

6th Educational IgCLL Workshop – Uppsala 2016

Session II Practical exercises

Vasilis Bikos

Session II - Practical exercises

- An Introduction of IMGT / V-QUEST annotation tool
 - get familiar with the sequence submission process*
 - review basic parameters and options of the tool*
- Hands-on exercises
 - two sets of IGH and IGL rearrangement sequences*
 - annotate the sequences and comment on their features*
 - and/or their interpretation*
- Sequence interpretation and results

Session II - Practical exercises

Our tool of choice : IMGT / V-QUEST

WELCOME!
to the IMGT Home page



THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®

IMGT
References and News
Contacts & Legal notices

IMGT®, the International ImMunoGeneTics Information system® <http://www.imgt.org>, is the global reference in Immunogenetics and Immunoinformatics, created in 1989 by Marie-Paule Lefranc (Université de Montpellier) and CNRS. IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T-cell receptors (TR), major histocompatibility (MH) of human and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MHSF) and related proteins of the immune system (RPI) of vertebrates and invertebrates. IMGT® provides a common access to sequence, genome and structure Immunogenetics data, based on the concepts of IMGT-ONTOLOGY and on the IMGT Scientific chart rules. IMGT® works in close collaboration with EBI (Europe), DBJ (Japan) and NCBI (USA). IMGT® consists of sequence databases, genome database, structure database, and immunological antibodies database, Web resources and interactive tools.

IMGT founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr), Université de Montpellier, CNRS, IGM, IGH, SFR, Montpellier (France).

The 2015 IMGT® Customer Satisfaction Survey. The Quality Management System of IMGT® Montpellier France has been approved by Lloyd's Register Quality Assurance France SAS to the following Quality Management System Standard: ISO 9001:2008

IMGT databases

IMGT/IGIMDB (doc) IGM, Montpellier France
Nucleotide sequences of IG and TR from 351 species (177 316 entries)
IMGT/MH-DB ANRL BPRC, hosted at EBI
Sequences of the human MH (HLA)
IMGT/PRIMER-DB (doc) IGM, Montpellier France
Oligonucleotides (primers) of IG and TR from 11 species (1 884 entries)
IMGT/ICLL-DB (db/aux) IGM, Montpellier France
IG sequences from CLL, an initiative of the IMGT/ICLL-DB group

IMGT/GENE-DB (doc) IGM, Montpellier France
International nomenclature for IG and TR genes from human, mouse, rat and rabbit
(3 825 genes, 6 827 alleles)

IMGT/3Dstructure-DB and IMGT/2Dstructure-DB (doc) IGM, Montpellier France
3D structures (IMGT Colliers de Perles) of IG antibodies, TR, MH and RPI (4 768 entries)
Source: PDB, INN, Kabat

IMGT/mAb-DB (doc) IGM, Montpellier France
Monoclonal antibodies (IG, mAb) and fusion proteins for immune applications (FPIA) (840 entries)

IMGT tools

IMGT/V-QUEST (doc) (sequence alignment software for IG and TR)
IMGT/HighV-QUEST (doc) (NGS High-Throughput analysis of IG and TR)
IMGT/StatClonotype (doc) (Statistical analysis from IMGT/HighV-QUEST output) NEW!
IMGT/JunctionAnalysis (doc) (for human and mouse IG and TR)
IMGT/Alele-Align
IMGT/PhyloGene (doc)
IMGT/DomainDisplay (doc) (Amino acid sequences)

IMGT/CDR3-Tool (doc) (IMGT/GenoView, IMGT/CloneSearch, IMGT/CloneSearch (doc) for human IGK, IG, IgH, TRα, TRβ, TRγ, mouse TRα/TRδ and human MH)
IMGT/GeneInfo (doc) (IMIC and ICH, Grenoble, IGM, Montpellier)
IMGT/GeneFrequency (doc)

IMGT/DomainGapAlign (doc)
IMGT/Collier-de-Perles (doc)
IMGT/DomainSuperimpose (doc)
IMGT/StructuralQuery (doc)

IMGT Web resources

IMGT Repertoire (IG and TR, MH and RPI)
IMGT Scientific chart (Sequence and 3D structure identification and description, Numbering, Nomenclature, Representation rules)
IMGT Index (FactBook, IMGT-ONTOLOGY, Sequence submission, Taxonomy,...)
IMGT Biblio-notes (Interesting links, PubMed, Meeting announcements, Postdoctoral positions and jobs, Messages, Search engines,...)
IMGT Education (IMGT Lexique, Aide-mémoire, Tutorials, Questions and answers, Enseignement,...)
IMGT Posters and diaporama
The IMGT Medical page
The IMGT Veterinary page
The IMGT Biotechnology page
The IMGT Immunoinformatics page

IMGT other accesses

IMGT Other accesses (ARSA, SRS, MRB)
Compare your sequence against IMGT (BLAST, FASTA)
IMGT/IGIMDB Sequence submission
IMGT downloads

IMGT Latest news

New version of IMGT/StatClonotype (1.0.1) (Fri, 09 Sep 2016 10:33:10 +0200)
New version of IMGT/GeneFrequency (2.0.0) (Wed, 24 Aug 2016 17:38:29 +0200)
New version of IMGT/VQUEST (program version 3.4.2, reference directory release: 20163141) (Thu, 4 Aug 2016 08:00:00 +0200)
New release of IMGT/V-QUEST (reference directory release: 201638-5) (Fri, 08 Jul 2016 08:00:00 +0200)

Access: <http://www.imgt.org/>

IMGT tools

IMGT/V-QUEST (doc) (sequence alignment software for IG and TR) ←
IMGT/HighV-QUEST (doc) (NGS High-Throughput analysis of IG and TR)
IMGT/StatClonotype (doc) (Statistical analysis from IMGT/HighV-QUEST output) NEW!
IMGT/JunctionAnalysis (doc) (for human and mouse IG and TR)
IMGT/Alele-Align
IMGT/PhyloGene (doc)
IMGT/DomainDisplay (doc) (Amino acid sequences)

Select “human”

on the tab for “immunoglobulin (IG) or antibody nucleotide sequences”

Session II - Practical exercises - *Submit sequence(s)*

Analyse your Ig or antibody nucleotide sequences

Your selection: Human

Your sequences will be compared to the Human (*Homo sapiens*) Ig set from the IMGT/V-QUEST reference directory sets.

Sequence sets to test IMGT/V-QUEST are available here

Sequence submission

Submit a single or a batch of sequences (up to 50)

*manual copy/paste
or file upload*

In FASTA format

Advanced parameters

<p>Selection of IMGT reference directory set</p> <p>Search for insertions and deletions in V-REGION</p> <p>Parameters for IMGT/JunctionAnalysis</p>	<p>F+ORF+ in-frame P</p> <p><input checked="" type="radio"/> Yes (slower, the nb of submitted sequences in a single run is limited to 10)</p> <p>Nb of accepted D-GENE in IGH JUNCTION (default is 1)</p> <p>Nb of accepted mutations:</p>	<p><input checked="" type="radio"/> With all alleles <input type="radio"/> With allele *01 only</p> <p><input type="radio"/> No</p> <p><input type="checkbox"/> default ▾</p> <p><input type="checkbox"/> default ▾ in 3'V-REGION <input type="checkbox"/> default ▾ in D-REGION <input type="checkbox"/> default ▾ in 5'J-REGION</p>		
<p>Parameters for "Detailed view"</p> <table border="0" style="width: 100%;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 9 and 10)</p> <p><input type="text" value=""/></p> </td> <td style="width: 50%; vertical-align: top;"> <p>Nb of nucleotides to add (or exclude) in 3' of the V-REGION for the evaluation of the alignment score (in results 1)</p> <p><input type="text" value=""/></p> </td> </tr> </table>			<p>Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 9 and 10)</p> <p><input type="text" value=""/></p>	<p>Nb of nucleotides to add (or exclude) in 3' of the V-REGION for the evaluation of the alignment score (in results 1)</p> <p><input type="text" value=""/></p>
<p>Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 9 and 10)</p> <p><input type="text" value=""/></p>	<p>Nb of nucleotides to add (or exclude) in 3' of the V-REGION for the evaluation of the alignment score (in results 1)</p> <p><input type="text" value=""/></p>			

Session II - Practical exercises – *Retrieve results*

Display results

A. Detailed view

HTML Text

Nb of nucleotides per line in alignments: ▾

Nb of aligned reference sequences: ▾

- 1. Alignment for V-GENE
- 2. Alignment for D-GENE
- 3. Alignment for J-GENE
- 4. Results of IMGT/JunctionAnalysis
 - with full list of eligible D-GENE
 - without list of eligible D-GENE
- 5. Sequence of the JUNCTION ('nt' and 'AA')

- 6. V-REGION alignment
- 7. V-REGION translation
- 8. V-REGION protein display
- 9. V-REGION mutation and AA change table
- 10. V-REGION mutation and AA change statistics
- 11. V-REGION mutation hotspots

- 12. Sequences of V-, V-J- or V-D-J- REGION ('nt' and 'AA') with gaps in FASTA and access to IMGT/PhyloGene for V-REGION ('nt')
- 13. Annotation by IMGT/Automat
- 14. [IMGT Collier de Perles](#)
 - link to IMGT/Collier-de-Perles tool
 - IMGT Collier de Perles (for a nb of sequences < 5)
 - no IMGT Collier de Perles

[Check all](#) | [Uncheck all](#) | [Default](#)

B. Synthesis view

HTML Text

Nb of nucleotides per line in alignments: ▾

Summary table sequence order: ▾

- 1. Alignment for V-GENE
- 2. V-REGION alignment
- 3. V-REGION translation
- 4. V-REGION protein display

- 5. V-REGION protein display (with AA class colors)
- 6. V-REGION protein display (only AA changes displayed)
- 7. V-REGION most frequently occurring AA
- 8. Results of IMGT/JunctionAnalysis

[Check all](#) | [Uncheck all](#) | [Default](#)

C. Excel file

Open in a spreadsheet Download in a zip archive Display 1 CSV file in your browser

- 1. Summary
- 2. IMGT-gapped-nt-sequences
- 3. nt-sequences
- 4. IMGT-gapped-AA-sequences
- 5. AA-sequences
- 6. Junction

- 7. V-REGION-mutation-and-AA-change-table
- 8. V-REGION-nt-mutation-statistics
- 9. V-REGION-AA-change-statistics
- 10. V-REGION-mutation-hotspots
- 11. Parameters
- 12. scFv (only for option "Analysis of single chain Fragment variable (scFv)")

[Check all](#) | [Uncheck all](#) | [Default](#)

Information about the type of files provided for each option:

IMGT/V-QUEST Documentation

http://www.imgt.org/IMGT_vquest/share/textes/imgtvquest.html

Practical exercises – Solutions – Batch I

>Sequence A1

```
caggtcagctacagcagtggggcgcaggactgttgaagcctcggagaccctgtc  
cctcacctgcgctgtctatggtggtccttcagtggttactactggagctggatccgcc  
agcccccaggaaagggctggagtggattgggaaatcaatcatagtggaaagcac  
caactacaaccgtccctcaagagtcgagtcaccatatcagtagacacgtccaaga  
accagttctccctaaagctgagctctgtgaccgccggacacggctgtgtattactg  
tgcgagaggtctacctttggagtggattgggtccttactactactacggta  
tggacgtctggggccaagggaccacggcaccgtctcct
```

Sequence A1

Sequence: 1 A1

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>A1
caggtgcagctacagcagtgggcgcaggactgttgaagccttcggagaccctgtccctc
acatcgctgtctatgggggtccttcagtggtaactactggagctggatccggccagcccc
ccagggaaaggggctggagttgggaaatcaatcatatgttggaaagcaccactacaac
ccgtccctcaagagtcgagtcaccatatacgttagacacgttccaaaaccaggttctcccta
aagctgagctctgtgaccggccggacacggctgttattactgtgcgagaggttctacct
ctttggagtggttattgggtccttactactactacgttatggacgtctggggccaa
gggaccacggtcaccgttcct
```

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV4-34*01 F	score = 1411	identity = 99.65% (284/285 nt)
J-GENE and allele	HomsapIGHJ6*02 F	score = 291	identity = 98.33% (59/60 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD3-3*01 F		D-REGION is in reading frame 1
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.10]	[8.7.22]	CARGLPLLEWLLGPYYYYYYGMDVW

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
AB019439 Homsap IGHV4-34*01 F	1411	99.65% (284/285 nt)
M99684 Homsap IGHV4-34*02 F	1402	99.30% (283/285 nt)
X92255 Homsap IGHV4-34*03 F	1402	99.30% (282/284 nt)
M95113 Homsap IGHV4-34*08 F	1402	99.30% (283/285 nt)
X56591 Homsap IGHV4-34*12 F	1402	99.30% (283/285 nt)

Sequence A1

6. V-REGION alignment according to the IMGT unique numbering

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56581 Homsap IGHV4-34*12 F

