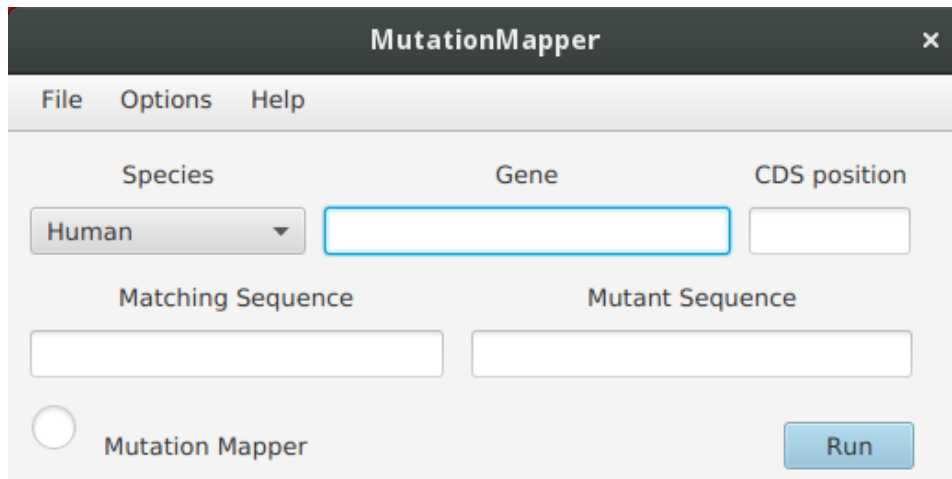


# MutationMapper Instructions

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.



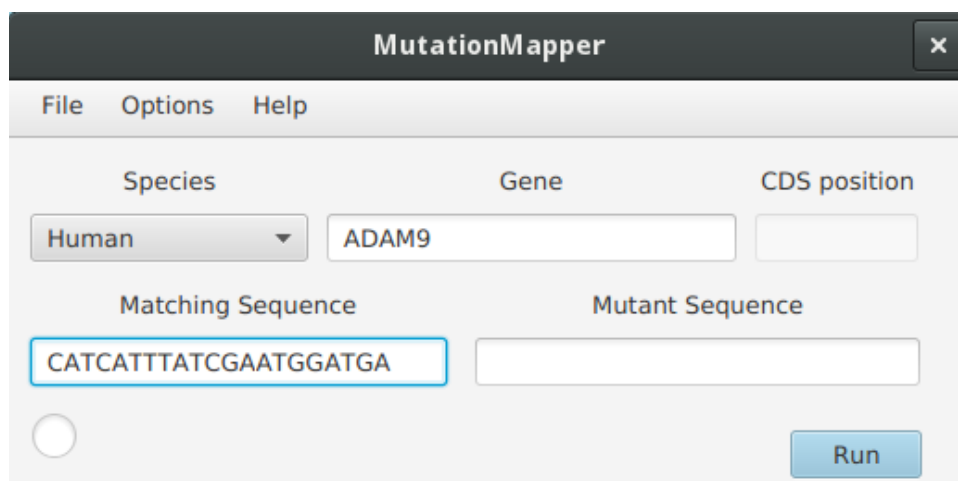
The screenshot shows the MutationMapper application window. It has a title bar with 'MutationMapper' and a close button. Below the title bar is a menu bar with 'File', 'Options', and 'Help'. The main area contains three input fields: 'Species' (a dropdown menu showing 'Human'), 'Gene' (an empty text box), and 'CDS position' (an empty text box). Below these are two more text boxes: 'Matching Sequence' and 'Mutant Sequence'. At the bottom left is a radio button labeled 'Mutation Mapper', and at the bottom right is a blue 'Run' button.

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.



This screenshot shows the MutationMapper application window with the 'Matching Sequence' field populated with the text 'CATCATTTATCGAATGGATGA'. The 'Species' dropdown remains 'Human', and the 'Gene' field now contains 'ADAM9'. The 'CDS position' field is still empty. The 'Run' button is still present at the bottom right.

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.

