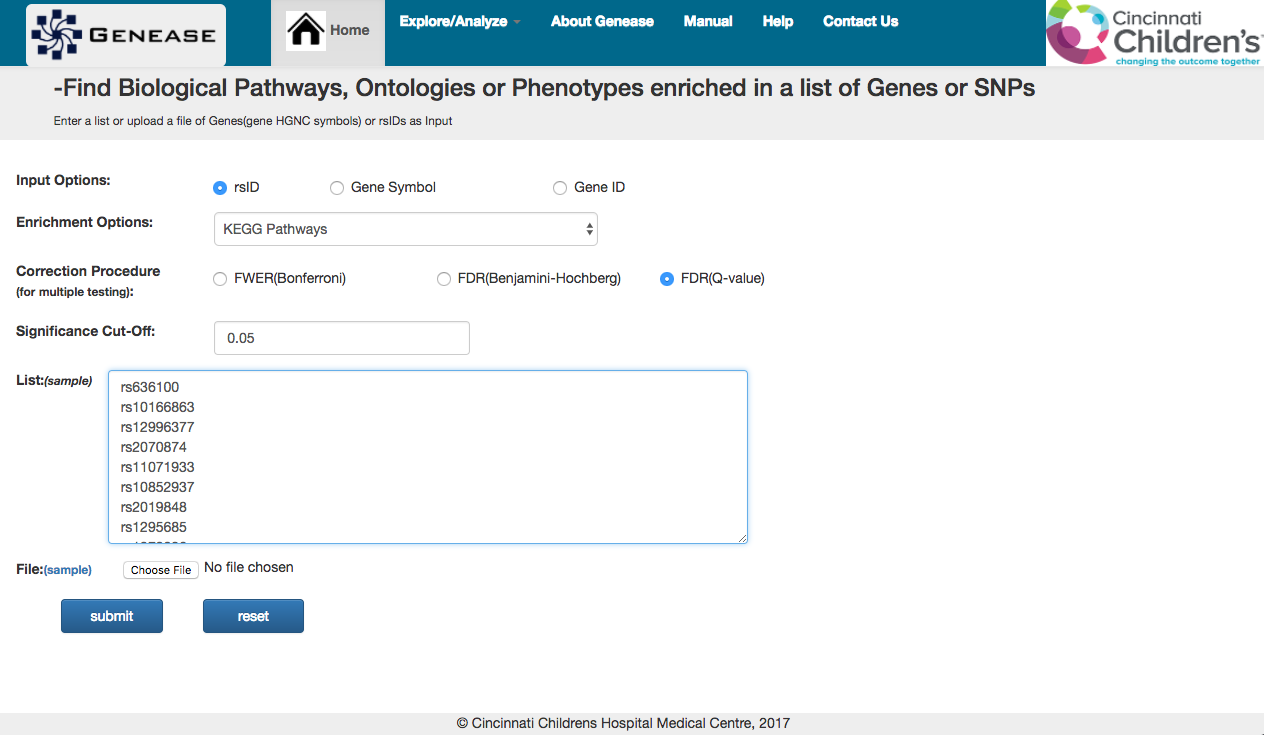
Counts of gene symbols identified and pathways mapped are listed in the “summary” section just above the result table including any gene symbols that were not identified.

Clicking on a gene symbol from the list in the output table re-directs to single gene exploration result for the gene.

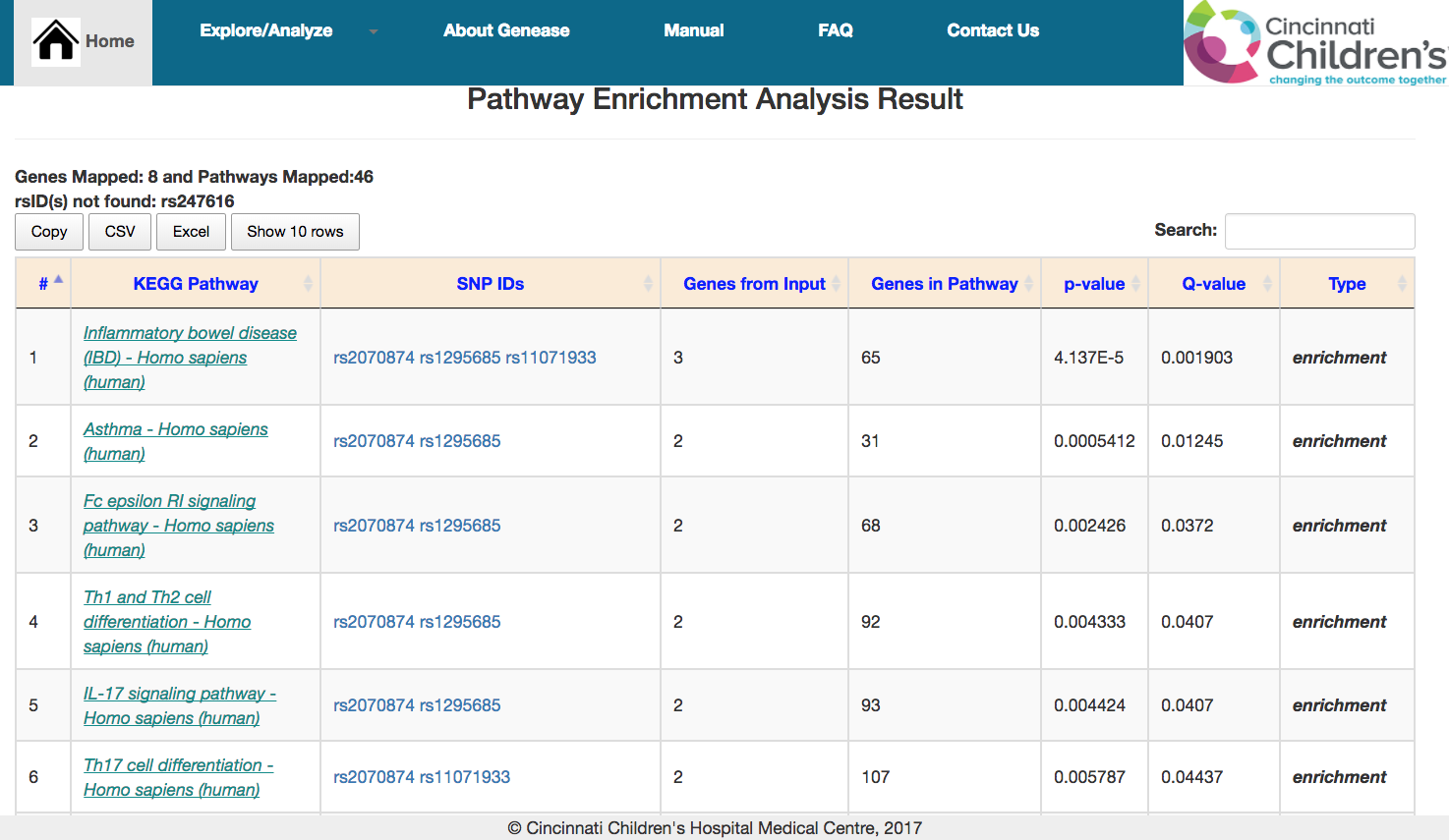
*Pathway enrichment analysis result – gene list*

