

Mousepost 2.0 User Guide

Steven Timmermans
Claude Libert

Contents

1	Introduction	2
2	Home: the main page	2
2.1	Data overview table	2
2.2	Compare to reference	4
2.3	Deviations in C57BL/6J	4
2.4	Pairwise	4
3	Lists	4
3.1	Stop Gained	5
3.2	Stop Lost	7
3.3	Start codon	7
3.4	Non-synonymous variants	7
3.5	Additional information	8
4	Search	8
4.1	By gene(s)	9
4.2	By region	11
4.3	By function	11
5	C57BL/6J variants	11
6	Pairwise comparisons	11

1 Introduction

This document will provide an overview of the functions of the Mousepost web resource along with several use examples. The overview will follow the options as they are available on the main menu (top) bar of the web tool.

2 Home: the main page

The main landing page, accessed directly from <https://mousepost.be>, provides a short description of the tool as well as a data summary table and links to all other functionality.

Home Lists Search C57BL/6J Pairwise Manual

A searchable website for mouse polymorphic sequence tags in protein coding genes: mousepost.be



With this website we aim to provide an overview of the non-synonymous mutations, both SNP and small insertions/deletions, that affect the protein coding transcripts of 52 often used laboratory strains of mice. We identified mutations on a per codon basis and classified them in 3 main groups: Stop Gained (SG), for mutations leading to an early stop codon, Stop Loss (SL), for mutations removing a stop codon and Mutated (MUT), for all others. Mutations not affecting a stop codon were analyzed with PROVEAN in order to predict the effect on the final protein function.

When using data retrieved from this website, please refer to one of the following papers: [Complete overview of protein-inactivating sequence variations in 36 sequenced mouse inbred strains](#) or [Overview of inactivating mutations in the protein-coding genome of the mouse reference strain C57BL/6J](#)

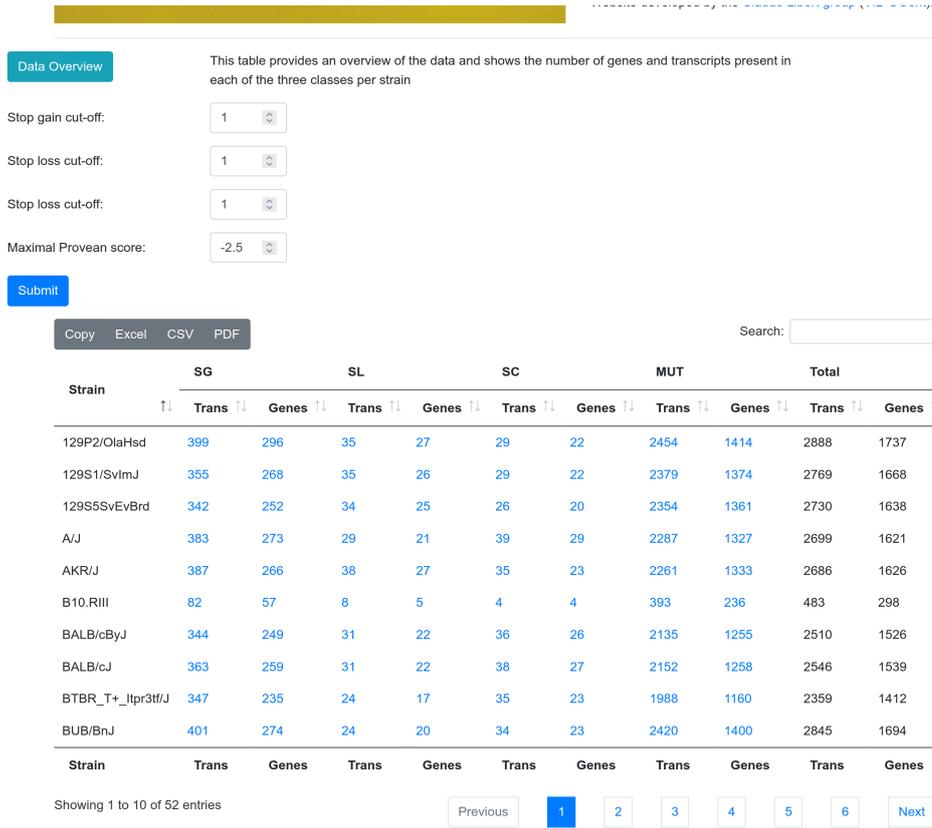
Website developed by the [Claude Libert group \(VIB-UGent\)](#).