The screenshot shows the MutationMapper application window. The title bar is 'MutationMapper' with a close button. The menu bar contains 'File', 'Options', and 'Help'. The main interface has three input fields at the top: 'Species' (a dropdown menu set to 'Human'), 'Gene' (a text box containing 'ADAM9'), and 'CDS position' (an empty text box). Below these are two text boxes: 'Matching Sequence' containing 'CATCATTTATCGAATGGATGA' and 'Mutant Sequence' containing 'CATCATTTATGAATGGATGA'. The 'G' at the 10th position of the mutant sequence is highlighted with a blue box. At the bottom left is a radio button, and at the bottom right is a 'Run' button.

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

The screenshot shows the MutationMapper application window. The title bar is 'MutationMapper' with a close button. The menu bar contains 'File', 'Options', and 'Help'. The main interface has three input fields at the top: 'Species' (a dropdown menu set to 'Human'), 'Gene' (a text box containing 'CNNM4'), and 'CDS position' (an empty text box). Below these are two text boxes: 'Matching Sequence' containing 'CTCAAGACTATCACTCG' and 'Mutant Sequence' containing 'CCTCAAGACTATCACTCG'. The 'C' at the beginning of the mutant sequence is highlighted with a blue box. At the bottom left is a radio button, and at the bottom right is a 'Run' button.

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

The screenshot shows the MutationMapper application window. The 'Species' dropdown is set to 'Human'. The 'Gene' text box contains 'GSC'. The 'CDS position' text box is empty. The 'Matching Sequence' text box contains 'GGCCGGGAGGCCCGCGCC'. The 'Mutant Sequence' text box contains 'G'. A radio button is selected next to the 'Mutation Mapper' label. The 'Run' button is visible at the bottom right.

## Mapping a Mutation using CDS Coordinates

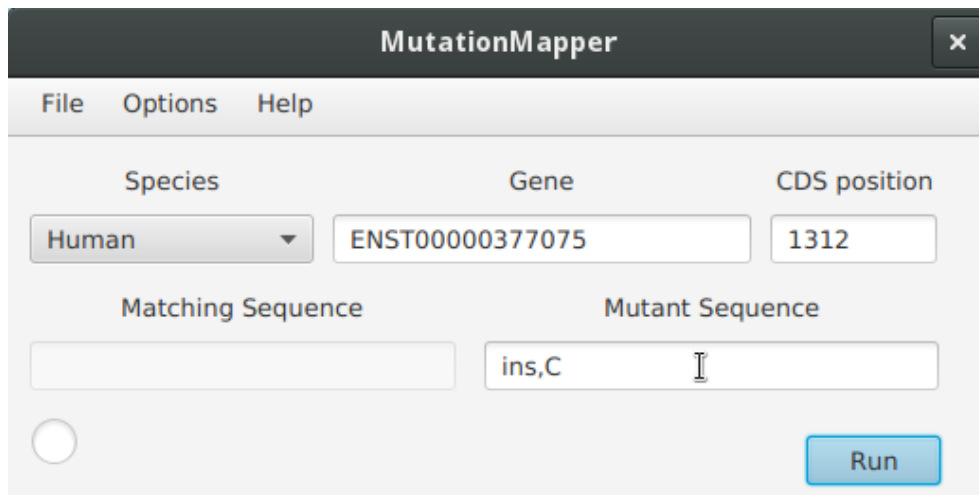
The examples below show equivalent uses where instead of using a matching and mutant sequence to determine the consequence of a variant, a CDS position can be used.

The screenshot shows the MutationMapper application window. The 'Species' dropdown is set to 'Human'. The 'Gene' text box contains 'ADAM9'. The 'CDS position' text box contains '490'. The 'Matching Sequence' and 'Mutant Sequence' text boxes are empty. A radio button is selected next to the 'Mutation Mapper' label. The 'Run' button is visible at the bottom right.

