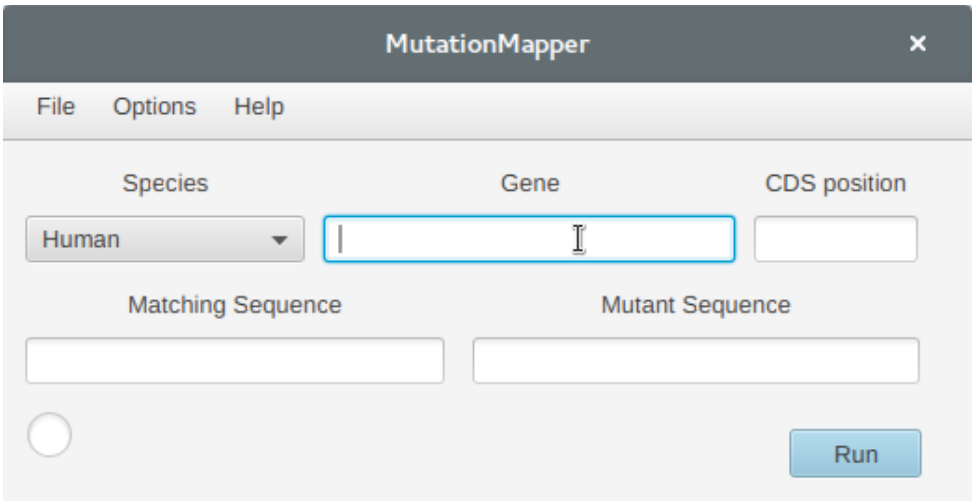
MutationMapper Instructions

MutationMapper is built to help simplify determining the functional consequences of mutations discovered using low-throughput sequencing methods (e.g. Sanger sequencing) and to map CDS coordinates to genomic positions. An **internet connection is required** in order to connect to Ensembl's REST API.

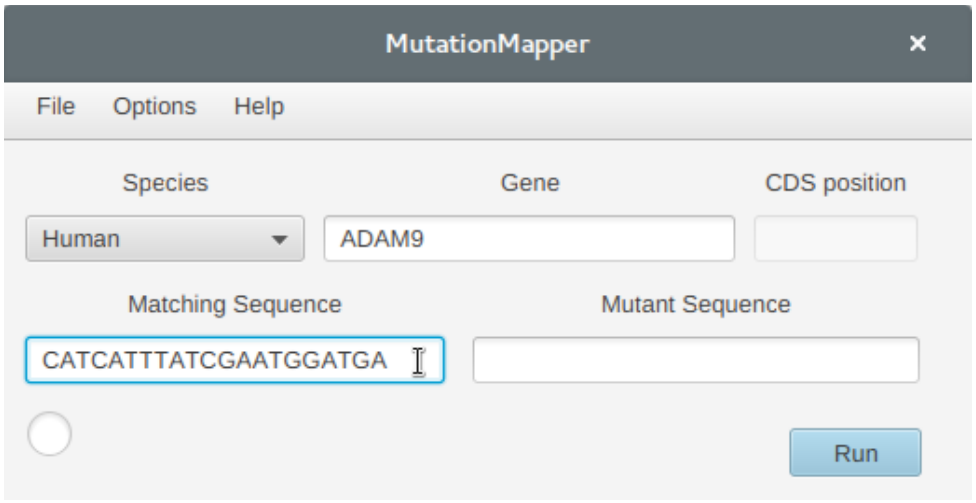


After choosing the desired species using the 'Species' dropdown box, you may enter either a gene symbol, Ensembl gene ID or Ensembl transcript ID to identify the desired gene or transcript to search. RefSeq IDs can also be used to attempt to search for the corresponding Ensembl transcript.

You may search for either the position of a matching sequence DNA sequence or for a CDS position. Typing in the 'Matching Sequence' text field disables the 'CDS position' text field and vice versa. To enable the 'CDS position' text field again clear any text in the 'Matching Sequence' text field and similarly, if there is text in the 'Matching Sequence' text field clear this text to re-enable the 'CDS position' text field.

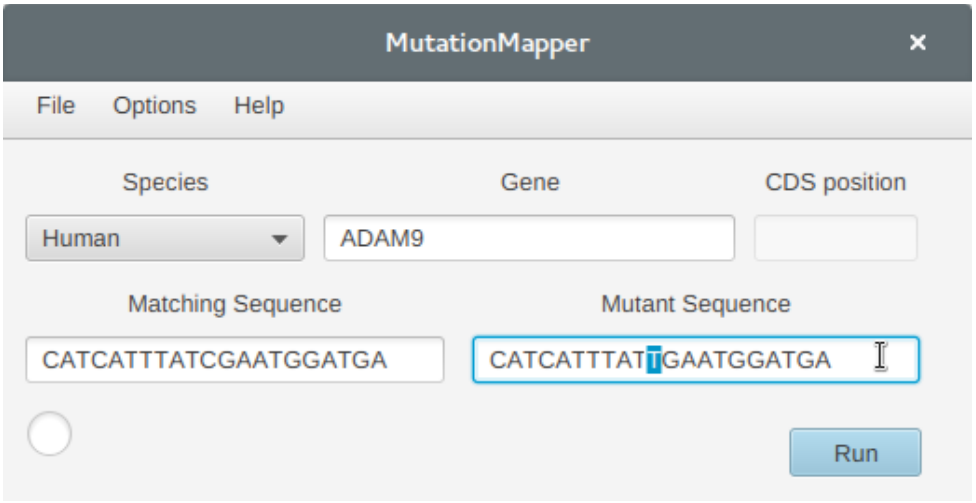
Mapping a Mutation using Matching and Mutant Sequence

The primary purpose of MutationMapper is to use short reference sequences and a mutant sequence (e.g. identified through Sanger sequencing) to determine the functional consequence of a mutation.