Data Overview	This table provides an overview of the data and shows the number of genes and transcripts present in each of the three classes per strain
Compare to reference	Obtain information of proteins different from the C57BL/6J reference strain
Deviations in C57BL/6J	Information on proteins that are different in C57BL/6J (reference) compared to other strains
Pairwise	Pairwise comparisons between included strains

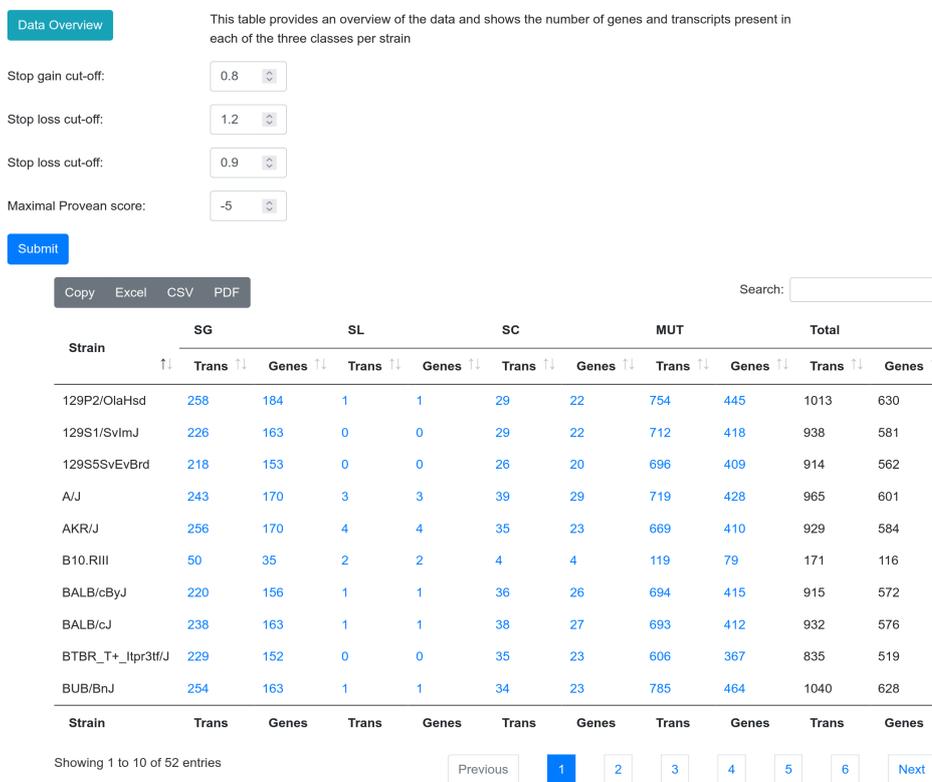
Figure 1: The mousepost.be main homepage, all functions are directly accessible from here

2.1 Data overview table

Available from the *data overview* button is an overview in table format of the number of transcripts and corresponding genes present per variant class for each included inbred mouse line. The table can be customized by using the cut-off selectors above the table, please not that it may take some time for the table to update after submitting the request. The numbers in the table serve as links to the **Lists** page of mousepost, and clicking them will allow you to obtain a list of the actual transcripts that make up the displayed number in this table, with the active applied filters to the table.



(a) Default table as shown on the page



(b) Table with various filters settings adjusted

Figure 2: The overview table: number of variants per class and per strain

2.2 Compare to reference

This section provides link to all the strain of interest vs reference (C57Bl/6J) functions of the mousepost tool: the **Lists** through the *Per strain info* button and **Search**, by means of the *Search genes, locations or functions* button, functions. These functions are also available through the top menu, however here a short description is provided to help guide you to the mousepost function that is most appropriate for the query of interest.

2.3 Deviations in C57BL/6J

Here the links to the mutation that were identified in the C57BL/6J reference genome can be found. Mutations are grouped par class: stop gain, stop loss, start codon and mutated. Like the comparisons to the reference, these data can also quickly be accessed from the tom menu: each link is also present as a item in the submenu of the **C57BL/6J** item.

2.4 Pairwise

A link to the pairwise comparison functionality. Here any two strains may be compared to obtain a list of their differences.