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:

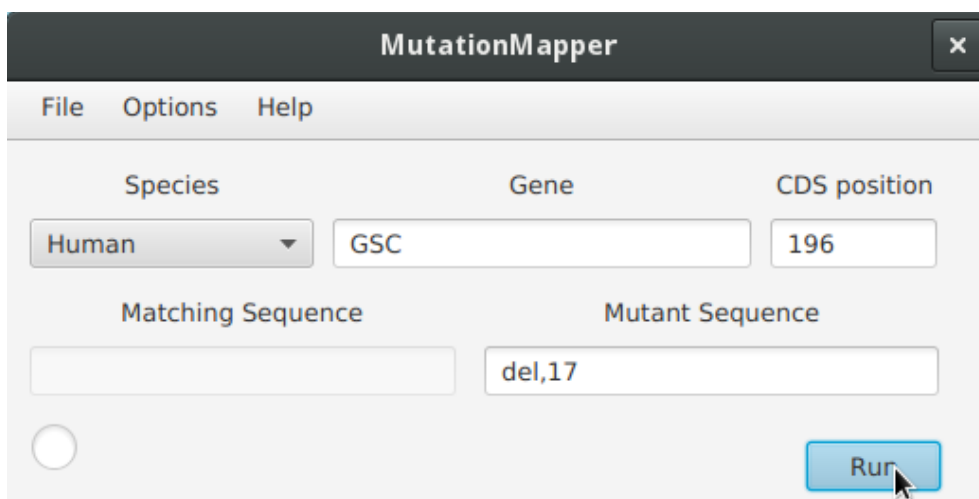
The screenshot shows the MutationMapper application window. The 'Species' dropdown is set to 'Human'. The 'Gene' text box contains 'ADAM9'. The 'CDS position' text box contains '490'. The 'Matching Sequence' text box is empty. The 'Mutant Sequence' text box contains 'T'. A radio button is selected next to the 'Mutation Mapper' label. The 'Run' button is visible at the bottom right.

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.



The screenshot shows the MutationMapper application window. The title bar is "MutationMapper" with a close button. The menu bar contains "File", "Options", and "Help". The main interface has three input fields at the top: "Species" (a dropdown menu set to "Human"), "Gene" (a text box containing "ENST00000377075"), and "CDS position" (a text box containing "1312"). Below these are two text boxes: "Matching Sequence" (empty) and "Mutant Sequence" (containing "ins,C" with a cursor at the end). At the bottom left is a radio button, and at the bottom right is a blue "Run" button.

For modelling 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).



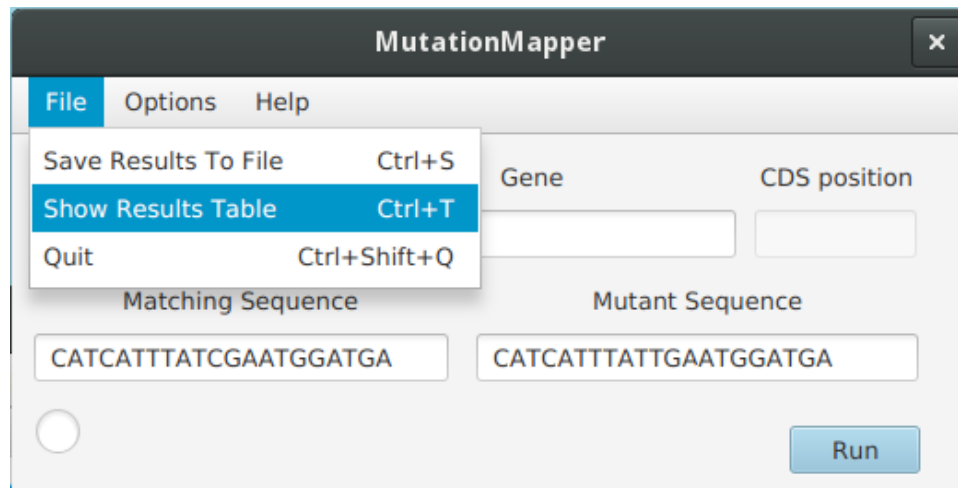
The screenshot shows the MutationMapper application window. The title bar is "MutationMapper" with a close button. The menu bar contains "File", "Options", and "Help". The main interface has three input fields at the top: "Species" (a dropdown menu set to "Human"), "Gene" (a text box containing "GSC"), and "CDS position" (a text box containing "196"). Below these are two text boxes: "Matching Sequence" (empty) and "Mutant Sequence" (containing "del,17"). At the bottom left is a radio button, and at the bottom right is a blue "Run" button with a mouse cursor hovering over it.

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*). The 'Genomic Coordinate' column gives the start and end coordinates of the matching sequence used as well as the genome reference build used. 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). The 'Ref' and 'Var' columns give the reference and variant alleles for a mutation - 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>).