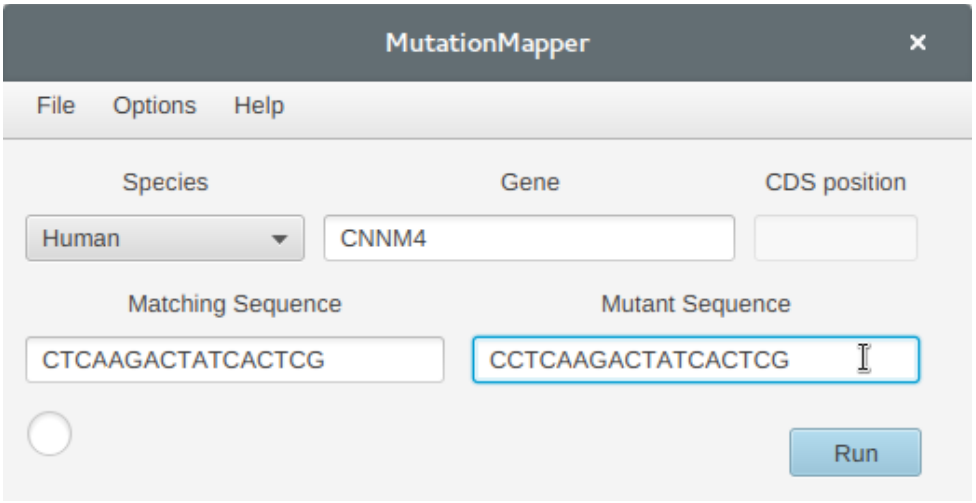


In the example above the position of the DNA in the 'Matching Sequence' text field will be searched against all transcripts of the human gene *ADAM9*. The genomic DNA sequence will be used to search for the matching sequence from the transcription start site to the transcript termination site. The **genomic and CDS positions** of any matching transcripts will be reported in the results.

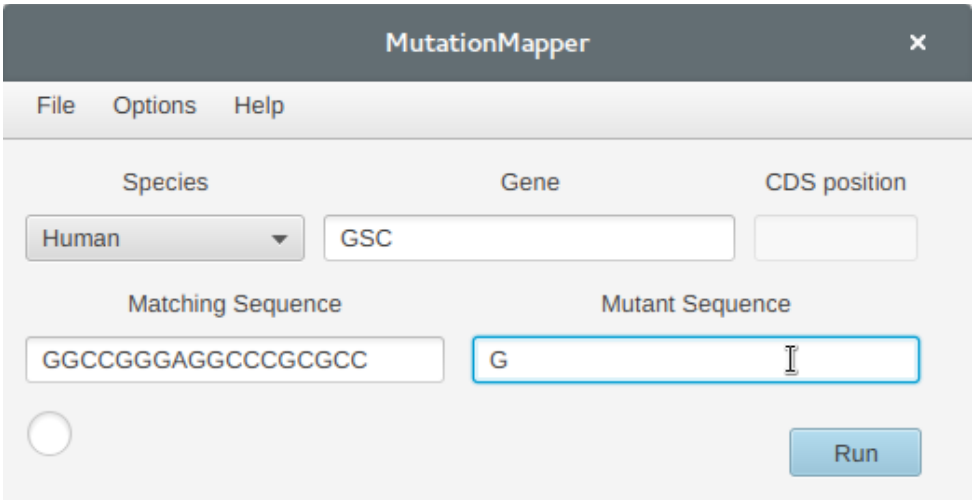
To determine the consequence of a mutation paste the mutant sequence into the 'Mutant Sequence' text field as shown below. You must ensure that the **only difference between the matching and mutant sequences is the mutation you want to determine the consequence of**. Any differences between the two sequences (e.g. if one sequence is longer than the other) will be interpreted as a mutation. In the example below, the consequences of a **single nucleotide variant** (SNV) is determined. The position of that SNV is highlighted in the mutant sequence.



In the example below, the consequence of an **insertion** is being determined. Note that the mutant sequence has an extra 'C' at the beginning.

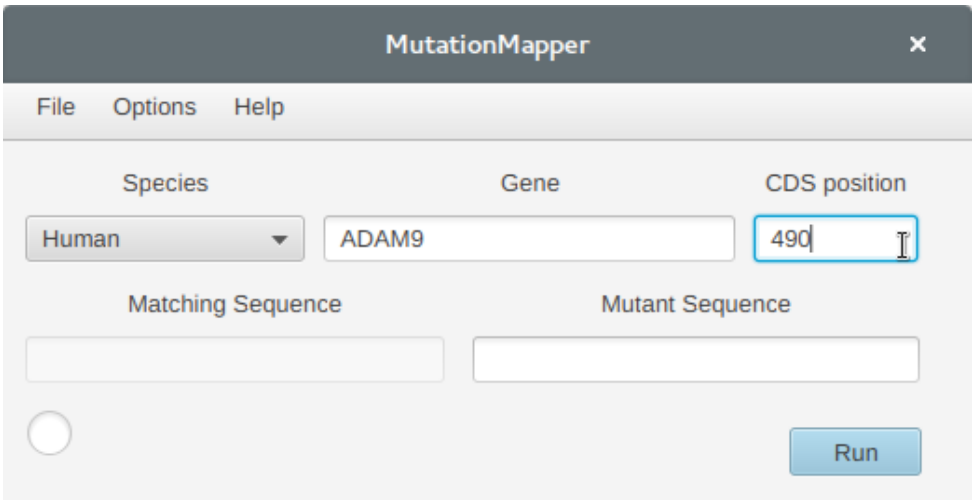


In the last example shown below, a **deletion** of 17 bp is being modelled.