Entrez gene IDs can also be used in the analysis. “Gene ID” should be selected as the input option instead. The remaining process remains the same as above.

1. **Pathway enrichment analysis (SNPs)**

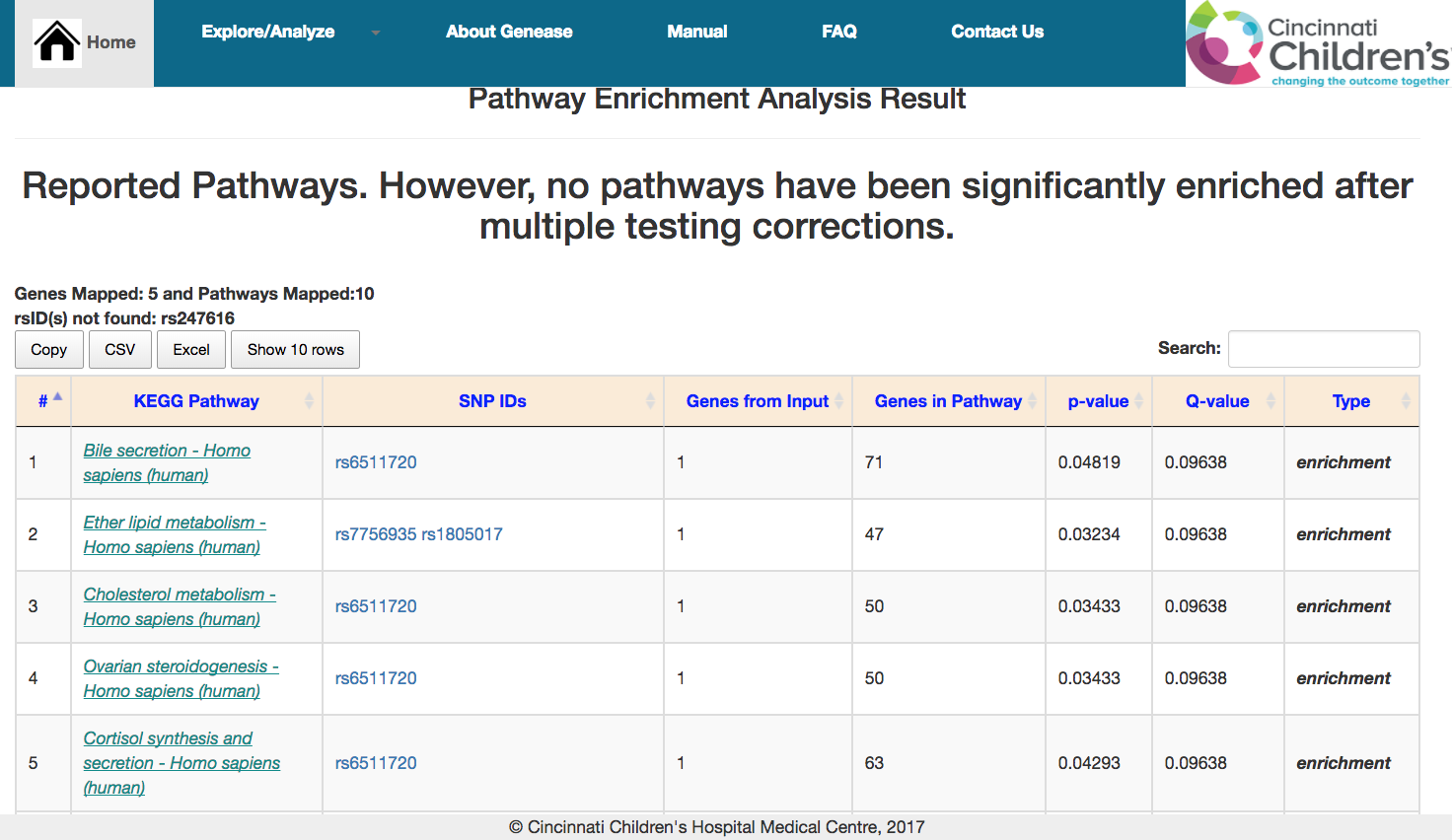
Firstly, select “rsID” as the input option and KEGG Pathways as enrichment options. Input rsIDs can be entered into the text-area or uploaded in a file.

