EuroGentest Unit 5

New technologies

Decision support for molecular diagnostic laboratories using Interactive Biosoftware Alamut v1.2

November 2007- February 2008

Prepared by:

NGRL Manchester

As part of the activities of:

EuroGentest Unit 5

within the work package WP5.4.

EuroGentest Project Co-ordinator: Professsor Jean-Jaques Cassiman
K.U.Leuven
Center for Human Genetics/ Centrum Menselijke Erfelijkheid
Gasthuisberg O&N
Herestraat 49
Box 602
3000 Leuven
Belgium

Table of contents

Summary

Alamut is a decision-support software application developed by Interactive Biosoftware for mutation diagnostics in medical molecular genetics. It is a client-server application that integrates genetic information from different sources to describe variants using HGVS nomenclature and to help interpret their pathogenic status. We evaluated Alamut in four areas: its user interface and usability; the suitability of its data sources; its applicability to diagnostic testing; and its validity and accuracy. Laboratories found the software intuitive and easy to use, and well adapted to diagnostic testing. Over 400 variants from 14 genes were tested. Alamut was found to provide accurate and high quality nomenclature, often better than manually generated nomenclature. The only problems found were in the protein-level nomenclature for a few complex variants, and with some cases of the interpretation of data from external data sources. Genomic and HGVS coordinates of genes were found to be in complete agreement with other data sources. We conclude that, with the usual professional care and critical assessment of the data it provides, Alamut is potentially a great asset to the diagnostic laboratory.

Acknowledgements

This evaluation was undertaken by Bert Bakker at LUMC, Jo Campbell at Guy's Hospital London, Andrew Devereau at NGRL Manchester, UK, and Jana Camajova at Prague. Thanks to Andre Blavier at interactive Biosoftware for making Alamut available for evaluation.

Introduction

Background

There are many online sources of data available covering a whole range of aspects of genetics and genomics. Diagnostic molecular laboratories have particular needs for data to help them in their work: having found a variant in a patient's sample they need to describe it accurately and then assess whether it is a pathogenic variant by looking at its effect on the gene and seeing if it has been found and reported before. The Clinical Molecular Genetics Society (CMGS) in the UK has recently published best practice guidelines to help standardise this process (www.cmgs.org). Laboratories face several problems though. The recommended nomenclature for the description of variants is that described by the Human Genome Variation Society (HGVS) (www.hgvs.org/rec) but this is not widespread in existing databases and it can be difficult to describe a variant correctly. Assessing the effect of a variant will often mean looking at many different sources of data, often using different nomenclature systems. And using in-silico tools such as splice-site finders or missense analysis tools like SIFT and PolyPhen can require assembly of input data such as species alignments and presentation of variants in certain formats such as FASTA. These require laboratory staff to visit many different applications, to translate coordinates and to assemble data, which is time consuming, open to error and difficult to standardise.

Alamut is a decision-support software application developed by Interactive Biosoftware (IBS) for mutation diagnostics in medical molecular genetics. It is designed to address the problems outlined above and thereby help interpret mutations by bringing together relevant molecular data and prediction methods inside a consistent and convenient environment.

Description of tool

Alamut is a client/server application, i.e. is has a client application that runs on the users' PC and which provides the user interface, but which connects via the internet to a central server which supplies it with data (Figure 1). The server is maintained by IBS and is regularly updated. Alamut obtains annotated genome and transcript information from Ensembl (www.ensembl.org) then integrates data from sources such as SwissProt, dbSNP and UCSC into one display. Variants and their effects can be modelled by editing the transcripts. Prediction tools such as splice-site tools can be used to gain further information about pathogenicity. Links are provided to other browsers and data sources such as publication and locus-specific databases. Figure 2 shows the user interface, for more information see http://www.interactive-biosoftware.com/index.html.