```
<----- FRI - IMGT  
1      5          18         15  
cag gtg cag cta cag cag tgg agg gca ... gga ctg ttg agg cct  
-----  
-----  
-----
```

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56581 Homsap IGHV4-34*12 F

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56581 Homsap IGHV4-34*12 F

CDR1 - IMGT
35 40 45
tgt tgt tac tac tgg acc tag atc cgc cag ccc

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56591 Homsap IGHV4-34*12 F

FR2 - IMGT> CDR
50 55 68
ccg ggg aag ggg ctc gag tgg att agg gaa atc aat cat agt ...

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56591 Homsap IGHV4-34*12 F

- IMGT _____ <-
..... 65 70 75
.... gga agc acc sac tac ssc ccc tcc ctc sag ... agt cgg
.....
.....
.....
.....

A1
AB019439 Homosp IGHV4-34*01 F
M99684 Homosp IGHV4-34*02 F
X92255 Homosp IGHV4-34*03 F
M95113 Homosp IGHV4-34*08 F
X56501 Homosp IGHV4-34*12 F

----- FR3 - IMGT -----
80 85 90
gtc acc ata tca gta gac acg tcc aag aac cag ttc tcc cta aat
----- ----- g
----- ----- g
----- ----- g
----- ----- g

A1
AB019439 Homosp IGHV4-34*01 F
M99684 Homosp IGHV4-34*02 F
X92255 Homosp IGHV4-34*03 F
M95113 Homosp IGHV4-34*08 F
X56501 Homosp TGHV4-34*12 F

95 100 104
ctg agc tct gtc acc gcc gcg gag acg acc gct gtc tat tac tgt gca
.....
.....

A1
AB019439 Homsap IGHV4-34*01 F
M99684 Homsap IGHV4-34*02 F
X92255 Homsap IGHV4-34*03 F
M95113 Homsap IGHV4-34*08 F
X56591 Homsap TGHV4-34*12 F

Sequence A1

7. V-REGION translation

- A single silent mutation
- VH CDR3 anchors and VH FR4 GXG motif are detected

A1 AB019439 Hom sapiens IGHV4-34*01 F

FR1 - IMGT
1 Q V Q L Q Q W G A 10 G L L K P
cag gtg cag cta cag cag tgg ggc gca ... gga ctg ttg aag cct

20 S E T L S L T C A V Y G G S F
tcg gag acc ctg tcc acc tgc gct gtc tat ggt ggg tcc ttc

CDR1 - IMGT
35 S G Y Y W S W I R Q P
agt ggt tac tac tgg agc tgg atc cgc cag ccc

FR2 - IMGT CDR2
50 P G K G L E W I G E I N H S
cca ggg aag ggg ctg gag tgg att ggg gaa atc sat cat agt ...

- IMGT
65 G S T N Y N P S L K S R
ggg agc acc acc tac acc ccc tcc ctc aag ... agt cga

FR3 - IMGT
80 V T I S V D T S K N Q F S L C
gtc acc ata tca gta gac acg tcc aag acc cag ttc tcc cta aag
85 ...
95 L S S V T A A D T A V Y Y C A
ctg agc tct gtg acc gcc gcg gac acg gct gtc tat tac tgt gcg
100 ...
184 ...
CDR3 - IMGT
R G L P L L E W L L G P Y Y Y
aga ggt cta cct ctt ttg gag tgg tta ttg ggt cct tac tac tac
A1 AB019439 Hom sapiens IGHV4-34*01 F
Y Y G H D V W G Q G T T V T V
tac tac ggt atg gac gtc tgg ggc cca ggg acc acg gtc acc gtc
A1 S
tcc t

Sequence A1

IGHV4-34*01

IGHD3-3*01

IGHJ6*02

identity 99,65%

in-frame

VH CDR3 length 22 aa

→ productive, unmutated IGH

Practical exercises – Solutions – Batch I

>Sequence A2

```
caggtgcagctggtgcagtctggagctgaagtgaagaaggcctgggcctcagtcaa  
ggtctcctgcaaggcttctggttacaccttacgaattatggtatcggtcggtgcga  
caggcccctggacaaggcgttgagtggatggatggatcagcggttacaatggta  
tacaaactatgcacagaagttccaggacagagtaccatgaccacagacacatcca  
cgagcacagcctatatggacctggggagcctgagatctgacgacacggccgttat  
tactgtgcgagagttagtggctacgctccccccgtctactggggccagggAACCTG  
gtcaccgtctcctca
```

Sequence A2

Sequence: 1 A2

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>A2
cagggtgcagctgggtgcagtctggagctgaagtgaaaggaaacccctggggcctcagtggatgc
ccctgcaaggcttctgggttacaccctttacgaattatggatcggttcggctgggtgcgcacaggcc
cctggacaaggcttggatggatggatggatcagcggttacaatggtgatacacaaactat
gcacagaaggttccaggacagagtccaccatgaccacagacacateccacgagcacaaggcttat
atggacacctggggggctgagatctgacgacacggccgtgttattactgtgcggagatgtg
gctacgtccccccgtctactggggccaggggacccctggtcaccgtctctca
```

Result summary:		Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele		<u>Homsap IGHV1-18*01 F</u>	score = 1309	identity = 95.14% (274/288 nt)
J-GENE and allele		<u>HomsapIGHJ4*02 F</u>	score = 172	identity = 85.11% (40/47 nt)
D-GENE and allele by IMGT/JunctionAnalysis		<u>HomsapIGHD5-12*01 F</u>	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.11]		CARWATLPPVYW

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
M99641 Homsap IGHV1-18*01 F	1309	95.14% (274/288 nt)
HM855463 Homsap IGHV1-18*03 F	1300	94.79% (273/288 nt)
KC713938 Homsap IGHV1-18*04 F	1300	94.79% (273/288 nt)
X60503 Homsap IGHV1-18*02 (F)	1245	94.57% (261/276 nt)
M99637 Homsap IGHV1-8*01 F	1147	88.89% (256/288 nt)

Sequence A2

6. V-REGION alignment according to the IMGT unique numbering

A2
N99641 Homsap IGHV1-18*01 F
H1855463 Homsap IGHV1-18*03 F
KC1713938 Homsap IGHV1-18*04 F
X66503 Homsap IGHV1-18*02 (F)
M99637 Homsap IGHV1-8*01 F

A2
M99641 Homsap IGHV1-18*01 F
HM1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X60503 Homsap IGHV1-18*02 (F)
NM0637 Homsap IGHV1-8*01 F

A2
M98641 Homsap IGHV1-18*01 F
HM1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X60503 Homsap IGHV1-18*02 (F)
M98637 Homsap IGHV1-8*01 F

```
— CDR1 - IMGT      <          45  
..... 35       40          45  
acg ast tat ggt atc gcc tgg  gtg cga cag gcc  
--c -gc ---     a-  
--c -gc ---     a-  
--c -gc --c --- a-  
--c -gc ---     a-  
--c -gc ---     a-  
--c -gc ---     a-
```

A2
M99641 Homsap IGHV1-18*01 F
HM1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X66503 Homsap IGHV1-18*02 (F)
M98637 Homsap IGHV1-8*01 F

A2
M99541 Homsap IGHV1-18*01 F
HM855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X60503 Homsap IGHV1-18*02 (F)
N99637 Homsap IGHV1-8*01 F

A2
M99641 Homsap IGHV1-18*01 F
HV1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X60503 Homsap IGHV1-18*02 (F)
M99637 Homsap IGHV1-8*01 F

```

A2
M99641 Homsap IGHV1-18*01 F
H1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X68503 Homsap IGHV1-18*02 F
M99637 Homsap IGHV1-8*01 F

```

95	100	104
ctg ggg agc ctg aga tct gag gag acg gcc gtg tat tac tgt ggc		
--- a-----		
--- a-----	-----t-	-----
--- a-----	-----	
--- a-----a	-----	
--- a-----a	-----	

A2
M99641 Homsap IGHV1-18*01 F
H1855463 Homsap IGHV1-18*03 F
KC713938 Homsap IGHV1-18*04 F
X60503 Homsap IGHV1-18*02 F
M99637 Homsap IGHV1-8*01 F

```
aga gta gtc gct acg ctc ccc ccc gtc tac tgg ggc cag gga acc  
--- -a  
--- -a  
--- -a  
  
--- -a
```

Sequence A2

9. V-REGION mutation and AA change table

FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
t6>g g33>a	c105>g g107>a, S36>N (- - -) c108>t, S36>N (- - -)	a118>g, S40>G (+ + -)	c173>g, A58>G (- + -) a190>g, N84>D (+ + -) c192>t, N84>D (+ + -)	c211>t, L71>F (+ - -) g221>a, G74>D (- - -) c264>t g270>c, E90>D (+ - +) a274>g, R92>G (- - -)	a320>t

http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/IMGTclasses.html

Amino acid classification

IMGT 'Physicochemical' classes of the 20 common amino acids *

Eleven IMGT 'Physicochemical' classes of the 20 common amino acids have been defined by the physicochemical properties of their side chains (Figure 1) [1]. These standardized classes are used in IMGT® databases and tools, for the description of amino acid class changes in mutations and protein engineering.

- Hydropathy
- Volume
- Chemical
- Charge
- Hydrogen donor/acceptor
- Polarity

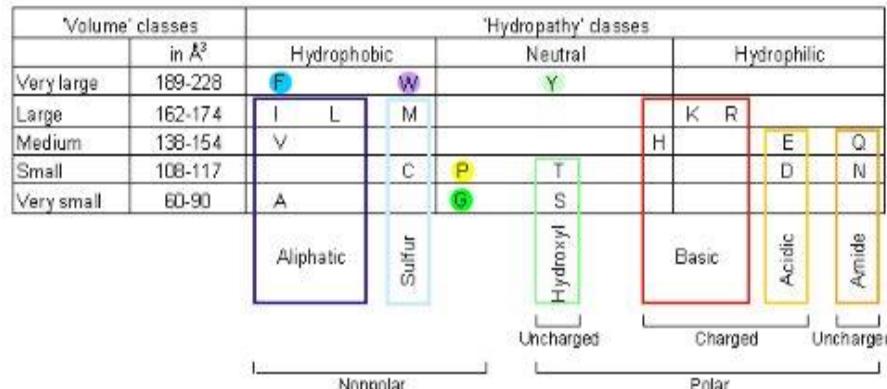


Figure 1: The 11 IMGT 'Physicochemical' classes of the 20 common amino acids [1].

Sequence A2

IGHV1-18*01

IGHD5-12*01

IGHJ4*02

identity 95,14%

in-frame

VH CDR3 length 11 aa

→ productive, mutated IGH

Practical exercises – Solutions – Batch I

>Sequence A4

```
caggtcagtcaggagtccccaggactggtaagcctcgaaaaaccctgtc  
cctcacctgcgctgtctggctccatcagcagtactaactggtgaggatgg  
cgccagccccaggaaaggggctggaggatggagaaatctatcatatgg  
gcaccaactacaaccgtccctcaagagtgcgactcaccatata  
aagaaccaggttctcgctaatctgaggcgtgaccggccggacacggccgttat  
tactgtcgagaggtagtggtagctggacttgcattactggccaggaa  
accctggtcaccgtctcc
```

Sequence A4

Sequence: 1 A4

Sequence compared with the human IG set from the IMGT reference directory

>A4
gggtgcagctgggggtctggggggcctggtcggccctgggggtccctgagactc
tcctgtgcggcctctggattcaccttcagtagctatgtatggactgggtccggccaggct
ccggggggggctgggtgggtctcatccatggcggtagtactatgtttacaaaaactac
gcggactcggtgaaaaaaacggccggatccacatctccgggggggggggggggggggg
ctgcggatgg
gggggtatgg

Result summary:		Productive IGH rearranged sequence: (no stop codon and in-frame junction)	
V-GENE and allele	Homsap IGHV3-21*01 F	score = 1372	identity = 97.57% (281/288 nt)
J-GENE and allele	HomsapIGHJ6*02 F	score = 215	identity = 90.38% (47/52 nt)
D-GENE and allele by IMGT/JunctionAnalysis	No results	-	-
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.9]	CARDANGMDWW

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
AB019439 Homsap IGHV3-21*01 F	1372	97.57% (281/288 nt)
M99658 Homsap IGHV3-21*02 F	1363	97.22% (280/288 nt)
HM855323 Homsap IGHV3-21*03 F	1363	97.22% (280/288 nt)
HM855688 Homsap IGHV3-21*04 F	1363	97.22% (280/288 nt)
HM855336 Homsap IGHV3-48*04 F	1300	94.79% (273/288 nt)

Sequence A4

No D-gene has been identified. But the sequence itself is more important than gene names

4. Results of IMGT/JunctionAnalysis

Maximum number of accepted mutations in: 3'V-REGION = 2, D-REGION = 4, 5'J-REGION = 2
Maximum number of accepted D-GENE = 1

Analysis of the JUNCTION

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N	5'J-REGION	J name	Vmut	Jmut	Ngc
A4	Homsap IGHV3-21*01	tgtgcgagaga	cgcgaat	ggatataggacgtctgg	Homsap IGHJ6*02	0	0 4/7

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	114	115	116	117	118	Frame	CDR3- IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	A	R	D	N	G	H	D	V	K		+	9	1,237.38	4.44	CARDANGHDVH