*Pathway enrichment analysis query – SNP list*

The result interface is similar to pathway enrichment of genes shown above. In this case the HTML table displays the SNPs from the input list that belong to the pathway term via the mapped genes.

*Pathway enrichment analysis result – SNP list*

Counts of mapped genes, mapped pathways are listed in the “summary” section above the table.

In the scenario where no enrichment was observed, all the mapped pathways are displayed instead and an appropriate message is shown.

*Pathway enrichment analysis result where no enrichment of pathways was observed.*

***Note: The analysis procedures followed for other enrichment types (gene ontologies and human phenotype ontologies) are the same as above.***

**Overlap analysis**

GENEASE overlap analysis result based on two or more sets of HGNC gene symbols, Entrez IDs or rsIDs displays a table containing the counts of unique elements in each set and the calculated similarity metrics between the sets. A sample use-case of overlap analysis could be:

*How similar are the two gene lists? What is the level of similarity?*

*What are the enriched phenotypes shared by them?*

*What is the level of inclusion between the lists?*

If either “KEGG Pathways”, “Biological Processes” or “Phenotypes” are selected as overlap criteria, then enrichment analysis is involved.

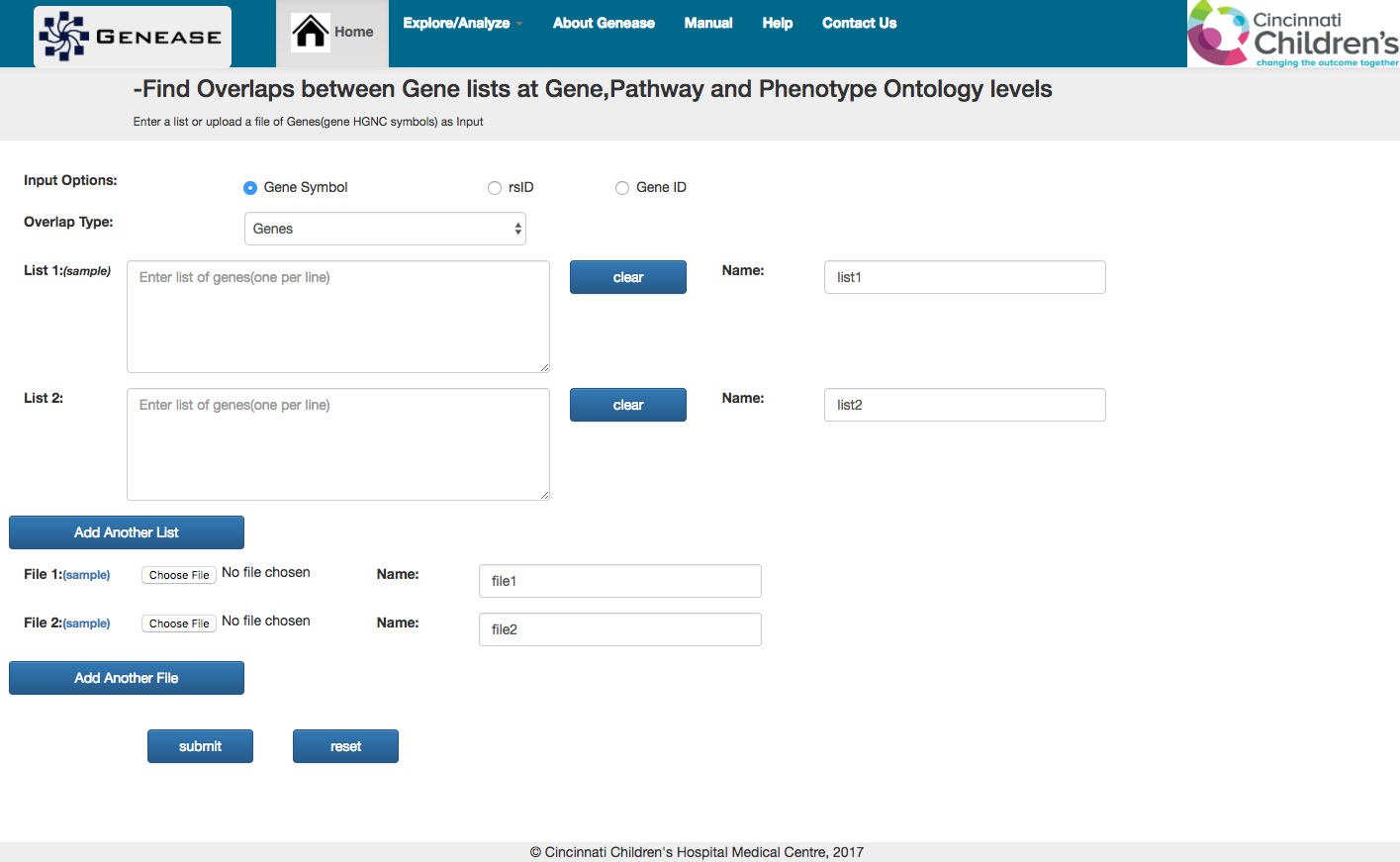
In this case users can also select the multiple correction procedure using the “Correction Procedure” checkboxes and the significant cut-off to be used can be entered in the text-box provided.

Input gene/SNP can be entered using “List” text-area or upload a file containing them.“Sample” links provided both input types allows the users to use a sample list of inputs.