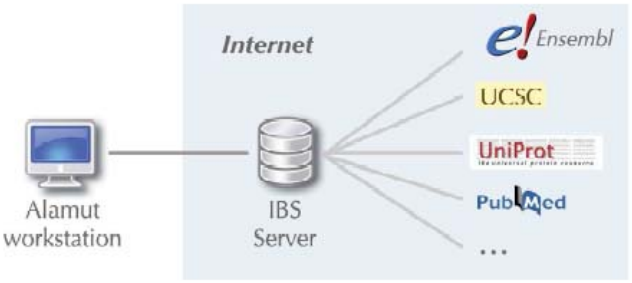

Figure 1 Alamut is a client/server application

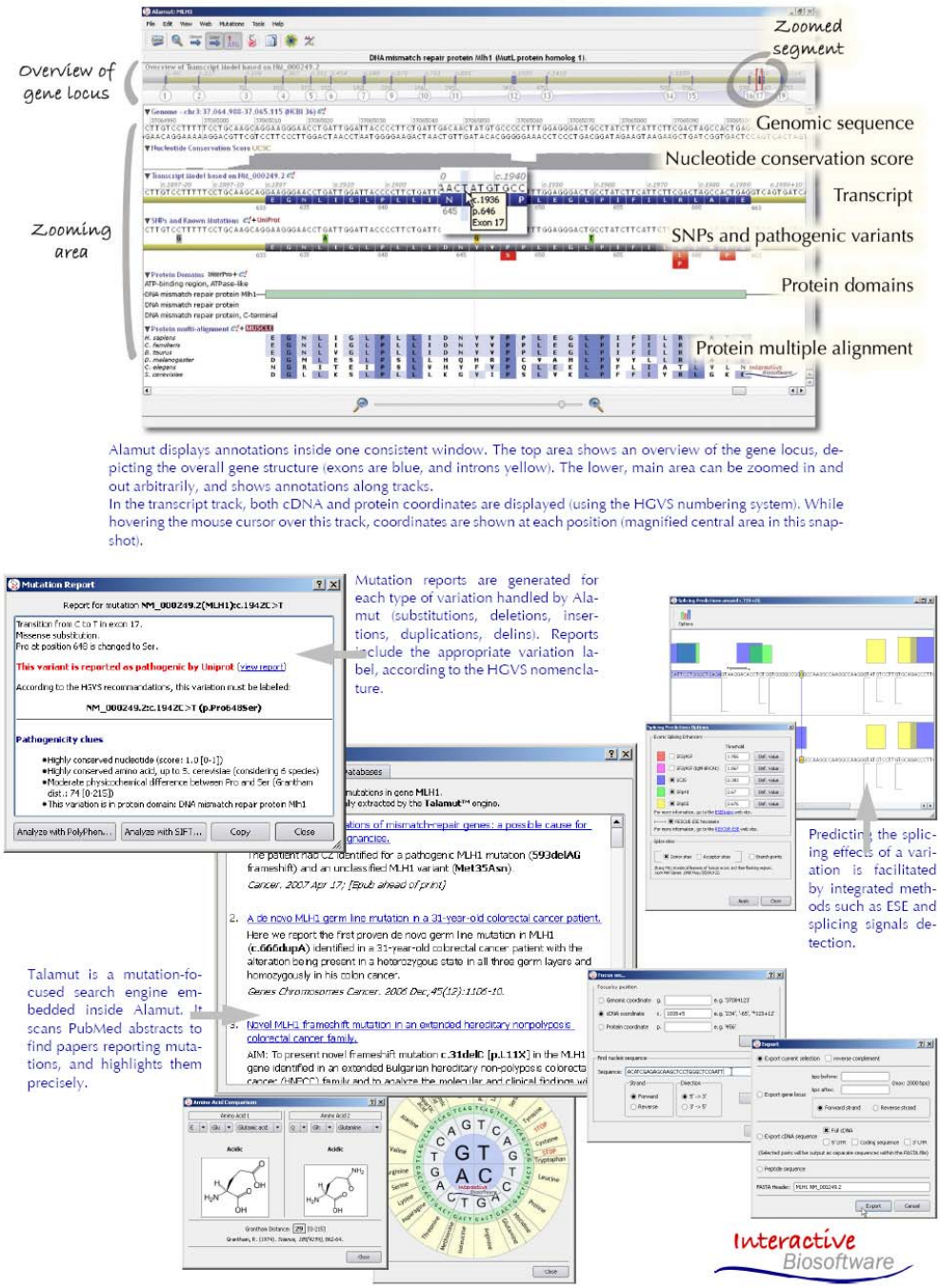

Figure 2 **This page from the Alamut brochure shows the user interface and available information**

Evaluation

Aim

The aim of this evaluation was to test the suitability of the software for use in diagnostic molecular genetic testing. This included the suitability of the user interface, ease of installation and use, the suitability of the data sources it relies upon, an assessment of where it fits into the diagnostic testing laboratory and a test of how accurate its outputs are. When adopting any new technique or tool a diagnostic laboratory needs to be sure that it will not introduce errors into their results and interpretations, and they need to know how to apply the tool and what it is for. This assessment was designed to address these concerns, and not as a general assessment of how well the package is designed and implemented unless this impacted on these concerns.

Methods

The software was assessment in four areas:

a. Its user interface and usability

b. The suitability of its data sources

c. It applicability to diagnostic testing

d. Its validity and accuracy

Evaluation of the tool was undertaken by four centres in order to gain as broad an experience as possible. The centres tested a set of genes including both 'common' and 'rare' genes and evaluated a., c. and d. by firstly installing and learning to use the tool, by using the tool within their and their colleagues' normal work, and then by comparing the output of the tool to sets of variant data for each of their genes. In the latter cases they looked at the exon numbering, variant location and nomenclature, and variant reports including splicing effect predictions, links to other tools and data sources and protein domain data.

The genes tested and the centres that tested them are shown in Table 1. The CF variants tested by NGRL Manchester were the set used for laboratory quality assessment by UKNEQAS (www.ukneqas.org.uk). The ability of Alamut to generate the same base and protein level HGVS nomenclature as those agreed by UKNEQAS was tested.

Table 1 Evaluation centres and the genes tested

Laboratory	Gene	Transcript used by Alamut	Transcript used by lab	No of variants tested
London – Guy's and St Thomas'	BRCA1	NM_007294.2	U14680.1	28
	BRCA2	NM_00059.3	U43746.1	32
	DMD	NM_004006.1		31
	COL4AS	NM_000495.3		21
	SPAST	NM_014946.3		21
	LAMA2	NM_000426.3		60
Prague – CF Centre	CF	NM_000492.3		34
Leiden - LUMC-LDGA	DMD	NM_004006.1		27
	BRCA1	NM_007294.2		30
	BRCA2	NM_00059.3		19
	NOTCH3	NM_000435.2		38
	DAX1	NM_000475.3		10
	EXT1	NM_000127.2		30
	EXT2	NM_000401.2		28
Manchester - NGRL	CF	NM_000492.3		38

In addition NGRL Manchester in the UK evaluated the data sources and versions of data used by the programme, ensuring that links were correct and that displayed data such as genomic sequences matched original and alternative data sources.

NGRL Manchester also tested validity further by looking at the coordinate outputs of Alamut for the five genes tested at Guy's hospital. Genomic coordinates were checked by finding the start and end of the coding sequence and an arbitrary mid point for each gene using Alamut and comparing these to the same genes displayed in the UCSC (genome.ucsc.edu) and Ensembl browsers. HGVS coordinates were tested by finding the start and end coordinates of the first and last exons and three other exons within each gene using Alamut and comparing these with HGVS coordinates given by the NGRL Universal Browser (ngrl.man.ac.uk/Browser). Note that Alamut uses Genbank transcripts whereas the Universal Browser uses Ensembl transcripts. Specific links to other data sources were tested by finding known SNPs and variants displayed by Alamut within the tested genes, following the links and ensuring that they connect to the correct information.

Results

User interface and usability

Installation of Alamut versions 1.1 to 1.3 was straightforward at Manchester. Manchester's hospital network is protected by a proxy server, in common with many hospital systems, but Alamut is able to take this into account. This worked well though it does require knowledge of the correct server settings: as we had experience of this from previous work this was not difficult but this could require support for some users. At Guy's hospital in London the software was installed without difficultly but did not work on the hospital network – this was thought to be due to the hospital firewall. However an alternative network was available which allowed the software to be used.

In Leiden, The Netherlands, Alamut version 1.2 was tested by the Laboratory for Diagnostic Genome Analysis (LDGA). Downloading of the software through the firewall was not possible, but Alamut was successfully installed on over 20 PCs from a CD, and accessing the Alamut server through the hospital network and firewall gave no problems.

All laboratories found Alamut intuitive and easy to use, with no training required, and found that it saves scientists' time.

Data sources

Table 2 summarises the data sources used by Alamut. All are well known sources used within the diagnostic community and are maintained and hosted by well-established organizations. No broken or incorrect links were found. Some issues with data taken from SwissProt and dbSNP were found which are discussed below. It is worth noting the importance of observing versions numbers of data sources: data sources such as dbSNP are regularly re-built but this can take time to propagate to the linked resources such as Alamut and Ensembl. This could lead to inconsistencies which users need to be aware of.

Table 2 Data sources used by Alamut. See the Glossary for explanation of acronyms.

