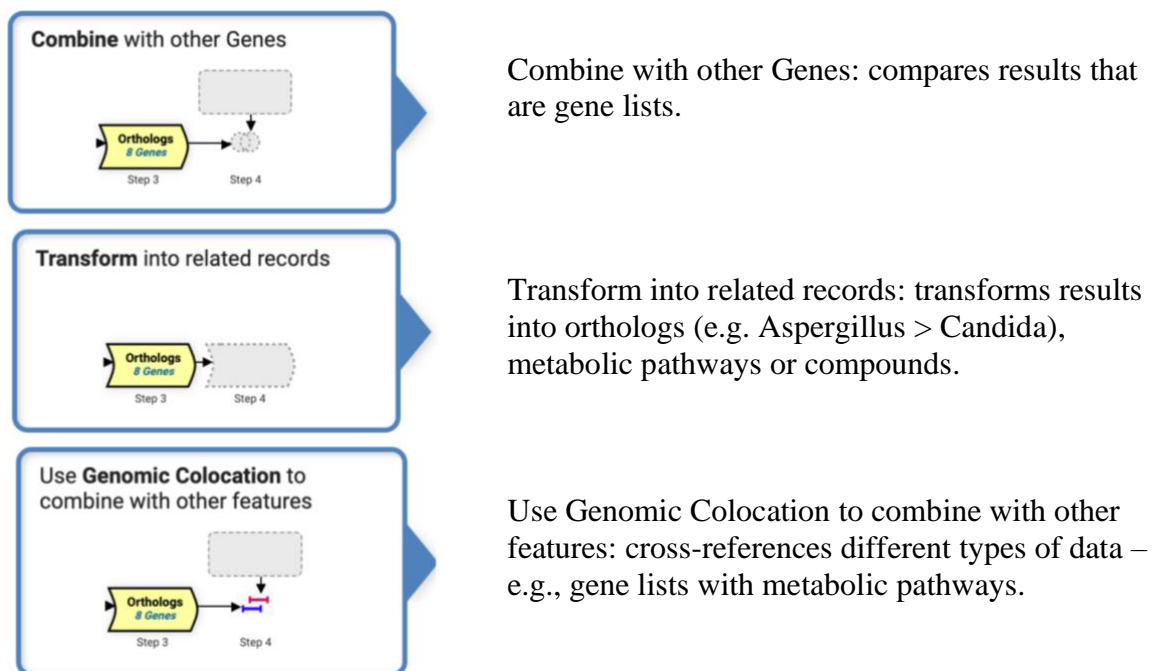


## Advanced Search Strategies

The strategy system offers a unique system of structured searches that can be combined to create multi-step *in-silico* experiments. As seen above, searches can be deployed from the site search, or the ‘Search For...’ menu on the home page, and from the ‘Searches’ dropdown menu in the header of every page.

Searches listed under the “Genes” category will return a list of gene IDs, while searches listed under the ‘SNPs’ or ‘Metabolic Pathways’ will return record relevant to SNPs data (e.g., sequences) and metabolic pathways, respectively.

When creating multi-step search strategy, the search strategy steps can be combined via three methods:



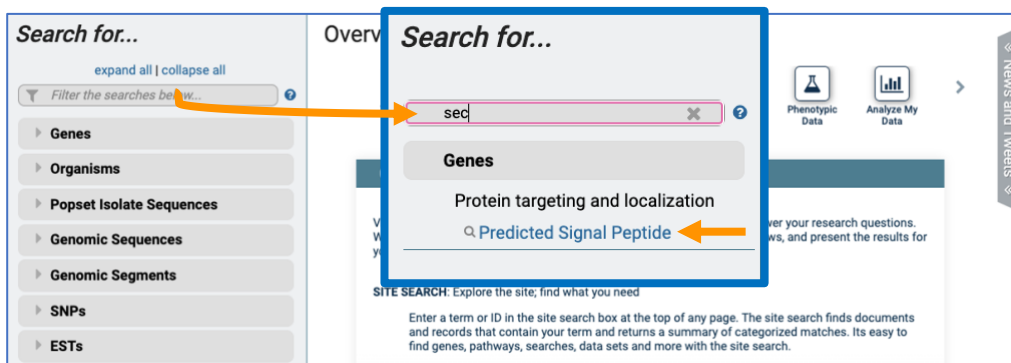
Within the search strategy, each step is connected via the system of Boolean operators that can intersect, unite, or subtract similar records (e.g., gene lists) and cross-references different types of data via the genomic colocation option. Additionally, steps can be masked off from the strategy with the help of “ignore step” Boolean operators that allow quick strategy modification without the need of step deletion.

