

Protocol

IMGT/V-QUEST: IMGT Standardized Analysis of the Immunoglobulin (IG) and T Cell Receptor (TR) Nucleotide Sequences

Véronique Giudicelli, Xavier Brochet, and Marie-Paule Lefranc¹*IMGT, the international ImMunoGeneTics information system, Laboratoire d'ImmunoGénétique Moléculaire LIGM, Université Montpellier 2, Institut de Génétique Humaine IGH, UPR CNRS 1142, 34396 Montpellier cedex 5, France*

INTRODUCTION

IMGT/V-QUEST is the integrated online IMGT tool for the standardized analysis of the immunoglobulin (IG) or antibody and T cell receptor (TR) rearranged nucleotide sequences. The analysis of these antigen receptors represents a crucial challenge for the study of the adaptive immune response in normal and disease-related situations. The expressed IG and TR repertoires represent a potential of 10^{12} IG and 10^{12} TR per individual. This huge diversity results from mechanisms that occur at the DNA level during IG molecular synthesis, including the combinatorial rearrangement of the variable (V), diversity (D), and joining (J) genes, the N diversity (V, D, and J exonuclease trimming and addition at random of nucleotides by the terminal deoxynucleotidyl transferase during the V-(D)-J rearrangement) and, for IG, somatic hypermutations. IMGT/V-QUEST, as described here, identifies the V, D, J genes and alleles by alignment with the germline IG and TR gene and allele sequences of the IMGT reference directory. The tool describes the V-REGION mutations and identifies the hot spot positions in the closest germline V gene. IMGT/V-QUEST integrates IMGT/JunctionAnalysis for a detailed analysis of the V-J and V-D-J junctions and IMGT/Automat for a complete sequence annotation of the sequences, and provides IMGT Collier de Perles.

RELATED INFORMATION

IMGT/V-QUEST (Lefranc 2004; Giudicelli and Lefranc 2005, 2008; Brochet et al. 2008) is part of IMGT, the international ImMunoGeneTics information system, <http://www.imgt.org> (Lefranc et al. 2009). Standardization and IMGT Scientific chart rules are based on the IMGT-ONTOLOGY concepts of identification, classification, description (Giudicelli and Lefranc 1999; Lefranc et al. 2004, 2005a, 2008), and numerotation (Lefranc 1997, 1999; Lefranc et al. 2003, 2005b,c) generated from the axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope (Duroux et al. 2008). The concepts of identification led to the IMGT standardized keywords, the concepts of classification to the IMGT standardized gene and allele names, the concepts of description to the IMGT standardized labels, and the concepts of numerotation to the IMGT unique numbering.

A detailed description of IMGT is provided in **IMGT, the International ImMunoGeneTics Information System** (Lefranc 2011a). Information is also available on **From IMGT-ONTOLOGY IDENTIFICATION Axiom to IMGT Standardized Keywords: For Immunoglobulins (IG), T Cell Receptors (TR), and Conventional Genes** (Lefranc 2011b), **From IMGT-ONTOLOGY DESCRIPTION Axiom to IMGT Standardized Labels: For Immunoglobulin (IG) and T Cell Receptor (TR) Sequences and Structures** (Lefranc 2011c), **From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT Standardized Gene and Allele Nomenclature: For Immunoglobulins (IG) and T Cell Receptors (TR)** (Lefranc 2011d), **IMGT Unique Numbering for the Variable (V), Constant (C), and Groove (G) Domains of IG, TR, MH, IgSF, and MhSF** (Lefranc 2011e), and **IMGT Collier de Perles for the Variable (V), Constant (C), and Groove (G) Domains of IG, TR, MH, IgSF, and MhSF** (Lefranc 2011f).

In addition, protocols are available for **IMGT/JunctionAnalysis: IMGT Standardized Analysis of the V-J and V-D-J Junctions of the Rearranged Immunoglobulins (IG) and T Cell Receptors (TR)** (Giudicelli and Lefranc 2011), **IMGT/Collier de Perles: IMGT Standardized Representation of Domains (IG, TR, and IgSF Variable and Constant Domains, MH and MhSF Groove Domains)** (Ehrenmann et al. 2011), **IMGT/DomainGapAlign: IMGT Standardized Analysis of Amino Acid Sequences of**

¹Corresponding author (Marie-Paule.Lefranc@igh.cnrs.fr).

Cite as: Cold Spring Harb Protoc; 2011; doi:10.1101/pdb.prot5633

www.cshprotocols.org

Variable, Constant, and Groove Domains (IG, TR, MH, IgSF, MhSF) (Ehrenmann and Lefranc 2011a), and **IMGT/3Dstructure-DB: Querying the IMGT Database for 3D Structures in Immunology and Immunoinformatics (IG or Antibodies, TR, MH, RPI, and FPIA)** (Ehrenmann and Lefranc 2011b).

MATERIALS

It is essential that you consult the appropriate Material Safety Data Sheets and your institution's Environmental Health and Safety Office for proper handling of equipment and hazardous materials used in this protocol.

Equipment

Computer (Internet-connected)

METHOD

1. Using any modern web browser, access the IMGT Home page, <http://www.imgt.org> and click, in the "IMGT tools" section, on the link to IMGT/V-QUEST.

This gives access to the IMGT/V-QUEST Welcome page.

IMGT/V-QUEST Welcome Page

2. In the IMGT/V-QUEST Welcome page (Fig. 1), locate the following sections:

- i. "Analyze your immunoglobulin (IG) or antibody nucleotide sequences"

This section shows the species for which the analysis of IG sequences is available.

There is no need to specify more precisely the type of IG chain or locus, as the tool will automatically identify whether sequences are IGH, IGK, or IGL by selecting one species or taxon in that section. This works for species with complete IMGT reference directories, such as human (Lefranc 2000a; Lefranc and Lefranc 2001a) and mouse (Giudicelli et al. 2005a). This also holds for the other species or taxons with incomplete IMGT reference directories; however, in those cases, results should be analyzed considering the status of the IMGT reference directory (information on the updates on the web site).

- ii. "Analyze your T cell Receptor (TR) nucleotide sequences"

This section shows the species for which the analysis of TR sequences is available.

There is no need to specify more precisely the type of TR chain or locus, as the tool will automatically identify whether sequences are TRA, TRB, TRG, or TRD by selecting one species or taxon in that section. This works for species with complete IMGT reference directories, such as human (Lefranc 2000b; Lefranc and Lefranc 2001b) and mouse (Giudicelli et al. 2005a). This also holds true for the other species or taxons with incomplete IMGT reference directories; however, in those cases, results should be analyzed considering the status of the IMGT reference directory (information on the updates on the web site).

- iii. "Download IMGT reference directory in FASTA format (IG and TR)"

This section shows the species for which the IMGT reference directory is available.

Clicking on one of the links gives access to the data (overview and IMGT/V-QUEST reference directory per species or taxon), if needed.

3. Choose the IG or TR section (see Step 2), depending on your sequences.
4. Click on the species (e.g., "Human") or taxon that corresponds to the IMGT reference directory against which you want your sequences to be analyzed.

This will open the IMGT/V-QUEST Search page. That page has a similar display regardless of the choice IG or TR (Step 3) and the species (Step 4).

IMGT/V-QUEST Search Page

5. In the IMGT/V-QUEST Search page (Fig. 2), and below the title that reflects your choice IG or TR (Step 3), locate the following.
 - i. "Your selection" that recalls your selection, for example "Human" (Step 4)

WELCOME ! to IMGT/V-QUEST

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



Citing IMGT/V-QUEST: Brochet, X. et al., Nucl. Acids Res. 36, W503-508 (2008). PMID: 18503082

IMGT/V-QUEST programme version: 3.2.18 (21 March 2011) - IMGT/V-QUEST reference directory release: 201110-1 (7 March 2011)

Analyse your immunoglobulin (IG) or antibody nucleotide sequences

☐ Human	☐ Teleostei
☐ Mouse	☐ - Atlantic cod
	☐ - Channel catfish
	☐ - Channel catfish
	☐ - Rainbow trout
☐ Rat	☐ Chondrichthyes
☐ Rabbit	☐ Sheep

Analyse your T cell Receptor (TR) nucleotide sequences

☐ Human	☐ Nonhuman primates
☐ Mouse	☐ Bovine
☐ Camel	☐ Dog
☐ Rainbow trout	

Download IMGT reference directory in FASTA format (IG and TR)

☐ Human	☐ Teleostei
☐ Mouse	☐ - Atlantic cod
	☐ - Channel catfish
	☐ - Rainbow trout
☐ Rat	☐ Chondrichthyes
☐ Rabbit	☐ Sheep
☐ Dog	☐ Nonhuman primates
☐ Camel	

FIGURE 1. IMGT/V-QUEST Welcome page.

A note indicates the IMGT reference directory set against which your sequences will be analyzed, e.g., "Human (Homo sapiens) IG set," as determined by the tool from Steps 3 and 4. A link provides access to the IMGT reference directory sets.

A link for sequence sets allows one to test IMGT/V-QUEST, if needed (only for human IG sequences).

ii. The sequence submission section

That section comprises:

Two windows to either type (or copy/paste) your sequence(s) or upload a file, in FASTA format (proceed to Step 6 for entering your sequences).

The "Start" and "Clear the form" buttons.

iii. "Selection for results"

IMGT/V-QUEST provides the choice between three types of selection for the results:

"A. Detailed view." This displays the results of the analysis for each sequence, individually (proceed to Step 8).

"B. Synthesis view." This displays the alignments of sequences that express the same V gene and allele and allows one to compare the localization of the mutations and the composition of their junctions (proceed to Step 14 for choice of the parameters).

"C. Excel file." This allows the users to get the results of the analysis in a spreadsheet with the choice of 11 sheets (proceed to Step 21 for choice of the parameters).

