

# IMGT/V-QUEST and IMGT/JunctionAnalysis:

## The standardized approach for IG and TR rearrangement analysis

Xavier Brochet

Laboratoire d'ImmunoGénétique Moléculaire  
Université Montpellier, UPR CNRS 1142, IGH

First Immunomics Summer School  
24 August - 3 September 2007



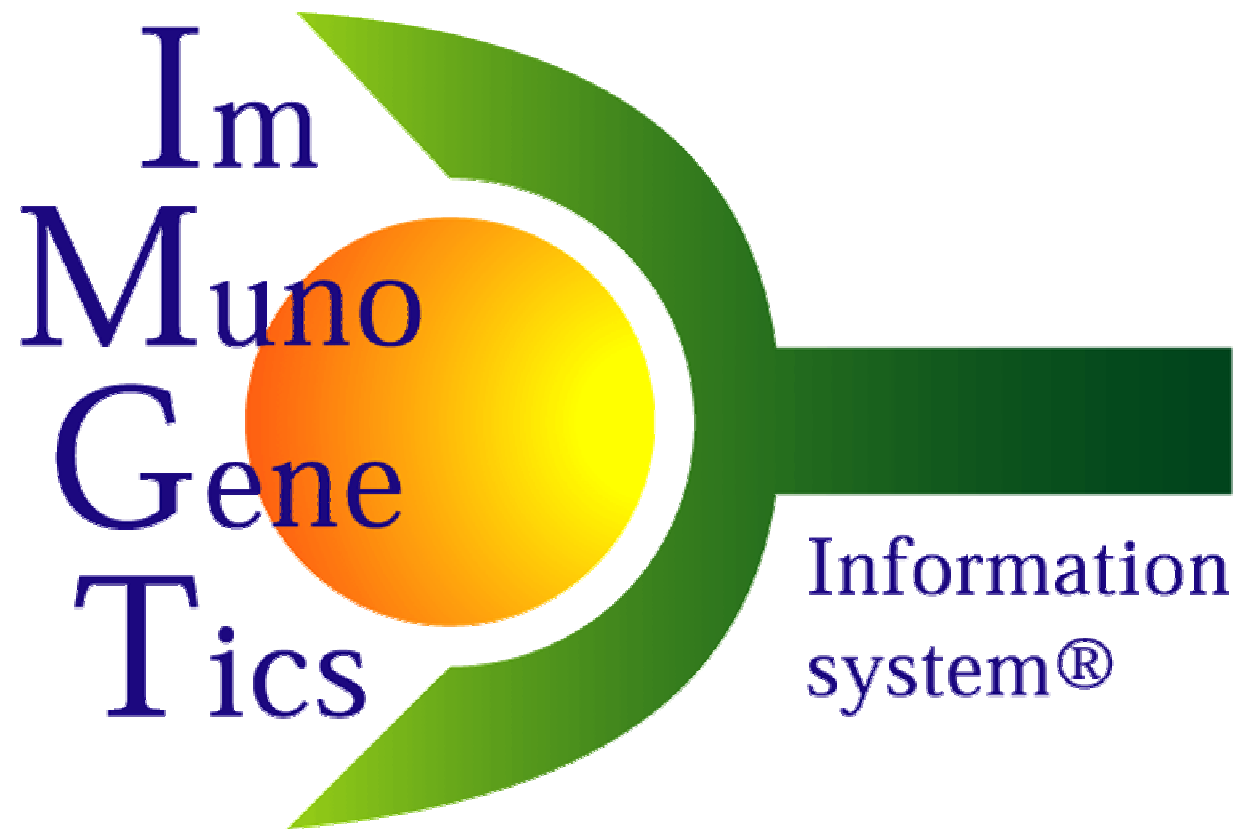
The international ImMunoGeneTics information system®

Founder and director : M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>



IMGT was created in June **1989**, during the Human Genome Meeting HGM 10



IMGT® was created to answer the need to manage the immunoglobulin (IG) and T cell receptor (TR) related data.