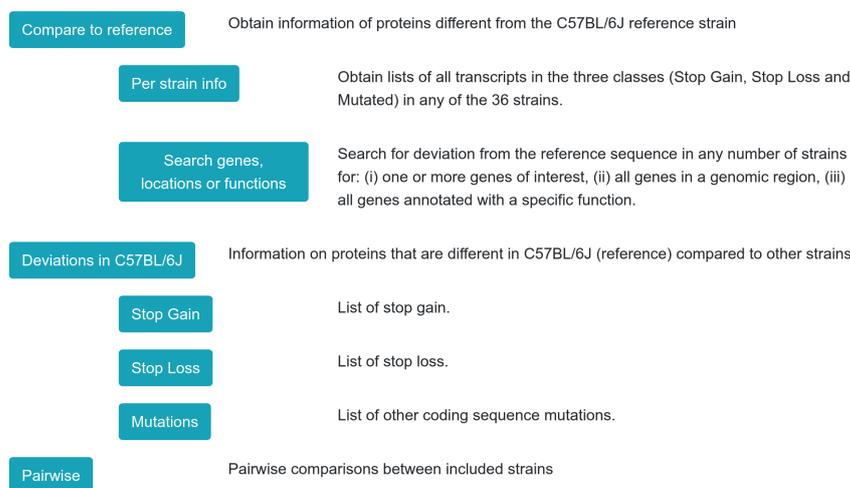


Figure 3: All links on the main page

3 Lists

The **Lists** functions allow the you to obtain per strain and per mutation type gene lists. The strain of interest can be selected from the drop down menu, and a mutation type can be selected by using one of the selector radio buttons.

Home Lists Search C57BL/6J Pairwise Manual

List of transcripts

Obtain a list of Stop Gain, Stop Loss, star codon or Mutated transcripts per strain and optionally filter by chromosome. Transcripts in bold are annotated as protein coding in the reference (C57BL/6J)

Strain:

Stop gain -- Maximal ratio: domains

Stop loss -- Minimal ratio:

Start Codon -- Minimal ratio:

Mutation -- Maximal provean score:

Optional:

chromosome

Figure 4: The list function: main user interface

3.1 Stop Gained

This will return the stop gain variants in the selected strain, compared to C57BL/6J, that are reduced to a least the length ration chosen. By default a value of 1 is used for the maximal length ratio, which disables filtering and will return all stop gain transcripts. Optionally, it is possible to also request that protein domains that are in the truncated region, and thus (partially) lost are included in the result set.

Results are reported in table format. The transcript id, gene name, length in C57BL/6J (ref length), length in the strain of interest, SOI/ref length ratio and a visual representation of the truncation are shown. The vial ratio is a bar representing the ref length as 100% with the green bar showing the length of the SOI sequence and the lost portion shown in red. If the domain checkbox is selected 2 more columns are added: the number of domains in the truncated region and the interpro domain ids. Finally the table include external links in the final column, these are present in all tables on the tool and are explained in section.

By using this tool it is possible to recover the Nlrp1b stop gain variant in the AKR/J strain: selecting AKR/J and a stop gain cut-off of 0.8, a total of 256 transcripts are returned (as also shown in the overview table in figure 2). The Nlrp1b gene has 4 transcripts with a stop gain mutation, reducing the length to just over 75% of the normal (C57BL/6j) length.

List of transcripts

Obtain a list of Stop Gain, Stop Loss, star codon or Mutated transcripts per strain and optionally filter by chromosome. Transcripts in bold are annotated as protein coding in the reference (C57BL/6J)

Strain:

Stop gain -- Maximal ratio: domains

Stop loss -- Minimal ratio:

Start Codon -- Minimal ratio:

Mutation -- Maximal provean score:

Optional: : -

chromosome

(a) AKR/j stop gain selection with 0.8 as cut-off

Copy Excel CSV PDF Search:

Transcript	Gene	Ref length	AKR_J length	Ratio	Graphic ratio	links
ENSMUST00000179472	<i>Mro</i>	246	34	0.1380		U E P
ENSMUST00000119239	<i>Mro</i>	245	34	0.1390		U E P
ENSMUST0000022245	<i>Mrps30</i>	442	22	0.0500		U E P
ENSMUST0000081777	<i>Mug2</i>	1451	290	0.2000		U E P
ENSMUST00000111393	<i>Ncor2</i>	2332	6	0.0030		U E P
ENSMUST0000094046	<i>Nlrp1b</i>	1174	886	0.7550		U E P
ENSMUST00000108516	<i>Nlrp1b</i>	1174	886	0.7550		U E P
ENSMUST00000108514	<i>Nlrp1b</i>	1177	890	0.7560		U E P
ENSMUST00000108515	<i>Nlrp1b</i>	1177	890	0.7560		U E P
ENSMUST00000107844	<i>Nme1</i>	117	79	0.6750		U E P

Showing 91 to 100 of 256 entries

(b) Table with various filters setting adjusted

Figure 5: Results table for AKR/J stop gain: the Nlrp1b transcripts

3.2 Stop Lost

Here you can query the stop loss variants in a strain of interest. Analogous to the stop gain set a length ratio cut-off can be set where the minimal extension compared to the reference is enforced. The results table is identical to the table from stop gained, with one main difference in the meaning of the visual ratio: the total length is the length of the sequence in the strain of interest. The green part represents the C57BL/6J length and the red the relative extension of the sequence in the selected strain.