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

Type (or copy/paste) your nucleotide sequence(s) in FASTA format

```
>seq1
caggtgcagctggtgcagctctgggctgaagtgaagaagcctgggtcctcggggaaggtc
tcctgcaaggtctttggagtcaccttcagtagttacgtatcagctgggtgcgacaggcc
cctggacaagggtcctgagtgatgggagggatcatcctctgttcggaaaggcaactac
gcacagaagttccaggggcaggtcacgattaccgggacgcatccacgaagcaggtctac
atggaggtgagcagctgagatctgaggacacggcgtgtattattgtgcgagacaatatt
ggtagtgtggttattacgctactggggccacggaaacctggtcacgctct
>seq2
caggtgcagctggtggagtcctggggagggcgtggtccagcctgggaggtccctgaaactc
tcctgtgcagcctctggattcaccttcagtagttatggcatgcactgggtccgccaggtc
ccaggcaagggtcggagtggtggcagttatcatcatgatggaagtaataaatctat
```

Or give the path access to a local file containing your sequence(s) in FASTA format

Display results

A. Detailed view ☒ HTML ☐ Text Nb of nucleotides per line in alignments:

1. ☒ Alignment for V-GENE	6. ☒ V-REGION alignment	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')
2. ☐ Alignment for D-GENE	7. ☒ V-REGION translation	13. ☐ Annotation by IMGT/Automat
3. ☒ Alignment for J-GENE	8. ☒ V-REGION protein display	14. IMGT Collier de Perles
4. ☒ Results of IMGT/JunctionAnalysis	9. ☐ V-REGION mutation and AA change table	☒ link to IMGT/Collier-de-Perles tool
☐ with full list of eligible D-GENE	10. ☐ V-REGION mutation and AA change statistics	☐ IMGT Collier de Perles (for a nb of sequences < 5)
☒ without list of eligible D-GENE	11. ☐ V-REGION mutation hot spots	☐ no IMGT Collier de Perles
5. ☐ Sequence of the JUNCTION ('nt' and 'AA')		

B. Synthesis view ☐ HTML ☒ Text Nb of nucleotides per line in alignments:

1. ☒ Alignment for V-GENE	5. ☒ V-REGION protein display (with AA class colors)
2. ☒ V-REGION alignment	6. ☒ V-REGION protein display (only AA changes displayed)
3. ☒ V-REGION translation	7. ☒ V-REGION most frequently occurring AA
4. ☒ V-REGION protein display	8. ☒ Results of IMGT/JunctionAnalysis

C. Excel file

1. ☒ Summary	7. ☒ V-REGION-mutation-and-AA-change-table
2. ☒ IMGT-gapped-nt-sequences	8. ☒ V-REGION-nt-mutation-statistics
3. ☒ nt-sequences	9. ☒ V-REGION-AA-change-statistics
4. ☒ IMGT-gapped-AA-sequences	10. ☒ V-REGION-mutation-hot-spots
5. ☒ AA-sequences	11. ☒ Parameters
6. ☒ Junction	

Advanced parameters

Selection of IMGT reference directory set ☒ With all alleles ☐ With allele *01 only

Search for insertions and deletions in V-REGION ☐ Yes (slower, the nb of submitted sequences in a single run is limited to 10) ☒ No

Parameters for IMGT/JunctionAnalysis Nb of accepted D-GENE in IGH JUNCTION (default is 1) Nb of accepted mutations: in 3'-REGION in D-REGION in 5'-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)

FIGURE 2. The IMGT/V-QUEST Search page. The title “Analyze your IG or antibody nucleotide sequences” and the selection “Human” reflect the user’s choice.

iv. "Advanced parameters"

"Advanced parameters" allows one to modify the parameters for sophisticated queries or unusual sequences (proceed to Step 11 for choice of "Advanced parameters").

Entering Nucleotide Sequences for Analysis

6. Copy and paste your sequence(s) in the text area of "Nucleotide sequences," or upload a text file by selecting the option "Or give the path access to a local file containing your sequence(s) in FASTA format" and using the "Browse" button to select the file.

The analyzed sequences are the V-J or V-D-J regions that encode the variable domain (V-DOMAIN) of the antigen receptors. The analysis is performed for sequences from genomic DNA (gDNA) or complementary DNA (cDNA) without need for the user to specify the molecule type. If the submitted sequences are in reverse orientation, the tool will automatically "complementary reverse" them and will provide the user sequences and the results in the sense orientation. The online IMGT/V-QUEST can analyze up to 50 sequences in a single run. The sequences must be formatted in FASTA format. Note that FASTA headers longer than 30 characters will be truncated in the results.

7. Depending on the selection you want for your results, proceed to:
 - i. Step 8 for A. Detailed view
 - ii. Step 14 for B. Synthesis view
 - iii. Step 21 for C. Excel file

IMGT/V-QUEST Detailed View

8. Select the option "A. Detailed view" (selected by default).
9. Select the format of the results, "HTML" (by default) or "Text," and the "Nb of nucleotides per line in alignments" (60 by default).
10. Select, in the light orange box, the displays of the results to be shown.

There is a choice of 14 different result displays for the analysis of the V-DOMAIN (Lefranc 2004; Giudicelli and Lefranc 2005, 2008; Brochet et al. 2008). The most commonly requested are already selected by default. You can check or uncheck the result displays as needed.

- i. Alignment for V-GENE
- ii. Alignment for D-GENE
- iii. Alignment for J-GENE
- iv. Results of IMGT/JunctionAnalysis with full list of eligible D-GENE or without list of eligible D-GENE
- v. Sequence of the JUNCTION (nucleotide ["nt"] and amino acid ["AA"])
- vi. V-REGION alignment
- vii. V-REGION translation
- viii. V-REGION protein display
- ix. V-REGION mutation and AA change table
- x. V-REGION mutation and AA change statistics
- xi. V-REGION mutation hot spots
- xii. 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")
- xiii. Annotation by IMGT/Automat
- xiv. IMGT Colliers de Perles: link to IMGT/Collier de Perles tool **IMGT/Collier de Perles: IMGT Standardized Representation of Domains (IG, TR, and IgSF Variable and Constant Domains, MH and MhSF Groove Domains)** (Ehrenmann et al. 2011), IMGT Collier de Perles (for a number [nb] of sequences less than five), or no IMGT Collier de Perles

11. Go to “Advanced parameters” at the bottom of the search page and modify the parameters according to the particularities of the sequences, if needed.

Default values of “Advanced parameters” have been set for classical analysis. They should be modified for sophisticated queries or for unusual sequences (for example, with insertions and/or deletions).

- i. “Selection of IMGT reference directory set,” with a choice of four sets (“F+ORF,” “F+ORF+in frame P” (by default), “F+ORF including orphans,” or “F+ORF+in frame P including orphans”) where F is functional, ORF is open reading frame, P is pseudogene. This allows sequences to be compared with only relevant sequences (e.g., orthon sequences are relevant for genomic, but not for expressed repertoire studies).

*The selected set can also be chosen either “With all alleles” or “With allele *01 only.”*

- ii. “Search for insertions and deletions in V-REGION” (“Yes” or “No”).

If “Yes,” the search is slower and the number of submitted sequences in a single run is limited to 10.

- iii. “Parameters for IMGT/JunctionAnalysis:” “Nb of accepted D-GENE” (provided for the IGH junctions and for the TRB and TRD junctions, respectively) or “Nb of accepted mutations” in 3’V-REGION, D-REGION, and 5’J-REGION.

- iv. “Parameters for Detailed view:” “Nb of nucleotides to exclude in 5’ of the V-REGION for the evaluation of the nb of mutations” (to avoid, e.g., counting primer specific nucleotides) or “Nb of nucleotides to add (or exclude) in 3’ of the V-REGION for the evaluation of the alignment score” (e.g., in case of low or high exonuclease activity).

12. Click on the “Start” button to launch the analysis.

“Detailed view” results are displayed in “A. Detailed results for the IMGT/V-QUEST analyzed sequences.” The top of this page indicates the number of analyzed sequences with links to individual results (Fig. 3).

13. Use the link associated with a sequence name to go directly to the individual results.

The results include the sequence in FASTA format and a Result summary (Fig. 4 and Fig. 5) and, if selected in Step 10:

- The Alignment for V, D, and J genes and alleles (Fig. 6)
- The results of IMGT/JunctionAnalysis and the sequence of the JUNCTION (“nt” and “AA”) (Fig. 7)
- Different displays of the V-REGION (Fig. 8)
- The analysis of the mutations (Fig. 9)



THANK YOU for using **IMGT/V-QUEST**

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://www.imgt.org>

MGTV-QUEST programme version: 3.2.18; IMGT/V-QUEST reference directory release: 201110-1

A. Detailed results for the IMGT/V-QUEST analysed sequences

Number of analysed sequences: **7**

[seq1](#) [seq2](#) [seq3](#) [seq4](#) [seq5](#) [seq6](#) [seq7](#)

This release of IMGT/V-QUEST uses [IMGT/JunctionAnalysis](#) for the analysis of the JUNCTION

Hyphens (-) show nucleotide identity, dots (.) represent [gaps](#)

FIGURE 3. IMGT/V-QUEST “Detailed view” result page: *top* of the page. The number of analyzed sequences is indicated. The link associated to a sequence name allows one to go directly to the individual sequence result.

Sequence number 1: seq1

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

```
>seq1
gaggtgcagctgttggagctctgggggagggcgtggtccagctgggaggtccctgagactc
tcctgtatagcctctggatccaccttcagtagctatccatgacctgggtccgccaggtc
ccaggcaaggggctggagtggtggcaagtatatcatatgacggaagtataaatataag
gtagactccatgaaggccgactcaccatctccagagacaattccaagaacacgctgtat
ttggaaatgaacagcctgacagctgaggacacggctgtgtattactgtgcgaggacagct
ttctttaacgctatgacttctggggccagggaacctggtcaccgtctcctcagcctcc
accaagggcccatcggtcttccccctggcaccctctccaagagcaccctctgggggcaca
gcggccctgggctgcctggtaaggactactccccgaaccggtgacggtgtcgtggaac
tcaggcgcctgaccagcggcgtgcacacctccccggtgtcctacagtcctcaggactc
tactcctcagcagcgtggtgacctgacctccagcagcttgggacccagacctacatc
tgcaacgtgaatcacagccagcaacaccaagtggaacaagaagttagcccaaatct
tgtgacaaaactcacaca
```

Result summary:	Productive IGH rearranged sequence (no stop codon and in-frame junction)		
V-GENE and allele	IGHV3-30*04	score = 1255	identity = 93.06% (268/288 nt)
J-GENE and allele	IGHJ4*02	score = 204	identity = 91.67% (44/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	IGHD5-5*01	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.11]	CARTAFFNAYDFW