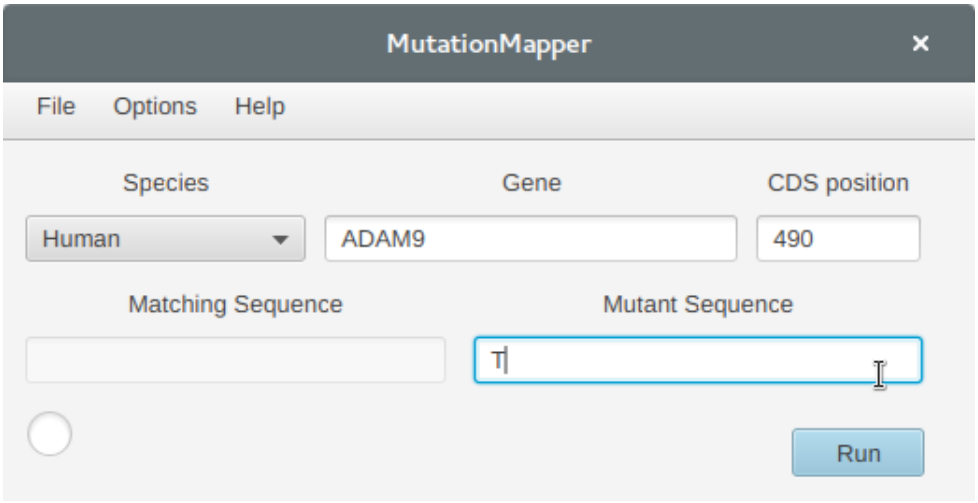


Mapping a Mutation using CDS Coordinates

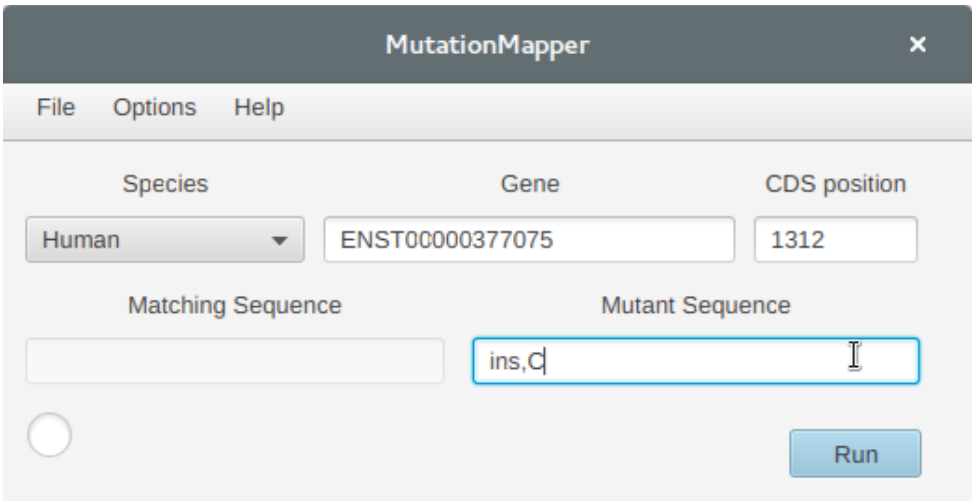
The examples below show equivalent uses where instead of using a matching and mutant sequence to determine the consequence of a variant, a CDS position can be used.



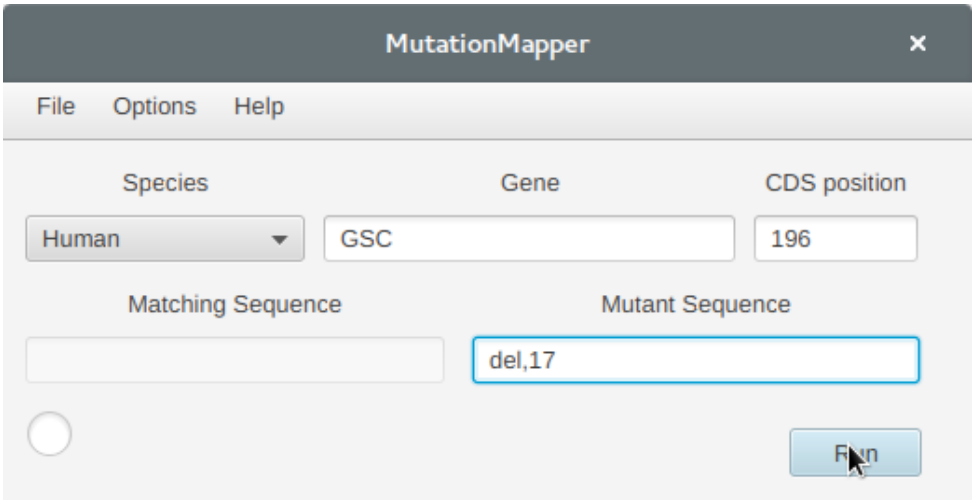
In the above example the genomic position of coding position 490 will be determined for all transcripts of the gene *ADAM9*. In this simplest use case, the genomic coordinate of a given transcript position can be determined. To determine the consequence of a substitution we can enter the mutant base in the 'Mutant Sequence' box as show below:



This example will give report the results of mutating the 490th coding nucleotide of the relevant transcripts to T (i.e. c.490C>T). Entering more than one nucleotide in the 'Mutant Sequence' box will model multi-nucleotide variants (MNVs) - i.e. substitutions of several adjacent nucleotides. **To model insertions or deletions, precede your mutation with either 'ins,' or 'del,' respectively.** In the example below, the insertion of a 'C' after CDS position 1312 is modelled for a single Ensembl transcript.