Revise as a boolean operation											
<input checked="" type="radio"/>		1 INTERSECT 2	<input type="radio"/>		1 UNION 2	<input type="radio"/>		1 MINUS 2	<input type="radio"/>		2 MINUS 1
Revise as a span operation											
<input type="radio"/>		1 RELATIVE TO 2, using genomic colocation									
Ignore one of the inputs											
<input type="radio"/>		IGNORE 2	<input type="radio"/>		IGNORE 1						
<button>Revise</button>											

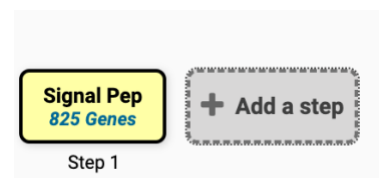
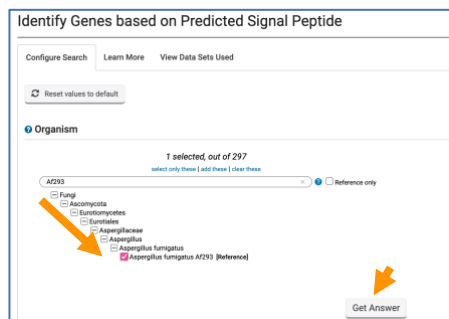
## Creating advanced search strategies in FungiDB.

**In silico experiment:** Identify *Aspergillus fumigatus* Af293 genes that have a signal peptide and non-synonymous mutations identified by whole genome sequencing (WGS) of clinical isolates. Next, determine which genes may be putative vaccine targets (are known epitopes that do not have orthologs in humans).

- **Deploy the “Predicted Signal Peptide” search**
  1. Select the search from the “Search for...” panel (shown below) or the “Searches” menu from the top menu.  
Hint: use the filter box to quickly bring up relevant search.



2. Set the organism to “Aspergillus fumigatus Af293 [Reference]” and click on the “Get Answer” button.



- **Find genes with non-synonymous SNPs using WGS data of clinical isolates aligned to the reference genome *Aspergillus fumigatus* Af293.**
  1. Click on the “Add a step” button with the search strategy.
  2. Within the “Combine with other Genes”, use the “1 INTERSECT 2” Boolean operator and filter available searches to identify and deploy the “SNP Characteristics” search.
  3. Click on the “SNP Characteristics” link in blue to deploy the search.

← Add a step to your search strategy ?

**Combine with other Genes**

Step 1 Step 2

**Transform into related records**

Step 1 Step 2

**Use Genomic Colocation to combine with other features**

Step 1 Step 2

- Choose how to combine with other Genes
  - ☒ 1 INTERSECT 2
  - ☐ 1 UNION 2
  - ☐ 1 MINUS 2
  - ☐ 2 MINUS 1
- Choose which Genes to combine. From...
  - ☒ A new search
  - ☐ An existing strategy
  - ☐ My basket

Genetic variation

[SNP Characteristics](#)

Find genes which have SNPs generated by HTS sequencing meeting specific criteria (Non-synonymous / synonymous SNP ratio, SNP density, etc).

Next, specify parameters of the SNP characteristics search:

4. Select *Aspergillus fumigatus* Af293 as the genome of interest.

Search for Genes by SNP Characteristics

The results will be ☒ intersected with ☐ the results of Step 1.