FIGURE 4. IMGT/V-QUEST “Detailed view” result page: *top* part of an individual result with the sequence and the “Result summary.” The *top* part of each individual result indicates the IMGT reference directory set (here, human IG set) against which the sequence was analyzed, and displays the user sequence in FASTA format (a sequence submitted in antisense orientation will be shown as a complementary reverse sequence, that is in V gene sense orientation). The “Result summary” provides a crucial feature that is the evaluation of the user sequence functionality performed by IMGT/V-QUEST: productive (if no stop codon and in-frame junction) or unproductive (if stop codons and/or out-of-frame junction). It also summarizes the main characteristics of the analyzed sequence, which includes (1) the names (with species in the six-letter abbreviation, and the functionality added in the version 3.2.19) of the closest “V-GENE and allele” (e.g., Homsap IGHV3-30*04 F) and “J-GENE and allele” (e.g., Homsap IGHJ4*02 F) with the alignment score, the percentage of identity and the ratio “number of identical nucleotides/number of aligned nucleotides,” (2) the name (with species in the six-letter abbreviation, and the functionality) of the closest “D-GENE and allele” (e.g., Homsap IGHJ4*02 F) determined by IMGT/JunctionAnalysis (Giudicelli and Lefranc 2011) with the D-REGION reading frame, (3) the framework region (FR)-IMGT lengths (e.g., [25.17.38.11]), the complementarity determining region (CDR)-IMGT lengths (e.g., [8.8.11]), and the amino acid (AA) JUNCTION sequence that characterizes a V domain. Information shown in orange (here, “Productive IGH rearranged sequence” and “93.06%”) are used by clinicians analyzing the level of mutations in chronic lymphocytic leukemia (CLL) IGHV rearranged sequences as a prognostic criterion (Giudicelli and Lefranc 2008). Note that IMGT/V-QUEST provides warnings (data not shown) that appear as notes in red to alert the user if potential insertions or deletions are suspected in the V-REGION (sequences with <85% of identity and/or with different CDR1-IMGT and/or CDR2-IMGT lengths compared with the closest germline V-REGION), or whether other possibilities for the J gene and allele are identified. The seq1 accession number is AB012909 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

- The localization of the mutation hot spots (Fig. 10)
- The 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”) (Fig. 11)
- The annotation by IMGT/Automat (Fig. 12)
- The IMGT Collier de Perles or, by default, the link to the IMGT Collier de Perles tool **IMGT/Collier de Perles: IMGT Standardized Representation of Domains (IG, TR, and IgSF Variable and Constant Domains, MH and MhSF Groove Domains)** (Fig. 13; Ehrenmann et al. 2011)

IMGT/V-QUEST Synthesis View

14. Select the option “B. Synthesis view.”
15. Select the format of the results, “HTML” (by default) or “Text,” and the “Nb of nucleotides per line in alignments” (60 by default).

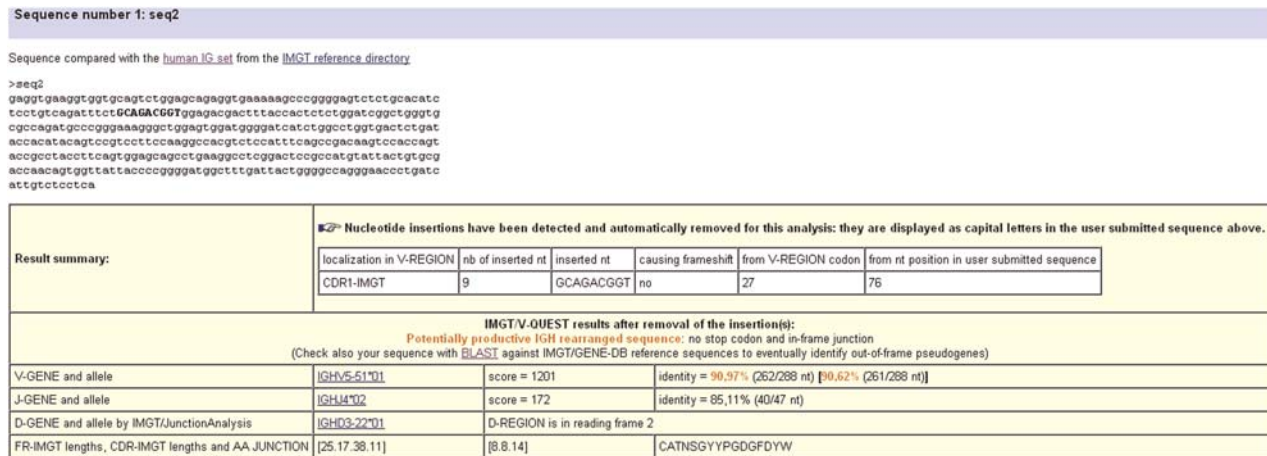


FIGURE 5. IMGT/V-QUEST “Detailed view” result page: example of “Result summary” for a sequence with insertion analyzed with the option “Search for insertions and deletions.” The detection and detailed description of insertions and/or deletions are shown in the “Result summary” first row to capture the user attention. Insertions appear as capital letters in the FASTA sequence. In the “V-GENE and allele” line of the “Result summary,” the percentage of identity and the ratio are those obtained after removal of the insertion(s) and/or deletion(s) by IMGT/V-QUEST. Between brackets are shown the percentage of identity and ratio considering each insertion or deletion as one mutational event. This information is used by clinicians analyzing the level of mutations in CLL IGHV rearranged sequences (Giudicelli and Lefranc 2008). The seq2 accession number is L26531 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

16. Check, in the light orange box, the displays of the results to be shown.

There are seven different result displays for the alignments of sequences that express the same V gene and allele, and the results of IMGT/JunctionAnalysis (Lefranc 2004; Giudicelli and Lefranc 2005, 2008; Brochet et al. 2008). They are all selected by default, but you can uncheck or check them as needed:

*Alignment for V-GENE
V-REGION alignment
V-REGION translation
V-REGION protein display
V-REGION protein display (with AA class colors)
V-REGION protein display (only AA changes displayed)
V-REGION most frequently occurring AA
Results of IMGT/JunctionAnalysis*

17. Go to “Advanced parameters” at the bottom of the search page and modify the parameters according to the particularities of the sequences, if needed (for details, see Step 11).

18. Click on “Start” to launch the analysis.

“Synthesis view” results are displayed in “B. Synthesis for the IMGT/V-QUEST analyzed sequences.” The top of this page indicates the number of analyzed sequences and displays a “Summary table” (Fig. 14).

19. Below the “Summary table,” use the links associated with the locus name(s) (e.g., IGH) to display Results of IMGT/JunctionAnalysis for sequences identified by the tool as belonging to the same locus (Fig. 15).

20. Use the links associated with the gene and allele names to display the synthesis results of input sequences identified by the tool as using the same closest germline V gene and allele.

The synthesis results provide seven different displays (if all were selected):

- “Alignment for V-GENE” (Fig. 16)
- “V-REGION alignment according to the IMGT numbering” (Fig. 16)
- “V-REGION translation”
- V-REGION protein displays in three different formats (Fig. 17)
- “V-REGION most frequently occurring AA per position and per FR-IMGT and CDR-IMGT” (Fig. 18)

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
<u>L06615</u>	IGHV3-30*04	1255	93,06% (268/288 nt)
<u>M83134</u>	IGHV3-30*01	1246	92,71% (267/288 nt)
<u>M77300</u>	IGHV3-30*09	1246	92,71% (267/288 nt)
<u>M77324</u>	IGHV3-30*14	1246	92,71% (267/288 nt)
<u>X92283</u>	IGHV3-30-3*01	1246	92,71% (267/288 nt)

Alignment with FR-IMGT and CDR-IMGT delimitations

```

seq1      gaggtgcagctggtggagctctggggga...ggcgtgggccagcctgggagggtccctgagactctcctgtatagcctctggattcaccttc
L06615   IGHV3-30*04                -g-----g-----..-----gc-----
M83134   IGHV3-30*01                c-----g-----..-----gc-----
M77300   IGHV3-30*09                c-----g-----..-----gc-----
M77324   IGHV3-30*14                c-----g-----..-----gc-----
X92283   IGHV3-30-3*01                c-----g-----..-----gc-----

          CDR1-IMGT    <-----FR2-IMGT----->          CDR
seq1      .....agtagctatcctatgacctgggtccgccaggctccaggcaaggggctggagtgggtggcaagtatatcatatgac...
L06615   IGHV3-30*04                ..-g--ca-----gt-----t...
M83134   IGHV3-30*01                ..-g--ca-----a-----gt-----t...
M77300   IGHV3-30*09                ..-g--ca-----gt-----t...
M77324   IGHV3-30*14                ..-g--ca-----gt-----t...
X92283   IGHV3-30-3*01                ..-g--ca-----gt-----t...

2-IMGT    <-----FR3-IMGT----->
seq1      ..ggaagttataaatataaggtagactccatgaag...ggccgactcaccatctccagagacaattccaagaacacgctgtatttggaa
L06615   IGHV3-30*04                ..-a---ct-c-c-----g-----t-----c-c-c-
M83134   IGHV3-30*01                ..-a---ct-c-c-----g-----t-----c-c-c-
M77300   IGHV3-30*09                ..-a---ct-c-c-----g-----t-g-----c-c-c-
M77324   IGHV3-30*14                ..-a---ct-c-c-----g-----t-----c-tc-
X92283   IGHV3-30-3*01                ..-ca---ct-c-c-----g-----t-----c-c-

          -----CDR3-IMGT-----
seq1      atgaacagcctgacagctgaggacacggctgtgtattactgtgcaggacagcttctttaacgcctatgactctctggggccagggaacc
L06615   IGHV3-30*04                -----g-----aga
M83134   IGHV3-30*01                -----g-----aga
M77300   IGHV3-30*09                -----g-----aga
M77324   IGHV3-30*14                -----g-----aga
X92283   IGHV3-30-3*01                -----g-----a

seq1      ctggtcaccgctctectcagcctcccacaaagggccccatcggtcttccccctggcaccctcctccaagagcacctctgggggcaca
L06615   IGHV3-30*04
M83134   IGHV3-30*01
M77300   IGHV3-30*09
M77324   IGHV3-30*14
X92283   IGHV3-30-3*01

```

3. Alignment for J-GENE and allele identification

Closest J-REGIONS

	Score	Identity
<u>X86355</u>	204	91,67% (44/48 nt)
<u>J00256</u>	195	89,58% (43/48 nt)
<u>M25625</u>	177	85,42% (41/48 nt)
<u>J00256</u>	174	82,35% (42/51 nt)
<u>X86355</u>	174	82,35% (42/51 nt)

Alignment

```

seq1      gagacgtcttctttaacgcctatgactctctggggccagggaacccctggtcacctgtctcctcagcctccaccaaggggcccatcggtcttccc
X86355   IGHJ4*02      .....-ta-t-----a-----
J00256   IGHJ4*01      .....-ta-t-----a-----
M25625   IGHJ4*03      .....g-ta-t-----a-----g-----
J00256   IGHJ5*01      .....ac-----tgg-tc-----c-----a-----
X86355   IGHJ5*02      .....ac-----tgg-tc-----cc-----

```

FIGURE 6. IMGT/V-QUEST “Detailed view” result page: “Alignment for V-GENE and allele identification” and “Alignment for J -GENE and allele identification.” If selected in the Search page, the alignments for the V- and J-GENE and allele identification are displayed with the alignment score, the percentage of identity for the five closest genes and alleles and the ratio number of identical nucleotides/number of aligned nucleotides. Dashes indicate identical nucleotides. Dots indicate gaps. FR-IMGT and CDR-IMGT are delimited according to the IMGT unique numbering (Lefranc 1997, 1999; Lefranc et al. 2003). The seq1 accession number is AB012909 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

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 1.