To model a deletion you can enter 'del,' and the number of nucleotides deleted or the sequence of the deleted nucleotides (these nucleotides won't be checked to see if they match the reference sequence, they will only be used to determine the length of the deletion). The example below shows how to determine the consequence of a 17 bp deletion at CDS position 196 (c.196-212del).

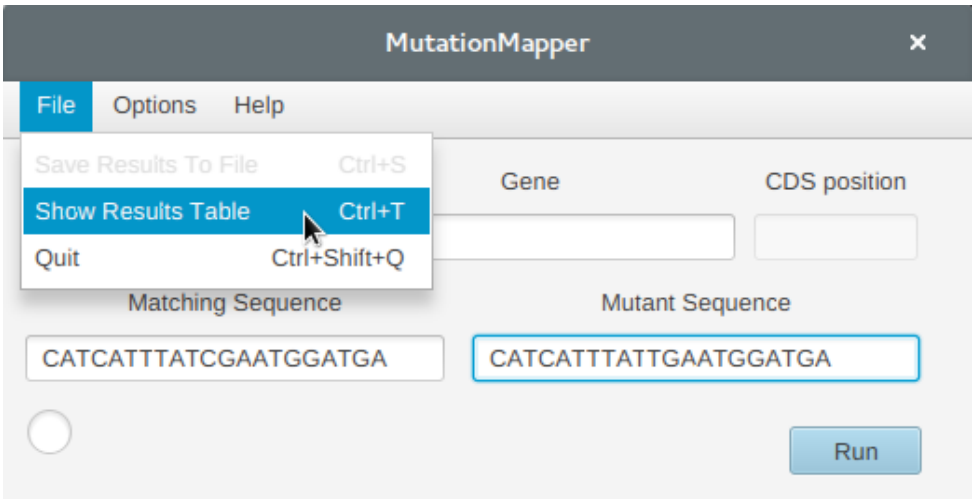


You may also use **intronic coordinates** in CDS positions if your mutation of interest is in an intron. For example, if you want to determine the result of mutating first base preceding an exon you could put something like "100-1" in the 'CDS position' text field where 100 is the CDS coordinate of the first base of the exon. Similarly if the mutation of interest is two bases after an exon you could use something like "150+2" where 150 is the CDS coordinate of the last base the exon. MutationMapper **will not check whether the CDS coordinate used is corresponds to the first/last base of an exon**, it will merely shift the

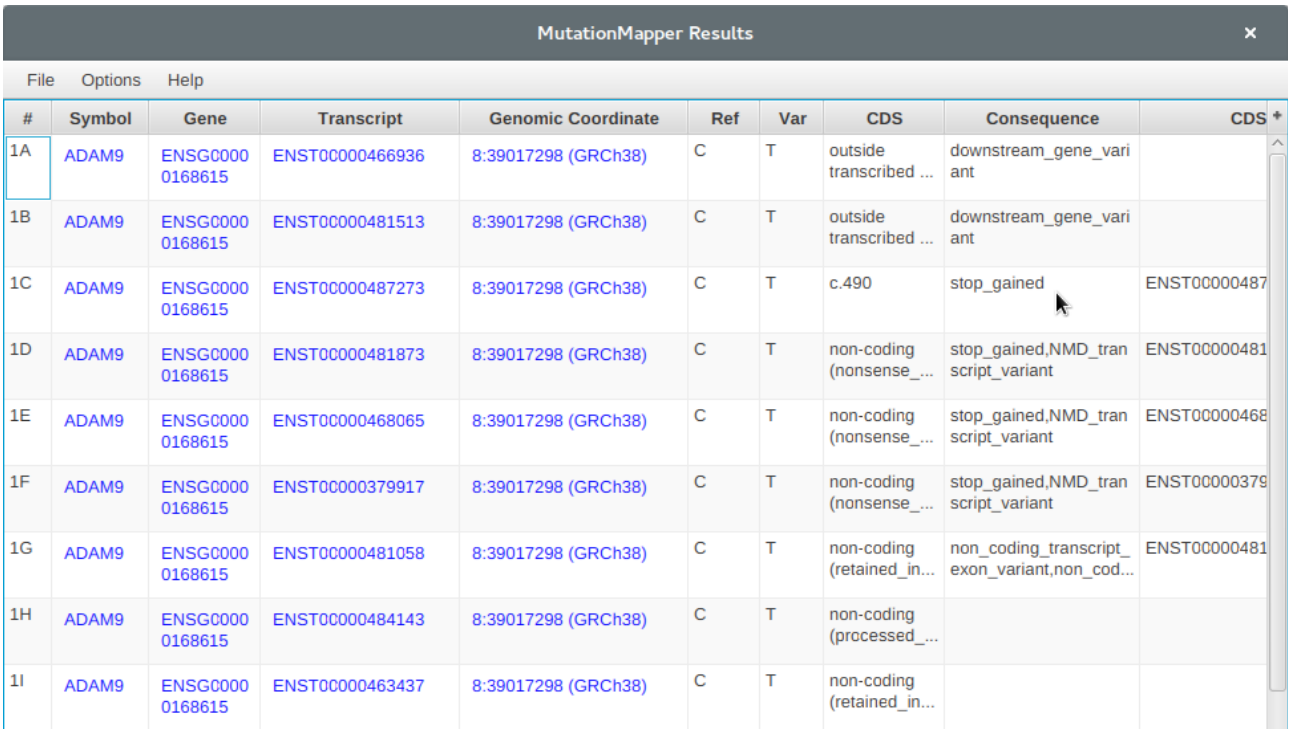
genomic coordinate mutated up or downstream relative to the CDS position according to the orientation of the gene.

