

Exploring the gene record page

Learning objectives

- Become familiar with gene page structure and content
- Navigate to and from the gene pages
- Use the site search to navigate to the gene record page of [Afu2g13260](#), which is a gene known to be important for the virulence of *Aspergillus fumigatus*.

The screenshot shows the gene record page for Afu2g13260. The top navigation bar includes 'My Strategies', 'Searches', 'Tools', 'My Workspace', 'Data', 'About', 'Help', and 'Contact'. A search bar contains 'Afu2g13260' with a magnifying glass icon and a 'Run a new search' button. Below the navigation bar, the page title is 'Genes matching Afu2g13260'. On the left is a 'Filter results' panel with sections for 'Filter Gene fields' (External links, Gene ID) and 'Filter organisms' (Fungi, Ascomycota). The main content area shows one result for 'Gene - Afu2g13260' with a description: 'Putative regulator of adherence, host cell interactions and virulence'. It also lists 'Gene name or symbol: medA' and 'Organism: Aspergillus fumigatus Af293'. A button 'Export as a Search Strategy' is visible in the top right.

- 1 The panel on the left provides a summary of all record types that match Afu2g13260.
- 2 Click on the gene link to navigate to the gene record page for Afu2g13260.
- 3 Clicking on this button will transform your search into a search strategy.
Note: If the button is shaded/inactive, limit your search to a single data type using the Filter results panel on the left.

Gene page components

The top section of the gene record page provides a snapshot of the information available for this gene and offers several shortcuts:

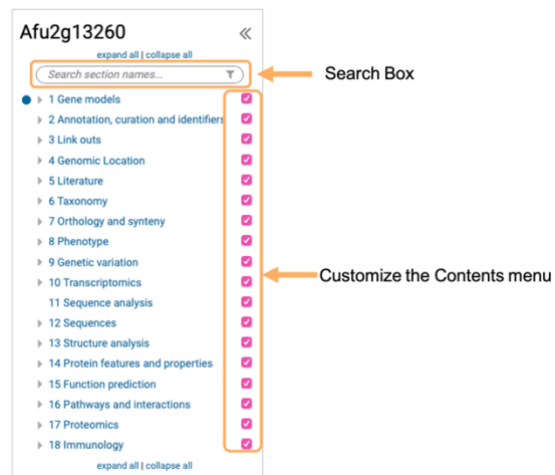
The screenshot displays the gene record page for **Afu2g13260**, a developmental regulator medA, putative. The page is organized into several sections:

- Top Navigation:** Three orange circular icons labeled 1, 2, and 3. Below them are links: "Remove from basket" (with a basket icon), "Remove from favorites" (with a star icon), and "Download Gene" (with a download icon).
- Gene Information:** A table-like structure showing details about the gene:
 - Name:** medA
 - Gene Type:** protein coding gene
 - Biotype Classification:** protein_coding
 - Chromosome:** 2
 - Location:** Chr2_A_fumigatus_Af293:3,408,135..3,413,423(-)
 - Species:** *Aspergillus fumigatus*
 - Strain:** Af293 (with a "Data set" link)
 - Status:** Reference Strain
 - User Comments:** View 1 / Add a new one (with an orange circle 4 next to the link)
 - Community Annotations:** View / Update in Apollo editor
 - Model Organism Database(s):** CGD C3.01180C_A
- Shortcuts Panel:** A horizontal bar with five tabs: Synteny, Alignments, SNPs, Transcriptomics, and Protein Features. Below these tabs are icons for Proteomics and a link to "Also see Afu2g13260 in the Genome Browser or Protein Browser". An orange circle 5 is placed above this panel.

- 1 Add to basket: Save to basket if you want to download gene-specific information for selected genes.
- 2 Add to favorites: Saves genes in the private My Favorites section, where you can add notes or keep track of your project.
- 3 Download Gene: Redirects to a download options page where gene records can be exported in text, FASTA, and GFF3 formats.
- 4 Submit a comment or annotate a gene in Apollo, a web-based structural and functional gene annotation platform.
- 5 The shortcuts panel provides quick access to the selected section within the gene record page.