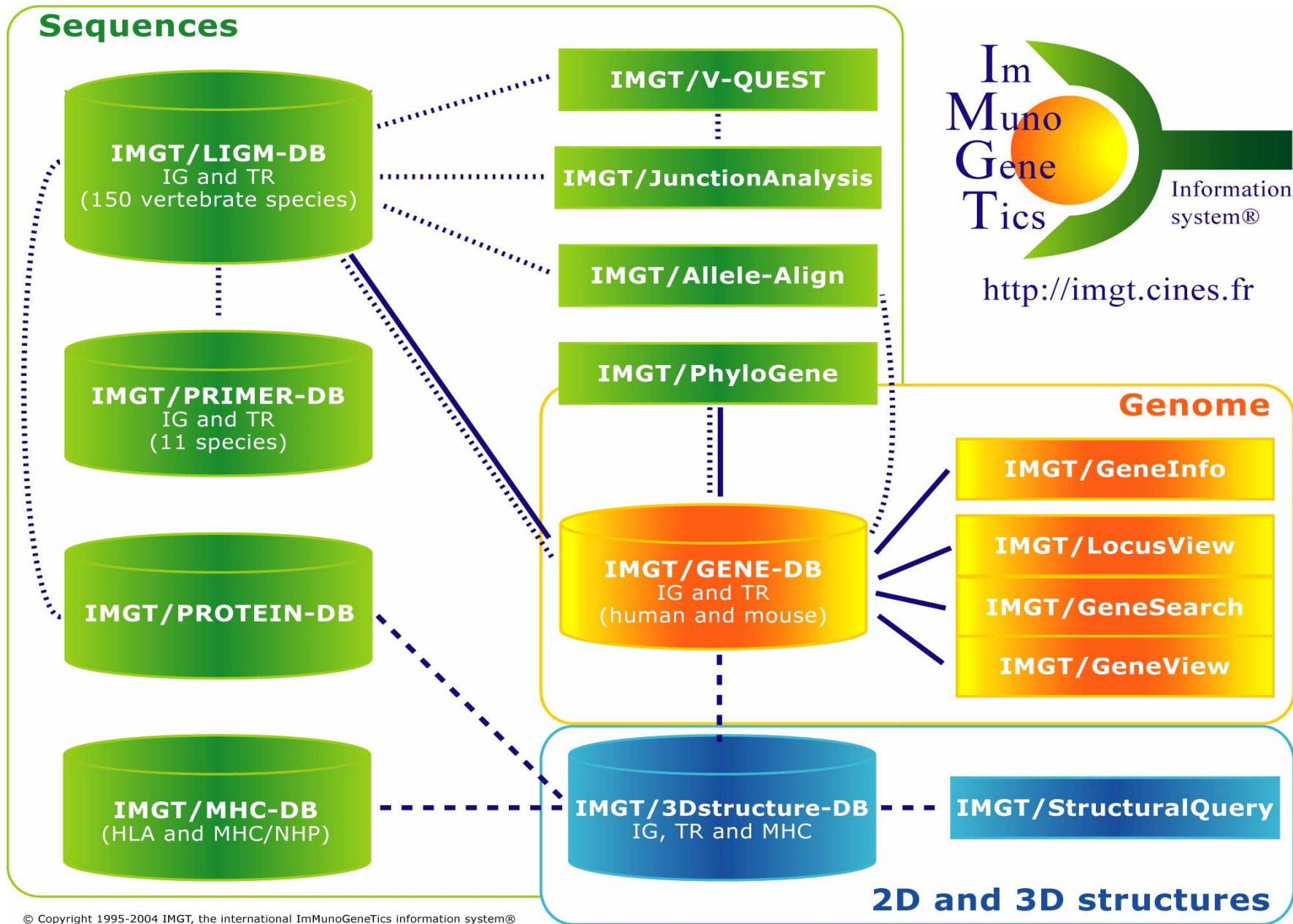


The international ImMunoGeneTics information system®

Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>





© Copyright 1995-2004 IMGT, the international ImMunoGeneTics information system®



The international ImMunoGeneTics information system®  
 Founder and director: M.-P. Lefranc, Montpellier, France