A4 tgt gcg aga gac gac atg ggt atg gac gtc tgg +

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

Subset #2 features

- IGHV3, mostly IGHV3-21
- MostlyIGHJ6
- Mostly mutated
- VH CDR3 : 9 aa
- Mostly D or E at position VH CDR3-107

This rearrangement represents a CLL subset#2 case. Thus it is related with poor prognosis and this should be stated clearly in the report

Sequence A4

IGHV3-21*01

IGHD ND

IGHJ6*02

identity 97,57%

in-frame

VH CDR3 length 9 aa

→ productive, borderline mutated IGH,
subset #2 case

Practical exercises – Solutions – Batch I

>Sequence A5

```
gaggtcagctggaggactggccctggatcacccatgcactgggtccgc  
acttcctgtgcagcctctggattcacccatgcactgggtccgc  
aggctccaggaaaggactggaatatgtttcagctatttagtagtaatggggtagc  
acatattatgcaaactctgtgaaggcagattcaccatctccagagacaattccaag  
aacacgctgtatcttcaaattggcagcctgagagctgaggacatggctgttattac  
tgtgcgagcgctggagtccttatggtgccacgagaggactactactacta  
catggacgtctgggccaaggaccctggtcac
```

Sequence A5

Sequence: 1 A5

Sequence compared with the human Ig set from the IMGT reference directory

Result summary:	Unproductive IGH rearranged sequence (out-of-frame junction)		
V-GENE and allele	Homsap IGHV3-64*01 F	score = 1435	identity = 100.00% (288/288 nt)
J-GENE and allele	HomsapIGHJ6*03 F	score = 215	identity = 90.38% (47/52 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD5-12*01 F	D-REGION is in reading frame 2	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.7]	[8.8.X]	CASAGSPSPMMATR#Y#YYYYYYMDVV

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
M99682	Homsap IGHV3-64*01 F	1435	100.00% (288/288 nt)
AB019437	Homsap IGHV3-64*02 F	1417	99.31% (286/288 nt)
KC713941	Homsap IGHV3-64D*06 F	1363	97.22% (280/288 nt)
M77301	Homsap IGHV3-64*05 F	1354	96.88% (279/288 nt)
M77298	Homsap IGHV3-64*03 F	1345	96.53% (278/288 nt)

Sequence A5

from IMGT scientific chart:

<http://www.imgt.org/IMGTScientificChart/Numbering/IMGThumbering.html>

IMGT unique numbering for all IG and TR V-REGIONS of all species: interest for structure and evolution

Citing IMGT unique numbering for V-REGION: Lefranc, M.-P. et al., Dev. Comp. Immunol., 27, 55-77 (2003) PMID: 12477501 [\[PubMed\]](#)

In order to easily compare V-REGION sequences of IG and TR from all species, a unique numbering has been defined by Marie-Paule Lefranc [1-8].

The 'IMGT unique numbering' concept is part of the 'NUMEROPTION' concept of IMGT-ONTOLOGY [1].

The IMGT unique numbering for all IG and TR V-REGIONS of all species relies on the high conservation of the structure of the variable region [1-8]. This numbering, set up after aligning more than 5 000 sequences, takes into account and combines the definition of the framework (FR) and complementarity determining regions (CDR) [8], structural data from X-ray diffraction studies [9], and the characterization of the hypervariable loops [10]. The delimitations of the FR-IMGT and CDR-IMGT regions have been defined and Correspondence between the IMGT unique numbering and the other numberings has been established [8].

The IMGT unique numbering has many advantages:

- It allows an easy comparison between sequences coding the variable regions, whatever the antigen receptor (IG or TR), the chain type (heavy or light chains for IG [4,11], alpha, beta, gamma or delta chains for TR [8]), or the species.
- A unique numbering is chosen. In the same way, the FR and CDR are numbered and can be compared. Moreover, the IMGT unique numbering is descriptive of the structure of the variable region, belonging to a group, a subgroup and/or a gene [1-8].
- Framework amino acids (and codons) located at the same position in different sequences can be compared without requiring sequence alignments [11,12]. This also holds for amino acids belonging to CDR-IMGT of the same length.
- The IMGT unique numbering has allowed a standardized IMGT description of mutations for the IMGT description of allele polymorphisms and somatic hypermutations of the variable regions. The mutations and allelic polymorphisms are described by comparison to the germline V-REGION, D-REGION, J-REGION and to the C-REGION from the IMGT reference directory. Based on these criteria, Alignments of alleles [4,13] and Tables of alleles have been set up.
- The IMGT unique numbering has allowed to set up the IMGT standardized criteria for the statistical analysis of IG and TR V-REGION amino acid properties [11].
- The unique numbering is used as the output of IMGT/QUEST integrated sequence alignment tool which analyzes your IG and TR variable (germline or rearranged) sequences according to IMGT criteria. In IMGT/QUEST, a variable rearranged sequence is compared to the appropriate sets of V-REGION, D-REGION and J-REGION alleles from the IMGT reference directory. The results show, aligned with the input sequence, the sequences of the most homologous V-REGION alleles and, if appropriate, D-REGION (for heavy chains) and J-REGION alleles. The aligned V-REGION sequences are displayed according to the IMGT unique numbering and with the FR-IMGT and CDR-IMGT delimitations.

By facilitating comparisons between the sequences and the descriptions of alleles and mutations, the IMGT unique numbering represents a big step forward in the analysis of the IG and TR sequences of all species. Moreover, it gives in-sight in the structural configuration of the variable domain and opens interesting views on the evolution of the sequences of the V-set, since this numbering has been applied with success to all the sequences belonging to the V-set of the immunoglobulin superfamily, including nonrearranging sequences in vertebrates (CD4, Xenopus CTX, etc.) and in invertebrates (Drosophila *Amalgam*, *Drosophila Fasciclin*, etc.) [1,2,8].



In the IMGT unique numbering, the conserved amino acids always have the same position, for instance **Cystein 23 (1st-CYS), Tryptophane 41 (CONSERVED-TRP), Leucine 89, Cystein 104 (2nd-CYS)**. The hydrophobic amino acids of the framework regions are also found in conserved positions.

Sequence A5

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	110	111	111.1	111.2	111.3	111.4	112.5	112.4	112.3	112.2	112.1	112	113	114	115	116	117	118	Frame	CDR3- IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	A	S	A	G	S	P	S	H	V	A	T	R	#	#	Y	Y	Y	Y	Y	H	D	V	W					
A5	tgt	ccg	agc	gtt	ggg	agt	cct	tct	atg	gtg	ggc	acg	aga	gg	ac	tac	tac	tac	tac	atg	gac	gtc	tgg	-	NR	NR	NR	NR

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele "01" for the analysis of the JUNCTION.

What to do when this is the single rearrangement amplified?

The amplification reaction should be **repeated**:

using **different sets of primers** in order to

- (i) confirm the presence of the unproductive rearrangement
- (ii) look for the presence of a second productive rearrangement.

Sequence A5

IGHV3-64*01

IGHD5-12*01

IGHJ6*03

identity 100,00%

out-of-frame

VH CDR3 length X

→ unproductive, unmuted IGH

Practical exercises – Solutions – Batch I

>Sequence A7

```
gaggtcagctggagtctgggtgaggcttgtacagcctggagggtccctgag  
acttcctgtgcagcctctggattcacccatcgtagctcctggatgcactgggtctgcc  
aggctccggagaagggctggagtgggtggccgacataaagtgtgacggaagtga  
caaatactatgttagactctgtgaagggccgattgaccatctccagagacaatgcca  
gaactcccttatctgcaagtgaacagcctgagagctgaggacatgaccgtgtatta  
ctgctgggggtcttacgatatttgactggttattacgccggaccgctcacaactgg  
ttcgaccctggggccagggAACCTGGTACCGTCTCCT
```

Sequence A7

Sequence: 1 A7

Sequence compared with the human Ig set from the IMGT reference directory.

>A7
gaggtgcagctgggtggagttctgggtggggcttgggtacagccctggaggggtccctgagactc
tcctgtgcggcctctggattcaccttcggatgtacgtctctggatgtgcactgggtctggcaggct
ccggagaaaggggctggagttgggtggccgacataatgtgtgacggaaagtgcacaaatactat
gttagactctgtgaaggccgattgaccatctccagagacaatggcaagaactccctctat
ctgcaagtggacacagccgtggagactggggacatggaccgtgtattactgtgggggtttac
gatattttgactgggtattacggccggggaccgctcacaaactggttcgaccccctggggccag
ggaaaccttggtcacccgtctcc

Result summary:	Unproductive IGH rearranged sequence (stop codons)		
V-GENE and allele	<u>Homsap IGHV3-52*01 P</u>	score = 1426	identity = 99.31% (286/288 nt)
J-GENE and allele	<u>HomsapIGHJ5*02 F</u>	score = 240	identity = 100.00% (48/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	<u>HomsapIGHD3-9*01 F</u>	D-REGION is in reading frame 2	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.10]	[8.8.21]	CWGSYDILTGYYAGTAHNWFDPW

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

			Score	Identity	
M99678	Homsap	IGHV3-52*01	P	1426	99.31% (286/288 nt)
Z17388	Homsap	IGHV3-52*02	P	1417	98.96% (285/288 nt)
J00237	Homsap	IGHV3-52*03	P	1417	98.96% (285/288 nt)
HM1855666	Homsap	IGHV3-7*03	F	1201	98.62% (261/288 nt)
M99649	Homsap	IGHV3-7*01	F	1192	98.28% (260/288 nt)

Sequence A7

from IMGT scientific chart discussing functionality:

<http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGTfunctionality.html>

Functionality

The definition of functionality is based on the sequence analysis [1].

The 'Functionality' concept is part of the 'IDENTIFICATION' axiom of IMGT-ONTOLOGY[2].

For IG and TR "Germline" entities and for "C-GENE"

A germline entity ([V-GENE](#), [D-GENE](#) or [J-GENE](#)) or a [C-GENE](#) can be functional, ORF or pseudogene.

FUNCTIONAL

A germline entity (V-GENE, D-GENE or J-GENE) or a C-GENE is functional if the coding region has an open reading frame without stop codon, and if there is no described defect in the splicing sites, recombination signals and/or regulatory elements.

ORF (Open Reading Frame)

A germline entity (V-GENE, D-GENE or J-GENE) or a C-GENE is qualified as ORF (Open Reading Frame) if the coding region has an open reading frame, but:

- alterations have been described in the splicing sites, recombination signals and/or regulatory elements.
- and/or changes of conserved amino acids have been suggested by the authors to lead to uncorrect folding.
- besides the entity is no protein.

PSEUDOGENE

A germline entity (V-GENE, D-GENE or J-GENE) or a C-GENE is qualified as pseudogene if the coding region has stop codon(s) and/or frameshift mutation(s).

In particular, a V-GENE is considered as pseudogene if these defects occur in the [L-PART1](#) and/or [V-EXON](#), or if there is a mutation in the [L-PART1 INIT-CODON](#) arg.

A J-GENE is considered as pseudogene if it has been identified by the presence of a recombination signal upstream of an open reading frame, but it has no donor splice site in 5' or the donor splice is not in the expected [sift](#) or if it has no conserved Phe(or Trp)-Gly-X-Gly motif. If the defects are important, pseudogenes can eventually also be qualified as

vestigial (vg), for example: a germline V-GENE which cannot be assigned to a given subgroup because it is too divergent from the other pseudogenes and has too many stop codons and frameshifts.

Only vestigial pseudogenes reported in [Gene tables](#) are qualified as pseudogenes in the [List of IG and TR genes](#) and in the [Potential germline repertoires](#).