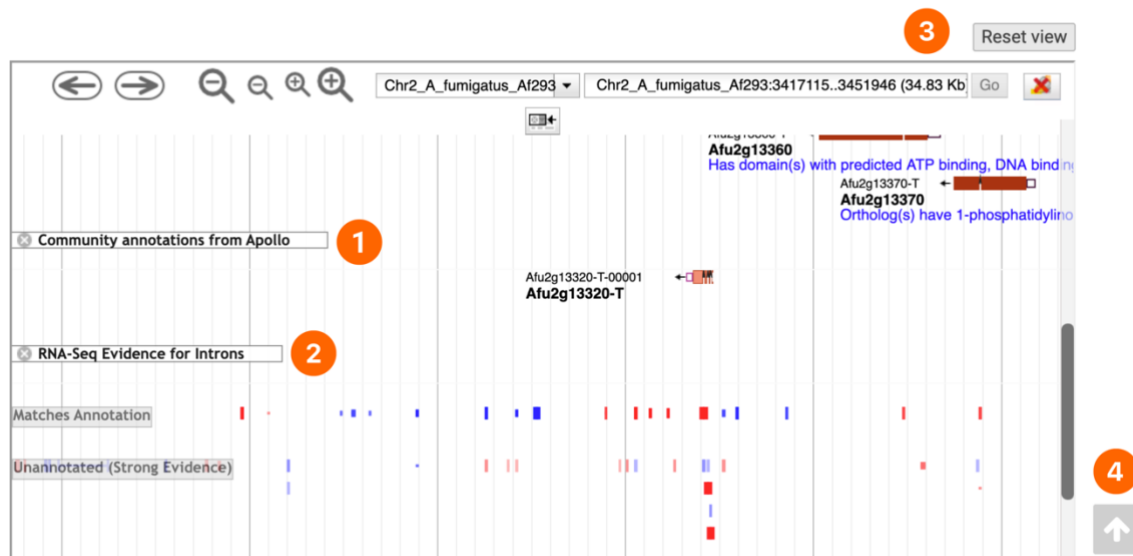
- **Explore the content of this gene record page.** Below are the navigational highlights.

The **Contents** section contains a list of links to various sections in the gene record page and is searchable. Sub-sections of the Contents menu can be hidden by checking the box to the right.



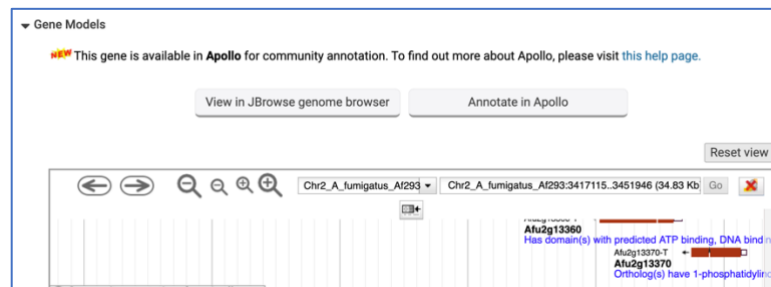
- **Explore the gene model section.**

The **Gene Models** section is the first section of the gene record page, and it contains information about the structure of the gene (e.g. exon count, transcript number, annotated UTRs, community gene annotation in Apollo, introns, etc.) displayed within the genome browse JBrowse wrapper.



- 1 Community annotation from Apollo provides the latest annotation updates (within 24-48hrs after the changes were submitted to the curation team in Apollo)
- 2 Introns that are matching transcript annotation for which there is an abundance of supporting data from aligned RNA-Seq reads are displayed in bold colours (Blue for a forward gene and red for a gene located on the reverse strand).
- 3 If you navigated away from the gene on interest while scrolling, click on the “Reset view” button to return to the default position within the JBrowse wrapper.
- 4 Click this button to navigate to the top of the gene record page.

The “View in JBrowse genome browser” and “Annotate in Apollo” buttons open in separate tabs. In JBrowse, you can activate additional tracks and build custom evidence views. In Apollo, you can modify and create new genes to improve the genome annotation.



- **Explore the RNA-Seq Evidence for Introns track.**

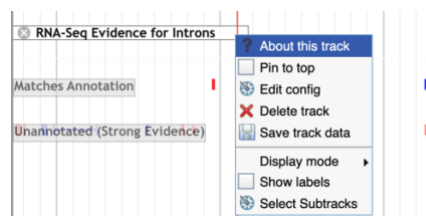
Click on the intron. The pop-up window contains a table showing all experiments and samples that provide evidence for this intron junction (generated by in-house automated pipelines):

Intron Spanning Reads (ISR): The total number of uniquely mapped reads (all samples) which map across the junction and are on the appropriate strand. GSNAP uses splice site consensus sequences to determine the strand of the mapped read.

