IMGT tools for interpretation of IG gene sequence data

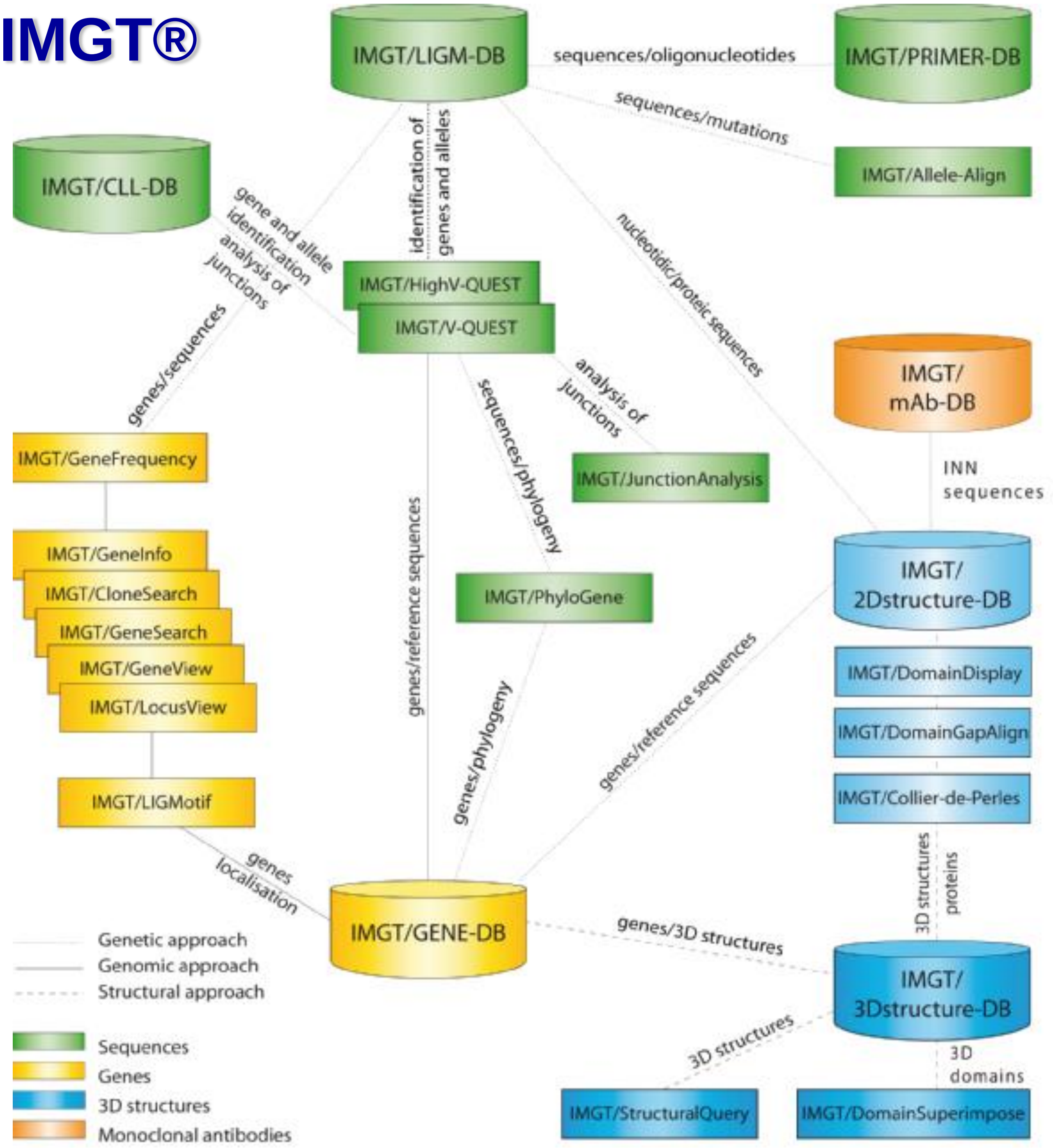
Véronique Giudicelli

IMGT®, the international ImMunoGeneTics information system®
IGH UMR 9002, CNRS – Université de Montpellier,
Montpellier, France

EDUCATIONAL WORKSHOP
Immunoglobulin Gene Sequence Analysis in CLL
March 28, 2019 Bethesda, MD

http://www.imgt.org

IMGT®



ImMunoGeneTics Information system®

http://www.imgt.org

Created in 1989 by
Pr. Marie-Paule Lefranc
(CNRS, Montpellier Univ)
Since 2015, Dir. Emeritus
& Dir. Pr Sofia Kossida

- 7 databases
- 17 online tools
- 20,000 web pages

  - Sequences
  - Genes
  - Structures

- Immunoglobulins (IG) (or antibodies)
- T cell receptors (TR)
- MH
- IgSF and MhSF

Lefranc et al., Nucleic Acids Res. 2015; 43:D413-22 http://www.imgt.org

IMGT-ONTOLOGY seven axioms:

To share, reuse and represent knowledge
in Immunogenetics and Life Sciences

IDENTIFICATION

CLASSIFICATION

DESCRIPTION

NUMEROTATION

LOCALIZATION

ORIENTATION

OBTENTION

Giudicelli and Lefranc, Front. Genet. 2012;3:79

Concepts of CLASSIFICATION

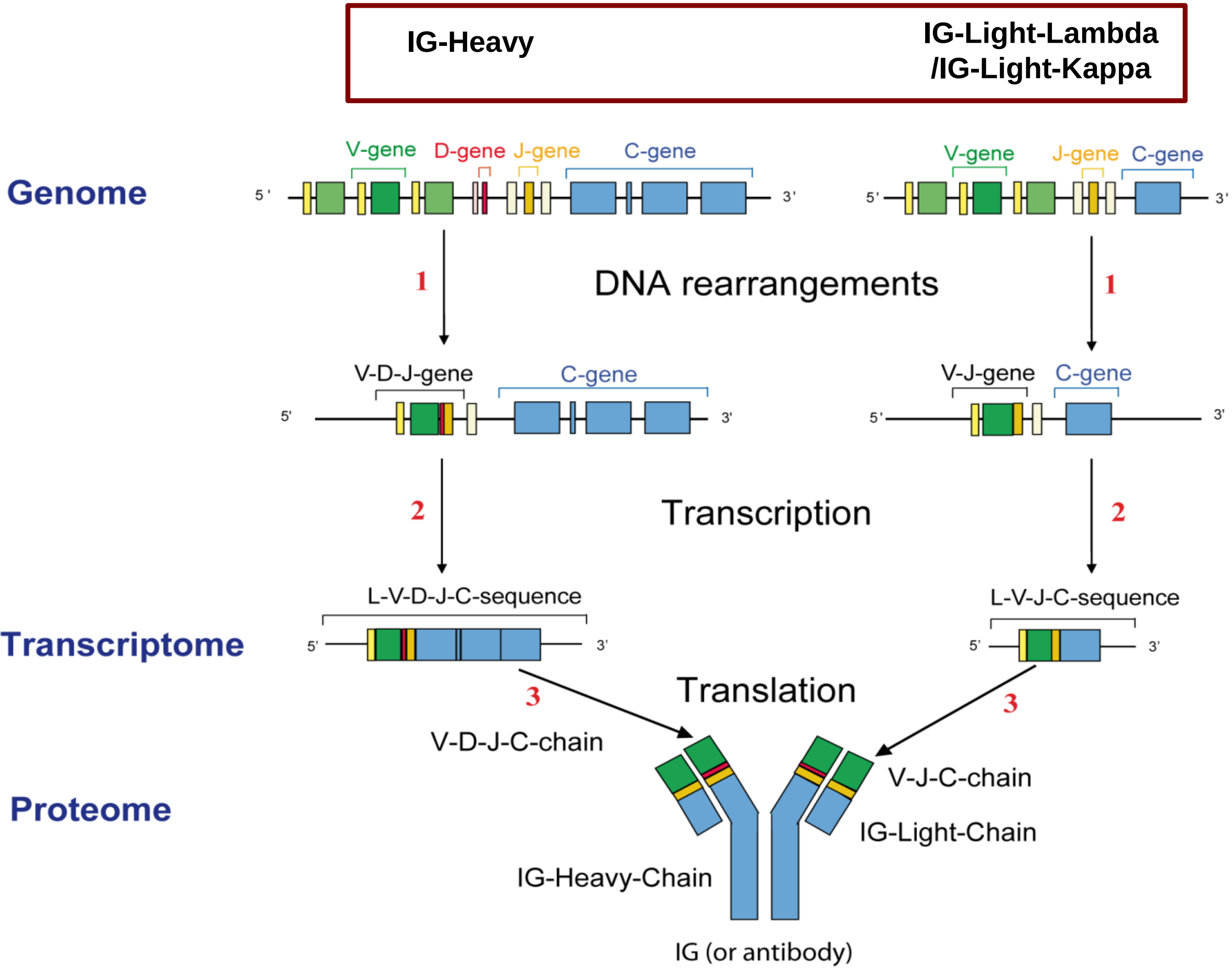
1. Immunoglobulin (IG) and T cell receptor (TR) genes
2. Nomenclature for the V, D, J and C genes
3. Group, subgroup, gene, allele
4. IMGT gene names approved by HUGO Gene Nomenclature Committee (HGNC) in 1999, and endorsed by NCBI in 2000
5. IMGT alleles validated by WHO-IUIS/IMGT-NC

IMGT/GENE-DB: international reference database for IG and TR genes (direct links from NCBI Gene) and alleles.

(462 human IG genes and 997 alleles, 249 TR human genes and 448 alleles)

Concepts of IDENTIFICATION : IMGT standardized keywords

http://www.imgt.org

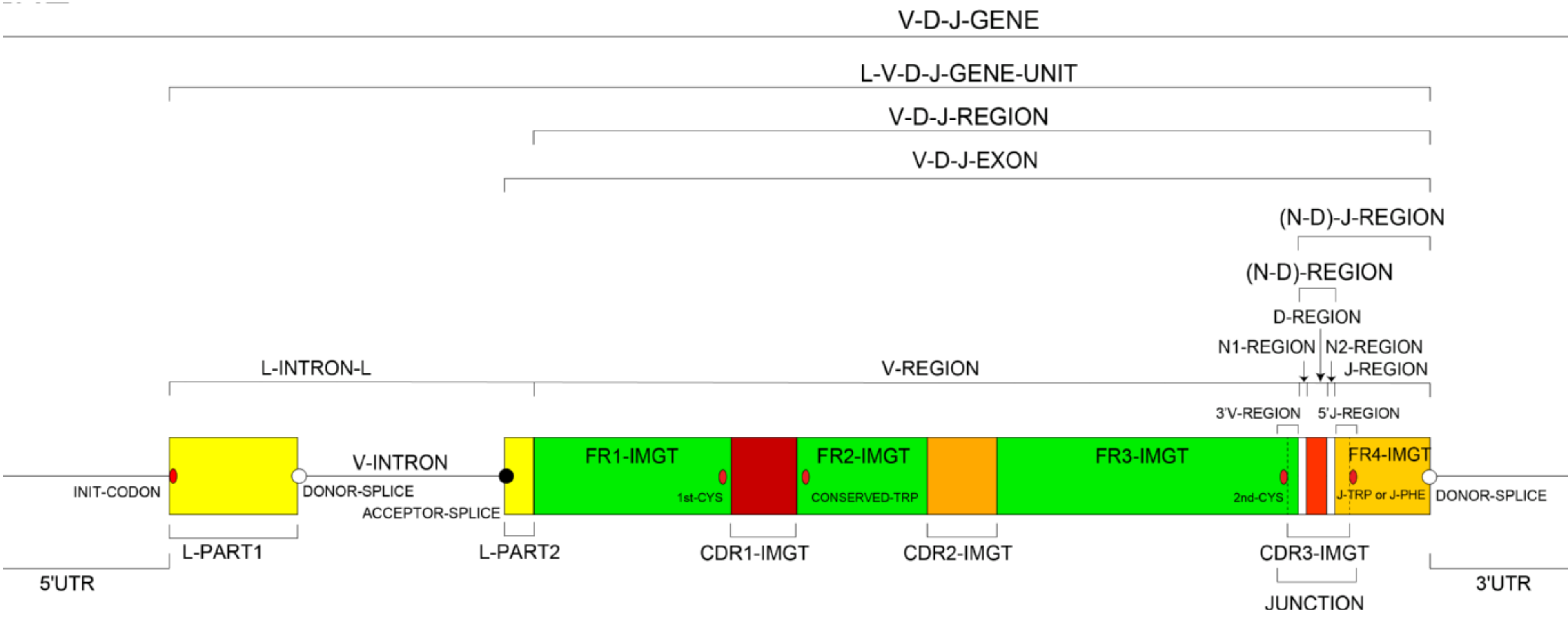


"MoleculeType" concept	"ConfigurationType" concept
gDNA	germline
gDNA	rearranged
mRNA (or *in vitro* cDNA)	rearranged
protein	rearranged