A V-gene is considered as a pseudogene if:

- there are defects along its Leader region or the V-region
- there is a mutation at the initial codon of the Leader-Part1 region

Sequence A7

IGHV3-52*01 is a pseudogene because it encodes for a **stop codon** at **VH FR1-9** position

7. V-REGION translation

A7

M99678 Hom sapiens IGHV3-52*01 P

FR1 - IMGT
1 S * 10 G L V Q P
E V Q L V E S G * 10 G L V Q P
gag gtg cag ctc gtc gag tct ggg tga ... ggc ttg gta cag ctc
----->
20 25 30
G G S L R L S C A A S G F T F
ggg ggg tcc ctc aga ctc tcc tgt gca gcc tct ggg ttc acc ttc
----->
CDR1 - IMGT
35 40 45
S S S W N H W V C Q A
... agt agc tcc tgg atg cac tgg gtc tgc cag gct
----->
FR2 - IMGT
50 55 60
P E K G L E W V A D I K C D
ccg ggg agg ggg ctc gag tgg gtc gcc gac ata agg tgt gac ...
----->
IMGT
65 70 75
G S D K Y Y V D S V K G R
... ggg agt gac aaa tac tat gta gac tct gtg agg ... ggc cga
E
----->
FR3 - IMGT
80 85 90
L T I S R D N A K N S L Y L Q
ttg acc atc tcc aga gac att gcc agg acc tcc ctc tat ctg cca
----->
95 100 104
V N S L R A E D N T V Y Y C W
gtg acc agc ctc aga gct gag gag atg acc gtg tat tac tgc tgg
V
-----> -t gt-
CDR3 - IMGT
G S Y D I L T G Y Y A G T A H
ggc tct tac gat att ttg act ggt tat tac gcc ggg acc gct cac
R
-a- gg
----->
N W F D P W G Q G T L V T V S
acc tgg ttc gac ccc tgg ggc cag gga acc ctc gtc acc gtc tcc

Sequence A7

IGHV3-52*01 P

IGHD3-9*01

IGHJ5*02

identity 99,31%

in-frame

VH CDR3 length 21

→ unmuted IGH, unproductive due
to IGHV pseudogene utilization

Practical exercises – Solutions – Batch I

>Sequence A9

```
gacatccagatgaccagtctccatcctccctgtctgcattgttaggagacagagtca  
ccatcaactgccaggcgagtcaggacattagcaactattaaattggtatcagcaga  
aaccagggaaagcccctaagctcctgatctacgatgcattcaatttggaaacaggg  
gtccccatcaagggttcagtggaaagtggatctggacagattttactttcaccatcagca  
gcctgcagcctgaagatattgcaacatattactgtcaacagcttgggtttccccgtg  
gaccttcggccaagggaccaagctggaaatcaaacgta
```

Sequence A9

Sequence: 1 A9

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

>A9
gacatccagatgacccaggctccatccctccctgtctgcacatctgttagggacacagactcacc
atcacttgcaggcgaggctggacattagcaactatttaaatttgtatcagcagaaaaacc
ggggaaaggcccctaagctctgtatcgtatccaaatttggaaaacagggtccccatca
gggttcaggatggatctggggacagatttactttcacccatcagcagctgtcagcc
ggggatattgcacatattactgtccaaacagcttggatgtttccctgtggacccattcggccaa
ggggaccaagctggaaatccaaacgtt

Result summary:	Productive IGK rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGKV1-33*01 F, or Homsap IGKV1D-33*01 F	score = 1345	identity = 98.21% (274/279 nt)
J-GENE and allele	Homsap IGKJ1*01 F	score = 172	identity = 94.74% (38/38 nt)
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[26.17.36.10]	[6.3.9]	CQQLGGSPWTF

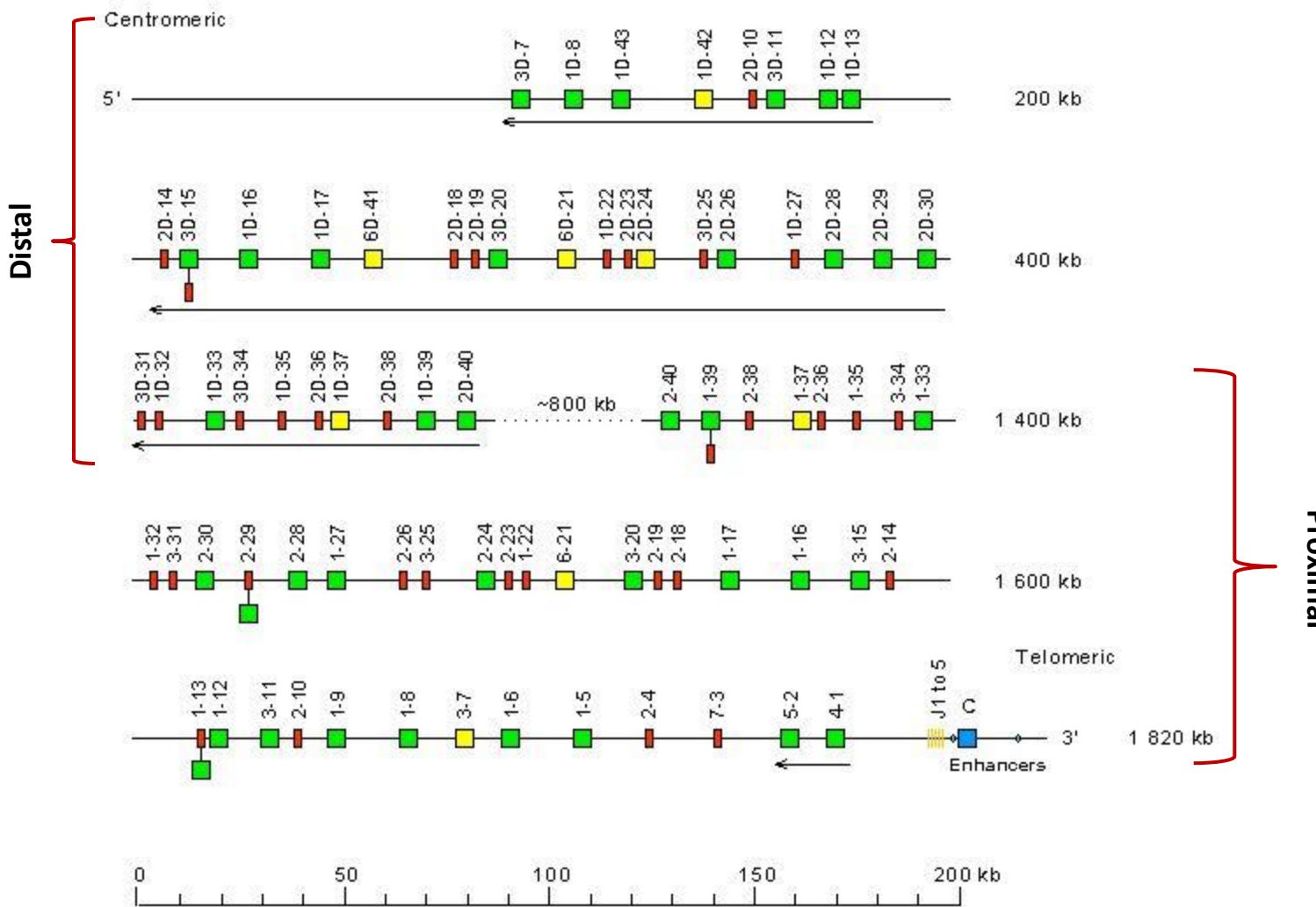
1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon plus 15 nt of the CDR3-IMGT)

	Score	Identity
M64856 Homsap IGKV1-33*01 F	1345	98.21% (274/279 nt)
M64855 Homsap IGKV1D-33*01 F	1345	98.21% (274/279 nt)
X59315 Homsap IGKV1-39*01 F	1165	91.04% (254/279 nt)
X59312 Homsap IGKV1D-39*01 F	1165	91.04% (254/279 nt)
X63398 Homsap IGKV1-27*01 F	1147	90.32% (252/279 nt)

Sequence A9

The sequence is aligned to both IGKV1-33 and IGKV1D-33



Distal V-CLUSTER: Duplication of proximal V-CLUSTER

Sequence A9

- All the identified mutations are **clustered at the end of the rearrangement**, along the VK CDR3 region
 - (%) percentage of identity is **evaluated** from the 1st IGKV nucleotide to the 2nd-CYS codon **plus 15nt of the VK CDR3**

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon plus 15 nt of the CDR3-IMGT)

			Score	Identity
M64856	Homsap	IGKV1-33*01	F	1345
M64855	Homsap	IGKV1D-33*01	F	1345
X59315	Homsap	IGKV1-39*01	F	1165
X59312	Homsap	IGKV1D-39*01	F	1165
X63398	Homsap	IGKV1-27*01	F	1147

Alignment with FR-IMGT and CDR-IMGT delimitations

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F
X63398 Homsap IGKV1-27*01 F

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F
X63398 Homsap IGKV1-27*01 F

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F
X59318 Homsap TGHV1-27*01 F

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F

A9
M64856 Homsap IGKV1-33*01 F
M64855 Homsap IGKV1D-33*01 F
X59315 Homsap IGKV1-39*01 F
X59312 Homsap IGKV1D-39*01 F

Sequence A9

4. Results of IMGT/JunctionAnalysis

Maximum number of accepted mutations in: 3'V-REGION = 2, 5'J-REGION = 2
Maximum number of accepted D-GENE = 0

Analysis of the JUNCTION

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N	P	5'J-REGION	J name	Vmut	Jmut	Ngc
A9	Homsap IGKV1-33*01	tgtcaacag <u>g</u> ttg.....	gtggttccc	c	gtggac <u>g</u> ttc	Homsap IGKJ1*01	2	1	6/9

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	114	115	116	117	118	Frame	CDR3- length	IMGT molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	Q	Q	L	G	G	S	P	N	T	F	+	9	1,223.38	5.49	COOLGGSPNTF

A9 tgt caa cag ctt ggt ggt tcc ccg tgg acg ttc

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

- case of **extreme exonuclease activity** in the 3' end of the V gene
- after 2nd-CYS codon **only 10 nucleotides belong to the V gene**

Sequence A9

Advanced parameters

Selection of IMGT reference directory set <input type="radio"/> F+ORF+ in-frame P	<input checked="" type="radio"/> With all alleles <input type="radio"/> With allele *01 only
Search for insertions and deletions in V-REGION <input checked="" type="radio"/> Yes (slower, the nb of submitted sequences in a single run is limited to 10) <input type="radio"/> No	
Parameters for IMGT/JunctionAnalysis Nb of accepted D-GENE in IGH JUNCTION (default is 1) <input type="button" value="default"/> Nb of accepted mutations: <input type="button" value="default"/> in 3'V-REGION <input type="button" value="default"/> in D-REGION <input type="button" value="default"/> in 5'J-REGION	
Parameters for "Detailed view" Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 9 and 10) <input type="text" value="-5"/> Nb of nucleotides to add (or exclude) in 3' of the V-REGION for the evaluation of the alignment score (in results 1)	

Sequence: 1 A9

Sequence compared with the human Ig set from the IMGT reference directory

Result summary:	Productive IGK rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGKV1-33*01 F , or Homsap IGKV1D-33*01 F	score = 1347	identity = 99.27% (272/274 nt)
J-GENE and allele	Homsap IGKJ1*01 F	score = 172	identity = 94.74% (36/38 nt)
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[26,17,36,10]	[6,3,8]	CQQLGGSPWTF

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon plus 10 nt of the CDR3-IMGT)

			Score	Identity
M64856	Homsap	IGKV1-33*01	F	1347
<u>M64855</u>	Homsap	IGKV1D-33*01	F	1347
X59315	Homsap	IGKV1-39*01	F	1176
<u>X59312</u>	Homsap	IGKV1D-39*01	F	1176
X63398	Homsap	IGKV1-27*01	F	1158

Practical exercises – Solutions – Batch II

>Sequence B1

```
gaggaggagggtttggagtctgggggaggcttggtacagcctgggggtccctgag  
actctcctgtgttagcctctggattcaccttagtaattatgtcatgagttggtccgcc  
aggctccagggagggactggagtggtctccgttattactgatagtggtggtcata  
catactatgcagactccgtgaagggccggttaccatctccagagacaattccaaga  
acaccctatatcttcaaatacagcctgagactcgaggacacggccagatattact  
gtgcgagagactgggggatattgtgggggaggcaccctggtcaccgtctccaca  
t
```