3.3 Start codon

The start codon variant class was newly added in the 2.0 update of mousepost. Here it is possible to query the transcripts that have a mutation in the canonical stop codon. The result table returned is identical to the stop gain results table. The reported length in the selected strain is now from the protein that is obtained after starting translation from the first following AUG start codon. If this is not in the same frame as the lost start codon, there will also usually be an *early* stop codon resulting a very small peptide. The visual ratio indicates were in the normal protein the start codon loss one should start in order to obtain a protein of the species length found.

(a) AKR/j start codon selection, no filtering applied

Transcript	Gene	Ref length	AKR_J length	Ratio	Graphic ratio	links
ENSMUST0000056955	4933428G20Rik	101	1	0.0100		U E P
ENSMUST0000086248	Aurkc	315	19	0.0600		U E P
ENSMUST00000160505	C1s1	694	42	0.0610		U E P
ENSMUST00000162443	C1s1	694	42	0.0610		U E P
ENSMUST0000084830	Chrna10	447	25	0.0560		U E P
ENSMUST0000030815	Cort	109	90	0.8260		U E P
ENSMUST0000023056	Deptor	397	396	0.9970		U E P
ENSMUST0000027826	Dusp23	150	52	0.3470		U E P
ENSMUST0000074783	Hmcn1	5634	28	0.0050		U E P
ENSMUST00000137197	Hmcn1	5517	28	0.0050		U E P

(b) Table with various filters setting adjusted

Figure 6: Results table for AKR/J start codon

3.4 Non-synonymous variants

Non-synonymous variants not belonging to the previous 3 classes can be queried using the "Mutation" selector. Here a maximal PROVEAN score must be entered as a filter, by default this is set to -2.5, the value proposed by the authors of PROVEAN for an optimal balanced accuracy. In the example in figure 7 a score cut-off of -5 is used, which is more stringent (but less sensitive). Using these settings a total of 669 transcripts are found in the AKR/J strain. The result set here includes the transcript and gene names, the chromosome, to make filtering

of very large datasets easier, the number of supporting sequences found by PROVEAN to calculate the score. PROVEAN scores obtained from less than 40 sequences may be considered unreliable, which is why the number is included. Next there is the number of mutated positions in the protein sequence as well as the PROVEAN score of the lowest scoring variant.

Obtain a list of Stop Gain, Stop Loss, star codon or Mutated transcripts per strain and optionally filter by chromosome. Transcripts in bold are annotated as protein coding in the reference (C57BL/6J)

Strain:

Stop gain -- Maximal ratio: domains

Stop loss -- Minimal ratio:

Start Codon -- Minimal ratio:

Mutation -- Maximal provean score:

Optional:
 chromosome : -

(a) AKR/j mutated selection, PROVEAN score of -5 or lower applied

Copy Excel CSV PDF Search:

Transcript	Gene	Chr	#sequences	#mutations	Lowest Score	links
ENSMUST00000061293	Olfir223	chr11	400	1	-33.704	U E P
ENSMUST00000072977	Olfir924	chr9	423	1	-31.825	U E P
ENSMUST000000217350	Olfir924	chr9	423	1	-31.825	U E P
ENSMUST000000210571	Olfir573-ps1	chr7	515	10	-30.816	U E P
ENSMUST00000064830	Olfir573-ps1	chr7	515	11	-30.812	U E P
ENSMUST000000134389	4930519G04Rik	chr5	7	2	-29.048	U E P
ENSMUST000000031547	4930519G04Rik	chr5	7	2	-27.381	U E P
ENSMUST000000215626	Olfir223	chr11	400	1	-25.465	U E P
ENSMUST00000088523	Samhd1	chr2	74	8	-24.656	U E P
ENSMUST00000054871	Fpr3	chr17	129	1	-22.353	U E P

Showing 1 to 10 of 669 entries

(b) Table with various filters setting adjusted

Figure 7: Results table for AKR/J mutated

3.5 Additional information

It is possible to apply a filtering on genomic location as an optional extra setting, which is applicable to all the mutation classes. If set, only genes overlapping the specified region will be returned in the list. This will return smaller and more directly useful result sets if a genomic region of interest (e.g. from a QTL) is known.