Concepts of DESCRIPTION: IMGT labels

ImMunoGeneTics Information system®
http://www.imgt.org

35 labels for the description of the rearranged V-D-J-GENE in gDNA



Concepts of NUMEROTATION

http://www.imgt.org

1. **IMGT unique numbering**

2. **IMGT Collier de Perles**
   (first one in **1997**)

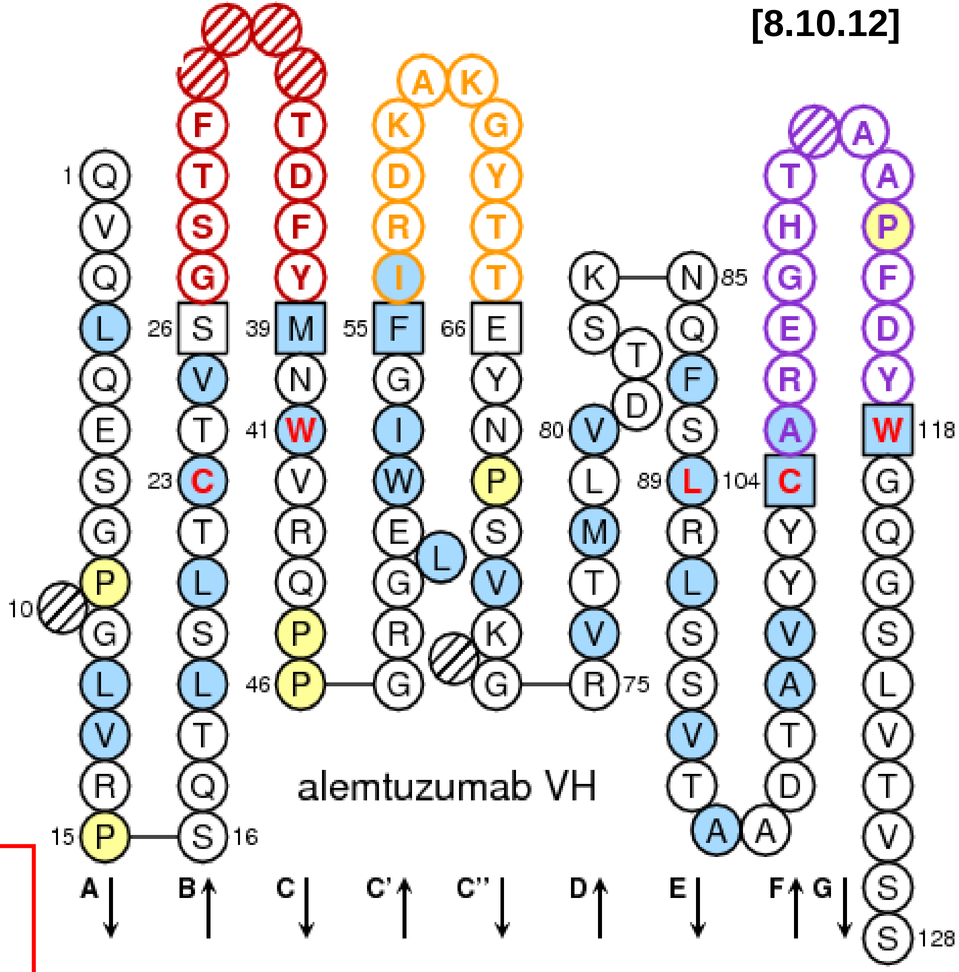
- conserved AA (and codons) always at the **same positions**:

23	1st-CYS
41	CONSERVED-TRP
89	hydrophobic
104	2nd-CYS
118	J-PHE, J-TRP

- six anchors: delimitation of the **FR-IMGT** and **CDR-IMGT**

CDR-IMGT lengths are crucial information

CDR-IMGT lengths [8.10.12]



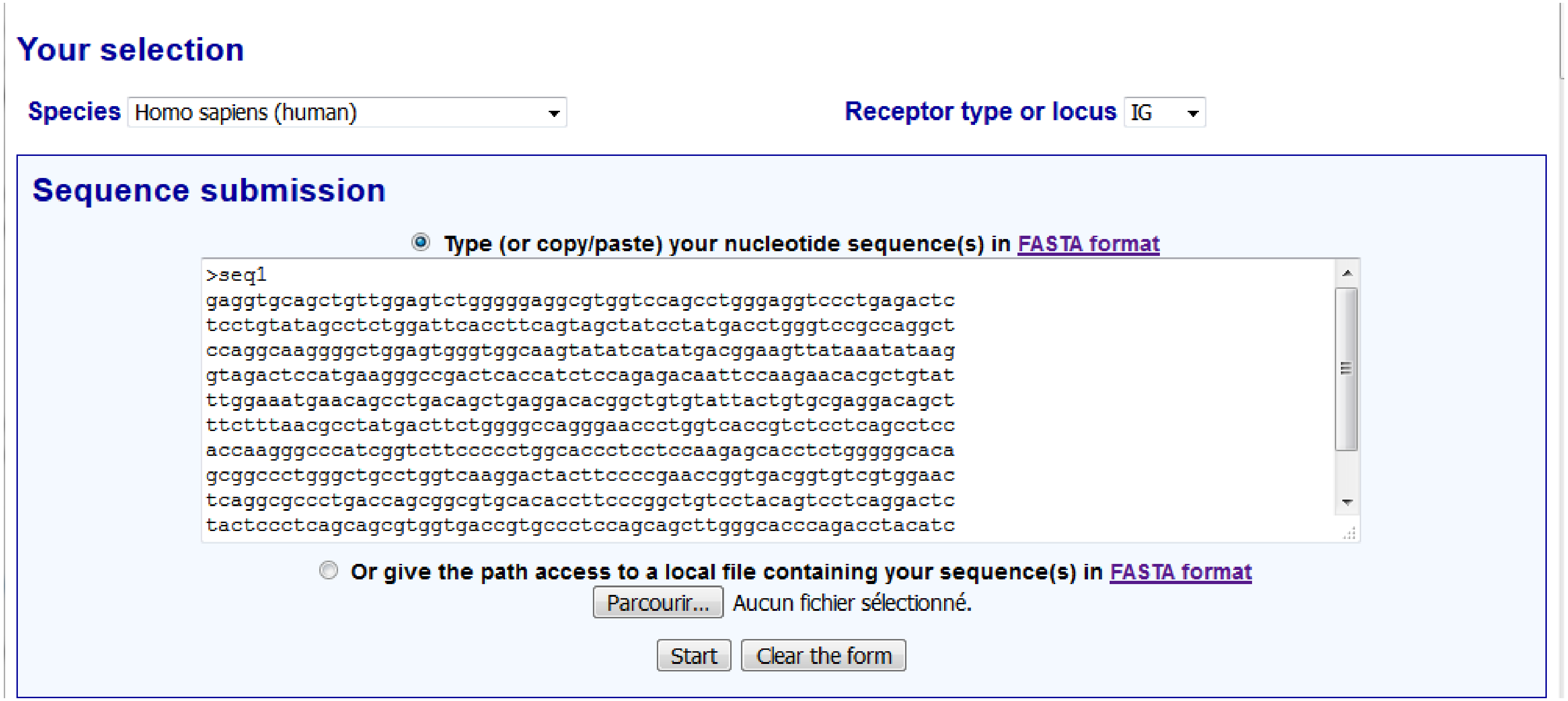
Lefranc et al. Dev. Comp. Immunol. 27, 55-77 (2003)