MutationMapper Results									
File Options Help									
#	Symbol	Gene	Transcript	Genomic Coordinate	Ref	Var	CDS	Consequence	+
1A	ADAM9	ENSG00000168615	ENST00000466936	8:39017288-390173...	C	T	outside transcribe...	downstream_gene_v	
1B	ADAM9	ENSG00000168615	ENST00000481513	8:39017288-390173...	C	T	outside transcribe...	downstream_gene_v	
1C	ADAM9	ENSG00000168615	ENST00000487273	8:39017288-390173...	C	T	c.480-c.500	stop_gained	ENST0
1D	ADAM9	ENSG00000168615	ENST00000481873	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra	ENST0
1E	ADAM9	ENSG00000168615	ENST00000468065	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra	ENST0
1F	ADAM9	ENSG00000168615	ENST00000379917	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra	ENST0
1G	ADAM9	ENSG00000168615	ENST00000481058	8:39017288-390173...	C	T	non-coding (retained_i...	non_coding_transcrip	ENST0
1H	ADAM9	ENSG00000168615	ENST00000484143	8:39017288-390173...	C	T	non-coding (processe...		
1I	ADAM9	ENSG00000168615	ENST00000463437	8:39017288-390173...	C	T	non-coding (retained_i...		

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 colocated variants).

MutationMapper Results				
File Options Help				
CDS Consequence	Protein Consequence	Exon/Intron	Colocated Variation	+
			rs137853041 (clin_sig=pathogenic)/CM093426	
			rs137853041 (clin_sig=pathogenic)/CM093426	
ENST00000487273.4:c.490C>T	ENSP00000419446.2:p.Arg164Ter	exon 6/22	rs137853041 (clin_sig=pathogenic)/CM093426	
a ENST00000481873.5:c.490C>T	ENSP00000418437.3:p.Arg164Ter	exon 6/21	rs137853041 (clin_sig=pathogenic)/CM093426	
a ENST00000468065.3:c.490C>T	ENSP00000418737.1:p.Arg164Ter	exon 6/20	rs137853041 (clin_sig=pathogenic)/CM093426	
a ENST00000379917.5:c.490C>T	ENSP00000369249.3:p.Arg164Ter	exon 6/21	rs137853041 (clin_sig=pathogenic)/CM093426	
o ENST00000481058.1:n.263C>T		exon 1/3	rs137853041 (clin_sig=pathogenic)/CM093426	

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.

MutationMapper Results									
File Options Help									
			Transcript	Genomic Coordinate	Ref	Var	CDS	Consequence	+
			ENST00000466936	8:39017288-390173...	C	T	outside transcribe...	downstream_gene_v...	
			ENST00000481513	8:39017288-390173...	C	T	outside transcribe...	downstream_gene_v...	
			ENST00000487273	8:39017288-390173...	C	T	c.480-c.500	stop_gained	ENSTO
1D	ADAM9	ENSG00000168615	ENST00000481873	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra...	ENSTO
1E	ADAM9	ENSG00000168615	ENST00000468065	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra...	ENSTO
1F	ADAM9	ENSG00000168615	ENST00000379917	8:39017288-390173...	C	T	non-coding (nonsense...	stop_gained,NMD_tra...	ENSTO
1G	ADAM9	ENSG00000168615	ENST00000481058	8:39017288-390173...	C	T	non-coding (retained_i...	non_coding_transcrip...	ENSTO
1H	ADAM9	ENSG00000168615	ENST00000484143	8:39017288-390173...	C	T	non-coding (processe...		
1I	ADAM9	ENSG00000168615	ENST00000463437	8:39017288-390173...	C	T	non-coding (retained_i...		