Viewing Results

Results for runs are kept in a table which appears after each run. If you close the window after a run you can open it again using the 'Show Results Table' menu item.

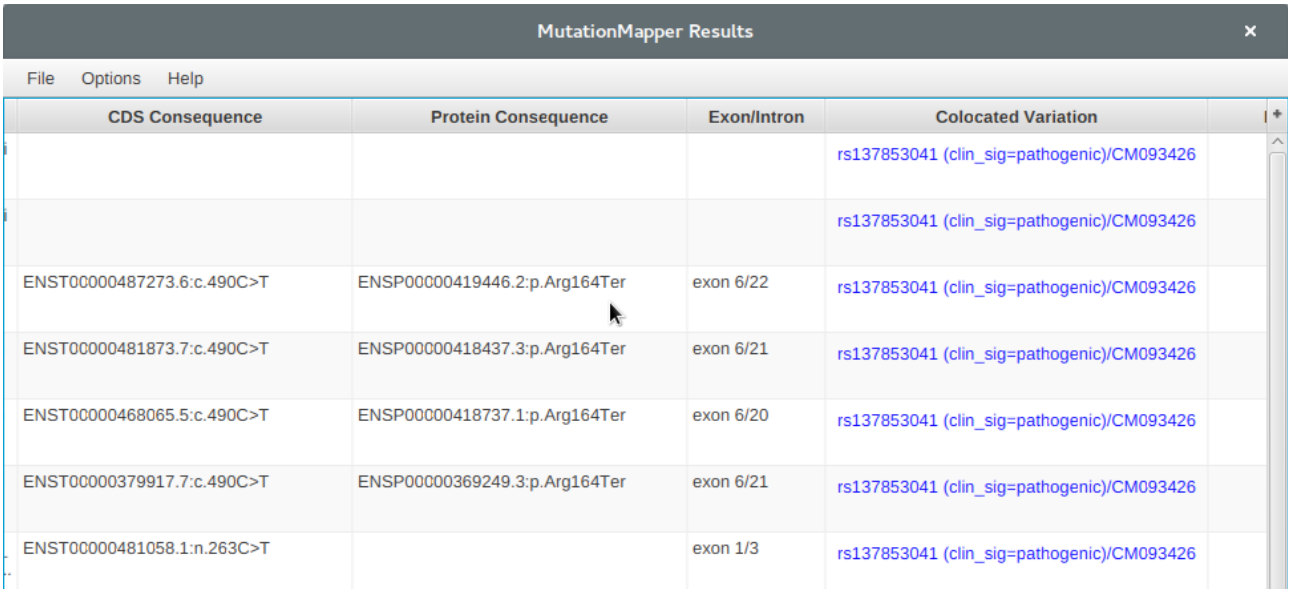


The image below shows part of a results table for a single run (in fact for the first example of an SNV in *ADAM9*). When mapping sequences *without a mutant sequence*, the **'Genomic Coordinate' column** gives the start and end coordinates of the matching sequence used. However, *if a mutant sequence is used,* this column reports the position of the 'Reference' allele, as shown in the 'Ref' column. Similarly, the 'CDS' columns gives the start and end CDS coordinates of the matching sequence for each transcript (if the sequence has mapped to a coding region) or of the 'Reference' allele if a mutant sequence was used. The 'Ref' and 'Var' columns give the reference and variant alleles for a mutation reduced to its simplest possible representation - these are genomic alleles on the '+' strand. The 'Consequence' column gives the functional consequence of the mutation for each transcript as determined by Ensembl's Variant Effect Predictor (http://www.ensembl.org/info/docs/tools/vep/index.html).

