

# UMD

The Universal Mutation Database

## UMD MANUAL

*June 2013 version*



Faculté  
de Médecine  
Aix-Marseille Université



**Inserm**

Institut national  
de la santé et de la recherche médicale

UMD is used by some major international projects in the field of human mutations:



# Table of Contents

I- DOWNLOAD OF THE UMD SOFTWARE.....	6
I-a How to install your UMD software? .....	6
I-b How to update your UMD software? .....	6
II- CREATE A NEW UMD-LSDB .....	9
II-a Set-up your own mutation database using the UMD® software?.....	9
II-b Integrate intronic sequences and SNPs from a genomic contig.....	11
III- HOW TO ENTER A NEW MUTATION? .....	16
III-a - Point mutations .....	17
1) c.9568C>T (p.Arg3190X) .....	17
2) c.3935T>A (p.leu1312X) .....	18
3) c.4996C>T (p.Arg1666X) .....	18
III-b- Small deletions .....	19
1) c.372delG.....	19
2) c.481_483delACC .....	20
4) c.3010_3011delAA .....	21
5) c.4918delA.....	21
III-c- Insertions .....	22
1) c.9319insC .....	22
2) c.9319insAA .....	23
3) c.3267insC .....	23
4) c.2929dup.....	24
5) c.2624_2638dup.....	24
III-d- Large deletions .....	25
III-e- Large duplications .....	26
III-f- Intronic mutations .....	27
1) c.9649+1G>A (c.IVS66+1G>A) .....	27
III-g- Missense mutations .....	32
1) c.10018T>C (p.Cys3340Arg) .....	32
2) c.831G>T (p.Gln277His).....	33
IV- IMPORT MOLECULAR DATA INTO UMD® .....	35
V- IMPORT CLINICAL DATA INTO UMD®.....	37
V-a- Various options to import clinical data into UMD® .....	39
1) Manual input during mutation's record creation .....	39
2) Import of clinical data for a single patient.....	41
Symptom.....	41
Severity [Age].....	41
3) Global import of individual records .....	42
4) Global import of individual records .....	42
VI DISPLAY FORMATS.....	44
VI-a- Creation of display formats .....	44
VI-b- Modification of display formats.....	45
VII SEARCH INTERFACES .....	48
VII-a- UMD-web search interface .....	48
VII-b- Regular search interface.....	49
VIII DESCRIPTION OF ALL FUNCTIONS .....	50
VIII-a- The "File" menu.....	50
1) Import Polymorphic markers.....	50
2) Import Epitopes .....	50
3) Import Data.....	50
4) Export Data.....	50

5) Export molecular data.....	51
6) Export clinical data.....	51
7) Import molecular data.....	51
8) Import clinical data.....	51
9) Clear clinical data.....	51
10) Import clinical data sheets.....	51
11) Import clinical from patients.....	51
12) Print records.....	51
13) Modify Structure.....	51
14) Structure preferences.....	52
15) Modify Gene symbol.....	53
16) Define organism.....	53
17) Genatlas ID.....	53
18) Display Genetic Code.....	53
19) Modify sequence.....	53
20) Modify exons.....	53
22) Exon phasing.....	54
23) Display coding sequence.....	55
24) Display genomic organization.....	55
25) Display splice sites.....	56
26) Modify polymorphic markers.....	56
27) Modify epitopes.....	56
28) Modify SIFT data.....	56
29) Define links.....	58
30) Create a new database.....	58
31) Gene type.....	58
32) Update clinical data.....	58
33) Clinical data type.....	58
35) Numeric value ranges.....	59
36) Search preference.....	59
37) Splice sites calculation.....	59
38) Pathway for dnamlk application.....	59
39) Activate webserver.....	59
41) Activate specific options.....	59
43) Restricted access.....	59
44) Web license.....	59
46) Password.....	60
47) Import Passwords.....	60
48) Quit.....	60
<i>VIII-b- The "References" menu.....</i>	<i>61</i>
1) Add new reference.....	61
2) Show all references.....	61
<i>VIII-c- The "Mutations" menu.....</i>	<i>62</i>
1) Add a new record.....	62
4) Deletion analysis.....	64
5) Insertion analysis.....	65
<i>VIII-d- The "Pedigree" menu.....</i>	<i>67</i>
1) Add a new pedigree.....	67
2) Modify a pedigree.....	67
1) View-Modify mutations.....	68
2) Define a display format.....	68
3) View-Modify display formats.....	68

4) Free search .....	68
VIII-f- The "Statview" menu .....	69
1) Position .....	70
2) Potential stop codons .....	70
3) Nonsenses and read-through .....	71
4) Mutational events .....	72
5) Detailed mutational events .....	73
6) Frequency of mutations .....	73
7) Frequency of events .....	74
8) Distribution of mutations .....	74
9) Del & dup distribution .....	75
11) Mutation map [2 groups] .....	77
12) Deletion map .....	77
14) Geographic distribution .....	79
15) Binary comparison .....	80
16) Stat exons .....	81
17) Distribution by exon .....	81
18) ESE map of exon .....	82
19) Sequence analysis .....	82
20) Splice mutations .....	82
21) Missense mutations .....	83
22) All missense variations .....	84
23) Missenses and hydrophobicity .....	84
24) Structure .....	85
25) Global analysis .....	85
26) Free Graph .....	86
27) Specific tools .....	86
VIII-g- The "Phenotype" menu .....	87
1) Phenotype-genotype analysis .....	87
2) Expressivity analysis .....	88
3) Onset of symptoms .....	89
4) Survival rate .....	90
5) Define an analysis matrix .....	91
6) Modify matrices .....	91
7) Analysis using a Matrix .....	92
8) Difference - Delay Graph .....	93
9) LVEF or FVC graph .....	94
10) Weight graph .....	94
11) Graph per symptom .....	95
12) 3D graph with two symptoms .....	95
VIII-h- The "Genotype" menu .....	98
1) Genotype-Phenotype analysis .....	98
3) 3D Graph .....	100
VIII-i- The "Haplotype" menu .....	102
1) Haplotypes distribution .....	102
VIII-j- The "Introns" menu .....	106
1) Add intronic sequences .....	106
2) View-modify intronic sequences .....	106
3) Polymorphism map .....	107
4) Splice sites map .....	108
5) ESE map .....	108
6) Branch point [one intron] .....	108

7) Branch point [free selection] .....	109
8) Branch point [all introns].....	109
9) Branch point [user sequence].....	109
1) Define module .....	110
2) View-modify module.....	111
3) Module analysis .....	112
1) Add data for a variation .....	113
2) Variation data list.....	113
3) Add variation interpretation.....	114
4) Variation interpretations list .....	114
IX- CONCLUSION .....	115

## I- DOWNLOAD OF THE UMD SOFTWARE

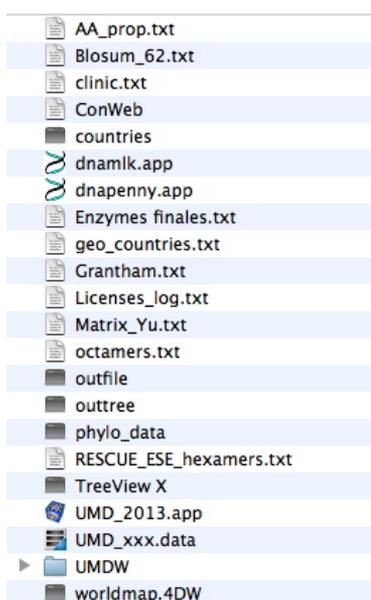
To download the UMD<sup>®</sup> software, go to the [UMD website](http://www.umd.be) at <http://www.umd.be>. On the home page, various information are given:

It was developed as a generic software to create locus-specific databases (LSDBs) with the 4th Dimension<sup>®</sup> package from 4D. The UMD software includes an optimized structure to assist and secure data entry and to allow the input of a wide range of clinical data. In addition various analyzing tools have been specifically designed to assist **clinicians** (phenotype-genotype correlations...), **geneticists** (distribution and frequency of mutations...) and **research biologists** (structural domains, molecular epidemiology...). Thanks to the flexible structure of the UMD software, it has been successfully adapted to many genes either involved in cancer (APC, BRCA1, BRCA2, TP53, RB1, MEN1, SUR1, VHL, WT1...) or in genetic diseases (FBN1, LDLR, DMD, VLCAD, MCAD, LMNA, EMD, FKRP, SGCG, SGCA, ATP7B...). This tool is freely available. To [download](#) the software please visit the download policy webpage.

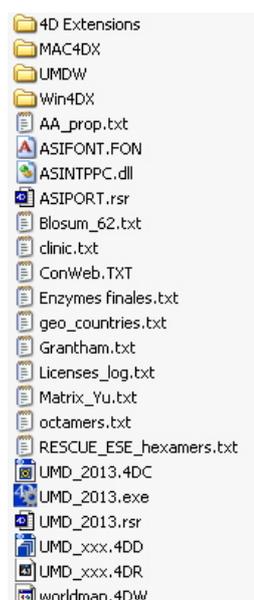
Click on the [download](#) hyperlink to access the download policy and register to the UMD software users' database. You will shortly receive an e-mail with a password that will allow you to download the various versions of UMD (Mac/PC) as well as documents.

### I-a How to install your UMD software?

Download the file of your choice and copy it to your hard drive: UMD\_2013\_PC1.zip (PC version) or UMD\_2013\_MAC.zip (MAC version). Now extract the file (double click on the corresponding zip file). You should get a directory with the followings items:



Mac directory



PC Director

Note that all files and sub-directories should remain in the same directory. In this documentation, all screen shots are from the Mac OSX version of the software but the various steps are identical for the PC version.

### I-b How to update your UMD software?

To update the UMD software, download the new UMD<sup>®</sup> software (see chapter I) and extract the file (see chapter I-a). Copy the data file (UMD\_xxx.data for Mac or UMD\_xxx.4DD and UMD\_xxx.4DR for PC) from your previous version of UMD into this new directory.

## II- How to create a new LSDB with UMD?

Before you can start the UMD software, you need to collect additional information such as the cDNA of your gene and its genomic organization. The cDNA should begin by the ATG (translation initiation site) and end by the stop codon.

The image shows a text editor window titled "MYO7A cDNA.txt" containing a long sequence of nucleotide bases (A, T, C, G). The sequence starts with "atggtgattcttcaagcaggggaccatgtgtggatggacctgagattggggcaggagtttgaogtgcacatogggggggtgg" and ends with "aagtga".

Overlaid on the right side of the text editor is a software menu with the following items:

- File
- Edit
- References
- Mutation
- Import Polymorphic markers
- Import Epitopes
- Export Data
- Print records
- Modify Structure
- Structure preferences
- Modify Gene symbol
- Modify Sequence
- Modify Exons
- Modify polymorphic markers
- Modify Epitopes
- Define links
- Create a new database
- Gene type
- Update clinical data

The "Gene type" option is highlighted in blue. At the bottom of the text editor, a status bar shows "Page 1 Sec 1 1/2 Å 2,4 cm Li 1 Col 1 0/2" and icons for "ENR" and "REV".

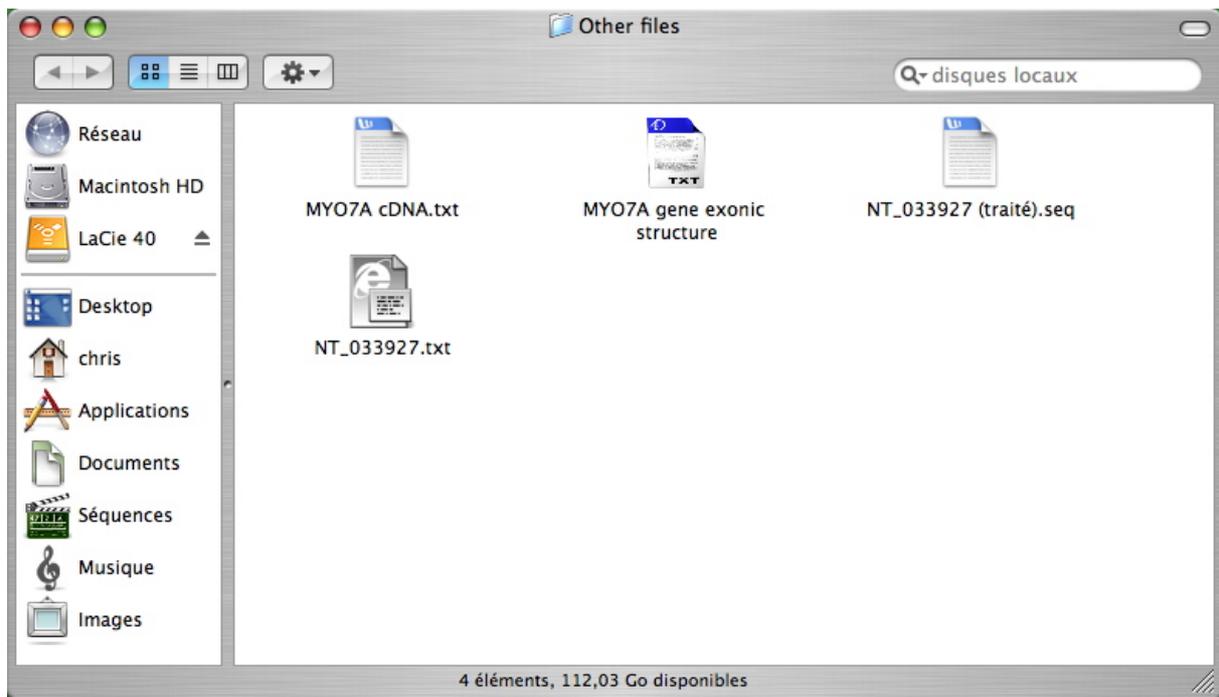
Example of the MYO7A cDNA. Note that this sequence does not include spaces, numbers or return symbols and was saved as a text file (MYO7A.txt).

You also need to know where are located each exon. For this, you need to define the first nucleotide of each exon using the cDNA reference sequence.

	A	B	C	D	E
1	1	1	18		
2	2	19	132		
3	3	133	285		
4	4	286	470		
5	5	471	590		
6	6	591	735		
7	7	736	849		
8	8	850	1003		
9	9	1004	1080		
10	10	1081	1200		
11	11	1201	1343		
12	12	1344	1554		
13	13	1555	1690		
14	14	1691	1797		
15	15	1798	1935		
16	16	1936	2094		
17	17	2095	2187		
18	18	2188	2281		
19	19	2282	2366		
20	20	2367	2586		
21	21	2587	2694		
22	22	2695	2904		
23	23	2905	3108		
24	24	3109	3285		
25	25	3286	3375		
26	26	3376	3503		
27	27	3504	3630		
28	28	3631	3750		
29	29	3751	3924		
30	30	3925	4152		
31	31	4153	4323		
32	32	4324	4441		
33	33	4442	4568		
34	34	4569	4852		
35	35	4853	5043		
36	36	5044	5168		
37	37	5169	5326		
38	38	5327	5480		
39	39	5481	5636		
40	40	5637	5742		
41	41	5743	5856		
42	42	5857	5944		
43	43	5945	6051		
44	44	6052	6237		
45	45	6238	6354		
46	46	6355	6438		
47	47	6439	6558		
48	48	6559	6648		
49					
50					
51					
52					
53					

*Exonic structure of the MYO7A cDNA. The position of each exon (first and last nucleotides) is indicated using the cDNA as reference (the A nucleotide of the ATG codon being the first nucleotide).*

With the almost completion of the human genome project, genomic contigs have been assembled and intronic sequences of most genes are available. To import these intronic sequences, you can download from the NCBI web site the full contig sequence. You also can download information about SNPs. To do that, please login to the Locus link (<http://www.ncbi.nlm.nih.gov/LocusLink/>) server and search for your gene. Search for the NCBI Reference Sequences (RefSeq) section and the Category: NCBI Genome Annotation. Follow the link to the Genomic Contig: (for example NT\_033927). Save the web page as a text file (using the “save as” menu from your browser). Choose the display FASTA option to get the contig sequence. Save now this sequence as a txt file (use the “save as” menu from your browser). Finally, you need to remove all comments and return symbols from this file.



The various reference sequences necessary to build a new LSDB using the UMD software. **MYO7A cDNA.txt** includes the reference cDNA sequence as a text file; **MYO7A gene exonic structure** the exonic organization of the MYO7A gene according to the reference cDNA; **NT\_033927.txt** is a text file which includes annotations of the genomic contig and **NT\_033927 (traité).seq** is the sequence of the genomic contig.

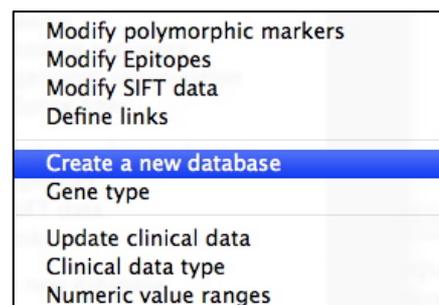
**You are now ready to build your LSDB using the UMD software.**

## II- CREATE A NEW UMD-LSDB

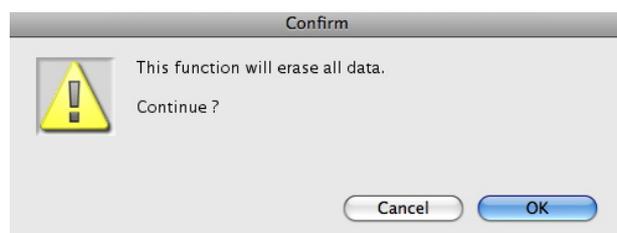
### II-a Set-up your own mutation database using the UMD® software?

Beware that you can only build a new UMD database from an existing one. Indeed the data file is not empty but contains the genetic code as well as other useful data that can not be directly modified by the user.

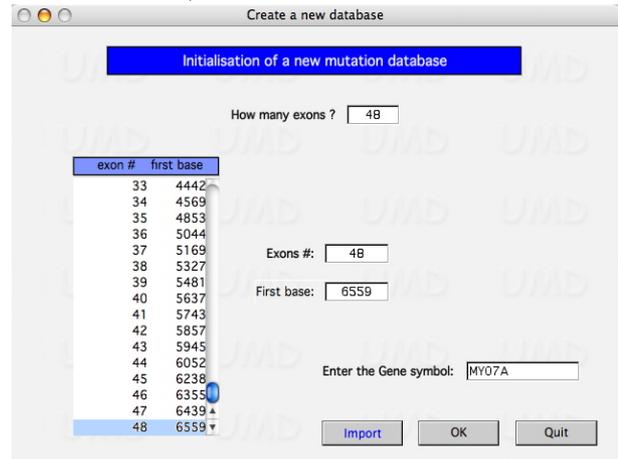
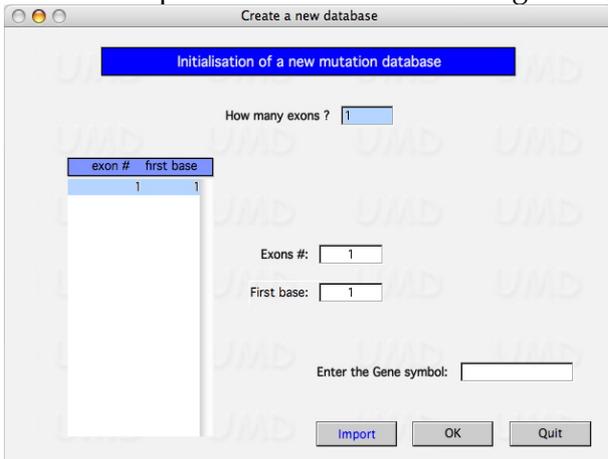
Therefore before the creation of a new database, you need to copy the directory of an existing UMD-LSDB and rename it for your new gene. When the new directory is ready, launch the UMD® software and choose the “Create a new database” option from the “File” menu.



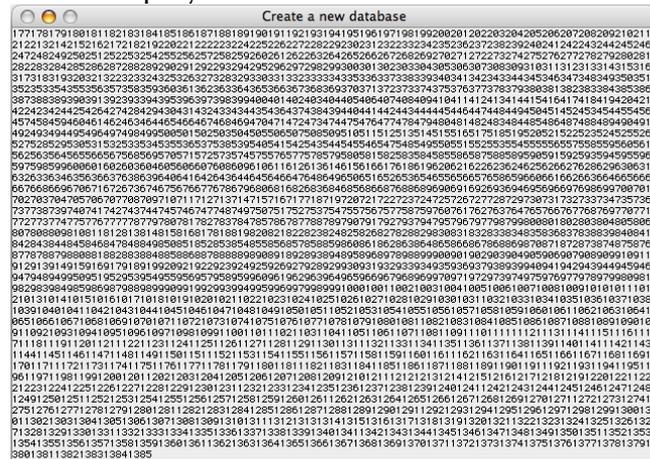
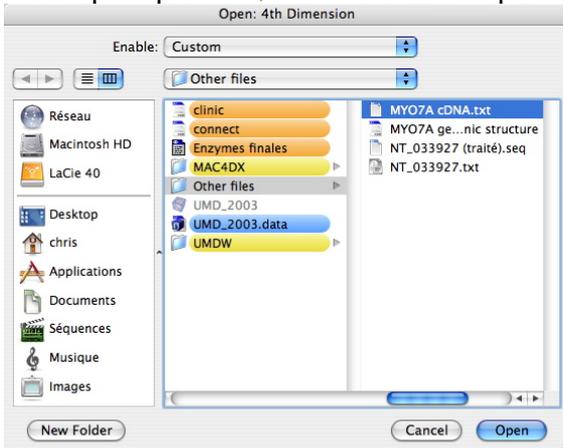
As this option will erase all data related to a specific gene (mutations, reference sequences as well as clinical data), an alert message will ask you to confirm this action.



You now need to specify the exonic organization of your gene (how many exons and the first nucleotide position of each according to the reference cDNA).

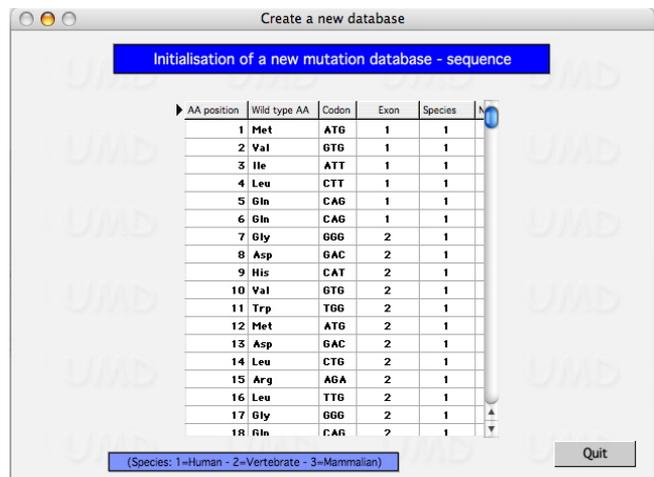


Enter the gene symbol (please refer to the nomenclature system for correct gene symbol name <http://www.gene.ucl.ac.uk/nomenclature/>). You now need to import the reference cDNA sequence. For this select the corresponding text file created in the previous section. During the import process, all amino acids positions will be displayed in a flush window.



When the process is over, the coding reference sequence is displayed as a table. For each amino acid, the corresponding codon and exon will be displayed with the species value set to 1. The species value represents the conservation of the amino acid with the following code:

- 1 = AA only found in human
  - 2 = AA conserved within mammals
  - 3 = AA conserved within vertebrates
- You can also set your own codes (4, 5 ...). All data can be modified by a double click on the corresponding amino acid position.



Now click on the “Quit” button in order to validate all modifications. Re-launched the UMD software, all new data are now integrated and UMD® will generate the reference sequence of your gene and will create reference pictures such as the exon phasing and the reference sequence.

At this stage the UMD® software will ask you to specify the “images4D directory” position in your hard disk. It is located in the UMDW directory.



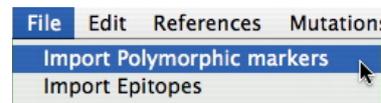
As UMD® is a generic software, it can be adapted to genes involved in cancer or genes involved in genetic diseases. To set-up this parameters please use the “**Gene type**” option from the “**File**” menu

The following message will appear “*This gene is involved in cancer. Do you confirm it?*” If your gene of interest is involved in cancer, just click “Yes” if not just click “No”. If you incorrectly set this value to the wrong option, just select again the “**Gene type**” option from the “**File**” menu. The following message will appear: “*This gene is not involved in cancer. Do you confirm it?*” Just click “Yes” or “No”.

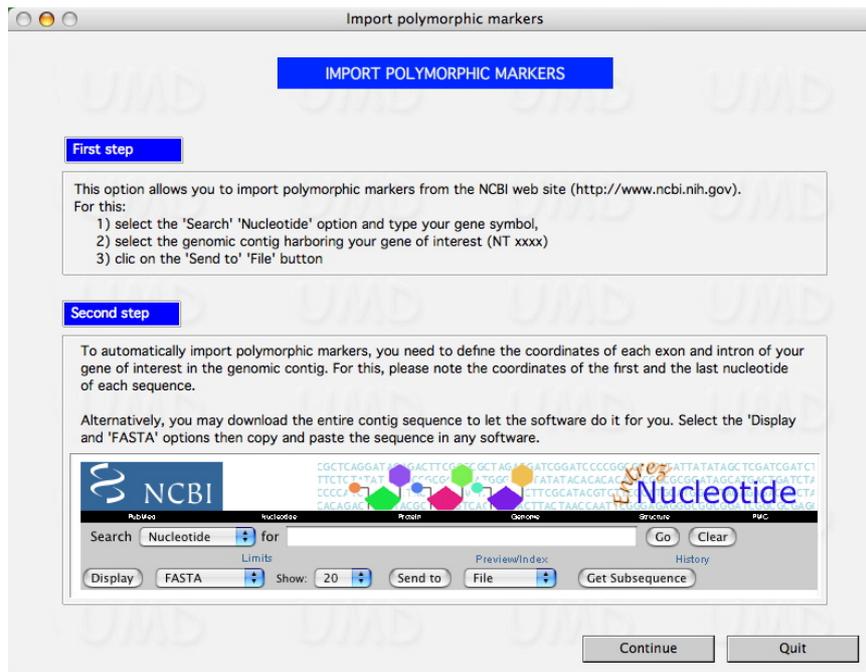
Your database is now ready to be used. If intronic sequences of your gene are known and localized in a genomic contig, we suggest that directly include intronic sequences. Nevertheless this process is optional and can be performed later.

### ***II-b Integrate intronic sequences and SNPs from a genomic contig***

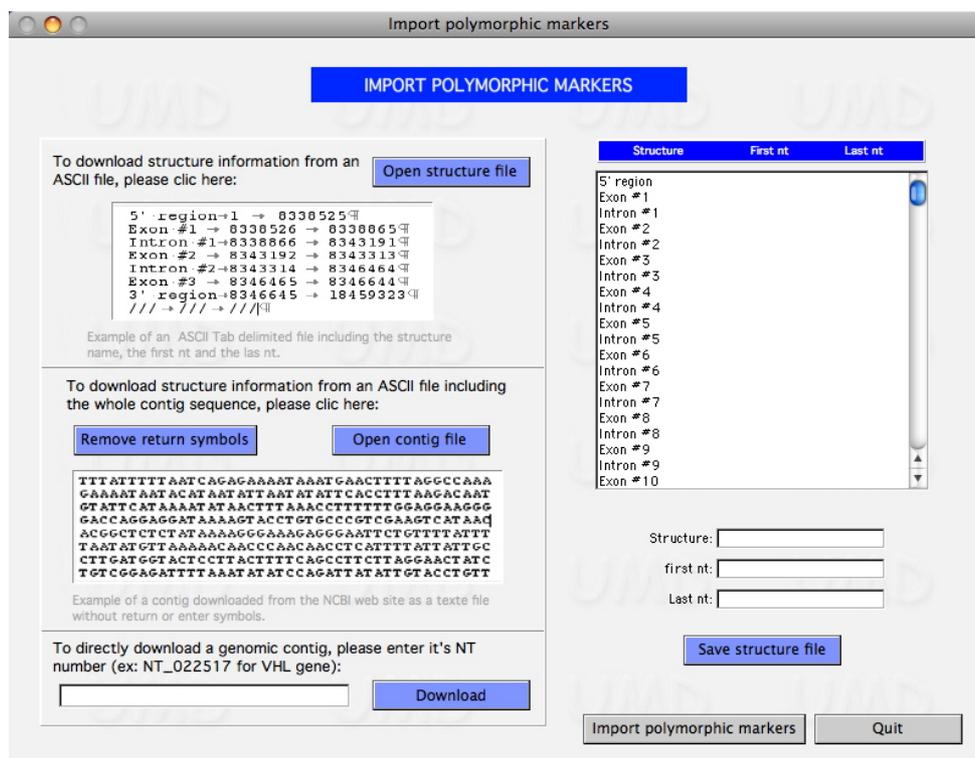
Select the “Import polymorphic markers” option from the file menu.



The following screen appears:



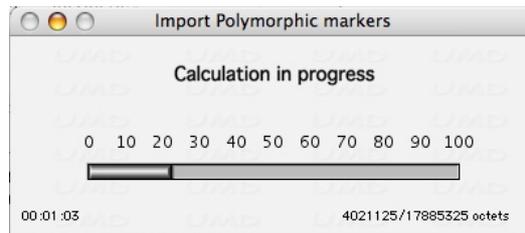
After downloading data from the NCBI (cf. II) click on "Continue". A second screen appears that allows you to use three alternative ways to import the data:



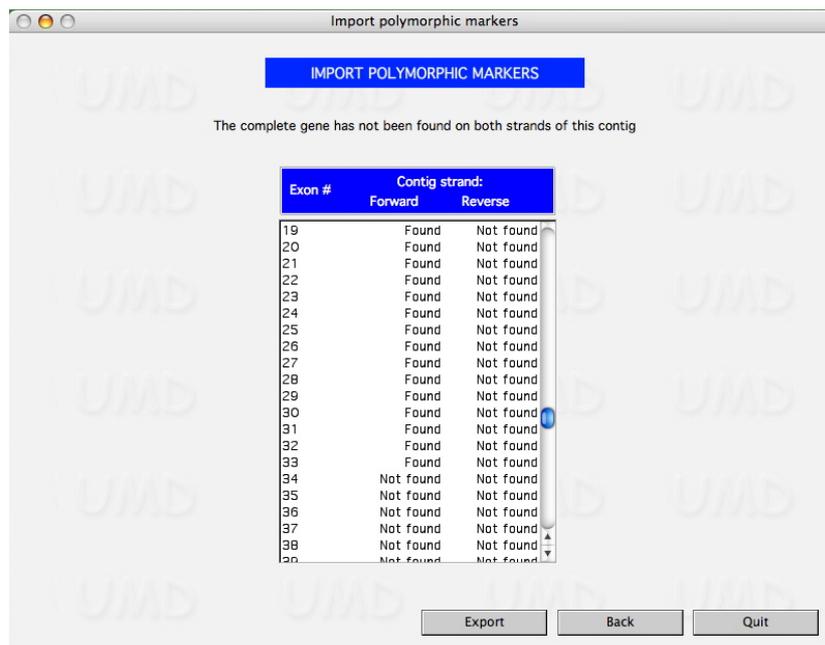
The first one is to use a tab-delimited file (ASCII file) describing the position of exons and introns in the contig (top). The second allows the import of the contig sequence file (only coding sequence) and the third one allows you to directly download through the Internet the contig using its NT code.

For option #2, if you got troubles to remove return symbols from your text editor software, you can use a specific tool to do that (press the remove return symbols button). If your text

files are ready to use, press the “Open contig file” button and select the contig sequence text file. The software will then search on both strands of the contig for exonic sequences.

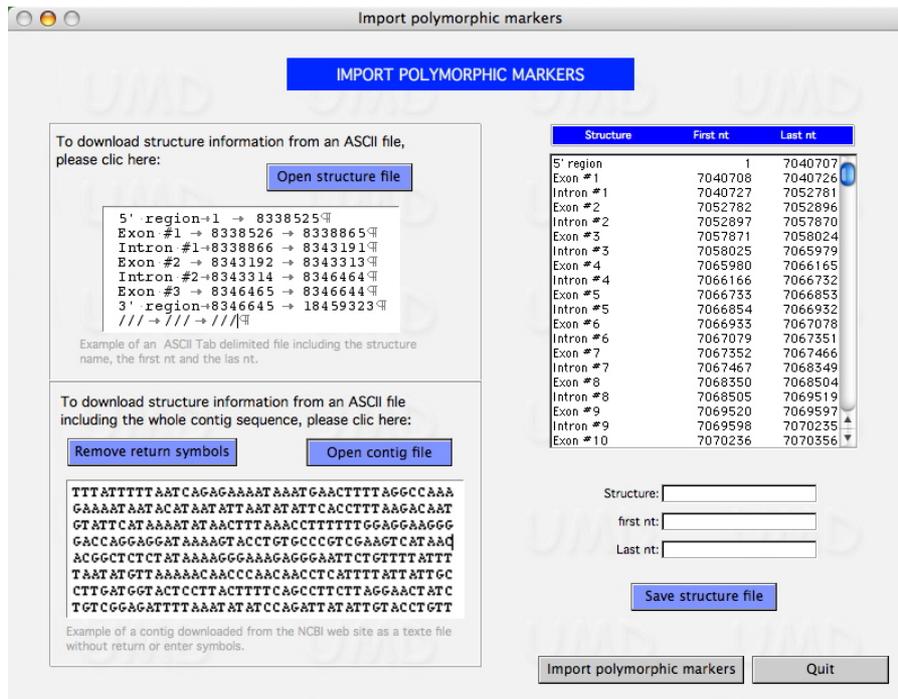


When this process is over UMD<sup>®</sup> can either have found all exonic sequences or only few of them. This is mainly due to polymorphisms in the contig sequence.

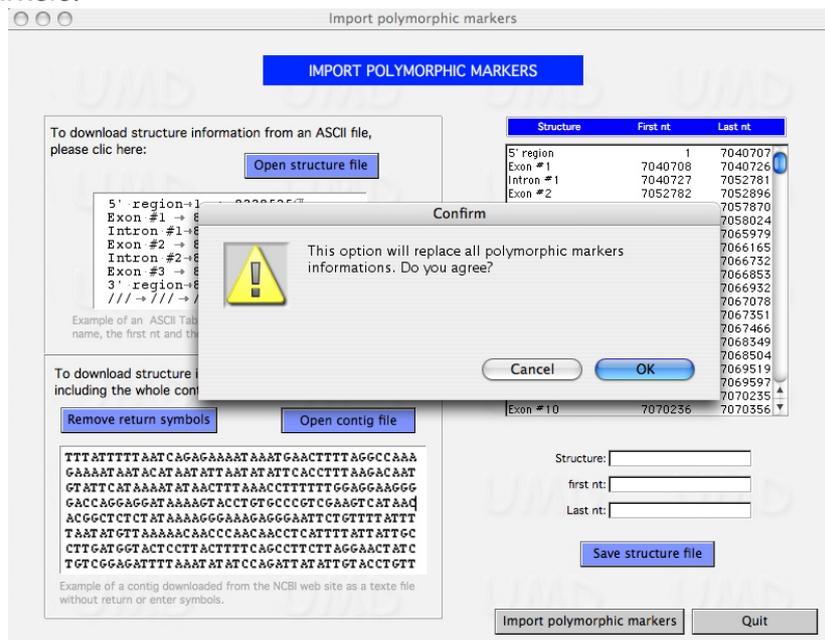


When no exon has been found, this screen displays partial results. You can export exonic sequences with the “Export” button and look for discrepancies between your coding reference sequence and the contig.

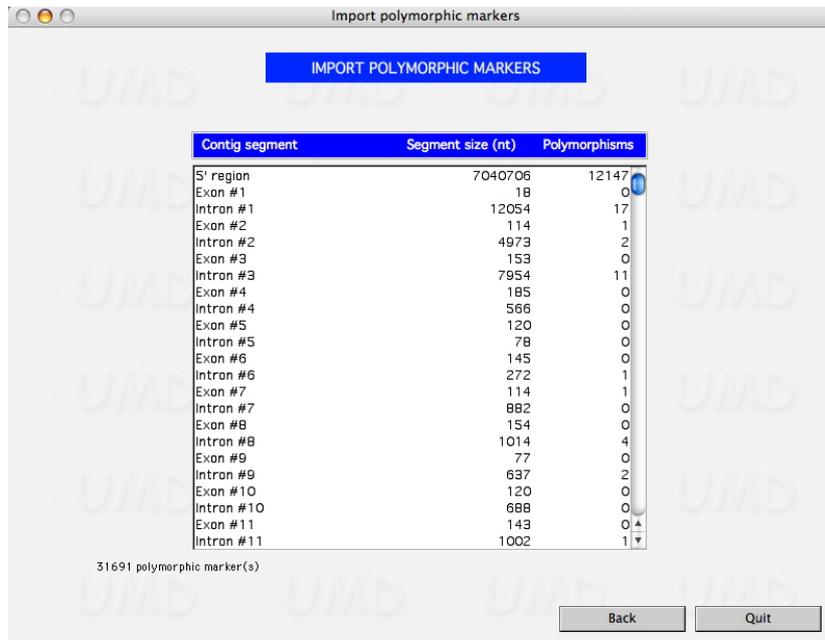
In this other example the UMD<sup>®</sup> software has found all exons on the forward strand of the contig. The intronic sequences are then automatically generated and included in UMD<sup>®</sup>.



You can now import polymorphic markers information with the corresponding button. A confirmation message will remind you that this process will update all information about polymorphic markers:



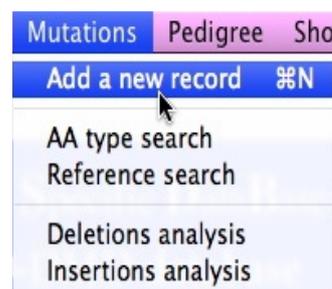
When the process is completed, a screen displays how many markers have been imported and their distribution in the various portions of the gene.



**Note that even if you import manually non-coding sequences, you should use the naming rule: intron #xx and 5' region or 3' region. If you use other names, the software will not be able to import polymorphic markers neither to search for intronic reference sequences.**

### III- HOW TO ENTER A NEW MUTATION?

The UMD-database now includes the genetic code as well as the reference sequence of your gene of interest. Select in the “Mutations” menu the “Add a new record” option:



The UMD-database can have different input formats (contact us to get a specific design) but all formats share common elements that allow the description of the mutation:

The image shows a web browser window titled 'New mutation' with a date of '25/08/2008' and a user name of 'concepteur'. The main heading is 'MUTATION: NEW RECORD'. There are radio buttons for 'Large', 'Small', and 'Rearrangement', with 'Small' selected. Below this are several tabs for mutation types: 'Large del', 'Large dup', 'indels', 'Intronic mutation', 'Correcting splices', 'Monoclonal Antibodies', and 'Polymorphic markers'. The 'Large del' tab is active. On the left, there are input fields for 'NUCLEOTIDE' (value '0'), 'AA POSITION' (value '0'), 'WT CODON', 'MUTANT CODON', and 'INSERTION'. There are also buttons for 'EXON', 'SIFT', 'ESE analysis', and 'Splice site?'. On the right, there are input fields for 'Event', 'CpG', 'Py-Py', 'Structure', and 'HCD', along with a button for 'Impact on isoforms'.

The left part (A) allows the input of various types of mutation while the right part (B) includes data that are automatically computed by the software. Below this common part of the input screen, most formats also include the following fields that allow the input of information related to the sample and/or the patient:

The image shows a form for entering patient and sample information. It includes the following fields: 'Sample ID', 'Patient ID', 'Family ID', 'Gender', 'Age of onset', 'Age of death', 'Date of birth' (with a default value of '00/00/00'), 'Last follow-up' (with a default value of '00/00/00'), 'Transmission', 'Geographic origin', 'Phenotypic group', and 'Mutation status'. At the bottom, there are radio buttons for 'Proband' (selected) and 'Relative'.

The “Sample ID” field is a unique identifier. You can set it using your own criteria. We suggest that you use a country code followed by a lab code a family and a patient code as well as a gender code. This will allow you to rapidly give enough information to the laboratory that reported the mutation to let it go back to the patient if necessary.

### III-a - Point mutations

In the following examples, we will use the DMD gene (OMIM 300377) involved in Duchenne and Becker Muscular Dystrophies (cf. [UMD-DMD LSDB](#)).

#### 1) c.9568C>T (p.Arg3190X)

In this example, we will show you how to enter a point mutation at nucleotide position #9568 resulting in an Arg3190X mutation. You can enter the mutation's position either at the nucleotide level (9568 in nucleotide field) or at the amino acid level (3190 in the AA position field).

The software will then display the wild type codon sequence (CGA), the corresponding wild type amino acid (Arg) and the exon where is localized this amino acid (Exon #66). Note also that the structural domain in which is localized the amino acid (Cystein-rich domain) is displayed.

The screenshot shows the 'New mutation' window with the following fields and values:

- MUTATION: NEW RECORD** (blue button)
- Radio buttons:  Large,  Small,  Rearrangement
- Buttons: Large del, Large dup, indels, Intronic mutation, Correcting splices, Monoclonal Antibodies, Polymorphic markers
- NUCLEOTIDE: 9568
- AA POSITION: 3190, EXON: 66, SIFT: [button]
- WT CODON: CGA, Arg, ESE analysis: [button]
- MUTANT CODON: [empty], Splice site?: [button]
- INSERTION: [empty]
- Event: [empty]
- CpG: [empty], Type: [empty]
- Py-Py: [empty]
- Structure: Cystein-rich domain
- HCD: [empty]
- Impact on isoforms: [button]

Now you need to enter the mutant codon "TGA". The software will automatically calculate and display the nomenclature name (c.9568C>T and p.Arg3190X) as well as the consequences of the mutation: "Molecular event = C->T"; "CpG = Yes"; "Mutation type = Transition (Ts)"; "Pyrimidine double (Py-Py) = No".

The screenshot shows the 'New mutation' window with the following fields and values:

- MUTATION: NEW RECORD** (blue button)
- Radio buttons:  Large,  Small,  Rearrangement
- Buttons: Large del, Large dup, indels, Intronic mutation, Correcting splices, Monoclonal Antibodies, Polymorphic markers
- NUCLEOTIDE: 9568
- AA POSITION: 3190, EXON: 66, SIFT: [button]
- WT CODON: CGA, Arg, ESE analysis: [button]
- MUTANT CODON: TGA, Stop, Splice site?: [button]
- INSERTION: [empty]
- Event: C->T
- CpG: Yes, Type: Ts
- Py-Py: No
- Structure: Cystein-rich domain
- HCD: [empty]
- Mutation: [button]
- Impact on isoforms: [button]
- Theoretical molecular mass of truncated protein: 369 kDa

If you proceed with other nonsense mutations you should get the following results:

### 2) c.3935T>A (p.leu1312X)

25/08/2008 concepteur

c.3935T>A  
p.Leu1312X

**MUTATION: NEW RECORD**

3061

Theoretical molecular mass of truncated protein: 151 kDa

Large  Small  Rearrangement

Large del	Large dup	indels	Intronic mutation
-----------	-----------	--------	-------------------

NUCLEOTIDE	3935			
AA POSITION	1312	EXON	29	SIFT
WT CODON	TTG	Leu	ESE analysis	
MUTANT CODON	TAG	Stop	Splice site?	
INSERTION				

Correcting splices	Monoclonal Antibodies	Polymorphic markers
--------------------	-----------------------	---------------------

Event	T->A		
CpG	No	Type	Tv
Py-Py	Yes, coding strand		
Structure	CRD-repeat #9		
HCD			
Mutation	Impact on isoforms		

### 3) c.4996C>T (p.Arg1666X)

25/08/2008 concepteur

c.4996C>T  
p.Arg1666X

**MUTATION: NEW RECORD**

3061

Theoretical molecular mass of truncated protein: 192 kDa

Large  Small  Rearrangement

Large del	Large dup	indels	Intronic mutation
-----------	-----------	--------	-------------------

NUCLEOTIDE	4996			
AA POSITION	1666	EXON	35	SIFT
WT CODON	CGA	Arg	ESE analysis	
MUTANT CODON	TGA	Stop	Splice site?	
INSERTION				

Correcting splices	Monoclonal Antibodies	Polymorphic markers
--------------------	-----------------------	---------------------

Event	C->T		
CpG	Yes	Type	Ts
Py-Py	Yes, coding strand		
Structure	CRD-repeat #12		
HCD			
Mutation	Impact on isoforms		

### III-b- Small deletions

#### 1) c.372delG

The mutation position selection is performed as for point mutations. To precisely define the deleted sequence you need to follow the UMD naming system. In this example you will create various deletions starting in codon #124 (first nucleotide is at the position 372).

Codon	A	T	G
Nucleotide position	372	373	374
UMD codon position	a	b	c

In the mutant codon section, you need to enter:

- The generic symbol “del”,
- The number of deleted nucleotides,
- The “UMD codon position” of the first deleted nucleotide.