IMGT/V-QUEST Sequence submission

http://www.imgt.org

For the analysis and interpretation of human IG rearranged sequences

- Online for up to 50 sequences
- Identifies the germline V, D and J genes and alleles,
- Analyses the V-(D)-J junction (IMGT/JunctionAnalysis)
- Characterizes the nucleotide mutations and amino acid changes
- Full annotation of the V-DOMAIN



Your selection

Species Homo sapiens (human)

Receptor type or locus IG

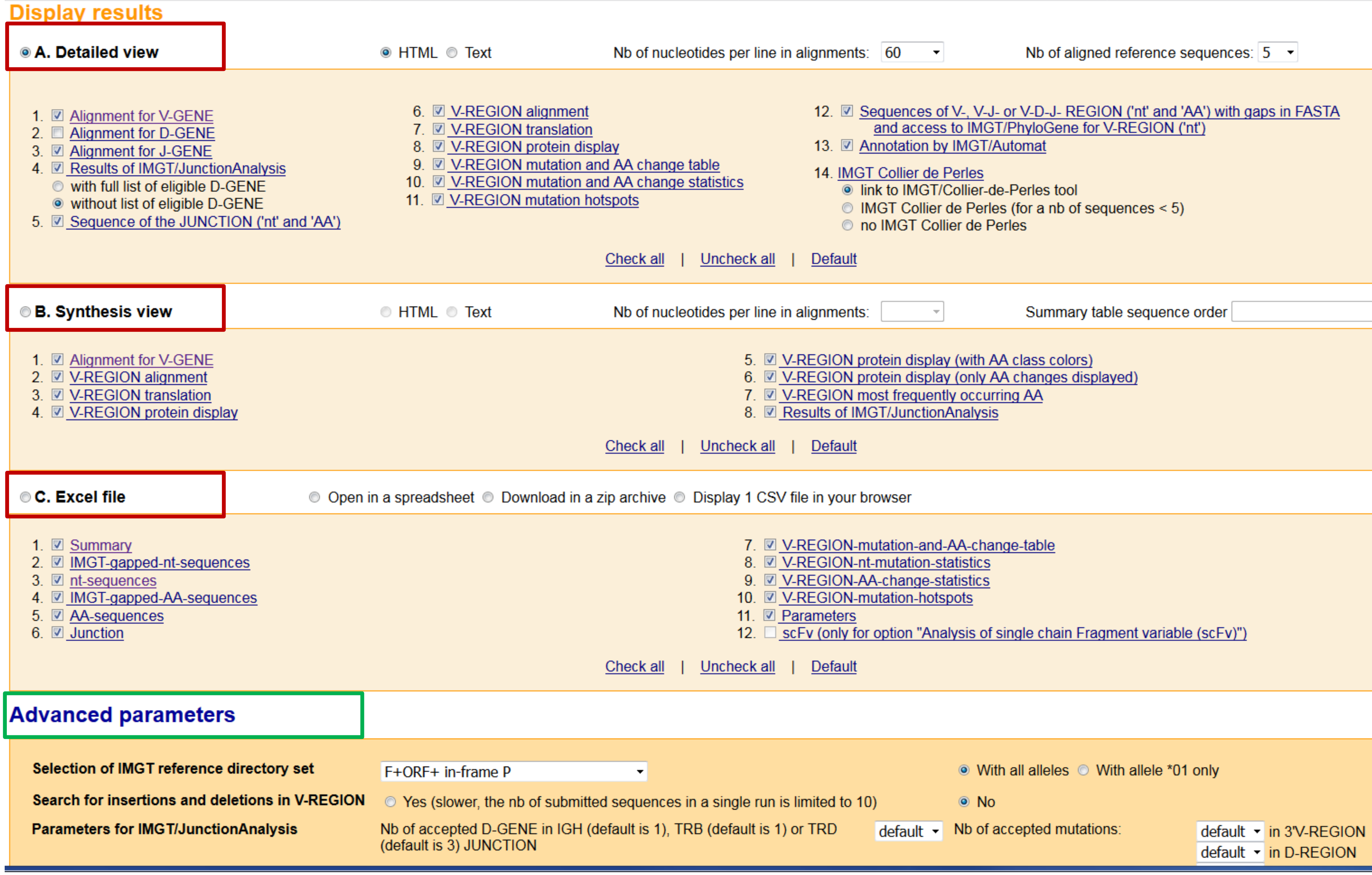
Sequence submission

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

```
>seq1
gaggtgcagctgttggagtctgggggaggcgtggtccagcctgggaggtccctgagactc
tcctgtatagcctctggattcaccttcagtagctatcctatgacctgggtccgccaggct
ccaggcaaggggctggagtgggtggcaagtatatcatatgacggaagttataaatataag
gtagactccatgaagggccgactcaccatctccagagacaattccaagaacacgctgtat
ttggaaatgaacagcctgacagctgaggacacggctgtgtattactgtgcgaggacagct
ttctttaacgcctatgacttctggggccagggaaccctggtcaccgtctcctcagcctcc
accaagggcccatcggtcttcccccctggcaccctcctccaagagcacctctgggggcaca
gcggccctgggctgcctggtcaaggactacttccccgaaccggtgacggtgtcgtggaac
tcaggcgccctgaccagcggcgtgcacaccttcccggctgtcctacagtcctcaggactc
tactccctcagcagcgtggtgaccgtgccctccagcagcttgggcacccagacctacatc
```

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

IMGT/V-QUEST Displays & Advanced parameters



Display results

● **A. Detailed view** — ● HTML ○ Text — Nb of nucleotides per line in alignments: 60 — Nb of aligned reference sequences: 5

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

Advanced parameters

Selection of IMGT reference directory set — F+ORF+ in-frame P — ● With all alleles ○ With allele *01 only