Data source	Developer/Host	Used for	Tool type	Version used	Linked to sources	Comments
UCSC genome browser	Genome Bioinformatics Group at University of California, Santa Cruz	Genomic sequence	Genome database	NCBI Human v36 (March 2006)	NCBI	Used to access the human genome assembly provided by NCBI
		Evolutionary conservation scores	Multiple alignment and conservation scores			
		Transcript information			NCBI RefSeq	
Ensembl	EMBL - EBI and the Wellcome Trust Sanger Institute	SNP data	Genome database	47 (October 2007)	dbSNP version 127	Used to access NCBI's dbSNP data
		Protein information	Genome database	47 (October 2007)	InterPro version 16.1	Used to access InterPro data
		Genome browser	Genome database	47 (October 2007)		Linked browser
dbSNP	NCBI	SNP information	Core variation database	127 (March 2007) via Ensembl		dbSNP was re-built to 128 in October 2007

Data source	Developer/Host	Used for	Tool type	Version used	Linked to sources	Comments
InterPro	EBI	Protein domain information	Protein database	16.1 (October 2007) via Ensembl		
NCBI RefSeq database	NCBI	Transcript information	Reference sequence database	25 (Sept. 2007) via UCSC		
SwissProt	SIB, EBI	Missense variant information	Core protein database	54.5 (Nov 2007)		
MEDLINE	NLM, COS	Publication information	Bibliographic database			Updated regularly during the week. Searches are done through the Talamut engine.
HGVS LSDB list	HGVS	LSDB database links	LSDB catalogue	13 March 2007		
NCBI browser	NCBI	Genome browsing	Genome browser			Linked browser
HGNC database	HGNC	Gene information	Gene database			
Google search tool	Google	Variant information	Web search engine			Variant names are used to generate a search across any indexed web site

Applicability to diagnostic testing

Guy's noted that checking the potential pathogenic significance of sequence variants is a frequent task and that Alamut provided a rapid and easy way of accessing a large amount of information.

The staff members at Leiden are used to determining the proper nomenclature and pathogenic status in relation to a disorder (potential pathogenic or not) for each mutation or variant detected. The amount of work involved depends on several points such as the gene, the type of mutations involved and the available information on the internet. Depending on how often the work has to be done, clinical scientists have developed their own set of special links, alignments, and datasets to ease the job. However even the well trained and well-organised staff members found that they saved time by using Alamut.

Groups in the UK, The Netherlands and Ireland collaborated in 2007 to produce best practice guidelines for the interpretation of unclassified variants (www.cmgs.org.uk). These guidelines cover the use of variant nomenclature, online data sources and interpretation tools: Alamut fits very closely into these guidelines.

It may therefore be concluded that Alamut is highly applicable to diagnostic testing.

Validity and accuracy

Of the 38 CF variants tested at Manchester only one issue was found, this was with the variant c.3717+10kBC>T which is not accurately characterised in HGVS nomenclature and therefore cannot be located on the sequence displayed by Alamut. It is also worth noting that Alamut provides a more complete description of frameshift consequences than is often used, e.g. c.948delT is predicted by Alamut as p.Phe316LeufsX12 rather than p.Phe316fs. Alamut can therefore be said to provide higher quality variant descriptions.

Of the 193 variants tested at Guy's hospital only two gave rise to nomenclature issues, and these were in the predicted protein effects rather than the base level nomenclature – these are described below. Furthermore, Alamut discovered errors in the test set nomenclature, at both protein and base level, in three cases. Other members of the laboratory looked at variants in genes PS1, SOD1, RYR1 and SMN1 with no reported problems. Appendix A contains a full list of the variants tested.

The first nomenclature issue was with the Dystrophin variant c.1218_1245delins AAGCTGATTCTACAATTGG, p.Ile407fs, which Alamut predicted as p.Ile407_Met3685delinsSer.

Table 3. The transcriptional effect of Dystrophin variant p.1218_1245delins AAGCTGATTCTA CAATTGG, p.Ile407fs; deleted bases are shown in blue, inserted bases in yellow.

1213											
AAG	CTG	ATT	GGA	ACA	GGA	AAA	TTA	TCA	GAA	GAT	GAA
Lys	Leu	Ile	Gly	Thr	Gly	Lys	Leu	Ser	Glu	Asp	Glu
405											

1213											
AAG	CTA	AGC	TGA	TTC	TAC	AAT	TGG	GAA			
Lys	Leu	Ser	X								
405											

The effect of this variant is shown in Table 3, which shows that the Alamut prediction is not inaccurate but that it does not conform to the HGVS guidelines.

The other issue was in LAMA2 for the variant c.5550_5562+8del21insG. This variant changes the end of an exon and gives rise to two possible new splice sites in consequence (p.Ser1851fs and p.Asn1850_Asp1854delinsLysTrp). Alamut predicts a single change of p.Asn1850_Asp1854delinsLys.

Splicing predictions for BRCA1 variants tested at Guy's hospital were found to generally agree with the systems normally used (http://violin.genet.sickkids.on.ca/~ali/splicesitescore.html and http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.htm) though scores did not correlate. The splicing prediction diagram and scores are only generated for substitution variants: the deletion c.302-2delA is predicted to alter the acceptor site of the exon and to be very likely to cause an exon skip but no further information is available.

Mutation reports in Alamut include data taken from SwissProt and dbSNP. Some BRCA variants that Alamut shows as pathogenic based on data in SwissProt were found to be annotated differently in SwissProt. For example, Alamut reports that c.9038C>T in BRCA2 is marked as pathogenic in SwissProt, but although SwissProt gives its status as 'disease' (from the options 'disease', 'polymorphism' and 'unclassified') it also has comments which state it is of unknown pathological significance. Alamut also reports it as 'annotated as a possible polymorphism' in dbSNP. NCBI state that dbSNP makes no assumptions about minimum allele frequency, so the presence alone of a variant in its database is not an indication that a variant is pathogenic or non-pathogenic. The Breast Cancer Information Core (BIC, http://research.nhgri.nih.gov/bic/) classifies this particular variant as 'not clinically important'. Similarly, c.2521C>T in BRCA1 is marked by Alamut as pathogenic based on SwissProt data but is annotated as a possible rare polymorphism in that database and is again marked as a possible polymorphism due to presence in dbSNP. It is reported in BIC as of unknown clinical importance. There are two issues highlighted by these examples. Firstly, there appear to be errors in SwissProt for these genes, and Alamut has passed these on to its users. Secondly, as data sources like SwissProt and dbSNP have different aims, and categorise their data differently, it is possible for Alamut to interpret data from these sources in ways that are not valid. Alamut does not attempt to provide an overall conclusion for particular variants but sensibly provides all the evidence it can with appropriate links. To avoid the issues discussed here Alamut needs to concentrate on this approach and avoid interpreting or summarising data so that users will not inspect data at its source. The problem with the SwissProt data is out of Alamut's control, but as the quality of Alamut's results are dependent on the quality of its data sources users need to be made aware of the potential for errors or omissions and be encouraged to inspect data at source. Information for users about the data sources used by Alamut, such as that in Table 2, and how the data are used, might be a useful addition. It is worth noting that it is the job of all users to alert data curators to errors in data such as that found in SwissProt.