If there is only one nucleotide deleted

- ❖ GTA **ATG** AAA -> GTA TGA AA      del1a
- ❖ GTA **ATG** AAA -> GTA AGA AA      del1b
- ❖ GTA **ATG** AAA -> GTA ATA AA      del1c

If there are two nucleotides deleted

- ❖ GTA **ATG** AAA -> GTA GAA A      del2a

- ❖ GTA ATG AAA -> GTA AAA A                      del2b
- ❖ GTA ATG AAA -> GTA ATA A                      del2c
- ❖

If there are three nucleotides deleted

- ❖ GTA ATG AAA -> GTA AAA                      del3a
- ❖ GTA ATG AAA -> GTA AAA                      del3b
- ❖ GTA ATG AAA -> GTA ATA                      del3c

**New mutation**  
 26/08/2008                      concepteur  
 c.372delG  
 p.Met124IlefsX18  
 MUTATION: NEW RECORD  
 1  
 Theoretical molecular mass of truncated protein: 16 kDa  
 The largest in frame cDNA corresponds to splice from exon 5 to exon 9

Small rearrangement

Large del | Large dup | indels | Intronic mutation

Correcting splices | Monoclonal Antibodies | Polymorphic markers

NUCLEOTIDE: 372  
 AA POSITION: 124 | EXON: 6 | SIFT  
 WT CODON: ATG | Met | ESE analysis  
 MUTANT CODON: del1c | Fs- | Splice site?  
 INSERTION:

Event: Stop at 141  
 CpG: | Type: Fr.  
 Py-Py:  
 Structure: Actin binding domain  
 HCD:  
 Mutation: | Impact on isoforms

## 2) c.481\_483delACC

**New mutation**  
 26/08/2008                      concepteur  
 c.481\_483delACC  
 p.Thr161del  
 MUTATION: NEW RECORD  
 1  
 Theoretical molecular mass of truncated protein: 426 kDa

Small rearrangement

Large del | Large dup | indels | Intronic mutation

Correcting splices | Monoclonal Antibodies | Polymorphic markers

NUCLEOTIDE: 481  
 AA POSITION: 161 | EXON: 6 | SIFT  
 WT CODON: ACC | Thr | ESE analysis  
 MUTANT CODON: del3a | InF | Splice site?  
 INSERTION:

Event: In frame del  
 CpG: | Type: InF  
 Py-Py:  
 Structure: Actin binding domain  
 HCD:  
 Mutation: | Impact on isoforms

NB: Remember that the international nomenclature system states that for all descriptions the most 3' possible position is arbitrarily assigned to have been changed (especially when repeated sequences are involved).

Other examples are given below:

### 3) c.2638delC

New mutation

26/08/2008 concepteur

c.2638delC  
p.Leu880TyrfsX11

Large  Small  Rearrangement

Small rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 103 kDa  
The largest in frame cDNA corresponds to splice from exon 20 to exon 23

### 4) c.3010\_3011delAA

New mutation

26/08/2008 concepteur

c.3010\_3011delAA  
p.Lys1004ArgfsX9

Large  Small  Rearrangement

Small rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 117 kDa  
The largest in frame cDNA corresponds to splice from exon 22 to exon 24

### 5) c.4918delA

New mutation

26/08/2008 concepteur

c.4918delA  
p.Thr1640IlnfsX17

Large  Small  Rearrangement

Small rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 191 kDa  
The largest in frame cDNA corresponds to splice from exon 34 to exon 36

### III-c- Insertions

#### 1) c.9319insC

The mutation position selection is performed as for point mutations. To define precisely where will be inserted the sequence you need to follow the UMD naming system:

Codon	A	G	G
Nucleotide position	9319	9320	9321
UMD codon position	a	b	c

In the mutant codon section, you need to enter:

- The generic symbol “ins”,
- The number of inserted nucleotides,
- The “UMD codon position” of the **first nucleotide localized after the insertion.**

Insertion of one nucleotide

❖ AGG ACT GCC	ins1a	1 nucleotide (T)	ATG GAC TGC C
❖ AGG ACT GCC	ins1b	1 nucleotide (T)	AGT GAC TGC C
❖ AGG ACT GCC	ins1c	1 nucleotide (T)	AGG TAC TGC C

Insertion of two nucleotides

❖ AGG ACT GCC	ins2a	1 nucleotide (TT)	ATT GGA CTG CC
❖ AGG ACT GCC	ins2b	1 nucleotide (TT)	AGT TGA CTG CC
❖ AGG ACT GCC	ins2c	1 nucleotide (TT)	AGG TTA CTG CC

## Insertion of three nucleotides

- ❖ AGG ACT GCC      ins3a 1nucleotide (TTT)    ATT TGG ACT GCC
- ❖ AGG ACT GCC      ins3b 1nucleotide (TTT)    AGT TTG ACT GCC
- ❖ AGG ACT GCC      ins3c 1nucleotide (TTT)    AGG TTT ACT GCC

New mutation

27/08/2008      concepteur

c.9319insC  
p.Arg3107GlnfsX25

Large    Small    Rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 362 kDa  
The largest in frame cDNA corresponds to splice from exon 63 to exon 65

Large del   Large dup   indels   Intronic mutation

Correcting splices   Monoclonal Antibodies   Polymorphic markers

NUCLEOTIDE: 9319

AA POSITION: 3107   EXON: 64   SIFT

WT CODON: AGG   Arg   ESE analysis

MUTANT CODON: ins1a   Fs.   Splice site?

INSERTION: C

Event: Stop at 3131

CpG:   Type: Fr.

Py-Py:   

Structure: CRD-hinge region #4

HCD:   

Mutation   Impact on isoforms

Other examples are given below:

### 2) c.9319insAA

New mutation

26/08/2008      concepteur

c.9319insAA  
p.Arg3107LysfsX5

Large    Small    Rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 360 kDa  
The largest in frame cDNA corresponds to splice from exon 63 to exon 65

Large del   Large dup   indels   Intronic mutation

Correcting splices   Monoclonal Antibodies   Polymorphic markers

NUCLEOTIDE: 9319

AA POSITION: 3107   EXON: 64   SIFT

WT CODON: AGG   Arg   ESE analysis

MUTANT CODON: ins2a   Fs.   Splice site?

INSERTION: AA

Event: Stop at 3111

CpG:   Type: Fr.

Py-Py:   

Structure: CRD-hinge region #4

HCD:   

Mutation   Impact on isoforms

### 3) c.3267insC

New mutation

27/08/2008      concepteur

c.3267insC  
p.Lys1089AsnfsX9

Large    Small    Rearrangement

**MUTATION: NEW RECORD**

Theoretical molecular mass of truncated protein: 127 kDa  
The largest in frame cDNA corresponds to splice from exon 23 to exon 25

Large del   Large dup   indels   Intronic mutation

Correcting splices   Monoclonal Antibodies   Polymorphic markers

NUCLEOTIDE: 3267

AA POSITION: 1089   EXON: 24   SIFT

WT CODON: AAA   Lys   ESE analysis

MUTANT CODON: ins1c   Fs.   Splice site?

INSERTION: C

Event: Stop at 1097

CpG:   Type: Fr.

Py-Py:   

Structure: CRD-repeat #7

HCD:   

Mutation   Impact on isoforms

#### 4) c.2929dup

New mutation

26/08/2008 concepteur

c.2929dup  
p.Gln977ProfsX37

**MUTATION: NEW RECORD**

Large 
  Small 
  Rearrangement

Theoretical molecular mass of truncated protein: 117 kDa  
The largest in frame cDNA corresponds to splice from exon 21 to exon 51

#### 5) c.2624\_2638dup

New mutation

26/08/2008 concepteur

c.2624\_2638dup  
p.Leu880MetfsX16

**MUTATION: NEW RECORD**

Large 
  Small 
  Rearrangement

Theoretical molecular mass of truncated protein: 103 kDa  
The largest in frame cDNA corresponds to splice from exon 20 to exon 23

### III-d- Large deletions

Large deletions correspond to the deletion of one or more exons of the gene. In order to facilitate the description of large deletions, you can use the “**Large del**” button. This will display a specific window where you can specify the first and last deleted exons. The software shows the deleted cDNA sequence:

**Large deletion**

First deleted exon

Last deleted exon

Deleted sequence (cDNA) Size = 176 bp

```
GAACTCCAGGATGGCATTGGGCAGCGCAAACTGTTGTC
AGAACATTGAATGCAACTGGGGAAAGAAATAATTCAGCA
ATCCTCAAAAACAGATGCCAGTATTCTACAGGAAAAAT
TGGGAAGCCTGAAATCTGCGGTGGCAGGAGGTCTGCAAAC
AGCTGTCAGACAGAAAAAAGAG
```

OK Quit

After validation, the software will fill all fields and calculate the position of the premature termination codon when a frameshift is produced:

New mutation

25/08/2008 concepteur

c.6439\_6614del  
p.Glu2147AlafsX17

**MUTATION: NEW RECORD** 3061

Large  Small  Rearrangement

deletion from exon 45 to 45

Theoretical molecular mass of truncated protein: 250 kDa  
The largest in frame cDNA corresponds to splice from exon 44 to exon 47

Correcting splices Monoclonal Antibodies Polymorphic markers

Event

CpG  Type

Py-Py

Structure

HCD

Mutation  Impact on isoforms

NUCLEOTIDE

AA POSITION  EXON  SIFT

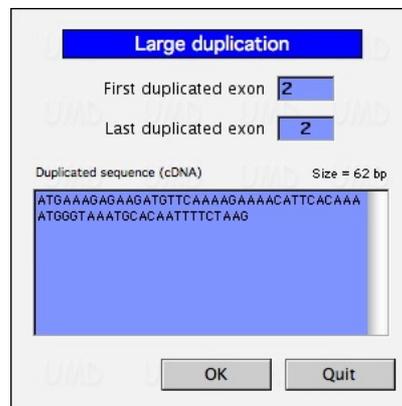
WT CODON   ESE analysis

MUTANT CODON   Splice site?

INSERTION

### III-e- Large duplications

As for large deletions, large duplications correspond to the duplication of one or more exons of the gene. In order to facilitate the description of large duplications, you can use the “**Large dup**” button. This will display a specific window where you can specify the first and last duplicated exons. The software shows the duplicated cDNA sequence:



Large duplication

First duplicated exon

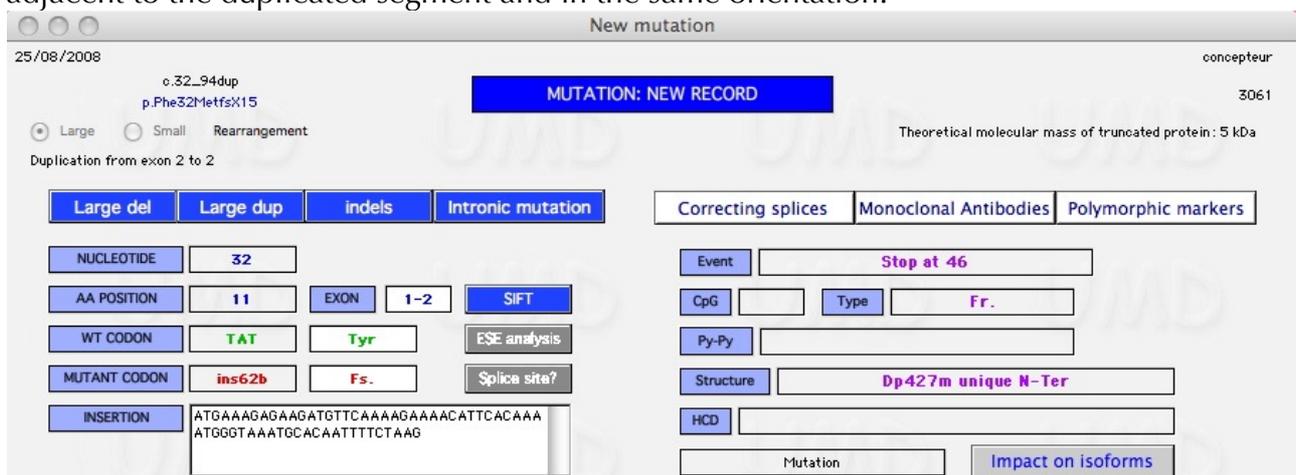
Last duplicated exon

Duplicated sequence (cDNA) Size = 62 bp

```
ATGAAAGAGAAGATGTTCAAAAGAAAACATTCACAAA
ATGGGTAAATGCACAATTTTCTAAG
```

OK Quit

After validation, the software will fill all fields and calculate the position of the premature termination codon when a frameshift is produced. This implies that the duplication is adjacent to the duplicated segment and in the same orientation:



New mutation

25/08/2008 c.32\_94dup p.Phe32MetfsX15 concepteur 3061

MUTATION: NEW RECORD

Large  Small  Rearrangement

Theoretical molecular mass of truncated protein: 5 kDa

Duplication from exon 2 to 2

Large del Large dup indels Intronic mutation

Correcting splices Monoclonal Antibodies Polymorphic markers

NUCLEOTIDE

AA POSITION  EXON  SIFT

WT CODON   ESE analysis

MUTANT CODON   Splice site?

INSERTION

Event

CpG  Type

Py-Py

Structure

HCD

Mutation  Impact on isoforms

### III-f- Intronic mutations

Beware that you can only enter intronic mutations (also called in UMD splice mutations) if you have previously defined the reference sequence of each intron using the proper nomenclature (cf. chapter 2). Because in most clinical diagnostic laboratories the naming system for such mutations is the alternative nomenclature (IVS), UMD integrate both nomenclatures (for more see [HGVS website](#)).

#### 1) c.9649+1G>A (c.IVS66+1G>A)

You can enter an intronic mutation using the “**Intronic mutation**” button. A new window is displayed:

The screenshot shows a window titled "Intronic mutation" with a blue button labeled "MUTATION: NEW RECORD". The window contains the following elements:

- 1) Select an intron: A dropdown menu.
- 2) Select a mutation type (+/-): A dropdown menu.
- 3) Enter a position: A text input field containing "0".
- 4) Verify the sequence context: A table with 11 columns and 2 rows labeled "Intron" and "Position".
- 5) Enter mutational event: A text input field.
- 6) Mutation type: A text input field.
- Right panel: Fields for "Intron number" (0), "Intron size (bp)" (0), and "Intron sequence" (empty text area).
- Bottom right: "Quit" and "OK" buttons.

To enter a mutation you need to process in three steps. For example, for a mutation of the first nucleotide of the intron (position +1):

- Select the intron # (here #66)
- The mutation type (+ for mutation of a donor splice site, - for an acceptor splice site)
- The intronic position (here #1)

The screenshot shows the "Intronic mutation" window with the following updates:

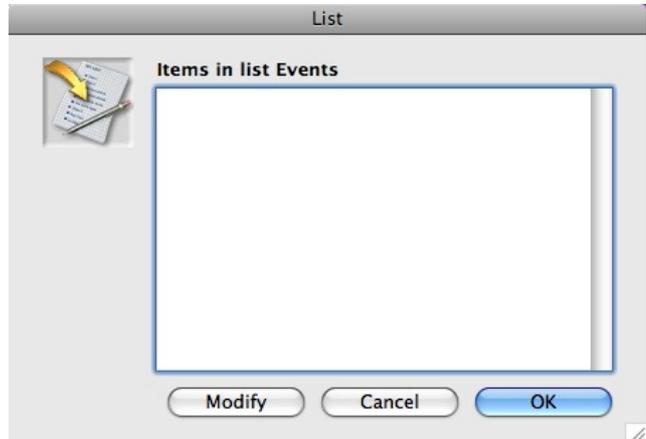
- 1) Select an intron: Dropdown menu set to "Intron #66".
- 2) Select a mutation type (+/-): Dropdown menu set to "+".
- 3) Enter a position: Text input field set to "1".
- 4) Verify the sequence context: The table below is populated with nucleotide data.
- Right panel: "Intron number" is 66, "Intron size (bp)" is 2464, and "Intron sequence" contains a long nucleotide sequence.
- Bottom right: "Quit" and "OK" buttons.

Intron	A	G	T	A	C	A	G	A	T	g	t	a	a	g	t	c	g	t	g	t
Position	9641	9642	9643	9644	9645	9646	9647	9648	9649	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11

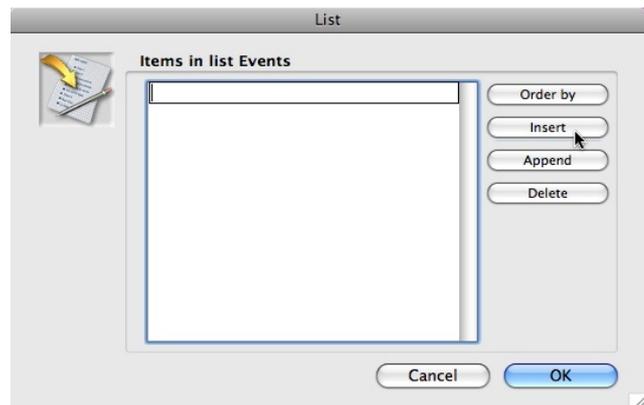
The UMD<sup>®</sup> software will display at the bottom of the window the reference sequence surrounding the selected position. Intronic sequence is displayed in small letters and exonic sequence in capital letters.

A window appears to allow a rapid selection of the mutational event:

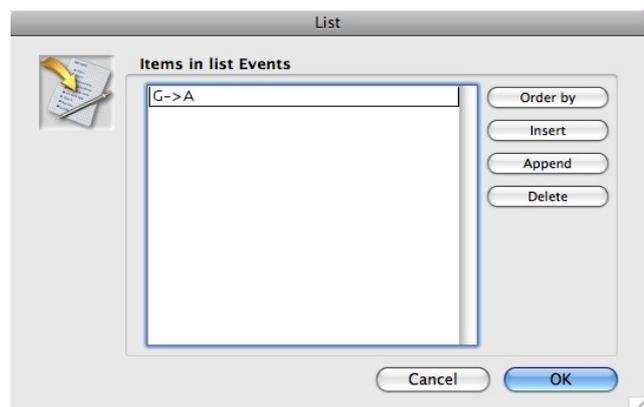
If the corresponding event is not available in the list, click the “**Modify**” button. At this stage, you can add, remove, insert or delete an event:



For the mutation c.9649+1G>A, the molecular event is a transition G to A. Click on the “**Insert**” button and add “**G->A** » in the text area :



Then press the “**OK**” button



The international name of the mutation is automatically calculated by UMD<sup>®</sup>. Note that for intronic mutations it also includes the alternative naming system as mentioned before.

New mutation

25/08/2008 concepteur

c.IV966+1G>A (c.9649+1G>A) 3061

**MUTATION: NEW RECORD**

Large
  Small
  Rearrangement

NUCLEOTIDE:

AA POSITION: 
 EXON:

WT CODON:

MUTANT CODON:

INSERTION:

Event:

CpG: 
 Type:

Py-Py:

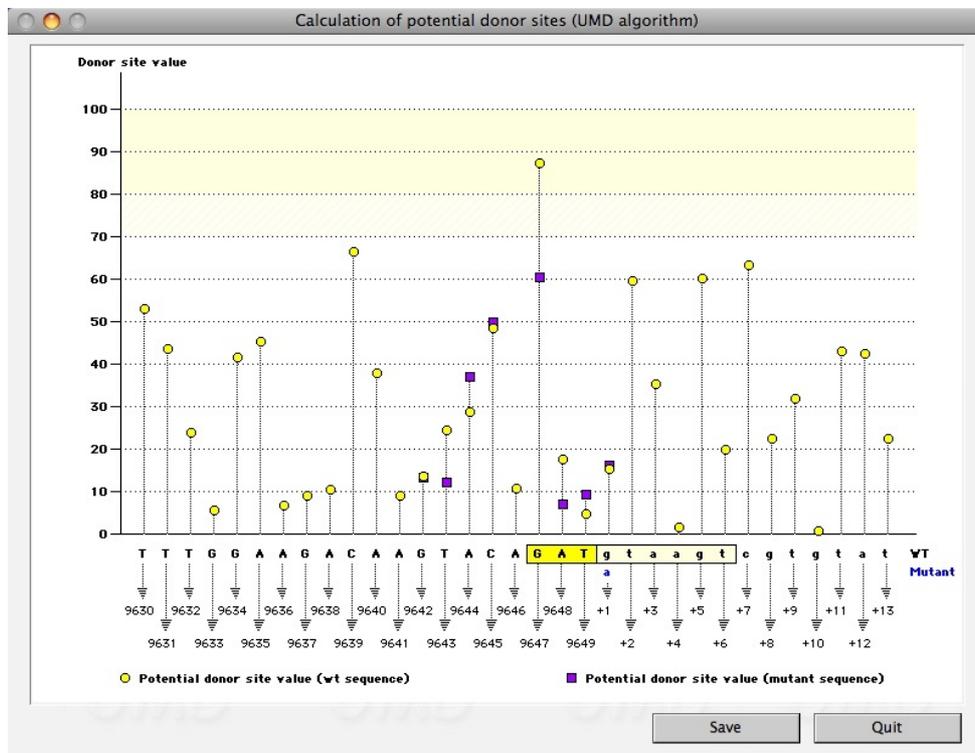
Structure:

HCD:

For all intronic mutations, two windows appear to display the possible impact of the mutation on donor (example below) and acceptor splice sites.

Exonic sequence are displayed in capital letters with corresponding nucleotide numbers (9630 to 9649) while intronic sequence are displayed in small letters with corresponding nucleotide numbers (+1 to +13). The normal donor splice site is shown in yellow (dark yellow for exonic nucleotides, light yellow for intronic nucleotides). The CV for each potential donor splice site (nine-nucleotide sequence) in the vicinity of the mutation is displayed according to the UMD rule (80-100 strong splice site; 70-80 = weaker splice site; 0-70 not a splice site).

Wild-type sequence CV are displayed as yellow square and mutant sequence CV as purple square. Each square represents the CV of the potential splice site beginning with the corresponding nucleotide.



### Disruption of a donor splice site

2) c.4519-5C>G (new nomenclature) / c.IVS32-5C>G (old nomenclature)

To enter this mutation, proceed as in the previous example, you should get:

**MUTATION: NEW RECORD**

1) Select an intron

2) Select a mutation type (+/-)

3) Enter a position

4) Verify the sequence context

Intron number   
 Intron size (bp)   
 Intron sequence  
 GTATGTATTCTGGTGGCAAAATACGCAGGTACCCCTTGACTTTCCTCATTAAAGAAGTAACT  
 GCTCTTTTATAAGAGAGAATTGTTTTTCAGATAACCATATAAATTATACTATGTAATTTTAA  
 GATTGAGAACAAAAATGGCACATGTATTGTGGCCTATTCTGTGTGCTTTTTTGGAGTCTA  
 TAATTCTGCTAAGAAGTGTGCAGAAATATAAAAGGATGTTGCTTGTGTTAATTTTATGAT  
 TCTTTAAATATATTTTAAATGGAAATCTAGAAAGGAAATATATTAATCTAAAAATAGTAAC  
 AACAAACAGTAGTTGCAATAATCCTAACCACTTCAGAAATAATGGATGCCAGGTAACACG  
 Sequence (3036 nt)

Intron	a	t	g	t	t	t	a	a	a	c	t	t	a	g	A	A	C	T	T	G
Position	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	4519	4520	4521	4522	4523	4524

5) Enter mutational event

6) Mutation type

After validation of these data, UMD® calculates the consequences of this mutation and give it a name:

**MUTATION: NEW RECORD**

25/08/2008 concepteur 3061

c.IVS32-5C>G (c.4519-5C>G)

Large  Small  Rearrangement

**Large del** **Large dup** **indels** **Intronic mutation** **Correcting splices** **Monoclonal Antibodies** **Polymorphic markers**

NUCLEOTIDE  AA POSITION  EXON  SIFT

WT CODON   ESE analysis

MUTANT CODON   Splice site?

INTEGRATION

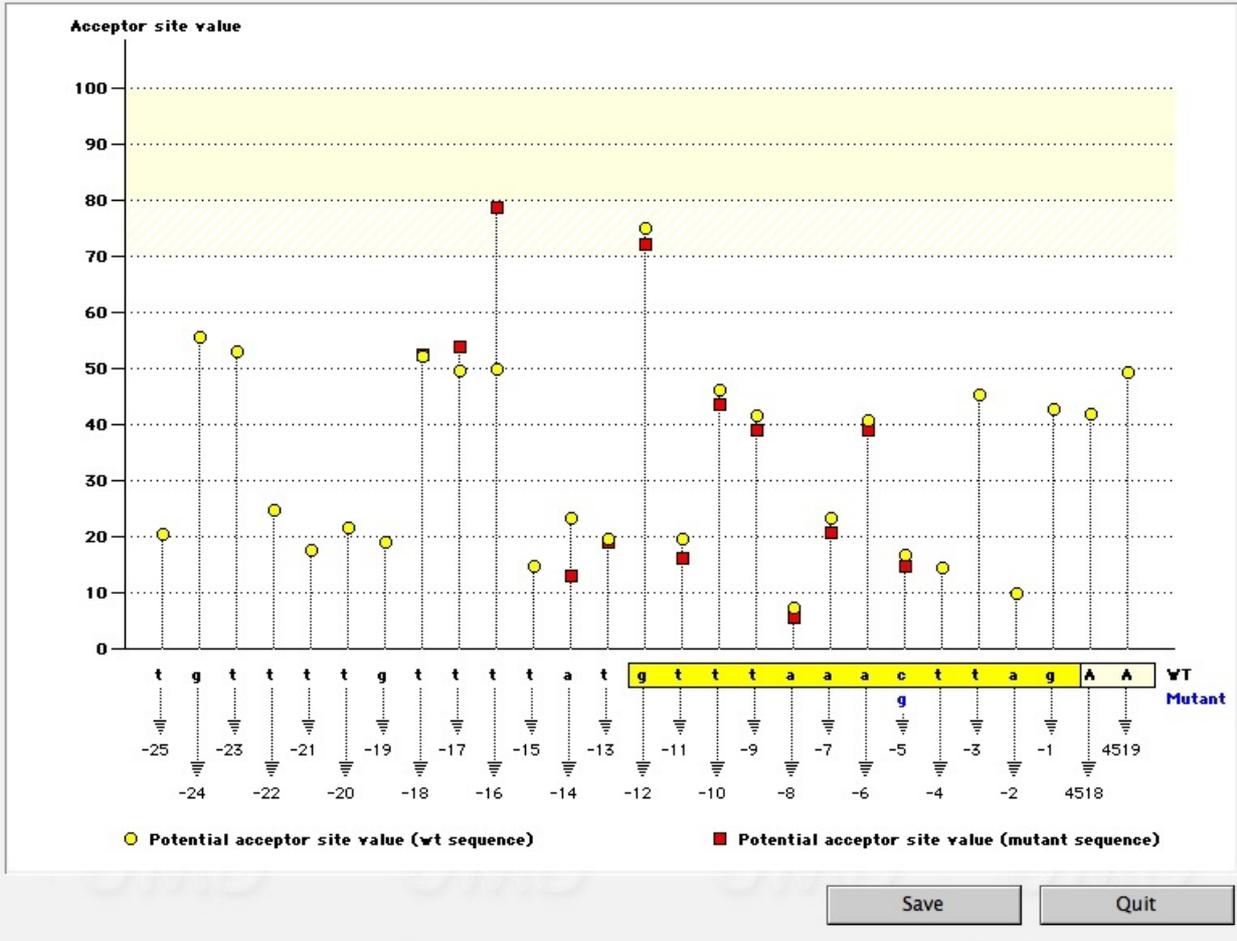
Event  CpG  Type  Py-Py

Structure

HCD

Mutation

Here the mutation at position -5, does not affect significantly the wt acceptor site (value drops from 75 to 72, pink circle) but instead creates a cryptic splice site (CV = 78.95) beginning at position -16 (wt site begins at position -12) leading to the incorporation of 4 nt and therefore to a premature stop codon and presumably to NMD (experimentally demonstrated).



### III-g- Missense mutations

To enter missense mutations, select the position using either the nucleotide or the codon number and enter the mutant codon. The software will automatically calculate the mutational event, the involvement of a CpG and/or Pyrimidine doublet, the structural domain that harbors the mutation as well as the exon and the mutant amino acid.

#### 1) c.10018T>C (p.Cys3340Arg)

New mutation

25/08/2008

c.10018T>C  
p.Cys3340Arg

MUTATION: NEW RECORD

concepteur 3061

Theoretical molecular mass of abnormal protein: 426 kDa

Large del Large dup indels Intronic mutation

Correcting splices Monoclonal Antibodies Polymorphic markers

Event T->C

CpG No Type Ts

Py-Py Yes, coding strand

Structure Carboxy-terminal region

HCD

Mutation Impact on isoforms

A click on the SIFT button, allow you to display SIFT predictions for this specific position:

AA position **3340 (Cys)** Conservation **0,42**

A (Ala)	C (Cys)	D (Asp)	E (Glu)	F (Phe)	G (Gly)	H (His)	I (Ile)	K (Lys)	L (Leu)
0,01	1,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00
M (Met)	N (Asn)	P (Pro)	Q (Gln)	R (Arg)	S (Ser)	T (Thr)	V (Val)	W (Trp)	Y (Tyr)
0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,00	0,00

**General information**  
SIFT is a multistep procedure that:  
(1) searches for and chooses similar sequences  
(2) makes an alignment of these sequences  
(3) calculates scores based on the amino acids appearing at each position in the alignment.

**Conservation**  
Is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

**Probabilities**  
Amino acids with probabilities < .05 are predicted to be deleterious.  
Substitutions predicted to be intolerant are highlighted in red.

SIFT: <http://blocks.fhcrc.org/sift/SIFT.html>

Quit

## 2) c.831G>T (p.Gln277His)

New mutation

25/08/2008 concepteur 3061

c.831G>T  
p.Gln277His

**MUTATION: NEW RECORD**

Large  Small  Rearrangement
 Theoretical molecular mass of abnormal protein: 426 kDa

AA position **277 (Gln)** Conservation **0,52**

A (Ala)	C (Cys)	D (Asp)	E (Glu)	F (Phe)	G (Gly)	H (His)	I (Ile)	K (Lys)	L (Leu)
0,93	0,19	0,38	0,79	0,74	0,43	0,56	0,44	0,55	0,84
M (Met)	N (Asn)	P (Pro)	Q (Gln)	R (Arg)	S (Ser)	T (Thr)	V (Val)	W (Trp)	Y (Tyr)
0,26	0,44	0,28	0,97	0,80	0,80	0,54	0,57	0,18	1,00

**General information**

SIFT is a multistep procedure that:

- (1) searches for and chooses similar sequences
- (2) makes an alignment of these sequences
- (3) calculates scores based on the amino acids appearing at each position in the alignment.

**Conservation**

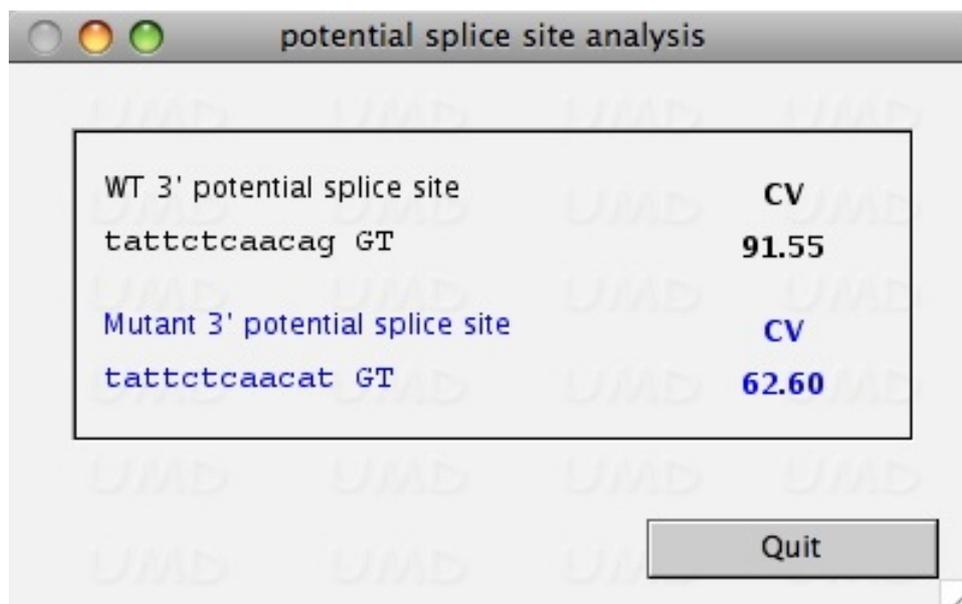
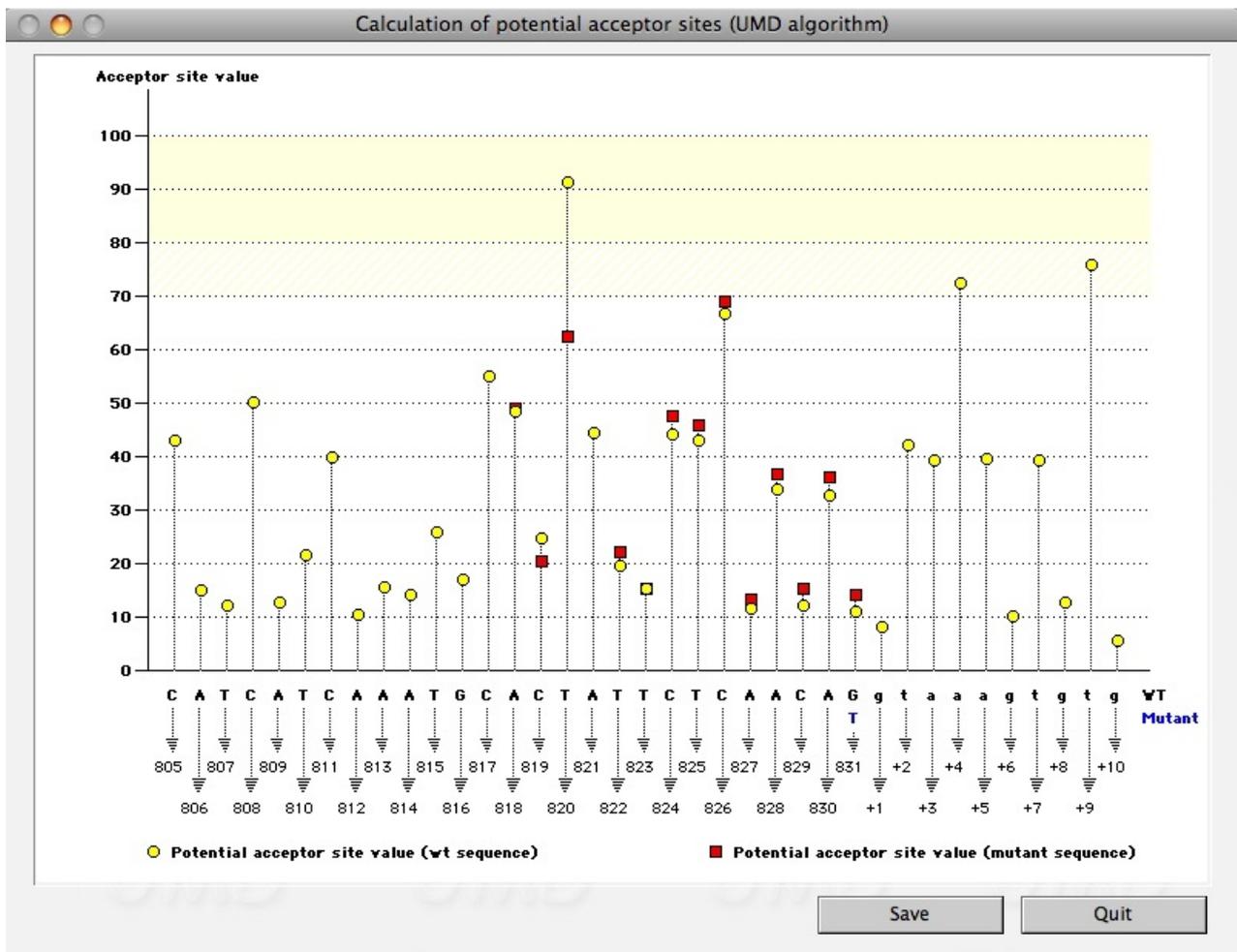
Is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

**Probabilities**

Amino acids with probabilities < .05 are predicted to be deleterious.  
Substitutions predicted to be intolerant are highlighted in red.

SIFT: <http://blocks.fhrc.org/sift/SIFT.html>

Note that SIFT does not predict this mutation as being pathogenous while UMD Predictor® predicts that it is indeed a pathogenous mutation as it involves the last nucleotide of the exon and therefore disrupt the donor splice site. To access this level of information, click on the "**Splice site?**" button. This will give you access to the same graphical display used for intronic mutations.



## IV- IMPORT MOLECULAR DATA INTO UMD®

When setting up a new UMD-LSDB, the curator has sometimes data already stored into a tab-delimited file format. In this situation, he can import these data directly into UMD® using the “**Import data**” option from the “**File**” menu.

**Import molecular data**

Note that this option allows only the import of molecular and biological data. To import clinical data use the appropriate option (import clinical data sheets or import clinic from patients)

	Mutation name	Nomenclature (Protein)	Sample ID	Gender	Mutation status	Disease	Proband-Relative (Y/N)	Comments	Reference	Variant type	Transmission	On line	Patient ID	Family ID	Allele
1	c.1521_1523delCCT	p.Phe508del	FR020-00017-09552	Male	Heterozygous	CBAVD	N		123	Mutation	Autosomal recessive	Y	17	9552	
2	c.3151A>G	p.Ile1051Val	FR020-00017-09552	Male	Heterozygous	CBAVD	N	Belgium	123	Mutation	Autosomal recessive	Y	17	9552	
3	c.1521_1523delCCT	p.Phe508del	FR020-00010-0000C	Female	Homozygous	CF	Y	Mexico	122	Mutation	Autosomal recessive	N	10	0000C	
4	c.1521_1523delCCT	p.Phe508del	FR020-00011-0000C	Female	Heterozygous	CF	Y	Mexico	122	Mutation	Autosomal recessive	N	11	0000C	
5	c.3484C>T	p.Arg1162X	FR020-00011-0000C	Female	Heterozygous	CF	Y	Mexico	122	Mutation	Autosomal recessive	N	11	0000C	
6	c.2562T>G	p.Thr854Thr	FR020-00011-0000C	Female	Heterozygous	CF	Y	Mexico	122	Polymorphism	Autosomal recessive	N	11	0000C	
7															
8															
9															
10															
11															
12															

field delimiter:  End of line:

**Select columns to import**

(\*) Mutation nomenclature (c.)

(\*) Sample ID

Gender

Age of onset

Age of death

Date of birth

Date of last follow-up

Mutation status

Mutation type

Transmission

Geographic origin

Phenotypic group

Proband/relative (Y/N)

Comments

Reference #

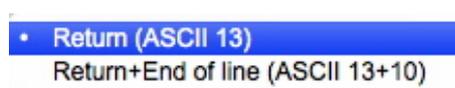
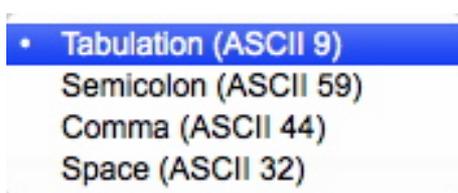
Available on-line (Y/N)

Disease

Allele

(\*) Mandatory items

This option allows the user to select a data file and to display its content into UMD®: The content of the original record is displayed as a table (top). If the file is not displayed correctly, the user can define different parameters for field delimiters and end of line.



When the parameters have been correctly set-up, press the “**Reanalyze file**” button. If the file content is correctly displayed, you can now use the bottom part of the screen to select the various columns to import and define field correspondence. Two fields are mandatory “**Mutation nomenclature (c.)**” and “**Sample ID**”. The other fields are optional but highly recommended. For each field, a pop-up menu allows you to select one of the columns from your file. If you select multiple times the same column, only the last selection will remain active all previous selections will be removed. After the column selection process, press the “**OK**” button. The UMD® software will then check if the international nomenclature name is correct (mutation name will appear in green) or not (mutation name will appear in red). At this stage you can either confirm the import or modify the table content.

Import molecular data

**Import molecular data**

Note that this option allows only the import of molecular and biological data. To import clinical data use the appropriate option (import clinical data sheets or import clinic from patients)

Sample ID	Gender/Phenotypic group	Mutation name	Exon #	Missing exons	Geographic origin/Reference #	Mutation origin	Mutation status	Probands-Relative (Y/N)	Variant type	Date of birth			
1	F3407004561	Male	BMD	c.3150_4071del	11	deletion from exon 11 to 29	FRANCE	106	Familial	Hemizygous	Proband	Mutation	01.03.00
2	F3407004791	Male	BMD	c.3150_4071del	11	deletion from exon 11 to 29	FRANCE	106	Familial	Hemizygous	Proband	Mutation	
3	F5420274261181	Male	BMD	c.94_265dup	3	duplication from exon 3 to 4	FRANCE	117	Unknown	Hemizygous	Proband	Mutation	
4	F671119900941	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 48	FRANCE	121	De novo	Hemizygous	Proband	Mutation	
5	F67939900021	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 47	FRANCE	121	De novo	Hemizygous	Proband	Mutation	
6	F671014700021	Male	BMD	c.129del	1	deletion from exon 1 to 1	FRANCE	121	Unknown	Hemizygous	Proband	Mutation	
7	F54093913981	Male	BMD	c.361_433del	10	deletion from exon 10 to 30	FRANCE	117	Familial	Hemizygous	Proband	Mutation	
8	F3441220891	Male	BMD	c.2792_2945dup	19	Duplication from exon 19 to 51	FRANCE	106	Familial	Hemizygous	Proband	Mutation	
9	F67163300421	Male	BMD	c.6300_7000del	45	deletion from exon 45 to 49	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
10	F67163300431	Male	BMD	c.6430_7000del	45	deletion from exon 45 to 49	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
11	F67344600381	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 48	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
12	F540260241	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 49	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
13	F541579620781	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 47	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
14	F541585148751	Male	BMD	c.94_265dup	3	duplication from exon 3 to 4	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
15	F548018103701	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 47	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
16	F542047626381	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 47	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
17	F5424302801	Male	BMD	c.94_265dup	3	duplication from exon 3 to 7	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
18	F541720222971	Male	BMD	c.6430_7877del	45	deletion from exon 45 to 53	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
19	F5413097027141	Male	BMD	c.6300_7000del	48	deletion from exon 48 to 49	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
20	F6301600071	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 48	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
21	F6301600081	Male	BMD	c.6430_7036del	45	deletion from exon 45 to 48	FRANCE	121	Familial	Hemizygous	Proband	Mutation	
22													

**Confirmer**

21 records out of 22 can be imported correctly.  
Do you want to import these records?

Select columns to import

(\*) Mutation nomenclature (-) Mutation name Transmission Mutation origin  
 (\*) Sample ID Sample ID Geographic origin Geographic origin  
 Gender Gender Phenotypic group Phenotypic group  
 Age of onset Probands-Relative (Y/N) Probands-Relative (Y/N)  
 Age of death Comments Comments  
 Date of birth Date of birth Reference # Reference #  
 Date of last follow-up Available on-line (Y/N) Available on-line (Y/N)  
 Mutation status Mutation status Mutation type Variant type Variant type

(\*) Mandatory items

Validator Quitter

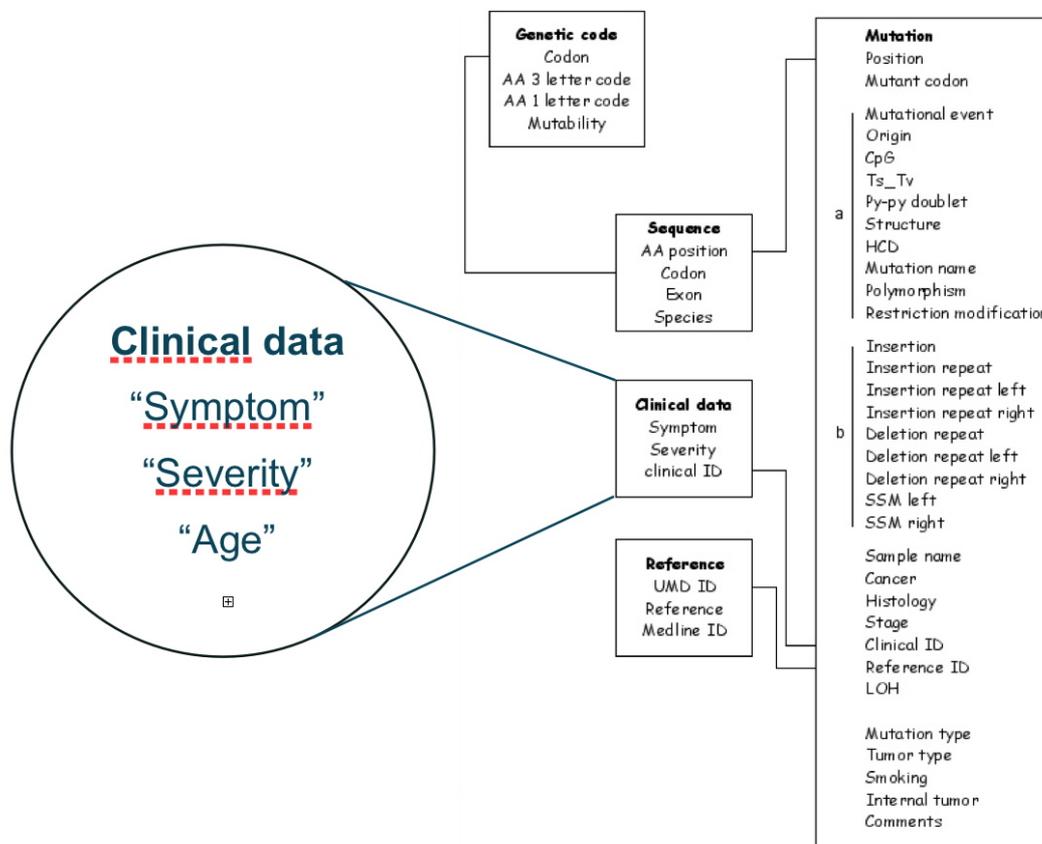
After the import process, mutations can be checked using the **“View-Modify mutations”** option from the **“Show all mut.”** menu.