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

MutationMapper Results						
File Options Help						
CDS	Consequence	CDS Consequence	Protein Consequence	Exon/Intron	Seq Input	+
10-c.500	stop_gained	ENST00000487273.4:c.490C>T	ENSP00000419446.2:p.Arg164Ter	exon 6/22	CATCATTATCGAATG ATGA/CATCATTATT..	✓ # ✓ Symbol ✓ Gene ✓ Transcript ✓ Genomic Coordinate ✓ Ref ✓ Var ✓ CDS ✓ Consequence ✓ CDS Consequence ✓ Protein Consequence ✓ Exon/Intron Colocated Variation Polyphen Sift ✓ Seq Input
	missense_variant	ENST00000377075.2:c.971T>C	ENSP00000366275.2:p.Leu324Pro	exon 1/7		
2	frameshift_variant	ENST00000377075.2:c.1312dupC	ENSP00000366275.2:p.Leu438ProfsTer9	exon 1/7		
	frameshift_variant	ENST00000238558.4:c.196_212delG GCGCGGGCCTCCCGGC	ENSP00000238558.3:p.Gly66ArgfsTer98	exon 1/3		
13-c.196	frameshift_variant	ENST00000238558.4:c.196_212delG GCGCGGGCCTCCCGGC	ENSP00000238558.3:p.Gly66ArgfsTer98	exon 1/3	GGCCGGGAGGCCCGG GCC/G	
2	frameshift_variant	ENST00000377075.2:c.1313_1332del ITCAAGACTATCACTCGCTTC	ENSP00000366275.2:p.Lys439Ter	exon 1/7		
8	frameshift_variant	ENST00000377075.2:c.1309_1328del ICCCCTCAAGACTATCACTCG	ENSP00000366275.2:p.Pro437LeufsTer3	exon 1/7		
112-c.13	frameshift_variant	ENST00000377075.2:c.1312dupC	ENSP00000366275.2:p.Leu438ProfsTer9	exon 1/7	CTCAAGACTATCACTC /CCTCAAGACTATCAC	

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.

MutationMapper Results									
File Options Help									
#	S	Genomic Coordinate		Ref	Var	CDS	Consequence		
1A	A	8:39017288-390173...		C	T	outside transcribe...	downstream_gene_v...		
1B	ADAM9	ENSG00000168615	ENST00000487273			outside transcribe...	downstream_gene_v...		
1C	ADAM9	ENSG00000168615	ENST00000487273			c.480-c.500	stop_gained	ENST0	
1D	ADAM9	ENSG00000168615	ENST00000481873			non-coding (nonsense...	stop_gained,NMD_tra...	ENST0	
1E	ADAM9	ENSG00000168615	ENST00000468065			non-coding (nonsense...	stop_gained,NMD_tra...	ENST0	
1F	ADAM9	ENSG00000168615	ENST00000379917			non-coding (nonsense...	stop_gained,NMD_tra...	ENST0	
1G	ADAM9	ENSG00000168615	ENST00000481058			non-coding (retained_i...	non_coding_transcrip...	ENST0	
1H	ADAM9	ENSG00000168615	ENST00000484143			non-coding (processe...			
1I	ADAM9	ENSG00000168615	ENST00000463437			non-coding (retained_i...			



MutationMapper Results									
File Options Help									
#	Symbol	Gene	Transcript	Genomic Coordinate	Ref	Var	CDS	Consequence	+
1C	ADAM9	ENSG00000168615	ENST00000487273	8:39017288-390173...	C	T	c.480-c.500	stop_gained	ENST0
2A	CNNM4	ENSG00000158158	ENST00000377075	2:96761970 (GRCh38)	T	C	971	missense_variant	ENST0
3A	CNNM4	ENSG00000158158	ENST00000377075	2:96762311 (GRCh38)	C	CC	1312	frameshift_variant	ENST0
4A	GSC	ENSG00000133937	ENST00000238558	14:94769820 (GRCh...	GGCC GGG...	G	196	frameshift_variant	ENST0 GCGCC

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.

MutationMapper Results					
File Options Help					
Protein Consequence	Exon/Intron	Colocated Variation	Polyphen	Sift	+
ENSP00000419446.2:p.Arg164Ter	exon 6/22	rs137853041 (clin_sig=pathogenic)/CM093426			C A
ENSP00000366275.2:p.Leu324Pro	exon 1/7	rs74552543 (clin_sig=pathogenic)/CM090741	probably_damaging (1)	deleterious (0)	
ENSP00000366275.2:p.Leu438ProfsTer9	exon 1/7				
ENSP00000238558.3:p.Gly66ArgfsTer98	exon 1/3				

## 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

MutationMapper is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <<http://www.gnu.org/licenses/>>.