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
seq1	IGHV3-30*04	tgtgcgag...	g	acagcttctctta	acgc	ctatgactctctgg	IGHJ4*02	IGHD5-5*01	0	3	2	4/5

Eligible D genes:

D name	D length	tgtgcgaggacagcttctcttaacgcctatgactctctgg	Score#	Mutation#	Location
IGHD1-1*01	17	---a---	5	1	d[4-9],s[10-15]
IGHD2-2*01	31	-----	5	0	d[19-23],s[11-15]
IGHD2-2*02	31	-----	5	0	d[19-23],s[11-15]
IGHD2-2*03	31	-----	5	0	d[19-23],s[11-15]
IGHD3-3*01	31	--c---	4	1	d[10-14],s[16-20]
IGHD3-3*02	31	---a--c---	7	2	d[6-14],s[12-20]
IGHD4-4*01	16	-----	4	0	d[10-13],s[21-24]
IGHD5-5*01	20	-----a-gg---	10	3	d[7-19],s[10-22]
IGHD6-6*01	18	-----	5	0	d[9-13],s[11-15]
IGHD1-7*01	17	-----	4	0	d[5-8],s[21-24]
IGHD2-8*01	31	--a---	4	1	d[25-29],s[18-22]
IGHD2-8*02	31	--a---	4	1	d[25-29],s[18-22]
IGHD3-9*01	31	---g---	5	1	d[13-18],s[19-24]
IGHD3-10*01	31	-----	4	0	d[28-31],s[21-24]
IGHD3-10*02	30	--a-g---	5	2	d[7-13],s[14-20]
IGHD4-11*01	16	-----	4	0	d[10-13],s[21-24]
IGHD5-12*01	23	---acga---	6	4	d[13-22],s[13-22]
IGHD6-13*01	21	-----	5	0	d[12-16],s[11-15]
IGHD1-14*01	17	-----	4	0	d[5-8],s[21-24]
IGHD2-15*01	31	-----	4	0	d[20-23],s[12-15]
IGHD3-16*01	37	--c-a-gc---	7	4	d[23-33],s[12-22]
IGHD3-16*02	37	--c-a-gc---	8	3	d[23-33],s[12-22]
IGHD5-18*01	20	-----a-gg---	10	3	d[7-19],s[10-22]
IGHD1-20*01	17	-----	4	0	d[5-8],s[21-24]
IGHD2-21*01	28	--g---	4	1	d[19-23],s[15-19]
IGHD2-21*02	28	---c---	6	1	d[17-23],s[9-15]
IGHD3-22*01	31	--a---	4	1	d[22-26],s[16-20]
IGHD4-23*01	19	-----	4	0	d[13-16],s[21-24]
IGHD5-24*01	20	---caaa---	6	4	d[10-19],s[13-22]
IGHD6-25*01	18	-----	4	0	d[9-12],s[11-14]
IGHD1-26*01	20	-----	4	0	d[12-15],s[12-15]
IGHD7-27*01	11	-----	4	0	d[2-5],s[21-24]

Translation of the JUNCTION

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

☐

	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass	pI
	C	A	R	T	A	F	F	N	A	Y	D	F	U				11	1,611.8	6.44
seq1	tgt	gcy	agg	aca	gct	ttc	ttt	aac	gcc	tat	gac	ttc	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.

5. Sequence of the JUNCTION ('nt' and 'AA')

104	105	106	107	108	109	110	111	112	113	114	115	116	117	118
C	A	R	T	A	F	F	N	A	Y	D	F	U		
tgt	gcy	agg	aca	gct	ttc	ttt	aac	gcc	tat	gac	ttc	tgg		

Input for IMGT/JunctionAnalysis

```
>seq1_IGHV3-30*04_IGHJ4*02
tgtgcgaggacagcttctcttaacgcctatgactctctgg
```

FIGURE 7. IMGT/V-QUEST “Detailed view” result page: “Results of IMGT/JunctionAnalysis” and “Sequence of the JUNCTION (‘nt’ and ‘AA’).” The “Results of IMGT/JunctionAnalysis” (Giudicelli and Lefranc 2011) comprise the detailed “Analysis of the JUNCTION” (Yousfi Monod et al. 2004) with the D-REGION reading frame, “Eligible D genes” if selected (D genes and alleles that match at least four or more nucleotides with the junction, allowing the user to visualize the result among other close solutions), and “Translation of the JUNCTION” with AA colored according to the 11 IMGT physicochemical classes (Pommié et al. 2004). The “Sequence of the JUNCTION (‘nt’ and ‘AA’)” provides the JUNCTION in nucleotides and amino acids according to the IMGT numbering and in the FASTA format required as input by IMGT/JunctionAnalysis online. The seq1 accession number is AB012909 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

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

[illegible]

7. V-REGION translation

	<-----FR1 - IMGT----->	
seq1	l 1 5 10 15 20 25 30 E V Q L L E S G G G V V V Q P G R S L R L S C I A S G F Y F gag gtc cag ctg ttg gag tct ggg gga ... ggc gtc gtc cag cct ggy agy tcc ctg aga tcc tcc tgt ata gcc tct gga ttc acc ttc Q V A gc- c-g-	
L06615 IGHV3-30*04		
	--- CDR1 - IMGT ---<-----FR2 - IMGT----->CDR2	
seq1	35 40 45 50 55 60 S S Y P M T U V V R Q A P G K G L E W V A S I S Y D agt agc tat cct atg acc ttc gtc cgc cag gct cca ggc aag ggy ctg gag tgy gty gca agt ata tca tat gac ... A H V g-ca- gt- --t ...	
L06615 IGHV3-30*04		
	- IMGT ---<-----FR3 - IMGT----->	
seq1	65 70 75 80 85 90 G S Y R Y K V D S H K G R L T I S R D N S K N T L Y L E ... gya agt tat aaa tat aag gra gac tcc atg aag ... ggc cga ctc acc atc tcc aga gac aat tcc aag aac acy ctg tat ttg yaa N Y A V F a--c t-c-c- g- ... t- c-	
L06615 IGHV3-30*04		
	-----> CDR3 - IMGT -----	
seq1	95 100 104 M N S L T A E D T A V Y Y C A R T A F F N A Y D F W C Q G T atg aac agc ctg aca gct gag gac acy gct gtg tat tac tct gcy agg aca gct ttc ttt aac gcc tat gac ttc tgy ggc cag gga acc R atg ... -g- --a ga	
L06615 IGHV3-30*04		
seq1	L V T V S S A S T K G P S V F P L A P S S K S T S G G Y ctg gtc acc gtc tcc tca gcc tcc acc aag ggc cca tcy gtc ttc ccc ctg gca ccc tcc tcc aag aac acc tct ggy ggc aca	
L06615 IGHV3-30*04		

8. V-REGION protein display

[illegible]

FIGURE 8. IMGT/V-QUEST “Detailed view” result page: three displays of the V-REGION. “V-REGION alignment according to the IMGT unique numbering” displays the nucleotide sequences with the FR-IMGT and CDR-IMGT delimitations according to the IMGT unique numbering (Lefranc 1997, 1999; Lefranc et al. 2003). The “V-REGION translation” displays the nucleotide sequence and deduced amino acid translation of the input sequence, aligned with the closest germline V-REGION, and with the FR-IMGT and CDR-IMGT delimitations. “V-REGION protein display” displays the deduced amino acid translation of the input sequence, aligned with the V-REGION of the closest germline V-GENE and with the FR-IMGT and CDR-IMGT delimitations, and on the third line of the alignment and shown in bold, the amino acids of the input sequences that are different from the closest germline V-REGION. The seq1 accession number is AB012909 from the IMGT/IGHM-DB database (Giudicelli et al. 2006).

9. V-REGION mutation and AA change table

FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
c1>g, Q1>E (+ + -)	g112>c, A38>P (- -)	c118>a, H40>T (+ - -)	t177>c	c198>t	a318>g
g13>t, V5>L (+ - +)		a119>c, H40>T (+ - -)	a190>t, N64>Y (- -)	t199>a, Y67>K (- -)	g319>a
g70>a, A24>l (+ - +)		g163>a, V55>S (- -)		c201>g, Y67>K (- -)	a320>c
c71>t, A24>l (+ - +)		t164>g, V55>S (- -)		c203>t, A68>V (+ - +)	
				g211>a, V71>M (+ - -)	
				t226>c, F76>L (+ - -)	
				c265>t	
				c268>g, G90>E (+ - +)	
				g284>c, R95>T (- -)	

10. V-REGION mutation and AA change statistics

Nucleotide (nt) mutations

IMGT labels	V-REGION	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
Nb of positions including IMGT gaps (nt)	317 (320)	78	36	51	30	117	5 (8)
Nb of nucleotides	293 (296)	75	24	51	24	114	5 (8)
Nb of identical nucleotides	273	71	23	47	22	105	5
Nb of mutations	20 (23)	4	1	4	2	9	0 (3)
Mutations							
Silent	3 (6)	0	0	0	1	2	0 (3)
Nonsilent	17	4	1	4	1	7	0
Transitions							
a>g	0 (1)	0	0	0	0	0	0 (1)
g>a	3 (4)	1	0	1	0	1	0 (1)
c>t	4	1	0	0	0	3	0
t>c	2	0	0	0	1	1	0
Transversions							
a>c	1 (2)	0	0	1	0	0	0 (1)
c>a	1	0	0	1	0	0	0
a>t	1	0	0	0	1	0	0
t>a	1	0	0	0	0	1	0
g>c	2	0	1	0	0	1	0
c>g	3	1	0	0	0	2	0
g>t	1	1	0	0	0	0	0
t>g	1	0	0	1	0	0	0

Amino acid (AA) changes

IMGT labels	V-REGION	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
Nb of positions including IMGT gaps (AA)	105 (106)	26	12	17	10	39	1 (2)
Nb of AA	97 (98)	25	8	17	8	38	1 (2)
Nb of identical AA	84 (85)	22	7	15	7	32	1 (2)
Nb of AA changes	13	3	1	2	1	6	0
AA changes							
Very similar (+ + +)	0	0	0	0	0	0	0
Similar (+ + -)	2	1	0	0	0	1	0
(+ - +)	3	2	0	0	0	1	0
(+ - -)	3	0	0	1	0	2	0
Dissimilar (- + -)	0	0	0	0	0	0	0
(- - +)	0	0	0	0	0	0	0
Very dissimilar (- - -)	5	0	1	1	1	2	0