100 records found.

Sample ID	Family ID	Gender	Phenotypic group	Mutation name	Exon #	Missing exons	Geographic origin	Reference #
F5912300111	123	Male	DMD	c.32_94dup	1-2	Duplication from exon 2 to 2		128
F59111011111	111	Male	BMD	c.32_650dup	1-2	Duplication from exon 2 to 7		128
F59234012341	234	Male	DMD	c.32_1603dup	1-2	Duplication from exon 2 to 13		128
F5945612341	456	Male	DMD	c.32_94dup	1-2	Duplication from exon 2 to 2		128
F5945512221	455	Male	DMD	c.32_94dup	1-2	Duplication from exon 2 to 2		128
F5945612231	456	Male	DMD	c.32_94dup	1-2	Duplication from exon 2 to 2		128
F5946762131	467	Male	DMD	c.32_2293dup	1-2	Duplication from exon 2 to 18		128
F5974514251	745	Male	BMD	c.94_265dup	3	Duplication from exon 3 to 4		128
F5987845631	878	Male	BMD	c.94_265dup	3	Duplication from exon 3 to 4		128
F5936512361	365	Male	DMD	c.94_650dup	3	Duplication from exon 3 to 7		128
F5985478961	854	Male	BMD	c.94_650dup	3	Duplication from exon 3 to 7		128
F5985296321	852	Male	DMD	c.32_94dup	1-2	Duplication from exon 2 to 2		128
F5932114781	321	Male	DMD	c.94_2169dup	3	Duplication from exon 3 to 17		128
F5955425631	554	Male	BMD	c.187_265dup	4	Duplication from exon 4 to 4		128
F5966625871	666	Male	DMD	c.187_650dup	4	Duplication from exon 4 to 7		128
F5999845631	998	Male	DMD	c.650_1483dup	7-8	Duplication from exon 8 to 12		128

## V- IMPORT CLINICAL DATA INTO UMD®

The structure of the UMD® LSDB is composed of various tables linked by unique fields (relational database). The central table in UMD® is the mutation that is linked to reference sequences, as well as the clinical data as shown below:



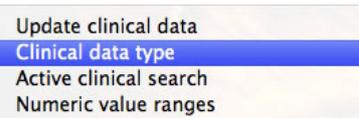
In order to have a very flexible structure that can be adapted to any data set, we designed the clinic table with 3 fields: "**Symptom**", "**Severity**" and "**Age**". The two first are alphanumeric fields that can handle text or numeric values while the third includes only numeric values. As a virtually unlimited number of records can be associated to a single sample (mutation), the user can easily add new clinical descriptions without any modification of the UMD® structure.

One possible drawback of this approach is that it is harder to query numeric values that are included in an alphanumeric field. To circumvent this we have design specific parameters available through the "**File**" menu:

- Clinical data type
- Numeric value ranges

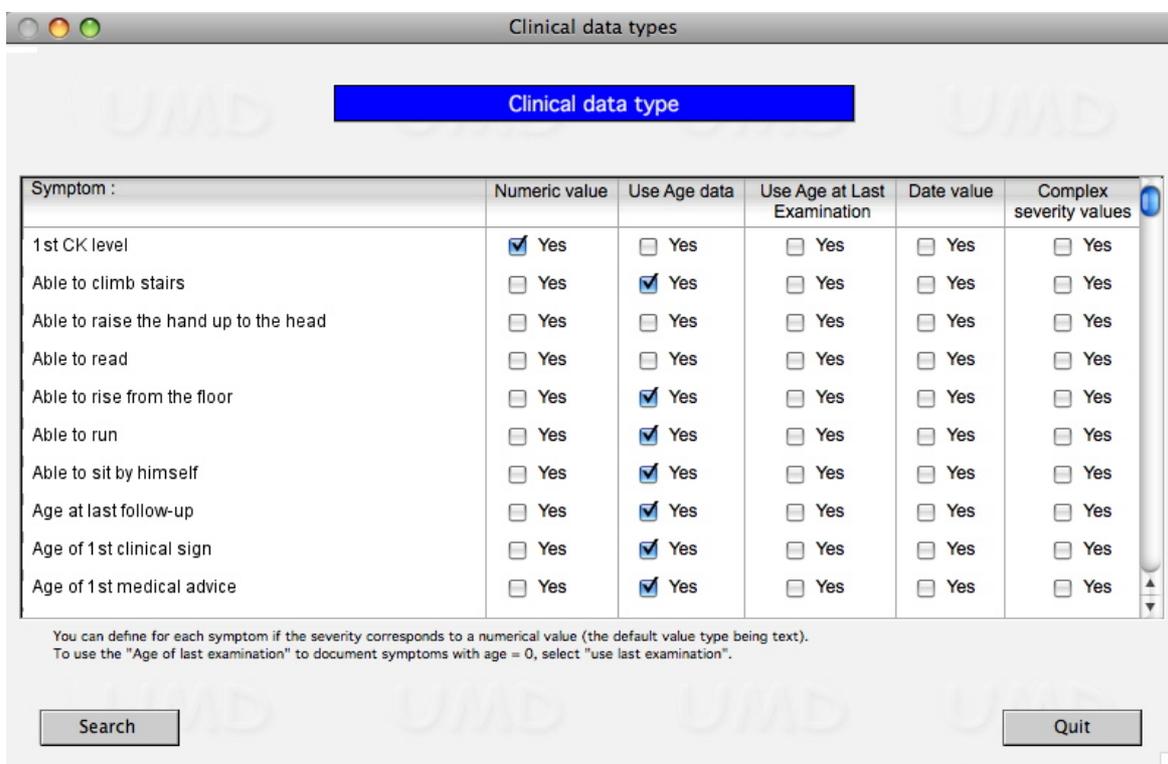
## Clinical data type

This option allows the user to define for each symptom if the associated severities are either numerical values, dates or simple text. In addition, the user can choose to use the “Age” field or the “Age at last examination” symptom. In fact many users created an “Age at last examination” symptom in order to record when was evaluated for the last time the patient. Finally, when complex answers are entered in the severity field, the user can specify it to the software “Complex severity values”. This is particularly useful for symptoms such as “Initial symptom(s)”. In this situation the user can enter multiple answers. To do so he should follow a simple rule:



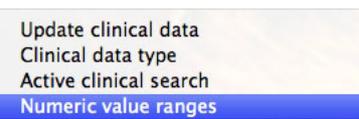
- Each category is separated by a coma
- Sub categories are included in parenthesis and separated by comas

For example: “Motor abnormalities (walk, weakness), Toe walking” means that the patient presented with 2 motor abnormalities (walk and weakness) and toe walking.

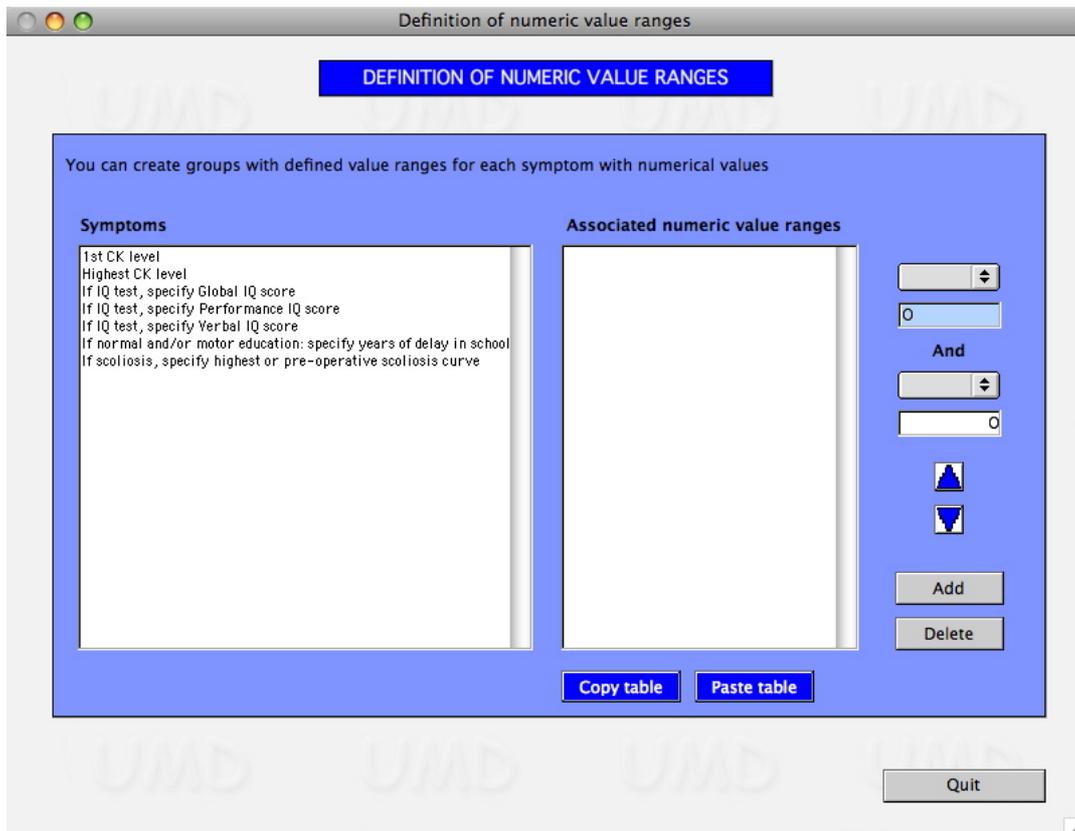


## Numeric value ranges

When using numerical values, it is often necessary to define groups in order to analyze data. The “**Numeric value ranges**” option from the “**File**” menu allows to define these groups for symptoms for which the user has defined as numerical values (1st CK level in the previous picture).



The user has access to the following display:



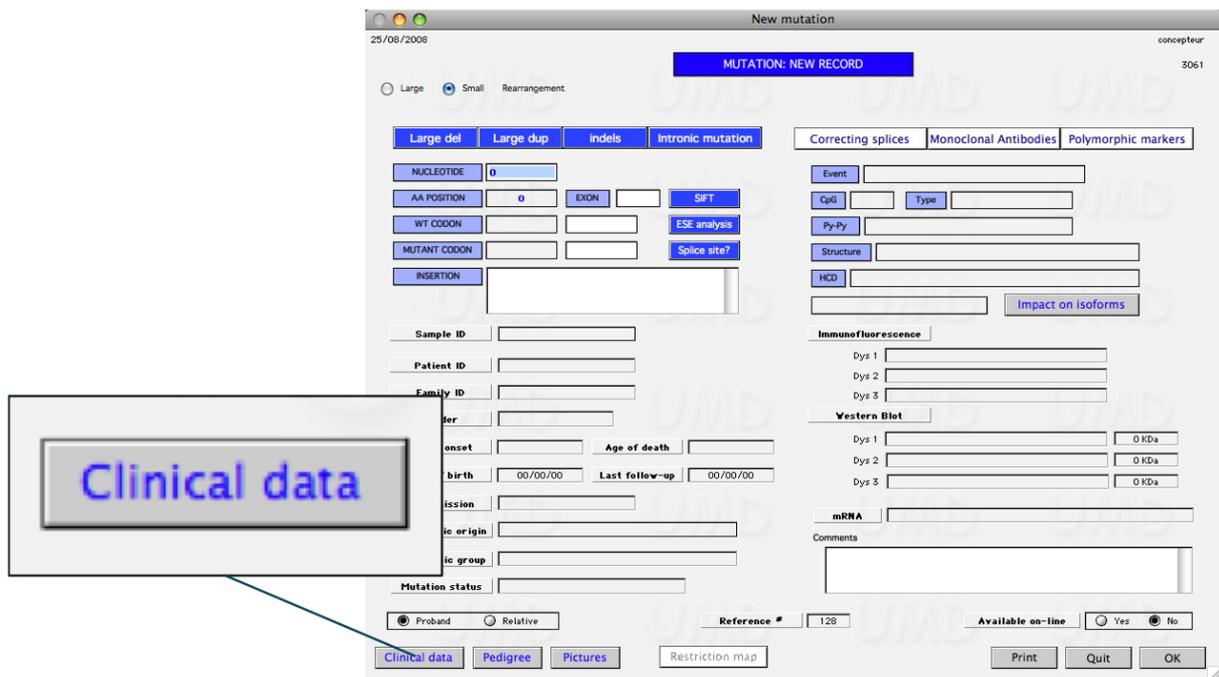
Click on the symptom of interest (left) and define the various groups with the **“Add”** button. For each group define the limits with the 2 pop-up menus (right). If you want to use the same groups for various symptoms, just use the **“Copy table”** and **“Paste table”** buttons. The various groups will be available for statistical analysis (see below).

### ***V-a- Various options to import clinical data into UMD®***

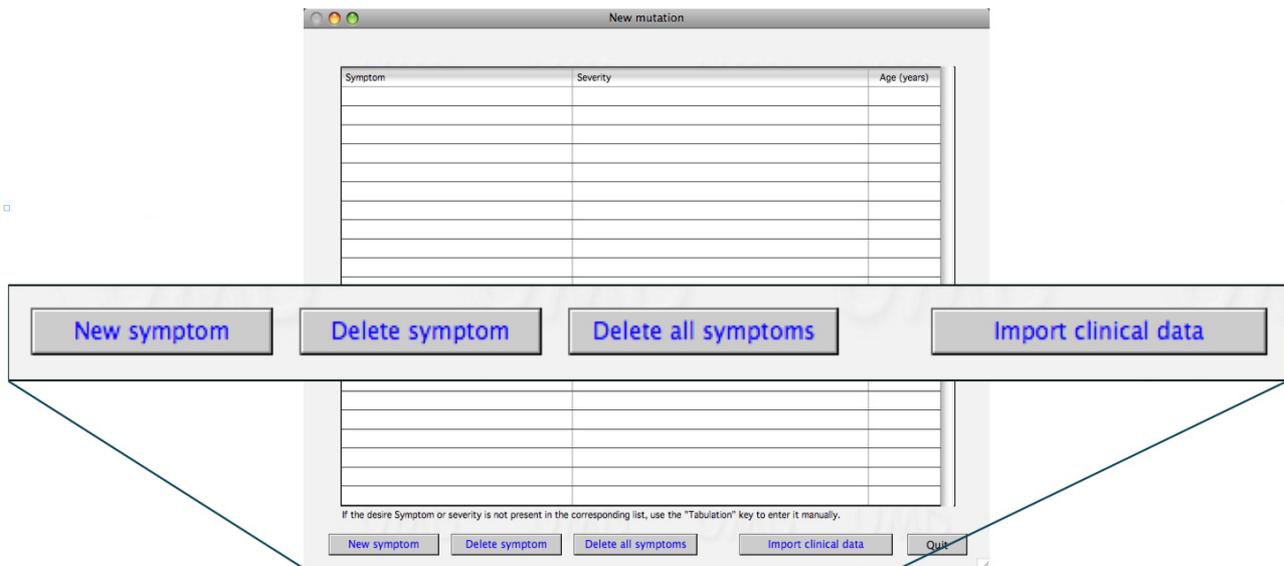
Clinical data can be imported into the UMD® software using four different options:

- Manually during the mutation’s record creation
- Via a file (automatic import) during the mutation’s record creation
- Via multiple individual files
- Via a single file containing data from multiple patients

#### **1) Manual input during mutation’s record creation**



On the first screen of the mutation input, a **“Clinical data”** button allows to access to the



clinical data screen.

The user has now access to a simple table containing the three columns: “Symptom”, “Severity” and “Age”. Four buttons can be used to create **“New symptom”**, delete one symptom **“Delete symptom”** or all symptoms **“Delete all symptoms”**. A fourth button **“Import clinical data”** can also be used (cf. 2).

When the user clicks on the **“New symptom”** button, he creates a new line in the table. This can be noticed by the 0 value that appears in the “Age” column. He can now double click in the “Symptom” box to get the “Symptom list” in which he can add, delete, sort symptoms. This list is automatically updated when UMD® is launched and will only contain data present in records. It is therefore not necessary to build a list at first as non-used data will be discarded when the software will be re-started. Just add a new symptom when necessary.

When a symptom has been selected from the list, a click on the severity box will display the “Severity list”. This list is updated based on the selected symptom and data already present in records. As for symptoms, it is unnecessary to create a long list of severity as unused choices will be discarded when the software will be re-started.

## 2) Import of clinical data for a single patient

Clinical data are often collected via collaborators. The easiest way is therefore to create ascii files (text tab-delimited files) that can be easily imported during the record creation. To do so, create a table with 2 columns (symptom and severity/age). Each symptom has a label and can include notes, such as the various severity, in parenthesis. These notes will be removed during the import process. An example is given below:

Symptom	Severity [Age]
Phenotype ( <i>DMD, BMD, DMC, IMD, symptomatic carrier, asymptomatic carrier</i> )	DMD
Age of 1st clinical sign ( <i>Birth /[age] in years /UK</i> )	[3.5]
Initial symptom ( <i>Motor DD (developmental delay) /Cognitive DD /Motor and cognitive DD /Motor abnormalities (congenital hypotonia, weakness, walk, climbing stairs, running) /Fatigue /Pain: cramping and/or myalgia /Muscle hypertrophy /Toe walking /Cardiomyopathy /UK</i> )	Motor abnormalities (weakness)
Age of 1st medical advice ( <i>specify [age] in years/UK</i> )	[3.75]
Diagnostic circumstances ( <i>Clinical sign /Elevated CK /Familial context /Familial context and clinical sign /UK</i> )	Clinical sign
1st CK level ( <i>in X normal values if available /if no in UI /I /elevated CK /normal CK /UK, specify [age in years]</i> )	
Highest CK level ( <i>in X normal values if available /if no in UI/I /elevated CK /normal CK /UK, specify [age of measurement in years]</i> )	
Sitting acquisition ( <i>Y /Never /UK, if Y: specify [age of acquisition in months]</i> )	Y
Walking alone ? ( <i>Y /Never /UK, if Y: specify [age of acquisition in months]</i> )	Y [24]
Able to run ? ( <i>Y or MFM 30th item = 2-3 or Walton score = 0 /N or MFM 30th item = 0-1 or Walton score ≥ 1 /Never /UK, if N: specify [age of losing the ability]</i> )	Never

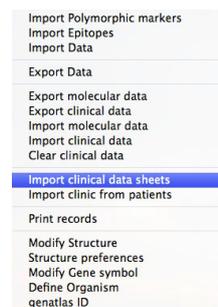
This file structure allows to give directives and help to clinicians and curators during the data collection.

To import such a file into UMD<sup>®</sup>, use the “**Clinical import data**” button (see upstairs).

### 3) Global import of individual records

If multiple data are collected from collaborators, it is easier to store each individual data file in the same directory and use the **“Import clinical data sheets”** option from the **“File”** menu.

With this option, the user selects a directory that contains all individual data files. The name of each file should correspond to the sample ID used in the mutation record.



### 4) Global import of individual records

When a single curator is collecting clinical data, he usually likes to store all data in a single file. To do so, he should create a text tab-delimited file that has the following structure:

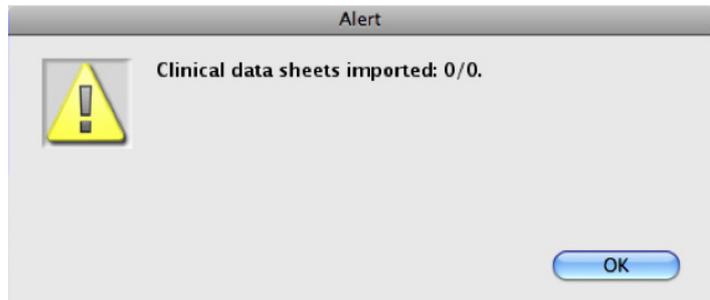


- The title of each symptom is in the first row
- Each column contains data for a single symptom
- The first column contains the sample ID
- Each row contains data from a single patient

Sample ID	Diagnosis (Phenotype: Duchenne Muscular Dystrophy (DMD) / Becker Muscular Dystrophy (BMD) / Intermediate Muscular Dystrophy (IMD) / Dilated cardiomyopathy (DCM) / symptomatic carrier / asymptomatic carrier / Unknown / I don't know)	Currently able to walk without support (10 steps without another person's help, with or without orthoses: Y or Walton score $\leq 6$ or MFM 29th item = 2-3 /N or Walton score $\geq 7$ or MFM 29th item = 0-1 /Never /UK, if N: specify [age of losing the ability])	Currently able to sit without support (Able to stand from sitting position without help: Y or MFM 13th item = 2-3 /N or MFM 13th item = 0-1 /Never /UK, if N: specify [age of losing the ability])	Wheelchair use (if over 3 years of age: Y (permanent) / Y (intermittent) / Never / UK, if Y: specify [age])	Scoliosis Surgery (Y /N /UK, if Y: specify [age at surgery])
F754006011	DMD	N [11]	Y	Y (permanent) [8,44]	N
F1300100931	DMD	N [10,33]	Y	Y (permanent) [9]	N
F1300210751	DMD	N [9]	Y	Y (permanent) [7,75]	N
F1300307401	DMD	N [11,25]	Y	Y (permanent) [10,25]	N
F1300410231	IMD	N [14,5]	Y	Y (permanent) [12,25]	N
F1301401051	DMD	N [9]	N [14]	Y (permanent) [10]	Y [14,44]

After selecting the ***“Import clinic from patients”*** option of the ***“File”*** menu, the user is prompted to select a text file containing clinical data. The UMD® software will then check for the sample ID and if such a sample exist will import the clinical data. If previous data are available for this sample, they are discarded. This process should therefore be used carefully. At the end of the process, a message will inform the user of the number of correctly imported records:

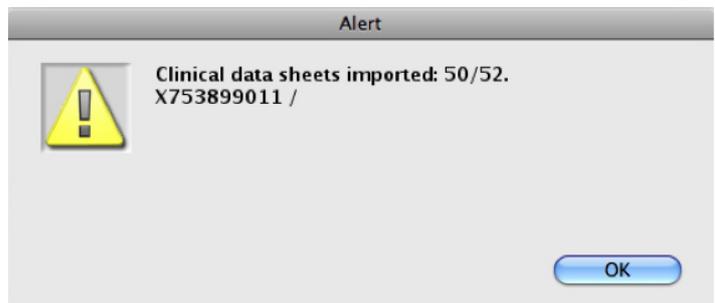
No mutation record has been found with a sample ID corresponding to data from the imported file.



The first column of the imported file is not labelled sample ID and therefore the import process has failed.



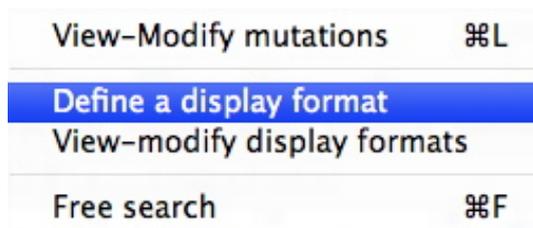
Only 50 out of the 52 rows with clinical data from patients have been successfully imported. The rows corresponding to sample ID ***“X753899011”*** and ***“”*** do not match any mutation record.



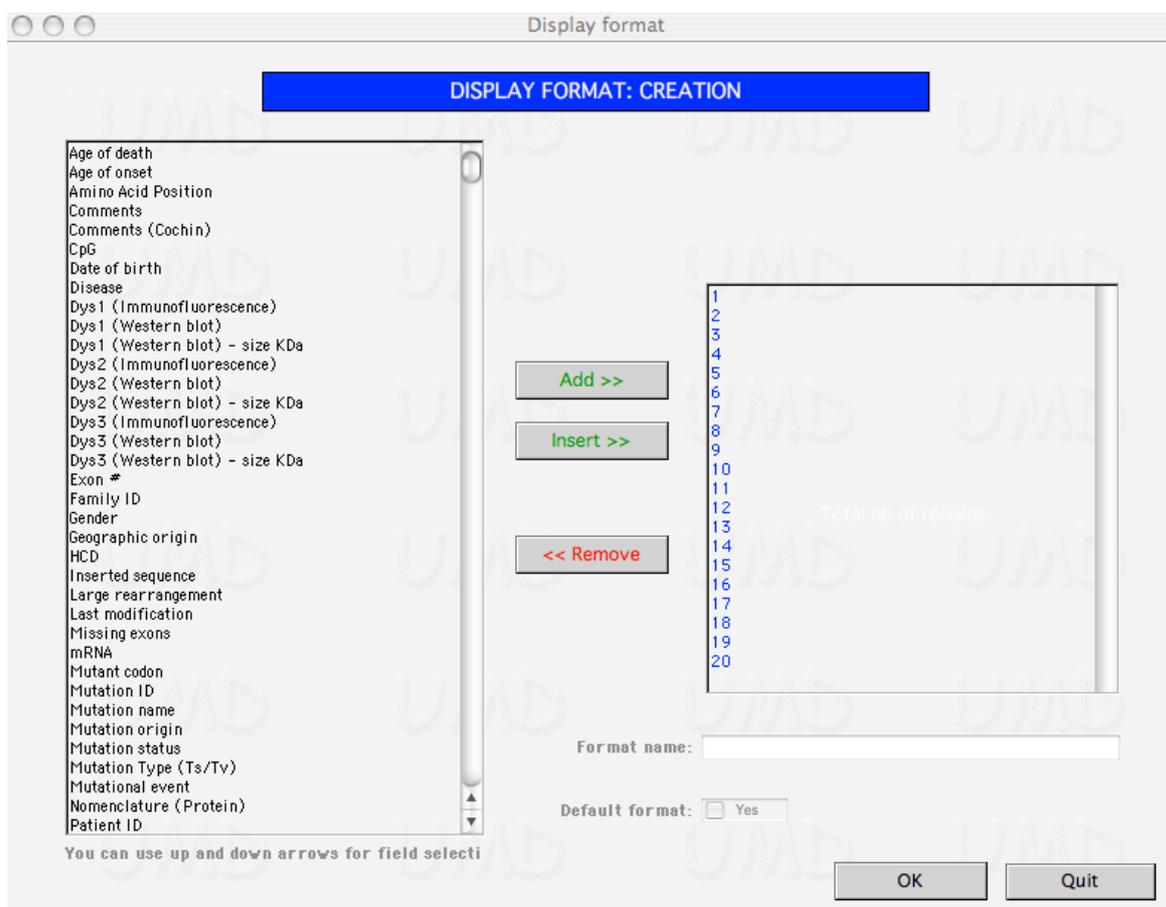
## VI DISPLAY FORMATS

### VI-a- Creation of display formats

Because all users have different objectives, they would like to visualize data in different formats. Before using the “**View-Modify mutations**” option of the “**Show all mut.**” menu, you then first need to define a display format. To do so, use the “**Define a display format**” option of the “**Show all mut.**” menu.



The following screen is then displayed:

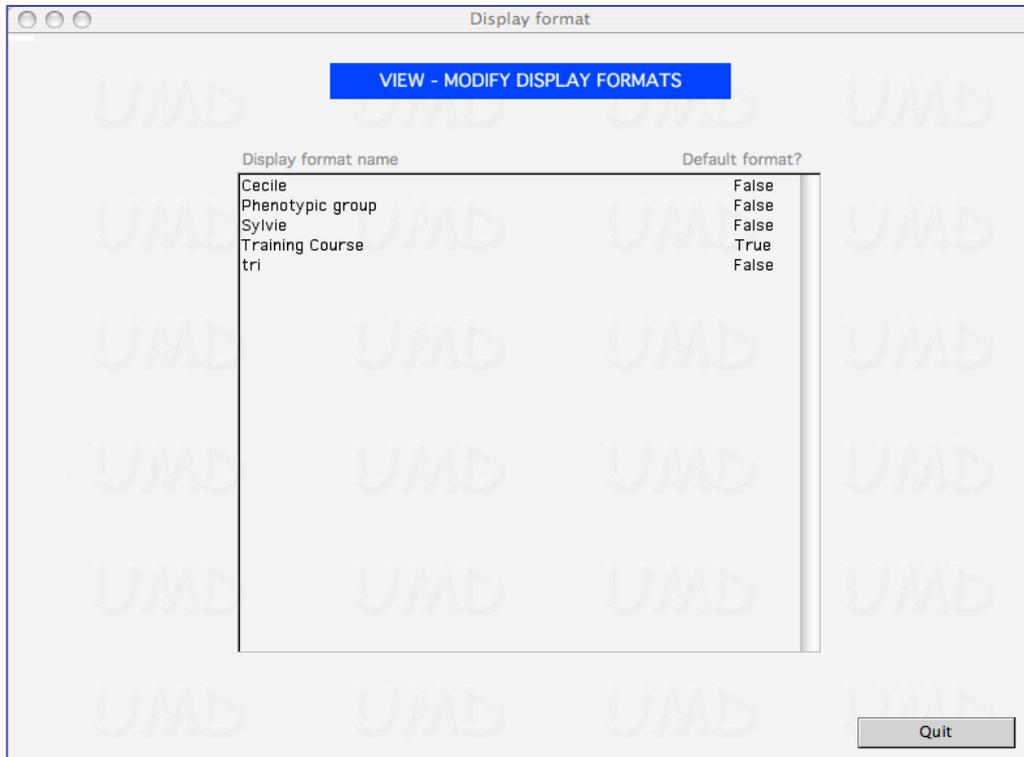


The left table contains the various available fields (except clinical data). You can select up to 20 of these fields to be inserted in a specific display format. Use the “**Add**”, “**Insert**” or “**Remove**” buttons to add fields in the right table.

Give a name to this format and decide if it needs to be the default format for display. You can create many formats but only one is active at a time.

## VI-b- Modification of display formats

To modify a display format use the “**View-modify display formats**” option from the “**Show all mut.**” menu. A list of all available display formats is available. The default format is labelled “True”.



To modify a format, double-click on the corresponding line. This allows a rapid change of the default format or a more complex change of the format itself. Note that if no default format is selected the following generic format is used:

Show mutations / generic display

Record	Nomenclature	Position	codon			AA		Event/ Consequence	Sample ID	Geographic origin	CpG	Type	Ref
			WT	Mut	exon	WT	Mut						
1987	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F671014700211	FRANCE	Fr.	121	
1484	c.1_7542del	1	ATG	del75	1	Met	Inf	In frame del	F3501803711	FRANCE	Inf	110	
1558	c.1_960del	1	ATG	del96	1	Met	Inf	In frame del	F3412607791	FRANCE	Inf	106	
1592	c.1_11055del	1	ATG	del11	1	Met	Inf	In frame del	F3422412861	FRANCE	Inf	106	
1594	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F3422612531	FRANCE	Fr.	106	
1863	c.1_649del	1	ATG	del64	1	Met	Fs.	Stop at 13	F1307409661	FRANCE	Fr.	115	
1909	c.1_649del	1	ATG	del64	1	Met	Fs.	Stop at 13	F3804695dm071		Fr.	113	
2415	c.1_831del	1	ATG	del83	1	Met	Inf	In frame del	F9401901691	FRANCE	Inf	126	
2430	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F9403405291	FRANCE	Fr.	126	
2529	c.1_6290del	1	ATG	del62	1	Met	Fs.	Stop at 3	F9408107731	FRANCE	Fr.	126	
2476	c.1_11055del	1	ATG	del11	1	Met	Inf	In frame del	F5903624991	FRANCE	Inf	127	
2562	c.1_649del	1	ATG	del64	1	Met	Fs.	Stop at 13	F67734000071	FRANCE	Fr.	121	
366	c.-648_Ddup	1	ATG	ins64	1	Met	Fs.	Stop at 228	F750642011		Fr.	1	
1820	c.1_6290del	1	ATG	del62	1	Met	Fs.	Stop at 3	F1302912311	FRANCE	Fr.	115	
703	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F755715011		Fr.	1	
727	c.1_11055del	1	ATG	del11	1	Met	Inf	In frame del	F750205011		Inf	3	
2914	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F750481011		Fr.	12	
2915	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F75048106		Fr.	12	
3007	c.1_31del	1	ATG	del31	1	Met	Fs.	Stop at 15	F755656011	FRANCE	Fr.	1	
3014	c.1_6438del	1	ATG	del64	1	Met	Inf	In frame del	F750171011	FRANCE	Inf	1	

2411 record(s) found.

Sort    Search in selection    Print    Quit

When you use a display format, you create a list that includes selected columns. All formats share a common set of buttons:



Whose functions are:

	To delete selected records and associated data (clinics, pictures).		To print the list of records using the columns from the display format
	To display all records.		To create or load a list of records.
	To reduce the list to selected records.		To quit the list.
	Gives access to the search interface to select mutations using multiple criteria.	 Ref # <input type="text"/> Medline # <input type="text"/>	To select records associated to a specific reference.

 Sort	To sort records using one or multiple criteria.	 Unique Mutations	To search unique mutations associated to a selected reference.
 Export	To export data limited to the columns from the display format.	 On-line	To set the status of all records from the list to "available on-line".

Note that you can select multiple lines using the "Command" (Mac) or "Control" (PC) buttons. This allows you to either reduce the selection to these selected records or to delete them.

## VII SEARCH INTERFACES

In order to facilitate the selection of records, two search interfaces can be used within the UMD<sup>®</sup> software: the "**UMD-web search interface**" and the "**Regular search interface**".

### VII-a- UMD-web search interface

This interface gives access to a limited list of fields from the "Mutation" table and to the 3 fields from the "Clinical data" table. The user can combine criteria and perform multiple searches. The use of the "Search Database (mutations criteria)" button use the fields from the "Mutation" table while the "Search Database (clinical data criteria)" button use the fields from the "Clinical data" table. The number of records corresponding to the request is display as a "Temporary selection". If the user wants to use this set of records, he can transfer the temporary records into the "Final selection".

The list of available fields from the "Mutation" table is listed below:

Age of death  
Age of onset  
Amino Acid Position  
Comments  
CpG  
Date of birth  
Date of last follow-up  
Disease  
Dys1 (Immunofluorescence)  
Dys1 (Western blot)  
Dys1 (Western blot) - size KDa  
Dys2 (Immunofluorescence)  
Dys2 (Western blot)  
Dys2 (Western blot) - size KDa  
Dys3 (Immunofluorescence)  
Dys3 (Western blot)  
Dys3 (Western blot) - size KDa  
Exon #  
Family ID  
Gender  
Geographic origin  
HCD  
Inserted sequence  
Large rearrangement  
Last modification  
Missing exons  
mRNA  
Mutant codon  
Mutation ID  
Mutation name  
Mutation origin  
Mutation status  
Mutation Type (Ts/Tv)  
Mutational event  
Nomenclature (Protein)  
Patient ID  
Pedigree ID  
Phenotypic group  
Proband-Relative (Y/N)  
Protein  
Py-Py doublet  
Reference #  
Reference keyword  
Reference Medline ID  
Restriction enzymes  
Sample ID  
Species  
Structure  
Variant type  
Wild Type Codon

The "Final AND Temporary", "Final OR temporary", "Final EXCEPT Temporary" and "CLEAR Final Selection" buttons allow a progressive selection of records combining results from the "Final" and the "Temporary" selections.

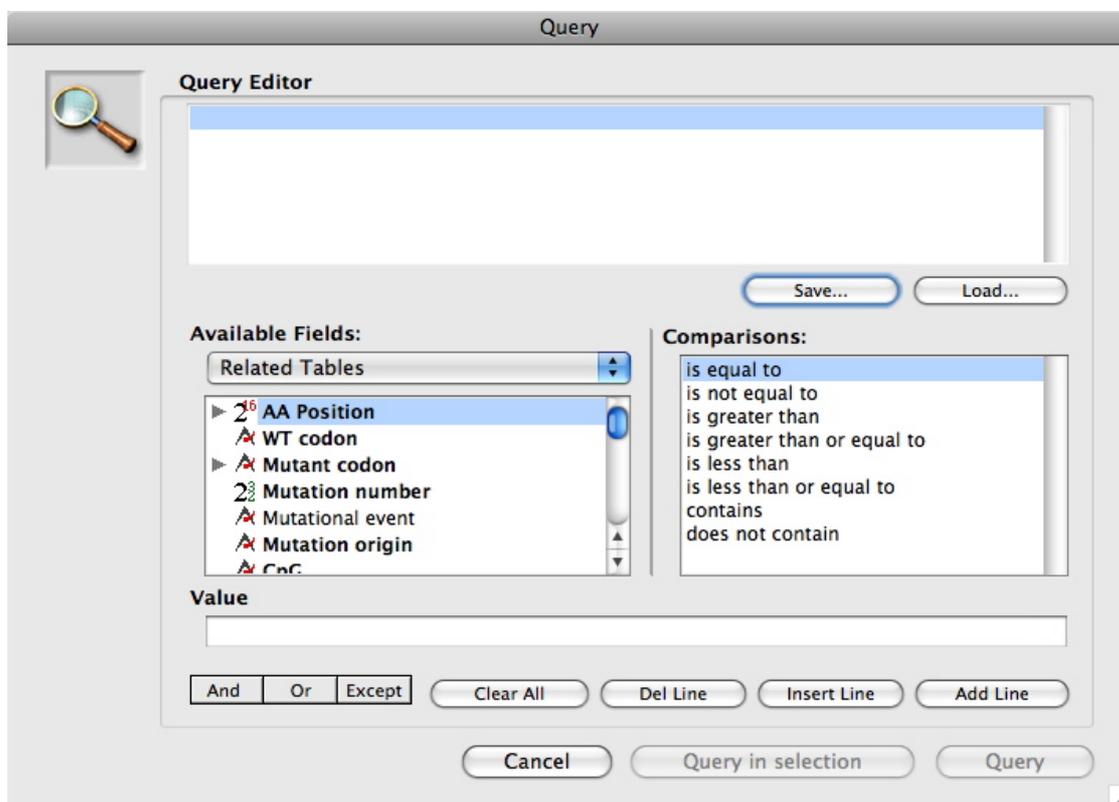
"Final AND Temporary": Records should belong to the "Final" and the "Temporary" group of records to be selected;

"Final OR temporary": Records should belong to either the "Final" or the "Temporary" group of records to be selected;

"Final EXCEPT Temporary": Records should belong to the "Final" but not to the "Temporary" group of records to be selected;

### VII-b- Regular search interface

This interface gives access to all fields from the "Mutation" and related tables from the UMD database. Note that all fields are available even the ones that are not relevant for your



database. We therefore recommend that you use the "**UMD-web search interface**" instead.

If you are familiar with the 4<sup>th</sup> Dimension software you may choose to use this regular interface. Note that you can add as many queries as necessary in the Query box using the "Add Line", "Insert Line" and "Del Line" button. You can combine queries using the "And", "Or" and "Except" buttons.

In addition, you can "Save" and "Load" your queries with the corresponding buttons.

## VIII DESCRIPTION OF ALL FUNCTIONS

All functions of the UMD® software are available through eleven menus: "**File**", "**References**", "**Mutations**", "**Pedigree**", "**Show all mut.**", "**Statview**", "**Phenotype**", "**Genotype**", "**Haplotype**", "**Introns**", "**Modules**" and "**Variations**".

### VIII-a- The "File" menu

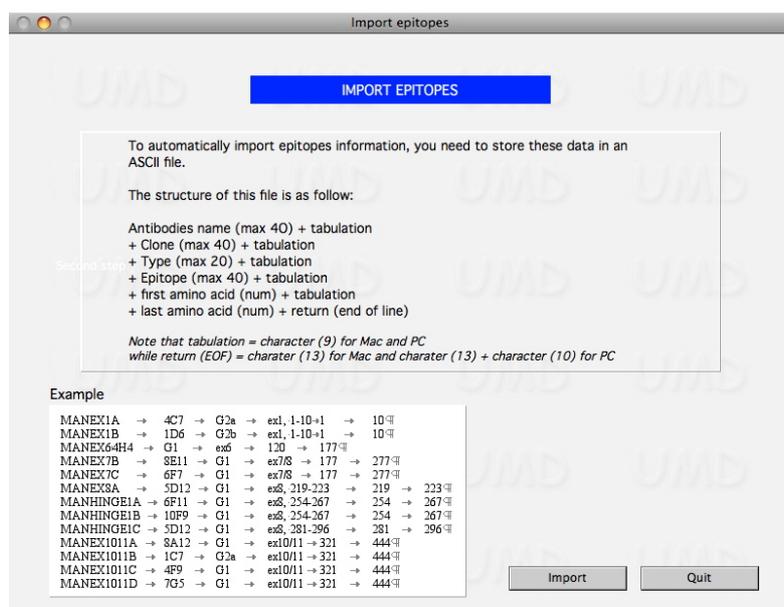
This menu contains 48 functions, which are mainly use to set preferences, import/export data and activate specific tools:

#### 1) Import Polymorphic markers

This function is used to import SNP data (cf. II-b).

#### 2) Import Epitopes

This function is used to import data about antibodies. The file must



contain 6 columns as indicated below:

After the initial import, data can be modified using the "Modify epitopes" function (cf. VIII-a-x).

#### 3) Import Data

This function allows the import of molecular data from any source (cf. IV).

#### 4) Export Data

This function allows the export of all data (molecular and clinical) from a specific set of records freely chosen by the user. Note that clinical data should have been correctly defined within the UMD software by using the "Update clinical data" function (cf. VIII-a-x).

Import Polymorphic markers
Import Epitopes
Import Data
Export Data
Export molecular data
Export clinical data
Import molecular data
Import clinical data
Clear clinical data
Import clinical data sheets
Import clinic from patients
Print records
Modify Structure
Structure preferences
Modify Gene symbol
Define Organism
genatlas ID
Display Genetic Code
Modify Sequence
Modify Exons
Gene annotation
Exon phasing
Display coding sequence
Display genomic organization
Display Splice Sites
Modify polymorphic markers
Modify Epitopes
Modify SIFT data
Define links
Create a new database
Gene type
Update clinical data
Clinical data type
Activate clinical search
Numeric value ranges
Search preference
Splice sites calculation
Pathname for dnamlk application
Activate Web server
Authorize Web Services
Activate Specific options
Specific Web Parameters
Restricted access
Web license
4D_molecules folder
Password
Import passwords
Quit

## 5) Export molecular data

This function is used to combine data from multiple curators. It should be used with the "*Import molecular data*" function.

## 6) Export clinical data

This function is used to combine data from multiple curators. It should be used with the "*Import clinical data*" function.

## 7) Import molecular data

This function is used to combine data from multiple curators. It should be used only with a file generated by the "*Export molecular data*" function.

## 8) Import clinical data

This function is used to combine data from multiple curators. It should be used only with a file generated by the "*Export clinical data*" function.

## 9) Clear clinical data

This function is used to clear all clinical data from the database.

## 10) Import clinical data sheets

This function is used to import multiple individual data sheets (cf. V-a-3).

## 11) Import clinical from patients

This function is used to import clinical data from multiple patients stored in a single file (cf. V-a-4).

## 12) Print records

This function is used to print all records from the database. A specific password is required to activate this function. Please contact [Christophe Bérout](#) to get this specific password.