To find overlap among three lists, “Add Another List” or “Add Another File” button can be used to either add a new list or upload a new file.

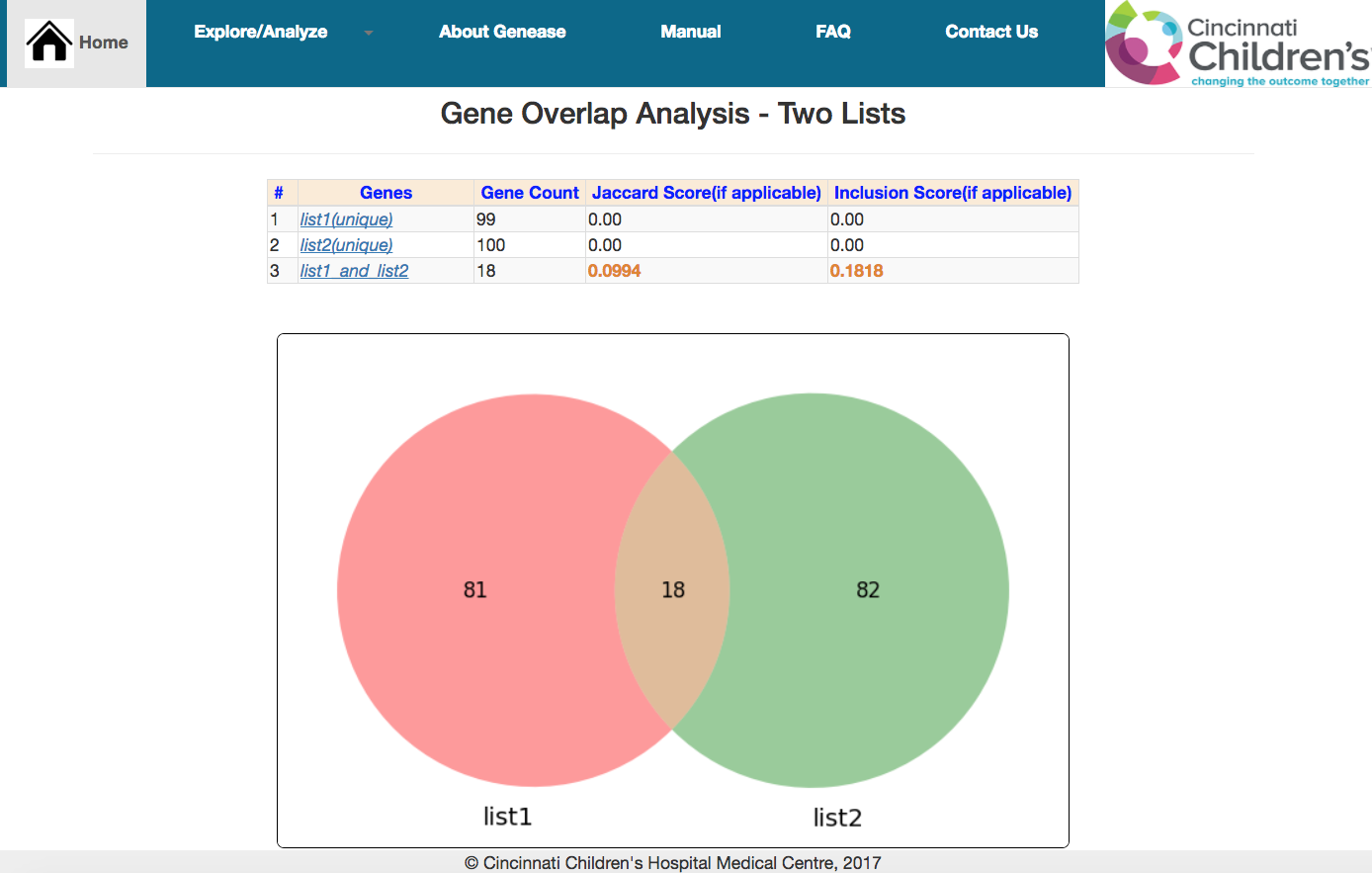
“Name” text boxes can be used to supply customized names to the lists entered. These names would be reflected in the plot shown in the result interface.

1. **Gene overlap (direct):**

To find “simple” overlap between two gene lists, select either Gene Symbol(default) or “Gene ID” as the input options and “Genes” as the overlap type.

*Direct gene overlap query*

Either copy the the gene lists in the text-areas or upload the files and click “Submit”

The result interface displays a Venn diagram plot showing the overlap levels between the lists.

*Direct gene overlap result*

Result screen also contains a table in which the filenames are listed in the first row. The table also contains an entry which indicates the “intersection” of each list.