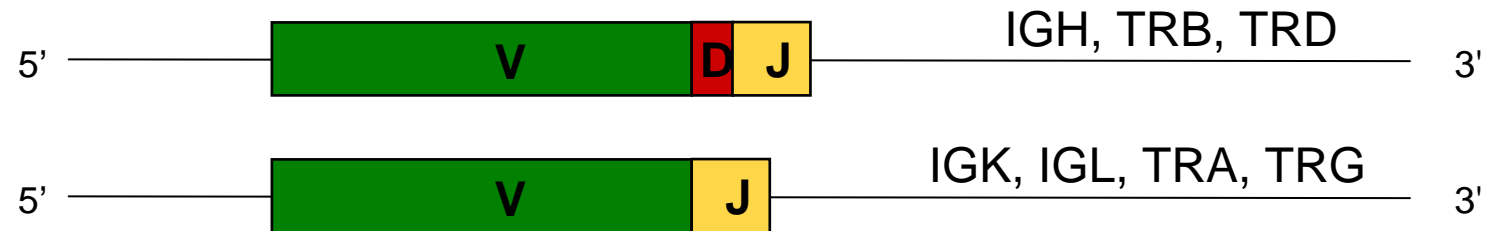
<http://imgt.cines.fr>



# IMGT/V-QUEST and IMGT/JunctionAnalysis

**IMGT/V-QUEST** and **IMGT/JunctionAnalysis**, part of IMGT®, the international ImMunoGeneTics information system®, analyse the IG and TR rearranged V-D-J and V-J sequences.

**Rearranged DNA**



- 1) identify the V, (D) and J genes and alleles involved in the rearrangements.
- 2) provide a detailed and accurate description of the sequences.

Results are according to the IMGT Scientific chart rules, based on the IMGT-ONTOLOGY axioms.

# IMGT/V-QUEST and IMGT/JunctionAnalysis

## IMGT/V-QUEST

- compares the user's sequences with the IMGT reference sequences (IMGT reference directory: CLASSIFICATION axiom)
- displays the nucleotide and protein alignments according to the IMGT Scientific Chart rules (IMGT labels: DESCRIPTION axiom)
- provides an extensive analysis of the mutations (IMGT unique numbering: NUMEROTATION axiom)

## IMGT/JunctionAnalysis

- analyses accurately the junctions of IG and TR rearranged sequences: identifies IGHD, TRBD and TRDD genes and alleles, N- and P REGION, "gc" content, amino acid physicochemical properties, PI, etc.

*Giudicelli et al. Nucleic Acids Res 32:435-440 (2004)*  
*Yousfi Monod et al. Bioinformatics 20:379-385 (2004)*



The international ImMunoGeneTics information system®

Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>

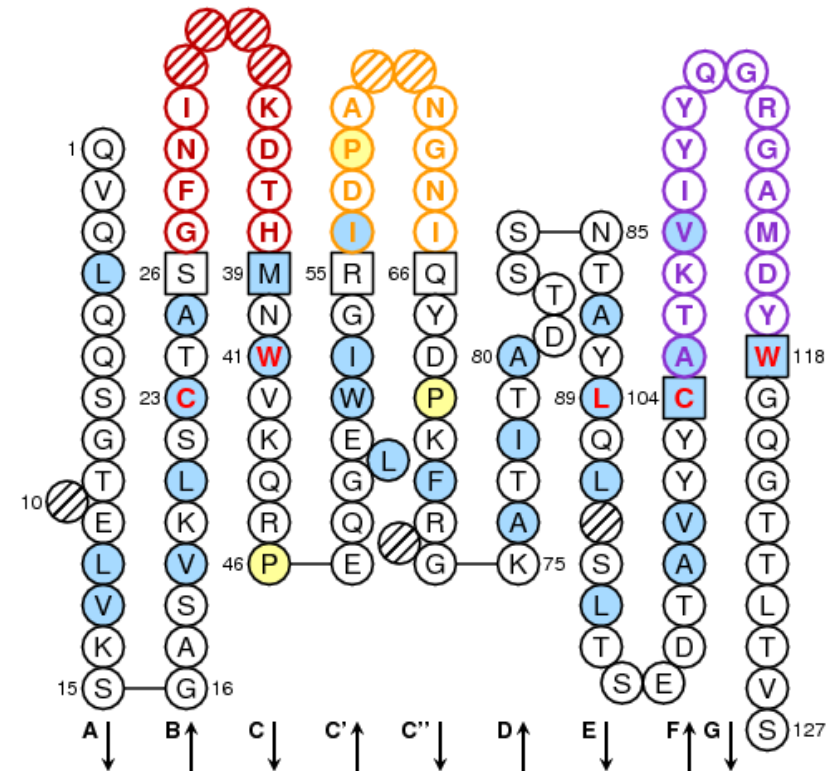


# Specificities of IMGT/V-QUEST algorithm

## IMGT unique numbering rules

- maintain conserved AA, and therefore codons, at the same positions
- standardize the delimitations of FR-IMGT and CDR-IMGT