Sequence B1

Sequence: 1 B1

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>B1
ccatctccggcggacagtgctcttaccacactgctgctttggaaactggatcaggcagtccc
catcgagaggcccttgagtgctggggaaaggacatactacaggaaactccaaactggtaaatg
atattgcagttttctgtcaaaaggtcgaatgtcatctatccggacacatccaaaggaaacctgt
tctccctggacctgtgaaactctgtgactccggggacacggctgtgttattactgtgcaaggg
gggttgggggggggattccaggggccatggctgtgactactggggccggggaccctgg
tcaccgtttccctcaggta
```

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction) (a)		
V-GENE and allele	Homsap IGHV6-1*01 F, or Homsap IGHV6-1*02 F	score = 341	identity = 60.94% (142/233 nt)
J-GENE and allele	HomsapIGHJ4*02 F	score = 159	identity = 81.25% (39/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD2-15*01 F		D-REGION is in reading frame 3
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[3,17,38,11]	[10,9,16]	CARGVGWEIPEAHGCDYW

(a) Low V-REGION identity (60.94%): this may indicate potential nucleotide insertion(s) and/or deletion(s). Try 'Search for insertions and deletions' in 'Advanced parameters' at the bottom of the Search page

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
Z04097 Homsap IGHV6-1*01 F	341	60.94% (142/233 nt)
Z14223 Homsap IGHV6-1*02 F	341	60.94% (142/233 nt)
Z75355 HomsapIGHJ4-4*06 F	178	53.24% (115/216 nt)
Z75356 HomsapIGHV4-34*13 F	160	53.05% (113/213 nt)
Z75359 HomsapIGHV4-59*09 F	158	52.11% (111/213 nt)

Warning message!!

Sequence B1

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
J04097	Homsap IGHV6-1*01 F	341	60.94% (142/233 nt)
Z14223	Homsap IGHV6-1*02 F	341	60.94% (142/233 nt)
Z75355	Homsap IGHV4-4*06 F	178	53.24% (115/216 nt)
Z75356	Homsap IGHV4-34*13 F	160	53.05% (113/213 nt)
Z75359	Homsap IGHV4-59*09 F	158	52.11% (111/213 nt)

Alignment with FR-IMGT and CDR-IMGT delimitations

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> FR1-IMGT -----
caggtacagctgcagcagtcaggtcca...ggactgtgtaaagccctcgcaagaccctctca
caggtacagctgcagcagtcaggtcca...ggactgtgtaaagccctcgcaagaccctctca
.....
.....
.....
.....
.....

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> CDR1-IMGT -----
.....ccatctccggcgcacgtgtcttacc.....aacgtgtgtttggaaactgg
cttacccgtgtc-at-tc--gggacagtgtct-t.....-g--acag---gcttggaaac
cttacccgtgtc-at-tc--gggacagtgtct-t.....-g--acag---gcttggaaac
.....tct-gtg-ctc-ata-g.....-ag-acac---tgtta-t
.....tat-gtg-gtc-tc.....-g-tac-actgg-a-c
.....tct-gtg-ctc-ata.....-ag-tac-actgg-a-c

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> FR2-IMGT -----> CDR
atcaggcagtccccatcgagggccttggatggctggggaggcacatcatcaggaaactcc
tgg-tcag-cagt-cc-atacg-aaggcattga-tg-ctgg-a-ggaca---tac-gg--
tgg-tcag-cagt-cc-atacg-aaggcattga-tg-ctgg-a-ggaca---tac-gg--
tgggtc-ccca-ag-gannngact-ga-tg-attg-ga-at---tcat-gt...
tgg-tc-ccca-ag-gannngact-ga-tg-attg-ga-at-a-tcat-gt...
tgg-tc-ccca-ag-gannngact-ga-tg-attg-ttatat-ttac-gt...

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> 2-IMGT -----
...aactgttttaatgattatcgtgtttgtcaaa...ggtcgtatgtcatttatca
....-g----at.....-a----g----.a-----ac----a-c---
....-g----at.....-a----g----.a-----ac----a-c---
....-a-ccca-c--caacccg--cc--g...-a-cac---g-cgt-
....-a-ccca-c--caacccccc--cc--g...-a-cac---a-cgt-
....-a-ccca-c--caacccc--cc--g...-a-cac---a-cgt-

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> FR3-IMGT -----
gcacacatccaaagaaccaggatctccctggacctgaactctgtgactcccggggacacggct
.....-c-g-
.....-c-g-
.....-g-----a-g-----g-----cg--c-----c
.....-g-----a-g-----g-----cg--c-----c
.....-g-----a-g-----g-----cg-t-c-----c

B1
J04097 Homsap IGHV6-1*01 F
Z14223 Homsap IGHV6-1*02 F
Z75355 Homsap IGHV4-4*06 F
Z75356 Homsap IGHV4-34*13 F
Z75359 Homsap IGHV4-59*09 F

-----> CDR3-IMGT -----
gtgttattactgtgtcaagggcggtgggtggagatccagaggccatggctgtgactcc
.....
.....
.....
.....
.....

An overall bad alignment

Sequence B1

IMGT/V-QUEST provides **warnings** to alert the users, for the possibility that potential insertions or deletions are present in the sequence

http://www.imgt.org/IMGT_vquest/share/textes/IMGTvquest-warnings.pdf

Warning	Objective
V-GENE and allele: low V-REGION identity percentage with the closest V germline	In order to alert users that the sequence has less than 85% of identity with the closest germline V- GENE and allele: this may indicate that the V-GENE and allele name could be not reliable. This could be due to insertion(s) and/or deletion(s).
V-GENE and allele: 2nd-CYS 104 not identified	In order to alert users that the conserved 2nd-CYS at position 104 according to the unique IMGT numbering is not identified. This could be due to insertion(s) and/or deletion(s).
V-GENE and allele: different CDR1-IMGT and/or CDR2-IMGT amino acid lengths compared to closest V germline	In order to alert users that the sequence has different lengths of CDR1 and/or CDR2 compared with the closest V germline: this may indicate that the V gene and allele name could be not reliable. This could be due to insertion(s) and/or deletion(s).

Sequence B1

- Back to the submission form
- Enable the option at the “advanced parameters” tab
- Note that now you can simultaneously analyze a batch of max. 10 sequences

Advanced parameters

Selection of IMGT reference directory set F+ORF+ in-frame P

Search for insertions and deletions in V-REGION

Parameters for IMGT/JunctionAnalysis

Yes (slower; the nb of submitted sequences in a single run is limited to 10)

Nb of accepted D-GENE in IGH JUNCTION (default is 1) default ▾

With all alleles With allele *01 only

No

Nb of accepted mutations: default ▾ in 3'V-REGION
default ▾ in D-REGION
default ▾ in 5'J-REGION

Parameters for "Detailed view"

Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 9 and 10)

Nb of nucleotides to add (or exclude) in 3' of the V-REGION for the evaluation of the alignment score (in results 1)

Sequence B1

Sequence: 1 B1

Sequence compared with the human IG set from the [IMGT reference directory](#)

```
>B1
ccatctccggcgacagtgtcttaccacaaactgtgtcttggaaactggatcaggcgatccc
catcgagggctttgtgggtggaaaggacataactacggAAActtcactgtttaaatg
atattgcgttttgtcaaaagggtcgaaatgtcatatatggatccaaagaaacctgt
tctccctggactctgtgtactccccggggacacggctgtgttattactgtgtcaagag
gggtgggtggagatcccgagggcccatggctgtgactactggggccggggaaacctgg
tcacggctttccctcaggta
```

 Nucleotide insertions have been detected and automatically removed for this analysis: they are displayed as capital letters in the user submitted sequence above.

Result summary:

localization in V-REGION	nb of inserted nt	inserted nt	causing frameshift	from V-REGION codon	from nt position in user submitted sequence
CDR2-IMGT	3	AAC	no	60	102

IMGT/V-QUEST results after removal of the insertion(s):

Potentially productive IGH rearranged sequence: no stop codon and in-frame junction
(Check also your sequence with [BLAST](#) against IMGT/GENE-DB reference sequences to eventually identify out-of-frame pseudogenes)

V-GENE and allele	HomsapIGHV6-1*01 F, or HomsapIGHV6-1*02 F	score = 1019	identity = 93.91% (216/230 nt) [93.48% (215/230 nt)]
J-GENE and allele	HomsapIGHJ4*02 F	score = 159	identity = 81.25% (39/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD2-15*01 F		D-REGION is in reading frame 3
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[2, 17, 38, 11]	[10, 9, 18]	CARGVGWEIFPEAHGCDYW

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
J84997 HomsapIGHV6-1*01 F	1019	93.91% (216/230 nt)
Z14223 HomsapIGHV6-1*02 F	1019	93.91% (216/230 nt)
M29812 HomsapIGHV4-59*02 F	437	66.97% (146/218 nt)
AB019438 HomsapIGHV4-59*01 F	428	66.51% (145/218 nt)
N95117 HomsapIGHV4-59*04 F	428	66.51% (145/218 nt)

3nt insertion detected in VH CDR2 region

displayed in capital letters in the submitted sequence and is automatically removed for this analysis

not causing frameshift → Potentially productive IGH rearranged sequence

% identity to germline in brackets, is calculated considering the insertion as one mutational event

Sequence B1

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

			Score	Identity
J04097	Homsap	IGHV6-1*01	F	1019 93.91% (216/230 nt)
Z14223	Homsap	IGHV6-1*02	F	1019 93.91% (216/230 nt)
M29812	Homsap	IGHV4-59*02	F	437 66.97% (146/218 nt)
AB019438	Homsap	IGHV4-59*01	F	428 66.51% (145/218 nt)
M55117	Homsap	IGHV4-59*04	F	428 66.51% (145/218 nt)

Alignment with FR-IMGT and CDR-IMGT delimitations

B1	J04097 Homsap IGHV6-1*01 F	Z14223 Homsap IGHV6-1*02 F	M29812 Homsap IGHV4-59*02 F	AB019438 Homsap IGHV4-59*01 F	M59117 Homsap IGHM4-59*04 F	FR1-IMGT
	caggtagtcacgtcgacgacgtcaggttcca..	caggtagtcacgtcgacgacgtcaggttcca..	caggtagtcacgtcgacgacgtcaggttcca..	caggtagtcacgtcgacgacgtcaggttcca..	caggtagtcacgtcgacgacgtcaggttcca..	caggtagtcacgtcgacgacgtcaggttcca..
	ggacttgtaaagccctcgacgaccctctca	ggacttgtaaagccctcgacgaccctctca	ggacttgtaaagccctcgacgaccctctca	ggacttgtaaagccctcgacgaccctctca	ggacttgtaaagccctcgacgaccctctca	ggacttgtaaagccctcgacgaccctctca

```

B1 .....ccatctccggcgacagtgtctct....accacacgtgtctgtttggggc
J04097 Homsap IGHV6-1*01 F ctcacctgtg.....g.....g.....g
Z14223 Homsap IGHV6-1*02 F ctcacctgtg.....g.....g.....g
M29812 Homsap IGHV4-59*02 F ctcacctgtca-tg---t-t-g-tcc---...-gt---tactac---g
AB019438 Homsap IGHV4-59*01 F ctcacctgtca-tg---t-t-g-tcc---...-gt---tactac---g
M95117 Homsap IGHV4-59*04 F ctcacctgtca-tg---t-t-g-tcc---...-gt---tactac---g

```

```

B1                               FR2-IMGT > CDR
J04097 Homsap IGHV6-1*01 F      tggatcaggcagtcccccatacgagggccctgactggaaaggacatactacggtcc
Z14223 Homsap IGHV6-1*02 F      -
M29812 Homsap IGHV4-59*02 F      -
AB019438 Homsap IGHV4-59*01 F      c-----gg---a-g-----a-t---gtat-tc-t-t...
M95117 Homsap IGHV4-59*04 F      c-----gg---a-g-----a-t---gtat-tc-t-t...

```

```

B1          2-IMGT <----->
J04097 Homsap IGHV6-1*01 F   ...-aactggttaaatgattatgcagtttctgtcaaa...ggtcgaatagtcatctatcca
Z14223 Homsap IGHV6-1*02 F   ....-g---at-----a-----g...a.....ac---a-c-
M29812 Homsap IGHV4-59*02 F   ....-g---at-----a-----g...a.....ac---a-c-
AB019430 Homsap IGHV4-59*01 F   .....g---agc-ccca-c--caaaaa-cc--g...a...-g-cac--a-cagt-
M05117 Human  IGHV4-50*04 F   .....g---agc-ccca-c--caaaaa-cc--g...a...-g-cac--a-cagt-

```

```

B1          ----- FR3-IMGT -----
J04097 Homsap IGHV6-1*01 F   gacacatccaaagaaccaggttctccctggacacctgaaactctgtgactcccgaggacacggct
Z14223 Homsap IGHV6-1*02 F   -c-g-
M29812 Homsap IGHV4-59*02 F   -c-g- -c-g- -c-g- -c-g-t-c- -c
AB019438 Homsap IGHV4-59*01 F   -c-g- -c-g- -c-g- -c-g-t-c- -c
M051117 Human   TGHV4_59*04 F   -c-g- -c-g- -c-g- -c-g-t-c- -c

```

```

B1                               -----> CDR3-IMGT
J04097 Homsap IGHV6-1*01 F      gtgttattactgtgcggaggccgttgggtggagatccggggccatggctgtgactc
Z14223 Homsap IGHV6-1*02 F      -----a
M29812 Homsap IGHV4-59*02 F      -----a
AB013438 Homsap IGHV4-59*01 F      -----a
M95117 Homsap IGHV4-59*04 F      -----a

```

Sequence B1

Assignment of both *01 and *02 alleles of IGHV6-1

From IMGT Repertoires: Alignments of alleles

http://www.imgt.org/IMGTrepertoire/Proteins/taballeles/human/IGH/IGHV/Hu_IGHVall.html

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
J04097	IGHV6-1*01 F 6-1G1/VH6	Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	
		cag	gtt	cag	ctg	cag	cag	tca	gtt	ccs	...	ggg	ctg	gtg	agg	ccc	tcg	cag	acc	ctc	tca
A8019441	IGHV6-1*01 F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X92228	IGHV6-1*01 F VHVI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X14889	IGHV6-1*01 F VH-VI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56378	IGHV6-1*01 F VWHITE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56383	IGHV6-1*01 F VHINW	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Z12374	IGHV6-1*01 F DP-74	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56377	IGHV6-1*01 F VHVIIS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56379	IGHV6-1*01 F VHVIJB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56380	IGHV6-1*01 F VHVICH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56381	IGHV6-1*01 F VHVICH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X56382	IGHV6-1*01 F VHVIIBLK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
X92224	IGHV6-1*01 F 6-1G1/VH6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Z14223	IGHV6-1*02 F VHGL6.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Alleles *01 and *02 differ only the last position of codon VH FR1-9

Sequence B1

IGHV6-1*01

IGHD2-15*01

IGHJ4*02

identity 93.48%

in-frame

VH CDR3 length 16

→ **productive, mutated IGH
rearrangement**

Practical exercises – Solutions – Batch II

>Sequence B2