ISR per million (ISRPM): Intron Spanning Reads Per Million intron spanning reads and thus represents a normalized count of unique reads.

% of Most Abundant Intron (MAI): The percentage (ISRPM of this junction / ISRPM of maximum junction for this gene) of this junction over the maximum for this gene.

Note that the tracks within the JBrowse wrapper have a drop-down menu for further track customization:



Chr2_A_fumigatus_Af293_3434523_3434611_0 details

Intron Junction Details

Intron LocationChr2_A_fumigatus_Af293:3434523..3434611 (+ strand)

Intron Spanning Reads (ISR)7027


ISR per million (ISRPM)2140.85

Gene assignmentAfu2g13295 - annotated intron


% of Most Abundant Intron (MAI)100

Sample Details

Experiment	Sample	Unique	ISRPM	ISR/Cov	% MAI
Gene expression in WT, hrmA deletion, hrmA OE, hrmA_REV, EVOL under hypoxia and normoxia conditions	Delta hrmA hypoxia	64	33.07	5.61	100
	Delta hrmA normoxia	61	28.65	4.78	61.61

- ## 2 Annotation, curation and identifiers
- Community annotations from Apollo [Data sets](#)
- Product Descriptions [Download](#) [Data sets](#)
- Alternate Product Descriptions [Data sets](#)
- Gene Name or Symbol **medA**
- ▼ Names, Previous Identifiers, and Aliases [Download](#) [Data sets](#)
- 
- | Name/ID/Alias | Type | Source |
|-------------------------|--------------|------------------|
| 746128 CADAFUBP00002828 | alternate ID | STRING |
| EAL93620.1 | alternate ID | protein_id |
| Q4X0J5 | alternate ID | Uniprot/SPTREMBL |
| UPI000051EE09 | alternate ID | UniParc |
| XM_750565.1 | alternate ID | RefSeq_dna |
| XP_755658.1 | alternate ID | RefSeq_peptide |
| medA | name | N/A |
| AFUA_2G13260 | previous ID | N/A |
| CADAFUAG00004866 | previous ID | N/A |

.g., CGD, Ensembl, MycoCosm,

[illegible]

In the screenshot above, the syntenic genes are highlighted in grey.

The **Phenotype** section offers curated information, including annotations from the Pathogen-Host Interactions database, COFUN project (selected transcription factors knockouts) and other sources.

▼ *Aspergillus fumigatus* transcription factor KO collection [Data sets](#)

No data available

▼ PHI-base curated phenotypes [Download](#) [Data sets](#)

PHI-base entry	Essential gene	Multiple mutations	Pathogen species	Pathogen strain	Host species	Host strain	Tissue	Mutant phenotype	Disease
PHI2661	no	no	<i>Aspergillus fumigatus</i>	Af293	<i>Galleria mellonella</i> (related greater wax moth)	N/A	larva	reduced virulence	Invasive pulmonary aspergillosis

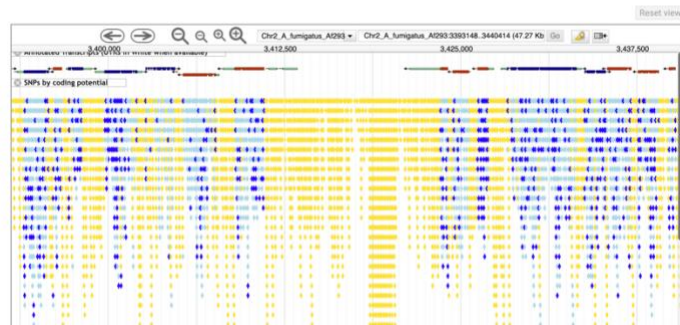
▼ Phenotype (qualities or directionality + entity or biological process) [Download](#) [Data sets](#)

Search this table.