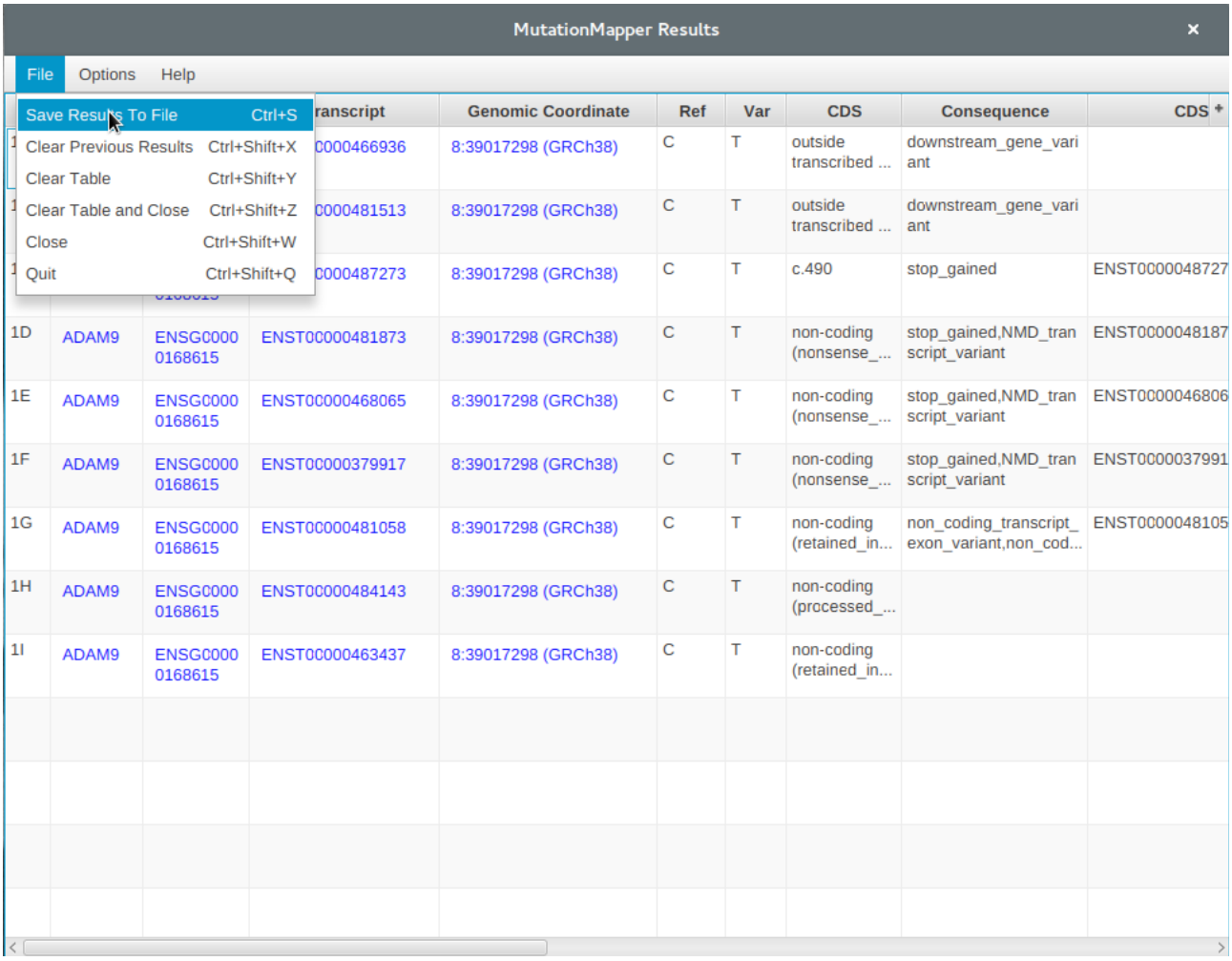


#	Symbol	Gene	Transcript	Genomic Coordinate	Ref	Var	CDS	Consequence	CDS
1A	ADAM9	ENSG0000 0168615	ENST00000466936	8:39017298 (GRCh38)	C	T	outside transcribed ...	downstream_gene_variant	
1B	ADAM9	ENSG0000 0168615	ENST00000481513	8:39017298 (GRCh38)	C	T	outside transcribed ...	downstream_gene_variant	
1C	ADAM9	ENSG0000 0168615	ENST00000487273	8:39017298 (GRCh38)	C	T	c.490	stop_gained	ENST00000487
1D	ADAM9	ENSG0000 0168615	ENST00000481873	8:39017298 (GRCh38)	C	T	non-coding (nonsense_...	stop_gained,NMD_transcript_variant	ENST00000481
1E	ADAM9	ENSG0000 0168615	ENST00000468065	8:39017298 (GRCh38)	C	T	non-coding (nonsense_...	stop_gained,NMD_transcript_variant	ENST00000468
1F	ADAM9	ENSG0000 0168615	ENST00000379917	8:39017298 (GRCh38)	C	T	non-coding (nonsense_...	stop_gained,NMD_transcript_variant	ENST00000379
1G	ADAM9	ENSG0000 0168615	ENST00000481058	8:39017298 (GRCh38)	C	T	non-coding (retained_in...	non_coding_transcript_exon_variant,non_cod...	ENST00000481
1H	ADAM9	ENSG0000 0168615	ENST00000484143	8:39017298 (GRCh38)	C	T	non-coding (processed_...		
1I	ADAM9	ENSG0000 0168615	ENST00000463437	8:39017298 (GRCh38)	C	T	non-coding (retained_in...		