FIGURE 9. IMGT/V-QUEST “Detailed view” result page: “V-REGION mutation and AA change table,” and “V-REGION mutation and AA change statistics.” The “V-REGION mutation and AA change table” lists the mutations (nt and AA) of the analyzed sequence compared with the closest V-REGION allele. They are described for the V-REGION and for each FR-IMGT and CDR-IMGT, with their nucleotide and codon position according to the IMGT unique numbering (Lefranc 1997, 1999; Lefranc et al. 2003) and for the AA changes according to the IMGT AA classes (Pommié et al. 2004). For example, c1>g, Q1>E (+ + -) means that the nucleotide mutation (c>g) at nt 1 leads to an AA change (Q>E) at codon 1 with the same hydrophathy (+) and volume (+) but with different physicochemical properties (-) classes (Pommié et al. 2004). It is the first time that such qualification of AA replacement is provided. This has led us to identify four types of AA changes: very similar (+ + +), similar (+ + -, + - +), dissimilar (- - +, - + -, + - -) and very dissimilar (- - -). The “V-REGION mutation and AA change statistics” comprises two tables, the first one for “Nucleotide (nt) mutations” described as silent/nonsilent, transitions/transversions, and the second one for “Amino acid (AA) changes” described as above. For both tables, results are given for the V-REGION and per FR-IMGT and CDR-IMGT. Statistics are calculated up to the 3' end of the V-REGION identified in the input sequence (this includes the 3' last two identical nucleotides with the closest germline V-REGION). The numbers in parentheses, in the V-REGION and CDR3-IMGT columns, correspond to the statistics calculated up to the 3' end of the closest germline V-REGION, and therefore may include nucleotide and amino acid differences due to the junction diversity. The accession number of the analyzed sequence is AB012909 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

11. V-REGION mutation hot spots (motifs and localizations in germline V-REGION)

(a/t) _a wa		t(a/t) tw		(a/g)g(c/t)(a/t) rgyw		(a/t)(a/g)c(c/t) wrgy	
Motif	Localization	Motif	Localization	Motif	Localization	Motif	Localization
ta	105-106 (CDR1)	tt	82-83 (CDR1)	agct	8-11 (FR1)	agct	8-11 (FR1)
ta	109-110 (CDR1)	tt	88-89 (CDR1)	agta	103-106 (CDR1)	agcc	41-44 (FR1)
ta	114-115 (CDR1)	ta	105-106 (CDR1)	ggct	132-135 (FR2)	agcc	72-75 (FR1)
aa	142-143 (FR2)	ta	109-110 (CDR1)	ggca	139-142 (FR2)	agct	106-109 (CDR1)
ta	165-166 (FR2-CDR2)	ta	114-115 (CDR1)	ggct	146-149 (FR2)	tgct	111-114 (CDR1)
ta	167-168 (CDR2)	tt	164-165 (FR2)	ggca	159-162 (FR2)	tact	196-199 (FR3)
ta	172-173 (CDR2)	ta	167-168 (CDR2)	agta	187-190 (CDR2)	agcc	277-280 (FR3)
aa	186-187 (CDR2)	ta	172-173 (CDR2)	agct	285-288 (FR3)	agct	285-288 (FR3)
ta	189-190 (CDR2)	ta	189-190 (CDR2)	ggct	297-300 (FR3)	tact	307-310 (FR3)
ta	192-193 (CDR2)	ta	192-193 (CDR2)				
aa	194-195 (CDR2)	ta	196-197 (FR3)				
ta	196-197 (FR3)	ta	199-200 (FR3)				
ta	199-200 (FR3)	tt	226-227 (FR3)				
aa	214-215 (FR3)	tt	246-247 (FR3)				
aa	244-245 (FR3)	ta	262-263 (FR3)				
aa	250-251 (FR3)	ta	304-305 (FR3)				
aa	253-254 (FR3)	tt	306-307 (FR3)				
ta	262-263 (FR3)						
aa	269-270 (FR3)						
aa	274-275 (FR3)						
ta	304-305 (FR3)						
ta	307-308 (FR3)						

FIGURE 10. IMGT/V-QUEST “Detailed view” result page: “V-REGION mutation hot spots (motifs and localization in germline V-REGION).” This table provides the hot spot motifs identified in the closest germline V-REGION and their localization. The hot spot patterns are (a/t)_a and (a/g)g(c/t)(a/t), and their complementary reverse motifs t(a/t) and (a/t)(a/g)c(c/t). The accession number of the analyzed sequence is AB012909 from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

12. 'seq1' V-REGION and V-D-J-REGION

V-REGION nucleotide sequence in FASTA format

```
>seq1
gaggtgcagctgttgagctctggggga...ggcgtgggtccagctgggaggtccctgaga
ctctctgtatagcctctggattcacttc.....agtagctatcctatgacc
tggttccggcaggtccaggcaagggtggagtggtggcaagtatatcatatgac...
...ggaagttataaatataaggtagactccatgaag...ggccgactcaccatctccaga
gacaattccaagaacacgctgtatttggaaatgaacagcctgacagctgaggacacggct
gtgtattactgtgcgag
```

Analyse this sequence with IMGT/PhyloGene

V-REGION amino acid sequence in FASTA format

```
>seq1
EVQLLESGG.GVVQPGRLRLSLSCIASGFTF...SSYPHTWVRQAPGKLEWVASISYD.
.GSYKYKVD SHK.GRLTISRDN SKNTLYLEHNSLTAEDTAVYYCA
```

V-REGION amino acid sequence on one line

```
>seq1
EVQLLESGG.GVVQPGRLRLSLSCIASGFTF...SSYPHTWVRQAPGKLEWVASISYD..GSYKYKVD SHK.GRLTISRDN SKNTLYLEHNSLTAEDTAVYYCA
```

V-D-J-REGION nucleotide sequence in FASTA format

```
>seq1
gaggtgcagctgttgagctctggggga...ggcgtgggtccagctgggaggtccctgaga
ctctctgtatagcctctggattcacttc.....agtagctatcctatgacc
tggttccggcaggtccaggcaagggtggagtggtggcaagtatatcatatgac...
...ggaagttataaatataaggtagactccatgaag...ggccgactcaccatctccaga
gacaattccaagaacacgctgtatttggaaatgaacagcctgacagctgaggacacggct
gtgtattactgtgcgaggacagcttctttaaagcctatgacttctggggccagggaacc
ctgggtcaccgtctctcag
```

V-D-J-REGION amino acid sequence in FASTA format

```
>seq1
EVQLLESGG.GVVQPGRLRLSLSCIASGFTF...SSYPHTWVRQAPGKLEWVASISYD.
.GSYKYKVD SHK.GRLTISRDN SKNTLYLEHNSLTAEDTAVYYCARTAFFNAYDFWGGQTLTVTS
```

V-D-J-REGION amino acid sequence on one line

```
EVQLLESGG.GVVQPGRLRLSLSCIASGFTF...SSYPHTWVRQAPGKLEWVASISYD..GSYKYKVD SHK.GRLTISRDN SKNTLYLEHNSLTAEDTAVYYCARTAFFNAYDFWGGQTLTVTS
```

FIGURE 11. IMGT/V-QUEST “Detailed view” result page: sequences of V-REGION and V-J- or V-D-J-REGION. This section provides nucleotide and amino acid sequences with gaps according to the IMGT unique numbering (Lefranc 1997, 1999; Lefranc et al. 2003) and includes the V-REGION nucleotide sequence in FASTA format with access to the IMGT/PhyloGene tool (Elemento and Lefranc 2003), “V-REGION amino acid on one line,” V-J or V-D-J-REGION nucleotide and amino acid sequences in FASTA format, V-J, or V-D-J-REGION AA sequence on one line.

13. Annotation by IMGT/Automat

Label	Location/Qualifiers
<u>V-D-J-REGION</u>	1..355 /CDR_length="[8.8.11]" /FR_length="[25.17.38.11]" /nucleotide sequence gaggtgcagctgttgagctctgggggagggcggtggtccagcctgggaggtccctgagactc tcctgtatagcctcttgattcaccttcagtagctatcctatgacctgggtccgcccaggtc ccaggcaaggggctggagtggtggcaagtatatcatatgacggaagttataataaag gtagactccatgaagggccgactcaccatctccagagacaattccaagaacacgctgtat ttggaaatgaacagcctgacagctgaggacacggctgtgtattactgtgcgagacagct ttctttaacgcctatgacttctggggccagggaacctggtcaccgtctcctcag /translation EVQLLESGGGVQPGRSLRLSCIASGFTFSSYPHTWVRQAPGKLEWVASISYDGSYKYK VDSMKGRLTISRDN SKNTLYLEHNSLTAE DTA VYVCARTAFFNAYDFWGQGLTVTVSS
<u>V-REGION</u>	1..293 /allele="IGHV3-30*04" /gene="IGHV3-30" /nucleotide sequence gaggtgcagctgttgagctctgggggagggcggtggtccagcctgggaggtccctgagactc tcctgtatagcctcttgattcaccttcagtagctatcctatgacctgggtccgcccaggtc ccaggcaaggggctggagtggtggcaagtatatcatatgacggaagttataataaag gtagactccatgaagggccgactcaccatctccagagacaattccaagaacacgctgtat ttggaaatgaacagcctgacagctgaggacacggctgtgtattactgtgcgag /translation EVQLLESGGGVQPGRSLRLSCIASGFTFSSYPHTWVRQAPGKLEWVASISYDGSYKYK VDSMKGRLTISRDN SKNTLYLEHNSLTAE DTA VYVCA
<u>FR1-IMGT</u>	1..75 /AA_IMGT="AA 1 to 26, AA 10 is missing" /nucleotide sequence gaggtgcagctgttgagctctgggggagggcggtggtccagcctgggaggtccctgagactc tcctgtatagcctct /translation EVQLLESGGGVQPGRSLRLSCIAS
<u>1st-CYS</u>	64..66 /nucleotide sequence tgt /translation C
<u>CDR1-IMGT</u>	76..99 /AA_IMGT="AA 27 to 38, AA 31, 32, 33, 34 are missing" /nucleotide sequence ggattcaccttcagtagctatcct /translation GFTFSSYP
<u>FR2-IMGT</u>	100..150 /AA_IMGT="AA 39 to 55" /nucleotide sequence atgacctgggtccgcccaggctccaggcaaggggctggagtggtggcaagt /translation MTWVRQAPGKLEWVAS
<u>CONSERVED-TRP</u>	106..108 /nucleotide sequence tgg /translation W
<u>CDR2-IMGT</u>	151..174 /AA_IMGT="AA 56 to 65, AA 60, 61 are missing" /nucleotide sequence atatcatatgacggaagttataaa /translation ISYDGSYK

FIGURE 12. IMGT/V-QUEST “Detailed view” result page: “Annotation by IMGT/Automat”: the results of the analysis of the input sequence by IMGT/Automat (Giudicelli et al. 2003, 2005b) provide a full automatic annotation for the V-J-REGION or V-D-J-REGION using IMGT standardized labels.