Gaps are inserted in IMGT reference sequences, taking into account the variable lengths of the CDR-IMGT



**IMGT proposes for the first time a standardized description of the V-DOMAIN, whatever the antigen receptor, the chain type and the species.**

# IMGT/V-QUEST algorithm

In order to respect the IMGT unique numbering rules:

- The IMGT/V-QUEST algorithm is based on a global alignment which does not accept any insertion and/or deletion
- Gaps are integrated into the user sequence, according to the IMGT unique numbering, by comparison with a model sequence of the IMGT reference directory.



The international ImMunoGeneTics information system®  
Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>




# IMGT/V-QUEST web interface (<http://imgt.cines.fr>)

## WELCOME ! to the IMGT/V-QUEST Search page

THE  
INTERNATIONAL  
IMMUNOGENETICS  
INFORMATION SYSTEM®

Citing IMGT/V-QUEST: Giudicelli, V. et al. Nucl. Acids Res. 2004, 32, W435-440 [PMID: 15215425](#) [PDF](#)

 You are in the new IMGT/V-QUEST, upgraded for multiple sequences and with new functionalities. **NEW!**

### Analyse your Immunoglobulin nucleotide sequences

-  [Human](#)
-  [Mouse](#)
-  [Rat \(test for IGHV2 and IGHV5 subgroups\)](#)
-  [Chondrichthyes](#)
-  [Teleostei](#)
  - [Atlantic cod](#)
  - [Channel catfish](#)
  - [Rainbow trout](#)
-  [Sheep](#)

### Analyse your T cell Receptor nucleotide sequences

-  [Human](#)
-  [Mouse](#)
-  [Non-human primates](#)



The international ImMunoGeneTics information system®

Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>





# IMGT/VQUEST+JCTA QUERY PAGE

Enter your sequence(s) in **FASTA format** (FASTA format is required):

Type (or copy/paste) your sequence(s) into the box below :

```
>AF184762
atggagtttgggctgagctggggttttctgttgctatittaaagggtgccactgtgag
gtgcagctgggtggagcttgggggaggttagtccagcctgggggatccctgaaactctc
ftgtcagcctctgggttcaccctcagtggtcctcaaatgtgcactgggtccgccaggcctc
gggaaaggctggagtggttggccgtatcaaaagggaatgctgagctgacgcgacagca
tatgtcgtcgtgatgagagcaggctcaccatctccagagatgattcaagaacacggcg
tttctgcaaatgaacagcctgaaaagcgtgacacggcctgattattgtgtgatccgg
ggagatgtttacaaccgacagtggggcccagggaaacctggtcaccgctcctcagcatcc
ccgaccagcccccaaggtcttcccgtgagcctctgcagcaccagccagat]
```

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

Browse...

Start Clear the form

Selection of parameters for the results

Display type : HTML

Nb of nucleotides per line in alignment: 60

**A. Detailed view**

- Alignment for V-GENE
- Alignment for D-GENE
- Alignment for J-GENE
- Results of IMGT/JunctionAnalysis
  - with full list of eligible D-GENES
  - without list of eligible D-GENES
- Sequence of the JUNCTION ('nt' and 'AA')
- V-REGION alignment according to the IMGT numbering
- V-REGION translation
- V-REGION mutation table
- V-REGION mutation statistics
- V-REGION mutation hot spots
- Sequences of V-, V-J- or V-D-J- REGION ('nt' and 'AA') with gaps in FASTA  
Access to IMGT/PhyloGene for V-REGION ('nt')
- Annotations by IMGT/Automat

11. [IMGT Collier de Perles](#)

- links to IMGT Collier de Perles
- IMGT Collier de Perles (PNG format, slow)
- no IMGT Collier de Perles

**Synthetic view per allele**

- Alignment for V-GENES
- V-REGION alignment according to the IMGT numbering
- V-REGION translation
- V-REGION protein display
- V-REGION protein display (with color)
- V-REGION protein display (mutations displayed)
- V-REGION most frequently occurring AA
- Results of IMGT/JunctionAnalysis