Only one of 34 variants tested at the CF Centre in Prague contained an error, which was the incorrect labeling of a TG deletion in a short TG repeated tract.

Of the 182 variants tested at Leiden the nomenclature generated by Alamut was found to be superior to the in-house, often hand generated, nomenclature. No mistakes by Alamut were detected and some mistakes in test data were able to be corrected. Other members of the laboratory looked at variants in genes APC, NSD1, MLH1, MSH2, MSH6, MUTYH and many more, with no reported problems apart from known difficult areas such as the CDKN2A gene with two overlapping (p16 and

p14ARF) transcripts: for more info see (http://www.lumc.nl/4080/DNA/FAMMM%20variants.pdf).

The only point where care was found to be necessary was with the predicted protein nomenclature (which often is much more detailed than manual predictions) where splicing still needs to be investigated separately. Examples are:

- the first or last base of an exon, such as dystrophin c.358delG p.Val120_Thr3684delinsSerLys (first base of exon 6) which leads to a splice defect;

- splice effects of deeply intronic variants, such as dystrophin c.9225 -647A>G, which are always difficult to predict and for which RNA should be analysed;

- exonic variants, such as a predicted dystrophin stop mutation c.3940 C>T p.Arg1314X, which is proven to influence splicing.

Appendix A contains a full list of the variants tested.

Testing of genomic and HGVS coordinates at Manchester showed no errors or disagreements with the data sources used. Table 4,Table **5**, Table **6** and

Table **7** show the coordinates used in these tests.

Table 4. Genomic coordinates of genes tested at Guy's found using Alamut. The midpoint was chosen arbitrarily and located within the UCSC and Ensembl genome browsers along with the start and end points of the coding sequence. Coordinates found using UCSC and Ensembl agreed with those from Alamut.

Gene	Transcript	Start	Midpoint	End
BRCA1	NM_007294.2	17:38,529,639	38,498,710 (AAC)	38,451,221
BRCA2	NM_000059.3	13:31,788,598	31,810,970 (AAA)	31,870,907
DMD	NM_004006.1	X:33,139,350	32,145,040 (GTC)	31,499,957
COL4A5	NM_033381.1	X:107,570,012	107,721,480 (CCC)	107,826,264
SPAST	NM_014946.3	2:32,142,405	32,194,750 (CAAG)	32,233,069

Table 5. HGVS coordinates of genes tested at Guy's found using Alamut. The start and end coordinates of the first and last coding exons and a number of exons lying within the gene were located using Alamut and with the NGRL Universal Browser*, which used the Ensembl transcripts shown. Alamut used the Genbank transcripts shown in Table 4. No differences were found.

Gene	Transcript	Exon	Start	End
BRCA1	ENST357654	2	1	80
		8	442	547
		11	671	4096
		16	4676	4986
		24	5468	5592
BRCA2	ENST380152	2	1	67
		6	476	516
		11	1910	6841
		16	7618	7805
		27	9649	10257
DMD	ENST357033	1	1	31
		6	358	530
		17	1993	2168
		46	6615	6762
		55	8028	8217
		79	11047	11058
COL4A5	ENST328300	1	1	81
		5	277	321
		21	1340	1423
		37	3247	3373
		53	4995	5076
SPAST	ENST315285	1	1	415
		4	587	682
		8	1099	1173
		14	1537	1616
		17	1729	1851

* https://ngrl.man.ac.uk/Browser/

No discrepancies were found between features displayed in Alamut and their linked sources, e.g. SNPs in dbSNP and variant information in SwissProt. One minor issue found was that residues can be described as full names (e.g. Lysine), three letter codes (e.g. Lys) or one letter codes (e.g. K) in different parts of Alamut and in the external sources. Alamut provides an amino acid code wheel and it would be useful there or elsewhere to show all three naming standards together.

Alamut version 1.2 includes a link to Google: this provides a more general search facility and can only be assessed for accuracy by ensuring that it generates the correct search data for the variant chosen. This was found to be the case. One minor issue was that it linked to the French Google interface rather than that of the country in which Alamut is being used.

Table 6. Genomic coordinates of genes tested at Leiden found using Alamut. The midpoint was chosen arbitrarily and located within the UCSC and Ensembl genome browsers along with the start and end points of the coding sequence. Coordinates found using UCSC and Ensembl agreed with those from Alamut.

Gene	Transcript	Start	Midpoint	End
DMD	NM_004006.1	X:33,139,350	32,145,040 (GTC)	31,499,957
BRCA1	NM_007294.2	17:38,529,639	38,498,710 (AAC)	38,451,221
BRCA2	NM_000059.3	13:31,788,598	31,810,970 (AAA)	31,870,907
NOTCH3	NT_011295.8	19:15,174,792	15,152,118 (GTC)	15,129,444
NR0B1 (DAX1)	NM_000475.3	X:30,239,413	30,234,960 (AAC)	30,230,506
EXT-1	NM_000127.2	8:119,195,239	119,037,019 (TCT)	118,878,783
EXT-2	NM_0207122.1	11:44,071,675	44,148.616 (AAT)	44,225,556

Table 7. HGVS coordinates of genes tested at LUMC found using Alamut. The start and end coordinates of the first and last coding exons and a number of exons lying within the gene were located using Alamut and Ensembl (transcripts numbers shown). Alamut used the Genbank transcripts shown in Table 6. No differences were found.

Gene	Transcript	Exon	Start	End
BRCA1	ENST357654	2	1	80
		8	442	547
		11	671	4096
		16	4676	4986
		24	5468	5592
BRCA2	ENST380152	2	1	67
		6	476	516
		11	1910	6841
		16	7618	7805
		27	9649	10257
DMD	ENST357033	1	1	31
		6	358	530
		17	1993	2168
		46	6615	6762
		55	8028	8217
		79	11047	11058
NR0B1 (DAX1)	ENST378970	1	1	1180
		2	1181	1425
NOTCH3	ENST263388	1	1	118
		5	680	802
		10	1493	1606
		23	3719	3837
		33	5914	6966
EXT1	ENST378204	1	1	962
		3	1057	1164
		7	1537	1632
		11	2056	2241
EXT2	ENST343631	2	1	536
		5	744	939
		10	1496	1662
		14	2019	5157

Conclusions

We found that Alamut is easy to use, saves time and is potentially a great asset to the diagnostic laboratory, helping to structure the laboratory's work. Tests for accuracy and validity showed that Alamut provided more accurate and higher quality variant nomenclature than manually generated data, and that it provided accurate genomic and HGVS coordinates. Users found that it identified mistakes in their work during testing. The only nomenclature problems were with predicted protein nomenclature for complex variants in which there are splicing effects. Users need to be aware that there are limitations in such difficult cases.