IMGT/V-QUEST Excel File

21. Select the option “C. Excel file.”

22. Select, in the light orange box, the displays of the results to be shown.

There are 11 sheets proposed for the result display. The “Summary” and “Parameters” sheets are always provided by default.

- Summary
- IMGT-gapped-nt-sequences

14. IMGT Collier de Perles for seq1

IMGT/V-QUEST identified region: V-D-J-REGION

CDR-IMGT lengths: [8.8.11]

FR-IMGT lengths: [25.17.38.11]

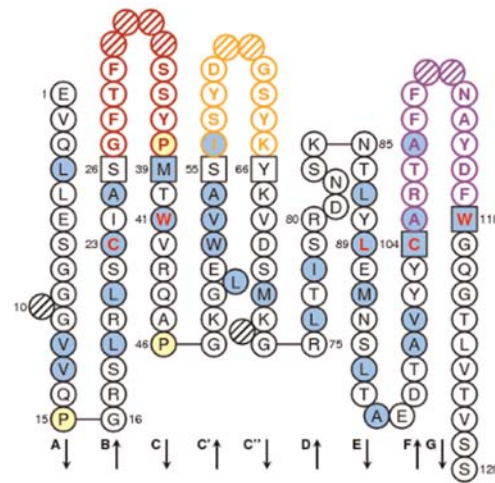
CDR-IMGT and FR-IMGT lengths are based on the [IMGT unique numbering for V-DOMAIN](#).

FIGURE 13. IMGT/V-QUEST “Detailed view” result page: IMGT Collier de Perles. This section recalls the identified region (here, V-D-J-REGION) and the CDR-IMGT lengths and FR-IMGT lengths. Depending on the choice selected in option 14 of the IMGT/V-QUEST Search page (see Fig. 2), the “IMGT Collier de Perles” (Ruiz and Lefranc 2002; Kaas et al. 2004, 2007; Kaas and Lefranc 2007) can be displayed either as a “link to IMGT/Collier de Perles tool” **IMGT/Collier de Perles: IMGT Standardized Representation of Domains (IG, TR, and IgSF Variable and Constant Domains, MH and MhSF Groove Domains)** (Ehrenmann et al. 2011) or as a direct “IMGT Collier de Perles (for a nb of sequences <5)” representation integrated into IMGT/V-QUEST results. The IMGT Collier de Perles shown was obtained by selecting this last option. Positions in blue mean that the amino acids of the user sequence at these positions is hydrophobic (hydropathy index with positive value) or is a tryptophan (W), as in 50% or more of analyzed V domains. Positions with red and bold letters indicate the five conserved positions of a V domain: 1st-CYS 23, CONSERVED-TRP 41, hydrophobic 89, 2nd-CYS 104, and J-TRP 118. Anchor positions are in squares. Hatched positions correspond to gaps according to the IMGT unique numbering for V domain (Lefranc 1997, 1999; Lefranc et al. 2003). Prolines are shown in yellow. Arrows indicate the β strands and their direction. The CDR-IMGT lengths of this domain are (8.8.11). The CDR-IMGT are colored as follows: CDR1-IMGT (red), CDR2-IMGT (orange), and CDR3-IMGT (purple) for the IGH (this figure), TRB and TRD V-D-J-REGION, and CDR1-IMGT (blue), CDR2-IMGT (green), and CDR3-IMGT (greenblue) (data not shown) for the IGH, IGL, TRA, and TRG V-J-REGION.

- *nt-sequences*
- *IMGT-gapped-AA-sequences*
- *AA-sequences*
- *Junction*
- *V-REGION-mutation-and-AA-change-table*
- *V-REGION-nt-mutation-statistics*
- *V-REGION-AA-change-statistics*
- *V-REGION-mutation-hot spot*
- *Parameters*

23. Go to “Advanced parameters” at the bottom of the search page and modify the parameters according to the particularities of the sequences, if needed (for details, see Step 11).

24. Click on “Start” to launch the analysis.

A small window will appear on the screen, giving the choice to “open” or “save” the Excel file.

B. Synthesis for the IMGT/V-QUEST analysed sequences

Number of analysed sequences: 7

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

Summary table:

Sequence ID	V-GENE and allele	Functionality	V-REGION score	V-REGION identity % (nt)	J-GENE and allele	D-GENE and allele	D-REGION reading frame	CDR-IMGT lengths	AA JUNCTION	JUNCTION frame
seq1	IGHV1-69*01	Productive	1273	93,75% (270/288 nt)	IGHJ4*01 or IGHJ4*02 or IGHJ4*03	IGHD3-22*01	2	[8.8.12]	CARQYGSSEYYAYW	in-frame
seq2	IGHV3-30*03, or IGHV3-30*18	Productive	1417	99,31% (286/288 nt)	IGHJ6*02	IGHD3-16*02	2	[8.8.23]	CAKDLDDYVWGSYRHSRYYYGMDFW	in-frame
seq3	IGHV3-30*04	Productive	1255	93,06% (268/288 nt)	IGHJ4*02	IGHD5-5*01	1	[8.8.11]	CARTAFFNAYDFW	in-frame
seq4	IGHV3-30*04	Productive	1435	100,00% (288/288 nt)	IGHJ4*02	IGHD6-19*01	2	[8.8.15]	CARDRSIAVAQYYFDYVW	in-frame
seq5	IGHV3-30*04	Productive	1408	98,96% (285/288 nt)	IGHJ4*02	IGHD6-13*01	2	[8.8.17]	CARDSSYPPIAAAGVYVW	in-frame
seq6	IGHV3-30*04	Productive	1426	99,65% (287/288 nt)	IGHJ1*01	IGHD3-10*01	2	[8.8.16]	CARGRTKSGRPGYFQHW	in-frame
seq7	IGHV3-9*01	Unproductive (stop codons)	1417	99,31% (286/288 nt)	IGHJ4*02	IGHD3-10*01	2	[8.8.17]	CAKDIGGKYSGSGSLD*W	in-frame

Results of IMGT/JunctionAnalysis for : IGH junctions

Alignment with the closest alleles:

The analysed sequences are aligned with the closest allele (with number of aligned sequences in parentheses):

[IGHV1-69*01\(1\)](#) [IGHV3-30*03\(1\)](#) [IGHV3-30*04\(4\)](#) [IGHV3-9*01\(1\)](#)

FIGURE 14. IMGT/V-QUEST “Synthesis view” result page: top of the page. The number of analyzed sequences is indicated. The Summary table displays one row for each input sequence, including: (1) the name of the sequence (Sequence ID), (2) the name of the closest V-GENE and allele (a note may appear with the V-GENE and allele name when the V-REGION score is very low (<200), and/or the percentage of identity is <85% and/or, compared with that of V, when input sequence has different CDR1-IMGT and/or CDR2-IMGT lengths: the alignment for this sequence has to be checked in “A. Detailed view,” using the advanced parameters “Search for insertions and deletions in V-REGION”), (3) the functionality of the sequence (when found, the presence of stop codons is indicated), (4) the V-REGION score, (5) the V-REGION percentage of identity and ratio number of identical nucleotides (nt)/number of aligned nt, (6) the name of the closest J-GENE and allele (notes may appear with the J-GENE and allele name to indicate that other possibilities exist for the choice of the J-GENE and allele name), and (7) provided according to the IMGT/JunctionAnalysis results, the D-GENE and allele name, the D reading frame, the CDR-IMGT lengths, the AA JUNCTION, and the JUNCTION frame. In the absence of results of IMGT/JunctionAnalysis, only the AA JUNCTION defined by IMGT/V-QUEST is displayed. The seq1, seq2, seq3, seq4, seq5, seq6, and seq7 correspond to the DQ100777, AB021524, AB012909, AB021511, AB021516, AB021514, and AB021539 accession numbers, respectively, from the IMGT/LIGM-DB database (Giudicelli et al. 2006).

25. To see the results, click on “open.”

The Excel file comprises 11 sheets. These can be viewed in the supplementary material (Results_Excel.doc).

DISCUSSION

IMGT/V-QUEST (<http://www.imgt.org>) provides a highly detailed and standardized analysis of the IG and TR rearranged sequences: it describes the main characteristics of user sequences according to the axioms and concepts of IMGT-ONTOLOGY (Giudicelli et al. 1999; Duroux et al. 2008). IMGT/V-QUEST is constantly updated based on the results of biological and clinical research. Improvements can be followed through the program version number and the reference directory release number indicated in the search pages and the list of the corresponding changes found in the corresponding documentations. New alleles annotated in IMGT are integrated in the IMGT reference directory as soon as they are confirmed experimentally and are publicly available in the generalist databases. IMGT/HighV-QUEST (Alamyar et al. 2010, 2011), the high-throughput version of IMGT/V-QUEST for Next Generation Sequencing (NGS) that analyzes up to 150,000 sequences per run, uses the IMGT/V-QUEST program and standards, and provides results with the same degree of resolution and high quality (results

Analysis of the JUNCTIONS

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



Input	V name	3' V-REGION	P	N1	P	D-REGION	N2	P	5' J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
seq1	IGHV1-69*01	tgtgcgaga..		caa		tatggtagtagtggttattac...	gc		ctactgg	IGHJ4*01	IGHD3-22*01	0	1	0	3/5
seq2	IGHV3-30*03	tgtgcgaaaga	tct	gg		atgattacgtttggggaggtatcgtcat...	tctagg		tactactacggtatggacttctgg	IGHJ6*02	IGHD3-16*02	1	1	1	5/8
seq3	IGHV3-30*04	tgtgcgag...		g		acagctttcttta.	acgc		ctatgacttctgg	IGHJ4*02	IGHD5-5*01	0	3	2	4/5
seq4	IGHV3-30*04	tgtgcgagaga	tc	gga		..gtatagcagtggc.....	cc	agt	actactttgactactgg	IGHJ4*02	IGHD6-19*01	0	0	0	4/5
seq5	IGHV3-30*04	tgtgcgagaga	t	tcctcctaccc	ccc	gggtatagcagcagctgg...	ag		ttgactactgg	IGHJ4*02	IGHD6-13*01	0	0	1	8/13
seq6	IGHV3-30*04	tgtgcgagag.		gaaggacgaag		gggtcggggag.....	acccggc		tacttccagcactgg	IGHJ1*01	IGHD3-10*01	0	0	0	12/18
seq7	IGHV3-9*01	tgtgcaaaagata	t	tgggggcaaa		tatggttcgggg.....	gggtcgggac		ttgactagtgg	IGHJ4*02	IGHD3-10*01	0	0	1	14/20