Analysis of several sequences (in FASTA format) in a single run

Selection of output parameters

Detailed view for each sequence

Synthetic view per allele



# 1. "Detailed view" results: Summary

Sequence number 1: AF184762

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

>AF184762

```
atggagtttgggctgagctggggtttccttggtgctatnttaaaagggtgtccactgtgag
gtgcagctgggtggagctcgggggaggcttagtcagcctgggggatccctgaaactctcc
tgtgcagcctctgggttcaccctcagtggtcctcaaatgtgcactgggtccgccaggcctcc
gggaaagggtggagtggttgccgtatcaaaaggaatgctgagctgacgacagca
tatgctgctgatgagaggcaggctcaccatctccagagatgattcaagaacacggcg
ttctgcaaatgaacagcctgaaaagcgatgacacggccatgtattattgtgtgatccgg
ggagatggttacaaccgacagtgggggccaggaaccctggtcaccgtctctcagcatcc
ccgaccagccccaaggctctcccgctgagcctctgcagcaccagccagat
```

Automatic interpretation

Result summary:	Productive IGH rearranged sequence (no stop codon and in frame junction)		
V-GENE and allele	<a href="#">IGHV3-73*01</a>	score = 1240	identity = <b>91,50%</b> (269/294 nt)
J-GENE and allele	<a href="#">IGHJ1*01 (a)</a>	score = 179	identity = 82,69% (43/52 nt)
D-GENE and allele by IMGT/JunctionAnalysis	<a href="#">IGHD3-10*01</a>	D-REGION is in reading frame 3	
CDR-IMGT lengths and AA JUNCTION	[8,10,10]	CVIRGDVYNRQW	

(a) Other possibilities: IGHJ4\*02 and IGHJ5\*02 (highest number of consecutive identical nucleotides)

## 2. "Detailed view" results: V, D and J alignments

### V-GENE

#### 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
<a href="#">Z27508</a> IGHV3-73*01	1240	91,50% (269/294 nt)
<a href="#">AB019437</a> IGHV3-73*02	1231	91,16% (268/294 nt)
<a href="#">X92206</a> IGHV3-72*01	1024	83,33% (245/294 nt)
<a href="#">X92216</a> IGHV3-15*01	979	81,63% (240/294 nt)
<a href="#">M99406</a> IGHV3-15*07	979	81,63% (240/294 nt)

Alignment with [FR-IMG1](#) and [CDR-IMG1](#) delimitations

```

AF184762                <----- FRL-IMG1 -----
Z27508  IGHV3-73*01    gggtgcagctggtggagtctggggga...ggcttagtccagcctggggatccctgaaa
AB019437  IGHV3-73*02  .....g.....g.....g.....g.....
X92206  IGHV3-72*01    .....g.....g.....g.....g.....
X92216  IGHV3-15*01    .....g-aa.....g---t-g-
M99406  IGHV3-15*07    .....g-aa.....g---t-g-
    
```

Score and nucleotide identity

```

AF184762                <----- t.....gtgcac----->
Z27508                 .....a.....
AB019437  IGHV3-73*02  .....t.....tgC.....a----
X92206  IGHV3-72*01    .....a.....t.....a-cact-c.....a--g--
X92216  IGHV3-15*01    .....a.....tt.....aa-g-ctgg.....a-ag-
M99406  IGHV3-15*07    .....t.....tt.....aa-g-ctgg.....a--a-
    
```

### D-GENE

#### 2. Alignment for D-GENE and allele identification

Closest D-REGIONS

	Score	Identity
<a href="#">X13972</a> IGHD3-10*01	28	72,73% (8/11 nt)
<a href="#">X93615</a> IGHD3-10*02	19	63,64% (7/11 nt)
<a href="#">J00256</a> IGHD7-27*01	14	60,00% (6/10 nt)
<a href="#">X97051</a> IGHD3-16*02	13	62,50% (5/8 nt)
<a href="#">X93614</a> IGHD3-16*01	10	54,55% (6/11 nt)

Alignment

```

AF184762                gtgatccgggg
X13972  IGHD3-10*01    a--g-t-----agttattataac
X93615  IGHD3-10*02    a--t-g---agttattataac
J00256  IGHD7-27*01    .ct-a-t-----a
X97051  IGHD3-16*02    --t--a-c
X93614  IGHD3-16*01    ac-t-tg---agttatgctta
    
```