Modification	Allele	Phenotype	Further information	PubMed	ChEBI Annotation Extension
null mutant	medA:hygR	abnormal colony color	Details: delay in brown pigmentation	19889083	N/A
null mutant	medA:hygR	decreased amount biological adhesion	Condition: fibronectin coated wells	19889083	N/A
null mutant	medA:hygR	decreased amount conidium formation	N/A	19889083	N/A
null mutant	medA:hygR	decreased amount virulence	Virulence model: immunosuppressed mouse pulmonary infection	19889083	N/A
null mutant	medA:hygR	decreased amount virulence	Virulence model: insect infection (<i>Galleria mellonella</i> larvae)	19889083	N/A
null mutant	medA:hygR	increased amount cell growth	Details: slightly larger conidia and conidiophores	19889083	N/A

The **Genetic variation** section summarizes integrated SNP data for a given region and classifies SNPs based on the effect on gene function:

- noncoding (yellow diamonds)
- non-synonymous (dark blue)
- synonymous (light blue)
- nonsense (red)



Note that you can interact with the SNP records by using left and right clicking options on your mouse/touch pad.

Left click brings up a pop-up window containing more information about a particular SNP:

NGS_SNP.Chr2_A_fumigatus_Af293.3411010 details

SNP:	NGS_SNP.Chr2_A_fumigatus_Af293.3411010
Location:	3411010
Gene:	Afu2g13260
Position in CDS:	80
Position in protein:	27
Type:	Coding (non-synonymous)
Number of strains:	840
	NA AA (frequency)
Af293 (reference):	C S
Major Allele:	C S (1)
Minor Allele:	A Y (0)

OK

The SNP record linked in blue is linked to the SNP record page, which contains summary of the SNP across different isolates and samples.

[Add to basket](#) [Add to favorites](#) [Download SNP](#)

SNP: NGS_SNP.Chr2_A_fumigatus_Af293.3411010

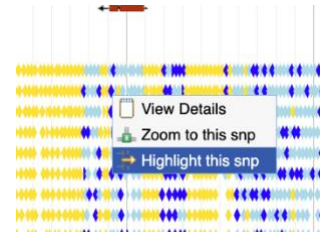
Organism: *Aspergillus fumigatus* Af293
 Location: Chr2_A_fumigatus_Af293: 3,411,010
 Type: coding
 Number of Strains: 840
 Gene ID: Afu2g13260
 Gene Strand: reverse
 Major Allele: G (1)
 Minor Allele: T (0)
 Distinct Allele Count: 2
 Reference Allele: G
 Reference Product: S 27
 Allele (gene strand): C
 SNP context: AGCCGATCCGTCCTGCTGCTATTTGCCAAAGAGCAGCAGTGCTCAAGGAAGAAAGAGGGCA
 SNP context (gene strand): TGCCCTCTTCTTCTTCTGAGCACTGCTGCTCTTTGGCAAATGCAGGCAGACGGATCGGCT

Major allele is the most common allele in the studied population/isolates.

Minor allele frequency is the frequency of the second most common allele. Minor allele

frequency is useful when looking for rare variants or disease-causing SNPs. However, there can be exceptions when minor allele frequency increases under selective pressure (e.g., development of drug resistance).

Right click provides more options for JBrowse view:



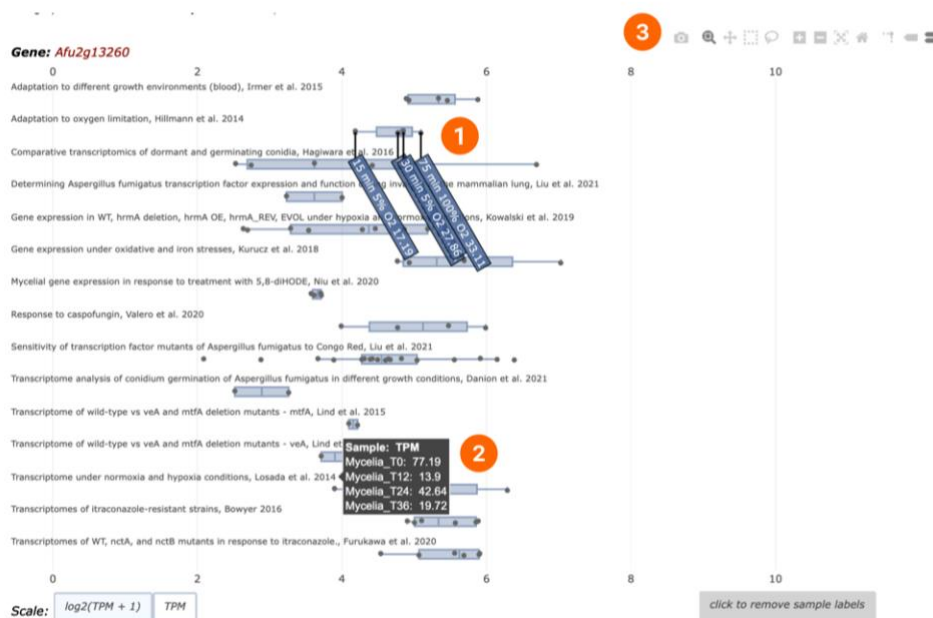
The **Transcriptomics** section (RNA-Seq and microarray data).