Configure Search Learn More View Data Sets Used

☒ Reset values to default

**Organism**

7 selected

☐ Reference only

- ☒ Ascomycota
- ☒ Eukaryotes
- ☒ Fungi
- ☒ Aspergillaceae
- ☒ Aspergillus
- ☒ Aspergillus fumigatus Af293 (Reference)

5. Select datasets.

Within the Set of Samples section, click on the Data Set category and select two datasets:

- SNP call on WGS of *Aspergillus fumigatus* drug-resistant clinical isolates
- SNP calls on *A. fumigatus* strains isolated from patients with PA and CNPA

#### Set of Samples

1,118 Set of Samples Total 39 of 1,118 Set of Samples selected [Data Set](#)

[expand all](#) | [collapse all](#)

Find a variable

- culture medium
- Fungal organism
- Data Set**
  - Sample
  - Sample collection
  - Geographic location
  - Sample source
  - Organism under investigation

☒ Keep checked values at top

1,118 (100%) of 1,118 Set of Samples have data for this variable

<input type="checkbox"/>	Data Set	Remaining Set of Samples	Set of Samples	Distribution	%
	Find items	1,118 (100%)	1,118 (100%)		
<input type="checkbox"/>	Aligned SNPs - <i>Aspergillus fumigatus</i> Af1163 strain	1 (< 1%)	1 (< 1%)		(100%)
<input type="checkbox"/>	<i>Aspergillus fumigatus</i> Af293 Genome Sequence and Annotation	1 (< 1%)	1 (< 1%)		(100%)
<input type="checkbox"/>	<i>Aspergillus fumigatus</i> LH-EVOL strains	4 (< 1%)	4 (< 1%)		(100%)
<input type="checkbox"/>	Genomic Context of Azole-Resistance Mutations in <i>Aspergillus fumigatus</i>	24 (2%)	24 (2%)		(100%)
<input checked="" type="checkbox"/>	SNP call on WGS of <i>Aspergillus fumigatus</i> drug-resistant clinical isolates	22 (2%)	22 (2%)		(100%)
<input checked="" type="checkbox"/>	SNP calls on <i>A. fumigatus</i> strains isolated from patients with PA and CNPA	17 (2%)	17 (2%)		(100%)

- Indicate specific SNP characteristics.  
Scroll down the parameter selection window and choose to deploy the search using the following:

**SNP Class = Non-Synonymous**

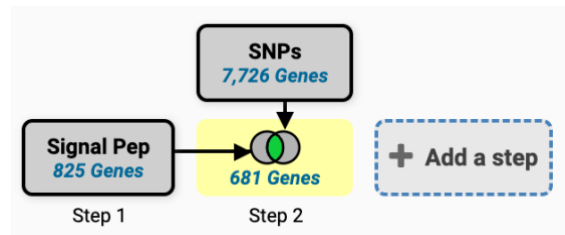
**Number of SNPs of above class >=1**

#### SNP Class

Non-Synonymous ▾

#### Number of SNPs of above class >=

1



- Identify known *Aspergillus* epitopes.**

Epitopes are recognized by immune system and can be used for vaccine development.

- Click on the “Add a step” button.
- Within the “Combine with other Genes”, use the “2 INTERSECT 3” Boolean operator and filter available searches for “epitope” to identify and deploy the “Epitope Presence (IEDB)” search.

**Combine with other Genes**

**Transform into related records**

**Use Genomic Colocation to combine with other features**

**1 Choose how to combine with other Genes**

☒ 2 INTERSECT 3
 ☐ 2 UNION 3
 ☐ 2 MINUS 3
 ☐ 3 MINUS 2

**2 Choose which Genes to combine. From...**

☒ A new search
 ☐ An existing strategy
 ☐ My basket

Immunology  
 Q Epitope Presence (IEDB)

- Set organism to *Aspergillus fumigatus* Af293.
- Set Confidence to “high” and “Medium” and click on the Run Step button.

## Organism

1 selected, out of 232

[select only these](#) | [add these](#) | [clear these](#)

Af293

☐ Reference only

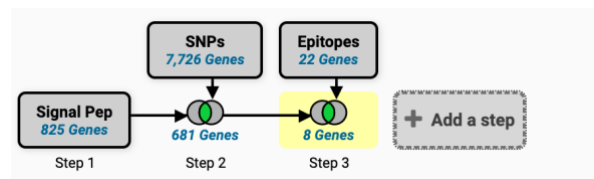
- ☐ Fungi
  - ☐ Ascomycota
    - ☐ Eurotiomycetes
      - ☐ Eurotiales
        - ☐ Aspergillaceae
          - ☐ Aspergillus
            - ☒ Aspergillus fumigatus Af293 [Reference]

## Confidence

- ☒ High
- ☒ Medium
- ☐ Low

[select all](#) | [clear all](#)

Run Step



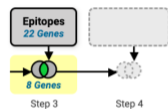
## Identify Aspergillus epitopes that do not have orthologs in humans.