#### 3. Alignment for J-GENE and allele identification

Closest J-REGIONS

	Score	Identity
<a href="#">J00256</a> IGHJ1*01	179	82,69% (43/52 nt)
<a href="#">X86355</a> IGHJ4*02	150	79,17% (38/48 nt)
<a href="#">J00256</a> IGHJ4*01	141	77,08% (37/48 nt)
<a href="#">X86355</a> IGHJ5*02	138	74,51% (38/51 nt)
<a href="#">M25625</a> IGHJ4*03	132	75,00% (36/48 nt)

Note that the highest number of consecutive identical nucleotides has been found in the alignment with IG

Alignment

```

AF184762                gtgatccgggggagatgtttacaaccgacagtggggccagggaaccctggtcaccgtctcc
J00256  IGHJ1*01         .....c--aa---tt--ag--c-----c-----
X86355  IGHJ4*02         .....ac---tttgact-c-----
J00256  IGHJ4*01         .....ac---tttgact-c-----a-----
X86355  IGHJ5*02         .....caac-gggt-gac-cc-----
M25625  IGHJ4*03         .....gc---tttgact-c-----a--g-----
    
```

### J-GENE

# 3. "Detailed view" results: IMGT/JunctionAnalysis

## 4. Results of IMGT/JunctionAnalysis

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

### 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
AF184762	<a href="#">IGHV3-73*01</a>	tgt.....	g	.....tgatcggggagatggtt.....	tacaaccga	.....cagtgg	<a href="#">IGHJ1*01</a>	<a href="#">IGHD3-10*01</a>	0	4	1	5/10

### Translation of the JUNCTION

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

	104	105	106	107	108	109	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass	pI
	C	V	<u>I</u>	R	G	<u>D</u>	<u>V</u>	Y	N	R	Q	W	+	10	1,508.72	9.24
AF184762	tgt	gtg	<u>atc</u>	cgg	gga	<u>gat</u>	<u>ggt</u>	tac	aac	cga	cag	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')

104 105 106 107 108 109 113 114 115 116 117 118  
 C V I R G D V Y N R Q W  
 tgt gtg atc cgg gga gat ggt tac aac cga cag tgg



The international ImMunoGeneTics information system®

Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>



# 4. "Detailed view" results: mutations analysis

## V-REGION mutation table

FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
g36>a	t88>c, F30>L (+ -)	a115>g, M39>V (+ -)	t168>c	g211>a, V71>M (+ -)	a313>g, T105>V (- -)
g51>a	t99>a	t135>c	g170>a, R57>K (+ + +)	a215>g, K72>R (+ + +)	c314>t, T105>V (- -)
	g100>a, A34>N (- -)		c174>g, S58>R (- -)	t226>c, F76>L (+ -)	t315>g, T105>V (- -)

## Mutation hotspot (motifs and positions in germline V-REGION)

(a/t)a wa		(a/g)g(c/t)(a/t) rgyw		(a/t)(a/g)c(c/t) wrsy		t(a/t) tw	
Motif	Positions	Motif	Positions	Motif	Positions	Motif	Positions
aa	58-59 (FR1)	agct	8-11 (FR1)	agct	8-11 (FR1)	tt	34-35 (FR1)
aa	142-143 (FR2)	ggct	31-34 (FR1)	agcc	41-44 (FR1)	tt	82-83 (CDR1)
ta	165-166 (FR2-CDR2)	ggtt	80-83 (CDR1)	aact	59-62 (FR1)	tt	88-89 (CDR1)
ta	168-169 (CDR2)	ggct	94-97 (CDR1)	agcc	72-75 (FR1)	tt	135-136 (FR2)
aa	171-172 (CDR2)	ggct	132-135 (FR2)	tgct	99-102 (CDR1)	tt	158-159 (FR2)
aa	175-176 (CDR2)	ggct	146-149 (FR2)	agct	177-180 (CDR2)	ta	165-166 (FR2-CDR2)
ta	180-181 (CDR2)	ggtt	156-159 (FR2)	tgct	201-204 (FR3)	tt	167-168 (CDR2)
ta	187-188 (CDR2)	agca	172-175 (CDR2)	agcc	277-280 (FR3)	ta	180-181 (CDR2)
ta	199-200 (FR3)	agct	177-180 (CDR2)	aacc	285-288 (FR3)	tt	186-187 (CDR2)
aa	214-215 (FR3)	agtt	184-187 (CDR2)	tact	307-310 (FR3)	ta	199-200 (FR3)
aa	249-250 (FR3)	agca	195-198 (CDR2)	tact	312-315 (FR3)	tt	226-227 (FR3)
aa	253-254 (FR3)	ggca	220-223 (FR3)			tt	246-247 (FR3)
ta	262-263 (FR3)	ggtt	224-227 (FR3)			ta	262-263 (FR3)
aa	269-270 (FR3)					ta	304-305 (FR3)
aa	274-275 (FR3)					tt	306-307 (FR3)
aa	283-284 (FR3)					ta	312-313 (FR3)
aa	285-286 (FR3)					ta	315-316 (CDR3)
ta	304-305 (FR3)						
ta	307-308 (FR3)						
ta	312-313 (FR3)						
ta	315-316 (CDR3)						