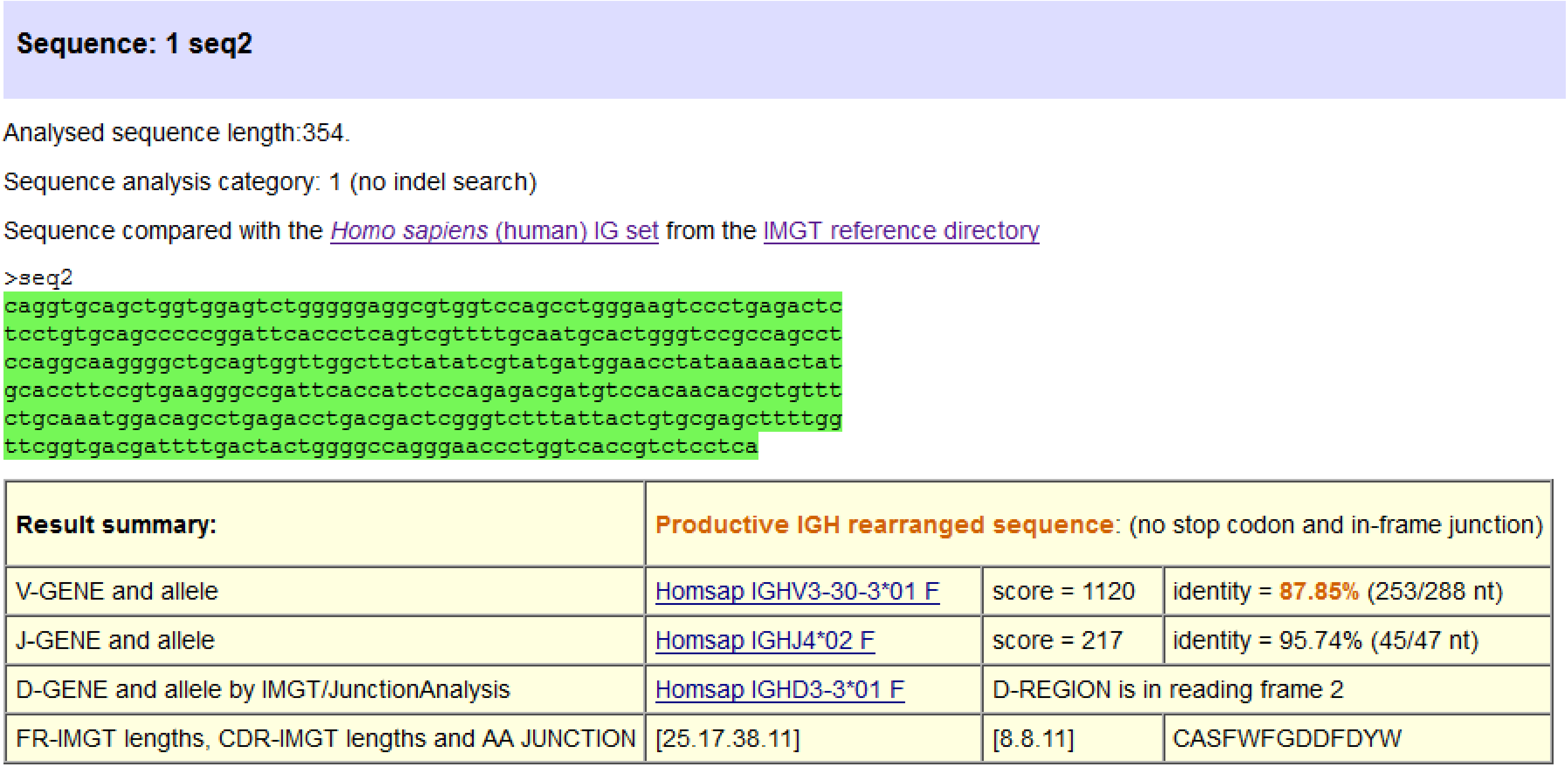
Search for insertions and deletions in V-REGION — ○ Yes (slower, the nb of submitted sequences in a single run is limited to 10) ● No

Parameters for IMGT/JunctionAnalysis — Nb of accepted D-GENE in IGH (default is 1), TRB (default is 1) or TRD (default is 3) JUNCTION: default — Nb of accepted mutations: default in 3'V-REGION; default in D-REGION

1. IMGT/V-QUEST Detailed view: Result summary

IMGT® http://www.imgt.org

- Evaluation of the functionality of rearranged sequences (Productive vs Unproductive)
- Identification of V (D) J genes and alleles, with score, % (/ nt)
- FR-IMGT and CDR-IMGT lengths
- Amino acid (AA) JUNCTION



Sequence: 1 seq2

Analysed sequence length:354.

Sequence analysis category: 1 (no indel search)