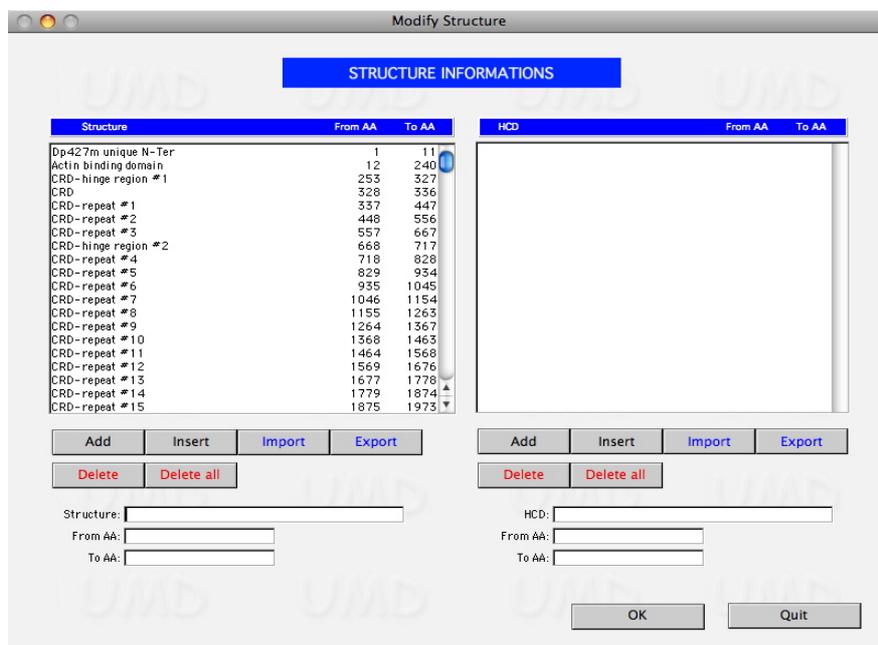
## 13) Modify Structure

This function is used to annotate the structural domains of the protein as well as Highly Conserved Domains (HCD), which represent key residues of the protein. These HCD are used to predict the pathogenicity of missense mutations and therefore a particular attention should be given to their annotation.

The user can either create the various Structural domains and HCD directly into UMD (see below) or import corresponding data from text tab-delimited files containing 3 columns: Structural domain name or HCD name; first amino acid; last amino acid.

Note that structural domains can not overlap and therefore they should be subdivided into unique entities. So a Nuclear localization signal (NLS) located at position 10 to 100 and a

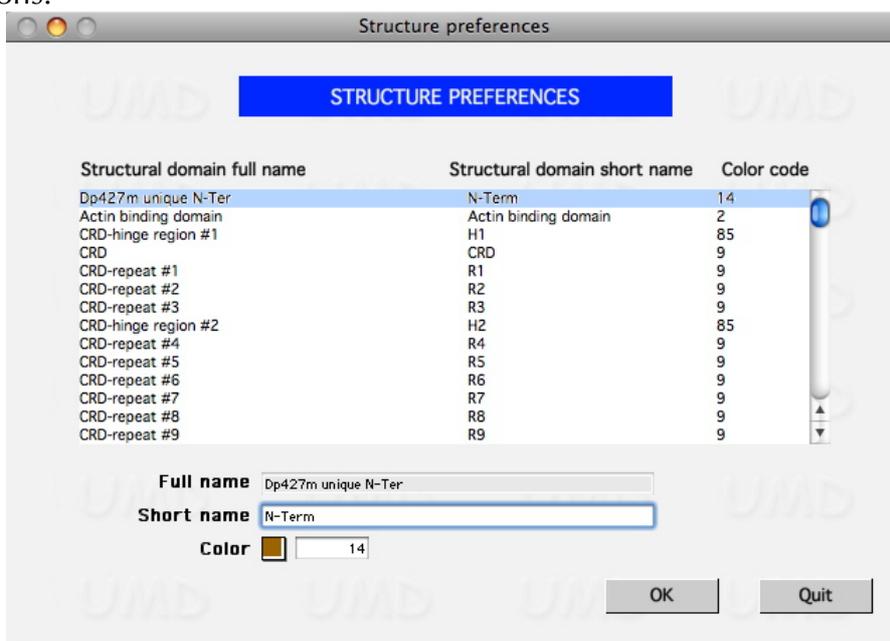
ring domain localized at position 20 to 35 should be annotated: a) NLS 10 to 19, b) NLS + Ring 20 to 35 and c) NLS 36 to 100.



Note that the "Add", "Insert", "Delete" and "Delete all" buttons allow the creation or suppression of rows in the corresponding table. After selection of a specific row, the corresponding data can be modified in the fields below the table (Structure/From AA/To AA fields or HCD/From AA/To AA).

#### 14) Structure preferences

This function is used to specify parameters of the structural domains used for graphical display functions.



The user can define a short name for each structural domain and use a color's range of 256 colors.

## 15) Modify Gene symbol

This function is used to specify the gene symbol during the creation process of a new database (cf. II-a).

## 16) Define organism

The user can choose between Human and Mouse organism. This will result in different matrices to search for Branch-point sequences.

## 17) Genatlas ID

This function is used to define the Genatlas ID associated to the Gene. This will be used in a future version of the UMD software to automatically collect annotations.

## 18) Display Genetic Code

This function is used to display the genetic code using the 3 letters- and the one-letter codes.

## 19) Modify sequence

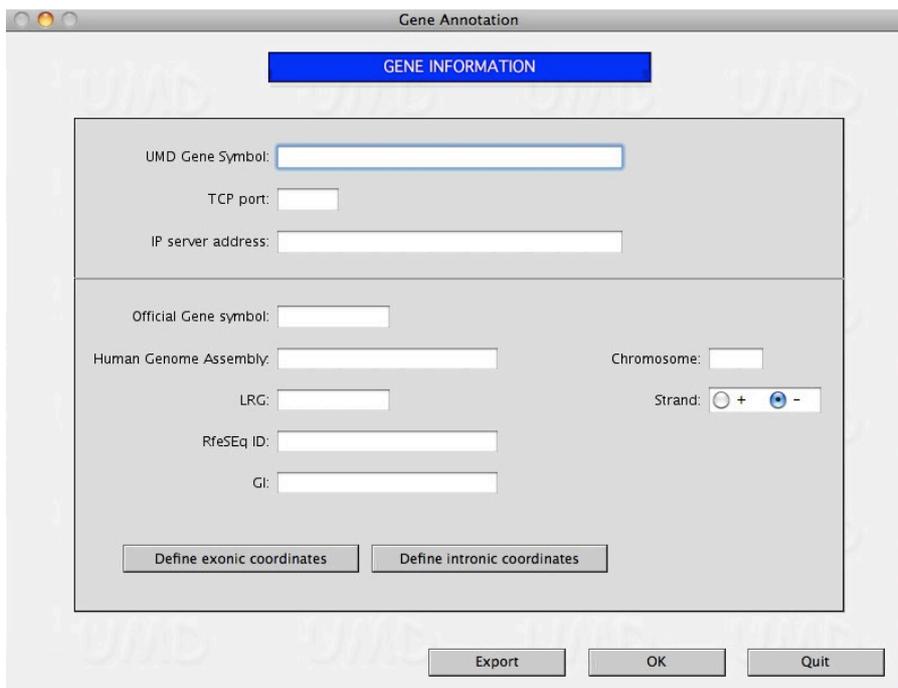
This function is used to display the cDNA reference sequence and eventually to modify it.

## 20) Modify exons

This function is used to display the various exons of the gene and their relative position in the cDNA reference sequence and eventually to modify these data.

## 21) Gene annotation

This function is used to specify gene information.



The screenshot shows a window titled "Gene Annotation" with a "GENE INFORMATION" tab. The form contains the following fields and controls:

- UMD Gene Symbol:
- TCP port:
- IP server address:
- Official Gene symbol:
- Human Genome Assembly:
- Chromosome:
- LRG:
- RfSeq ID:
- GI:
- Strand:  +  -
- Buttons: Define exonic coordinates, Define intronic coordinates, Export, OK, Quit

It has been designed to facilitate cross-references of the database using different reference systems such as: The Human Genome Assembly, the LRG and RefSeq systems as well as the GI number from ncbi.

Once the general information is provided, the user can define each coordinate using the “define exonic coordinates” and “define intronic coordinates” buttons.

**Gene Annotation**

**GENE INFORMATION**

**Exons' annotation**

exon	cDNA start	cDNA End	Chr. start	Chr. end	LRG start	LRG end	RefSeq start	RefSeq end
1	0	0	0	0	0	0	0	0
2	1	63	0	0	0	0	0	0
3	64	136	0	0	0	0	0	0
4	137	231	0	0	0	0	0	0
5	232	342	0	0	0	0	0	0
6	343	444	0	0	0	0	0	0
7	445	528	0	0	0	0	0	0
8	529	678	0	0	0	0	0	0
9	679	867	0	0	0	0	0	0
10	868	1053	0	0	0	0	0	0
11	1054	1260	0	0	0	0	0	0
12	1261	1353	0	0	0	0	0	0
13	1354	1467	0	0	0	0	0	0
14	1468	1644	0	0	0	0	0	0
15	1645	1812	0	0	0	0	0	0

To import data, use the format: exon number (tabulation) start position (tabulation) end position (end of line)

Import Chr. data    Import LRG data    Import RefSeq data    Back

Here is presented the screen linked to the “define exonic coordinates” button. The user can import data from TSV files or manually update each information

## 22) Exon phasing

Gives access to a graphical display of the various exons. This is particularly useful to evaluate the impact of the deletion of one or more exons on the framing.

**Exon phasing**

1 2 3 4 5 6 7 8 9 10

11 12 13 14 15 16 17 18 19 20

21 22 23 24 25 26 27 28 29 30

31 32 33 34 35 36 37 38 39 40

41 42 43 44 45 46 47 48 49 50

51 52 53 54 55 56 57 58 59 60

61 62 63 64 65 66 67 68 69 70

71 72 73 74 75 76 77 78 79

Save    Quit

## 23) Display coding sequence

The screenshot shows a window titled "File: Display coding sequence". It displays a cDNA sequence with 30 codons per line, grouped into seven exons (Exon #1 to Exon #7). Each codon is shown above its corresponding amino acid translation. For example, the first codon is ATG (Met) and the last is CCT (Pro). The amino acids are abbreviated (e.g., Met, Leu, Trp, etc.). The window includes "Save" and "Quit" buttons at the bottom right.

This function gives access to a graphical presentation of the reference cDNA sequence.

## 24) Display genomic organization

The user has access to the various exons of the gene with a freely selected portion of the

The screenshot shows a window titled "View exons" with a sub-header "View exonic and intronic sequences". On the left, there is a vertical list of exons numbered 1 through 20. Exon 6 is currently selected. The main area displays the "Full sequence of the selected exon" as a single line of text. The sequence is: `tccttgctcaa ggaatgcatt ttcttatgaa aatttatttc cacatgtagG TCAAAAATGT AATGAAAAAT ATCATGGCTG GATTGCAACA AACCAACAGT GAAAAGATTC TCCTGAGCTG GGTCCGACAA TCAACTCGTA ATTATCCACA GGTAAATGTA ATCAACTTCA CCACCAGCTG GTCTGATGGC CTGGCTTTGA ATGCTCTCAT CCATAGTCAT AGgtaagaag attactgaga cattaataaa cttgtaaaag tggtgattta ga`. A note at the bottom states: "Intronic sequences are displayed in small characters, Exonic sequences are displayed in large characters". "Export" and "Quit" buttons are at the bottom right.

surrounding introns. The corresponding sequences can thus be exported as text file.

Intronic sequences are shown in small letters while exonic sequences are displayed in capital letters.

## 25) Display splice sites

This function gives access to the natural acceptor and donor splice sites strengths. It could be used to compare the relative strength of cryptic and natural splice sites and facilitate the

	Exon #	Acceptor splice site	CV	Donor splice site	CV
1	1			GTTgtaagt	82,62
2	2	tttgcatttttagAT	82,2	AAgtaaga	96,71
3	3	ttttaattcagTT	89,22	CTGgtatgt	86,45
4	4	ttttgtctcagCC	89,83	AAgtaagt	88,18
5	5	tttctttaacagGT	93,02	CAGtaaga	97,66
6	6	ttccacatgtagCT	83,78	TAGtaaga	95,68
7	7	tatgttttttagGC	79,92	AAggttgg	85,68
8	8	atgtgcttacagAT	86,69	CAGtaaag	85,9
9	9	ccttctctcagAT	92,73	CAGgtctgt	86,17
10	10	tttattgtgcagCA	86,71	CAGtaaac	83,9
11	11	cttctttgtcagGG	90,99	CAGtaagt	89,42
12	12	tcaaatttcagTT	83,11	AAgtaggt	90,71
13	13	tttatcttcagGT	94,86	AAggtcaga	91,69
14	14	tctctctccagGT	95,94	CAGgtgtgt	90,04
15	15	aatgtcttcagTC	82,85	GCCgtatgt	73,67
16	16	ttgtttaacagGT	91,13	CAGgttagt	94,98
17	17	tacttctcacagAT	89,54	AAggtgaga	85,55
18	18	ttgctgtcttagGT	82,62	AAgtaggt	79,84
19	19	ctcatgctgcagCC	90,12	ATGtaatt	81,98
20	20	ttttcttttagAG	87,52	AAgtaaga	96,71
21	21	tctatggcacagGA	86	CAGtaagt	89,42
22	22	cttttgataagTT	77,86	CAGgtctgt	86,17
23	23	attgttttttagGC	82,68	CAGtaatt	87,68
24	24	gtctcgtttcagAA	88,3	ACAgtaaga	81,28
25	25	ttatttcttagCT	81,73	CAGgtatag	77,1
26	26	gttttgtggaagGT	79,32	AAgtaaaa	84,55

interpretation of mutations.

A color code is used to display the strength of splice sites: a) dark green for very strong splice sites; b) green for strong splice sites and c) light green for weaker splice sites. The corresponding sequences and strengths can also be exported in a text file.

## 26) Modify polymorphic markers

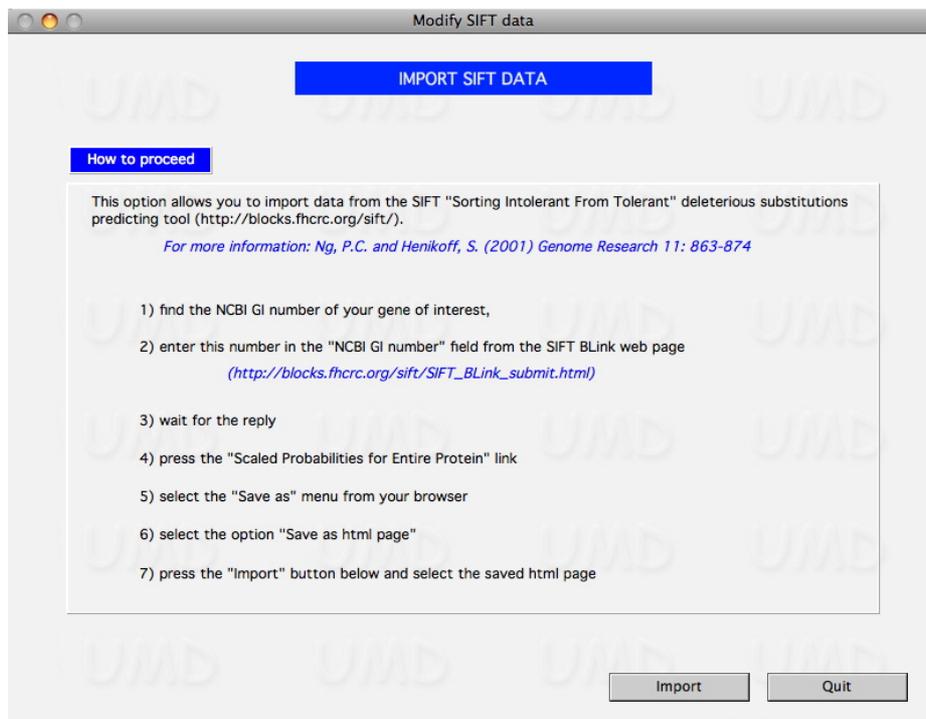
This function is used to modify polymorphic markers or to perform searches from this table.

## 27) Modify epitopes

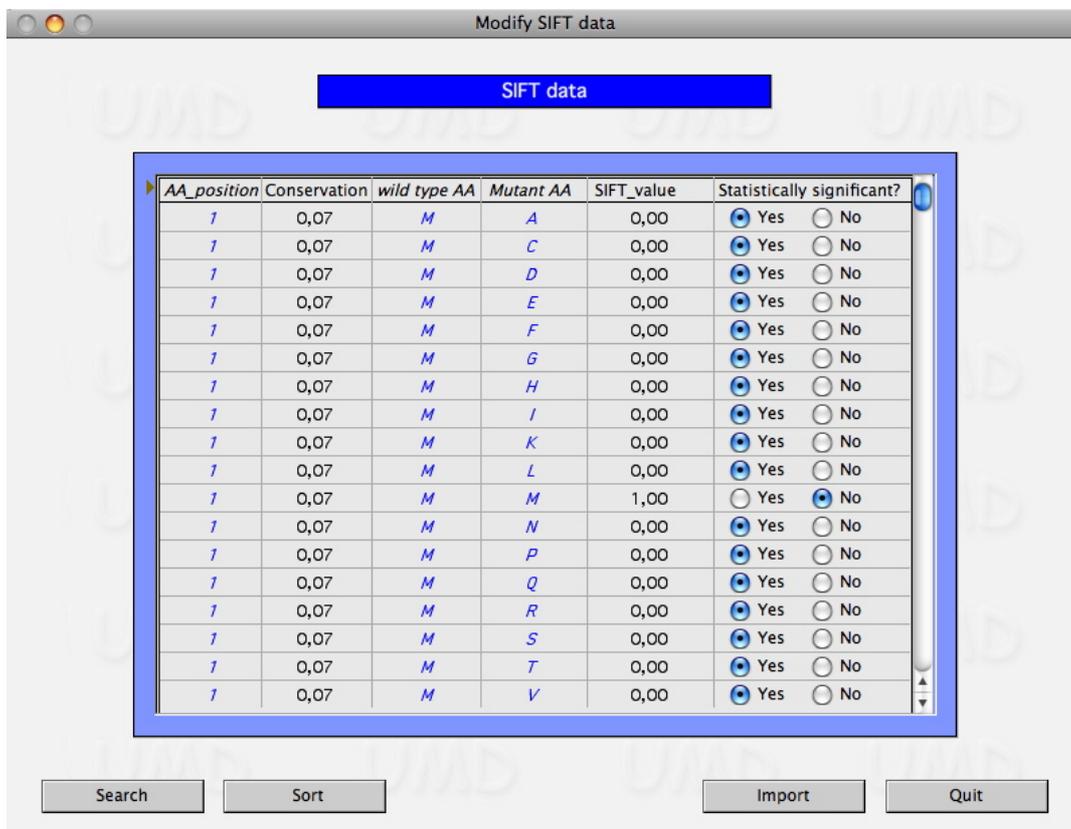
This function is used to modify epitopes or to perform searches from this table.

## 28) Modify SIFT data

This function gives access to a specific interface to download SIFT predictions. The various steps are available below:



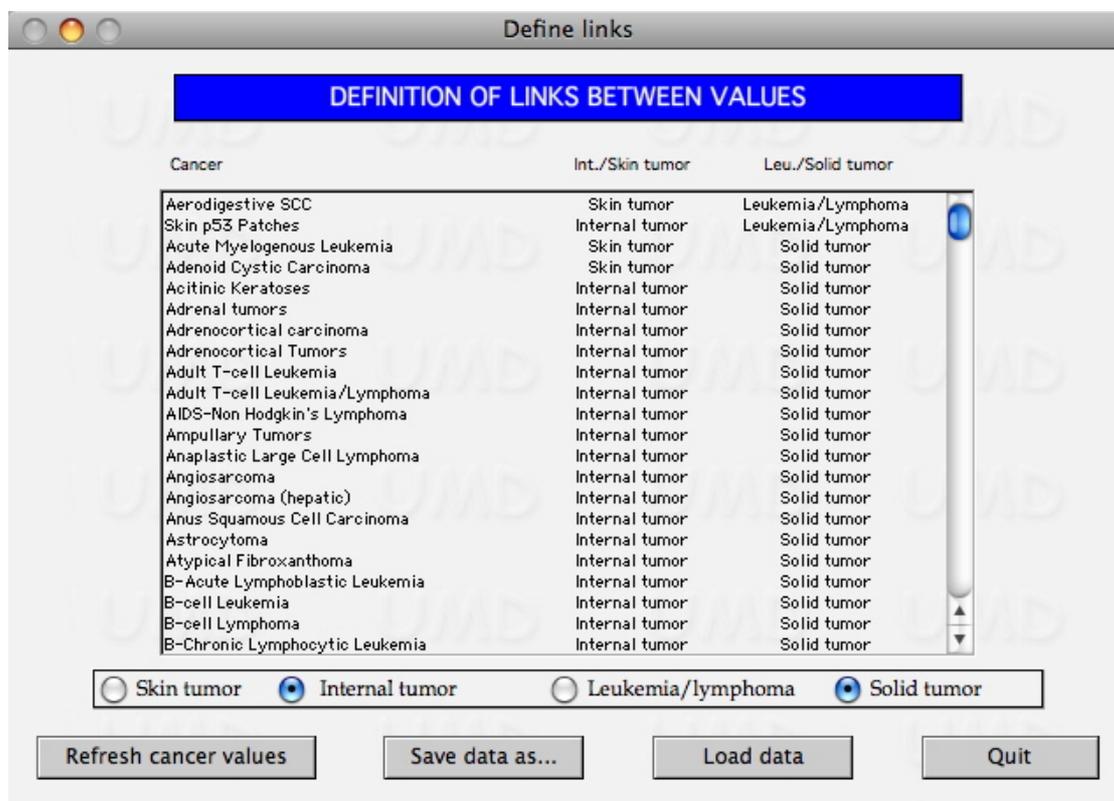
Note that the SIFT predictions are highly dependent of the alignment and a special care should be made to the selection of the various organisms used. The default option is to use the NCBI GI number but this usually gives less efficient predictions. When the proper SIFT file has been saved, a click on the import button allow the automatic



creation of the corresponding records in the UMD database. These data can be modified either directly or through another import. They are used by the UMD-Predictor tool to predict the pathogenicity of missense mutations and substitutions.

## 29) Define links

This function is only useful for genes involved in cancers. It allows to define if cancer entities



are internal or skin tumors and leukemia/lymphoma or solid tumors.

## 30) Create a new database

This function is used to create a new UMD database (cf. II).

## 31) Gene type

Allows to define if the gene is involved in cancers or not. If the gene is involved in cancers, this will activate the "Define links" function as well as a specific display during the mutation creation.

## 32) Update clinical data

This function allows to refresh the list of clinical symptoms and associated severities. It is necessary to activate this function prior to the use of the "Export data" function.

## 33) Clinical data type

This function is used to defined the various severities associated to each symptom (text, numeric values, date) (cf.V).

## 34) Activate clinical search

This function is used to turn on or off the web search module that includes clinical data. It is only useful for on-line UMD-LSDBs.

### **35) Numeric value ranges**

This function is used to defined numerical ranges associated to severities (cf.V).

### **36) Search preference**

This function is used to select the search interface (cf.VII).

### **37) Splice sites calculation**

The user can switch between 2 methods to calculate the acceptor and donor splice sites strength: the UMD algorithm or the Cartegni algorithm. This option can be modified anytime in order to compare predictions.

### **38) Pathway for dnamlk application**

Allows the user to define the directory containing the dnamlk application ued for phylogeny.

### **39) Activate webserver**

This function is used by the Montpellier's team to activate UMD databases through the Internet. Note that we offer free hosting of your UMD database. For more information, please contact [Christophe Bérout](#).

### **40) Authorize Web Services**

This function is used to activate or inactivate Web Services.

### **41) Activate specific options**

This function is restricted to curators for whom specific options have been activated. It requires specific passwords.

### **42) Specific Web Parameters**

This function is restricted to the UMD website administrator.

### **43) Restricted access**

This function is restricted to collaborators for whom specific options have been activated. It requires specific passwords.

### **44) Web license**

This function is used by the Montpellier's team to activate UMD databases through the Internet. Note that we offer free hosting of your UMD database. For more information, please contact [Christophe Bérout](#).

#### **45) 4D\_molecules folder**

This function is restricted to the UMD website administrator.

#### **46) Password**

This function is used to create users and associated passwords. They are used to limit the access to the UMD<sup>®</sup> software but also through the Internet.

#### **47) Import Passwords**

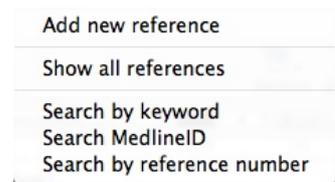
The login and passwords are stored in the UMD<sup>®</sup> software itself and therefore are lost when the UMD<sup>®</sup> software is updated. The "Import Passwords" function allows an easy update of passwords from a text file.

#### **48) Quit**

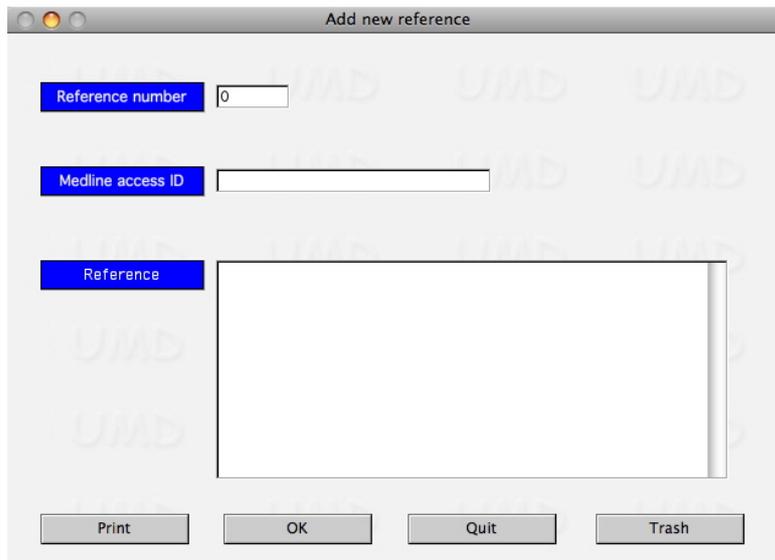
Allows to quit the UMD<sup>®</sup> software.

### VIII-b- The "References" menu

This menu contains 5 functions, which are related to references. In fact, we recommend to link every mutation to a reference in order to facilitate the curation process.



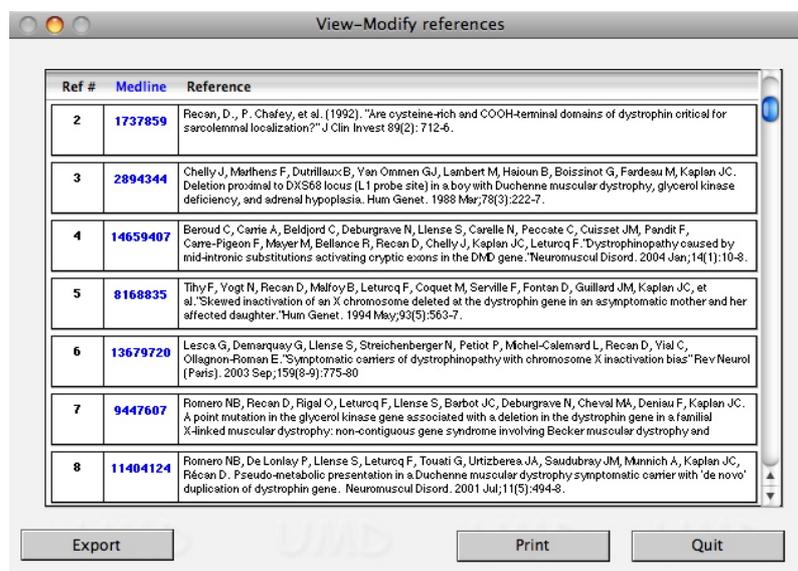
#### 1) Add new reference



This function is used to create a reference. Each reference is characterized by a unique reference number, a Medline access ID (used to establish links with the Medline website) and a reference (free text).

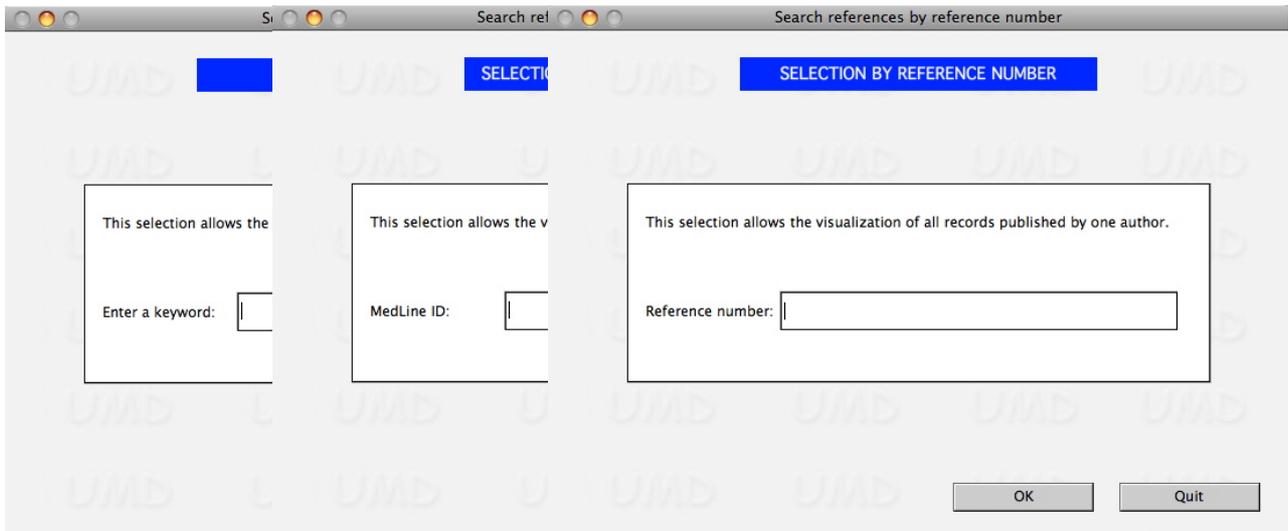
#### 2) Show all references

This function gives access to the list of all references.



#### 3) Search by keyword, Search Medline ID, Search by reference number

These three functions allow to search the reference table using the corresponding criteria (keyword, Medline ID or reference number).



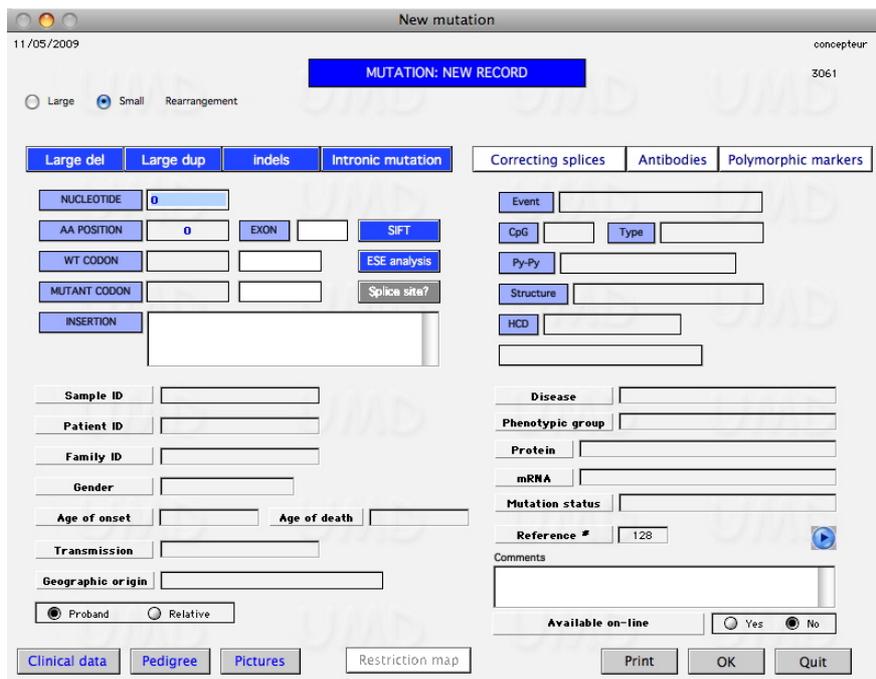
### VIII-c- The "Mutations" menu

This menu contains 5 functions, which are related to mutations.



#### 1) Add a new record

This is the central function of the UMD<sup>®</sup> software. It allows the creation of a new mutation. A generic format is used for new databases. Specific input formats could also be used (contact us to create a specific format).



The input format contains fields related to the mutation itself (top part, cf. III), to the patient/sample (bottom part). The Blue arrow gives access to the cancer related input format:

New mutation

11/05/2009 concepteur  
3061

**MUTATION: NEW RECORD**

Large  Small  Rearrangement

**Large deletion** **Large duplication**  Correcting splices  Monoclonal Antibodies  Polymorphic markers

NUCLEOTIDE   EXON

AA POSITION

WT CODON

MUTANT CODON

INSERTION

Event

CpG  Type

Py-Py

Structure

HCD

Cancer type

LOH

Histology

Smoking

Stage

Grade

Reference #

Comments

## 2) AA type search

Each amino acid is associated with a "Species" field. This data could be used to search for mutations involving amino acids conserved among vertebrates, mammals or specifically

AA type search

**AA TYPE**

AA Type number ?  (Species: 1=Human - 2=Mammalian - 3=Vertebrate)

Record	Nomenclature	Position	codon		AA		Event/ Consequence	Sample ID	Geographic origin	CpG	Type	Ref
			WT	Mut	WT	Mut						
1979	c.1150_4071del	384	GGG	del29	11	Gly	InF	F3407004561	FRANCE		InF	106
1980	c.1150_4071del	384	GGG	del29	11	Gly	InF	F3407004731	FRANCE		InF	106
2012	c.94_6438del	32	TTT	del63	3	Phe	InF	F5420274261181	FRANCE		InF	117
1982	c.5615_7542del	2205	AGG	del92	45-4	Arg	Fs	F671340100461	FRANCE		Fr	121
1983	c.7201_7309del	2401	AGG	del10	50	Arg	Fs	F671163400561	FRANCE		Fr	121
1984	c.6439_7098del	2147	GAA	del66	45	Glu	InF	F671119900941	FRANCE		InF	121
1985	c.6439_6912del	2147	GAA	del47	45	Glu	InF	F67939000321	FRANCE		InF	121
1986	c.6913_7098del	2305	GTT	del18	48	Yal	InF	F67759100141	FRANCE		InF	121
1987	c.1_31del	1	ATG	del31	1	Met	Fs	F671014700211	FRANCE		Fr	121
1994	c.961_4233del	321	CAT	del32	10	His	InF	F54003911981	FRANCE		InF	117
2346	c.2293_7543dup	765	GCC	ins52	19	Ala	InF	F3441220691	FRANCE		Fr	106
1990	c.6439_7200del	2147	GAA	del76	45	Glu	InF	F67163300421	FRANCE		InF	121
1991	c.6439_7200del	2147	GAA	del76	45	Glu	InF	F67163300431	FRANCE		InF	121
1992	c.6913_7309del	2305	GTT	del39	48	Yal	Fs	F67344800451	FRANCE		Fr	121

Sort 2411 file(s) found. Print Quit

human. This function uses the generic display format.

## 3) Reference search

This function allows a rapid display of mutations associated to a specific reference. It is frequently used for curation in order to avoid duplicates from a specific team.

**Reference search**

**SEARCH: CLASSIFICATION BY REFERENCE**

Reference number ?

Medline ID: 9447607

Reference: Ponomero NB, Recon D, Rigal O, Letarocq F, Liense S, Barbot JC, Debrugrave N, Chevrel M, Denis F, Kaplan JC. A point mutation in the glycerol kinase gene associated with a deletion in the dystrophin gene in a familial X-linked muscular dystrophy: non-contiguous gene syndrome involving Becker muscular dystrophy and glycerol kinase loci. *Neuromuscul Disord.* 1997 Dec;7(8):499-504.

Record	Nomenclature	Position	WT	Mut	exon	Yal	InF	Event/Consequence	Sample ID	Geographic origin	CpG	Type	Ref
450	c.1483_4071del	495	GTG	del25	13	Yal	InF	In frame del	F751863011			InF	7
2904	c.1483_4071del	495	GTG	del25	13	Yal	InF	In frame del	F751863311			InF	7
2905	c.1483_4071del	495	GTG	del25	13	Yal	InF	In frame del	F751863111			InF	7
2911	c.1483_4071del	495	GTG	del25	13	Yal	InF	In frame del	F751863411			InF	7
2912	c.1483_4071del	495	GTG	del25	13	Yal	InF	In frame del	F751863211			InF	7

Sort      5 record(s) found.      Print      Quit

#### 4) Deletion analysis

Small deletions frequently result from slippage of the polymerase during the replication (for example deletion of an A in a stretch of As) or to a more complex process that involves repeated sequence and lead to the deletion of the sequence localized between the repeats with or without the repeats themselves.

During a mutation's creation, the UMD<sup>®</sup> automatically searches for repeated sequences that are localized at the extremities of the deletion and for stretches of nucleotides that could explain the molecular mechanism that led to the mutation. These data are accessible via the "Deletion analysis":

**Deletion analysis**

**DELETIONS ANALYSIS: REPEATED SEQUENCES**

Deletion  
 CAGATTTTCAGAGCCTCTCACCACCACTGAGGCGATGCGTAAACAGAGCAACTGTATAAGGAA

Repeated sequences  
 CAGATTTTCAGAGCCTCTCACCACCACTGAGGCGAGCAGCCTATGCGTAAACAGAGCAACTGTATAAGGAA

POSITION	MUTANT CODON	NB OF RECORDS	REPEATED SEQUENCE
66	del1c	1	AAAAA
127	del2a	1	AT
161	del3b	1	CCA
226	del3a	1	AAG
412	del1b	1	TT
542	del1b	3	AAA
571	del1c	3	TTTTT
576	del1a	1	TTT
588	del1a	1	AA
604	del1c	1	CC
660	del2a	1	AA
673	del11a	1	CACT
744	del2a	1	AG
761	del5a	1	AAAAG
761	del4a	1	AAAAG
902	del1c	1	AA
1004	del2b	1	GA
1011	del1c	1	CCC
1053	del1a	1	AAAAA

Deletions with repeated sequences: **70** } Ratio: **59,32 %**  
 Deletions with known sequences: **118**  
 Total number of deletions: **1756**

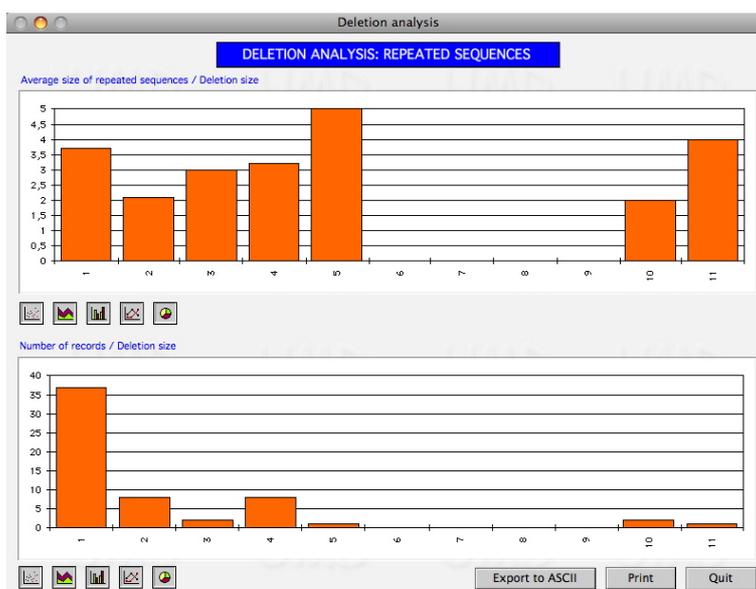
Chart    Print    Quit

In the upper part of the screen are displayed the reference and the mutant sequences. The deleted sequence is shown in white on a blue background while repeated sequences are

displayed in yellow and red on a black background. A simple click on the table will display the corresponding information on the upper part. Simple statistics are also available (bottom). A "Chart" button gives access to two graphics:

- The first one displays the average size of repeated sequences compared to the deletion size;
- The second one displays the number of records found for each deletion size.

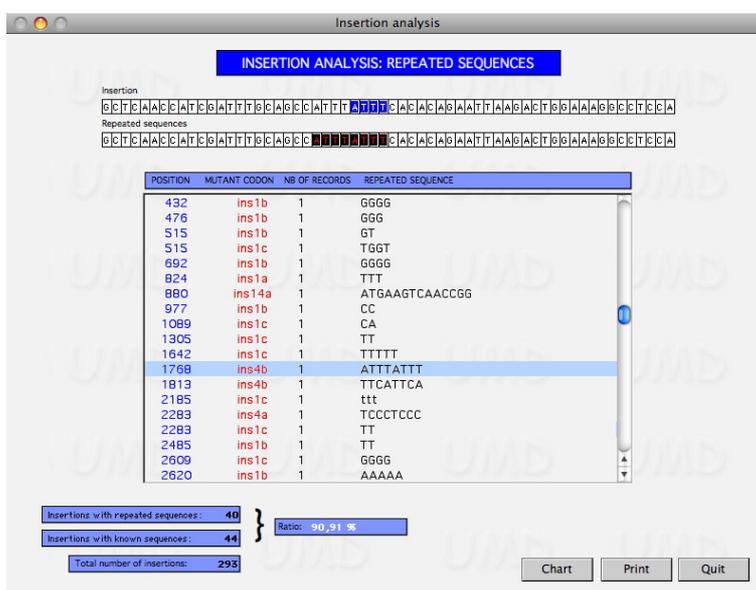
Both graphs can be modified using dedicated buttons, and corresponding data could be exported as an text tab-delimited file in order to be used in more sophisticated graphical



display tools.

### 5) Insertion analysis

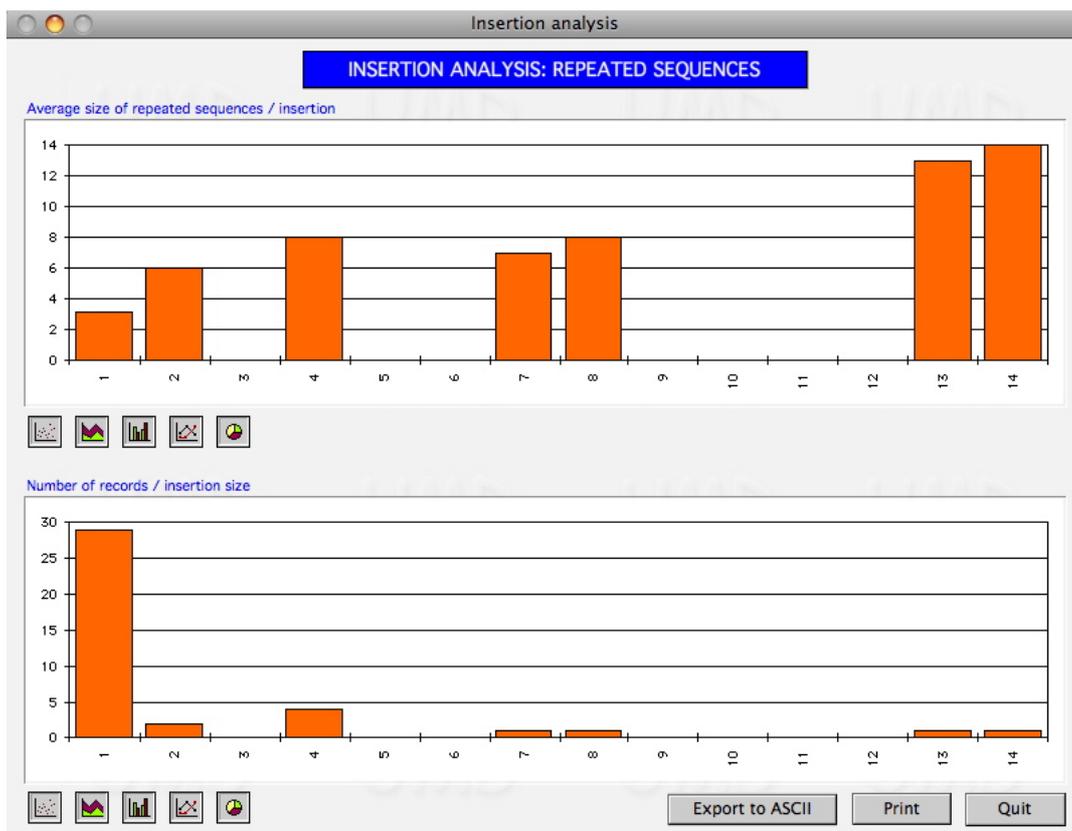
As for small deletions, small insertions frequently result from the creation of repeated sequences. Therefore during a mutation's creation, the UMD<sup>®</sup> automatically searches for repeated sequences generated by the mutation:



In the upper part of the screen are displayed the reference and the mutant sequences. The inserted sequence is shown in white on a blue background while repeated sequences are displayed in red on a black background. A simple click on the table will display the corresponding information on the upper part. Simple statistics are also available (bottom). A "Chart" button gives access to two graphics:

- The first one displays the average size of repeated sequences compared to the deletion size;
- The second one displays the number of records found for each deletion size.

Both graphs can be modified using dedicated buttons, and corresponding data could be exported as a text tab-delimited file in order to be used in more sophisticated graphical display tools.