Translation of the JUNCTIONS

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	111.5	112.5	112.4	112.3	112.2	112.1	112	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass	pI		
	C	A	R	Q	Y	E	S												S	G	Y	Y	A	Y	W						
#1 seq1	tgt	gcg	aga	caa	tat	ggg	agt	...	...	...	...	...	...	...	...	...	...	...	agt	ggt	tat	tac	gcc	tac	tgg	+	12	1,674.81	8.89		
	C	A	K	D	L	D	Y	V	W	G	S	Y	R	H	S	R	Y	Y	Y	C	M	D	F	W							
#2 seq2	tgt	gcg	aaa	gat	ctg	gat	tac	gtt	tgg	ggg	agt	tat	cgt	cat	tct	agg	tac	tac	tac	ggt	atg	gac	ctc	tgg	+	23	3,197.52	5.78			
	C	A	R	T	A	F	F												N	A	Y	D	F	W							
#3 seq3	tgt	gcg	agg	aca	gct	ctc	ctt	...	...	...	...	...	...	...	...	...	...	...	...	aac	gcc	tat	gac	ctc	tgg	+	11	1,611.8	6.44		
	C	A	R	D	R	S	L	A	V									A	Q	Y	Y	F	D	Y	W						
#4 seq4	tgt	gcg	aga	gat	cgg	agt	ata	gca	gtg	...	...	...	...	...	...	...	...	...	gcc	cag	tac	tac	ttt	gac	tac	tgg	+	15	2,127.37	6.58	
	C	A	R	D	S	S	Y	P	P	G								I	A	A	A	G	V	X	Y	W					
#5 seq5	tgt	gcg	aga	gat	tcc	tcc	tac	ccc	ccg	ggg	...	...	...	...	...	...	...	...	ata	gca	gca	gct	gga	gtt	gan	tac	tgg	+	17	1,884.1	6.44
	C	A	R	G	R	T	K	G	S										G	R	P	G	Y	F	Q	H	W				
#6 seq6	tgt	gcg	aga	gga	agg	acg	aag	ggg	tcg	...	...	...	...	...	...	...	...	...	ggg	aga	ccc	ggc	tac	ttc	cag	cac	tgg	+	16	2,064.32	11.72
	C	A	K	D	I	G	G	K	Y	G									S	G	G	S	C	L	D	*	W				
#7 seq7	tgt	gca	aaa	gat	att	ggg	ggc	aaa	tat	ggg	...	...	...	...	...	...	...	...	tcg	ggg	ggc	tcg	gga	ctt	gac	taa	tgg	+	17	1,770.94	6.58

FIGURE 15. IMGT/V-QUEST “Synthesis view” result page: results of IMGT/JunctionAnalysis. This section displays the “Analysis of the JUNCTIONS” and the “Translation of the JUNCTIONS” per locus, as provided by the IMGT/JunctionAnalysis tool for all sequences identified by IMGT/V-QUEST as belonging to the same locus (IGH, IGK, or IGL locus for IG sequences, TRA, TRB, TRG, or TRD locus for TR sequences) (Yousfi Monod et al. 2004; Brochet et al. 2008; Giudicelli and Lefranc 2011).

Sequences aligned with IGHV3.30*04

1. Alignment for V-GENE

[illegible]

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

	<-----FR1 - INGT----->	
L06615 IGHV3-30*04	1 5 10 15 20 25 30	
seq3	cag gtg cag <u>ctg</u> gtg gag tct ggg gga ... ggc gtg gtc <u>cag</u> <u>cct</u> ggg agg tcc ctg aga ctc tcc tgc <u>gcc</u> tct gga <u>tcc</u> acc <u>tcc</u>	
seq4	g-- --- cag <u>ctg</u> t-- gag tct ggg gga ... ggc gtg gtc <u>cag</u> <u>cct</u> ggg agg tcc ctg aga ctc tcc tgc <u>at-</u> <u>ccc</u> tct gga <u>tcc</u> acc <u>tcc</u>	
seq5	--- ---	
seq6	--- ---	
	CDR1 - INGT<-----FR2 - INGT----->CDR2	
L06615 IGHV3-30*04	35 40 45 50 55 60	
seq3	 <u>agt</u> <u>agc</u> <u>tac</u> <u>gct</u> atg cac tgg gtc cgc cag <u>gct</u> cca <u>ggc</u> <u>aag</u> <u>ggy</u> <u>ctg</u> gag tgg gtg <u>cca</u> <u>gtt</u> ata tca tat gat ...	
seq4	 --- --- --- c-- ac-- ac-- --- --- --- --- --- --- --- ag-- ag-- ata tca tat gat ...	
seq5	 ---	
seq6	 ---	
	- INGT<-----FR3 - INGT----->	
L06615 IGHV3-30*04	65 70 75 80 85 90	
seq3	... gga <u>agt</u> <u>aat</u> <u>aaa</u> <u>tac</u> <u>tac</u> gca gac tcc gtg <u>aag</u> ... ggc cga <u>ttc</u> acc atc tcc aga gac <u>aat</u> <u>tcc</u> <u>aag</u> <u>aac</u> acg ctg <u>tat</u> ctg <u>caa</u>	
seq4	... --- --- t-- --- --t a-g -t- ... a-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- t-- g--	
seq5	... ---	
seq6	... ---	
	CDR3 - INGT	
L06615 IGHV3-30*04	95 100 104	
seq3	atg <u>aac</u> <u>agc</u> <u>ctg</u> aga <u>gct</u> gag gac acg <u>gct</u> gtg <u>tat</u> <u>tac</u> <u>tgt</u> gcc aga ga	
seq4	--- --- --- --- c-- ---	
seq5	--- ---	
seq6	--- ---	
L06615 IGHV3-30*04		IGHJ4*02
seq3	ctg gtc acc gtc tcc tca g	IGHJ4*02
seq4	ggc cag gga acc ctg gtc acc gtc tcc tca g	IGHJ4*02
seq5	tac tgg ggc cag gga acc ctg gtc acc gtc tcc tca g	IGHJ4*02
seq6	tgg ggc ca	IGHJ1*01

FIGURE 16. IMGT/V-QUEST “Synthesis view” result page: “Alignment for V-GENE” and “V-REGION alignment according to the IMGT unique numbering.” The sequences are displayed with the delimitations of FR-IMGT and CDR-IMGT. Dashes indicate identical nucleotides. Dots indicate gaps. FR-IMGT and CDR-IMGT are delimited according to the IMGT unique numbering (Lefranc 1997, 1999; Lefranc et al. 2003). Displayed sequences are those of the input set, which have been identified as using the same closest germline gene and allele (here IGHV3-30*04). The sequence of the V-REGION of IGHV3-30*04 is shown on the first line and the input sequences are aligned, with nucleotide mutations indicated by comparison with IGHV3-30*04. Underlined nucleotides in IGHV3-30*04 correspond to mutation hot spot motif localizations.

similar to those of the Excel file, formatted in comma separated value [CSV]). Rules for sequence analysis are improved through scientific collaboration and with the constant feedback of scientists. IMGT/V-QUEST has been recommended by the European Research Initiative on chronic lymphocytic leukemia CLL (ERIC) for the analysis of the IGHV gene mutational status in CLL (Ghia et al. 2007). The level of mutations in IGHV genes of CLL patients is a prognostic criterion. Moreover, >20% of patients with

4. V-REGION protein display

[illegible]

5. V-REGION protein display with colored AA according to the AA IMGT classes

	FR1-IMGT (1-26)					CDR1-IMGT (27-38)					FR2-IMGT (39-55)					CDR2-IMGT (56-65)					FR3-IMGT (66-104)				
	1	10	20	30	40	40	50	60	70	80	90	100													
L06615	I	G	H	V	R	Q	A	G	K	G	L	E	V	A	V	I	S	D	.	G	N	K	.	A	
seq3	V	V	L	E	S	G	.	V	V	Q	G	R	L	R	S	.	A	A	S	.	G	F	T	.	
seq4	I	V	L	L	E	S	G	.	V	V	Q	G	R	L	R	S	.	A	A	S	.	G	F	T	
seq5	V	V	L	E	S	G	.	V	V	Q	G	R	L	R	S	.	A	A	S	.	G	F	T	.	
seq6	V	V	L	E	S	G	.	V	V	Q	G	R	L	R	S	.	A	A	S	.	G	F	T	.	

6. V-REGION protein display (only AA changes displayed)

[illegible]

FIGURE 17. IMGT/V-QUEST “Synthesis view” result page: V-REGION protein displays. Three different displays are shown. “V-REGION protein display,” “V-REGION protein display with colored AA according to the AA IMGT classes,” and “V-REGION protein display (only AA changes displayed).” The first line shows the AA sequence of the closest V-REGION allele against which are aligned the translations of the input sequences identified by IMGT/V-QUEST as using the same gene and allele (here IGHV3-30*04).

CLL carry stereotyped receptors with pathogenetic implications and clinical correlations (Stamatopoulos et al. 2007). Results obtained by the users have important implications in basic and medical research, antibody and TR specificity, and in biotechnology related to antibody engineering (Giudicelli and Lefranc 2008; Lefranc 2009; Ehrenmann et al. 2010). IMGT standards are approved by the World Health Organization-International Union of Immunological Societies (WHO-IUIS) subcommittee for IG and TR (Lefranc 2007, 2008).

7. V-REGION most frequently occurring AA per position and per FR-IMGT and CDR-IMGT

FR1-IMG1																											
Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
Occurrence	G	V	G	L	V	A	A	A	G	G	G	V	V	G	A	G	R	A	L	R	L	A	A	A	A	A	
Frequency	3/4	4/4	4/4	4/4	3/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	3/4	4/4	4/4	4/4	4/4	4/4	100	
	75%	100%	100%	100%	75%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	75%	100%	100%	100%	75%	100%	100	

Position	27	28	29	30	31	32	33	34	35	36	37	38
AA	G	F	I	F					G	A	A	A
Occurrence nb	4/4	4/4	4/4	4/4					4/4	4/4	4/4	3/4
Frequency	100%	100%	100%	100%					100%	100%	100%	75%

[illegible]

Position	56	57	58	59	60	61	62	63	64	65
AA	I	S					G		H	K
Occurrence	4/4	4/4	4/4	4/4			4/4	4/4	3/4	4/4
Frequency	100%	100%	100%	100%			100%	100%	75%	100%

[illegible]

CDR3-IMGT	
Position	105
AA	A
Occurrence nb	4/4
Frequency	100%

FIGURE 18. IMGT/V-QUEST “Synthesis view” result page: “V-REGION most frequently occurring AA per position and per FR-IMGT and CDR-IMGT.” This section shows, for each FR-IMGT and CDR-IMGT, and for each codon (or AA), the position of the most frequently occurring AA.