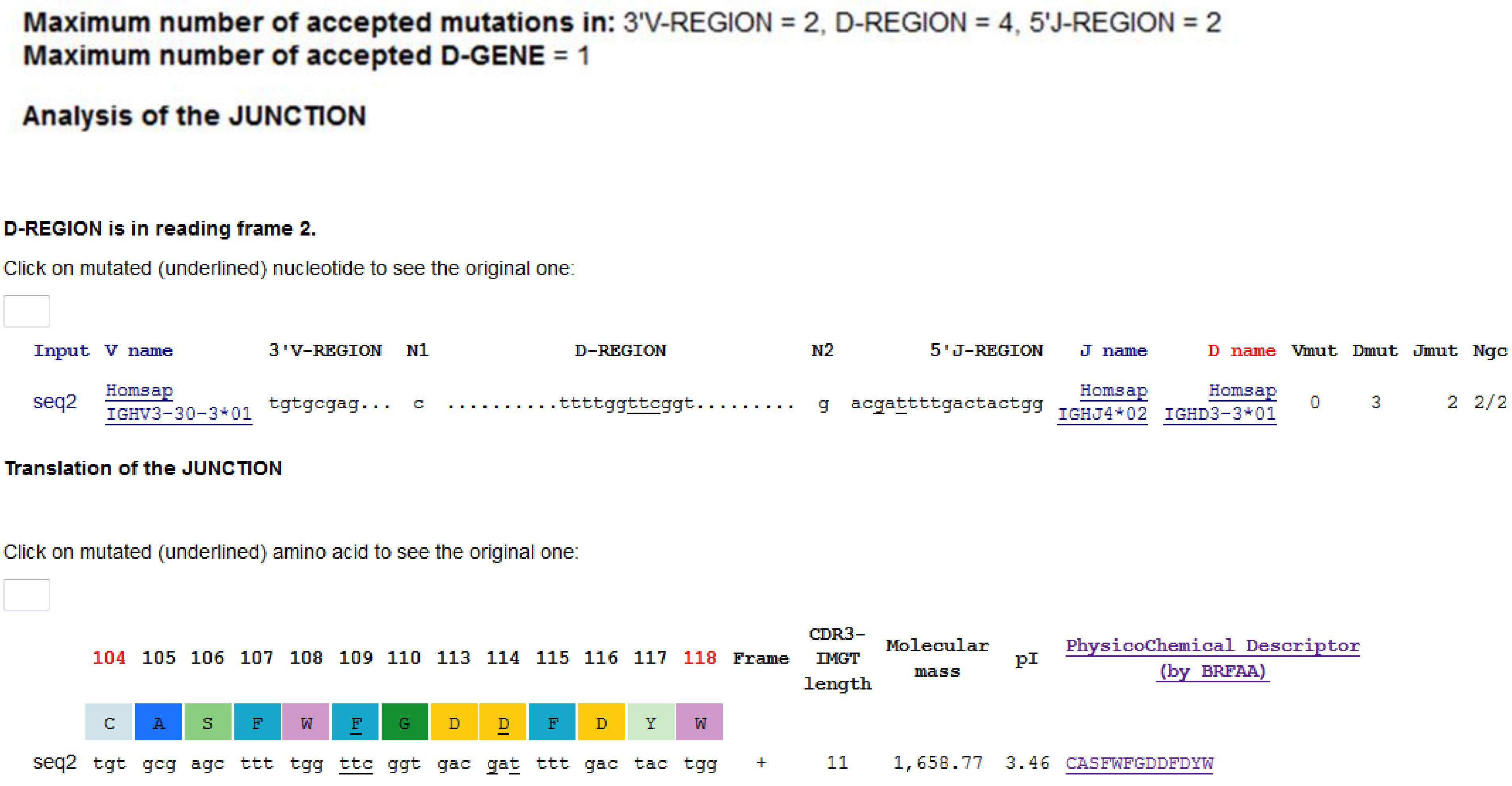
Sequence compared with the *Homo sapiens* (human) IG set from the IMGT reference directory

```
>seq2
caggtgcagctggtggagtctgggggaggcgtggtccagcctgggaagtccctgagactc
tcctgtgcagcccccggattcaccctcagtcgttttgcaatgcactgggtccgccagcct
ccaggcaaggggctgcagtggttggcttctatatcgtatgatggaacctataaaaactat
gcaccttccgtgaagggccgattcaccatctccagagacgatgtccacaacacgctgttt
ctgcaaatggacagcctgagacctgacgactcgggtctttattactgtgcgagcttttgg
ttcggtgacgattttgactactggggccagggaaccctggtcaccgtctcctca
```

Result summary:	**Productive IGH rearranged sequence**: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV3-30-3*01 F	score = 1120	identity = **87.85%** (253/288 nt)
J-GENE and allele	Homsap IGHJ4*02 F	score = 217	identity = 95.74% (45/47 nt)
D-GENE and allele by IMGT/JunctionAnalysis	Homsap IGHD3-3*01 F	D-REGION is in reading frame 2	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.11]	CASFWFGDDFDYW

2. IMGT/V-QUEST Results of IMGT/JunctionAnalysis

IMGT Information system®
http://www.imgt.org



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 2.

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
seq2	Homsap IGHV3-30-3*01	tgtgcgag...	c	..........ttttggttcggt.........	g	acgattttgactactgg	Homsap IGHJ4*02	Homsap IGHD3-3*01	0	3	2	2/2

Translation of the JUNCTION

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

	104	105	106	107	108	109	110	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
	C	A	S	F	W	F	G	D	D	F	D	Y	W					
seq2	tgt	gcg	agc	ttt	tgg	ttc	ggt	gac	gat	ttt	gac	tac	tgg	+	11	1,658.77	3.46	CASFWFGDDFDYW

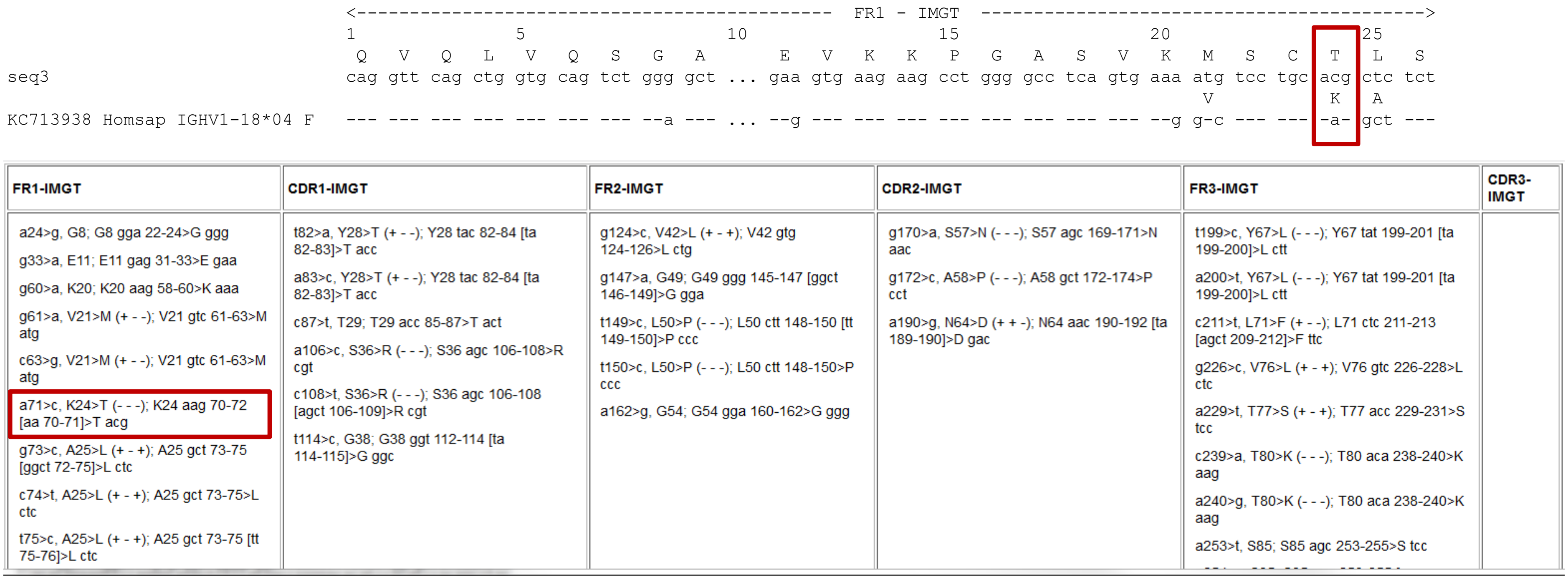
Yousfi Monod et al. Bioinformatics, 20, i379-i385 (2004)