## V-REGION mutation statistics

Nucleotides		V-REGION	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
Total nucleotides with gaps		312 (318)	78	36	51	30	117	0 (6)
Mutations	Total	25 (30)	2	4	2	9	8	0 (5)
	Silent	8	2	1	1	3	1	0
Transitions	Non silent	17 (22)	0	3	1	6	7	0 (5)
	a>g	3 (4)	0	0	1	1	1	0 (1)
Transitions	g>a	6	2	1	0	1	2	0
	c>t	1 (2)	0	0	0	0	1	0 (1)
	t>c	4	0	1	1	1	1	0
	a>c	0 (1)	0	0	0	0	0	0 (1)
	c>a	1	0	1	0	0	0	0
	a>t	3	0	0	0	2	1	0
	t>a	1	0	1	0	0	0	0

## V-REGION mutation statistics

### Amino acids

IMGT labels		V-REGION	FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT	
Total amino acids (AA) with gaps		104 (106)	26	12	17	10	39	0 (2)	
Identical AA		89	26	10	16	5	32	0	
AA changes	Total	15 (17)	0	2	1	5	7	0 (2)	
	Conserved IMGT AA classes (hydropathy, volume, chemical)	(- -)	3 (4)	0	1	0	2	0	0 (1)
		(+ + +)	2	0	0	0	1	1	0
		(+ - -)	7	0	1	1	2	3	0
		(- - +)	1 (2)	0	0	0	0	1	0 (1)
		(- - +)	0	0	0	0	0	0	0
		(+ + -)	0	0	0	0	0	0	0
		(+ - +)	2	0	0	0	0	2	0

# Who are the IMGT/V-QUEST users?

## 1. IG and TR genome diversity and evolution studies

- genome diversity and genome evolution studies of the adaptive immune responses
- veterinary research (analysis of IG and TR repertoires in farm, wild life and model animal species)

Comparison of IG and TR from newly sequenced genomes is done versus complete known loci. For example: non-human primates vs human, rat vs mouse, etc.

## 2. Biotechnology related to antibody engineering

- single chain Fragment variable (scFv)
- phage displays
- combinatorial libraries
- chimeric, humanized and human antibodies

Humanization of antibodies based on standardized human FR-IMGT and non-human CDR-IMGT allows to decrease immunogenicity and to maintain specificity.



The international ImMunoGeneTics information system®  
Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>



# Who are the IMGT/V-QUEST users?

## 3- Fundamental and medical research

Repertoire analysis of the IG antibody sites and of the TR recognition sites:

- in normal situations
- in pathological situations (autoimmune diseases, infectious diseases, AIDS, leukemias, lymphomas, myelomas)

Diagnostics

- clonalities in leukemias, lymphomas
- detection and follow up of residual diseases

Prognostic factor in Chronic Lymphocyte Leukemia (CLL)

- “unmutated” IGH V-REGION (identity > 98%): bad prognostic
- “mutated” IGH V-REGION (identity < 98%): good prognostic

**ERIC, the European Research Initiative on CLL ([www.ericll.org](http://www.ericll.org)) recommends IMGT/V-QUEST for such studies**



The international ImMunoGeneTics information system®

Founder and director: M.-P. Lefranc, Montpellier, France

<http://imgt.cines.fr>



# IMGT, the international ImMunoGeneTics information system®

<http://imgt.cines.fr>



The IMGT team at Montpellier, France

## Acknowledgements:

- IMGT team
- ImmunoGrid partners
- MARIE Network

*Many thanks  
to the First  
Immunomics  
Summer  
School organizers*