### VIII-d- The "Pedigree" menu

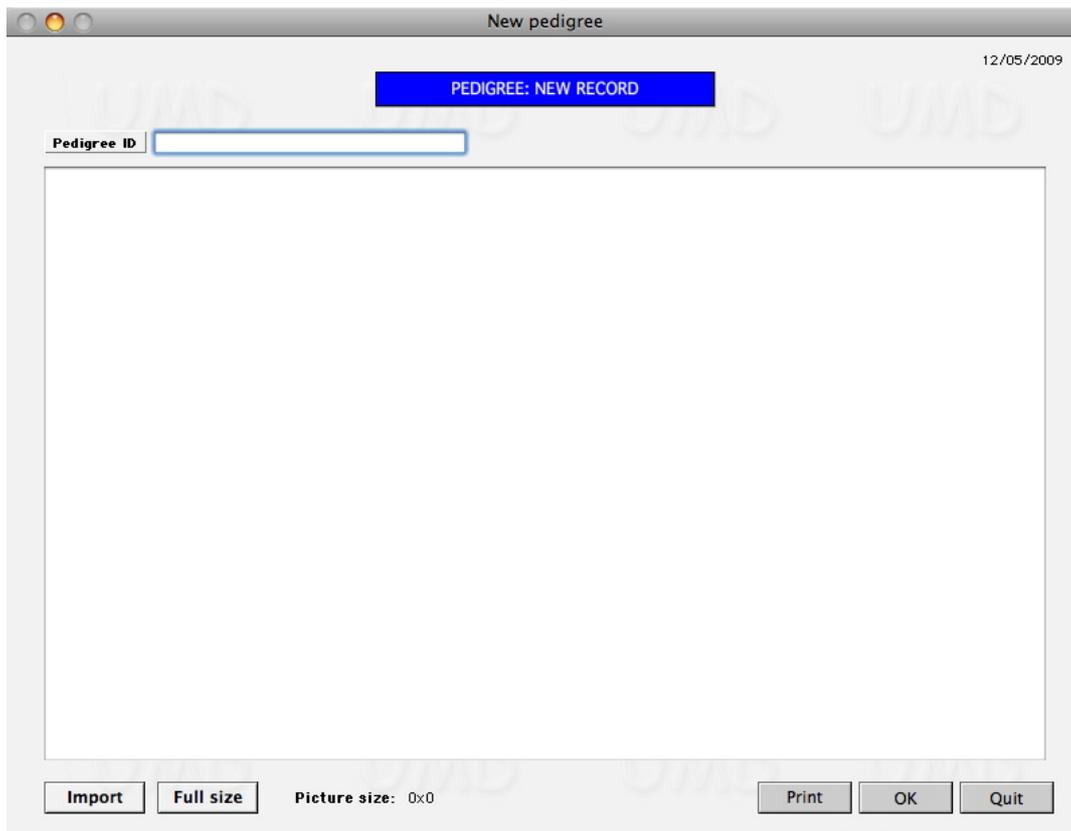
This menu contains 2 functions, which are related to the creation and modification of pedigrees.

Add a new pedigree

Modify a pedigree

#### 1) Add a new pedigree

This function allows the import of a pedigree file. Each pedigree should be labelled with a unique identifier (pedigree ID). If the pedigree picture is larger than the display format, a click on the "Full size" button will reformat the display screen in order to view the picture in



the most appropriate format.

#### 2) Modify a pedigree

This function is used to list all pedigrees and modify a specific record (update or delete).

### VIII-e- The "Show all mut." menu

This menu contains 4 functions, which are related to the display of mutations and queries.

View-Modify mutations	⌘L
Define a display format	
View-modify display formats	
Free search	⌘F
Create a report	

#### 1) View-Modify mutations

This function gives access to the list of all mutations recorded in the database (cf. VI-b).

#### 2) Define a display format

Multiple personalized display formats can be created in the UMD® software (cf. VI-a).

#### 3) View-Modify display formats

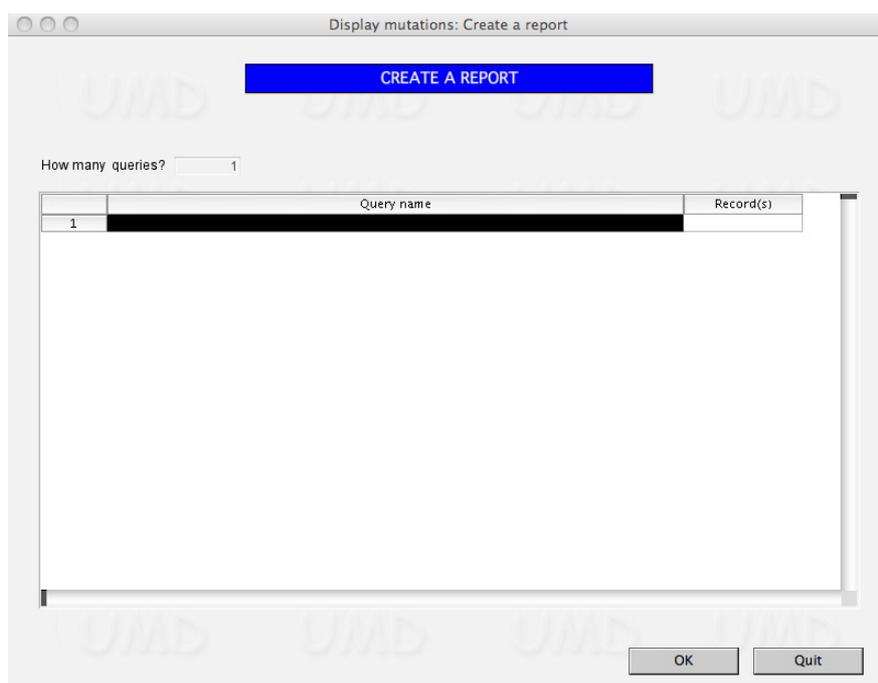
This function gives access to the modification of personalized display formats (cf. VI-b).

#### 4) Free search

With this function, the user can select mutations records using one or many criteria. He can use two search interfaces to perform the queries (cf. VII).

#### 5) Create a report

This function is used to compare the number of records corresponding to various selection criteria. The user can define the number of lanes and enter a title for each selection criteria. A click on the lane number gives access to the search interface and returns the number of records.

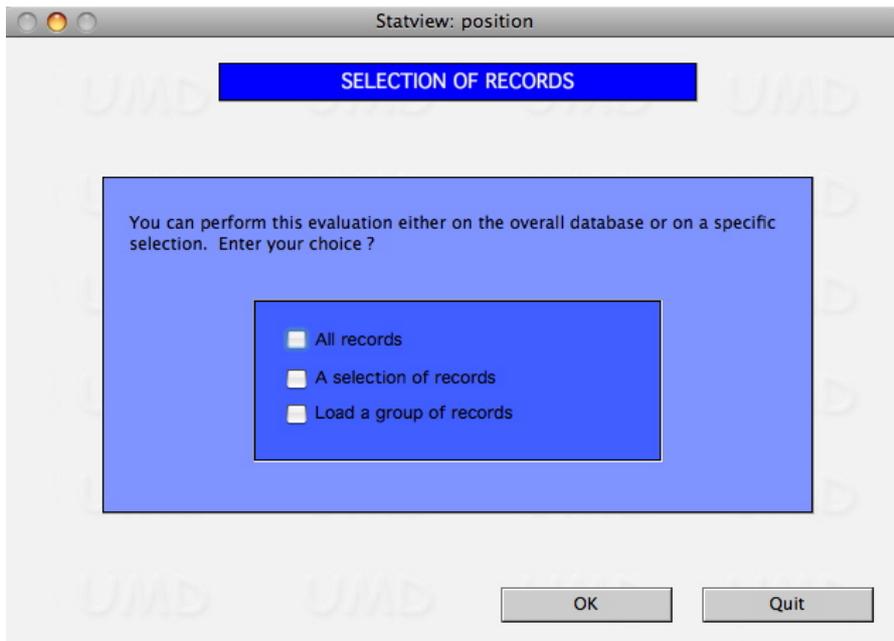


### VIII-f- The "Statview" menu

This menu contains 27 functions, which are related to the data analysis either using statistical or graphical presentations.

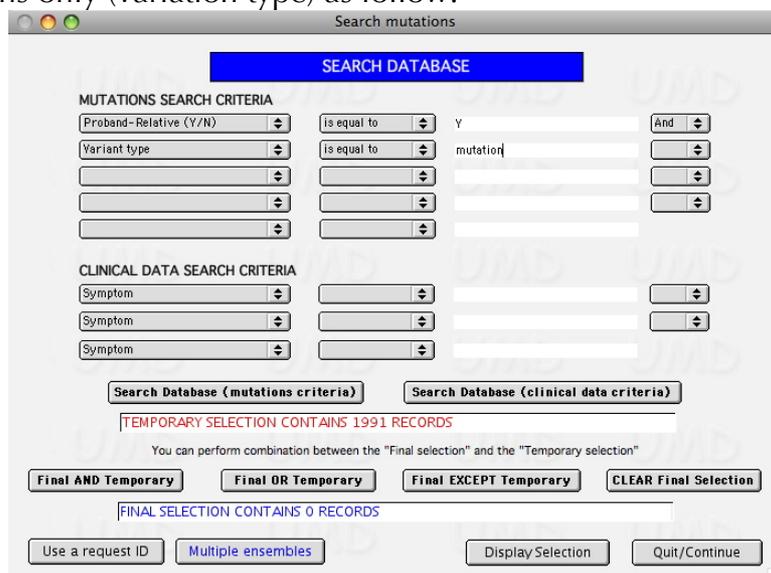
Because it is very useful to be able to perform analysis on a subset of records, all functions from this menu are linked to a "Selection of record" system.

- Position
  - Potential Stop codons
  - Nonsenses and read-through
  - Mutational events
  - Detailed mutational events
- Frequency of mutations
- Frequency of events
- Distribution of mutations
  - Del & dup distribution
  - Mutation map
  - Mutation map [2 groups]
  - Deletion map
  - Stop codons map
- Geographic distribution
- Binary comparison
- Stat exons
  - Distribution by exons
  - ESE map of exons
- Sequence analysis
- Splice mutations
- Missense mutations
  - All missense variations
  - Missenses & Hydrophobicity
- Structure
- Global analysis
- Free Graph
- Specific tools



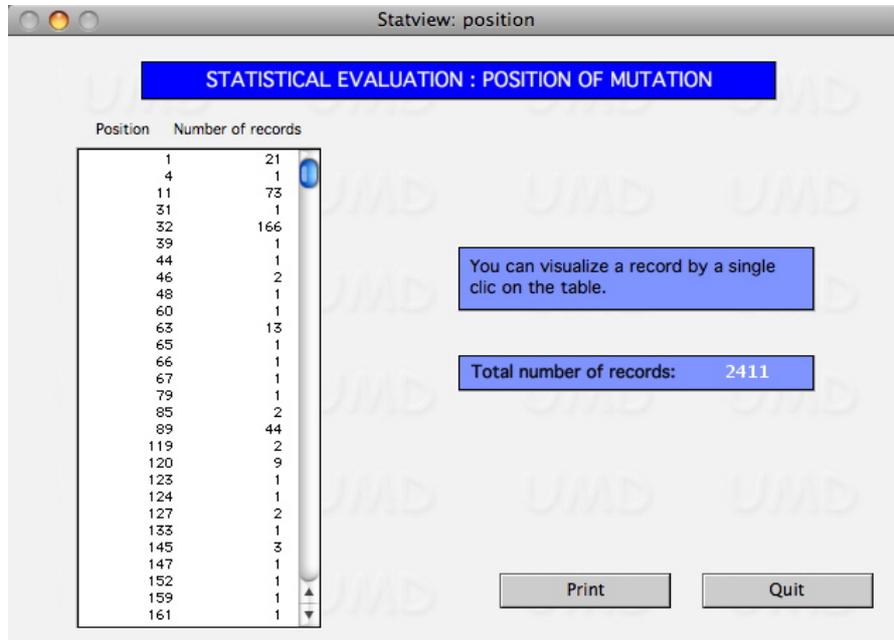
The user can work from "All records", freely select a group of records using the search interface when selecting the "A selection of records" option or use a previously saved ensemble of records using the "Load a group of records".

Note that working with all records should be avoided if you want to study molecular events. In this situation you need to select probands using the search interface. You can also wish to filtrate for mutations only (variation type) as follow:



## 1) Position

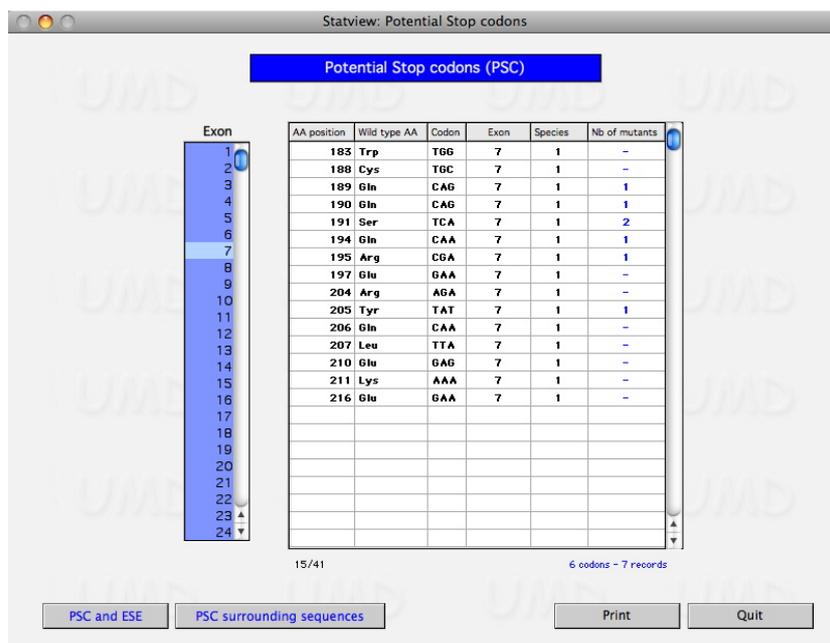
This function gives access to the number of records (mutations) localized at a specific amino acid residue. If large rearrangements are involved, they are associated to the first residue targeted by the mutation.



A simple click on one element of the list will open a window with corresponding records.

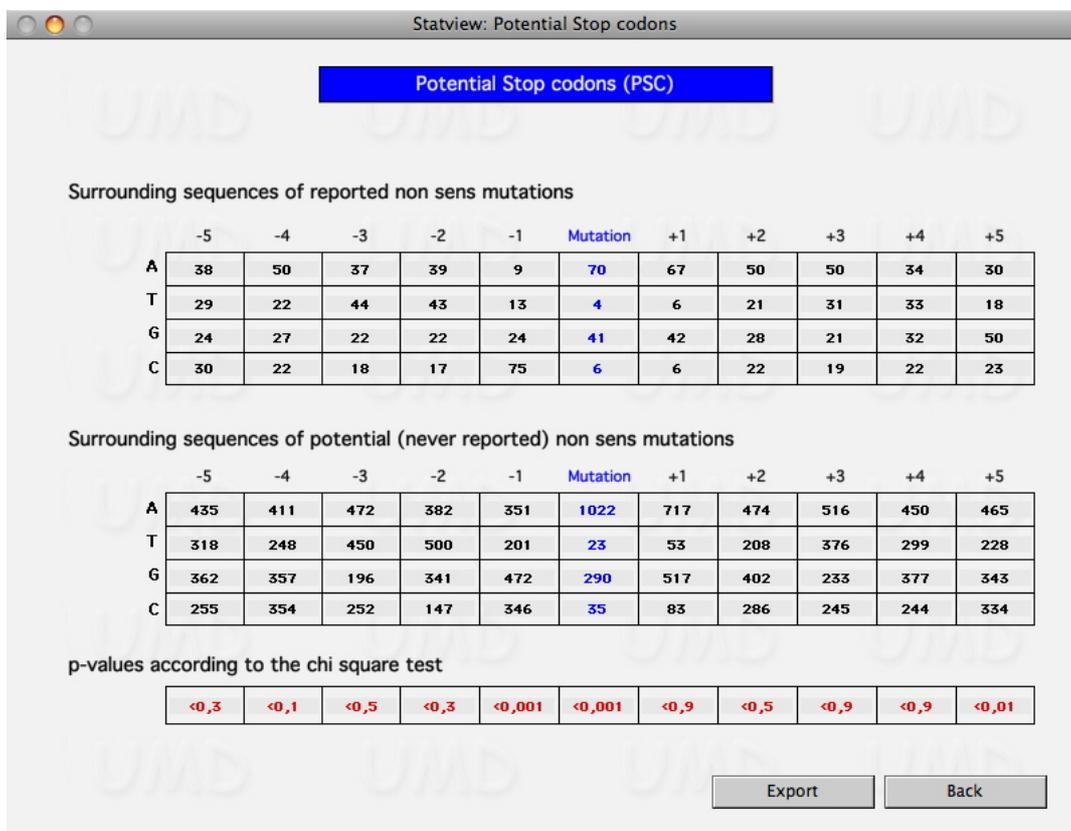
## 2) Potential stop codons

This function gives for each exon, the list of codons that can be mutated in premature stop codons (PTC) by a single substitution and the number of such mutations.



In this example, 15 out of 471 codons of the exon 7 of the DMD gene could lead to PTC. Among these 7 PTC, mutations have been reported for 6 of them. Two buttons lead to complementary data:

The "PSC surrounding sequences" option gives statistical data about the environment of observed PTC compared to potential stop codons for which no mutation has never been reported. This option is used to study the environment of PTC and find out if this context could play a role in PTC efficiency via codon recognition and read-through or other mechanisms.



### 3) Nonsenses and read-through

This function is used to evaluate the potential to correct a premature termination codon mutation by a read-through therapeutic strategy. In fact, read-through can be induced by various chemical compounds (aminoglycosides, PTC124) and leads to the incorporation of an amino acid at the premature stop codon position. This will therefore lead to the production of a quasi normal protein with an amino acid substitution at the position of the premature stop codon.

This function therefore lists all reported nonsense mutations with for each position, the number of pathogenous missense mutations and/or polymorphisms (non-pathogenic mutations) affecting the residue.

Only nonsense mutations affecting a residue without associated pathogenic missense mutations are good candidates for nonsense read-through strategies. Among these nonsense mutations, those associated to polymorphisms are even better candidates for these therapeutic strategies.

**Nonsense mutations and read-through**

	Nonsense mutation	Number of records	Missense mutation	Polymorphism
1	p.Ser4X	3		
2	p.Glu60X	1	p.Glu60Lys (x1)	
3	p.Arg75X	2		p.Arg75Gln (x2)
4	p.Tyr122X	5		
5	p.Gln220X	4		
6	p.Ser489X	1		
7	p.Gln493X	2		
8	p.Gly542X	54		
9	p.Arg553X	8		
10	p.Glu585X	4		
11	p.Glu656X	2		
12	p.Glu664X	1		
13	p.Arg709X	1		
14	p.Lys710X	6		
15	p.Leu732X	3		
16	p.Arg764X	2		
17	p.Ser776X	1		
18	p.Arg792X	5		
19	p.Lys830X	1		
20	p.Glu831X	1		
21	p.Cys832X	1		
22	p.Trp846X	2		
23	p.Arg851X	1		
24	p.Gln890X	1		
25	p.Gln1042X	1		
26	p.Trp1063X	1		

Export      Quit

In this example (CFTR gene), the p.Glu60X targets the Glu 60 residue that is also targeted by a missense mutation and therefore not a good candidate. Conversely, the p.Arg75Gln polymorphism targets the Arg 75 residue involved in the p.Arg75X and is therefore a better candidate for nonsense read-through.

#### 4) Mutational events

This function list mutational events reported at each codon. A summary table is also available. A click on one row of the tables allows the display of corresponding records using

**Statview: mutational events**

**STATISTICAL EVALUATION: MUTATIONAL EVENTS**

Position	Number of record	Type	Number of record
1	12	A->G	5
		G->A	1
		G->C	5
		T->A	1
4	5	C->A	3
5	1	C->T	1
17	1	Stop	1
18	10	A->T	1
		In frame del	6
		Stop	3
32	1	T->C	1
46	1	C->A	1
50	1	T->C	1
55	1	C->T	1
56	1	Stop	1
57	1	T->G	1
58	1	A->G	1
59	1	Stop	1
60	2	G->A	1
		G->T	1
67	1	C->T	1
74	8	C->T	8
75	4	C->T	2
		G->A	2
76	1	T->G	1
78	2	Stop	2
85	7	G->A	7
88	2	Stop	2
91	1	T->G	1
92	1	G->C	1
98	1	A->G	1
1005	1	T->G	1
1022	1	In frame del	1

You can visualize a record by a single clic on the table.

A->C	4
A->G	217
A->T	13
allele 3T	1
allele 5T	86
allele 6T	2
allele 7T	157
allele 9T	127
allele TG10	155
allele TG11	99
allele TG12	72
allele TG13	13
allele TG9	8
A[12]	1
C->A	30

Total nb of records: 2436

Print      Quit

the personalized default display format.

## 5) Detailed mutational events

This function displays all informations about mutational events. Positions are sorted by frequency.

Statview: detailed mutational events

**STATISTICAL EVALUATION : DETAILED MUTATIONAL EVENTS**

Number of records: 1450      Number of mutations: 223

Position	WT codon	Number of records	Mutant codon	Number of records
507	ATC	507	de13e	In frame del
470	ATG	144	de13a	In frame del
854	ACT	110	GTG	A->G
542	GGA	54	ACG	T->G
1463	CAG	39	TGA	G->T
1303	AAC	36	CAA	G->A
117	CGC	24	AAG	C->G
1152	GAT	20	CAC	A->C
997	TTG	20	CAT	G->C
206	TTG	17	TGC	C->T
668	CGT	15	CAT	G->C
1282	TGG	14	del379a	Stop at 1007
576	GGA	14	del1151a	Stop at 1009
443	GAT	14	TGG	T->G
685	CAA	13	TGT	C->T
1092	TAC	10	TGA	G->T
347	CGC	10	TGA	G->T
1162	CGA	9	GCA	G->C
508	TTT	9	GCA	G->C
334	CGG	9	TAT	G->T
18	AGC	9	indels	indels
553	CGA	8	ins1a	Stop at 688
201	GTG	8	TAA	C->A
74	CGG	8	CAC	G->A
			CCC	G->C
			TGA	C->T
			CTA	G->T
			TGT	T->G
			TGG	C->T
			del111e	In frame del
			del220e	Stop at 33
			TGA	C->T
			ATG	G->A
			del11a	Stop at 214
			TGG	C->T

You can visualize a record by a single clic on the table.

Export    Sort by number of records    Print    Quit

A click on the "Sort by number of records" button opens a second window that sorts mutational events by frequency. A second table also lists the number of positions with similar number of records.

Statview: detailed mutational events

**STATISTICAL EVALUATION : DETAILED MUTATIONAL EVENTS**

Number of records: 1450      Number of mutations: 223

Position	WT codon	Mutant codon	Number of records	Number of records	Number of positions
507	ATC	de13e	In frame del	501	34,55 %
470	ATG	GTG	A->G	144	9,93 %
854	ACT	ACG	T->G	110	7,59 %
542	GGA	TGA	G->T	54	3,72 %
1463	CAG	CAA	G->A	39	2,69 %
1303	AAC	AAG	C->G	35	2,41 %
117	CGC	CAC	G->A	23	1,59 %
1152	GAT	CAT	G->C	20	1,38 %
206	TTG	TGG	T->G	17	1,17 %
997	TTG	TTC	G->C	16	1,10 %
668	CGT	TGT	C->T	15	1,03 %
1282	TGG	TGA	G->A	14	0,97 %
576	GGA	GCA	G->C	14	0,97 %
443	GAT	TAT	G->T	14	0,97 %
1092	TAC	TAA	C->A	10	0,69 %
685	CAA	indels	indels	9	0,62 %
508	TTT	TGT	T->G	9	0,62 %
334	CGG	TGG	C->T	9	0,62 %
1162	CGA	TGA	C->T	8	0,55 %
553	CGA	TGA	C->T	8	0,55 %
74	CGG	TGG	C->T	8	0,55 %
347	CGC	CAC	G->A	7	0,48 %
201	GTG	ATG	G->A	7	0,48 %
1270	GAT	AAT	G->A	7	0,48 %
178	GGA	AGA	G->A	7	0,48 %
85	GGA	GAA	G->A	7	0,48 %
507	ATC	de13a	In frame del	6	0,41 %
18	AGC	del111e	In frame del	6	0,41 %
1066	CGT	TGT	C->T	6	0,41 %
952	ATG	ATC	G->C	6	0,41 %
710	AAA	TAA	A->T	6	0,41 %

You can visualize a record by a single clic on the table.

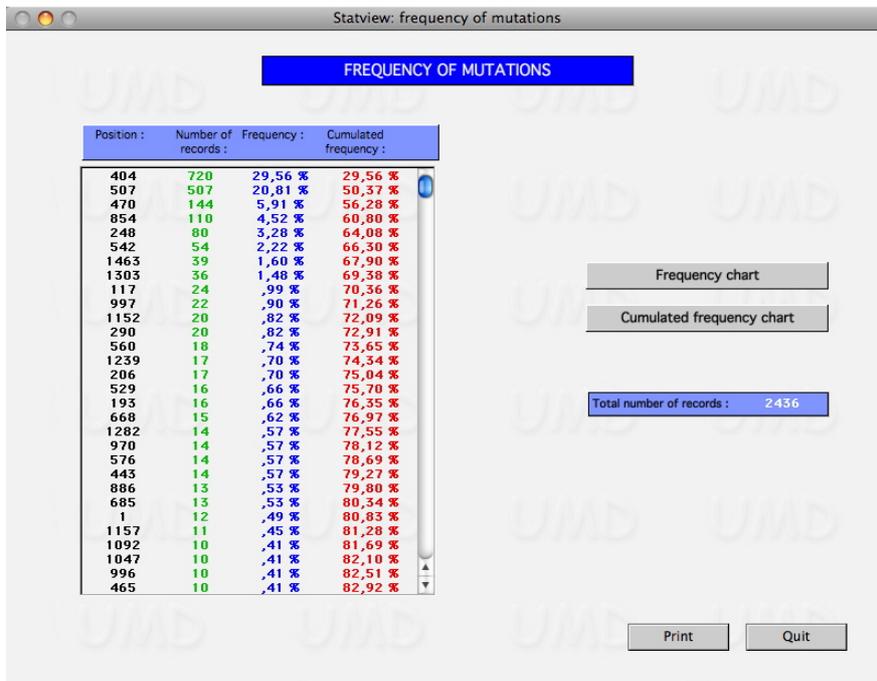
Number of records	Number of positions
501	1
144	1
110	1
54	1
39	1
35	1
23	1
20	1
17	1
16	1
15	1
14	1
14	3
10	1
9	3
8	3
7	5
6	6
5	6
4	10
3	14
2	29
1	132

Distribution of events    Distribution of mutations

Export    Back to previous table    Print    Quit

## 6) Frequency of mutations

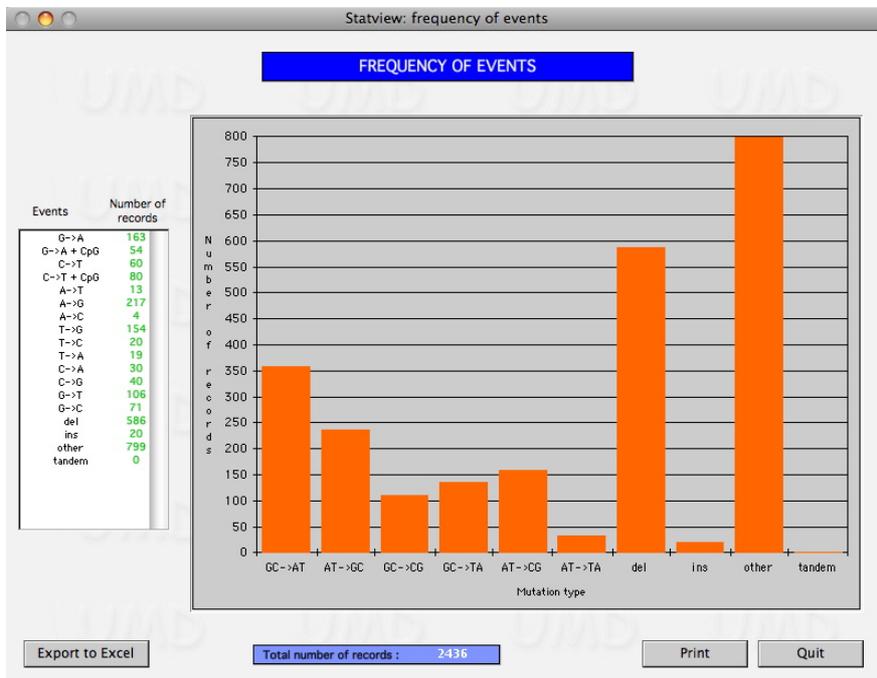
This function sorts the amino acid positions according to the number of records. In addition, a cumulative frequency value allows the evaluation of screening strategies and their relative efficiency. This function is frequently used on subsets of records selected by phenotypes.



Two graphical displays are also available to display positions by frequency of mutations or by cumulated frequency of mutations.

## 7) Frequency of events

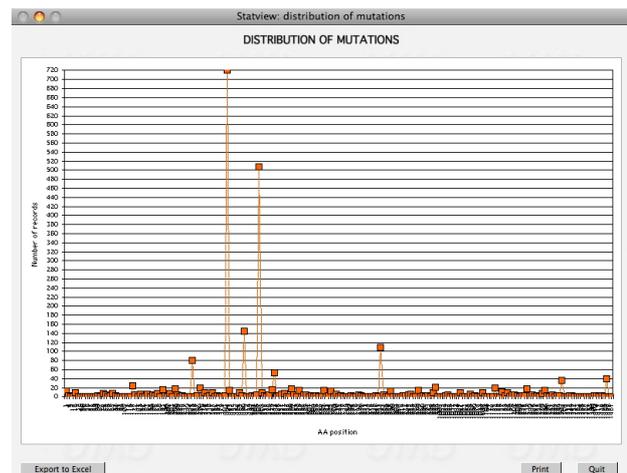
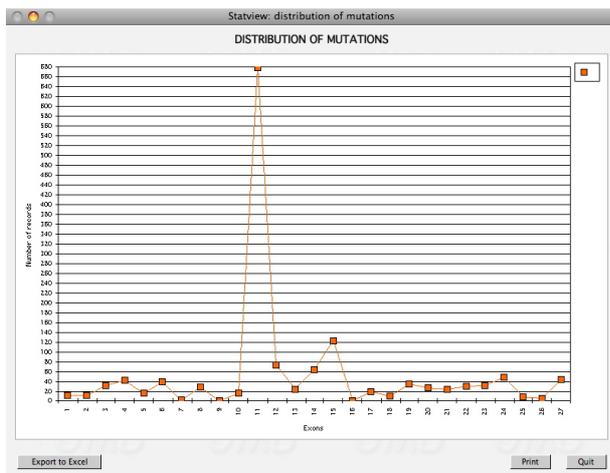
This function displays the distribution of mutational events.



## 8) Distribution of mutations

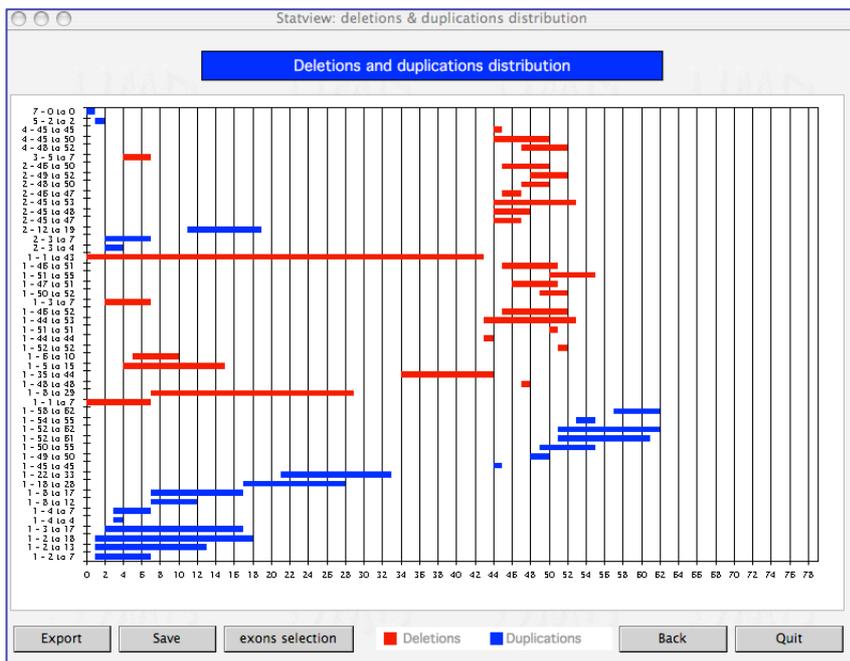
This function is designed to compare the mutations of up to 8 groups of records. Each group is selected using the regular search interface. The results are displayed in 2 formats:

- A graphical display of the distribution of mutation per exon;
- A graphical display of mutations by residue.



## 9) Del & dup distribution

This function is restricted to the study of large deletions and duplications ( $\geq$  one exon). The user can select the exons of interest prior to the display.



**Exons selection**

Limit the graphic display to large rearrangements involving exons

From  >=

To  <=

Large deletions

Large duplications

In the x-axis are displayed the various exons and in the y-axis the various deletions/duplications with the corresponding number of records. The "Exon selection" button allows the modification of display parameters.

## 10) Mutation map

After choosing the subset of records that should be used for this analysis, the user can choose between two display options:

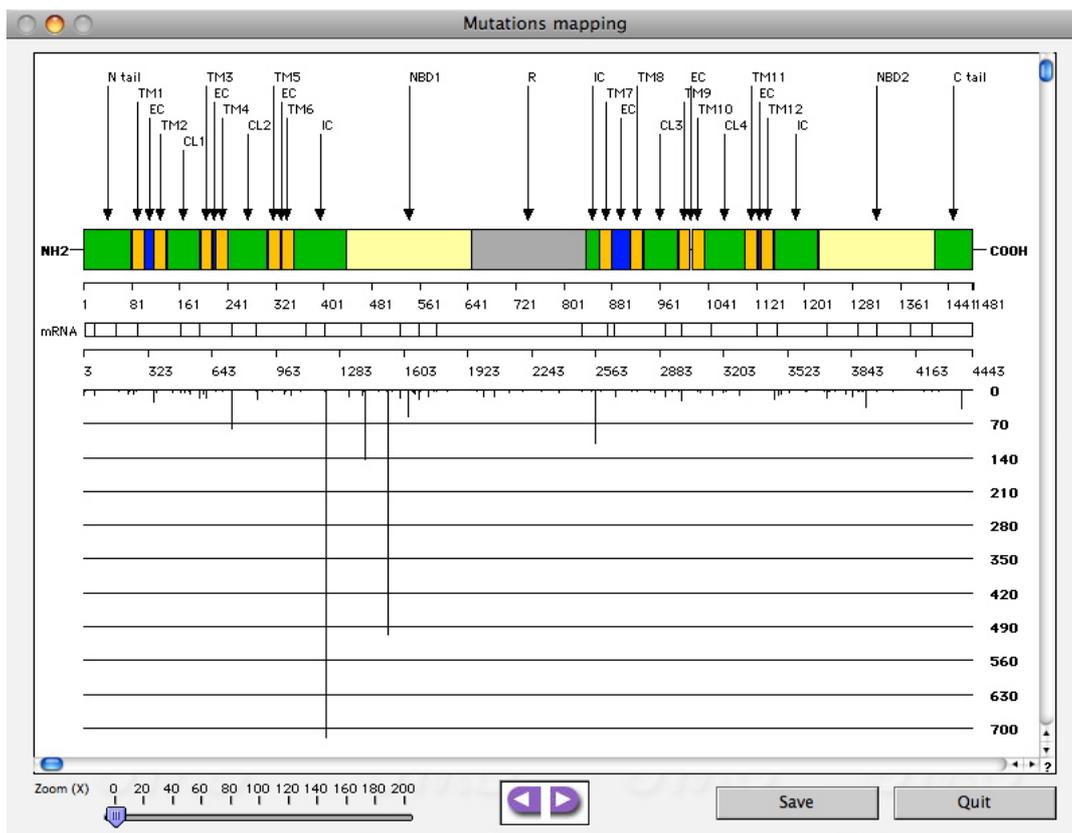
"All" is used to display all mutations and their relative numbers

"Phenotypes" is used to display residues harboring mutations for each phenotype. In order to avoid differences between phenotypes frequency, the number of mutations is not presented in this graph.

Both graphs use the same scales: Protein structural domains and associated amino acid scale (top) and the various eons and the associated nucleotide scale (middle).

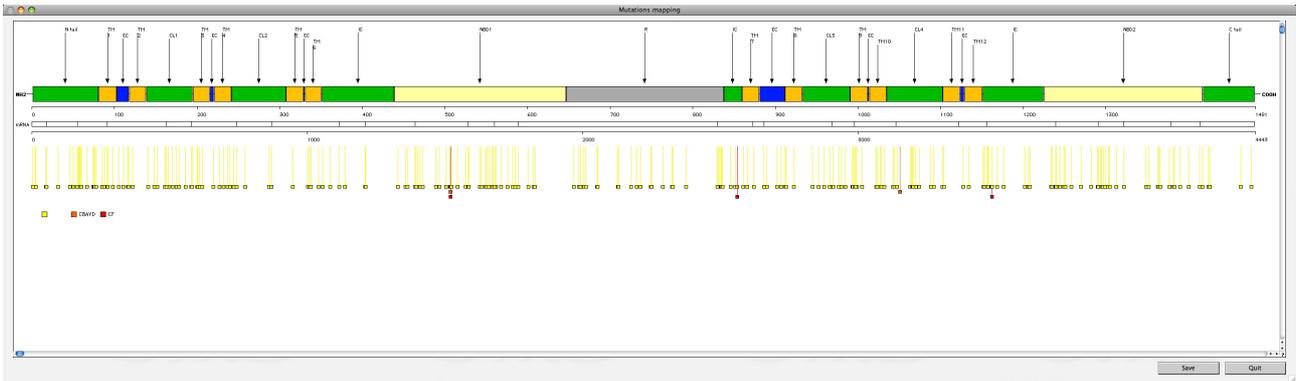
*Note the various structural domains are labelled with short names and used colors defined with the "Structure preference" function from the "File" menu (cf. ).*

Below are presented either the number of mutations at each residue or the presence or absence of a mutation at each residue for each phenotype.



Both pictures can be saved as pictures using the "Save" buttons.

In the "All" display format, zooming options are available (up to x200). The user can then click on the graph to select the region of interest or can navigate within the graph using the two violet arrows.

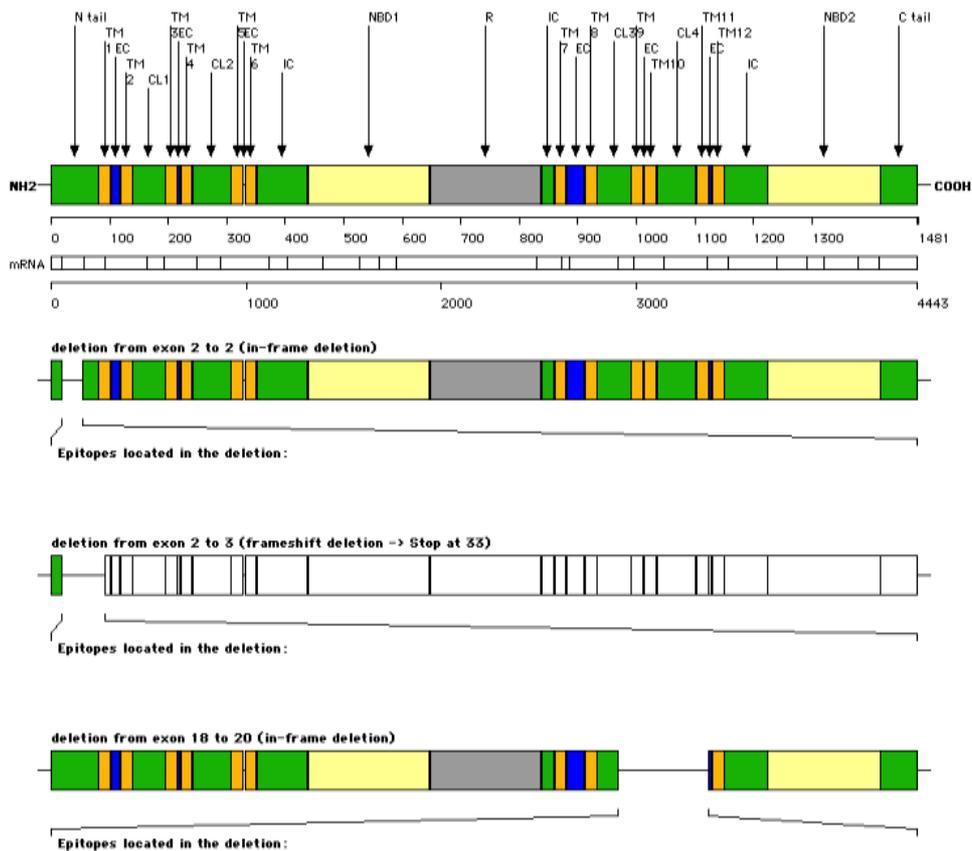


## 11) Mutation map [2 groups]

This function is similar to the “Mutation map” except that it allows the user to compare two subsets of records. Unlike the other function, the user cannot choose the display options but he can add a legend for both subsets of records.

## 12) Deletion map

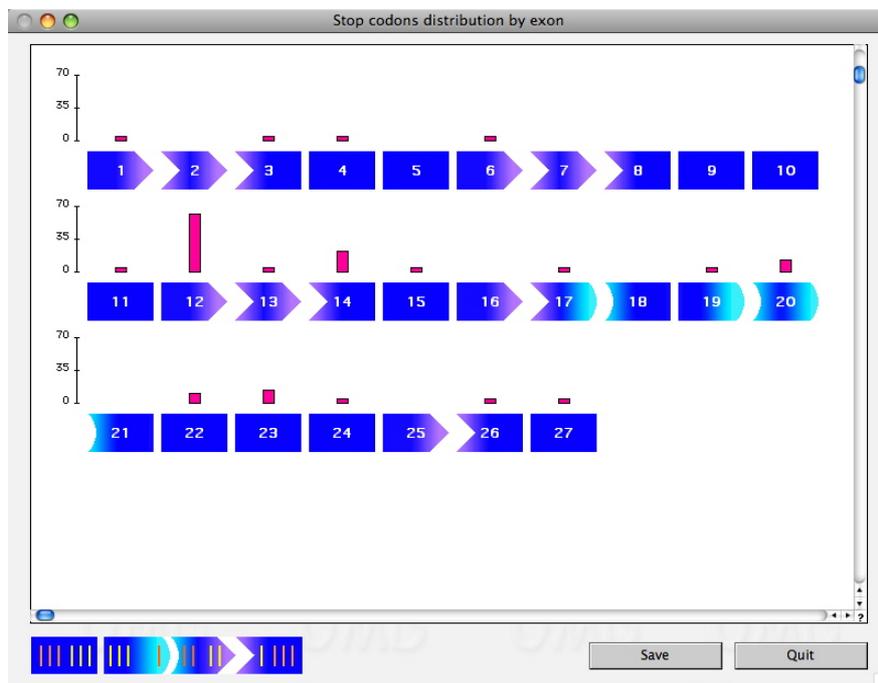
This function displays the consequence of large deletions on the reading frame and the various antibodies epitopes localized within the deletion.



In this example, the deletions of exon 2 and exons 18 to 20 of the CFTR gene are in-frame deletions and therefore lead to internal deletions. The various structural domains affected by these mutations could easily be determined with the color code system. The deletion of exons 2 and 3 is out-of-frame and probably lead to the absence of protein. Nevertheless, the predicted potential truncated protein is displayed in the graph.

### 13) Stop codons map

This function displays the exon phasing and the position and number of reported nonsense mutation (stop codons). this function was primarily designed to choose the most relevant



exons that could rescue nonsense mutations by the exon-skipping therapeutic approach.