Clicking on the gene name or transcript ID will lead to a detail page with all information available on the gene and transcript. For MUT class transcripts this includes a list of every variant present with the corresponding PROVEAN scores. For all the other classes the sequences of both C57BL/6J and the selected strain are provided.

4 Search

The third main function of mousepost is an advanced search function providing 3 ways to get information from the database: by gene name (+ filtering on strain/type), by location and by function.

In all cases results are returned in the same format: in a table that is divided in one section for each variant type. The variant type remains in the first row as the table is browsed at make clear which class the currently shown transcripts belong to. Information in the table includes the strain, gene name, transcript and the lowest PROVEAN score (MUT) or length ratio (SG, SL, SC)

The screenshot shows the mousepost.be main homepage. At the top left is a blue 'Search' button. Below it are buttons for 'Copy', 'Excel', 'CSV', and 'PDF'. To the right is a search input field labeled 'Search:'. Below these is a table with the following columns: Transcript, Gene, Chr, #sequences, #mutations, Lowest Score, and links. The table contains 10 rows of data. At the bottom, it says 'Showing 1 to 10 of 16 entries' and has navigation buttons for 'Previous', '1', '2', and 'Next'.

Transcript	Gene	Chr	#sequences	#mutations	Lowest Score	links
ENSMUST00000082088	Mamld1	chrX	52	1	-7.745	U E P
ENSMUST00000114629	Mamld1	chrX	51	1	-7.745	U E P
ENSMUST00000114546	Zfp185	chrX	50	1	-6.195	U E P
ENSMUST00000164800	Zfp185	chrX	50	1	-6.195	U E P
ENSMUST00000059003	Hsf3	chrX	166	1	-5.880	U E P
ENSMUST00000164693	Hsf3	chrX	50	1	-5.746	U E P
ENSMUST00000119035	Hsf3	chrX	50	1	-5.546	U E P
ENSMUST00000179832	Hsf3	chrX	50	1	-5.546	U E P
ENSMUST00000239162	Mamld1	chrX	52	1	-5.030	U E P
ENSMUST00000065932	Gripap1	chrX	50	1	-4.767	U E P

Figure 8: The mousepost.be main homepage, all functions are directly accessible from here

4.1 By gene(s)

The first search option is a straightforward search on gene name and any number of genes can be entered, separated by *new lines*. It is further possible to find mutations in the genes entered only in a subset of strains, or search only for a subset of variants. For ease of use, we have grouped the strains of the collaborative cross project in on set so that they can be selected or deselected easily. For example, if we want, by means of a test, check the known *Tlr3* variant in the CZECHII/EiJ, which has a P369L mutation, this search function can be used. Entering *Tlr3* and adding the filter mutated (optional) will give a result set of all strains having a mutated *Tlr3* gene. Indeed, three transcripts of the gene are mutated in CZECHII/EiJ, including the primary transcript (ENSMUST00000034056) with 2 mutations. Checking the full details of the transcript reveal that the variants T392A and P369L were annotated by mousepost.

Search function

Gene(s): Strain(s): Type(s):

Enter region: : -

GO term: Chr (optional):

Copy Excel CSV PDF Search:

Strain	↑↓	Ensgene	↑↓	Symbol	↑↓	Transcript	↑↓	Ratio/score	↑↓	links	↑↓		
SNP/indel Mutation													
ZALENDE/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000034056		-0.001		U	G	E	P
ZALENDE/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000167106		-0.001		U	G	E	P
ZALENDE/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000209772		-0.001		U	G	E	P
LEWES/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000034056		-0.920		U	G	E	P
LEWES/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000167106		-0.920		U	G	E	P
LEWES/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000209772		-0.920		U	G	E	P
CZECHII/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000034056		-1.453		U	G	E	P
CZECHII/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000167106		-1.453		U	G	E	P
CZECHII/EiJ		ENSMUSG00000031639		Tlr3		ENSMUST00000209772		-1.453		U	G	E	P
JF1/MsJ		ENSMUSG00000031639		Tlr3		ENSMUST00000034056		-1.453		U	G	E	P

Figure 9: Results from Tlr3 gene search

4.2 By region

Searching by region can be done by selecting the chromosome and optionally setting a start and stop region to further narrow the results. This is analogous to the location filter that can be used in the per strain lists, but will give results for **all** strains in the database.