The other area of concern was with the links and data provided from other sources such as SwissProt and dbSNP, where there is a possible danger of passing on incorrect information or allowing the significance of data to be interpreted incorrectly. We feel that Alamut should ensure that users follow appropriate links and should provide information for the user in order to address this issue. Proper disclaimers are incorporated at several levels of the software.

Alamut should therefore be used with the usual professional care and common sense, with the user critically assessing the information gathered.

Glossary

COS	Community of Science, www.cos.com/
dbSNP	Public domain archive for simple genetic polymorphisms, www.ncbi.nlm.nih.gov/projects/SNP/
EBI	European Bioinformatics Institute, www.ebi.ac.uk/
EMBL	European Molecular Biology Laboratory, www.embl-heidelberg.de/
FASTA	Similarity search application against protein databases
Genbank	Annotated collection of all publicly available DNA sequences, www.ncbi.nlm.nih.gov/Genbank/
HGNC	The Human Genomic Organisation (HUGO) Gene Nomenclature Committee, www.hugo-international.org/comm_genenomenclaturecommittee.php
HGVS	Human Genome Variation Society, www.hgvs.org/
HUGO	Human Genome Organisation, www.hugo-international.org/
IBS	Interactive Biosoftware, www.interactive-biosoftware.com/
Interpro	Database of protein families, www.ebi.ac.uk/interpro/
LDGA	Laboratory for Diagnostic Genome Analysis, www.lumc.nl/4080/DNA/patientcare_DNA.html
LUMC	Leiden University Medical Centre, www.lumc.nl/
MEDLINE	An online publication catalogue, medline.cos.com/
NCBI	National Centre for Biotechnology Information, www.ncbi.nlm.nih.gov/
NGRL	National Genetics Reference Laboratory, Manchester, UK, www.ngrl.org.uk/Manchester
NLM	US National Library of Medicine, www.nlm.nih.gov/
PolyPhen	Predictive tool of possible impact of an amino acid substitution on the structure and function of a human protein, genetics.bwh.harvard.edu/pph/

SIFT	Sorting Intolerant from Tolerant – an in-silico missense variant analysis tool, blocks.fhcrc.org/sift/SIFT.html
SIB	Swiss Institute for Bioinformatics, www.isb-sib.ch/
SwissProt	Annotated protein sequence database, www.psc.edu/general/software/packages/swiss/swiss.html
UCSC	University of California Santa Cruz, www.ucsc.edu/public/

Appendix A Tested Gene Variants

Variants tested at Guy's Hospital

BRCA1

Our nomenclature	Alamut nomenclature	Comments
c.1141A>T p.Lys381X		
c.2071delA; p.Arg691fs		
c.1A>G		
c.3087_3100dup14; p.Asn1034fs		
c.2681_2682delAA p.Lys894fs		
c.5152+1G>T		predicts loss of donor site
c.303T>G p.Tyr101X		
c.302-2delA		predicted to alter acceptor site, but can't see diagram as only available for substitutions.
c.1471C>T; p.Gln491X		
c.134+3A>T		consequence not predictable
c.4524G>A; p. Trp1508X		
c.4065_4068delTCAA p.Asn1355fs		
c.81-1G>C		predicts loss of acceptor site
c.5266dupC p.Gln1756fs		
c.3331_3334delCAAG p.Gln1111fs		
c.3119G>A p.Ser1040Asn		
c.5576C>G; p.Pro1859Arg		
c.5468-40T>A		
c.557C>A p.Ser186Tyr		
c.3418A>G, p.Ser1140Gly		
c.1486C>T; p.Arg496Cys		
c.-41_-40dupAT	c.-19-22_-19-21dup	Our mistake
c.2521C>T p.Arg841Trp		
c.3848A>G p.His1283Arg		
c.5096G>A p.Arg1699Gln		
c.53T>C p.Met18Thr		
c.3156T>A p.Asn1052Lys		
c.3640G>A p.Glu1214Lys		

BRCA2

Our nomenclature	Comments
c.8578_8579delAA p.Lys2860fs	incorrect protein nomenclature, but corrected by v1.2
c.5946delT p.Ser1982fs	
c.2330dupA; p.Asp777fs	
c.9157delG p.Glu3053fs	incorrect protein nomenclature, but corrected by v1.2
c.4478_4481delAAAG p.Glu1493fs	
c.5909C>A p.Ser1970X	
c.4965C>G; p.Tyr1655X	
c.5279C>G; p.Ser1760X	
c.6275_6276delTT; p.Leu2092fs	
c.6698delC; p.Ala2233fs	
c.8067delT; p.Cys2689fs	incorrect protein nomenclature, but corrected by v1.2
c.2514dupA; p.Tyr839fs	
c.9699_9702delTATG p.Cys3233fs	incorrect protein nomenclature, but corrected by v1.2
c.2606C>G p.Ser869X	
c.7480C>T p.Arg2494X	
c.8247_8248delGA p.Lys2750fs	incorrect protein nomenclature, but corrected by v1.2
c.8398C>T; p.Pro2800Ser	
c.9646C>T; p.Leu3216Leu	
c.3055C>G p.Leu1019Val	
c.7182A>G p.Arg2394Arg	
c.3516G>A; p.Ser1172Ser	
c.2803G>A p.Asp935Asn	
c.7988A>T p.Glu2663Val	
c.4187A>G p.Gln1396Arg	
c.4319A>G p.Lys1440Arg	
c.5455C>T p.Pro1819Ser	
c.7988A>T p.Glu2663Val	
c.68-7T>A	
c.6323G>A; p.Arg2108His	
c.9038C>T p.Thr3013Ile	
c.8168A>C p.Asp2723Ala	
c.1792A>G p.Thr598Ala	