### 14) Geographic distribution

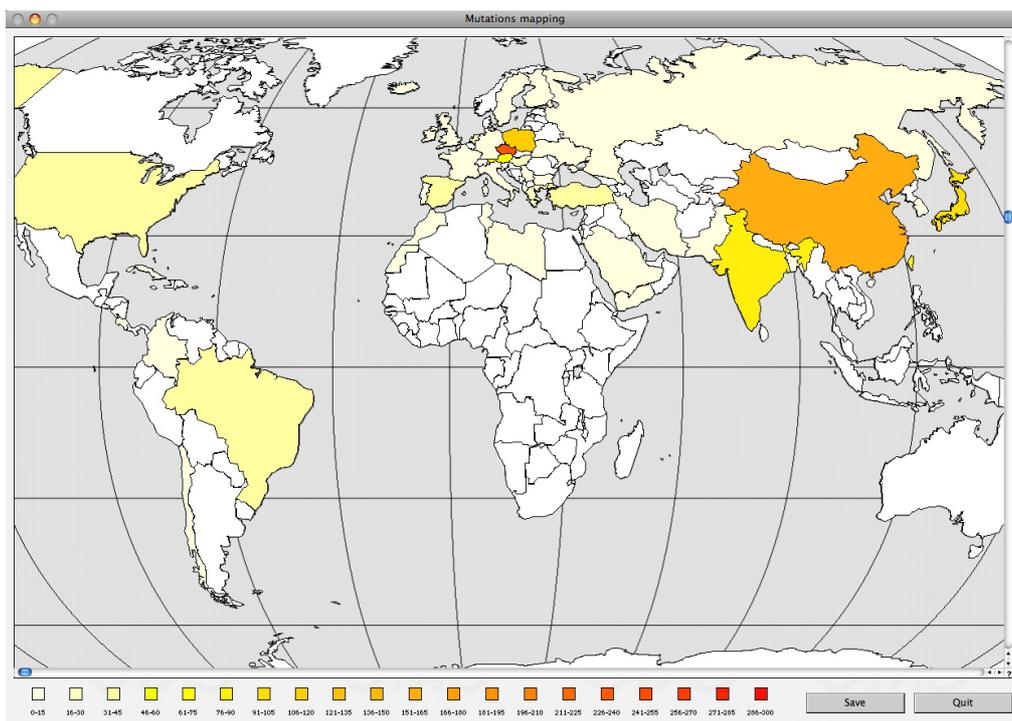
This function can only be used if data about geographic origin of patients have been collected. The user can choose all or a subset of records, select one or more mutation types and to display crude data or percentage relative to the selected records.

The screenshot shows a window titled "Geographic distribution of mutations". It contains a form with the following sections:

- Mutation(s) selection:** Includes a radio button for "All", a "Search" button, and a "Result" input field. Below it, a note states: "You can select one or more mutations using the search module."
- Values are presented as:** Includes radio buttons for "Crude data" and "Percentage relative to selections below".
- Select one or more mutation type(s):** Includes checkboxes for "Missense mutations", "Nonsens mutations", "Deletions", "Insertions", "Splice mutations", and "Truncated proteins".
- Select samples type(s):** Includes checkboxes for "All samples" and "Only probands".

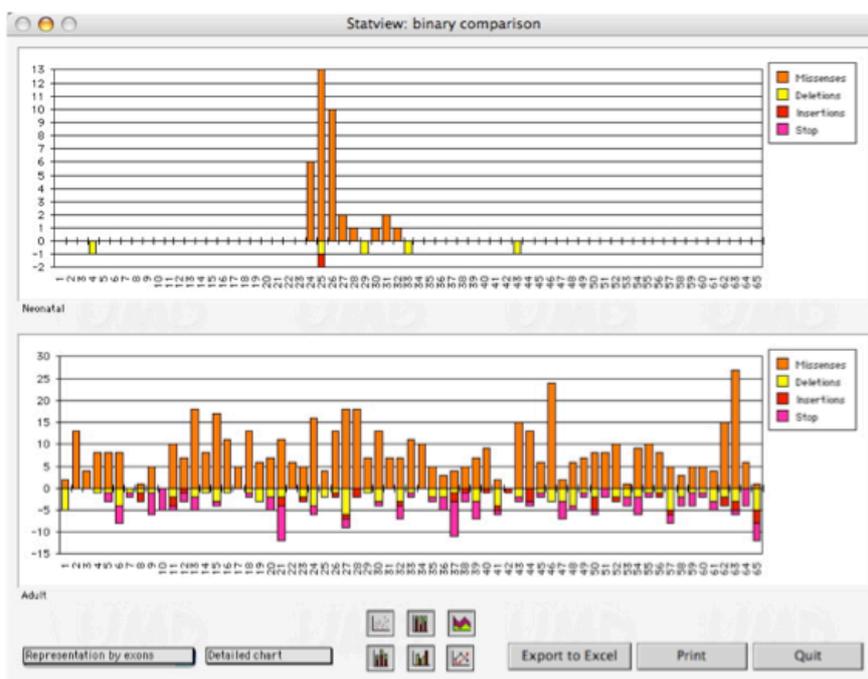
At the bottom, there are buttons for "Define links between geographic origins and countries", "OK", and "Quit".

The example below displays the distribution of missense mutations within the ATP7B database.



## 15) Binary comparison

This function is used to compare two sets of data. In this example the neonatal form of the



Marfan syndrome (top) to the Adult form of the disease.

Note that the user can use different parameters: representation by residues or exons; detailed (missense, deletions, insertions and nonsense mutations) or cumulated charts as well as various display options. Finally, the corresponding data can be exported "Export to Excel" button.

## 16) Stat exons

This function is used to look for hotspots associated with specific exons. The UMD® software analyzes each exon to define the expected mutation number corresponding to the hypothesis of random distribution of mutations. Are taken into account the exons lengths and

Exon	Observed mutations	Expected mutations	Significance
1	95	7,8	<0,001
2	1	13,7	<0,001
3	172	19,4	<0,001
4	19	16,4	<0,9
5	46	20,2	<0,001
6	29	37,9	<0,2
7	85	25,7	<0,001
8	6	40,3	<0,001
9	5	27,1	<0,001
10	47	41,4	<0,5
11	42	40,3	<0,9
12	3	33,7	<0,001
13	18	25,6	<0,2
14	24	22,5	<0,9
15	4	23,8	<0,001
16	6	39,1	<0,001
17	26	38,3	<0,05
18	6	27,0	<0,001
19	24	19,3	<0,3
20	9	53,5	<0,001
21	27	39,0	<0,1
22	8	32,2	<0,001
23	9	46,3	<0,001
24	3	24,8	<0,001
25	8	34,6	<0,001
26	16	38,7	<0,001
27	5	39,8	<0,001
28	6	29,7	<0,001
29	11	32,6	<0,001
30	8	35,6	<0,001
31	5	24,6	<0,001

Total number of records : 2411

their nucleotides compositions to take into account the degenerated code. After selection of records (see introduction of this chapter), the observed and expected number of mutations per exon are displayed as well as the statistical significance (Chi-square test).

## 17) Distribution by exon

This function is complementary to the previous one. It gives details about mutation types found in each exon (missense, insertions, deletions and nonsense mutations).

Exon	Total	Missenses	Insertions	Deletions	Nonsens
1	95	3,9 %	18 11,3 %	50 17,0 %	26 1,4 %
2	1	0,0 %	1 0,6 %	0 0,0 %	0 0,0 %
3	172	7,1 %	5 3,1 %	46 15,6 %	121 6,8 %
4	19	0,7 %	4 2,5 %	5 1,7 %	7 0,3 %
5	46	1,9 %	2 1,2 %	12 4,0 %	32 1,8 %
6	29	1,2 %	3 1,8 %	5 1,7 %	16 0,9 %
7	83	3,4 %	1 0,6 %	27 9,2 %	48 2,7 %
8	6	0,2 %	2 1,2 %	0 0,0 %	3 0,1 %
9	5	0,2 %	1 0,6 %	1 0,3 %	0 0,0 %
10	47	1,9 %	4 2,5 %	6 2,0 %	32 1,8 %
11	42	1,7 %	6 3,7 %	6 2,0 %	24 1,3 %
12	3	0,1 %	1 0,6 %	1 0,3 %	0 0,0 %
13	18	0,7 %	2 1,2 %	2 0,6 %	14 0,7 %
14	24	0,9 %	5 3,1 %	4 1,3 %	10 0,5 %
15	4	0,1 %	0 0,0 %	0 0,0 %	4 0,2 %
<b>TOTAL</b>	<b>2411</b>	<b>158 6,55 %</b>	<b>293 12,15 %</b>	<b>1756 72,83 %</b>	<b>204 8,46 %</b>

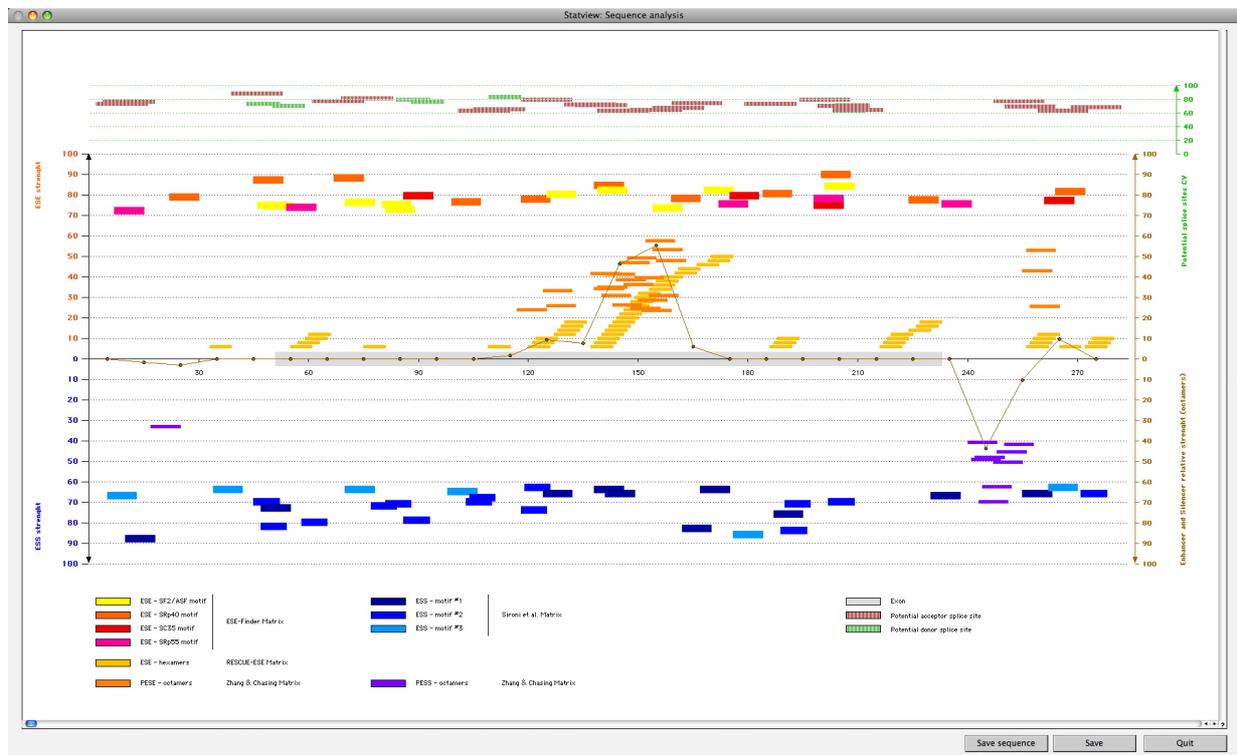
Total number of records : 2411

## 18) ESE map of exon

This function is used to predict the various auxiliary motifs (ESE) using the ESE-finder matrix. We recommend that you use the "Sequence analysis" function that gives more complete predictions.

## 19) Sequence analysis

This function is used to predict splicing signals including donor and acceptor splice sites as well as the various auxiliary motifs (ESE and ESS) using multiple matrices (for more information, view the Human Splicing Finder tool [Nucleic Acid Research, 2009, April](http://www.umd.be/HSF/) or visit the corresponding website <http://www.umd.be/HSF/>).



## 20) Splice mutations

This function is used to display the consequence of intronic mutations on splice sites and branch points.

The UMD algorithm is used for predictions and each result is displayed with a color code:

- Dark green = very strong splice site;
- Green = strong splice site;
- Light green = weaker splice site;
- Red = note a splice site.

The strength variation between the wild type and mutant potential splice sites are also given

Statview: splice mutations

**INTRONIC MUTATIONS: evaluation of splice-site consequences**

	Mutation	Splice site type	Wt sequence	CV	Mutant sequence	CV	Variation (%)
1	c.IVS1+1G>T (c.31+1G>T)	Donor	GTTgtaagt	82,62	GTTttaagt	55,79	-32.48
2	c.IVS1+5G>C (c.31+5G>C)	Donor	GTTgtaagt	82,62	GTTgtaact	70,61	-14.54
3	c.IVS1+36947G>A (c.31+36947G>A)	Cryptic Acceptor?	TTGGTTTTGCGGCTT	58,41	TTGGTTTTGCAGCTT	87,38	33.14
4	c.IVS1+4insT (c.31+4insT)	Donor	GTTgtaagt	82,62	GTTgtag	59,72	-27.72
5	c.IVS2+1del5 (c.93+1del5)	Donor	AAGgtaaga	86,71	AAGataaga	69,88	-27.75
6	c.IVS2-2A>T (c.94-2A>T)	Acceptor	ttttaattcagTT	89,22	ttttaattctgTT	60,27	-32.45
7	c.IVS2-1G>A (c.94-1G>A)	Acceptor	ttttaattcagTT	89,22	ttttaattcaaTT	60,27	-32.45
8	c.IVS3+1G>A (c.186+1G>A)	Donor	CTGtatgt	86,45	CTGatatgt	59,61	-31.04
9	c.IVS3-2A>G (c.187-2A>G)	Acceptor	ttttgttctcagCC	89,83	ttttgttctcggCC	60,88	-32.22
10	c.IVS3-1G>C (c.187-1G>C)	Acceptor	ttttgttctcagCC	89,83	ttttgttctcacCC	60,88	-32.22
11	c.IVS5+1G>C (c.357+1G>C)	Donor	CAGgtaaga	87,68	CAGctaaga	70,83	-27.48
12	c.IVS5+1G>A (c.357+1G>A)	Donor	CAGgtaaga	87,68	CAGataaga	70,83	-27.48
13	c.IVS5-1G>A (c.358-1G>A)	Acceptor	ttccacatgtagGT	83,78	ttccacatgtaaGT	54,83	-34.55
14	c.IVS5-2A>G (c.358-2A>G)	Acceptor	ttccacatgtagGT	83,78	ttccacatgtagGT	54,83	-34.55
15	c.IVS6-1G>T (c.531-1G>T)	Acceptor	tatgtgttttagGC	79,92	tatgtgttttatGC	50,97	-36.22
16	c.IVS7+2T>C (c.649+2T>C)	Donor	AAGgttggt	85,88	AAAGctggt	58,85	-31.32
17	c.IVS8-15A>G (c.832-15A>G)	Cryptic Acceptor?	CCAATCCCCAAACC	58,09	CCAATCCCCAGACC	87,04	33.26
18	c.IVS9-5925A>C (c.961-5925A>C)	Cryptic Acceptor?	CTTCATTTAAAGAGA	80,21	CTTCATTTACAGAGA	89,60	10.48
19	c.IVS9-5831C>T (c.961-5831C>T)	Cryptic Donor?	TTGgcaagt	66,43	TTGgtaagt	93,27	28.77
20	c.IVS11-9A>G (c.1332-9A>G)	Acceptor	tcaaatttcagTT	83,11	tcagatttcagTT	83,23	0.14
21	c.IVS11+1G>T (c.1331+1G>T)	Donor	CAAgtaagt	89,42	CAAttaagt	62,59	-30.01
22	c.IVS12+1G>T (c.1482+1G>T)	Donor	AAGtaggt	90,71	AAGttaggt	63,87	-29.58
23	c.IVS13+1G>A (c.1602+1G>A)	Donor	AAGgtcaga	81,89	AAGatcaga	64,85	-29.27
24	c.IVS12-1G>C (c.1483-1G>C)	Acceptor	tttatcttccagGT	94,86	tttatcttccacGT	65,91	-30.51
25	c.IVS13-2A>C (c.1603-2A>C)	Acceptor	tctcttccagGT	95,94	tctcttcccagGT	66,99	-30.17
26	c.IVS13-1G>A (c.1603-1G>A)	Acceptor	tctcttccagGT	95,94	tctcttcccaaGT	66,99	-30.17
27	c.IVS14+1G>T (c.1704+1G>T)	Donor	CAGgtgtgt	90,04	CAGttgtgt	63,20	-29.80
28	c.IVS18-1G>T (c.2293-1G>T)	Acceptor	ctcatgctcagGC	90,12	ctcatgctcatGC	61,18	-32.12

Export Quit

in %. Significant values (>10%) are indicated in red.

Note that deep intronic mutations from this example are predicted to activate acceptor or donor cryptic splice sites. All corresponding predictions have been proven experimentally. For more information [UMD-DMD database, Human Mutation](#).

## 21) Missense mutations

We have created the UMD-Predictor<sup>®</sup> tool to predict the pathogenicity of substitutions localized in the coding sequence. To have more information about how it works, read the [UMD-Predictor paper, Human Mutation](#).

Sasview: Missense mutations															
Prediction of deleterious missense variations															
Nonconserved C	Nonconserved L	Wild type AA	Mutant AA	Number of records	Reference	Structure	hCD	Conservation	SFF	Entropy	Biochemically	EE modification	Pathogenicity	Classification	
1	c.242A>G	p.R64L	Met	1	140	Signal peptide	0.5	1.00	-1.01	0.33	No impact	No impact	82	Pathogenic	
2	c.427T>A	p.L142I	Met	1	140	Signal peptide	0.5	0.19	-3.00	0.33	enable	No impact	88	Pathogenic	
4	c.558A>C	p.H190Y	Tyr	1	120	Signal peptide	0.5	0.19	-3.00	0.30	enable	Potential acceptor splice site (70-69)	71	Probably pathogenic	
5	c.1115G>C	p.Y373H	Thr	1	147	NAC-lesion region	0.5	1.00	-3.00	0.60	No impact	No impact	116	Pathogenic	
6	c.1175A>C	p.H390H	Ala	1	146	NAC-lesion region	0.5	0.00	-3.00	0.63	No impact	No impact	65	Probably pathogenic	
9	c.1845G>C	p.C615A	Gly	2	145	4-cystrine motif	0.5	0.00	-3.00	0.75	No impact	Donor splice site (intron) (72-73-55-91)	100	Pathogenic	
10	c.1884G>T	p.V628Y	Arg	8	64-95, 103, 145, 162	4-cystrine motif	0.67	0.00	-3.00	0.25	-SF2(A,F) (2, 63)	No impact	100	Pathogenic	
9	c.1888A>C	p.V628S	Tyr	1	96	4-cystrine motif	0.67	0.85	-2.00	0.38	No impact	No impact	71	Probably pathogenic	
8	c.2025A>T	p.C675S	Ser	1	145	4-cystrine motif	0.67	0.00	-3.00	0.63	No impact	Potential acceptor splice site (70-91)	84	Pathogenic	
11	c.2055G>C	p.V685S	Cys	1	96	4-cystrine motif	0.67	0.00	-3.00	0.63	No impact	No impact	82	Pathogenic	
12	c.2055G>T	p.C685I	Cys	1	154	4-cystrine motif	0.67	0.00	-3.00	0.46	No impact	No impact	100	Pathogenic	
13	c.2121T>C	p.H737R	Trp	1	149	4-cystrine motif	0.67	0.00	-3.00	0.38	No impact	No impact	100	Pathogenic	
14	c.2381C>G	p.C793G	Cys	1	145	4-cystrine motif	0.67	0.00	-3.00	0.25	No impact	No impact	100	Pathogenic	
15	c.2381C>A	p.C793T	Cys	1	127	4-cystrine motif	0.67	0.00	-3.00	0.46	No impact	No impact	100	Pathogenic	
16	c.2665A>A	p.C887Y	Cys	3	134, 145, 149	ECF-like #01	C in disulfide bonds 89-100	0.83	0.00	-2.00	0.39	No impact	No impact	100	Pathogenic
17	c.2665G>C	p.C887S	Cys	1	145	ECF-like #01	C in disulfide bonds 89-100	0.83	0.00	-2.00	0.46	-SF40 (2, 95) -SC35 (3, 35)	No impact	100	Pathogenic
18	c.2673C>G	p.C897P	Tyr	1	145	ECF-like #01	C in disulfide bonds 89-100	0.83	0.00	-2.00	0.39	-SF40 (2, 95)	No impact	100	Pathogenic
19	c.3035A>G	p.H1011A	Thr	1	144	ECF-like #01	0.83	0.00	-3.00	0.29	No impact	No impact	76	Pathogenic	
20	c.3337C>C	p.Y1114R	Gly	1	45	ECF-like #01	C in disulfide bonds 102-111	0.83	0.00	-3.00	0.25	-SF40 (2, 95)	No impact	100	Pathogenic
21	c.3346G>C	p.Y1114S	Ser	1	95	ECF-like #02	0.83	0.00	-3.00	0.63	-SC35 (2, 43)	No impact	71	Probably pathogenic	
22	c.3346G>T	p.Y1114Y	Arg	1	125	ECF-like #02	0.83	0.00	-3.00	0.25	-SC35 (2, 58)	No impact	100	Pathogenic	
23	c.3365A>G	p.Y1132C	Arg	8	10, 37, 89, 145, 154	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.36	-SC35 (2, 58)	No impact	100	Pathogenic
24	c.3365A>C	p.Y1132S	Gly	1	90	ECF-like #02	conserved AA in ECF-like	0.83	0.00	-3.00	0.29	-SF2(A,F) (1, 98)	No impact	100	Pathogenic
25	c.3365A>T	p.Y1132I	Trp	1	125	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.36	-SC35 (2, 58)	No impact	100	Pathogenic
26	c.4057C>C	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
27	c.4057C>G	p.C1314A	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
28	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
29	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
30	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
31	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
32	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
33	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
34	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
35	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
36	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
37	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
38	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
39	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
40	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
41	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
42	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
43	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
44	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
45	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
46	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
47	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
48	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
49	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
50	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
51	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
52	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
53	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
54	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
55	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
56	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
57	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
58	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
59	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
60	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
61	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
62	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
63	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
64	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
65	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
66	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
67	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
68	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
69	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
70	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
71	c.4057C>A	p.C1314T	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
72	c.4057C>G	p.C1314G	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
73	c.4057C>T	p.C1314C	Cys	1	149	ECF-like #02	C in disulfide bonds 123-134	0.83	0.00	-3.00	0.46	-SF40 (3, 7)	No impact	100	Pathogenic
74	c.40														

## 24) Structure

This function is used to list the various structural domains of the protein where are localized the mutations. The same analyze is also performed at the HCD level.

Structural domain	Number of records
Signal peptide	8 (0,45%)
HC unique region	5 (0,28%)
4-cys motif LTSP-like	18 (1,01%)
EGF-like #01	8 (0,45%)
EGF-like #02	20 (1,13%)
EGF-like #03	22 (1,24%)
Hydrol module#01	40 (2,25%)
cb EGF-like #01	5 (0,28%)
cb EGF-like #02	10 (0,56%)
TGFBRP #01	26 (1,46%)
Proline-rich	10 (0,56%)
EGF-like #04	19 (1,07%)
cb EGF-like #05	25 (1,39%)
cb EGF-like #06	45 (2,54%)
cb EGF-like #07	17 (0,96%)
cb EGF-like #08	40 (2,25%)
TGFBRP #02	29 (1,63%)
cb EGF-like #09	25 (1,39%)
cb EGF-like #10	27 (1,52%)
cb EGF-like #11	21 (1,18%)
Hydrol motif #02	40 (2,25%)
cb EGF-like #12	17 (0,96%)
TGFBRP #03	45 (2,54%)
cb EGF-like #13	40 (2,25%)
cb EGF-like #14	41 (2,31%)
cb EGF-like #15	41 (2,31%)
cb EGF-like #16	19 (1,07%)
cb EGF-like #17	19 (1,07%)
cb EGF-like #18	29 (1,63%)
cb EGF-like #19	27 (1,52%)
cb EGF-like #20	25 (1,39%)
cb EGF-like #21	18 (1,01%)

HCD	Number of records
N-Term Fur	2 (0,11%)
4-cysteine 1	11 (0,62%)
4-cysteine 2	11 (0,62%)
4-cysteine 3	11 (0,62%)
4-cysteine 4	11 (0,62%)
4-cysteine 5	11 (0,62%)
C-terminal	0 (0,00%)

Total nb of records: 1769

## 25) Global analysis

This function was designed to give a summary of mutations types related to a group of records. The "All mutations (homozygous)" button can be used to avoid bias when genes with recessive transmission are involved. In this situation homozygous mutations are counted twice. The results can be exported as a text-tab delimited file using the "Export to Excel" button.

TOTAL	1769
<b>Deletions and insertions</b>	<b>290 (16,39%)</b>
Deletions	222 (12,55%)
Out of frame deletions	196 (11,08%)
In frame deletions	26 (1,47%)
Insertions	68 (3,84%)
Out of frame insertions	63 (3,56%)
In frame insertions	5 (0,28%)
<b>Point mutations</b>	<b>1275 (71,96%)</b>
Missenses	1032 (58,34%)
Nonsens	241 (13,62%)
G->A	236
G->A at CpG	101
C->T	49
C->T at CpG	198
A->T	26
A->G	126
A->C	21
T->G	60
T->C	192
T->A	47
C->A	40
C->G	56
G->T	132
G->C	81
<b>Intronic mutations</b>	<b>206 (11,64%)</b>
<b>Complex mutations</b>	<b>5 (0,28%)</b>

All mutations (homozygotes)

Export to Excel

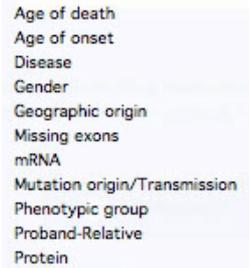
Print Quit

## 26) Free Graph

This analysis can be performed either on all patients or on a pre-selection of records. The field for the analysis can be chosen in a given list (depending of the gene. For a specific design, please contact us).

A graphical display of the field values is displayed. The user can choose the chart type.

This graph can be saved as picture using the "Save" button or printed using the "Print" button.



- Age of death
- Age of onset
- Disease
- Gender
- Geographic origin
- Missing exons
- mRNA
- Mutation origin/Transmission
- Phenotypic group
- Proband-Relative
- Protein

## 27) Specific tools

This function is activated on a collaborative base when a curator needs specific tools. For example tools related to exon skipping for the DMD gene...

If you need specific tools that are not yet available in the UMD software, please contact us.

## VIII-g- The "Phenotype" menu

This menu contains 13 functions, which are related to the clinical data analysis either using statistical or graphical presentations. Note that because clinical data are stored in an alpha-numeric field (severity), specific parameters should be used for proper analysis (cf. VIII-a-30 31 and 32).

Phenotype-Genotype analysis
Expressivity analysis
Onset of symptoms
Survival rate
Define an analysis matrix
Modify matrices
Analysis using a matrix
Difference - Delay Graph
LVEF or FVC Graph
Weight Graph
Graph per symptom
3D Graph with two symptoms
Graph per age

### 1) Phenotype-genotype analysis

Be sure that you first have used the "Update clinical data" function prior to use this function. A summary table is then displayed with, for each symptom, the corresponding severities and the number of patients.

Each symptom is displayed as a light grey row and the corresponding severities are listed below. The user can select a symptom by a double-click either on the symptom row or on any row corresponding to an associated severity. A triple star symbol (\*\*\*) will then appear in the selection column. A new double-click on the same lines will deselect the symptom. The analysis can then be performed with the "Display results" button.

Selection	Number of records	Symptom	Severity
231	*** 365	<b>Able to climb stairs</b>	
232	252	<i>Able to climb stairs</i>	N
233	5	<i>Able to climb stairs</i>	Never
234	12	<i>Able to climb stairs</i>	UK
235	92	<i>Able to climb stairs</i>	Y
236	*** 365	<b>Able to raise the hand up to the head</b>	
237	111	<i>Able to raise the hand up to the head</i>	N
238	48	<i>Able to raise the hand up to the head</i>	UK
239	202	<i>Able to raise the hand up to the head</i>	Y
240	365	<b>Able to read</b>	
241	79	<i>Able to read</i>	N
242	42	<i>Able to read</i>	UK
243	235	<i>Able to read</i>	Y
244	*** 365	<b>Able to rise from the floor</b>	
245	249	<i>Able to rise from the floor</i>	N
246	1	<i>Able to rise from the floor</i>	Never
247	10	<i>Able to rise from the floor</i>	UK
248	101	<i>Able to rise from the floor</i>	Y
249	365	<b>Able to run</b>	
250	214	<i>Able to run</i>	N
251	51	<i>Able to run</i>	Never
252	43	<i>Able to run</i>	UK
253	53	<i>Able to run</i>	Y
254	365	<b>Able to sit by himself</b>	
255	211	<i>Able to sit by himself</i>	N
256	14	<i>Able to sit by himself</i>	UK
257	136	<i>Able to sit by himself</i>	Y
258	365	<b>Age at last follow-up</b>	
259	365	<b>Age of 1st clinical sign</b>	
260	76	<i>Age of 1st clinical sign</i>	UK
261	365	<b>Age of 1st medical advice</b>	
262	18	<i>Age of 1st medical advice</i>	UK
263	365	<b>Anxiety disorder</b>	
264	277	<i>Anxiety disorder</i>	N
265	47	<i>Anxiety disorder</i>	UK
266	37	<i>Anxiety disorder</i>	Y

To select a Symptom for phenotype-genotype analysis, please double click on the corresponding row, symbol "\*\*\*" will then appear.

Export    Display results    Quit

The various combinations of symptoms and severities are then computed and the number of patients and associated genotypes are listed.

Phenotype-Genotype analysis

	Number of records	Able to climb stairs	Able to raise the hand up to the head	Able to rise from the floor	cDNA nomenclature	Genotypes	Number of records	Samples
197					cIV522+1C>T (c2949+1C>T)		1	F3427615591
198	1	N	Y	UK				
199								
200	12	N	Y	Y	c7099_7660del	deletion from exon 49 to 52	1	F5908016771
201					c2568T>A	p.Leu53X	1	F3423914111
202					c265_649del	deletion from exon 5 to 7	1	F1303812521
203					c3500_3501delCA	Small rearrangement	1	F3429145591
204					c4250T>A	p.Leu141L>Y	1	F3418210761
205					c6439_7872del	deletion from exon 45 to 53	1	F5901914341
206					c7201_7660del	deletion from exon 50 to 52	1	F6902000M91
207					c7201_8218dup	Duplication from exon 50 to 55	1	F5909304181
208					c7543_7660del	deletion from exon 52 to 52	1	F6906408911
209					c8217dup	p.Asp2740Argfs>S	1	F3436518771
210					c8608C>T	p.Arg2870X	1	F6921762MA1
211					c94_649del	deletion from exon 3 to 7	1	F3403201901
212					cIV563+2T>A (c9286+2T>A)		1	F69075072m1
213	1	Never	N	N				
214					c9287_9361del	deletion from exon 64 to 64	1	F3411813011
215	1	Never	N	Never				
216					c6439_7309del	deletion from exon 45 to 50	1	F1301911361
217	2	Never	Y	N				
218					c6615_6912del	deletion from exon 46 to 47	1	F3421012141
219					c6913_7309del	deletion from exon 48 to 50	1	F6308400471
220	1	Never	Y	Y				
221					c6439_7309del	deletion from exon 45 to 50	1	F1301911601
222	5	UK	UK	N				
223					c2293_6438del	deletion from exon 19 to 44	1	F3402601721
224					c6439_7309del	deletion from exon 45 to 50	1	F3403001811
225					c7201_7660del	deletion from exon 50 to 52	1	F3402001641
226					c7310_8217del	deletion from exon 51 to 55	1	F3406504051
227					c961_1331del	deletion from exon 10 to 11	1	F3402801751
228	4	UK	UK	UK				
229					c2301_3432del	deletion from exon 20 to 25	1	F3402901771
230					c2804_4845del	deletion from exon 22 to 34	1	F3403717761
231					c5923_6290del	deletion from exon 42 to 43	1	F3437519131
232					c961_1332dup	Duplication from exon 10 to 11	1	F5909404501
233	3	UK	Y	UK				
234					c6913_7200del	deletion from exon 48 to 49	1	F9407207011
235					c7099_7660del	deletion from exon 49 to 52	1	F1306208621
236					c9995T>G / c10018T>C	p.Phe3332Cys /	1	F3426416701
237	2	Y	Y	N				
238					c6439_6614del	deletion from exon 45 to 45	1	F1304407741
239					c94_649del	deletion from exon 3 to 7	1	F5902627461
240	2	Y	Y	UK				

Export Quit

In addition, the "Samples" column give an easy access to the corresponding patients. The results can be exported via the "Export" button, while the "Quit" button returns the user to the full list of symptoms and associated severities in order to perform another analysis.

## 2) Expressivity analysis

This function is used to study the expressivity of a specific symptom within one or multiple

Phenotype: Expressivity analysis

Expressivity analysis: definition of working groups

How many groups?

#	group	number of records
1	DMD	1435
2	BMD	709

Search

Which symptom?

Severity for 'Phenotype +'?

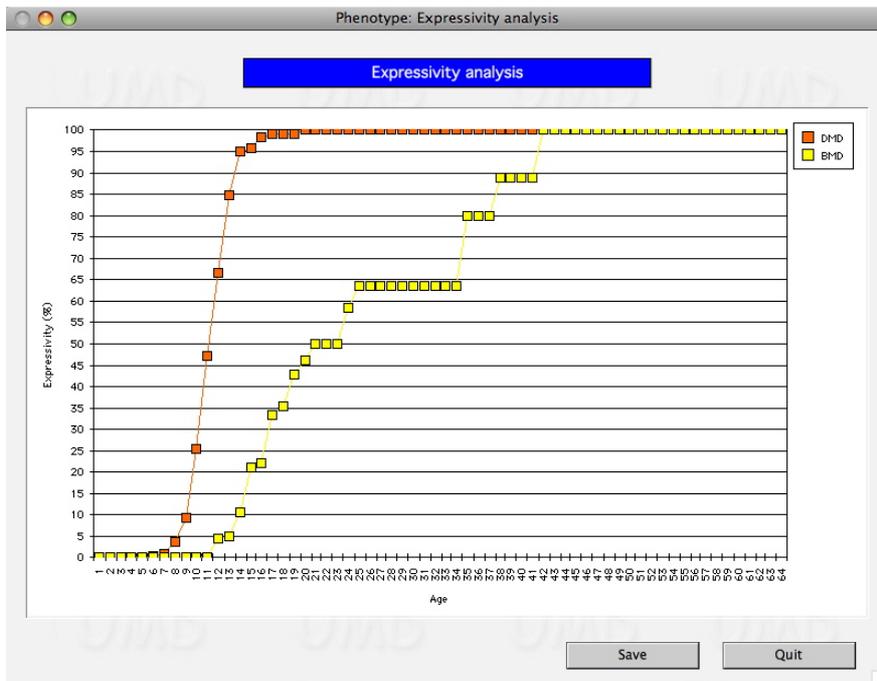
Severity for 'Phenotype -'?

The 0 value for 'Age' should be used as:  True data  Unknown age

OK Quit

groups (up to 8) of records defined by the user.

The user then selects a symptom and defines the severities associated to the "+" and "-" phenotype. The value 0 can either be used as a true data or as an unknown age as the default value for the "Age" field is 0 even if no data was available.



A graph with the expressivity (percentage of patients for which a data is available at a specific age and present the "+" phenotype) is then shown. The "Quit" button returns to the previous screen to perform other analysis.

### 3) Onset of symptoms

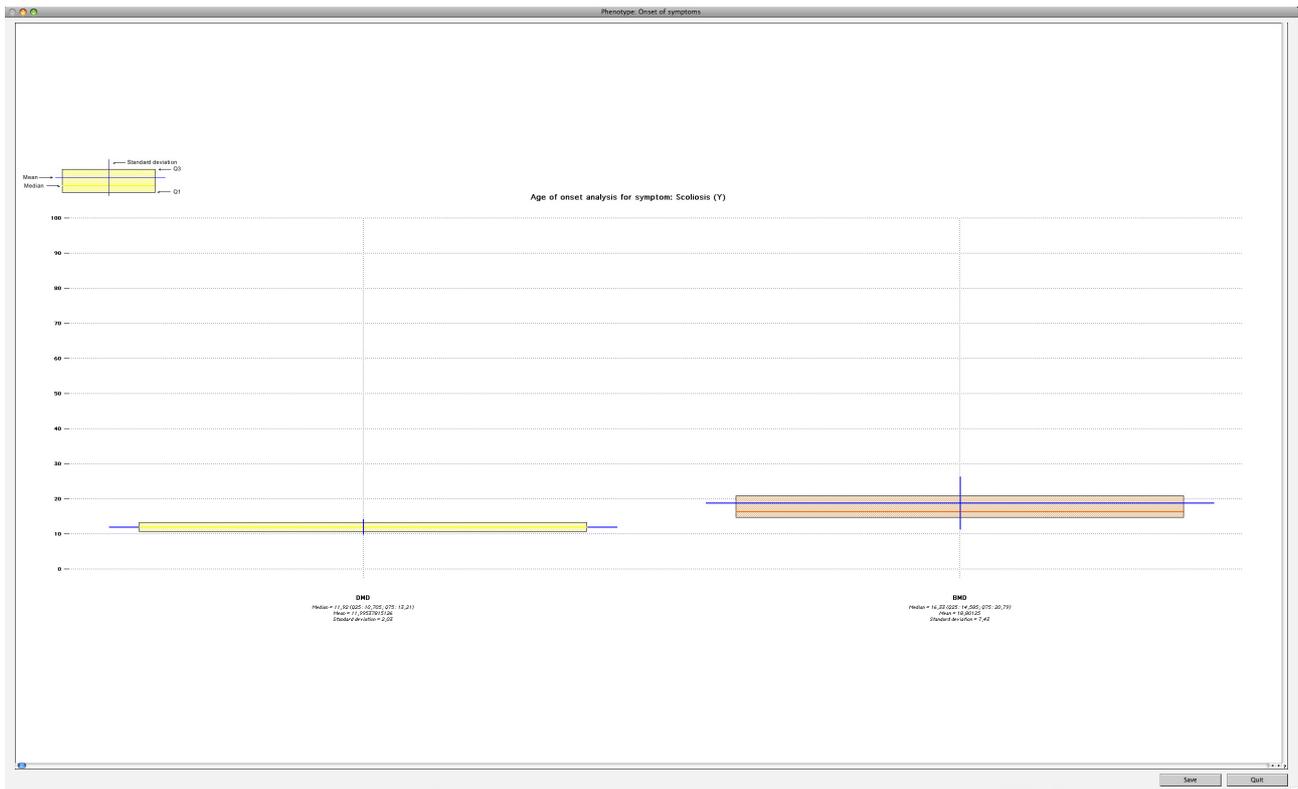
This function is used to study the age of onset of a specific symptom. As for the previous analysis, the user can define various groups of records and choose the symptom and the

The form is titled "Onset of symptoms analysis: definition of working groups". It includes a dropdown menu for "How many groups?" set to 2. Below is a table with columns for "#", "group", and "number of records".

#	group	number of records
1	DMD	1435
2	BMD	709

Below the table is a "Search" button and a text field for "Group name : BMD". At the bottom, there is a dropdown menu for "Which symptom?" set to "Scoliosis" and a text field for "Severity for 'Phenotype +'?" with the value "1". "OK" and "Quit" buttons are at the bottom right.

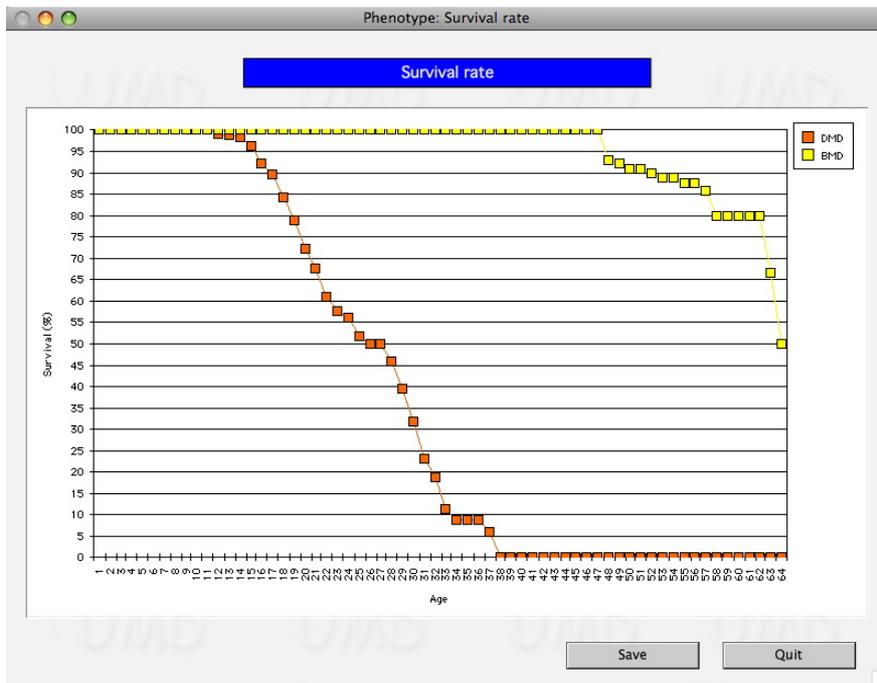
severity associated with the "+" phenotype. When the parameters are validated, a click on the "OK" button will open the result screen.



The following data are available for each group of records: Mean, median, standard deviation, first and third quartiles.

#### 4) Survival rate

The user first defines the various groups of records that should be used for this analysis (up to 8). The UMD<sup>®</sup> software will then analyze the data using the "Age of death" field and define the percentage of patients still alive at a specific age.



## 5) Define an analysis matrix

The analysis matrices are used to perform statistical analysis from clinical data. The user can create a virtually unlimited number of matrices. For each, he needs to select one or more "Analysis functions" from the list (left) and select one or more symptoms (right). When the

The window is titled "Phenotype: Define an analysis matrix". It contains a "Matrix title:" input field at the top. Below this are two columns: "Analysis function" and "Symptoms", each with a list of items and an "Order" column. A "Reset" button is located below each list. At the bottom right, there are "Save" and "Quit" buttons.

Analysis function	Order	Symptoms	Order
Available data		1st CK level	
Unknown data		Able to climb stairs	
Non available		Able to raise the hand up to the head	
Mean		Able to read	
Mean - 95% CI		Able to rise from the floor	
Minimum value		Able to run	
Maximum data		Able to sit by himself	
Standard deviation		Age at last follow-up	
Median		Age of 1st clinical sign	
Quartiles		Age of 1st medical advice	
Number of records		Anxiety disorder	
Table (values - frequencies) [95% CI]		Attention deficit / Hyperactivity disorder	
		Autistic disorder	
		Biliary lithiasis	
		Brain image	
		Calves hypertrophy	
		Cardiomyopathy	
		Chronic pain requiring the use of antalgic drug	
		Clinical symptoms of heart failure	
		Clinical symptoms of respiratory failure	
		Curative treatment with ACE inhibitors	
		Current cardiac medication	
		Current steroid therapy	
		Currently able to sit without support	
		Currently able to walk	

selections are completed, a unique matrix title should also be given (top).

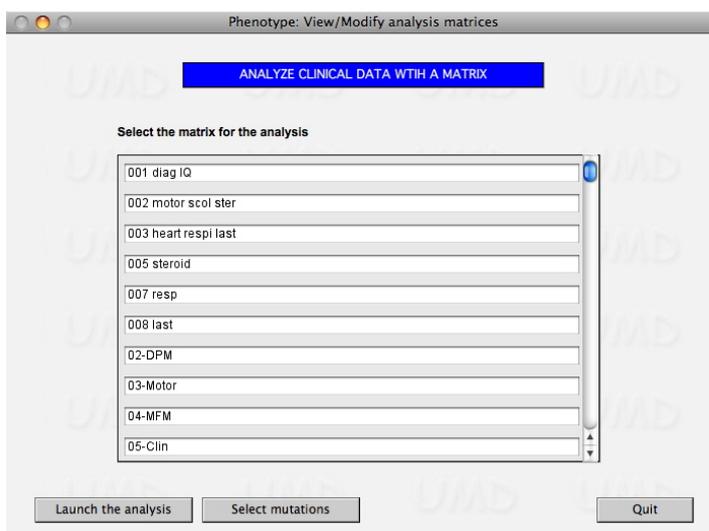
## 6) Modify matrices

This function lists all available matrices and allow their deletion or modification.

## 7) Analysis using a Matrix

The user can use a specific matrix from the list on all or a subset of clinical records. The "Select mutations" button gives access to the search interface in order to extract a group of records using any of the data fields (for more information about record extraction, cf. chapter VII).

A simple click on the matrices list will select the appropriate matrix and the "Launch the analysis" button will perform the analysis.