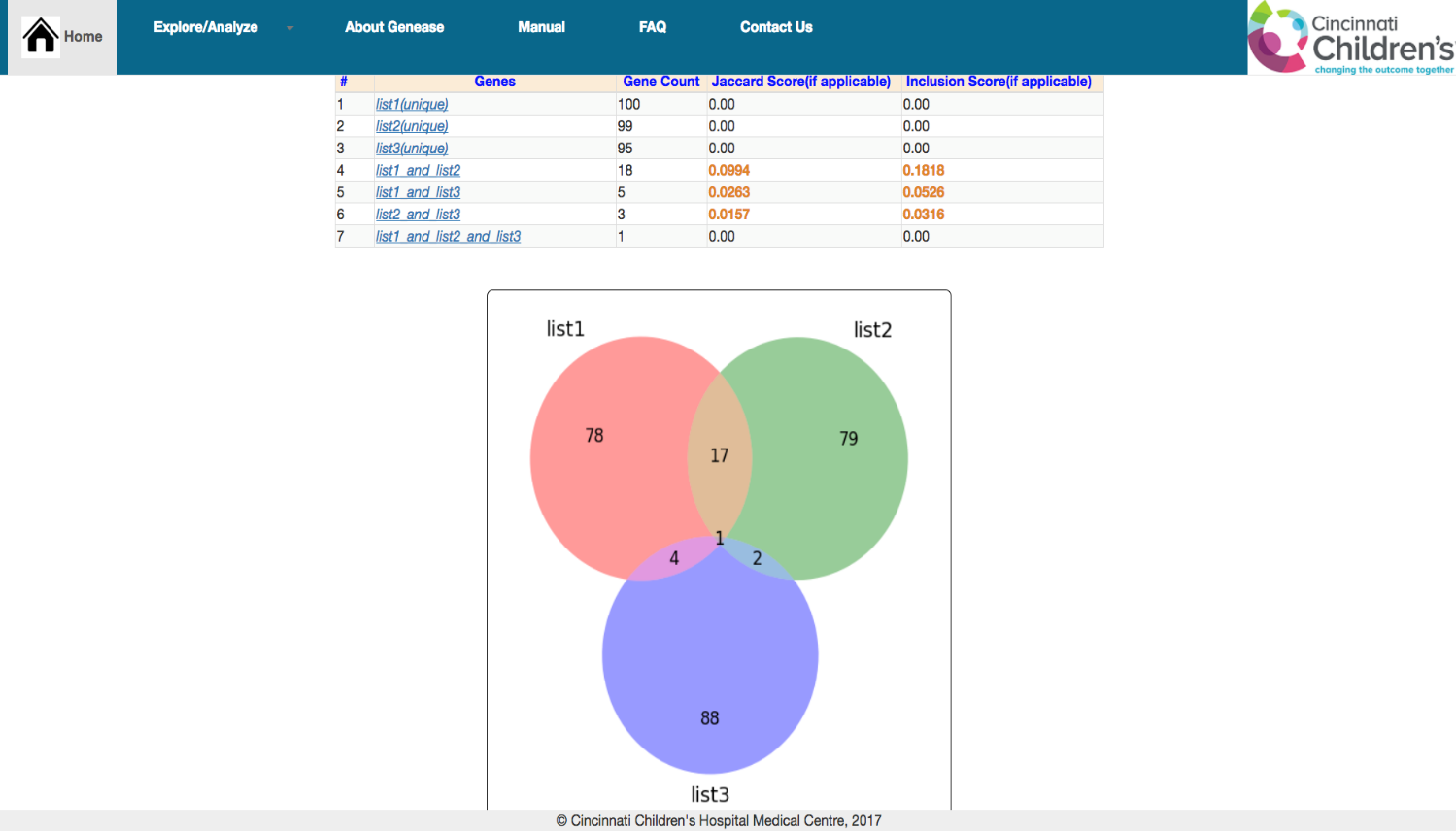
The table row containing the overlap result is formatted with a different color to distinguish from the other entries.

Also the row entry indicating the intersection of the files contains the jaccard and inclusion score metrics for the input lists.

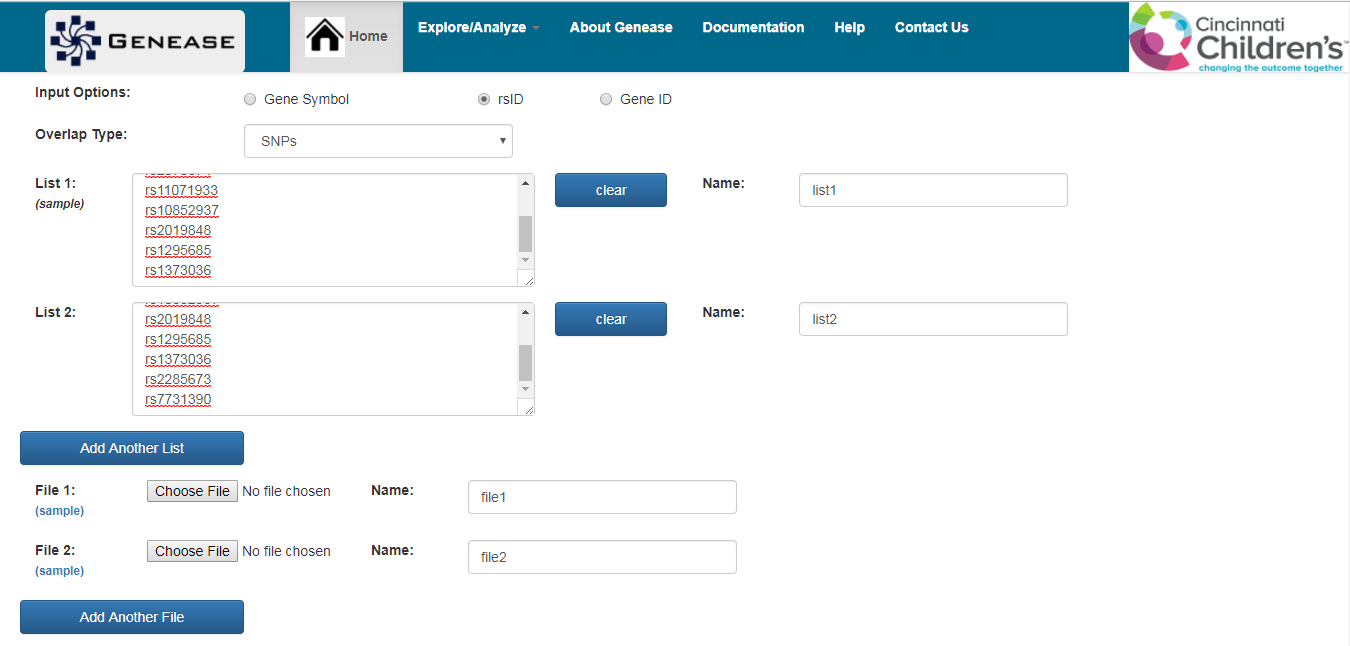
Unique entries in each list (including the intersection entries) can be downloaded by clicking on these names.