Dystrophin

Our nomenclature	Alamut	Comments
c.34_38delGAAAG; p.Glu12fs		
c.133C>T; pGln45X		
c.464A>G; p.Asn155Ser		
c.709delC; p.Gln237fs		
c.932dupA; p.Asp311fs		
c.1218_1245delinsAAGCTGATTCTACAATTGG; p.Ile407fs	p.Ile407_Met3685delinsSer	see notes in report
c.1555G>T; p.Glu519X		
c.1732A>T;p.Lys578X		
c.2163delG; p.Arg723fs		
c.2436dupG; pIle813fs		
c.2896C>T; p.Gln966X		
c.3365_3366delAG; p.Glu1122fs		
c.3892G>T; p.Gly1298X		
c.4515_4518delTGTG; p.Cys1505X		
c.4918_4919delinsTG; p.Thr1640X		
c.5637G>A; p.Trp1897X		
c.5983delT; p.Tyr1995fs		
c.6554dupT; p.Lys2185fs	c.6554dupT; p.Leu2185fs	our mistake
c.6951delA; p.Ala2317fs	c.6951delA; p.Ala2318fs	our mistake
c.7590_7596del; p.Ile2531fs		
c.7899G>A; p.Trp2633X		
c.8355delG; p.Lys2785fs		
c.8897_8898dupTC; p.Leu2967fs		
c.9210_9214delGCCCT; p.Pro3071fs		
c.9337C>T; p.Arg3113X		
c.9634G>T; p.Glu3212X		
c.9868C>T; p.Gln3290X		
c.10115_10116delTT; p.Phe3372fs		
c.10271C>A; p.Ser3424X		
c.10453dupC; p.Leu3485fs		
c.10922-2A>G		

SPAST

c.85dupC (p.Leu29fs)	
c.131C>T (p.Ser44Leu)	
c.397A>G (p.Ile133Val)	
c.422_425delAGAA (p.Gln141fs)	
c.451_454delAAAG (p.Lys151fs)	
c.447_448insTGAAGAACTGGAATGGTATTG (p.Lys150X)	
c.724_725delAC (p.Thr242X)	
c.937delG (p.Asp313fs)	
c.1095dupT (p.Glu366X)	
c.1104C>A,(p.Phe368Leu)	
c.1245+1G>C	
c.1252G>T (p.Glu418X)	
c.1352delGinsTC (p.Arg451fs)	
c.1453G>A (p.Ala485Thr)	
c.1535_1536+1delAGG (p.Glu512fs or p.Glu512del)	Depends on splice site options, Alamut gave p.Glu512fs
c.1553T>C (p.Leu518Pro)	
c.1635_1636insAA (p.Gly546fs)	
c.1676G>A (p.Gly559Asp)	
c.1728G>A (p.Glu576Glu)	
c.1741C>T (p.Arg581X)	
c.1844C>T (p.Thr615Ile)	

COL4A5

c.1117C>T p.Arg373X
c.465+2T>G
c.3266G>A p.Gly1089Glu
c.3731G>A p.Gly1244Asp
c.3319G>A p.Gly1107Arg
c.3427G>A p.Gly1143Ser
c.1295G>A p.Gly432Glu
c.2351G>A p.Gly784Asp
c.4780_4781delTG p.Trp1594fs
c.3922delC p.Gln1308fs
c.4045delG p.Glu1349fs
c.796C>T p.Arg266X
c.1633G>A p.Gly545Ser
c.5029C>T p.Arg1677X
c.4069G>A p.Gly1357Ser
c.1958G>T p.Gly653Val
c.973G>A p.Gly325Arg
c.4244G>A p.Gly1415Asp
c.2288G>A p.Gly763Glu
c.2509G>A p.Gly837Ser
c.1243G>A p.Gly415Arg

Variants tested at LUMC - LDGA

Dystrophin

Our nomenclature	Alamut nomenclature	Notes
c.897 T>G pTyr299X		
c.8713 C>T p.Arg2905X		
c.9262delA p.3088M fsX1	c.9262del p.Met3088X	
c.843 C>T p.Gln615X		
c.3622 C>T p.Gln1208X		
c.9225 -647 A>G p. N3075fsX2	c.9225-647A>G p.?	To deep intronic to predict p.
c.200dupG p.Gly67GlyfsX21	c.200dupG p.Ser68IlefsX21	Our mistake
c.2466_2467insGTGAGAGA p.Leu823ValfsX25	c.2459_2466dup p.Leu823ValfsX26	
c.3408_3420del p.Thr1136ThrfsX12	c.3408_3420del p.Gln1137CysfsX12	Our mistake
c.358delG p.Val120SerfsX3 of p.?	c.358delG p.Val120_Thr3684delinsSer Lys	First base of exon 6, Splice effect
c.1061 G>A p.Trp354X		
c.5341 A>T p.Lys1781X		
c.5899 C>T p.Arg1967X		
c.3217 G>T p. Glu1073X		
c.3433delG p.Val1145SerfsX	p.Val1145SerfsX8	
c.4996 C>T p.Arg1666X		
c.9897_9898del p.His3299GlnfsX14		
c.5699 T>G p.Leu1900X		
c.3940 C>T p.Arg1314X,	Correctly labeled, as a stop. However this mutation in exon 29 does shows, on splice prediction a loss of an ESE SRp55 site.	We looked at this because we knew that the whole exon was spliced out in the majority of the cells restoring the reading frame: r.3940 c>u, r.3922_4071del
c.5697dupA p.Leu1900IlefsX5		
c.9204-9207delCAAA p.Asn3068_3069LysfsX3088	c.9204-9207delCAAA p.Asn3068LysfsX20	Our mistake
c.9897_9898del p.His3299GlnfsX14		
c.4996 C>T p.Arg1666X		
c.5699 T>G p.Leu1900X		
c.5697dupA p.Leu1900IlefsX5		
9204-9207delCAAA p.Asn3068_3069LysfsX3088		
c.3328TCA>CC p. Cys1040Cysfs-1X1043		