```
gaggtgcacctggagtctggggaggcctggtaaccgggggtccctcag  
actctcctgtgcagcctctggattcacccatggatcatgacccatgggtccgc  
aggctccagggcagggctggagtgggtctcatccatggagtgggtctcatccatta  
gtatgtatgtatattacataactacgcagactcagtgaaggccgattcaccatct  
ccagagacaacgccaagaactcactgttatctgcaattgaacagcctgagaggcga  
ggacacggctgtttattactgtgcgagaatggtggcttccaagtaccaccatactac  
tttgacttctggggcccggaaaccctggtcaccgtctcctcag
```

Sequence B2

Sequence: 1 B2

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

>B2
gagggtgcacccatgggggtctggggggggcctggtcacccgggggggtccctcagactc
tcctgtcagccctcgatgttccatcttcgtatatttttagatgtacccgtgggtccggccaggct
ccaggcccggggctgggtctcatccatgggtgtgggtctcatccatggatgtatgtatgt
atgtatccatatactacgcagactcaagtgaagggccgattcacatctccagagacaac
gcggaaactactgtatctgcattgtggatcggcggggcggacacggctgttttat
tactgtgcggaaatgggtttccaaatccaccatactactttgtacttgcggggcccg
ggaaacctgtcaccgtctccctcag

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction) (a)		
V-GENE and allele	Homsap IGHV3-21*02 F	score = 490	identity = 63.54% (183/288 nt)
J-GENE and allele	HomsapIGHJ4*02 F	score = 222	identity = 95.83% (46/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD5-12*01 F	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25,17,38,11]	[8,8,16]	CARMMASKYPPYYFDFW

(a) Low V-REGION identity (63.54%): this may indicate potential nucleotide insertion(s) and/or deletion(s). Try 'Search for insertions and deletions' in 'Advanced parameters' at the bottom of the Search page

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
M99658 Homsap IGHV3-21*02 F	490	63.54% (183/288 nt)
AB019439 Homsap IGHV3-21*01 F	481	63.19% (182/288 nt)
AJ879485 Homsap IGHV3-69-1*02 P	475	63.16% (180/285 nt)
KC713940 Homsap IGHV3-11*06 F	472	62.85% (181/288 nt)
HW855323 Homsap IGHV3-21*03 F	472	62.85% (181/288 nt)

Sequence B2

Sequence: 1 B2

Sequence compared with the human Ig set from the IMGT reference directory

Result summary:						
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	FR-IMGT	[25.17.38.11]	[8.8.16]	CARMVASKYPPYYFDFW	FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	FR-IMGT
Localization in V-REGION	nb of inserted nt	inserted nt	causing frameshift	from V-REGION codon	from nt position in user submitted sequence	
FR2-IMGT	18	GAGTGGGTCTCATCCATG	no	51	138	
IMGT/V-QUEST results after removal of the insertion(s):						
Potentially productive IGH rearranged sequence: no stop codon and in-frame junction (Check also your sequence with BLAST against IMGT/GENE-DB reference sequences to eventually identify out-of-frame pseudogenes)						
V-GENE and allele	Homsap [IGHV3-21*01 F, or Homsap [IGHV3-21*02 F]		score = 1318	identity = 95.49% (275/288 nt) [95.14% (274/288 nt)]		
J-GENE and allele	Homsap [IGHJ4*02 F		score = 222	identity = 95.83% (46/48 nt)		
D-GENE and allele by IMGT/JunctionAnalysis	Homsap [IGHD5-12*01 F			D-REGION is in reading frame 1		

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

			Score	Identity
AB019439	Homsap	IGHV3-21*01	F	1318
M98658	Homsap	IGHV3-21*02	F	1318
HW855323	Homsap	IGHV3-21*03	F	1309
HW855688	Homsap	IGHV3-21*04	F	1309
HW855336	Homsap	IGHV3-48*04	F	1246

Sequence B2

Assignment of both *01 and *02 alleles of IGHV3-21

From IMGT Repertoires: Alignments of alleles

http://www.imgt.org/IMGTrepertoire/Proteins/taballeles/human/IGH/IGHV/Hu_IGHVall.html

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20				
AB019439	IGHV3-21*01	F	V-REGION	gDNA	E	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	S	L	R		
					gag	gtg	cag	ctg	gtc	gag	tct	ggg	ggg	...	ggc	ctg	gtc	aag	cct	ggg	ggg	tcc	ctg	aga
X62127	IGHV3-21*01	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
X62129	IGHV3-21*01	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
Z14073	IGHV3-21*01	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
AC245166	IGHV3-21*01	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
M99658	IGHV3-21*02	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
HM855323	IGHV3-21*03	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
HM855688	IGHV3-21*04	F	V-REGION	gDNA	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			

Alleles *01 and *02 differ only the last position of codon VH FR1-3

Sequence B2

Assignment of both *01 and *02 alleles of IGHV3-21

The position where the alleles differ is **mutated** in this rearrangement

6. V-REGION alignment according to the IMGT unique numbering

B2					
AB019439	Homsap	IGHV3-21*01	F		
M99658	Homsap	IGHV3-21*02	F		
HM1855323	Homsap	IGHV3-21*03	F		
HMW55688	Homsap	IGHV3-21*04	F		
HMW55336	Homsap	IGHV3-48*04	F		

B2
AB019439 Homosp IGHV3-21*01 F
M99658 Homosp IGHV3-21*02 F
HW1855323 Homosp IGHV3-21*03 F
HW1855688 Homosp IGHV3-21*04 F
HW1855336 Homosp IGHV3-48*04 F

28 > 25 30
ggg ggg tcc ctc aga ctc tcc tgt gca gcc tct gga ttc acc acc ttc
.....
.....
.....
.....
.....
.....

B2
AB019439 Homsap IGHV3-21*01 F
M99658 Homsap IGHV3-21*02 F
HM855323 Homsap IGHV3-21*03 F
HM855688 Homsap IGHV3-21*04 F
HM855336 Homsap IGHV3-48*04 F

CDR1 - IMGT 35 40 45
agt ast ttt agc atg acc tgg gtc cgc cag gct
"gc -a- --- -a-
"gc -a- --- -a-

```

B2
AB019439 Homsap IGHM3-21*01 F
M99658 Homsap IGHM3-21*02 F
HM855323 Homsap IGHM3-21*03 F
HM855608 Homsap IGHM3-21*04 F
HM855336 Homsap IGHM3-48*04 F

```

B2
AB019439 Homosp IGHM3-21*01 F
M99658 Homosp IGHM3-21*02 F
HM1855323 Homosp IGHM3-21*03 F
HM1855688 Homosp IGHM3-21*04 F
HM1855336 Homosp IGHM3-18*01

```

- IMGT _____ <----- 65 -----> 78 -----> 75
... agt sat tac ata tac tac gca gac tca gtg aag ... ggc cga
... -g- -----> -g- -----> -g- ----->

```

B2
AB019439 Hom sapiens IGHM3-21*01 F
M99658 Hom sapiens IGHM3-21*02 F
HM1855323 Hom sapiens IGHM3-21*03 F
HM1855688 Hom sapiens IGHM3-21*04 F
HM1855326 Hom sapiens IGHM3-21*04A

----- FR3 - INGT -----
88 85 98
ttc acc atc tcc aga gac aac gcc aag aac tca ctg tat ctg cca

B2
AB019439 Homsap IGHM3-21*01 F
M99658 Homsap IGHM3-21*02 F
HM855323 Homsap IGHM3-21*03 F
HM855688 Homsap IGHM3-21*04 F
HM855755 Homsap IGHM3-21*05 F

B2
AB019439 Homosp IGHV3-21*01 F
M99658 Homosp IGHV3-21*02 F
HM855323 Homosp IGHV3-21*03 F
HM855688 Homosp IGHV3-21*04 F

```
aga atg gtg gct tcc aag tac cca cca tac tac ttt gac ttc tgg  
--- ga  
--- ga  
--- ga  
--- ga
```

Sequence B2

4. Results of IMGT/Junction Analysis

Maximum number of accepted mutations in: 3'V-REGION = 2, D-REGION = 4, 5'J-REGION = 2

Maximum number of accepted D-GENE = 1

Analysis of the JUNCTION

D-REGION is in reading frame 1.

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N1	D-REGION	N2	P	5'J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
B2	<u>Homsap</u> IGHV3-21*01	tgtgcgaga..	atggtggt.....	tccsaagtacccacca	t	actactttgacttctgg	<u>Homsap</u> IGHJ4*02	<u>Homsap</u> IGHD5-12*01	0	1	1	8/15

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	110	111	111.1	112.1	112	113	114	115	116	117	118	Frame	CDR3- length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	A	R	I	V	A	S	K	Y	P	P	Y	Y	F	D	E	W					

B2 tgt gcg aga atg gtg gct tcc aag tac cca cca tac tac ttt gac ttc tgg + 15 2,144.51 8.19 CARNWASKYPPVVFDFW

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele '01 for the analysis of the JUNCTION.

Sequence B2

IGHV3-21*01

IGHD5-12*01

IGHJ4*02

identity 95.14%

in-frame

VH CDR3 length 15

→ **productive, mutated IGH
rearrangement**

Practical exercises – Solutions – Batch II

>Sequence B3

gagggtcacctggtagtctgggggagggcctggtaacccggggggccctcag
acttcctgtgcagcctctggattcaccttcagtaatttagcatgacactgggtccgc
aggctccagggcagggctggagtgggtctcatccatggagtgggtctcatccatta
gtagtagtagtaattacatatactacgcagactcagtgaagggccgattcaccatct
ccagagacaacgccaagaactcactgttatctgcaattgaacagcctgagaggcga
ggacacggctgtttattactgtgcgagaatggtggtccaaagtacccaccatactac
tttgacttctggggcccggAACCTGGTCAACGTCTCAG

Sequence B3

Sequence: 1 B3

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

>B3

```
gtgggttccttcagtggtaactatgggacttggatccgtcagcccccaaaaaaaatgg  
tgtggattggggaaaataatcatatgtggactgggggggggggggggggggggggggg  
tcgtcatatgtggaaactgcgaaactacaacccgtcccttcaggatgtcgaggcacatata  
tagacactgtccaaagaaaccgggttttttttttttttttttttttttttttttttttt  
ctgtatataactgtgggggggggggggggggggggggggggggggggggggggggggg  
gccggggaaacctggtcatcgatcttcctcag
```

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction) (a)		
V-GENE and allele	HomsapIGHV4-34*13 F	score = 547	identity = 73.24% (156/213 nt)
J-GENE and allele	HomsapIGHJ4*02 F	score = 188	identity = 87.50% (42/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD1-28*01 F	D-REGION is in reading frame 3	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[14.17,38.11]	[8.7.14]	CARGLQVGGTYSLDYW

(a) Low V-REGION identity (73.24%); this may indicate potential nucleotide insertion(s) and/or deletion(s). Try 'Search for insertions and deletions' in 'Advanced parameters' at the bottom of the Search page

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
Z75356 HomsapIGHV4-34*13 F	547	73.24% (156/213 nt)
Z75351 HomsapIGHV4-30-2*04 F	502	70.89% (151/213 nt)
Z75347 HomsapIGHV4-61*06 ORF	484	69.95% (149/213 nt)
Z75355 HomsapIGHV4-4*06 F	478	69.01% (147/213 nt)
Z75353 HomsapIGHV4-30-4*05 F	475	69.16% (148/214 nt)

Sequence B3

Sequence: 1 B3

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

 Nucleotide insertions have been detected and automatically removed for this analysis: they are displayed as capital letters in the user submitted sequence above.

localization in V-REGION	nb of inserted nt	inserted nt	causing frameshift	from V-REGION codon	from nt position in user submitted sequence
CDR2-IMGT	45	CTAGGGAAAGGGCTGGAGTGATTGCGAAAATCAGTCATAGTGAA	no	84	90

IMGT/V-QUEST results after removal of the insertion(s):

Potentially productive IGH rearranged sequence: no stop codon and in-frame junction
(Check also your sequence with BLAST against IMG/GENE-DB reference sequences to eventually identify out-of-frame pseudogenes)

V-GENE and allele	<u>HomsapIGHV4-34*01 F, or HomsapIGHV4-34*02 F</u>	score = 898	identity = 92.34% (193/209 nt) [91.87% (192/209 nt)]
J-GENE and allele	<u>HomsapIGHJ4*02 E</u>	score = 188	identity = 87.50% (42/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	<u>HomsapIGHD1-28*01 F</u>	D-REGION is in reading frame 3	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION [X,17,38,11]		[7,7,14]	CARGLQVGGTYSLDYW

1. Alignment for V-GENE and allele identification

Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

			Score	Identity
AB019439	Homsap	IGHV4-34*01	F	896 92.34% (193/209 nt)
M99684	Homsap	IGHV4-34*02	F	896 92.34% (193/209 nt)
X22255	Homsap	IGHV4-34*03	F	887 91.83% (181/208 nt)
M95113	Homsap	IGHV4-34*08	F	887 91.87% (192/209 nt)
X56591	Homsap	IGHV4-34*12	F	887 91.87% (192/209 nt)

This rearrangement is productive.

The insertion does not cause frameshift, but its length (45nt) may render it non functional

Sequence B3

IGHV4-34*01

IGHD1-26*01

IGHJ4*02

identity 91.87%

in-frame

VH CDR3 length 14

→ **productive, mutated IGH
rearrangement**

Practical exercises – Solutions – Batch II

>Sequence B4