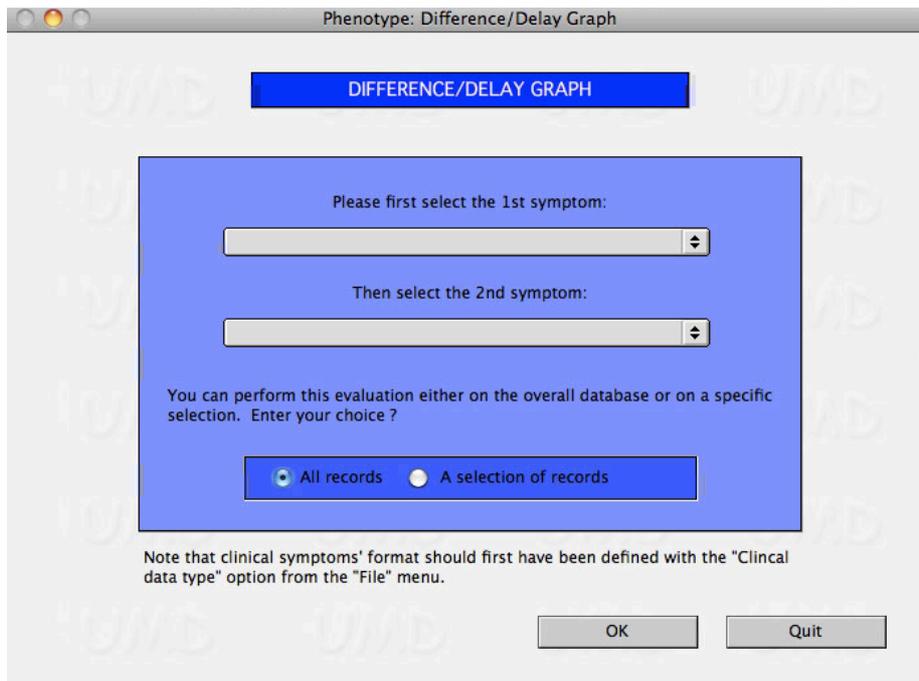
	Diagnosis	Age of 1st clinical sign - Age	Initial symptom	Age of 1st medical advice - Age
1 Available data	365	285	312	344
2 Number of records	365	365	365	365
3 Unknown data	0	80	53	21
4 Mean		5,53		7,63
5 Mean - 95% CI		[4,57 - 6,5]		[6,63 - 8,64]
6 Standard deviation		8,28		9,47
7 Minimum value		0,07		0,01
8 Maximum data		57,00		57,00
Table (values - frequencies) [95% CI]	BMD (84 - 23,01 %) [11,68% - 34,35%] DMD (256 - 70,14 %) [57,82% - 82,46%] IMD (10 - 2,74 %) [NA] pending (12 - 3,29 %) [NA] symptomatic carrier (3 - 0,82 %) [NA]	0,07 (1 - 0,35 %) 0,08 (1 - 0,35 %) 0,17 (1 - 0,35 %) 0,5 (4 - 1,4 %) 0,67 (1 - 0,35 %) 0,75 (6 - 2,11 %) 1 (16 - 5,61 %) 1,08 (1 - 0,35 %) 1,17 (1 - 0,35 %) 1,25 (4 - 1,4 %) 1,33 (1 - 0,35 %) 1,5 (24 - 8,42 %) 1,58 (2 - 0,7 %) 1,66 (1 - 0,35 %) 1,67 (4 - 1,4 %) 1,83 (4 - 1,4 %) 1,92 (2 - 0,7 %) 2 (38 - 13,33 %) 2,25 (1 - 0,35 %) 2,5 (11 - 3,86 %) 2,58 (1 - 0,35 %) 3 (32 - 11,23 %) 3,5 (14 - 4,91 %) 3,67 (1 - 0,35 %) 3,92 (1 - 0,35 %) 4 (29 - 10,18 %) 4,5 (5 - 1,75 %) 5 (10 - 3,51 %) 5,25 (1 - 0,35 %) 5,44 (1 - 0,35 %) 5,5 (2 - 0,7 %) 5,58 (1 - 0,35 %) 5,75 (1 - 0,35 %) 5,83 (1 - 0,35 %) 6 (10 - 3,51 %) 6,33 (1 - 0,35 %) 7 (5 - 1,75 %) 8 (3 - 1,05 %) 8,25 (1 - 0,35 %) 8,5 (1 - 0,35 %)	Motor abnormalities (223 - 71,47 %) [NA] ---->walk (160 - 51,45 %) [NA] ---->climbing stairs (31 - 9,97 %) [NA] ---->weakness (101 - 32,48 %) [NA] ---->congenital hypotonia (4 - 1,29 %) [NA] ---->running (15 - 4,82 %) [NA] DD (62 - 19,87 %) [NA] ---->cognitive (8 - 12,9 %) [NA] ---->motor and cognitive (33 - 53,23 %) [NA] ---->motor (21 - 33,87 %) [NA] Pain: cramping and/or myalgia (14 - 4,49 %) [NA] Toe walking (5 - 1,6 %) [NA] Muscle hypertrophy (5 - 1,6 %) [NA] Cardiomyopathy (3 - 0,96 %) [NA]	0,01 (1 - 0,29 %) 0,07 (2 - 0,58 %) 0,08 (21 - 6,1 %) 0,33 (1 - 0,29 %) 0,44 (3 - 0,29 %) 0,5 (3 - 0,87 %) 0,67 (1 - 0,29 %) 0,7 (1 - 0,29 %) 0,75 (1 - 0,29 %) 0,92 (1 - 0,29 %) 1 (2 - 0,58 %) 1,17 (1 - 0,29 %) 1,25 (1 - 0,29 %) 1,44 (3 - 0,87 %) 1,5 (2 - 0,58 %) 1,58 (2 - 0,58 %) 1,67 (2 - 0,58 %) 1,75 (3 - 0,87 %) 1,83 (2 - 0,58 %) 1,92 (3 - 0,87 %) 2 (9 - 2,62 %) 2,08 (1 - 0,29 %) 2,33 (1 - 0,29 %) 2,44 (1 - 0,29 %) 2,5 (3 - 0,87 %) 2,58 (1 - 0,29 %) 2,68 (1 - 0,29 %) 2,75 (2 - 0,58 %) 2,83 (2 - 0,58 %) 2,92 (2 - 0,58 %) 3 (6 - 1,74 %) 3,08 (3 - 0,87 %) 3,17 (2 - 0,58 %) 3,25 (4 - 1,16 %) 3,33 (4 - 1,16 %) 3,44 (2 - 0,58 %) 3,5 (4 - 1,16 %) 3,55 (1 - 0,29 %) 3,57 (1 - 0,29 %) 3,58 (2 - 0,58 %)

In this example 9 "analysis functions" have been selected and applied to various symptoms. Note that when severities are texts (Diagnosis, Initial symptom) only few functions could be applied. When numeric values or age are used (Age of first clinical sign, Age of first medical advice), all functions are used. Note that for all types of data are given: the various values, the number of occurrence, their frequency and the 95% confidence interval.

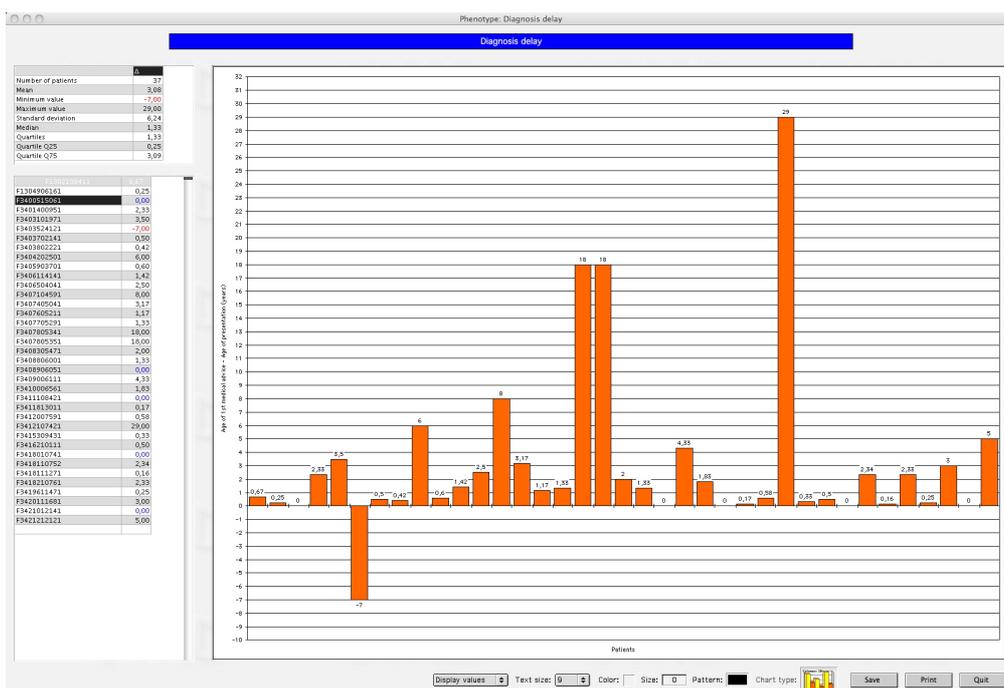
## 8) Difference - Delay Graph

This function was designed to evaluate the diagnosis delay, but can also be used for other purposes (differences between 2 dates or numerical values). It displays the difference between two “clinical items” chosen by the user for all records or for a selection of records. The chart type can be changed.

The user needs to select two symptoms: for a typical diagnostic delay graph, the first symptom is the “age of first medical advice” and the second symptom is the “age of presentation”.



The Y-axis represents the difference between the 2 values. In addition are presented the: mean, minimum and maximum values, standard deviation, median, quartiles, quartile Q25 and quartile Q75. The X-axis represents each individual.



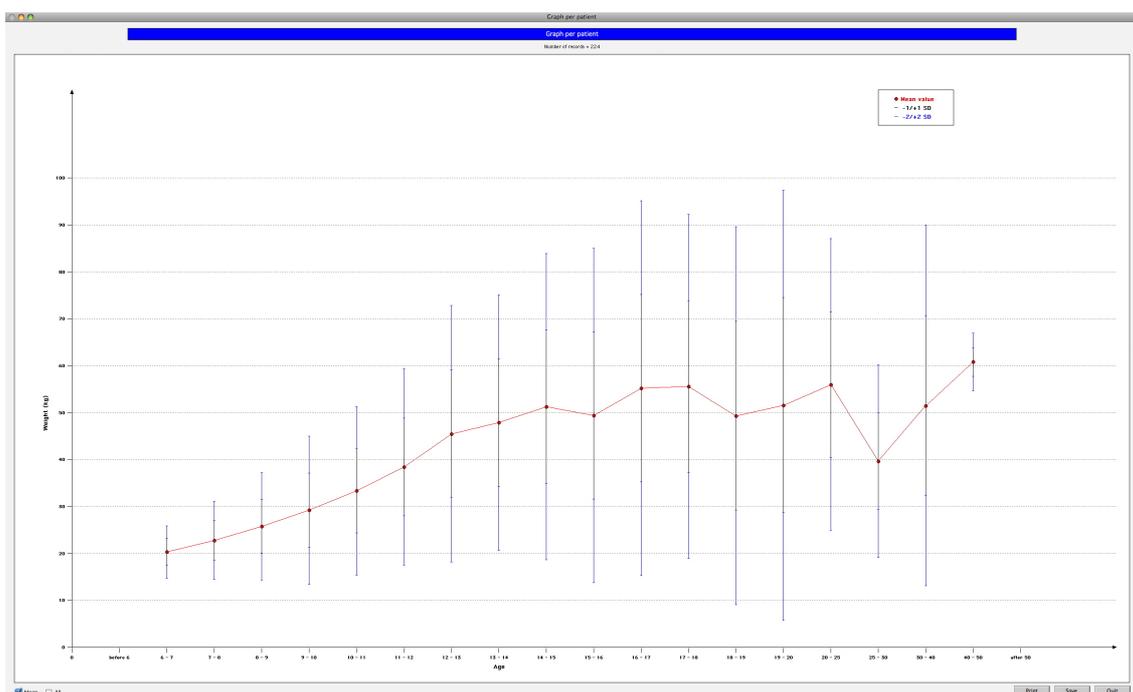
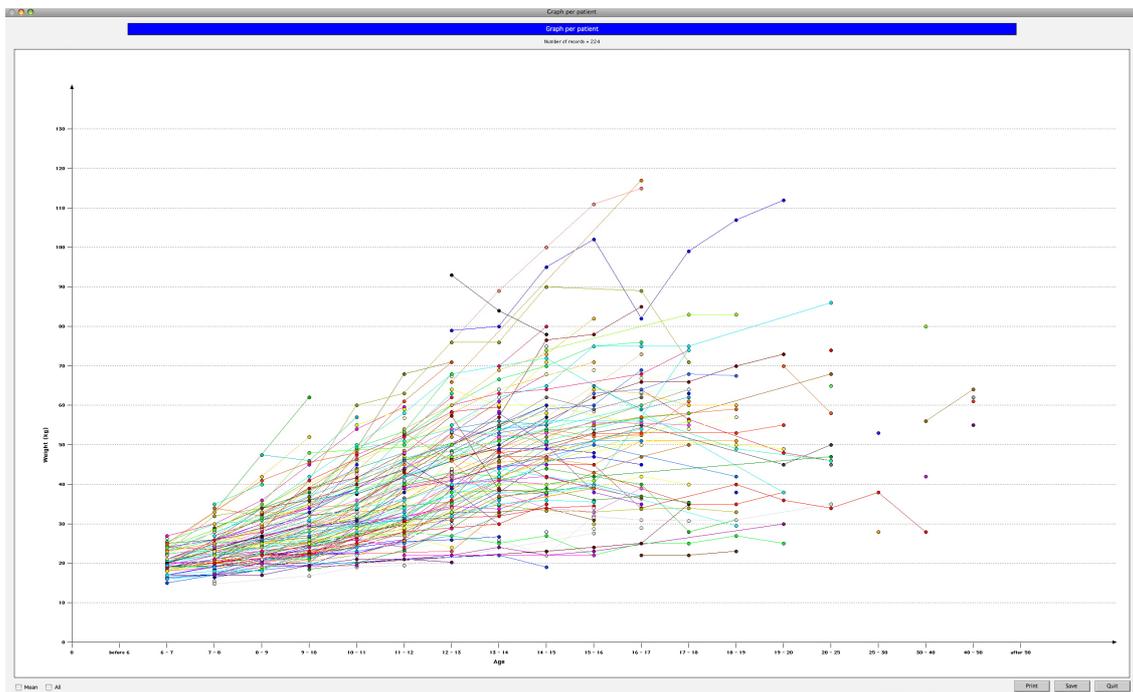
## 9) LVEF or FVC graph

These functions are only available for few curators. If you want to activate these functions, please contact us for more details.

## 10) Weight graph

This analysis can be performed either on one or all patients. A pre-selection of records can also be performed.

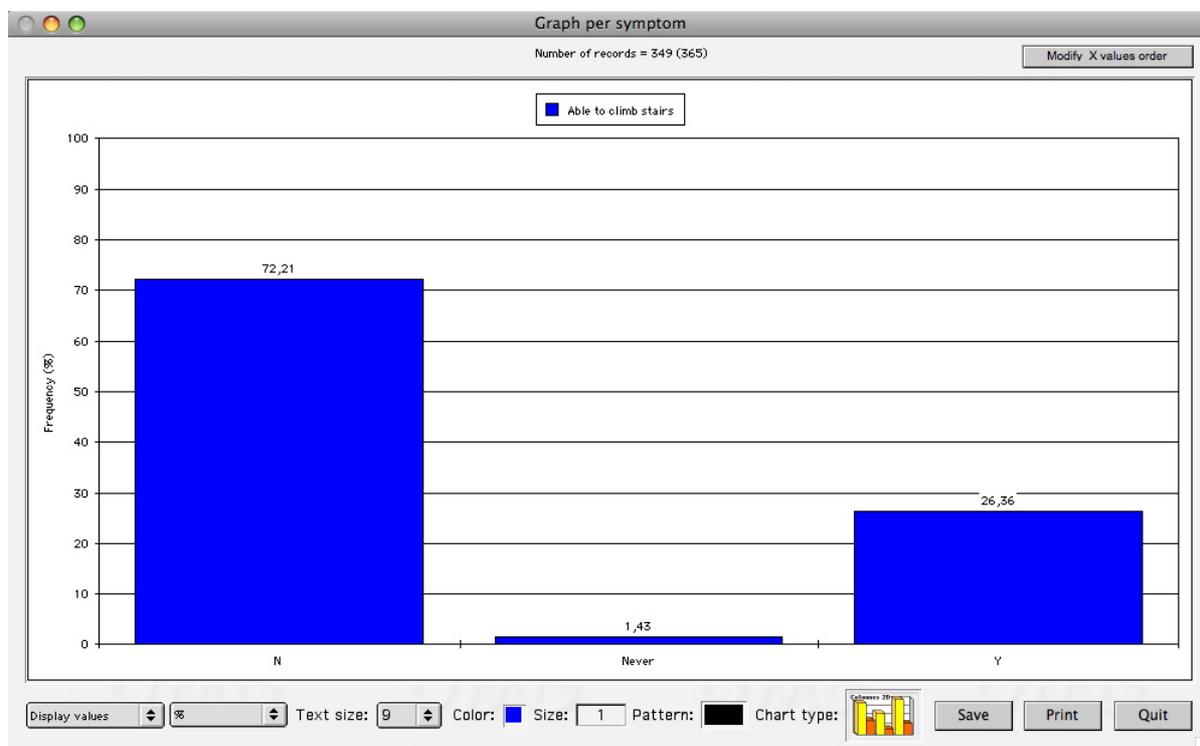
A graphical display of the weight evolution is displayed. The user can choose to combine all data and visualize mean values and standard deviations in a second graph.



Both graphs can be saved as pictures using the "Save" buttons or printed using the "Print" buttons.

## 11) Graph per symptom

This analysis can be performed on any symptom from the clinical data list. The user can



choose to perform it on all or only a subset of records.

The user can modify the graphics using the following tools:

- The "Display values" pop-up menu allows to hide or display the values;
- The "%" pop-up menu allows to display either the number of patients or the % values;
- The "Text size" to change the text size used in the graph;
- The "Color" can be changed using the 256 colors pallet;
- The "size" of border lines and the graph "Pattern";
- The "Chart type" to modify the graph type.

In addition, the "Modify X values order" button can be used to modify the order of severities within the graph.

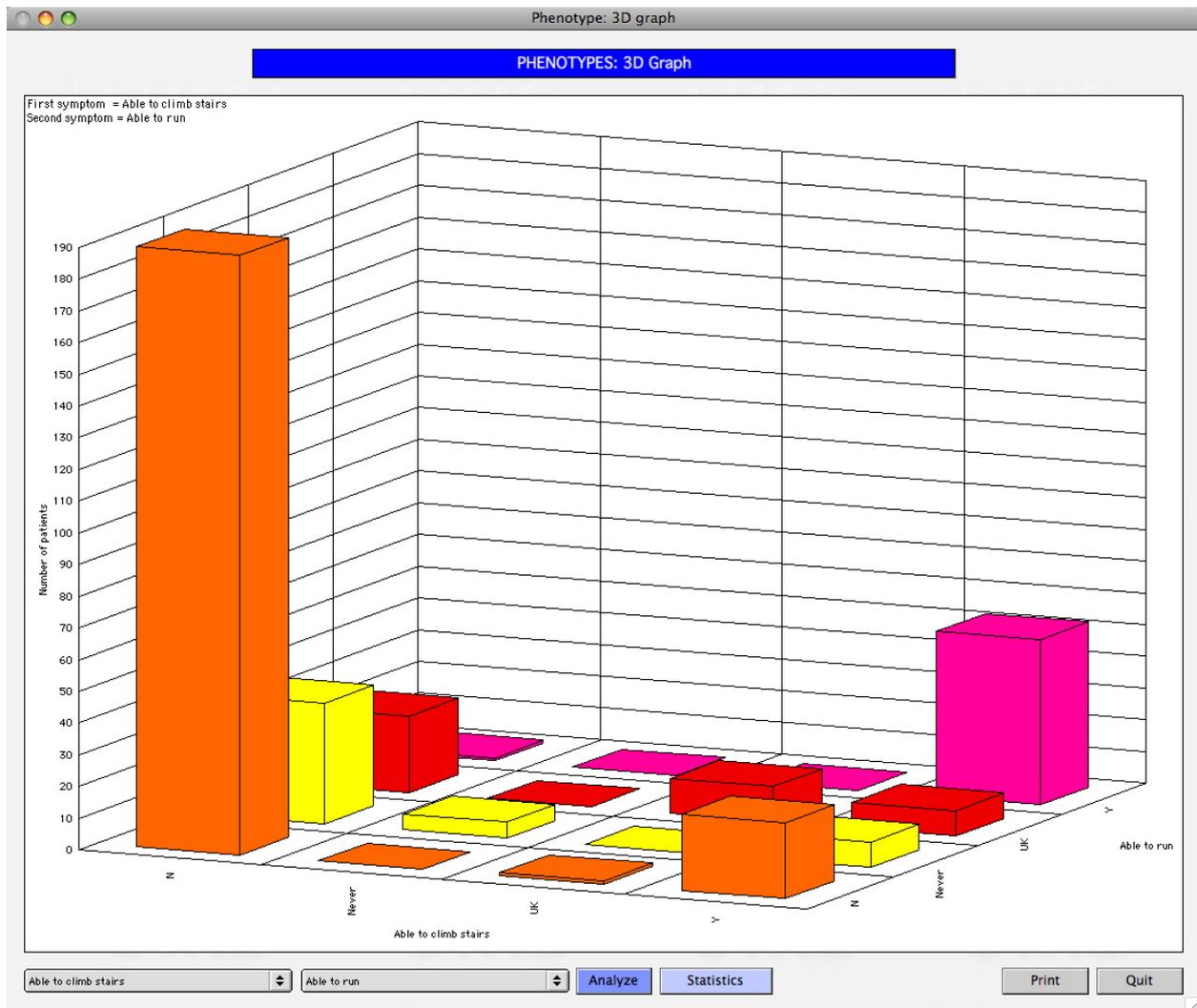
## 12) 3D graph with two symptoms

This function is used to combine 2 symptoms and display the various combinations of associated severities. This analysis can be performed on a selection of records such as a specific disease or phenotype.

The user first need to select 2 symptoms, he will be afterwards able to modify these symptoms directly from the graph.

The 3D graph vie can be modified using the left, right, up and down arrows.

Two pop-up menus allows the selection of different symptoms and the "Analyze" button will allow the update of the graph using these new parameters. Finally, the "Statistics" button will give access to a table with the numerical values used to draw the graph. These data can be exported.



	A	B	C	D	E	F
1			Able to run			
2			N	Never	UK	Y
3	Able to climb stairs	N	189	38	24	1
4		Never	0	5	0	0
5		UK	1	0	11	0
6		Y	24	8	8	52

### 13) Graph per age

This analysis can be performed on any symptom from the clinical data list. The user can choose to perform it on all or only a subset of records.

Two graphs will be displayed, the first one is similar to the one presented for the "*Graph per symptom*" function, while the second one has the same display than the "*Onset of symptoms*" function.

## VIII-h- The "Genotype" menu

This menu contains 3 functions related to genotypes.

Genotype-Phenotype analysis

Distribution by Class-Severity

3D graph

### 1) Genotype-Phenotype analysis

This function is used to display mutations (one or two alleles per patient), alleles and genotypes frequencies.

You can visualize a record by a single clic on the table.

1st mutation	2nd mutation	Sample ID	Geographic origin
c.350G>A	c.1521_1523deICTT	FR001-05174-01624	
c.350G>A	c.1521_1523deICTT	FR001-05200-01633	
c.366T>A	c.1521_1523deICTT	FR001-00115-00022	
c.366T>A	c.1521_1523deICTT	FR001-00800-00148	
c.366T>A	c.1521_1523deICTT	FR001-01257-00022	
c.366T>A	c.IVS4+3A>G (c.489+3A>G)	FR001-04127-01224	
c.422C>A	c.3454G>C	FR001-04184-01244	
c.442delA	c.3909C>G	FR001-01407-00282	
c.442delA	c.1521_1523deICTT	FR001-01166-00228	
c.442delA	c.1521_1523deICTT	FR001-01907-00438	
c.442delA	c.1624G>T	FR001-01935-00446	
c.454A>G	c.1521_1523deICTT	FR001-01990-00462	
c.IVS4+1G>T (c.489+1G>T)	c.IVS16+1G>A (c.2988+1G>A)	FR001-02488-00624	
c.IVS4+1G>T (c.489+1G>T)	c.1521_1523deICTT	FR001-00358-00047	
c.IVS4+1G>T (c.489+1G>T)	c.1521_1523deICTT	FR001-01025-00202	
c.IVS4+1G>T (c.489+1G>T)	c.1624G>T	FR001-01182-00217	
c.IVS4+1G>T (c.489+1G>T)	c.2125C>T	FR001-01676-00358	
c.IVS4+2T>G (c.489+2T>G)	c.2374C>T	FR001-00711-00129	
c.494T>C	c.IVS17a-26A>G (c.3140-26A>G)	FR001-02103-00490	
c.509G>A	c.1521_1523deICTT	FR001-03437-00969	
c.509G>A	c.3964_4440del	FR001-03563-01018	
c.509G>A	c.1521_1523deICTT	FR001-03608-01033	

Allele	Nb	Frequency
c.1521_1523deICTT	622	43,34 %
c.IVS8-12T[5] (c.1210-12T[5])	87	6,06 %
c.1624G>T	55	3,83 %
c.3909C>G	38	2,65 %
c.350G>A	25	1,74 %
c.3454G>C	20	1,39 %
c.3846G>A	17	1,18 %
c.IVS19+12191C>T (c.3717+12191C>T)	17	1,18 %
c.IVS5+1G>T (c.579+1G>T)	17	1,18 %
c.617T>G	17	1,18 %
c.2991G>C	15	1,05 %
c.IVS11+1634A>G (c.1679+1634A>G)	15	1,05 %
c.IVS10-1G>A (c.1585-1G>A)	13	0,91 %
c.IVS14b+5G>A (c.2657+5G>A)	12	0,84 %
c.2051_2052deInsG	10	0,70 %
c.IVS16+1G>A (c.2988+1G>A)	10	0,70 %
c.3276C>A	10	0,70 %

Genotype	Nb
c.3846G>A/c.1521_1523deICTT	5
c.1521_1523deICTT/c.1040G>A	4
c.3484C>T/c.1521_1523deICTT	4
c.1523T>G/c.1521_1523deICTT	4
c.3878_3881deITATT/c.3878_3881deITATT	4
c.1521_1523deICTT/c.1040G>C	3
c.1521_1523deICTT/c.1519_1521deIATC	3
c.2128A>T/c.1521_1523deICTT	3
c.2834C>T/c.1521_1523deICTT	3
c.IVS17a-26A>G (c.3140-26A>G)/c.1521_1523deICTT	3
c.2052dup/c.1521_1523deICTT	3
c.3208C>T/c.1521_1523deICTT	3
c.2856G>C/c.1521_1523deICTT	3
c.2374C>T/c.1624G>T	3
c.2991G>C/?	3
c.3846G>A/c.3454G>C	3
c.366T>A/c.1521_1523deICTT	3

Select a line for genotype-phenotype analysis

Draw graph

Total nb of patients: 757

Print

Quit

A simple click on the table gives access to the corresponding patient records.

To save the corresponding data of each table as a text-tab delimited file, click on the symbol.

The "Draw graph" button opens a graph that displays the various number of genotypes (see below) or alleles (from the upper screen).

A click on the "Complex allele" button opens a new window to display the complete genotype and therefore to complex alleles (see below). A simple click on the table gives access to the patient records.

Genotype: Genotypes distribution

**GENOTYPE: Genotypes distribution**

1st allele	2nd allele	Sample ID
p.Leu206Trp	p.Phe508del	FR001-04006-01176
p.Phe508del	p.Leu206Trp	FR001-04202-01290
p.Leu206Trp	c.IVS8-12T[5]	FR001-04441-01335
p.Phe508del	p.Gln220X	FR001-00199-00037
p.Gly542X	p.Gln220X	FR001-01434-00287
p.Lys684AsnfsX38	p.Gln220X	FR001-03118-00852
p.Phe508del	p.Leu227Arg	FR001-01525-00305
p.Phe508del	p.Val232Asp	FR001-01632-00343
p.Arg354Trp	p.Val232Asp	FR001-04274-01290
p.Arg1162X	p.Val232Asp	FR001-04333-01300
p.Phe508del	p.Gln237Glu	FR001-03745-01084
p.Phe508del	p.Gly241GlnfsX13	FR001-01709-00378
p.Phe508del	p.Met244Lys	FR001-03296-00917
c.IVS10-1G>A	p.Arg259Gly	FR001-02278-00539
p.Phe508del	p.Asn287Tyr / p.Gln685ThrfsX4	FR001-04642-01415
p.Ile507del	p.Asn287LysfsX19	FR001-01507-00298
c.IVS6+5G>A	p.Phe508del	FR001-03162-00871
p.Phe508del	p.Phe216LeufsX12	FR001-00151-00031
p.Phe508del	p.Phe216LeufsX12	FR001-00665-00120
p.Phe316LeufsX12	p.Lys710X	FR001-03003-00815
p.Asn1098LysfsX68	p.Arg534Trp	FR001-00000-00149
p.Ile507del	p.Arg334Trp	FR001-00000-00747
p.Arg334Trp	p.Phe508del	FR001-00185-00001
p.Phe508del	p.Arg334Trp	FR001-00784-00145
p.Arg334Trp	p.Phe508del	FR001-01029-00203
p.Phe508del	p.Arg334Trp	FR001-01081-00215
p.Phe508del	p.Arg334Trp	FR001-04784-01464
p.Ile536Lys	p.Phe508del	FR001-02031-00475
p.Thr338Ile	p.Phe508del	FR001-03625-01038
p.Thr338Ile	p.Gly542X	FR001-03657-01053
p.Phe508del	p.Arg347His	FR001-02454-00612
p.Phe508del	p.Arg347His	FR001-03235-00895
p.Phe508del	p.Arg347His	FR001-03257-00900
p.Arg347Pro	p.Tyr1092X	FR001-03303-00922
p.Phe508del	p.Arg347His	FR001-03917-01142
p.Arg347Pro	p.Arg347His	FR001-04499-01356
p.Phe508del	p.Phe508del	FR001-00000-00647
p.Phe508del	p.Arg347Pro	FR001-00075-00012
p.Arg347Pro	p.Phe508del	FR001-01903-00437
p.Phe508del	p.Arg352Gln	FR001-02753-00728
p.Trp361CysfsX8	p.Arg764GlnfsX7	FR001-00087-00167
p.Phe508del	p.Trp361CysfsX8	FR001-01472-00294

You can visualize a record by a single clic on the table.

Note that the "FR001-04642-01415" record harbors two alleles in the CFTR gene: The frequent p.Phe508del mutation and a complex allele p.Asn287Tyr/p.Gln6851ThrfsX4. A click on the "Tables" button opens the following window:

Genotype: Genotypes distribution

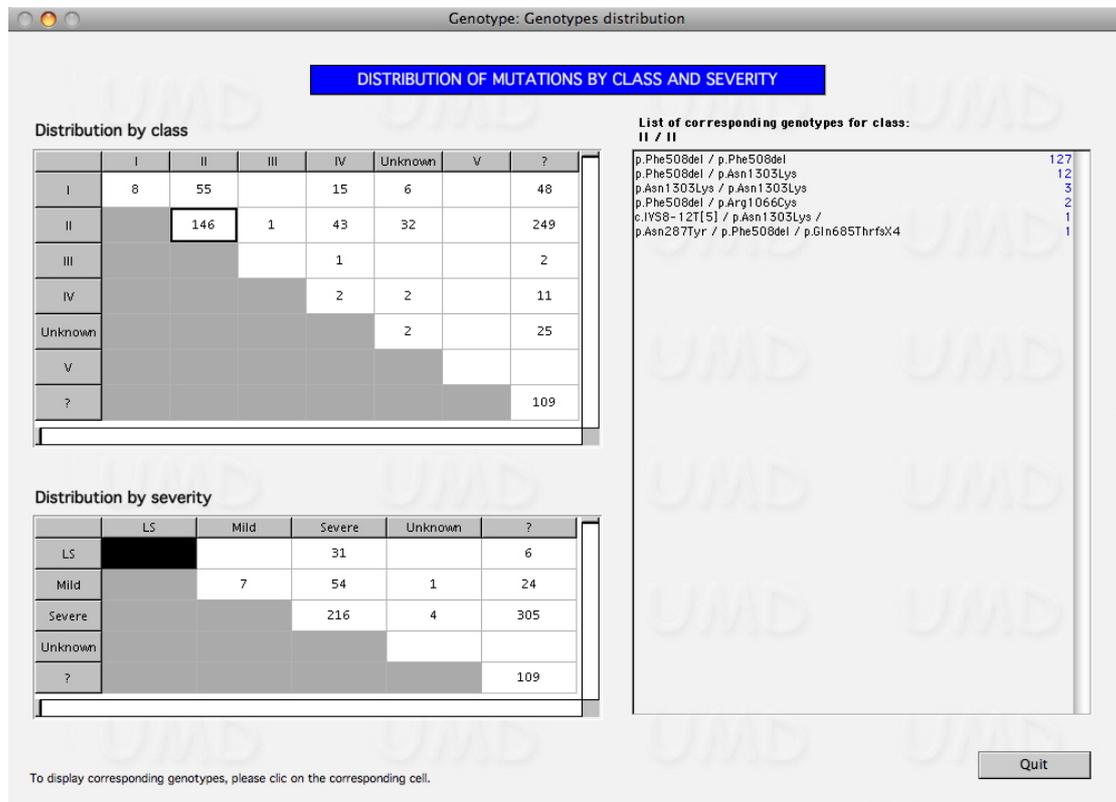
**GENOTYPE: Genotypes distribution**

Allele	Nb	Frequency	Genotype	Nb
p.Phe508del	498	40,42%	p.Phe508del	152
c.IVS8-12T[5]	72	5,84%	p.Phe508del / c.IVS8-12T[5]	44
p.Gly542X	54	4,38%	p.Phe508del / c.IVS19+12191C>T	11
p.Asn1303Lys	33	2,68%	c.IVS8-12T[5] (c.1210+12T[5])	11
p.Arg117His	23	1,87%	p.Phe508del / p.Asp1152His	9
p.Asp1152His	20	1,62%	p.Phe508del / p.Gly542X	9
p.Leu206Trp	17	1,38%	p.Gly542X / p.Phe508del	9
c.IVS19+12191C>T	15	1,22%	p.Phe508del / p.Arg117His	8
c.IVS11+1634A>G	15	1,22%	p.Gly542X	7
p.Leu997Phe	15	1,22%	p.Phe508del / c.IVS11+1634A>G	6
c.IVS5+1G>T	13	1,06%	p.Phe508del / p.Asn1303Lys	6
c.IVS10-1G>A	13	1,06%	p.Asn1303Lys / p.Phe508del	6
p.Trp1282X	13	1,06%	p.Phe508del / p.Leu206Trp	6
c.IVS14b+5G>A	11	0,89%	p.Leu206Trp / p.Phe508del	6
c.IVS8-12T[5] (c.1210+12T[5])	11	0,89%	p.Phe508del / p.Leu997Phe	5
p.Tyr1092X	10	0,81%	c.IVS10-1G>A / p.Phe508del	5
p.Lys684AsnfsX38	10	0,81%	c.IVS8-12T[5]	5
p.Arg334Trp	9	0,73%	p.Phe508del / c.IVS5+1G>T	5
c.IVS16+1G>A	8	0,65%	p.Phe508del / p.Gly178Arg	5
p.Arg553X	8	0,65%	p.Asn1303Lys	5
p.Phe508Cys	8	0,65%	p.Phe508del / p.Phe508Cys	4
p.Gly178Arg	7	0,57%	p.Phe508del / c.IVS14b+5G>A	4
p.Met117	7	0,57%	p.Phe508del / p.Tyr1092X	4
p.Gly550Glu	7	0,57%	p.Phe508del / p.Arg347His	4
p.Arg347His	7	0,57%	c.IVS5+1G>T / p.Phe508del	4
p.Lys710X	6	0,49%	p.Ile1295PhefsX32	4
p.Ile507del	6	0,49%	p.Gly542X / p.Arg792X	3
c.IVS17c+26A>G	6	0,49%	p.Phe508del / p.Arg1070Trp	3
p.Ser18_Glu54del	6	0,49%	p.Gln685ThrfsX4 / p.Phe508del	3
p.Gly551Asp	6	0,49%	p.Phe508del / p.Trp1282X	3
p.Arg1162X	6	0,49%	p.Phe508del / p.Ser945Leu	3
p.Arg1066Cys	6	0,49%	p.Phe508del / p.Arg1162X	3
p.Lys1177SerfsX15	5	0,41%	p.Phe508del / p.Lys710X	3
p.Arg792X	5	0,41%	p.Phe508del / p.Ile507del	3
p.Tyr122X	5	0,41%	c.IVS8-12T[5] / p.Phe508del	3
p.Ser945Leu	4	0,32%	p.Phe508del / p.Arg334Trp	3
p.Asp1202AlafsX9	4	0,32%	p.Trp1282X	3
p.Glu595X	4	0,32%	p.Leu997Phe	3
p.Arg170His	4	0,32%	p.Lys684AsnfsX38	3
c.IVS4+1G>T	4	0,32%	p.Trp1282X / p.Asp1152His	2
p.Ile148LeufsX5	4	0,32%	p.Gly551Asp / p.Lys1177SerfsX15	2
p.Gly1244Glu	4	0,32%	p.Gly542X / p.Arg1066Cys	2

All data can be exported via the corresponding symbol.

## 2) Distribution by Class-Severity

In the Variation table (see chapter VIII-l), each mutation can be associated to a mutation class and a mutation severity. These two annotations are used in this function for recessive



mutations.

For each patient, the two mutations are classified both at the class and at the severity levels in simple 2D tables. In the example presented here, 146 patients have 2 alleles harboring a class II mutation. A click on this box displays the list of corresponding genotypes on the large table on the right.

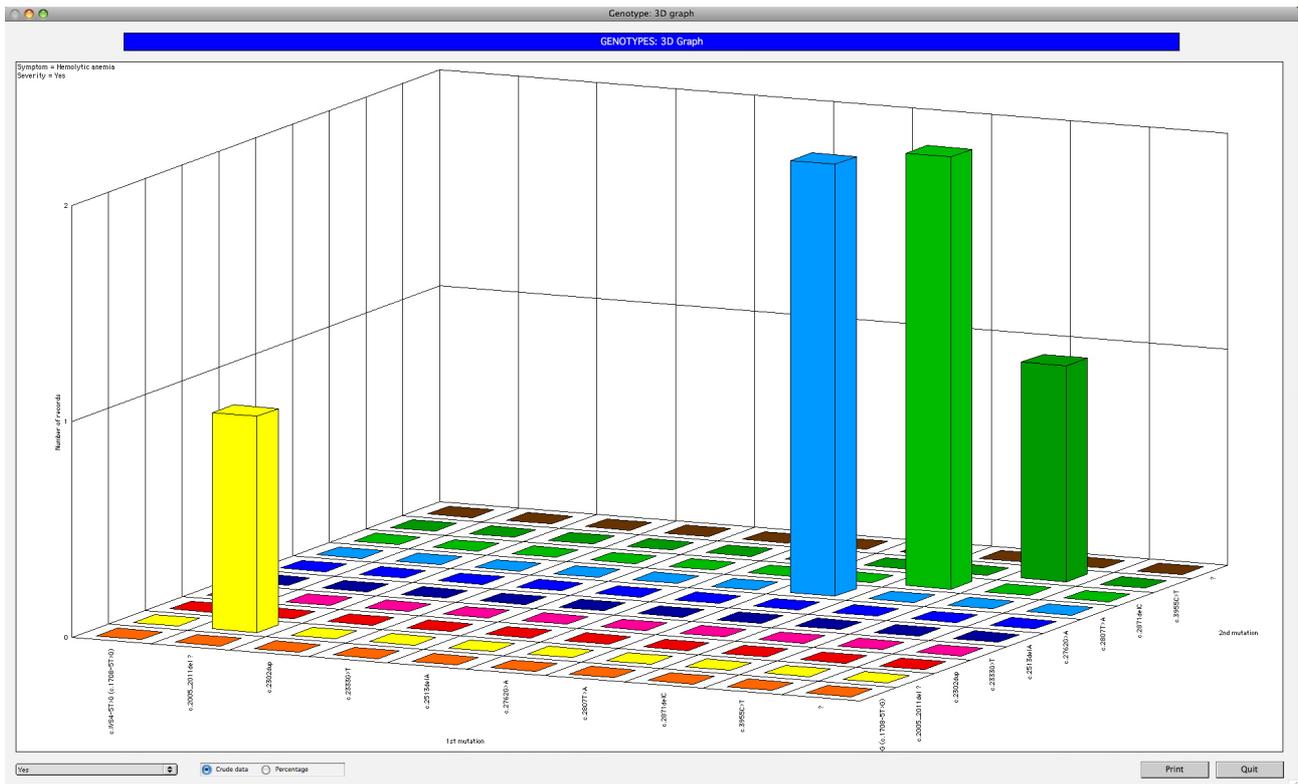
## 3) 3D Graph

As for the previous function, the "3D Graph" function can only be used for recessive mutations associated with clinical data.

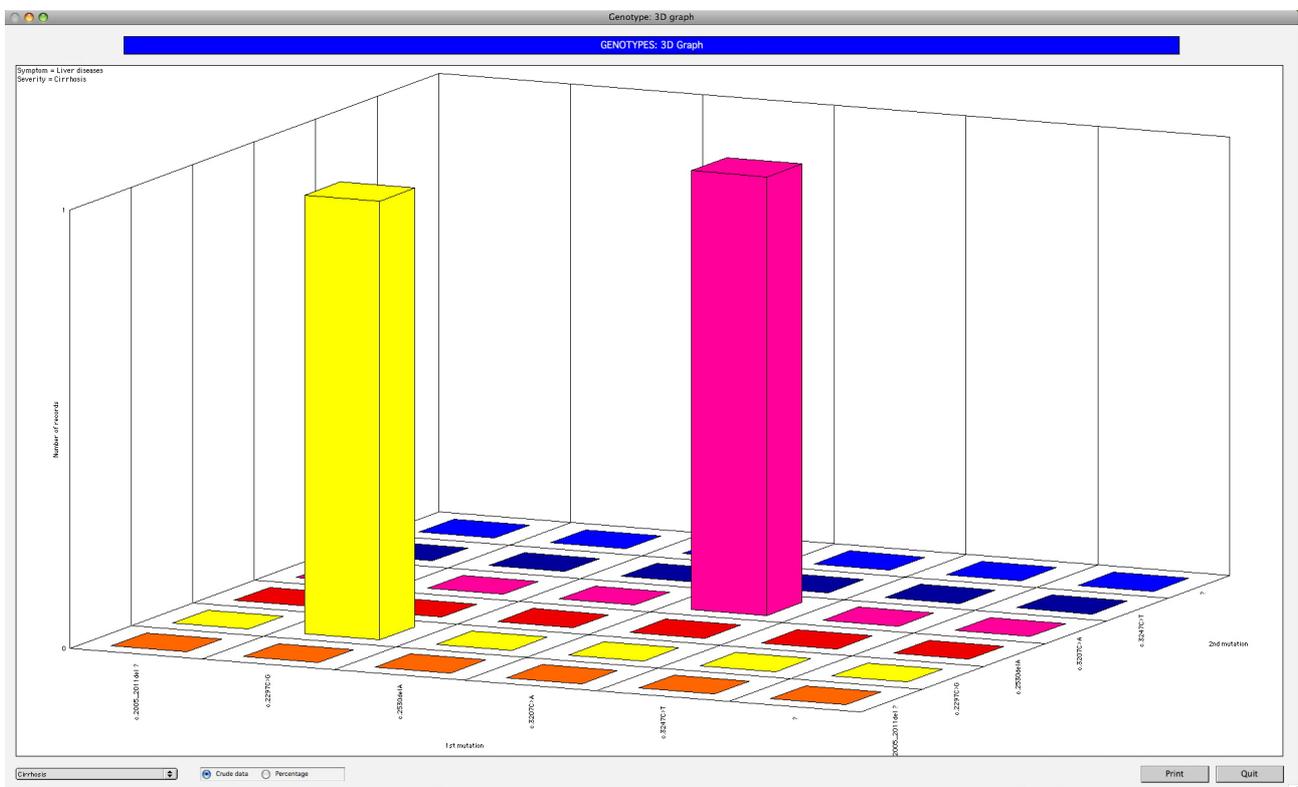
The user can first select a subset of records and then a specific symptom.

The UMD<sup>®</sup> software will then select all patients with data for the selected symptom. It will then extract the various severities reported for this symptom and then display a 3D Graph with the various genotypes for allele 1 (x-axis) and allele 2 (y-axis) and the number of patients with the corresponding severity (z-axis).