In case of three lists, click on the “Add Another List” or “Add Another File” buttons and either copy the third list or upload it.

The result layout is same except that there is a 3 circle venn diagram to illustrate the overlaps and the table contains entries for all possible combinations of overlaps between the 3 lists.

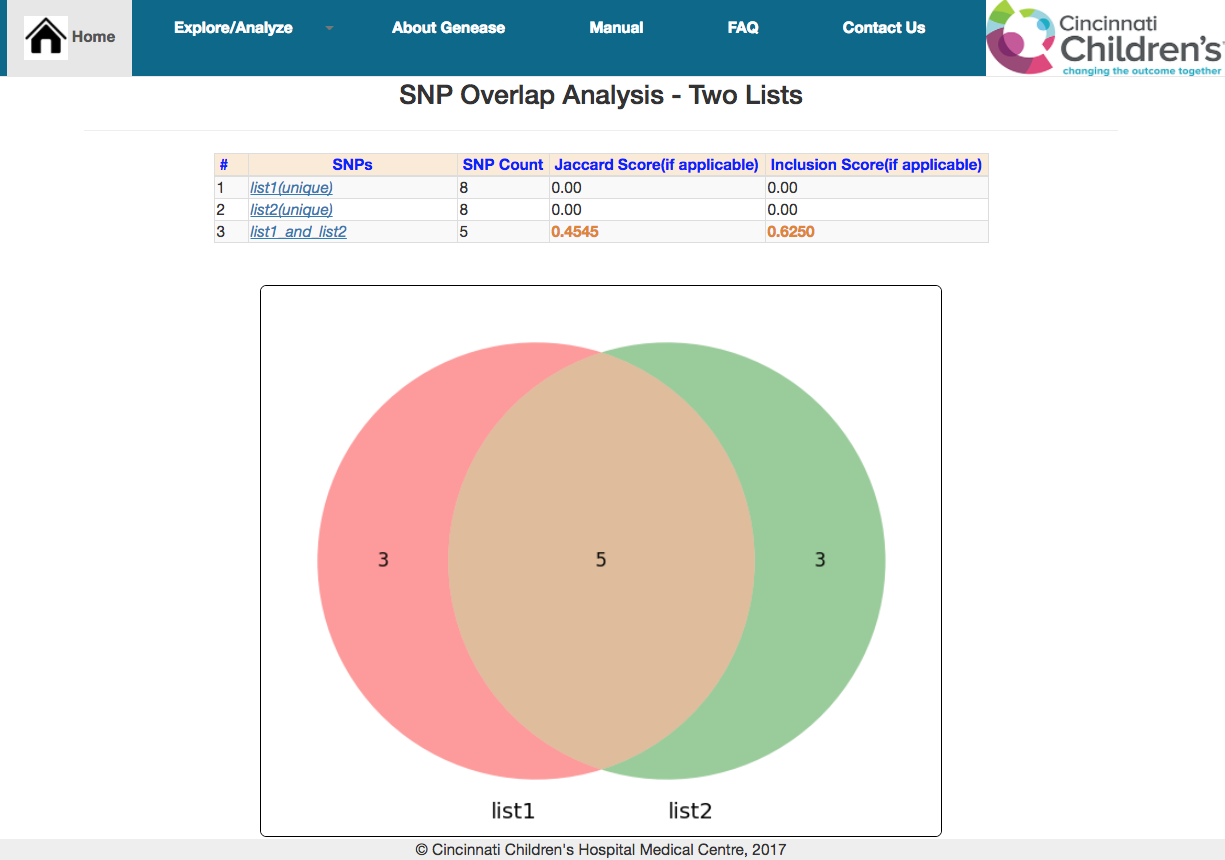
Steps for simple “SNP Overlap” are exactly the same except that “rsID” has to be selected as input option and overlap type is “SNPs”.