1. Click on the “Add a step” button.

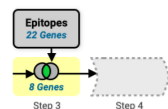
2. Select to deploy the “Orthology Phylogenetics Profile” search.

This search uses OrthoMCL algorithm to identify fungal orthologs across all species in VEuPathDB. Run this search is you want to explore species outside those supported in FunigDB.org.

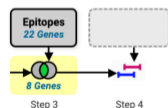
### Combine with other Genes



### Transform into related records



### Use Genomic Colocation to combine with other features



### 1 Choose how to combine with other Genes

- ☒ 3 INTERSECT 4
- ☐ 3 UNION 4
- ☐ 3 MINUS 4
- ☐ 4 MINUS 3

### 2 Choose which Genes to combine. From...

- ☒ A new search
- ☐ An existing strategy
- ☐ My basket

orthology

Orthology and synteny  
[Q Orthology Phylogenetic Profile](#)  
[Q Paralog Count](#)

3. Set parameters for the “Search for Gene by Orthology Phylogenetic Profile”:  
**Find genes in these organisms:** *Aspergillus fumigatus* Af293  
**Select orthology profile:** *Homo sapiens* REF (hsap) must not be in group

**Find genes in these organisms**

1 selected, out of 297  
[select only these](#) | [add these](#) | [clear these](#)

Af29

☐ Fungi  
☐ Ascomycota  
☐ Eurotiomycetes  
☐ Eurotiales  
☐ Aspergillaceae  
☐ Aspergillus  
☐ Aspergillus fumigatus  
☒ Aspergillus fumigatus Af293 [Reference]

☐ Reference only

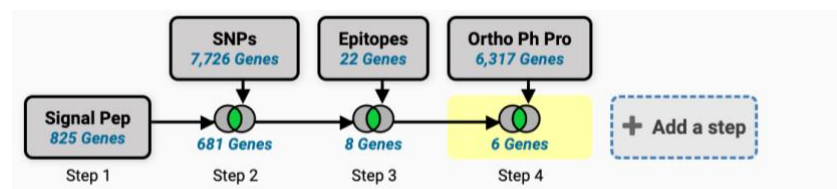
**Select orthology profile**

Click on to determine which organisms to include or exclude in the orthology profile.  
 ( = no constraints | = must be in group | = must not be in group | = mixture of constraints )

sapiens

\* All Organisms  
 \* Eukaryota (EUKA)  
 \* Metazoa (META)  
 \* Chordata (CHOR)  
 \* Mammalia (MAMM)  
 \* Homo sapiens REF (hsap)

How many genes did you get?



Search strategy links:

FGC2024 advanced search strategy 1:

<https://fungidb.org/fungidb/app/workspace/strategies/import/741f30239b2cc68d>

**Optional: Examining host-pathogen datasets in VEuPathDB.org**  
**The next block of exercises will be carried out in [HostDB.org](https://hostdb.org)**

- **Identify genes up-regulated in mice infected with SC5314 at 1d.**
  1. Navigate to the RNA-Seq Evidence search and filter RNA-Seq datasets for “Kirch” to examine the dataset by Kirchner et al. 2019.
  2. Click on the “DE” button.
  3. Choose to examine the sense strand.
  4. Select reference sample: naïve.
  5. Select comparator sample: SC5314\_infected\_1d.
  6. Look for up-regulated genes.
  7. Select magnitude of upregulation: 4 fold.

The screenshot illustrates the VEuPathDB.org RNA-Seq Evidence search interface, with numbered steps 1 through 7 indicating the workflow:

- Step 1:** The search bar contains "ma". The left sidebar shows the navigation menu with "RNA-Seq Evidence" selected.
- Step 2:** The "Filter Data Sets" section shows "kirch" entered, resulting in 1 result (filtered from a total of 44). The "DE" button is highlighted.
- Step 3:** The "Experiment" section shows the selected experiment: "Mouse transcriptomes during oropharyngeal candidiasis infection - Sense".
- Step 4:** The "naive" sample is selected as the reference sample.
- Step 5:** The "SC5314\_infected\_1d" sample is selected as the comparator sample.
- Step 6:** The "Direction" dropdown is set to "up-regulated".
- Step 7:** The "fold difference >=" input field is set to "4".