The user can use the "left", "right", "up" and "down" arrows in order to modify the 3D-graph display.



A pop-up menu on the left bottom corner gives access to the other severities associated with the symptom.



A second parameter (crude data / percentage) allows to modify the display z-axis format. In this example, only 2 homozygous ATP7B genotypes are associated with a cirrhosis.

## VIII-i- The "Haplotype" menu

This menu is only accessible for curators who use a specific format that includes the "Allele" field. In addition all variations should be reported (mutations and polymorphisms) with the allele annotation.

### 1) Haplotypes distribution

The user can select a specific group of records using the usual search interfaces. The UMD<sup>®</sup> software will then reconstruct the haplotypes for each patient when data about alleles are

Haplotypes	total	Allele #1	Allele #2
c.1000C>T	9	3	6
c.1007T>A	1	1	0
c.1013C>T - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[10] (c.1210-34TG[10])	2	2	0
c.1013C>T - c.IVS8-34TG[10] (c.1210-34TG[10])	1	1	0
c.1040G>A - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G	1	1	0
c.1040G>A - c.IVS8-12T[9] (c.1210-12T[9]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G	1	0	1
c.IVS6b+11C>T (c.869+11C>T) - c.1040G>A - c.IVS8-12T[9] (c.1210-12T[9]) - c.IVS8-34TG[10] (c.1210-34TG[10])	1	0	1
c.1040G>A - c.IVS8-12T[9] (c.1210-12T[9]) - c.IVS8-34TG[12] (c.1210-34TG[12])	1	0	1
c.1040G>A - c.IVS8-34TG[10] (c.1210-34TG[10])	1	1	0
c.1040G>A	2	0	2
c.1040G>C	3	2	1
c.1055G>A	1	0	1
c.1083delG	2	1	1
c.1135G>A - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11])	1	1	0
c.11C>A	3	1	2
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[10] (c.1210-34TG[10]) - c.1327G>T - c.1727G>C - c.2002C>T - c.2562T>G	1	1	0
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[10] (c.1210-34TG[10]) - c.1327G>T - c.1727G>C - c.2002C>T	10	2	8
c.IVS8-12T[7] (c.1210-12T[7]) - c.1327G>T - c.1727G>C - c.2002C>T	1	0	1
c.IVS8-34TG[10] (c.1210-34TG[10]) - c.1327G>T - c.1727G>C - c.2002C>T	2	0	2
c.1355A>C	1	0	1
c.1364C>A	1	0	1
c.137C>A	1	0	1
c.1399C>T - c.1521_1523delCTT	3	2	1
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.1477C>T	1	1	0
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.1523T>G	4	1	3
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.1545_1546delITA	1	1	0
c.1408A>G - c.1574_1590del	1	1	0
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.1631G>T	1	0	1
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[12] (c.1210-34TG[12]) - c.1408A>G - c.1657C>T	1	1	0
c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.1820_1903del	1	1	0
c.1A>G - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G	1	1	0
c.220C>T - c.801G>A - c.IVS8a-33GATT[7] (c.744-33GATT[7]) - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.38	1	0	1
c.220C>T - c.801G>A - c.IVS8-12T[7] (c.1210-12T[7]) - c.IVS8-34TG[11] (c.1210-34TG[11]) - c.1408A>G - c.3808G>A	5	2	3
c.220C>T - c.IVS8-12T[5] (c.1210-12T[5]) - c.IVS8-34TG[12] (c.1210-34TG[12]) - c.1408A>G	1	1	0

available. The first screen lists all haplotypes and the number of occurrences.

Seven buttons are available at the bottom of the screen:

- The "Allele-allele associations" button is used to build a table allowing to cross alleles. Various parameters can be used to color rows and columns corresponding to one or 2 alleles;
- The "Allele-haplotypes associations" button is used to build a table allowing to cross alleles with haplotypes. Various parameters can be used to color rows and columns corresponding to one or 2 alleles or only positive cells;
- The "Haplotype-haplotype associations" button is used to build a table allowing to cross alleles with haplotypes. Various parameters can be used to color rows and columns corresponding to one or 2 alleles or only positive cells;
- The "Haplotypes per patient" button is used to build a table allowing to cross patients haplotypes with alleles. The user can underline a specific variation with a color from the 256 pallet. He can also choose to display only mutations or all variations. In

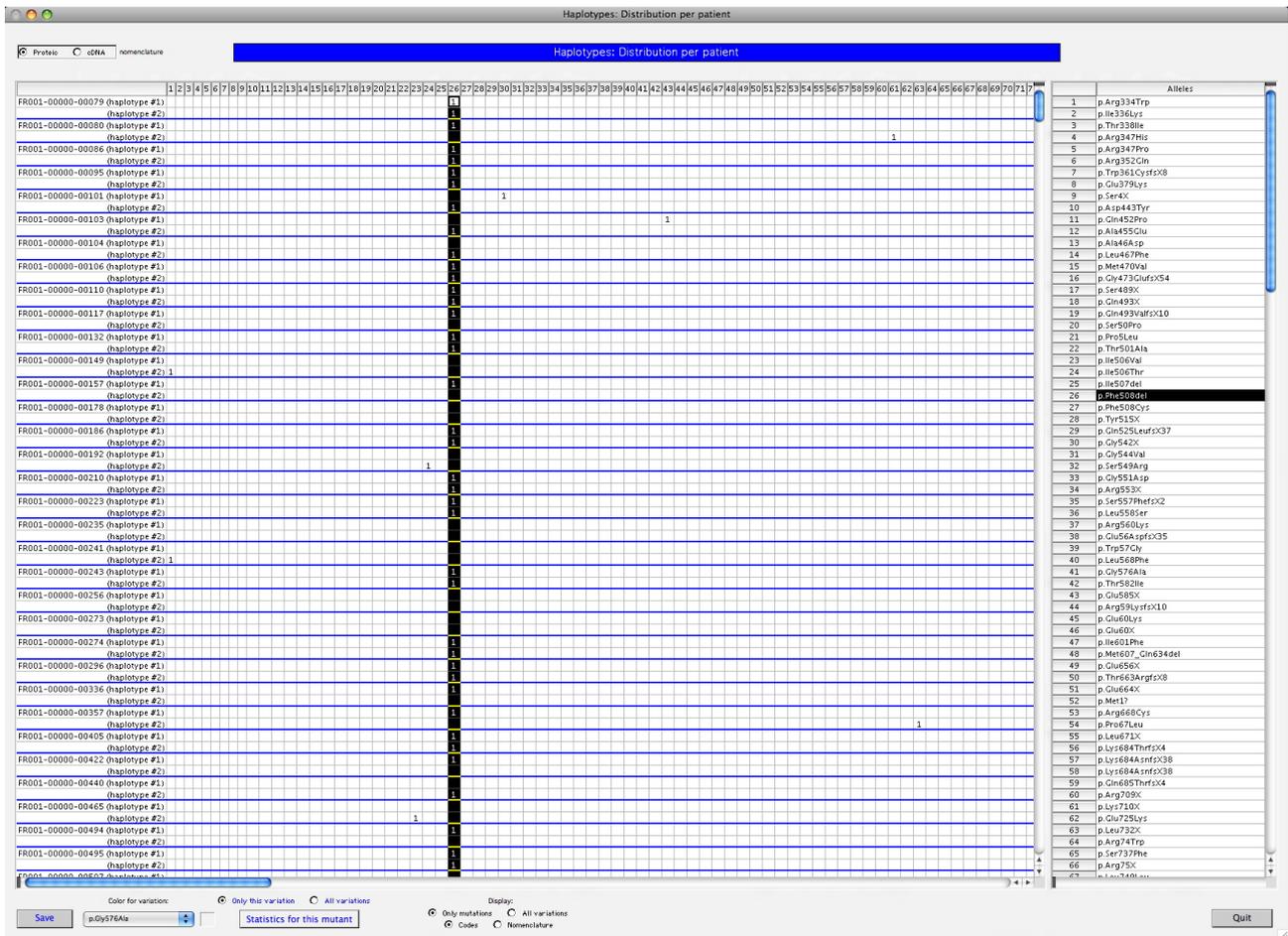
addition the c. or p. international nomenclature can be used or only a code to simplify the table. Finally, the "Statistics for this mutant" button allows to get details about this variation: number of homozygous and heterozygous and associated variations in cis or trans.

- The "Phylogeny (all haplotypes)" button is used to reconstruct the phylogenetic tree of the various chromosomes observed within patients. You need to install the dnamlk software prior to use this function. The UMD<sup>®</sup> software first prepare compatible files and launch the dnamlk application. An alert message ask you to indicate when the computation via the dnamlk is completed in order to import the results into UMD and display a phylogenetic tree with color codes within UMD. **Only available upon request.**
- The "Phylogeny (mutant haplotypes)" button is used to reconstruct the phylogenetic tree of the various chromosomes observed within patients. Only chromosomes harboring mutations are taken into account for this analysis. **Only available upon request.**
- The "Allele Frequency" button gives access to a simple table listing all variations and their frequency in the group of records selected by the user. The c. or p. nomenclature can be used for this analysis.

The screenshot shows a software window titled "Haplotype: Allele-allele associations". The window contains a table with columns for "Protein" and "cDNA" nomenclature. The table lists various mutations and their frequencies. A vertical pink bar highlights a column, and a horizontal yellow bar highlights a row. The table contains various mutation codes and their frequencies. At the bottom, there are controls for "All cells", "Positive cells", and color coding for columns and rows.

Protein	cDNA	homenclature
1	p.Arg334Trp	
2	p.Ile336Lys	
3	p.Thr338Ile	
4	p.Arg347His	
5	p.Arg347Pro	
6	p.Arg352Cln	
7	p.Trp361CysfsX8	
8	p.Glu379Lys	
9	p.Ser4X	
10	p.Asp443Tyr	
11	p.Cln452Pro	
12	p.Ala455Glu	
13	p.Ala46Aasp	
14	p.Leu467Phe	
15	p.Met470Val	2
16	p.Cly473GluufsX54	
17	p.Ser489X	
18	p.Cln493X	
19	p.Cln493ValfsX10	
20	p.Ser50Pro	
21	p.Pro5Leu	
22	p.Thr501Ala	2
23	p.Ile506Val	
24	p.Ile506Thr	
25	p.Ile507del	
26	p.Phe508del	
27	p.Phe508Cys	
28	p.Tyr515X	
29	p.Cln525LeufsX37	
30	p.Cly542X	
31	p.Cly544Val	
32	p.Ser549Arg	
33	p.Cly551Aasp	
34	p.Arg553X	
35	p.Ser557PhefsX2	
36	p.Leu558Ser	
37	p.Arg560Lys	
38	p.Glu56AaspfsX35	
39	p.Trp57Cly	
40	p.Leu588Phe	
41	p.Gly576Ala	
42	p.Thr592Ile	14
43	p.Cln595X	
44	p.Arg59LysfsX10	
45	p.Cln60Lys	
46	p.Cln60X	
47	p.Ile601Phe	
48	p.Met607_Cln634del	
49	p.Cln656X	
50	p.Cln656X	
51	p.Cln656X	
52	p.Cln656X	
53	p.Cln656X	
54	p.Cln656X	
55	p.Cln656X	
56	p.Cln656X	
57	p.Cln656X	
58	p.Cln656X	
59	p.Cln656X	
60	p.Cln656X	
61	p.Cln656X	
62	p.Cln656X	
63	p.Cln656X	
64	p.Cln656X	
65	p.Cln656X	
66	p.Cln656X	
67	p.Cln656X	
68	p.Cln656X	
69	p.Cln656X	
70	p.Cln656X	
71	p.Cln656X	
72	p.Cln656X	
73	p.Cln656X	
74	p.Cln656X	
75	p.Cln656X	
76	p.Cln656X	
77	p.Cln656X	
78	p.Cln656X	
79	p.Cln656X	
80	p.Cln656X	
81	p.Cln656X	
82	p.Cln656X	
83	p.Cln656X	
84	p.Cln656X	
85	p.Cln656X	
86	p.Cln656X	
87	p.Cln656X	
88	p.Cln656X	
89	p.Cln656X	
90	p.Cln656X	
91	p.Cln656X	
92	p.Cln656X	
93	p.Cln656X	
94	p.Cln656X	
95	p.Cln656X	
96	p.Cln656X	
97	p.Cln656X	
98	p.Cln656X	
99	p.Cln656X	
100	p.Cln656X	

Example of the "Allele-allele associations" function. Two mutations have been color-coded.



Example of the "Haplotypes per patient" function. The column corresponding the p.Phe508del mutation has been selected. Most patients are either heterozygous or homozygous for this mutation. Note that only the mutations are displayed. A click on the "Statistics for this mutant" button give the following results:

Haplotypes: Distribution per patient

Haplotypes: Distribution per patient

Homozygotes  Heterozygotes

	Variations in cis	Number	Variations in trans	Number
1	c.1399C>T	3	c.1000C>T	5
2	c.2562T>G	3	c.1007T>A	1
3	c.2719A>G	1	c.1013C>T	1
4	c.3080T>C	1	c.1040G>A	4
5	c.3199G>A	1	c.1040G>C	3
6	c.IVS6b+11C>T (c.869+11C>T)	1	c.1055G>A	1
7	c.IVS8-12T[9] (c.1210-12T[9])	100	c.1083delG	1
8	c.IVS8-34TG[10] (c.1210-34TG[10])	96	c.11C>A	1
9			c.1327G>T	6
10			c.1355A>C	1
11			c.1364C>A	1
12			c.1408A>G	56
13			c.1418delG	1
14			c.1477_1478delCA	1
15			c.148T>C	1
16			c.14C>T	1
17			c.1516A>G	1
18			c.1519_1521delATC	2
19			c.1523T>G	5
20			c.1545_1546delTA	1
21			c.1624G>T	17
22			c.1647T>G	1
23			c.1652G>A	1
24			c.1657C>T	2

Quit

## VIII-j- The "Introns" menu

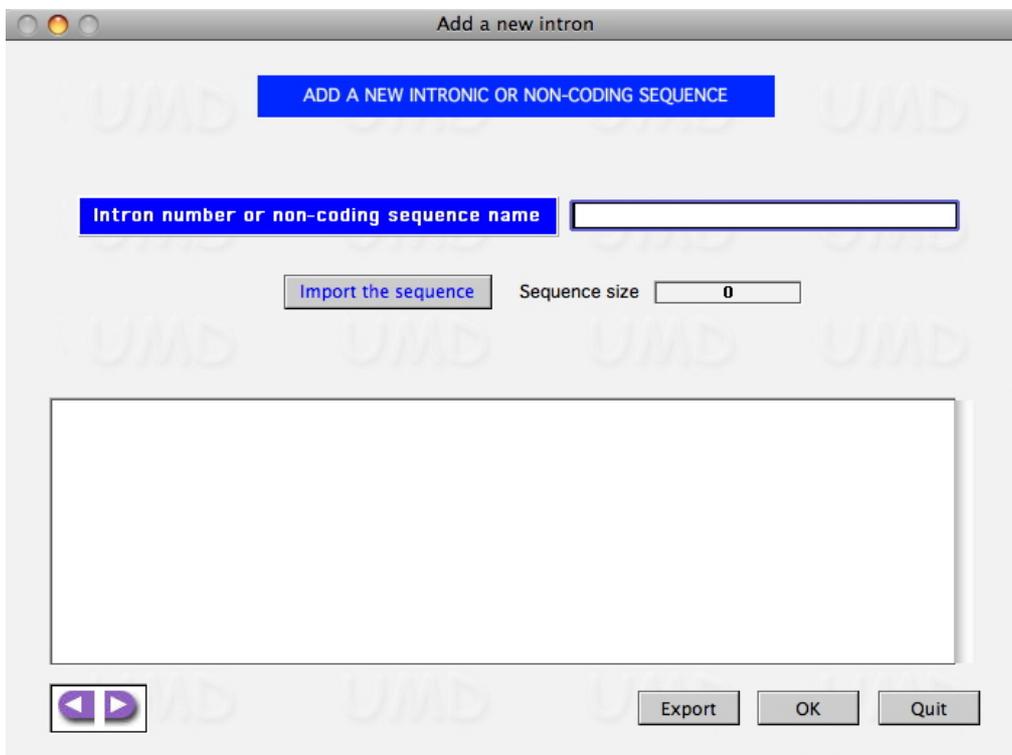
This menu contains 9 functions related to non coding sequences (introns and 5' and 3' regions).

Add intronic sequences
View-modify intronic sequences
Polymorphism map
Splice sites map
ESE map
Branch Point [one intron]
Branch Point [free selection]
Branch Point [all introns]
Branch Point [user sequence]

### 1) Add intronic sequences

This function is used to add manually non-coding sequences into the database. If your gene of interest is available via NCBI, UCSC or Ensembl, we recommend that you import data using the "Import polymorphic markers" function from the "File" menu.

**Important:** All introns should be labeled "Intron #xx" where xx stands for the intron number. Other regions should be labeled "3' region" and "5' region". If you use other names, these sequences will not be recognized by the software.

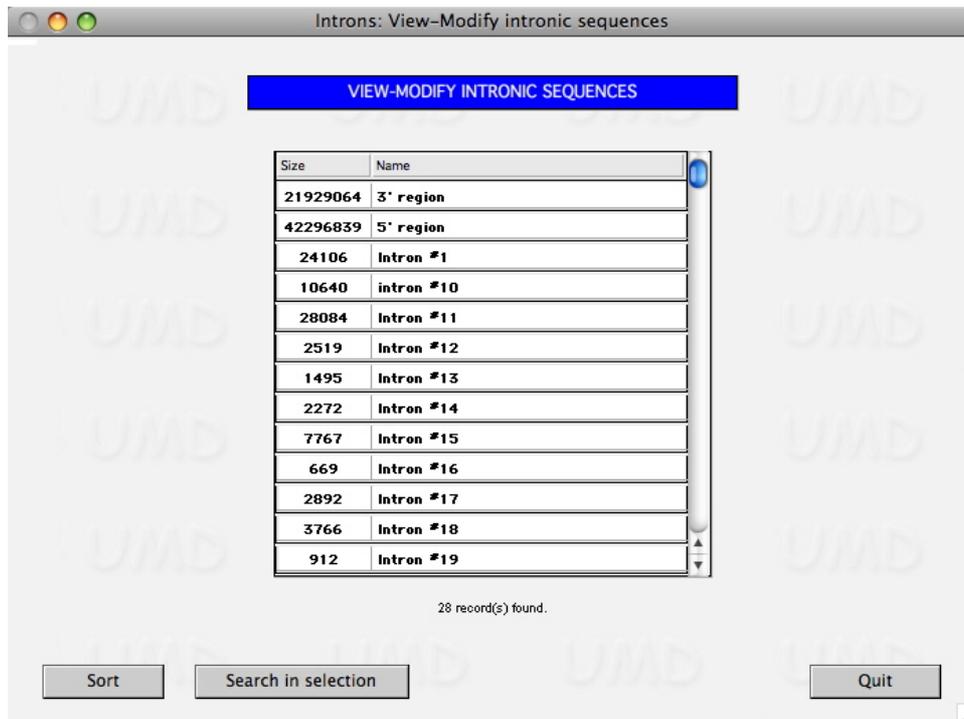


The "Import the sequence" button allows the import of a text sequence with only A, T, G or C characters (no numbers or end of line symbols). After the import the sequence size is displayed for control and the imported sequence can be visualized. Because only 32000 characters can be displayed at a time, 2 arrows allow to navigate within the sequence.

### 2) View-modify intronic sequences

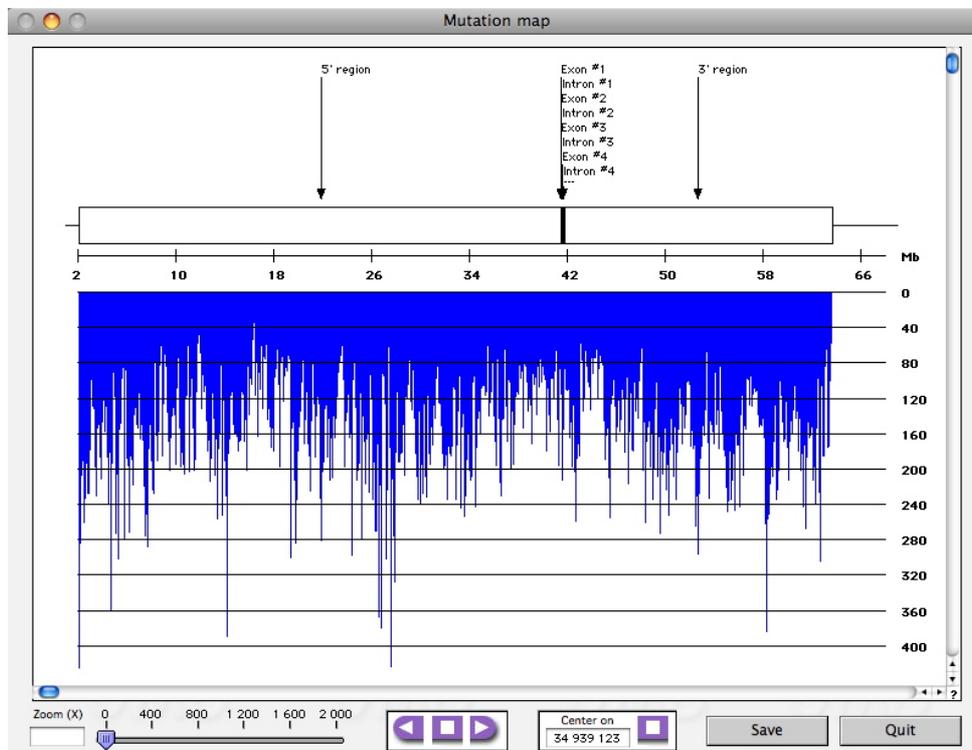
This function is used to display non-coding reference sequences used by the UMD® software and eventually modify these sequences.

A simple list of all records is presented. A click on the chosen non-coding sequence will open a screen similar to the one used for creation of intronic sequences (see previous function).



### 3) Polymorphism map

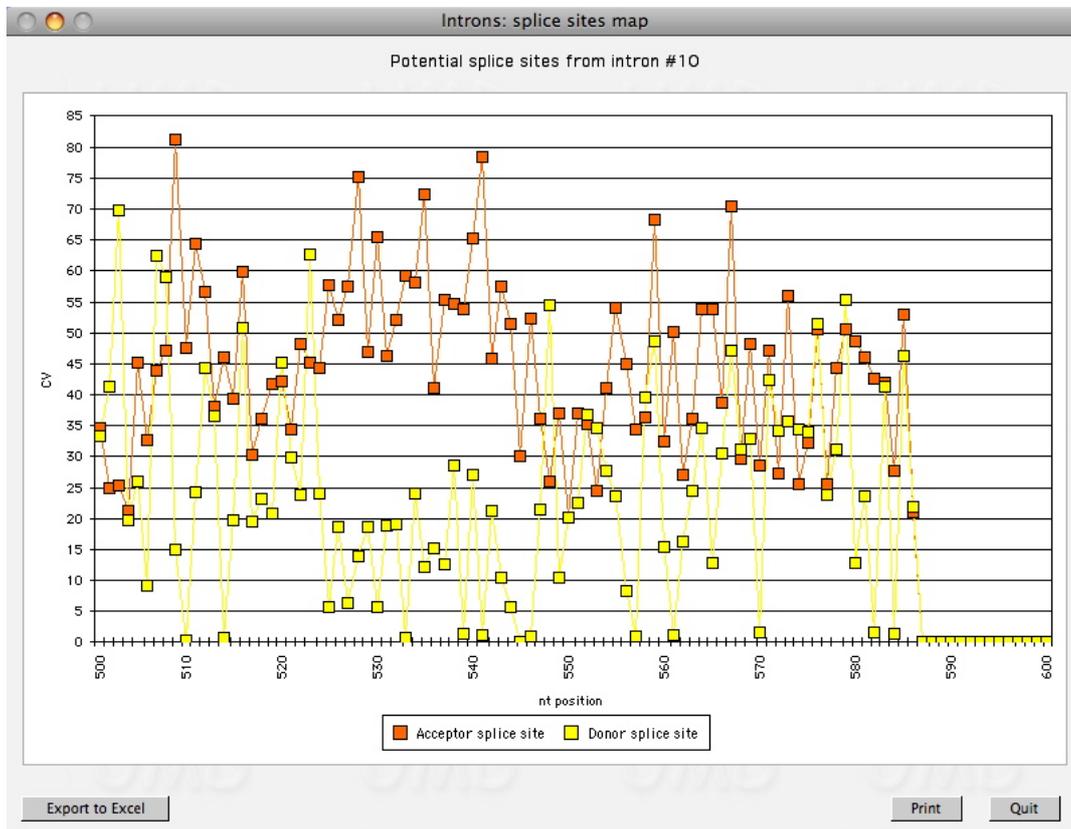
This function can be used only if polymorphic markers have been imported through the "Import polymorphic markers" function from the "File" menu. A graphical display of these markers is displayed:



A zoom option allow to get more details on a specific region of the genome. A zoom of x2000 will allow to click on the blue lines and display the corresponding information about SNPs and ESTs.

#### 4) Splice sites map

This function is used to search for potential acceptor and donor splice sites within a non-coding sequence. It is usually used when an intronic mutation is predicted to activate a cryptic splice site. The user can then search for a complementary splice site in order to predict the potential cryptic exon that could be inserted in the cDNA. Remember that potentially active splice sites have aCV >70 (for more information, refer to the Human



Splicing Finder tool [Nucleic Acid Research, 2009, April](#)).

#### 5) ESE map

This function is used to search for potential auxiliary splicing signals using the ESE-Finder matrices. For a more sophisticated analysis, we recommend that you use the [HSF tool](#).

#### 6) Branch point [one intron]

This function is used to search for the branch point of an intron. The user selects one intron and the UMD® software will search for branch point (for more information, refer to the Human Splicing Finder tool [Nucleic Acid Research, 2009, April](#)).



## VIII-k- The "Modules" menu

This menu contains 3 functions related to the definition and analysis of repeated modules within a single gene. It requires that structural domains have been defined in order to be used.

Define module  
View-Modify module  
Module Analysis

### 1) Define module

This function allows the alignment of multiple modules from the same protein. The user needs to convert the raw data into a consensus module.

The first step is to perform the alignment of the various modules using one of the numerous available tools.

The conserved and variant positions are then defined (step#2).

In the third step, the user will indicate to the software to position of each residue from a module. For example the module #3 has the "Q" conserved residue at position 1, a variant residue at position 2, the "E, E, Y, Y, M" conserved residues at positions 3 to 7, a variant residue at position 8...

Note that position 8 (variant residue) can correspond to a single residue (module #3) or 2

The screenshot shows a window titled "Modules definition" with a blue header bar. Below the header is a blue box with instructions: "This function allows to define the alignment of multiple modules. You need to enter each AA position of conserved residues from various modules in order to allow the analysis of the mutation distribution over the selected modules." The interface is divided into three steps:

**Step #1: Raw data**

```
Module #1 : Q S E E Y Y M Q A L I Q Q H
Module #2 : Q A E E Y Y M A M Q A A L I Q S H
Module #3 : Q S E E Y Y M A Q A L I Q S A H
```

**Step #2: Alignment of the various modules**

```
Module #1 : Q S E E Y Y M - - Q A - L I Q Q - H (10-23)
Module #2 : Q A E E Y Y M A M Q A A L I Q S - H (30-46)
Module #3 : Q S E E Y Y M A - Q A - L I Q S A H (60-75)
```

**Step #3: Definition of the AA positions for the conserved residues**

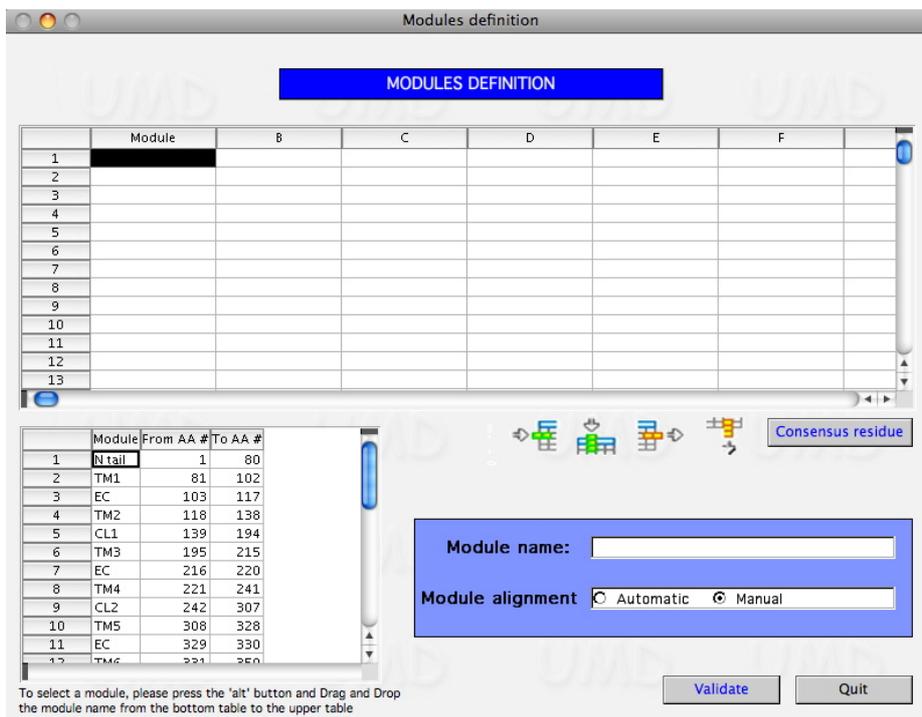
```
Consensus module : Q E E Y Y M Q A L I Q H
Module #1 : 1 3 4 5 6 7 8 9 10 11 12 14
Module #2 : 30 31 33 34 35 36 39 40 42 43 44 46
Module #3 : 60 62 63 64 65 66 68 69 70 71 72 75
```

At the bottom right, there are "Continue" and "Quit" buttons.

residues (module #2) or event be absent (module #1).

In the second screen, two tables are available to create the consensus module. The bottom table contains all available structural domains. They can be selected and drop into the upper table (Alt+mouse). Each column should contain either a conserved residue or a \* symbol to indicate that it is a non-conserved region.

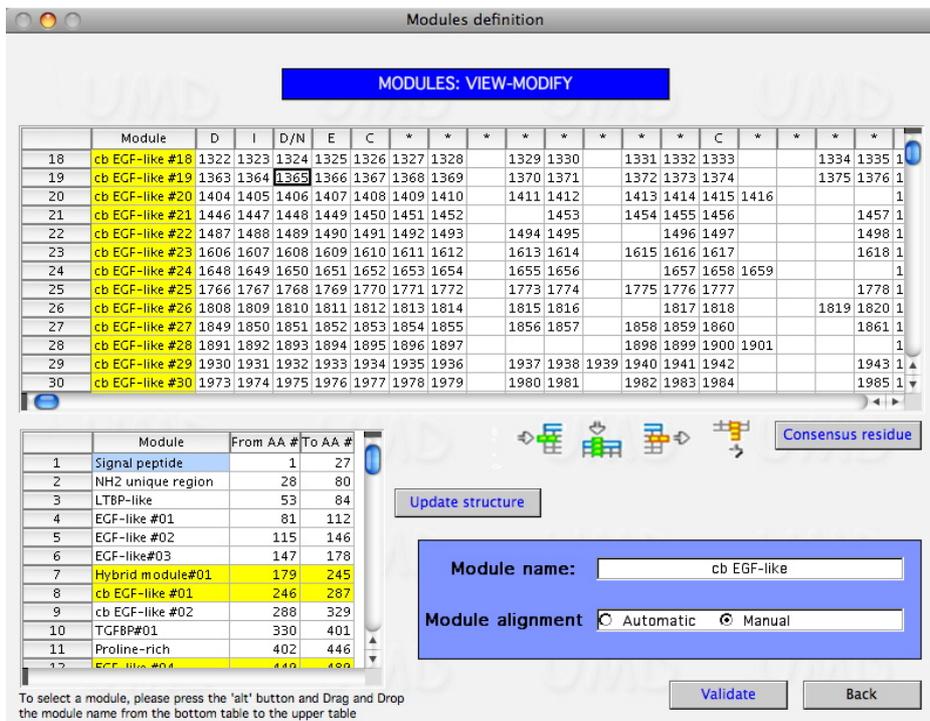
Use the "Consensus residue" button to annotate each column. Enter the position of the residue within the selected module in the corresponding cell (see the example from the "View-Modify modules" function).



Each module alignment should have a unique name "Module name".

## 2) View-modify module

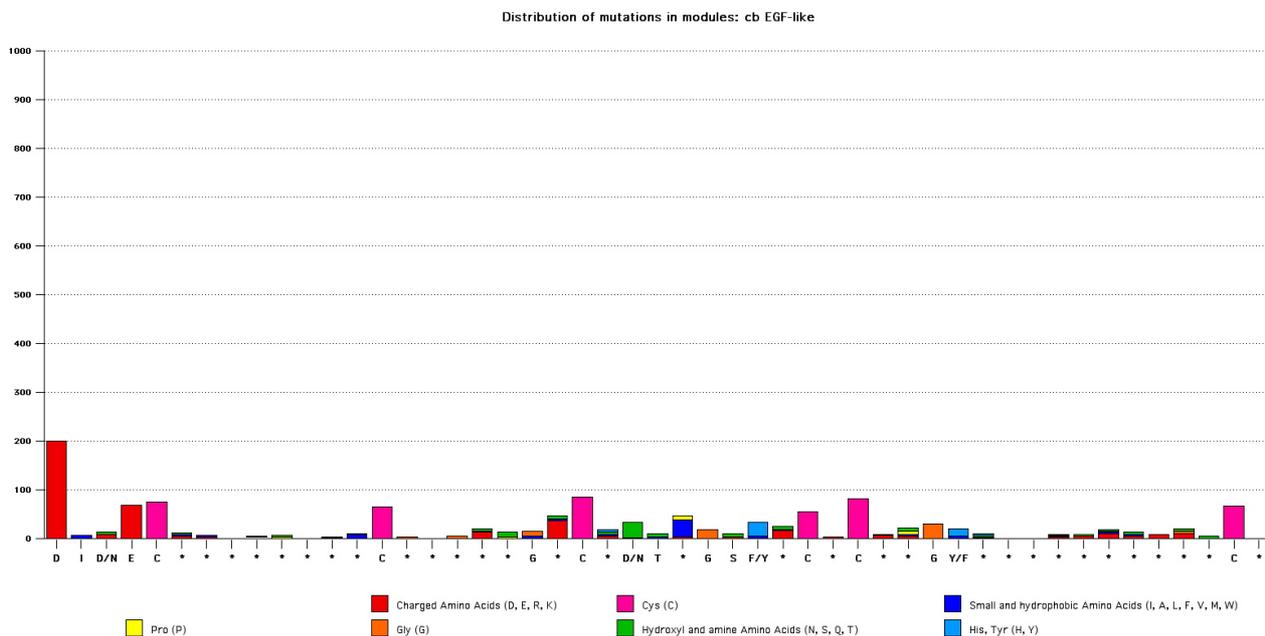
This function allows the modification of created modules alignments.



### 3) Module analysis

This function is used to study the distribution of mutations localized in different modules that have been aligned in a "Module" (see previous functions).

The user can select a subset of records prior to launch this analysis.

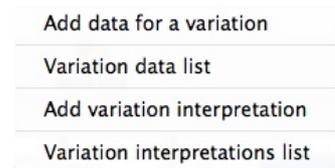


Amino acid with similar properties are grouped in one of the 7 groups and are displayed with a specific color code. The number of mutations affecting either a conserved residue or a variant regions are combined from all modules and displayed in the consensus module.

Note that in this example, mutations affect conserved Cys (C) residues and the first Asp (D) residue.

## VIII-I- The "Variation" menu

This menu contains 4 functions related to the annotation of variations.



### 1) Add data for a variation

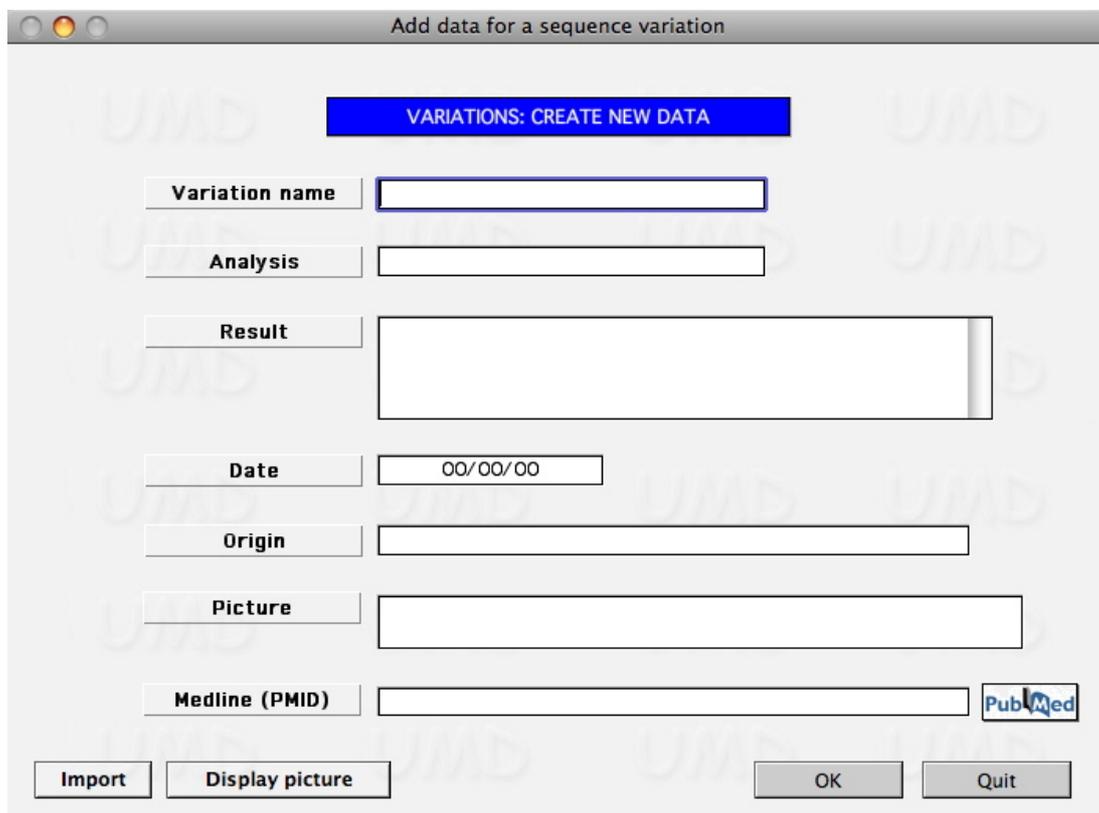
This option is used to collect evidences associated to the effect of a specific variation. Multiple evidences can then be collected for a single variation.

The variation should be described with the c. or p. international nomenclature systems in order to keep a coherence with data generated in the "Mutation" table.

The method of analysis can be stored in the "Analysis" field and the corresponding results in the "Result" text field.

In addition the date and the origin of the data can also be collected. If these data have been published, the PubMed ID can be store in the "Medline (PMID)" field and a direct link to the abstract is available via the "PubMed" button.

Finally a picture can be associated to this record using the "Import" and the "Display picture"



buttons.

### 2) Variation data list

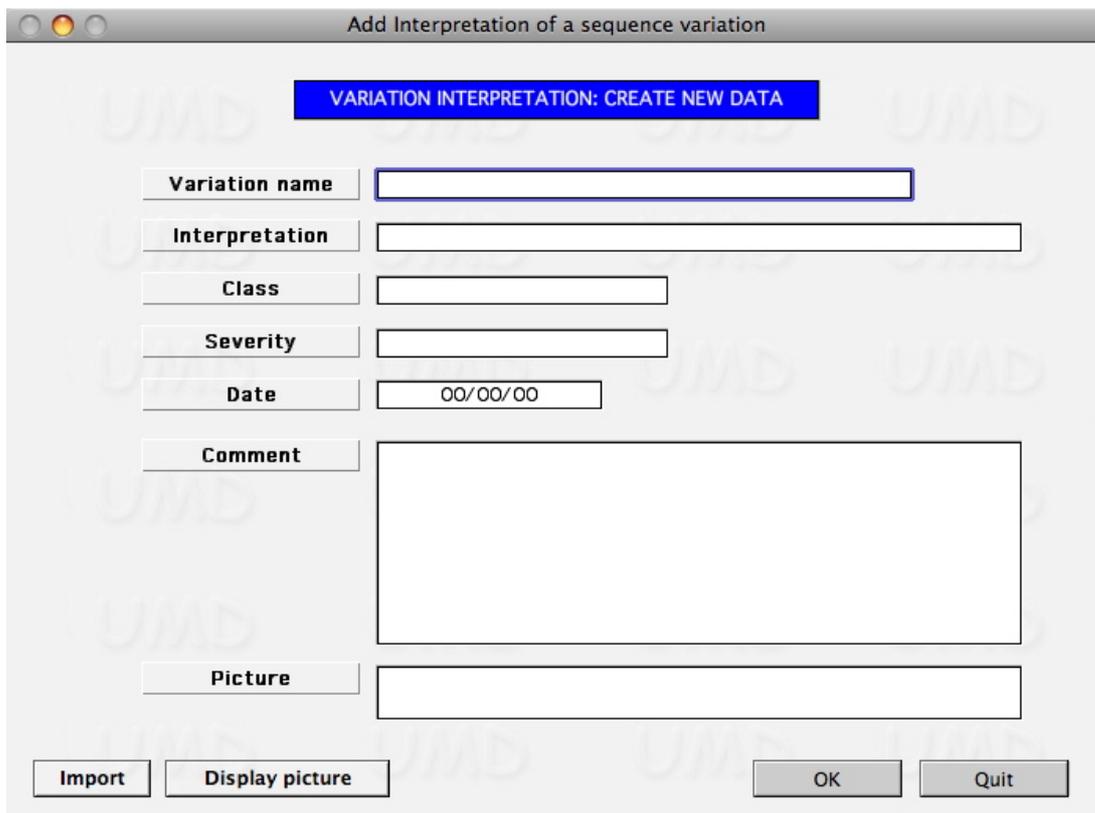
This option is used to view or modify the Variation records.

### 3) Add variation interpretation

While the previous "Variation data" table philosophy is to collect all evidences related to a specific variation, the philosophy of the "Variation interpretation" table is to conclude about the status of this variant. Therefore only one record per variant should be included in this table.

This table includes the following fields:

- The "*Variation name*" should be described with the c. or p. international nomenclature systems in order to keep a coherence with data generated in the "*Mutation*" table;
- The "*Interpretation*" should be done using the international consensus: "*Probably pathogenous*", "*Probably non pathogenous*", "*Strong evidence for being pathogenous*", "*Strong evidence for being non pathogenous*" and "*Unclassified variant*";
- The "*Class*" can be used when such system has been defined for a specific protein;
- The "*Severity*" can be used when such system has been defined for a specific protein in relation to a specific disease;
- The "*Date*" is used to store the date when the interpretation (consensus) was performed (reached);
- The "*Comment*" text field is used to describe specific features about this variation;



- The "*Picture*" field is used to store the picture file pathway. The "Import" and "Display picture" buttons are used to download and visualized this picture.

### 4) Variation interpretations list

This option is used to view or modify the Variation interpretations.

## IX- CONCLUSION

The UMD<sup>®</sup> software was designed by Dr. Christophe Bérout more than 17 years ago in order to fill the gap between core databases such as GenBank and the day-to-day benchwork leading to discoveries of mutations responsible for human genetic diseases. Thierry Soussi and Gwenaëlle Collod-Bérout played a major role in the evolution of this system since these early days and the creation of the first UMD-LSDBs.

UMD is now maintained and developed at the INSERM UMR827 directed by Pr. Mireille Claustres. A bioinformatics team directed by Christophe Bérout is regularly developing new tools, Dr. Dalil Hamroun playing a major role in this process.

This whole adventure has been made possible by the numerous collaborators and users who have chosen to use our tool and develop UMD-LSDBs. Because they have new questions and problems related to their genes, we are frequently confronted to research challenges that can benefit from large scale analysis and bioinformatics.

The emergence of new genotype-based therapies has led to the involvement of clinicians and the creation of new functions to increase our understanding of the natural history of diseases, the search for genotype-phenotype associations and the possibility to select group of patients for clinical trials.

Because the field of Locus Specific DataBases and Knowledgebases (to which belongs the UMD system) is rapidly growing, we are facing new challenges to better understand the relationships between mutations and human diseases and ultimately move towards the predictive medicine.

Finally, we would like to give special thanks to:

The **AFM** (Association Française contre les **Myopathies**) that has supported this work over the years;

The European Community Sixth Framework Program FP6 under grant agreement number 036825 - the **TREAT-NMD** Network of Excellence;

The European Community Seventh Framework Program FP7/2007-2013 under grant agreement number 200754- the **GEN2PHEN** project.