Direct gene overlap result – 3 sets



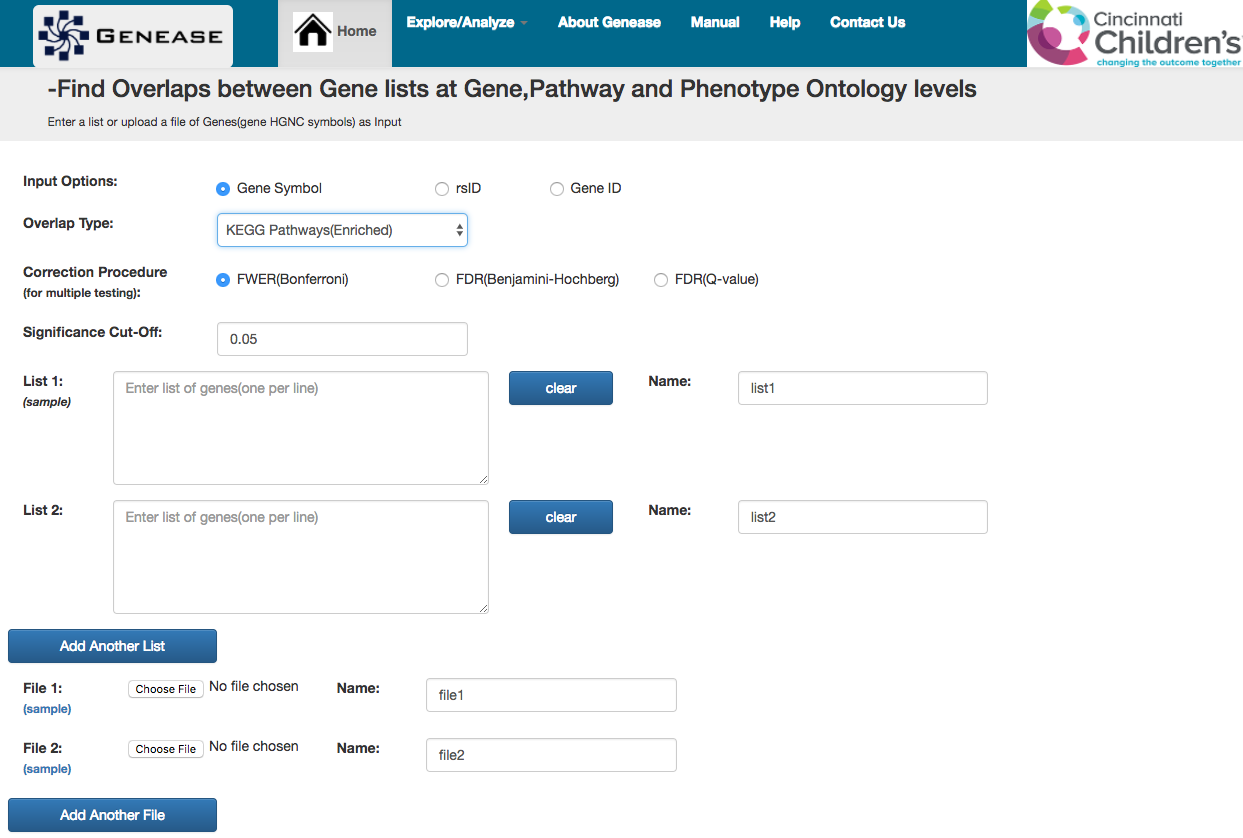
*Simple SNP overlap – 2 sets*

The result interface is similar to gene overlap screenshots from above.



*Simple SNP overlap – 2 sets*

1. **Enriched pathway overlap**

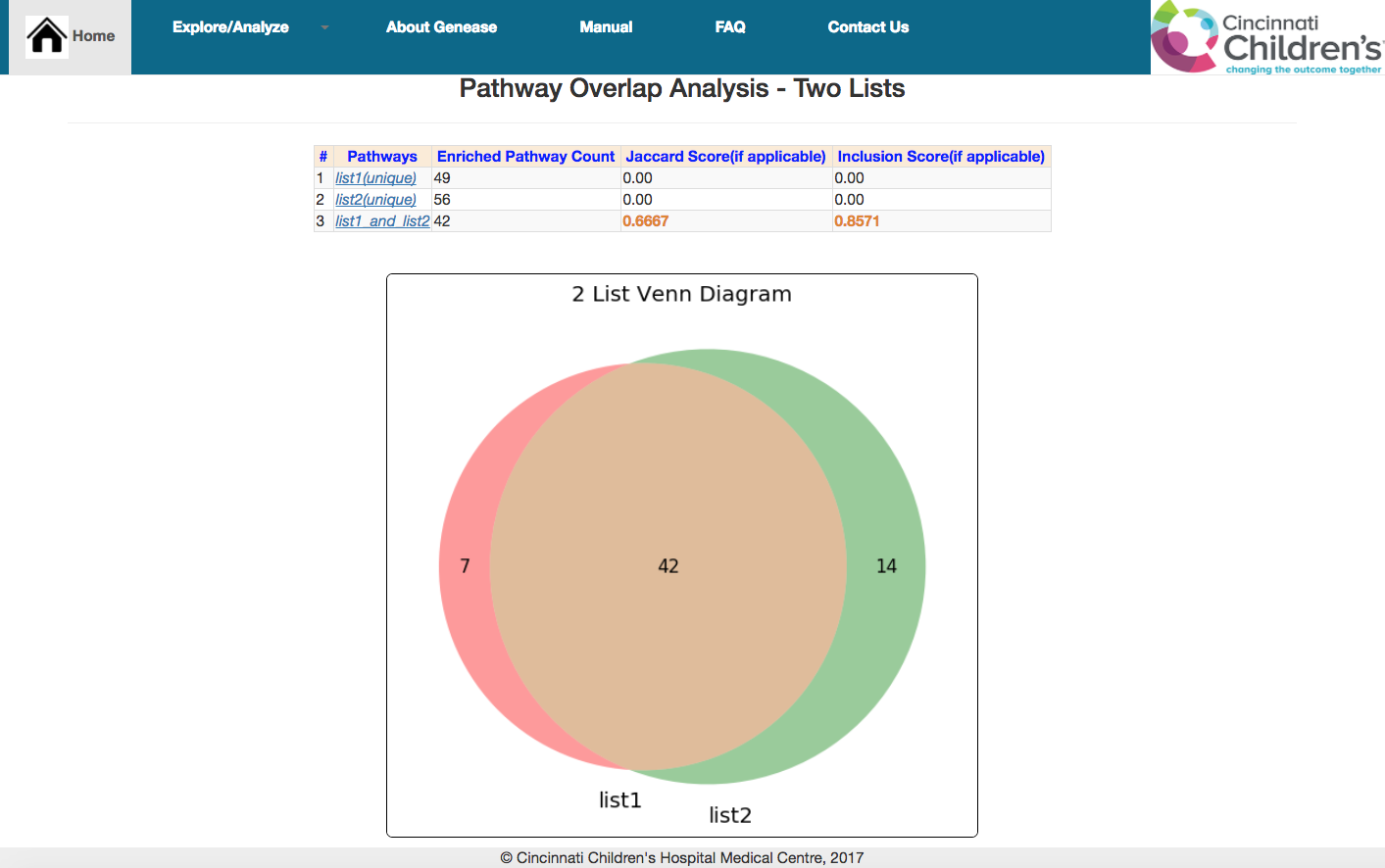
To find overlap using “enriched pathways”, select the appropriate input option and “KEGG Pathways(Enriched)” as the desired overlap type.