```
gaggaggagggtttggagtctgggggaggcttggtacagcctgggggtccctgag  
actctcctgtgttagcctctggattcaccttagtaattatgtcatgagttggtccgcc  
aggctccagggagggactggagtggtctccgttattactgatagtggtggtcata  
catactatgcagactccgtgaagggccggttaccatctccagagacaattccaaga  
acaccctatatcttcaaatacagcctgagactcgaggacacggccagatattact  
gtgcgagagactgggggatattgtgggggaggcaccctggtcaccgtctccaca  
t
```

Sequence B4

Sequence: 1 B4

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>B4
gaaaaaggaggtttggaggtctggggggggctggatcacggctgggggtccctgagactc
tcctgtgtggctctggattcaccttttagtaattatgtcatgagttgggtccggcaggct
ccagggggggggactgggggtctcggttattactgtatgtgggtgtcatacatactat
gcggactccgtgaaggggggggtcaccatctccagagacaaattccaaageacaccctatat
cttcggatgtggccgggggggggggggggggggggggggggggggggggggggggggggggg
ggggatattgtgggggggggggggggggggggggggggggggggggggggggggggggggggg
```

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV3-23*01 F, or Homsap IGHV3-23D*01 F	score = 1210	identity = 91.32% (263/288 nt)
J-GENE and allele	HomsapIGHJ5*01 F	score = 100	identity = 68.09% (32/47 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD7-27*01 F	D-REGION is in reading frame 3	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25, 17, 38, 11]	[8, 8, 7]	CARDWGDIV (TRP 118 not identified)

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
M99660 Homsap IGHV3-23*01 F	1210	91.32% (263/288 nt)
AC244492 Homsap IGHV3-23D*01 F	1210	91.32% (263/288 nt)
AJ879486 Homsap IGHV3-23*04 F	1201	90.97% (262/288 nt)
AC245166 Homsap IGHV3-23D*02 F	1201	90.97% (262/288 nt)
M35415 Homsap IGHV3-23*02 F	1192	90.62% (261/288 nt)

The 2nd VH CDR3 anchor is
not identified and the VH
CDR3 length is unusually
short

Sequence B4

4. Results of IMGT/Junction Analysis

Maximum number of accepted mutations in: 3'V-REGION = 2, D-REGION = 4, 5'J-REGION = 2
Maximum number of accepted D-GENE = 1

Analysis of the JUNCTION

D-REGION is in reading frame 3.

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N1	D-REGION	N2	5'J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
B4	Homsap IGHV3-23*01	tgtgcgagagactgggg.	ggatatgt	g	Homsap IGHJ5*01	HomsapIGHD7-27*01	1	0	0	3/9

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	115	116	117	118	Frame	CDR3- IMGT length	Molecular mass	pI	<u>PhysicoChemical Descriptor (by BRFAA)</u>
C	A	R	D	K	G	D	I	V	+ 7	1,034.16	4.44	CARDWGOIV	

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

Non identification of W118 is owed to extreme 5'-3' exonuclease activity

Sequence B4

The amplification reaction should be repeated using different sets of primers in an attempt to amplify a second rearrangement.

However, this mutated codon 118 is followed by G-Q-G motif (IMGT positions 119-121) in the FR4 encoded by the IGHJ5 gene,

The rearrangement is assumed to be productive and can be used for interpretation of the mutational status.

7. V-REGION translation

84

M99660 Homsap IGHV3-23*01 F

84

M99660 Hom sapiens IGHV3-23*01 F

84

M99660 Homsap IGHV3-23*01 F

三

M99660 Homsap IGHV3-23*01 F

34

M99660 Hom sapiens IGHV3-23*01 F

84

M99660 Homsap 1GHV3-23*01 F

14

Message Number 10HV3-23-01 F

10

1995000 Nonstop 101105-25 01

Sequence B4

IGHV3-23*01 or IGHV3-23D*01

IGHD7-27*01

IGHJ5*01

identity 91.32%

in-frame

VH CDR3 length 7

→ **productive, mutated IGH
rearrangement**

Practical exercises – Solutions – Batch II

>Sequence B5

```
gaggtcagctggagtctggggaggctggccagccgggggtccctgag  
acttcctgtgcagcctctggattcacctaataggtaggttggatgaggttgggtccgcc  
aggctccaggaaaggggctggagtgggtggccaacataaaggaagatggaagtg  
agaaaatactatgtggactctgtgaaggccgattcaccatctccagagacaacgac  
aagaactcacttatctgcaaatacgtctgagatcgaagacacggccgtgt  
ctactgtgcgagagtcctttcttggagaggttattgggggggtcactggttcgacc  
cccgccccagggAACCT
```

Sequence B5

Sequence: 1 B5

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

>B5
gggtgcagctgtggagtctggggggggcttgggtccggccgggggggtccctggggactc
tcctgtgcagccctctggattcacccatataggatattggatgggtgggtccggccaggct
ccggggaggggctggggatggggatggccacataaaggggatggggatggggatggggat
gtggactctgtggggccgattcacatctccggggatggggatggggatggggatggggat
ctggggatggggatggggatggggatggggatggggatggggatggggatggggatggggat
ttctttggggatggggatggggatggggatggggatggggatggggatggggatggggat

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV3-7*03 F	score = 1345	identity = 96.53% (278/288 nt)
J-GENE and allele	HomsapIGHJ5*02 F	score = 134	identity = 88.24% (30/34 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD3-3*01 F	D-REGION is in reading frame 3	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.5]	[8.8.18]	CARVLFFGEVIWGHHWFDP (TRP 118 not identified)

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
MW55666 Homsap IGHV3-7*03 F	1345	96.53% (278/288 nt)
M99649 Homsap IGHV3-7*01 F	1336	96.18% (277/288 nt)
X92288 Homsap IGHV3-7*02 F	1327	95.83% (276/288 nt)
M99678 Homsap IGHV3-52*01 P	1129	88.19% (254/288 nt)
Z17388 Homsap IGHV3-52*02 P	1120	87.85% (253/288 nt)

The 2nd VH CDR3 anchor is
not identified

Sequence B5

4. Results of IMGT/JunctionAnalysis

Maximum number of accepted mutations in: 3'V-REGION = 2, D-REGION = 4, 5'J-REGION = 2
Maximum number of accepted D-GENE = 1

Analysis of the JUNCTION

D-REGION is in reading frame 3.

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N1	D-REGION	N2	5'J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
B5	Homsap IGHV3-7*03	tgtgcgagag.	tccttttctttggagggttatt....	tggggggggtc	...actggttcgaccccggs	HomsapIGHJ3*02	HomsapIGHD3-3*01	0	1	1	11/18

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	110	111	111.1	111.2	112.3	112.2	112.1	112	113	114	115	116	117	118	Frame	CDR3- IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	A	R	V	L	F	F	G	E	V	I	W	G	G	H	W	F	D	P	R					

B5 tgt gcg agg gtc ctt ttc ttt agg gag gtt att tgg agg ggt cac tgg ttc gac ccc agg + 18 2,392.77 7.18 CARVLFFGEVINGGGHNFDFPR

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele '01 for the analysis of the JUNCTION.

Non identification of W118 is owed to SHM activity

Sequence B5

7. V-REGION translation

B5

HM1855666 Hom sapien IGHV3-7*03 F

<----- FR1 - IMGT ----->
1 E V Q L V E S G G 10 G L V Q P
gag gtc cag ctc gtc gag tct ggg gga ... ggc ttg gtc cag ccg

20 G G S L R L S C A S G F T P
ggg ggg tcc ctc aca ctc tgt gca gcc tct gga ttc acc ttt

— CDR1 - IMGT ————— <-----
35 N R Y W M S W V R Q A 45
aat agg tat tgg atg agt tgg gtc cgc cag gct
S S

FR2 - IMGT -----> CDR2
50 P G K G L E W V A N I K E D 60
cca ggg aag ggg ctc gag tgg atg gcc aac ata aag gaa gat ...
Q

— IMGT ————— <-----
65 G S E K Y Y V D S V K G R 75
... gga agt gag aca tac tat gtg gac tct gtg aag ... ggc cga

----- FR3 - IMGT ----->
80 F T I S R D N D K N S L Y L Q 90
ttc acc aac tcc aca gac aac gac aac tcc ctc tat ctg cca
A

95 M N S L R V E D T A V Y Y C A 100 104
atg aac agt ctg aca gtc gaa gac aca gcc gtc tac tac tat gca
A

— CDR3 - IMGT —————
R V L F F G E V I N G G H W F
aga gtc ctt ttc ttt gga gag gtt att tgg egg ggt cac tgg ttc
--- -a
D P R G Q G T
gac ccc cgg ggc cag gga acc ct

The VH FR4 GXG motif is present in the rearrangement

The rearrangement is assumed to be productive and can be used for interpretation of the mutational status.

Sequence B5

IGHV3-7*03

IGHD3-3*01

IGHJ5*02

identity 96.53%

in-frame

VH CDR3 length 18

→ **productive, mutated IGH
rearrangement**

Practical exercises – Solutions – Batch II

>Sequence B6

```
ctgtgttacgcagccgcctctgcatcagcctccctggagcctcggtcacactcac  
ctgcaccctgagcagcgactaaagtaattataaagtggactggtagccaagagagac  
cagggcagggcccccagttgtatgcgagtggcactggtaggtggatcc  
aaggggatggcatccctgatcgcttcagtctggactcaggcctgagtcggttgc  
tgaccatcaagaacatccaagaagaggatgagagtgactaccactgtgggcaga  
ccatggcgatggaaacaactcgtccctgggtttaggcggagggaccaagctga  
ccgtcctag
```

Sequence B6

Sequence: 1 B6

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>B6
ctgtgttgcgcagccgccttcgtcatcaggcccggggggccctcggtcacactcacct
gcacccctgagcgcgactaaatataaaatggactgggtaccaaagagagaccaggcc
ggggcccccaagtttgtgtatgcgactggggactggggatttgtggatccaagggggatg
gcatccctgatcgcttcagtctggactcaggcctgagtgccgtgtggatccatcaga
atcccaaaaaggaggatggagactaccactgtggggcagaccatggcgatggggaca
acttcgtccccctgggtttggcggggaccsaagctgaccgtccctag
```

Result summary:	Unproductive IGL rearranged sequence (stop codons)		
V-GENE and allele	Homsap IGLV9-49*01 F	score = 1316	identity = 94.54% (277/293 nt)
J-GENE and allele	Homsap IGLJ3*02 F	score = 172	identity = 94.74% (38/38 nt)
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[23.17.38.10]	[7.8.14]	CGADHGDGNNPVPWVL (PHE 118 not identified)

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon plus 18 nt of the CDR3-IMGT)

	Score	Identity
Z73675 Homsap IGLV9-49*01 F	1316	94.54% (277/293 nt)
D87016 Homsap IGLV9-49*02 F	1307	94.20% (276/293 nt)
U03869 Homsap IGLV9-49*03 F	1307	94.20% (276/293 nt)
X57828 Homsap IGLV4-3*01 F	527	64.83% (188/290 nt)
Z73648 Homsap IGLV4-69*01 F	482	62.76% (182/290 nt)

The 2nd VL CDR3 anchor is
not identified

Sequence B6

4. Results of IMGT/JunctionAnalysis

Maximum number of accepted mutations in: 3'V-REGION = 7, 5'J-REGION = 7
Maximum number of accepted D-GENE = 0

Analysis of the JUNCTION

Click on mutated (underlined) nucleotide to see the original one:

Input	V name	3'V-REGION	N	5'J-REGION	J name	Vmut	Jmut	Ngc
B6	Homsap IGLV9-49*01	tgtggggcagaccatggcgatggggacaacttcgt.....	cccc	.tgggtgtta	Homsap IGLJ3*02	3	1	4/4

Translation of the JUNCTION

Click on mutated (underlined) amino acid to see the original one:

104	105	106	107	108	109	110	111	112.1	112	113	114	115	116	117	118	Frame	CDR3- IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
C	G	A	D	H	G	D	G	N	N	F	V	P	W	V	L					

B6 tgt ggg gca gac cat ggc gat ggg gac gac ttc gtc ccc tgg gtg tta + 14 1,700.85 4.43 CGADHGDNNFVPWVL

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

Non identification of F118 is owed to SHM activity

Sequence B6

7. V-REGION translation

The rearrangement is rendered unproductive due to the presence of a stop codon in position VL CDR1-29

B6
Z73675 Hom sapiens IGLV9-49*01 F

FR1 - IMGT
1 V L T Q P P S 10 A S A S L
... .ct gtg ttg acg cag ccg cct tct ... gca tca gcc tcc ctg
Q P
cag c--- c--- t--- a--- ... ---
-----> ----->
20 G A S V T L T C T L S S D * 30
ggg gcc tcg gtc aca ctc acc tgc acc ctg agc agc gag taa act
G Y
----->
----->
CDR1 - IMGT
35 N Y K V D W Y Q E R
... aat tat aac gtg gag tgg tac caa gag aga
Q
----->
----->
FR2 - IMGT
50 P G Q G P Q F V H R V G T G 60
cca ggg cag ggc ccc cag ttt gtg atg cga gtg ggc act ggt ...
K R
----->
----->
- IMGT
65 G I V G S K G D G I P D R 75
... ggg att gtg gga tcc aag ggg gat ggc atc cct ... gat cgc
----->
----->
FR3 - IMGT
80 F S V L D 85 S G L S R L L T 90
ttc tca gtc ttg gac ... tca ggc ctg atg cgg ttg ctg acc
G N Y
----->
----->
95 I K N I Q E E D E S D Y H C G 100 104
atc aag aac atc caa gaa gag gat gag atg gac tac cac tgt ggg
----->
----->
CDR3 - IMGT
A D H G D G N N F V P W V L G
gca gac cat ggc gat ggg aac aac ttc gtc ccc tgg atg tta ggc
S S +
----->
----->
G G T K L T V L
ggg ggg acc aac gtc acc gtc ctc g

B6
Z73675 Hom sapiens IGLV9-49*01 F

Sequence B6

IGLV9-49*01

IGLJ3*02

identity 94.74%

in-frame

VH CDR3 length 18

→ unproductive, mutated IGL
rearranged sequence

Practical exercises – Solutions – Batch II

>Sequence B7