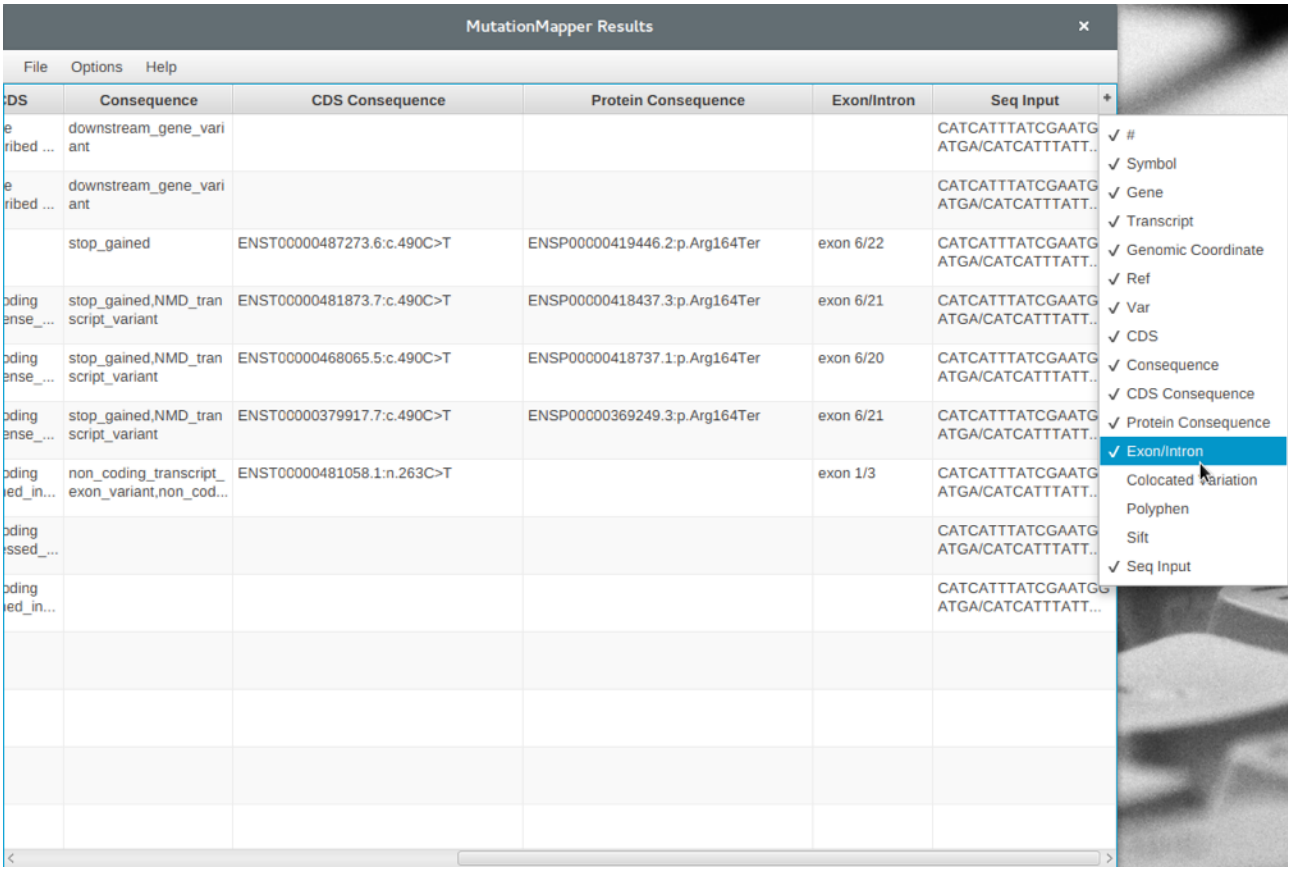
Other columns give the HGVS names for both the CDS and protein consequences of a mutation as well as Polyphen and SIFT predictions for missense variants. The ‘Colocated Variation’ column indicates whether any known variation lies at the same site as the mutation (NB the reference and mutant alleles are not checked to see if they match with any collocated variants).



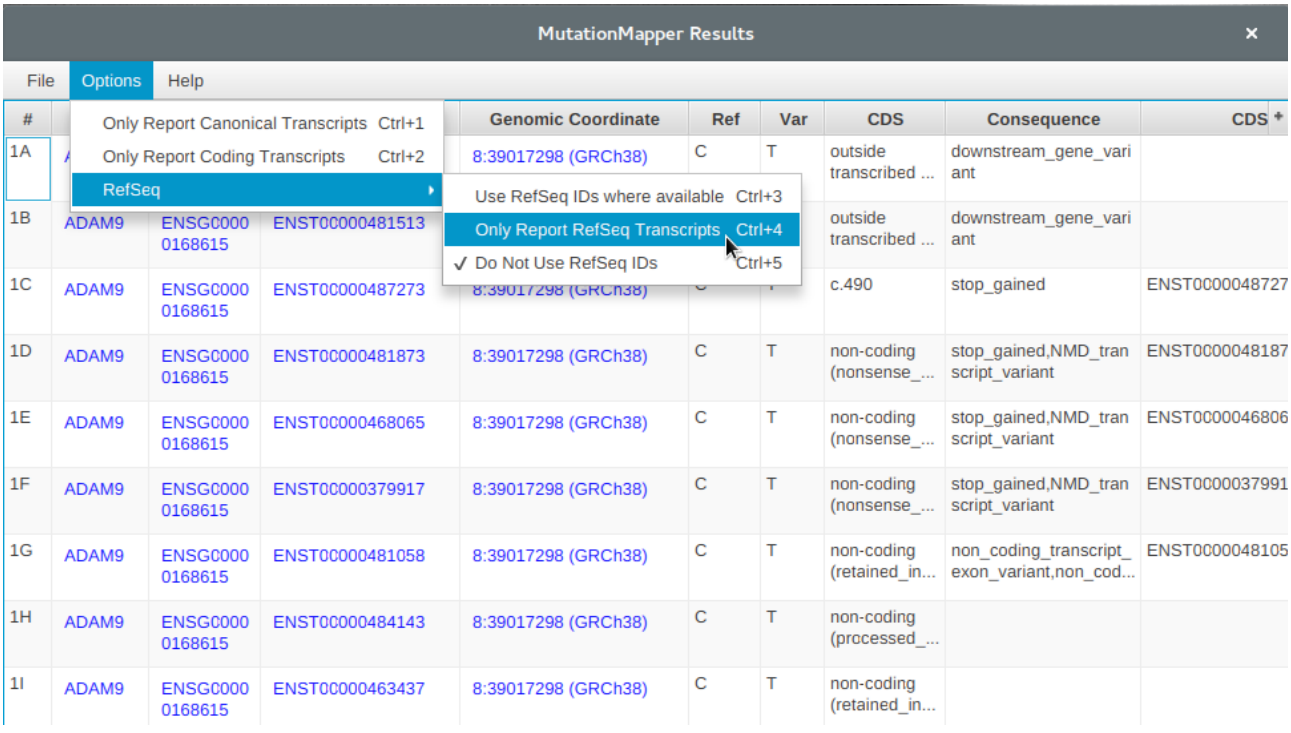
The results shown in the table can be saved using the ‘File’ menu as below. Results can also be cleared to only show the results from the last run or cleared completely to empty the table.