*Pathway overlap analysis query*

Since enrichment is involved in the overlap type, one has to select the appropriate “Correction Procedure” for multiple testing correction and enter the “Significance Cut-off” to be used in the analysis.

Either copy the the gene lists in the text-areas or upload the files and click “Submit”

The result interface displays a Venn diagram plot showing the overlap levels between the lists.

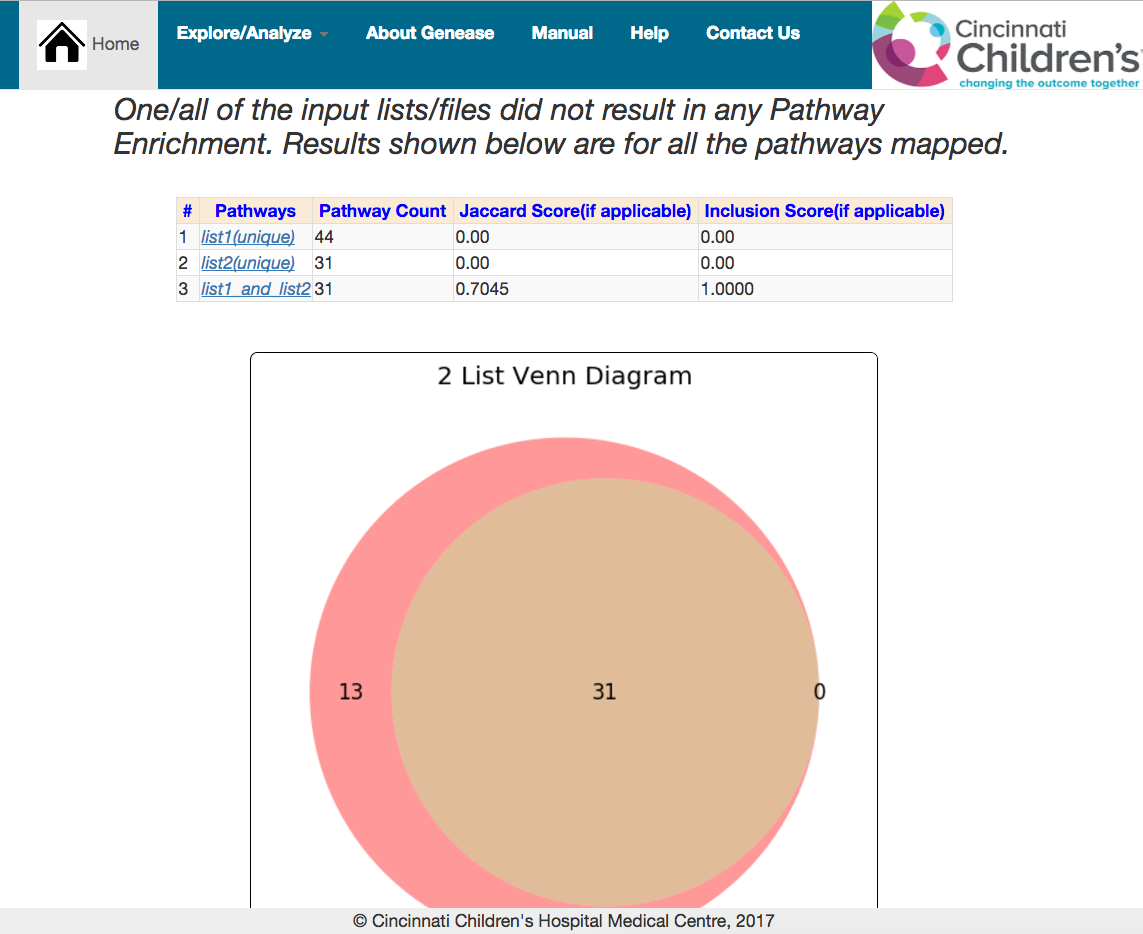


*Enriched pathway overlap analysis result – 2 sets*

Result screen also contains a table in which the filenames are listed in the first row. The table also contains an entry which indicates the “intersection” of enriched pathways from each list.

Unique enriched pathways belonging to each list (including the intersection entries) can be downloaded by clicking on these names.

In the scenario where NO pathways are enriched, the overlaps are found among ALL the pathways mapped to the list instead of just using the enriched ones and a suitable message is displayed.



*Enriched pathway overlap analysis result – 2 sets*

In case of three lists the steps are the same as described above. Use the “**Add Another List”** or “**Add Another File”** buttons to enter or upload the third list and click “**Submit”**

The result screen contains a three circle venn diagram to show the overlap levels.

Similarly overlap analysis can be performed using other criteria like “**Biological Processes(Enriched)”** and “**Phenotype Ontologies(Enriched)”**

The result layout remains the same across all the scenarios.

If “**rsID”** is used as an input option, the corresponding mapped genes are retrieved and the overlaps are found among the enriched annotations mapped to them.