3. IMGT/V-QUEST description of V-REGION mutations, AA changes & hotspots

http://www.imgt.org

IMGT/V-QUEST provides description of:
- nucleotide (nt) mutations
- amino acid (AA) changes

with hotspots localization in corresponding germline per FR-IMGT and CDR-IMGT of the V-REGION



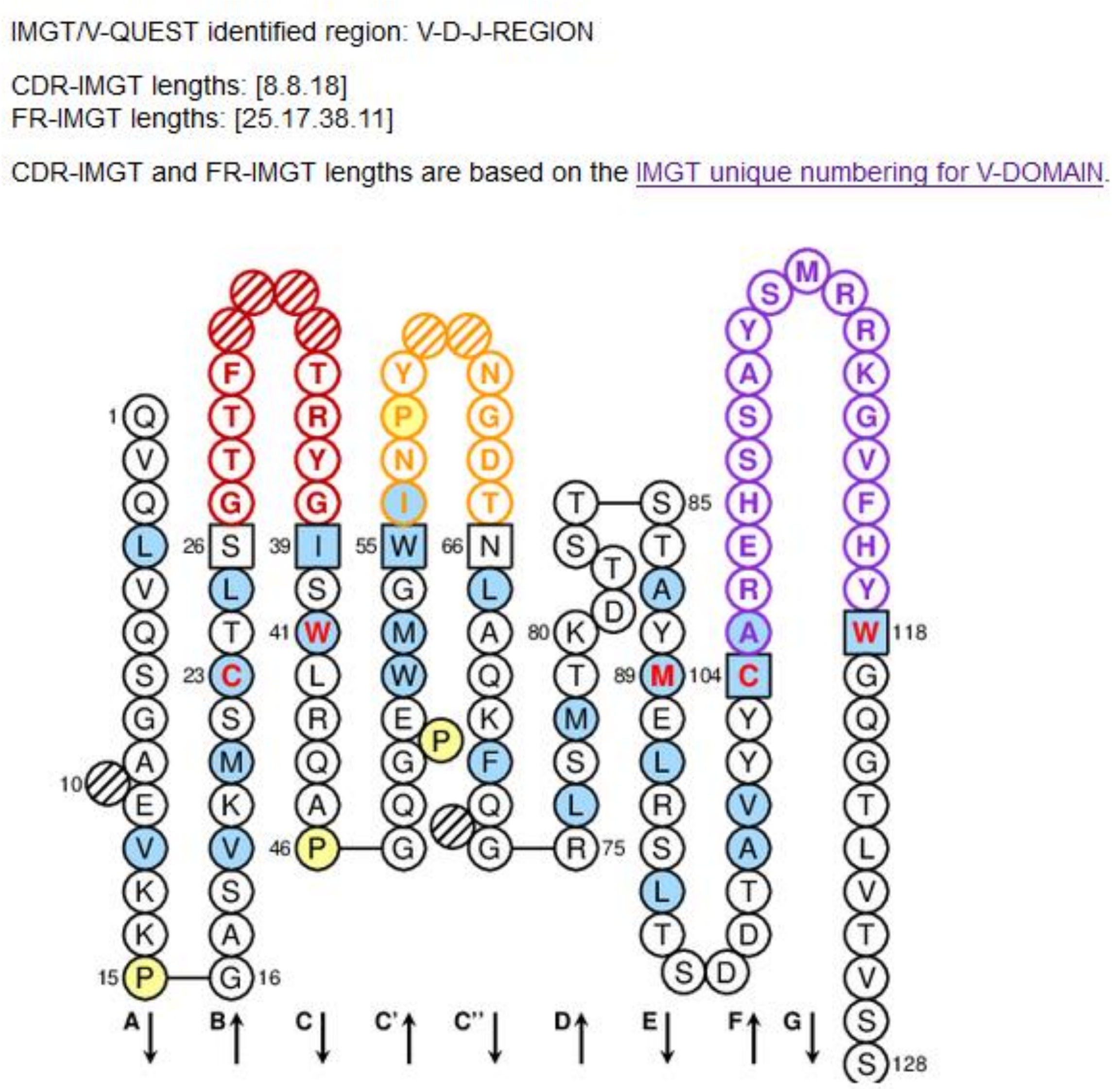
```
                                 <----------------------------------------- FR1 - IMGT ------------------------------------->
                                 1               5                   10                  15                  20                  25
                                  Q   V   Q   L   V   Q   S   G   A       E   V   K   K   P   G   A   S   V   K   M   S   C   T   L   S
seq3                             cag gtt cag ctg gtg cag tct ggg gct ... gaa gtg aag aag cct ggg gcc tca gtg aaa atg tcc tgc acg ctc tct
                                                                                                                  V           K   A
KC713938 Homsap IGHV1-18*04 F    --- --- --- --- --- --- --- --a --- ... --g --- --- --- --- --- --- --- --- --g g-c --- --- -a- gct ---
```

FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
a24>g, G8; G8 gga 22-24>G ggg g33>a, E11; E11 gag 31-33>E gaa g60>a, K20; K20 aag 58-60>K aaa g61>a, V21>M (+ - -); V21 gtc 61-63>M atg c63>g, V21>M (+ - -); V21 gtc 61-63>M atg a71>c, K24>T (- - -); K24 aag 70-72 [aa 70-71]>T acg g73>c, A25>L (+ - +); A25 gct 73-75 [ggct 72-75]>L ctc c74>t, A25>L (+ - +); A25 gct 73-75>L ctc t75>c, A25>L (+ - +); A25 gct 73-75 [tt 75-76]>L ctc	t82>a, Y28>T (+ - -); Y28 tac 82-84 [ta 82-83]>T acc a83>c, Y28>T (+ - -); Y28 tac 82-84 [ta 82-83]>T acc c87>t, T29; T29 acc 85-87>T act a106>c, S36>R (- - -); S36 agc 106-108>R cgt c108>t, S36>R (- - -); S36 agc 106-108 [agct 106-109]>R cgt t114>c, G38; G38 ggt 112-114 [ta 114-115]>G ggc	g124>c, V42>L (+ - +); V42 gtg 124-126>L ctg g147>a, G49; G49 ggg 145-147 [ggct 146-149]>G gga t149>c, L50>P (- - -); L50 ctt 148-150 [tt 149-150]>P ccc t150>c, L50>P (- - -); L50 ctt 148-150>P ccc a162>g, G54; G54 gga 160-162>G ggg	g170>a, S57>N (- - -); S57 agc 169-171>N aac g172>c, A58>P (- - -); A58 gct 172-174>P cct a190>g, N64>D (+ + -); N64 aac 190-192 [ta 189-190]>D gac	t199>c, Y67>L (- - -); Y67 tat 199-201 [ta 199-200]>L ctt a200>t, Y67>L (- - -); Y67 tat 199-201 [ta 199-200]>L ctt c211>t, L71>F (+ - -); L71 ctc 211-213 [agct 209-212]>F ttc g226>c, V76>L (+ - +); V76 gtc 226-228>L ctc a229>t, T77>S (+ - +); T77 acc 229-231>S tcc c239>a, T80>K (- - -); T80 aca 238-240>K aag a240>g, T80>K (- - -); T80 aca 238-240>K aag a253>t, S85; S85 agc 253-255>S tcc	