Below the search results, the "Get Answer" button is visible. The results section shows "Calb\_Galleria\_mouse (de)" with 857 genes identified. A "Step 1" label is present below the results.

- **Identify genes up-regulated in response to 101 persistent strain at 1d of infection.**
  1. Click on the “Add Step” button.
  2. Navigate to the RNA-Seq Evidence search, select “1 minus 2” Boolean operator, filter for “Kirch” to quickly identify the dataset and click on the “DE” button.
  3. Choose to examine the sense strand.
  4. Select reference sample: naïve.
  5. Select comparator sample: 101\_infected\_1d.
  6. Look for up-regulated genes.
  7. Select magnitude of upregulation: 4 fold.

**Step 1:** Click on the “Add Step” button.

**Step 2:** Navigate to the RNA-Seq Evidence search, select “1 minus 2” Boolean operator, filter for “Kirch” to quickly identify the dataset and click on the “DE” button.

**Step 3:** Choose to examine the sense strand.

**Step 4:** Select reference sample: naïve.

**Step 5:** Select comparator sample: 101\_infected\_1d.

**Step 6:** Look for up-regulated genes.

**Step 7:** Select magnitude of upregulation: 4 fold.

**Legend:** Similarity Differential Expression Fold Change Percentile Sense/Antisense

**Filter Data Sets:** 1 result (filtered from a total of 44)

**Organism:** Data Set

**Mouse transcriptomes during oropharyngeal candidiasis infection (Kirchne, et al. 2019)**

**Choose a Search:** [DE] [FC] [P] [S]

**3** ☒ Mouse transcriptomes during oropharyngeal candidiasis infection in mouse - Sense  
☐ Mouse transcriptomes during oropharyngeal candidiasis infection in mouse - Antisense

**4** **Reference Sample**

☒ naïve  
☐ 101\_infected\_9h  
☐ 101\_infected\_1d  
☐ 101\_infected\_3d  
☐ 101\_infected\_7d  
☐ SC5314\_infected\_9h  
☐ SC5314\_infected\_1d  
☐ SC5314\_infected\_3d  
☐ SC5314\_infected\_7d

**5** **Comparator Sample**

☐ naïve  
☐ 101\_infected\_9h  
☒ 101\_infected\_1d  
☐ 101\_infected\_3d  
☐ 101\_infected\_7d  
☐ SC5314\_infected\_9h  
☐ SC5314\_infected\_1d  
☐ SC5314\_infected\_3d  
☐ SC5314\_infected\_7d

**6** **Direction**

up-regulated

**7** **fold difference >=**

4

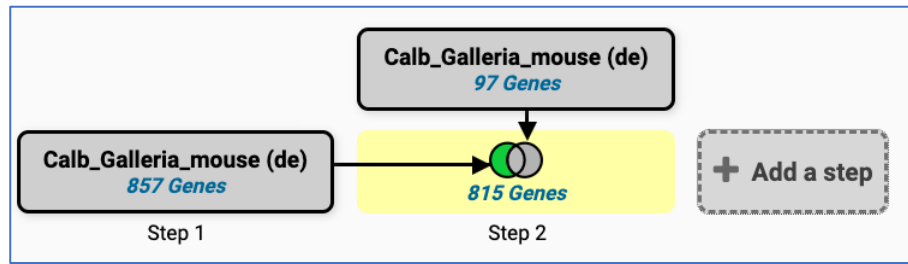
**adjusted P value less than or equal to**

0.1

**Get Answer**



- Modify the Boolean operator to determine genes that are upregulated in response to SC5314 but not 101 strain.