The Transcript Expression Summary section provides a big picture of gene expression across different samples and experiments, and helps identify experiments in which the current gene is highly regulated.

▼ RNA-Seq Transcription Summary Data sets

▼ **Summary of expression values.** Each row represents a distinct RNA-Seq experiment. [Click to read more...](#)

- Each black dot represents expression in a sample. A boxplot is shown, with the box representing the median and upper/lower quartiles, and the whiskers representing the max/min values (or 1.5 times the interquartile range, in which case values beyond the whiskers are considered outliers).
- Hover over the experiment name to show a table of sample names and values.
- Click on a boxplot to show labels and values for each sample in an experiment. Click again on the boxplot to hide labels and values. A button at the bottom-right removes labels and values for all experiments.
- Use the toggle button at the bottom-left to switch from Log Scale to Linear Scale.
- Log Scale values are $\log_2(\text{TPM}+1)$ for these reasons:
 1. TPM+1: to de-emphasize low noisy TPM values (i.e., <1).
 2. log2: so that each unit on the x-axis represents a 2-fold difference.
- Navigation buttons appear at the above-right when hovering over the graph. If the buttons do not appear, reload the page.
- Zoom in with the navigation button or click and drag within the graph. Zoom out with the navigation button or double-click within the graph.
- This graph was created with Plotly. [Get more help at their website.](#)