ACKNOWLEDGMENTS

We thank François Ehrenmann and Patrice Duroux for checking the protocol. We are grateful to Gérard Lefranc and to the IMGT team for their motivation and expertise.

REFERENCES

- Alamyar E, Giudicelli V, Duroux P, Lefranc M-P. 2010. IMGT/HighV-QUEST: A high-throughput system and Web portal for the analysis of rearranged nucleotide sequences of antigen receptors - High-throughput version of IMGT/V-QUEST. JOBIM 2010 Poster 60. <http://www.jobim2010.fr/?q=fr/node/55>.
- Alamyar E, Duroux D, Lefranc M-P, Giudicelli V. 2011. IMGT tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol* (in press).
- Brochet X, Lefranc M-P, Giudicelli V. 2008. IMGT/V-QUEST: The highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucl Acids Res* **36**: W503–W508.
- Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V. 2008. IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. *Biochimie* **90**: 570–583.
- Ehrenmann F, Lefranc M-P. 2011a. IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant and groove domains (IG, TR, MH, IgSF, MhSF). *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot5636.
- Ehrenmann F, Lefranc M-P. 2011b. IMGT/3Dstructure-DB: Querying the IMGT database for 3D structures in immunology and immunoinformatics (IG or antibodies, TR, MH, RPI, and FPIA). *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot5637.
- Ehrenmann F, Duroux P, Giudicelli V, Lefranc M-P. 2010. Standardized sequence and structure analysis of antibody using IMGT. In *Antibody engineering* (ed. R Kontermann, S Dübel), 2nd ed, Vol 2, pp. 11–31. Springer-Verlag, Berlin/Heidelberg, Germany.
- Ehrenmann F, Giudicelli V, Brochet X, Lefranc M-P. 2011. IMGT/Collier de Perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot5635.
- Elemento O, Lefranc M-P. 2003. IMGT/PhyloGene: An on-line tool for comparative analysis of immunoglobulin and T cell receptor genes. *Dev Comp Immunol* **27**: 763–779.
- Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson F, Davi F, Rosenquist R. 2007. ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* **21**: 1–3.
- Giudicelli V, Lefranc M-P. 1999. Ontology for immunogenetics: IMGT-ONTOLOGY. *Bioinformatics* **15**: 1047–1054.
- Giudicelli V, Lefranc M-P. 2005. Interactive IMGT on-line tools for the analysis of immunoglobulin and T cell receptor repertoires. In *New research on immunology* (ed. BA Veskler), pp. 77–105. Nova Science, Hauppauge, NY.
- Giudicelli V, Lefranc M-P. 2008. IMGT standardized analysis of immunoglobulin rearranged sequences. In *Immunoglobulin gene analysis in Chronic Lymphocytic Leukemia* (ed. P Ghia, R Rosenquist, F Davi), pp. 33–52. Wolters Kluwer Health Italy Ltd, Milan, Italy.
- Giudicelli V, Lefranc M-P. 2011. IMGT/JunctionAnalysis: IMGT standardized analysis of the V-J and V-D-J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb Protoc* doi: 10.1101/pdb.prot5634.
- Giudicelli V, Protat C, Lefranc M-P. 2003. The IMGT strategy for the automatic annotation of IG and TR cDNA sequences: IMGT/Automat. In *Proceedings of the european conference on computational biology*, pp. 103–104. Paris, France.
- Giudicelli V, Chaume D, Lefranc M-P. 2005a. IMGT/GENE-DB: A comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* **33**: D256–D261.
- Giudicelli V, Chaume D, Jabado-Michaloud J, Lefranc M-P. 2005b. Immunogenetics sequence annotation: The strategy of IMGT based on IMGT-ONTOLOGY. *Stud Health Technol Inform* **116**: 3–8.
- Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, Lefranc M-P. 2006. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucl Acids Res* **34**: D781–D784.
- Kaas Q, Lefranc M-P. 2007. IMGT Colliers de Perles: Standardized sequence-structure representations of the IgSF and MhSF superfamily domains. *Current Bioinformatics* **2**: 21–30.
- Kaas Q, Ruiz M, Lefranc M-P. 2004. IMGT/3Dstructure-DB and IMGT/StructuralQuery, a database and a tool for immunoglobulin, T cell receptor and MHC structural data. *Nucl Acids Res* **32**: D208–D210.
- Kaas Q, Ehrenmann F, Lefranc M-P. 2007. IG, TR and IgSF, MHC and MhSF: What do we learn from the IMGT Colliers de Perles? *Brief. Funct Genomic Proteomic* **6**: 253–264.
- Lefranc M-P. 1997. Unique database numbering system for immunogenetic analysis. *Immunol Today* **18**: 509. doi: 10.1016/S0167-5699(97)01163-8.
- Lefranc M-P. 1999. The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *The Immunologist* **7**: 132–136.
- Lefranc M-P. 2000a. Nomenclature of the human immunoglobulin genes. In *Current protocols in immunology* (ed. JE Coligan, BE Bierer, DE Margulies, EM Shevach, W Strober), pp. A.1P.1–A.1P.37. John Wiley and Sons, Inc, Hoboken, NJ.
- Lefranc M-P. 2000b. Nomenclature of the human T cell receptor genes. In *Current protocols in immunology* (ed. JE Coligan, BE Bierer, DE Margulies, EM Shevach, W Strober), pp. A.1O.1–A.1O.23. John Wiley and Sons, Inc, Hoboken, NJ.
- Lefranc M-P. 2004. IMGT, the international ImMunoGenetics information system. In *Antibody engineering methods and protocols* (ed. BKC Lo), 2nd ed, Vol 248, pp. 27–49. Methods in Molecular Biology, Humana Press, Totowa, NJ.
- Lefranc M-P. 2007. WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report. *Immunogenetics* **59**: 899–902.
- Lefranc M-P. 2008. WHO-IUIS Nomenclature Subcommittee for Immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev Comp Immunol* **32**: 461–463.
- Lefranc M-P. 2009. Antibody database and tools: The IMGT experience. In *Therapeutic monoclonal antibodies: from Bench to Clinic* (ed. A Zhiqiang), pp. 91–114. John Wiley Sons, Inc, Hoboken, NJ.
- Lefranc M-P. 2011a. IMGT, the International ImMunoGeneTics Information System. *Cold Spring Harb Protoc* doi: 10.1101/pdb.top115.
- Lefranc M-P. 2011b. From IMGT-ONTOLOGY IDENTIFICATION axiom to IMGT standardized keywords: For immunoglobulins (IG), T cell receptors (TR), and conventional genes. *Cold Spring Harb Protoc* doi: 10.1101/pdb.ip82.
- Lefranc M-P. 2011c. From IMGT-ONTOLOGY DESCRIPTION axiom to IMGT standardized labels: For immunoglobulin (IG) and T cell receptor (TR) sequences and structures. *Cold Spring Harb Protoc* doi: 10.1101/pdb.ip83.
- Lefranc M-P. 2011d. From IMGT-ONTOLOGY CLASSIFICATION axiom to IMGT standardized gene and allele nomenclature: For immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb Protoc* doi: 10.1101/pdb.ip84.

- Lefranc M-P. 2011e. IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb Protoc* doi: 10.1101/pdb.ip85.
- Lefranc M-P. 2011f. IMGT Collier de Perles for the Variable (V), Constant (C), and Groove (G) Domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb Protoc* doi: 10.1101/pdb.ip86.
- Lefranc M-P, Lefranc G. 2001a. *The Immunoglobulin FactsBook*. pp. 1–458. Academic Press, London, UK.
- Lefranc M-P, Lefranc G. 2001b. *The T cell receptor FactsBook*. pp. 1–398. Academic Press, London, UK.
- Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* **27**: 55–77.
- Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, et al. 2004. IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics. *In Silico Biol*, Epub 2003 Nov 22; **4**: 17–29.
- Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, et al. 2005a. IMGT-Choreography for Immunogenetics and Immunoinformatics. *In Silico Biol* **5**: 45–60.
- Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M, et al. 2005b. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* **29**: 185–203.
- Lefranc M-P, Duprat E, Kaas Q, Tranne M, Thiriot A, Lefranc G. 2005c. IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev Comp Immunol* **29**: 917–938.
- Lefranc M-P, Giudicelli V, Regnier L, Duroux P. 2008. IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief Bioinform* **9**: 263–275.
- Lefranc M-P, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, et al. 2009. IMGT, the international ImMunoGeneTics information system. *Nucl Acids Res* **37**: D1006–D1012.
- Pommié C, Levadoux S, Sabatier R, Lefranc M-P. 2004. IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J Mol Recognit* **17**: 17–32.
- Ruiz M, Lefranc M-P. 2002. IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics* **53**: 857–883.
- Stamatopoulos K, Belessi C, Moreno C, Boudjograh M, Guida G, Smilevska T, Belhouli L, Stella S, Stavroyianni N, Crespo M, et al. 2007. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood* **109**: 259–270.
- Yousfi Monod M, Giudicelli V, Chaume D, Lefranc M-P. 2004. IMGT/JunctionAnalysis: The first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics* **20**: i379–i385.



Cold Spring Harbor Protocols

IMGT/V-QUEST: IMGT Standardized Analysis of the Immunoglobulin (IG) and T Cell Receptor (TR) Nucleotide Sequences

Véronique Giudicelli, Xavier Brochet and Marie-Paule Lefranc

Cold Spring Harb Protoc; doi: 10.1101/pdb.prot5633

Email Alerting Service

Receive free email alerts when new articles cite this article - [click here](#).

Subject Categories

Browse articles on similar topics from *Cold Spring Harbor Protocols*.

[Alignment of Sequences](#) (46 articles)
[Bioinformatics/Genomics, general](#) (193 articles)
[Characterization of Proteins](#) (208 articles)
[Computational Biology](#) (100 articles)
[Genetics, general](#) (380 articles)
[Immunology, general](#) (128 articles)
[Phylogenetic Prediction](#) (19 articles)
[Sequence Database Searching](#) (38 articles)

To subscribe to *Cold Spring Harbor Protocols* go to:
<http://cshprotocols.cshlp.org/subscriptions>