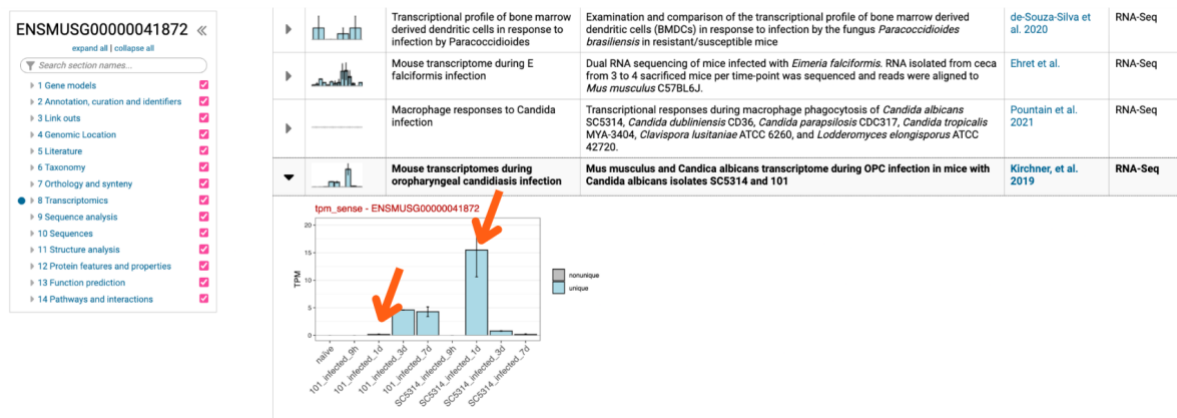


- Examine the results in HostDB:

1. Click on the [Gene ID](#) link for “interleukin 17F” and navigate to the transcriptomics expression section.

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	# Transcripts
ENSMUSG000000104379	ENSMUST000000194867	mmusCS78L6J_chr1:15,853,331..15,856,499(+)	predicted gene, 37509 [Source: MGI Symbol; Acc: MGI:5610737]	1
ENSMUSG000000067780	ENSMUST000000088476	mmusCS78L6J_chr1:17,601,901..17,630,939(+)	peptidase inhibitor 15 [Source: MGI Symbol; Acc: MGI:1934659]	1
ENSMUSG000000025929	ENSMUST000000027061	mmusCS78L6J_chr1:20,730,905..20,734,496(+)	interleukin 17A [Source: MGI Symbol; Acc: MGI:107364]	1
ENSMUSG000000041872	ENSMUST000000039046	mmusCS78L6J_chr1:20,777,146..20,790,617(+)	interleukin 17F [Source: MGI Symbol; Acc: MGI:2676631]	4
ENSMUSG000000041872	ENSMUST000000189301	mmusCS78L6J_chr1:20,777,146..20,790,617(+)	interleukin 17F [Source: MGI Symbol; Acc: MGI:2676631]	4
ENSMUSG000000041872	ENSMUST000000190692	mmusCS78L6J_chr1:20,777,146..20,790,617(+)	interleukin 17F [Source: MGI Symbol; Acc: MGI:2676631]	4
ENSMUSG000000041872	ENSMUST000000191111	mmusCS78L6J_chr1:20,777,146..20,790,617(+)	interleukin 17F [Source: MGI Symbol; Acc: MGI:2676631]	4
ENSMUSG0000000104358	ENSMUST000000192924	mmusCS78L6J_chr1:34,823,525..34,826,560(+)	predicted gene, 37127 [Source: MGI Symbol; Acc: MGI:5610355]	1
ENSMUSG0000000047180	ENSMUST000000056946	mmusCS78L6J_chr1:36,264,597..36,274,679(+)	neuralized E3 ubiquitin protein ligase 3 [Source: MGI Symbol; Acc: MGI:2429944]	2
ENSMUSG0000000047180	ENSMUST000000188666	mmusCS78L6J_chr1:36,264,597..36,274,679(+)	neuralized E3 ubiquitin protein ligase 3 [Source: MGI Symbol; Acc: MGI:2429944]	2
ENSMUSG0000000037447	ENSMUST000000097778	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000115029	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000115031	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000115032	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000115033	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000115034	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000124280	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000126413	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000137906	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000140218	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15
ENSMUSG0000000037447	ENSMUST000000141121	mmusCS78L6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source: MGI Symbol; Acc: MGI:2443039]	15

In summary, we identified genes upregulated in response to SC5314 infection. Notice that the interleukin 17F response is much stronger at 1d in response SC5314 infection. This is consistent with mouse response to *C. albicans* strain 101 being delayed compared to strain SC5314. Now, you may want to go back and look at gene enrichment signatures in fungi to learn more about SC5314 and 101-driven responses.



Search strategy links:

FGC2024 advanced search strategy 2:

<https://hostdb.org/hostdb/app/workspace/strategies/import/ec36d02df763b3d5>