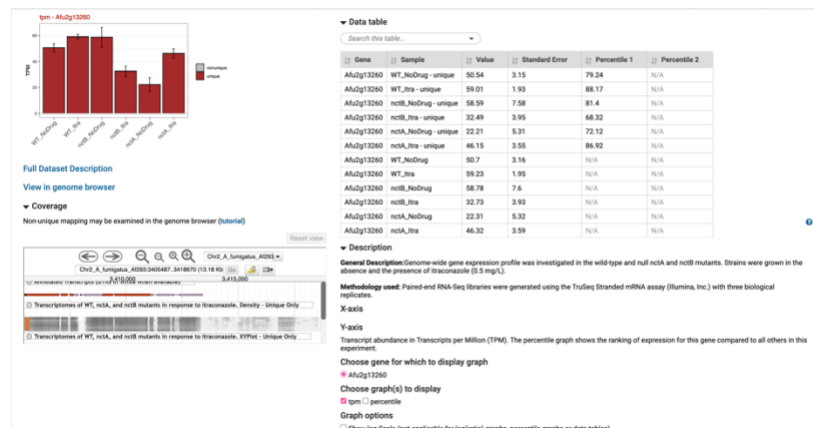


1 Clicking on the box plot will bring up sample labels.

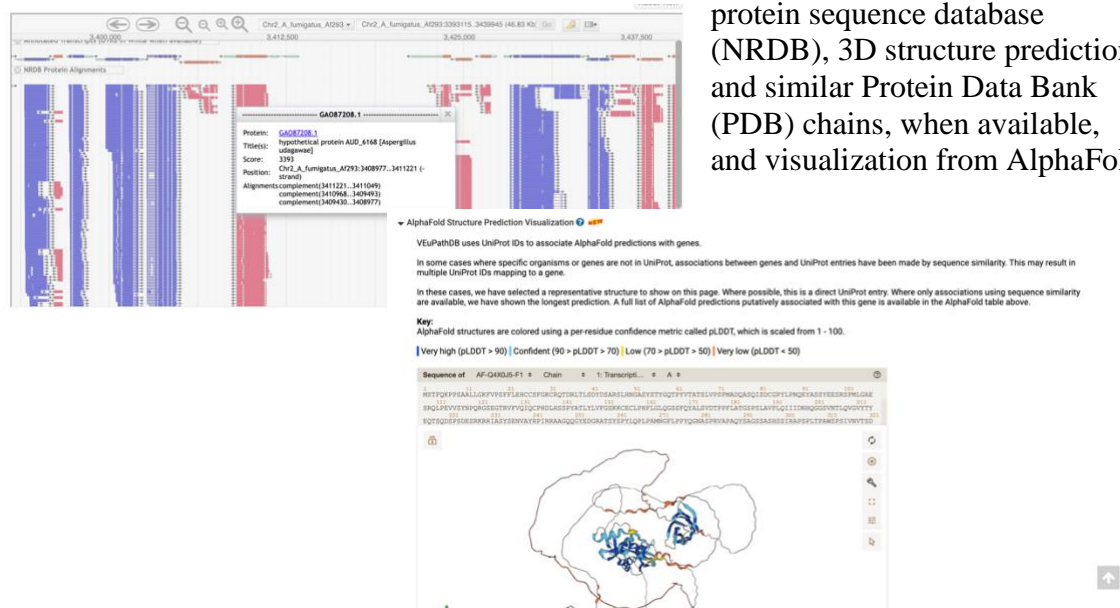
2 Hovering over the experiments will display sample names.

The buttons above the summary graph provide additional options (e.g., download data in PNG, zoom, pan, etc.).

The **Transcript Expression** section, which is located under the RNA-Seq summary section, can be expanded to view the expression graph (TPM), data table, full dataset description, coverage plots, a link to the dataset in JBrowse.



The **Sequence analysis**, **Sequences**, and **Structure analysis** sections offer sequence information (DNA, RNA, and protein), an interactive summary of EST alignments and BLAT hits against the nonredundant protein sequence database (NRDB), 3D structure predictions and similar Protein Data Bank (PDB) chains, when available, and visualization from AlphaFold.



The **Protein features and properties** section provides access to information about protein domains identifications (InterPro), signal and transmembrane predictions graphics, BLASTP hits, and other tools that can be deployed directly from the gene record page using the amino acid sequence of interest:

- ## 14 Protein features and properties
-
- ▶ Attributes and Protein Browser
 - ▶ BLASTP (protein-protein BLAST)
 - ▶ GPI anchor prediction: big-PI Predictor
 - ▶ InterPro Domains
 - ▶ InterProScan: Run on EBI site
 - ▶ MitoProt
 - ▶ STRING: functional protein association networks
 - ▶ WoLF PSORT

The **Function Prediction** section features Gene Ontology (GO) assignments that have been either downloaded from Gene Ontology databases and manually curated by FungiDB. GO Slim terms, Enzyme Commission (EC) numbers with links to EC number and GO terms description and relevant publications are available as well.

Gene ontology provides statements for describing the functions of genes along with three aspects: Molecular function, Biological process, and Cellular component, and it is organized in hierarchies. The GO terms table above provides GO IDs and terms associated with a particular gene and additional metadata that are available for these associations such as source, evidence code (e.g. IDA, IMP, etc.) and reference (PubMed ID).

There are three classes of GO terms in FungiDB:

1. Automatically assigned by InterPro2GO
2. Assigned by FungiDB curators
3. Obtained from external resources such as AspGD, MIPS, and others

For some genes, the **Pathways and interactions** section provides information about metabolic pathways loaded from the KEGG and MetaCyc repositories. Genes are linked to individual pathways by EC numbers when this data is available and clicking on any of the links redirects to an interactive metabolic pathway viewer where the user can explore individual reactions or export all known data for a given pathway. EC numbers are inferred by the OrthoMCL pipeline automatically. As with any automated analysis, use caution when interpreting this data.

For example, Afu2g13260 is not associated with metabolic pathways but its neighbor ([Afu2g13250](#)) does:

Glycine, serine and threonine metabolism

Pathway Source: KEGG
Pathway ID: ec00260
Total Pathway Enzymes: 81
Total Pathway Compounds: 48

1 Pathways and interactions §

Show Contents

▼ 1.1 Metabolic pathways

▼ Cytoscape Drawing

NOTE Click on nodes for more info. Nodes highlighted in orange are EC numbers that we have mapped to at least one gene. The nodes, as well as the info box, can be repositioned by dragging.