BRCA1

Our nomenclature	Alamut nomenclature	Notes
c.1-7 G>A	c.-7 G>A	Neutral; RNA no effect
c.81-6 T>A		
c135-15_-12delCTTT		
c.213-12A>G		
212+5A>G		
c.288C>T p.Asp96Asp	c.288C>T p.=	
c.302-3C>G		
c.441G>C p.Leu147Phe		
c.736T>G p.Leu246Val		
c.825C>T p.Gly275Gly	c.825C>T p.=	
c.1418A>G p.Asn473Ile		
c.1209_1210insT p.Glu404X	c.1209dup p.Glu404X	
c.1288insA p.Asp430GlufsX6	c.1288dup p.Asp430GlufsX6	
c.1292insT p.Leu431PhefsX5	c.1292dup p.Leu431PhefsX5	
c.1456T>C p.Phe486Leu		
c.1525A>G p.Thr509Ala		
c.1621C>T p.Gln541X		
c.1961delA p.Lys654SerfsX47	c.1961del p.Lys654SerfsX47	
c.2014A>T p.Lys672X		
c.2019delA p.Glu673AspfsX28	c.2019del p.Glu673AspfsX28	
c.3748G>T p.Glu1250X		
c.3820_3821insG p.Val1274GlyfsX13	c.3820dup p.Val1274GlyfsX13	
c.4483delA p.Arg1495GlyfsX10	c.4483del p.Arg1495GlyfsX10	
c.4986+5G>T		Donor site reduced; new donor site +16 bp in: 16bp in RNA
c.5030_5033delCTAA p.Thr1677IlefsX2	c.5030_5033del p.Thr1677IlefsX2	
c.5177_5180delGAAA p.Gln1826ArgfsX7	c.5177_5180del p.Gln1826ArgfsX7	
c.5177_5178delGA p.Arg1726LysfsX5	c.5177_5178del p.Arg1726LysfsX5	
c.5430insG p.Gln1811AlafsX19	c.5430dup p.Gln1811AlafsX19	
c.5407+5G>A	c.5406+5G>A	Our mistake
c.5503C>T p.Arg1835X		
c.5503_5564del62 p.Arg1835ThrfsX24	c.5503_5564del p.Arg1835ThrfsX24	

BRCA2

Our nomenclature	Alamut nomenclature	Notes
c.68-7 T>A		Acceptor site reduced; skip exon3
c.7617+2 T>G		Donor site lost; skip exon 15
c.469_470delAA,p.Lys157fs	c.469_470del p.Lys157ValfsX25	
c.793+1G>A		
c.1153A>T p.Lys385X		
c.3864_3865delTA,p.Asn1288fsX8	c.3864_3865del p.Asn1288LysfsX9	
c.4146_4148delAGA,p.1382delGlu	c.4146_4148del p.Glu1382del	
c.5682C>G, p.Tyr1894X		
c.5722_2723delCT, p.Leu1908fsX	c.5722_2723del p.Leu1908ArgfsX2	
c.6578_6584delAAATTGG p.Glu2193ValfsX11	c.6578_6584del p.Glu2193ValfsX11	
c.8067T>A p.Cys2689X		
c.8754+5G>A		
c.9097dupA, p.Thr3033AsnfsX11	c.9097dup p.Thr3033AsnfsX11	
c.9099_9100delTC, p.Thr3033fsX	c.9099_9100del p.Gln3034ValfsX9	
c.9117G>A, p.Pro3039Pro	c.9117G>A, p.=	
c.9256+28delT	c.9256+28delT	
c.9501+8T>G		
c.9509A>G,p.Asp3170Gly		
c.9672_9673insA p.Tyr3225fsX	c.9672 dup p.Tyr3225IlefsX30	

NOTCH3

Our nomenclature	Alamut nomenclature	Notes
c.206 G>T p.Cys43Phe	c.128G>T p.Cys43Phe	Correction old nomenclature (ATG not at c.1 but at -77)
c.346 C>T p.Arg90Cys	c.268C>T p.Arg90Cys	
c.406C>T p.Arg110Cys	c.328C<T p.Arg110Cys	
c.461G>T p.Cys128Phe	c.338G>T p.Cys128Phe	
c. 475 C>T p.Arg133Cys	c.397C>T p.Arg133Cys	
c.480 C>G p.Cys134Trp	c.402C>G p.Cys134Trp	
c.499 C>T p.Arg141Cys	c.421C>T p.Arg141Cys	
c.509 G>T p.Cys144Phe	c.431G>T p.Cys144Phe	
c.535 C>T p.Arg153Cys	c.457C>T p.Arg153Cys	
c.542 G>A p.Cys155Tyr	c.464G>A p.Cys155Tyr	
c.564 C>G p.Cys162Trp	c.486C>G p.Cys162Trp	
c.585 C>T p.Arg169Cys	c.507C>T p.Arg169Cys	
c.622 C>T p.Arg182Cys	c.544C>T p.Arg182Cys	
c.644 A>G p.Tyr189Cys	c.566A>G p.Tyr189Cys	
c.697 C>T p.Arg207Cys	c.619C>T p.Arg207Cys	
c.743 G>A p.Cys222Tyr	c.665G>A p.Cys222Tyr	
c.749 G>A p.Cys224Tyr	c.671G>A p.Cys224Tyr	
c.777 T>G p.Cys233Trp	c.699T>G p.Cys233Trp	
c.829 T>A p.Cys251Ser	c.751T>A p.Cys251Ser	
c.851 A>G p.Tyr258Cys	c.773A>G p.Tyr258Cys	
c.1072 C>T p.Arg332Cys	c.992C>T p.Arg332Cys	
c.1088 A>G p.Tyr337Cys	c.1010A>G p.Tyr337Cys	
c.1222 G>T p.Gly382Cys	c.1144G>T p382Cys	
c.1381 T>C p.Cys435Arg	c.1303T>C p.Cys435Arg	
c.1415 G>T p.Cys446Phe	c.1337G>T p.Cys446Phe	

NOTCH3 continued…

Our nomenclature	Alamut nomenclature	Notes
c.1423 C>T p.Arg449Cys	c. 1345C>T p.Arg449Cys	
c.1472 A>G p.Tyr465Cys	c.1394A>G p.Tyr465Cys	
c.1708 C>T p.Arg544Cys	c.1630C>T p.Arg544Cys	
c.1723 T>C p.Cys549Arg	c.1645T>C p.Cys549Arg	
c.1750C>T p.Arg558Cys	c.1672C>T p.Arg558Cys	
c.1810 C>T p.Arg578Cys	c.1732C>T p.Arg578Cys	
c.1897 C>T p.Arg607Cys	c.1819C>T p.Arg607Cys	
c.1996 C>T p.Arg640Cys	c.1918C>T p.Arg640Cys	
c.2260 C>T p.Arg728Cys	c. 2182C>T p.Arg728Cys	
c.3094 C>T p.Arg1006Cys	c.3018C>T p.Arg1006Cys	
c.3121 T>C p.Cys1015Arg	c.3043T>C p.Cys1015Arg	
c.3140 A>G p.Tyr1021Cys	c. 3062A>G p.Tyr1021Cys	
c.3304C>T p.Arg1076Cys	c.3226C>T p.Arg1076Cys	

DAX1

Our nomenclature	Alamut nomenclature	Notes
c.129C>A p.Cys43X		
c.404delG p.Cys135fs	c.404del p.Cys135LeufsX129	
c. 573_574insCCGG p.Ala191fs	c.573_574insCCGG p.Ala192ProfsX14	
c.708G>A p.Trp236X		
c.739dupG p.Ala247fs	c.739dup p.Ala247GlyfsX52	
c.1241_1253delACACCAGGATGAC(13 bp) p.His414fsX19	c.1241_1253del p.His414ArgfsX19	
c.1267delC p.His423fsX13		
c.1292delG p.Ser431fsX6		
c.1307T>G p.Leu436Arg		
c.1411T>C p.X471Gln +18aa	c.1411T>C p.X471GlnextX19	