```
cccagactccactctccttgttatcacccctggagagcaggcctccatgtcctgcag  
gtctagtcagagcctcctgcatagtgatggatacacctattgtattggttctgcaga  
aagccaggccagtctccacagcctgatctatgaagttccaaccgggtctctggag  
tgccagataggttcagtggcagcgggtcagggacagattcacactgaaaatcagc  
cggtggaggctgaggatgtggagtttattactgcatgcaagatgcacaagatcct  
cgcttcggcggagggaccaagctggagatcaaacgt
```

Sequence B7

Sequence: 1 B7

Sequence compared with the human Ig set from the IMGT reference directory

Result summary:	Unproductive IGK rearranged sequence (out-of-frame junction) (a)		
V-GENE and allele	Homsap IGKV2-29'03 F, or Homsap IGKV2D-29'02 F	score = 523	identity = 64.89% (183/282 nt)
J-GENE and allele	Homsap IGKJ4'01 F (b)	score = 148	identity = 91.43% (32/35 nt)
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[22,17,36,10]	[11.3,X]	CMQDAQDPR#F

(a) Low V-REGION identity (84.89%): this may indicate potential nucleotide insertion(s) and/or deletion(s). Try 'Search for insertions and deletions' in 'Advanced parameters' at the bottom of the Search page.

Other possibilities: Homsap IGKJ2*01 and Homsap IGKJ2*02 (highest number of consecutive identical nucleotides).

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon plus 15 nt of the CDR3-IMGT)

		Score	Identity		
AJ783437	Homsap	IGKV2-29*03	F	523	64.89% (183/282 nt)
U41644	Homsap	IGKV2D-29*02	F	523	64.89% (183/282 nt)
X63396	Homsap	IGKV2-29*01	P	514	64.54% (182/282 nt)
U41645	Homsap	IGKV2-29*02	F	514	64.54% (182/282 nt)
M31952	Homsap	IGKV2D-29*01	F	514	64.54% (182/282 nt)

Sequence B7

Sequence: 1 B7

Sequence compared with the human Ig set from the IMGT reference directory

 Nucleotide insertions have been detected and automatically removed for this analysis: they are displayed as capital letters in the user submitted sequence above.

localization in V-REGION	nb of inserted nt	inserted nt	causing frameshift	from V-REGION codon	from nt position in user submitted sequence
FR2-IMGT	1	A	yes	51	137

IMGT/V-QUEST results after removal of the insertion(s):

Unproductive IGK rearranged sequence (out-of-frame junction)

(Check also your sequence with BLAST against IMG/GENE-DB reference sequences to eventually identify out-of-frame pseudogenes)

V-GENE and allele	Homsap IGKV2D-28*03 F	score = 1391	identity = 99.64% (280/281 nt) [99.29% (279/281 nt)]
J-GENE and allele	Homsap IGKJ4*01 F (a)	score = 148	identity = 91.43% (32/35 nt)
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[21.17.38.10]	[11.3.X]	CMQDAQDPR#F

Other possibilities: Homsao_IGK-I2*01 and Homsao_IGK-I2*02 (highest number of consecutive identical nucleotides)

Repeat/check sequencing reaction/output to decide whether the insertion is true

If this is the single rearrangement amplified repeat the amplification with different sets of primers and/or try to amplify a lambda chain rearrangement in order to:

- (i) confirm the presence of the unproductive rearrangement
 - (ii) look for the presence of a second productive rearrangement

Sequence B7

IGKV2D-26*03

IGKJ4*01

identity 99.29%

out-of-frame

VH CDR3 length X

→ unproductive, unmutated IGK
rearranged sequence

Practical exercises – Solutions – Batch II

>Sequence B8

```
cagctgcagctgcaggagtccggcccaggactggcgagaccctgtccctcacctg  
cactgtctctggctccatcagcttatcgattactactggggctggatacgccag  
cccccaggaaaggggctggagtggattggagtgcttattatagtggagcaccta  
ctacaatccgtccctcaagagtcgagtcaccatgtccgtcgacacgtccaagaacca  
gttctccctgaagttgagctctgtgaccgccgcagacacggctgtgtattactgttg  
agacttccgggtatTCAGTGCCACAATACTACACTGTGGACGTCTGGGCAAGGG  
ACCACGGTCACCGTCTCCTCA
```

Sequence B8

Sequence: 1 B8

Sequence compared with the human IG set from the IMGT reference directory

```
>B8
cagctgcagctgcaggagtcggggccaggactggcgggagaccctgtccccctcacctgca
gtctctggatggccatcagcttatcgattactactggggctggatacgcggccccc
ggggaggggctggatggatggaggatgtcttatatagtggggacacctactacatccg
tccctcaagagtgcggatgtccgtcgacacgtccaaagaaccgttccctgaag
ttggatctgtggccggccggccggatgtgttatatactgtgtggacttccgggtatt
tcgggtccacaatactacactgtggatgtctggggccaaaggaccacggtcaccgtctcc
tca
```

Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV4-39*01 F	score = 1063	identity = 86.52% (244/282 nt)
J-GENE and allele	HomsapIGHJ6*02 F (a)	score = 193	identity = 79.03% (49/62 nt)
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD6-19*01 F	D-REGION is in reading frame 2	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[22.17.38.11]	[10.7.16]	CVRLPGISVPQYYTVDNW

(a) Other possibilities: Homsap_IGHJ3*01 (shorter alignment but highest percentage of identity)

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
A8019439 Homsap IGHV4-39*01 F	1063	86.52% (244/282 nt)
X05715 Homsap IGHV4-39*02 F	1054	86.17% (243/282 nt)
X92259 Homsap IGHV4-39*03 F	1054	86.12% (242/281 nt)
M95116 Homsap IGHV4-39*05 F	1054	86.17% (243/282 nt)
A01940222 Homsap IGHV4-39*07 F	1036	85.46% (241/282 nt)



**The rearrangement shows
low V-gene identity but
doesn't activate the IMGT
warning**

Sequence B8

Alignment with FR-IMGT and CDR-IMGT delimitations

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<----- FR1-IMGT ----->
cagctgcagctgagg...tcggccaggactggcgagacccgtcc
cagctgcag-t-a-g-tc-gcccc...ggacttgtgaag-ctt
cagctgcag-t-a-g-tc-gcccc...ggacttgtgaag-ctt
cagctgcag-t-a-g-tc-gcccc...ggacttgtgaag-ctt
cagctgcag-t-a-g-tc-gcccc...ggacttgtgaag-ctt
cagctgcag-t-a-g-tc-gcccc...ggacttgtgaag-ctt

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<-----> CDR1-IMGT <-----
ctcacctgcactgtctctgtggctccatcagc.....tctatcgattactggggc
.....ag--gtag.....ag--gtag.....ag--gtag.....ag--gtag.....ag--gtag

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<----- FR2-IMGT -----> CDR
tggatacgcagccccagggaaaggggctggagtggattggagtgcttattatagt...
-----C-----a
-----C-----a
-----C-----a
-----C-----a
-----C-----a

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<----- 2-IMGT ----->
.....ggagcacctactacaatccgtccctcaag...agtgcagtcaccatgtccgtc
.....C.....a.....a.....a.....a.....a.....a.....a.....a.....a.....a

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<----- FR3-IMGT ----->
gacacgtccaagaaccagtctccctgaagttgagctctgtgaccggccagacacggct
.....C.....C.....C.....C.....C.....C.....g.....C

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

<-----> CDR3-IMGT <-----
gtgttattactgtgtgagactccgggtatccatgtgccacaatactacactgtggacgtc
.....C----a
.....C----ga
.....C
.....C----ga

Poor quality alignment at the beginning of VH FR1 region resulting in mutational imbalance compared to the remaining V-region

Enable the “search for indels” option

Sequence B8

Sequence: 1 B8

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>B8
cagctcgaggctgcaggactggccaggactggcgagaccctgtccctcacctgcact
gtctctgggtggctccatcagcttatcgattacttggggctggatacgccagccccca
gggaaggggctggagttggggatgtctattatagtgggacacctacaatccg
tccctcaagagtgcaggatcaccatgtccgtcgacacgtccaagaaccaggatctccctgaag
ttgagctgtgaccgcgcagacacggctgtgtattactgtgtgagacttccgggtatt
tcagtgccacaataactacactgtggacgtctggggcaagggaccacggtaccgtctcc
tca
```

Result summary:	 Nucleotide deletions have been detected (shown by dots in the alignments):				
	localization	nb of deleted nt	causing frame shift	from V-REGION codon	from nt position in user submitted sequence
	FR1-IMGT	9	no	13	34
IMGT/V-QUEST results after filling the deletion(s) gap(s):					
Potentially productive IGH rearranged sequence : no stop codon and in-frame junction (Check also your sequence with BLAST against IMGT/GENE-DB reference sequences to eventually identify out-of-frame pseudogenes)					
V-GENE and allele	HomsapIGHV4-39*01 F	score = 1288	identity = 95.39% (269/282 nt)	[95.04% (268/282 nt)]	
J-GENE and allele	HomsapIGHJ6*02 F (a)	score = 193	identity = 79.03% (49/62 nt)		
D-GENE and allele by IMGT/JunctionAnalysis	HomsapIGHD6-19*01 F	D-REGION is in reading frame 2			
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[22.17.38.11]	[10.7.16]	CVRLPGISVPQYYTVVDWV		

(a) Other possibilities: [Homsap_IGHJ3*01](#) (shorter alignment but highest percentage of identity)

1. Alignment for V-GENE and allele identification

Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

	Score	Identity
AB019439 Homsap IGHV4-39*01 F	1288	95.39% (269/282 nt)
X05715 Homsap IGHV4-39*02 F	1279	95.04% (268/282 nt)
X92259 Homsap IGHV4-39*03 F	1279	95.02% (267/281 nt)
M95116 Homsap IGHV4-39*05 F	1279	95.04% (268/282 nt)
AM940222 Homsap IGHV4-39*07 F	1261	94.33% (266/282 nt)

Sequence B8

Gene assignment is not altered

The V-gene identity score and the alignment is improved

The 9nt deletion is shown in the alignment

Alignment with [FR-IMGT](#) and [CDR-IMGT](#) delimitations

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

```
cagctgcagctgcaggagtggggccca...ggactg.....gtgaagcctt  
-----  
-----  
-----  
-----  
-----  
-----
```

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

-----> CDR1-IMGT -----<
ctcacctgcactgtcttggctccatcagc.....tctatcgattactactggggc
.....ag--gtag
.....ag--gtag
.....ag--gtag
.....ag--gtag
.....ag--gtag

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

----- FR2-IMGT -----> _____ CDR
tggataccgcagccccccagggaaggggctggagtggattggagtgctattatag...
----C---- a
----C---- a
----C---- a
----C---- a
----C---- a

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

2-IMGT _____<-----
.....gggagcacctactacaatccgtcccctcaag...agtgcagtcaccatgtccgtc
.....-C-.....a.....a
.....-C-.....a.....a
.....-C-.....a.....a
.....-C-.....a.....a
.....-C-.....a.....a

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

----- FR3-IMGT -----
gacacgtccaaagaaccaggttctccctgaagttgagctctgtgaccggccgcagacacggct
----- C -----
----- C -----
----- C ----- C
----- C -----
----- C ----- C

B8
AB019439 Homsap IGHV4-39*01 F
X05715 Homsap IGHV4-39*02 F
X92259 Homsap IGHV4-39*03 F
M95116 Homsap IGHV4-39*05 F
AM940222 Homsap IGHV4-39*07 F

-----> CDR3-IMGT
gtgttattactgtgtgagacttccgggtatTCAGTGCCACAATACTACACTGTGGACGTC
-----C-----a
-----C-----ga

-----C-
-----C-----ga

Sequence B8

IGHV4-39*01

IGHD6-19*01

IGHJ6*02

identity 95.04%

in-frame

VH CDR3 length 16

→ **productive, mutated IGH
rearrangement**