4.3 By function

A search for variants that are annotated to a specific GO annotation is also possible. A GO term can be entered, with an autocomplete function providing suggestions, and genes that are annotated to this term will be used as if that had been entered in the "gene" search method.

5 C57BL/6J variants

Using a consensus based reference, we have also checked the 4 mutation classes in C57Bl/6J compared to this reference. One page was created for each of the classes and is available through the **C57BL/6J** submenu. However, the C57BL/6J start codon variants could not be determined as the strain belonged to the consensus sequence in all cases and this class was left out of the options. In order to check if a protein has mutations in C57BL/6J the result tables on all 3 pages may be filtered based on gene name. The main evaluation criterion is the *agreement score*, which shows how "unique" the C57BL/6J sequence is. It is comprised of the number of different variations at a position, and how many strains agree with the consensus or the C57BL/6J sequence. For example if 52/53 strains have the consensus sequence and C57BL/6J is different and unique the agreement score will be very high (=1). If the consensus sequence has 40/53 strains and all the other strain have the same sequence as C57BL/6J the score will be lower but still good (± 0.75). In the case that there would be 5 variants total and 30/53 strains make up the consensus and C57BL/6J has only a few strain that agree with it the score will be low. Low scores (from around 0.4 or lower) should be carefully interpreted. Clicking on the score will show the two components individually as well as all strains that agree with C57BL/6J.

As these are predictions, you should further check the quality especially for lower scores. It is possible that the C57BL/6J is predicted to have a mutation when compared to the consensus, however it must be considered that the majority of the other sequenced strain is wrong, only a small minority of all strain has been sequenced after all. Checking the sequence in a direct C57BL/6J comparison with one of the strains having the consensus sequence using the lists tool is advised as well as comparing the peptide sequence annotated in the UniProt database.

6 Pairwise comparisons

Finally, mouspost allows any two strains to be compared and the differences between them recovered. In order to do so, a strain A and a strain B must be selected from the strain sets. Optionally, only results for a specific genomic region can be filtered for. The query will give a result table with all transcripts giving different protein sequences in the selected strains. As there is no actual reference, no classification of the variants is (can be) performed. The results returned are the gene/transcript ids, the number of positions in the sequence that are different between the strains, and how many strains have the same sequence as the "A" (*Like A*) and "B" (*like B*) strain respectively. For example, comparing A/J (A) and AKR/J (B) will show a total of 7002 results in the table, the first, ENSMUST00000000049 (Aphoh) has 1 difference and 5 strains have the same sequence as A/J, while 42 other strains have the same sequence as AKR/J). Clicking on the Like A or Like B will return a list of all strains having the sequence found in strain A and strain B, respectively. Clicking on the number differences will give you the positions that are different, which for Aphoh means a valine at position 17 for A/J, but an alanine for AKR/J. In case that there are more than 500 differences, the complete sequences are shown instead.

Pairwise comparison

Strain A:

Strain B:

Optional:

chromosome

Search

Copy Excel CSV PDF Search:

Transcript	Ensg	Symbol	Strain A	Strain B	# diffs	Like A	Like B
ENSMUST00000000049	ENSMUSG00000000049	ApoH	A_J	AKR_J	1	30	1
ENSMUST00000000095	ENSMUSG00000000093	Tbx2	A_J	AKR_J	1	27	5
ENSMUST00000000096	ENSMUSG00000000094	Tbx4	A_J	AKR_J	2	24	4
ENSMUST00000000109	ENSMUSG00000055022	Cntn1	A_J	AKR_J	1	9	21
ENSMUST00000000206	ENSMUSG00000000202	Btbd17	A_J	AKR_J	1	4	29
ENSMUST00000000266	ENSMUSG00000026535	Ifi202b	A_J	AKR_J	3	1	20
ENSMUST00000000310	ENSMUSG00000000301	Pemt	A_J	AKR_J	1	12	4
ENSMUST00000000451	ENSMUSG00000000441	Raf1	A_J	AKR_J	2	25	9
ENSMUST00000000466	ENSMUSG00000028494	Plin2	A_J	AKR_J	1	16	14
ENSMUST00000000542	ENSMUSG00000000530	Acvrl1	A_J	AKR_J	2	12	14

Showing 1 to 10 of 7,002 entries

Previous **1** 2 3 4 5 ... 701 Next

```
Pos A_J AKR_J
17 A V
```

Figure 10: Pairwise comparisons, here A/J vs AKR/J