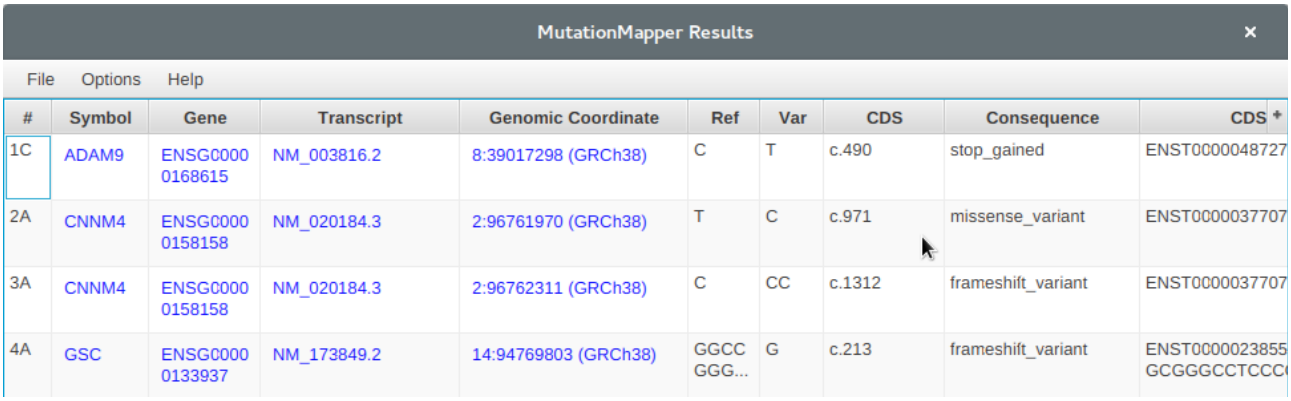


You can choose which columns are shown in the table using the '+' symbol at the edge of the table.

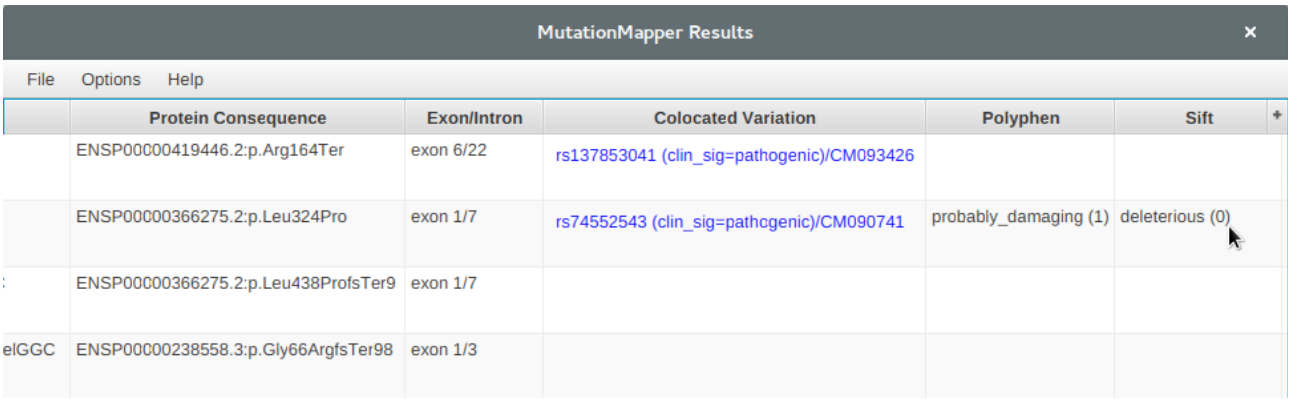


The 'Options' menu can be used to display only information from canonical transcripts, coding transcripts or transcripts with a RefSeq ID. There is also an option to show the RefSeq transcript ID instead of the Ensembl transcript ID if available.





In the example above several runs are shown where mutation consequences are only shown for canonical transcripts. Below shows more columns for the same runs, with an example of Polyphen and SIFT output.



GRCh37

For Human sequences, you can choose to use the GRCh37 reference (also equivalent to hg19) instead of the default GRCh38 using the options menu on the main window. Reference choices are not available for other species.

Credit

MutationMapper was written by David A. Parry and is available from:

https://github.com/gantzgraf/MutationMapper

or alternatively:

https://sourceforge.net/projects/MutationMapper/

It was originally available as a perl script and a perl/perl + objective C based GUI application for Windows and Mac OS X. This version is a complete rewrite using java and is available for Windows, Mac OS X and linux.

If you use MutationMapper for primer designs that are used in published work, please cite the URL 'https://github.com/gantzgraf/MutationMapper'.

License