EXT-1

Our nomenclature	Alamut nomenclature	Notes
c.2 T>C p.Met1?	c.2T>C p.?	
c.164G>A p.Trp55X		
c.212T>A, L71X	c.212T>A p.Leu71X	
357_358insAC, P120fsX	c.356_357dup p.Pro120ThrfsX17	
c.524_525delGA p.Arg175fsX13		
c.538_539delAG p.Ser180fsX7	c.538_539del p.Leu181ProfsX7	Our mistake
c.572delT p.Leu191X	c.572del p.Leu191X	
c.588T>A p.Tyr196X		
c.640_659del p.Ala214HisfsX4		
c.697 del T p.Ser232LeufsX	c.697del p.Ser233LeufsX19	
c.699T>C p.Ser233Ser	c.699T>C p.=	
c.739G>T p.Glu247X		
c.771 ins C p.Pro257SerfsX	c.773dup p.Leu259SerfsX30	Our mistake
c.798 ins T p.Lys266GlnfsX	c.798_799insT p.Lys267X	
c.812_813insCT,p.Tyr271fsX	c.812_813insCT p.Leu272SerfsX2	
c.958G>T p.Glu320X		
c.1018C>T p.Arg340Cys		
c.1019G>T p.Arg340Leu		
c.1027G>A p.Gly343Arg		
c.1056+1G>A		
c.1152delA,p.Arg384fsX	c.1152del p.Arg384SerfsX19	
c.1215_1218delACAG p.Arg405fsX18	c.1215_1218del p.Arg405SerfsX19	
c. 1225delCA,Q409fsX419	c.1225_1226del p.Gln409IlefsX11	
c.1410delT p.Leu472X	c.1410delT p.Asn471IlefsX2	
c.1469delT p.Leu490fsX8	c.1469del p.Leu490ArgfsX9	
c.1469insC L490fsX493	c.1469_1470insC p.Val491GlyfsX30	
c.1537 -14C>T		
c.1776C>A p.Tyr592X		
c.1784_1785delGC,p.Arg595fsX	c.1784_1785delG p.Arg595GlnfsX6	
c.1883+1G>C		

EXT-2

Our nomenclature	Alamut nomenclature	Notes
c.28C>A,p.Arg10Arg	c.28C>A p.=	
c.201insC,P67fsX70	c.201_202insC p.Val68ArgfsX3	
c.445_454 del CGGGCCTGTC p.Arg138CysfsX	c.445_454del p.Arg149CysfsX118	Our mistake
c.536+1G>A		
IVS2+1g>a	c.-87+1G>A	
c.519G>C,A173A	c.519G>C p.=	
c.626+1delG	c.626+1del	
IVS3+1delG	c.-31+1del	
IVS3+1g>a	c.-31+1G>A	
c.544C>T,R182X	c.544C>T p.Arg182X	
IVS4+1g>a	c.536+1G>A	
IVS4+1g>t	c.536+1G>T	
c.659G>A, W220X	c.659G>A p.Trp220X	
c.710 C>T p.Ser237Leu		
c.764T>C, L255P	c.764T>C p.Leu255Pro	
c.919 G>A p.Asp307Asn		
c.1112C>G,S371X	c.1112C>G p.Ser371X	
IVS6-2a>g	c.744-2A>C	
IVS7+1g>a	c.939+1G>A	
c.1185delC,p.Trp396fsX41	c.1185delC p.Trp396GlyfsX40	
c.1306 -23 C>T		
c.1588G>A p.Glu530Lys		
c.1641C>T,D547D	c.1641C>T p.=	
c.1916 C>T p.Thr625Met	c.1916 C>T p.Thr639Met	Our mistake
c.1936 -40 G>A		
c.2165c>g	c.*1143+1011C>G	
c.2213g>a	c.*1143+1070G>A	
c.2258g>c	c.*1143+1115C>A	

Variants tested at CF Center, Prague

CF

Our nomenclature	Alamut nomenclature	Notes
p.Arg347His	c.1040G>A	
c.711+3A>G (p.?)	c.579+3A>G	
c.641G>A (p.Arg170His)	c.509G>A	
c.2683C>T (p.Arg851X)	c.2551C>T	
c.3434T>G (p.Met1101Arg)	c.3302T>G	
c.4135C>T (p.Leu1335Phe)	c.4003C>T	
c.3271G>C (p.Gly1047Arg)	c.3139G>C	
c.1811+1G>C (p.?)	c.1679+1G:C	
c.3604C>T (p.Arg1158X)	c.3472C>T	
c.2988G>T (p.Met952Ile)	c.2856G>T	
c.355C>T (p.Arg75X)	c.223C>T	
c.2622+1G>A (p.?)	c.2490+1G>A	
c.211G>A (p.Gly27Arg)	c.79G>A	
c.3238del (p.Thr1036profsX24)	c.3106del	
c.296+1G>A (p.?)	c.164+1G>A	
c.3328C>T (p.Arg1066Cys)	c.3196C>T	
c.1249-1G>A (p.?)	c.1117-1G>A	
c.3944_3945del	c.3816_3817del, non official c.3948_3949del (p.Ser1273LeufsX28)	our inexact label
c.3123G>C (p.Leu997Phe)	c.2991G>C	
c.1092A>T (p.Leu320Phe)	c.960A>T	
c.356G>A (p.Arg75Gln)	c.224G>A	
c.185+1G>A (p.?)	c.53+1G>A	
c.665G>A (p.Gly178Glu)	c.533G>A	
c.1885G>T (p.Glu585X)	c.1753G>T	
c.3719C>G (p.Ser1196X)	c.3587C>G	
c.1525-1G>A (p.?)	c.1393-1G>A	
c.4558C>T (p.Gln1476X)	c.4426C>T	
c.257C>T (p.Ser42Phe)	c.125C>T	
c.575T>C (p.Ile148Thr)	c.443T>C	

CF continued…

Our nomenclature	Alamut nomenclature	Notes
c.711+1G>T (p.?)	c.579+1G>T	
c.4103T>C (p.Leu1324Pro)	c.3971T>C	
c.2837del (p.Ser902ThrfsX4)	c.2705del	
c.4136T>C (p.Leu1335Pro)	c.4004T>C	
c.2721_2731del (p.Ile864serfsX28)	c.2589_2599del	
c.3485C>T (p.Ser1118Phe)	c.3353C>T	