a71>c, K24>T (- - -); K24 aag 70-72 [aa 70-71]>T acg

3. IMGT Collier de Perles

http://www.imgt.org



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

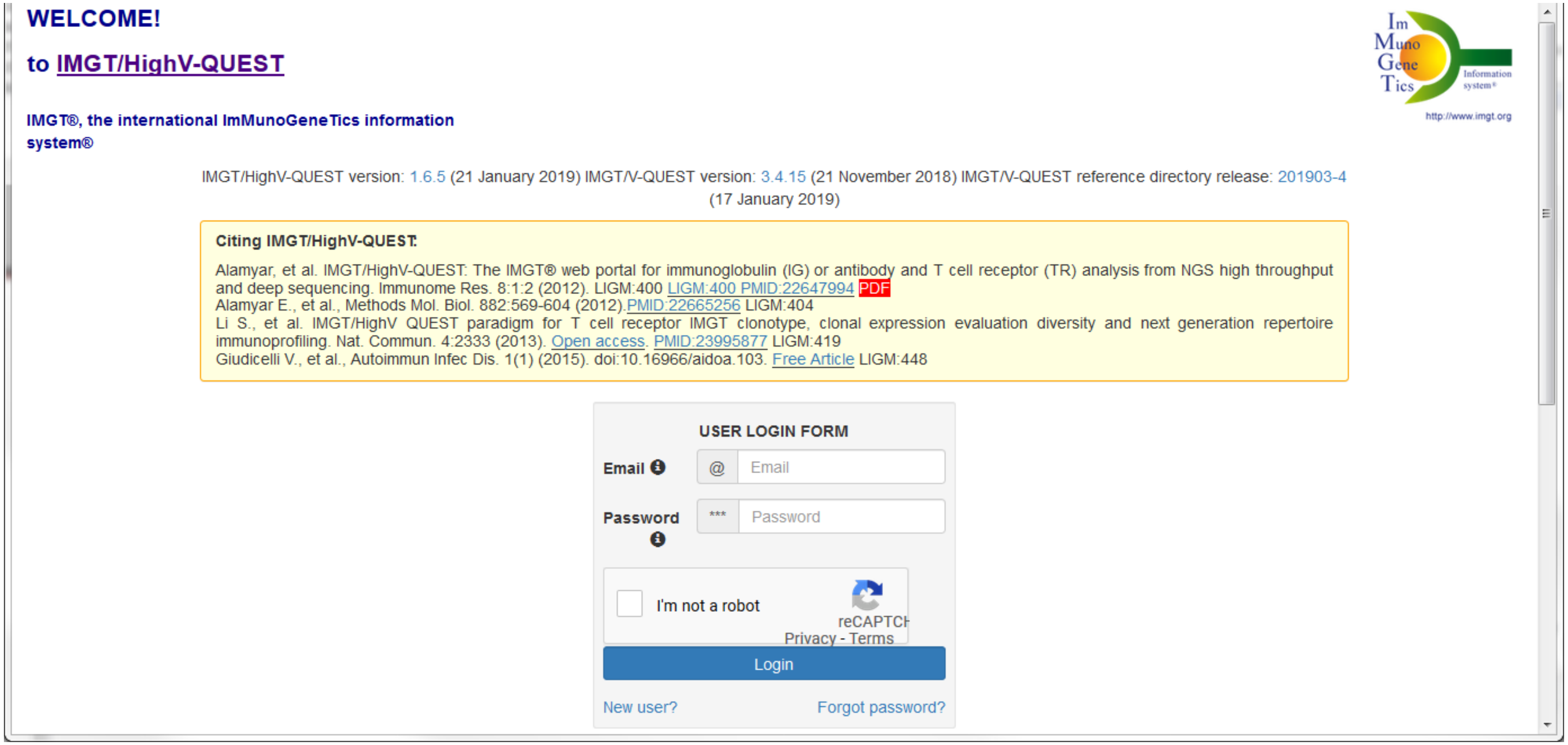
CDR-IMGT lengths: [8.8.18]
FR-IMGT lengths: [25.17.38.11]

CDR-IMGT and FR-IMGT lengths are based on the IMGT unique numbering for V-DOMAIN.

IMGT/HighV-QUEST

http://www.imgt.org

- High throughput version of IMGT/V-QUEST on the Web since 2010
- Analyses the IG and TR rearranged sequences from NGS
- Analyses up to 500,000 sequences per batch (results in 11-12 CSV files)
- Compares clonotypes for up 1,000,000 sequences
- Freely available for academics (user registration required)
- “Search for insertions and deletions in V-REGION” activated by default



WELCOME!

to IMGT/HighV-QUEST

IMGT®, the international ImMunoGeneTics information system®

http://www.imgt.org

IMGT/HighV-QUEST version: 1.6.5 (21 January 2019) IMGT/V-QUEST version: 3.4.15 (21 November 2018) IMGT/V-QUEST reference directory release: 201903-4 (17 January 2019)

Citing IMGT/HighV-QUEST:

Alamyar, et al. IMGT/HighV-QUEST: The IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Res. 8:1:2 (2012). LIGM:400 LIGM:400 PMID:22647994 PDF
Alamyar E., et al., Methods Mol. Biol. 882:569-604 (2012).PMID:22665256 LIGM:404
Li S., et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype, clonal expression evaluation diversity and next generation repertoire immunoprofiling. Nat. Commun. 4:2333 (2013). Open access. PMID:23995877 LIGM:419
Giudicelli V., et al., Autoimmun Infec Dis. 1(1) (2015). doi:10.16966/aidoa.103. Free Article LIGM:448

USER LOGIN FORM

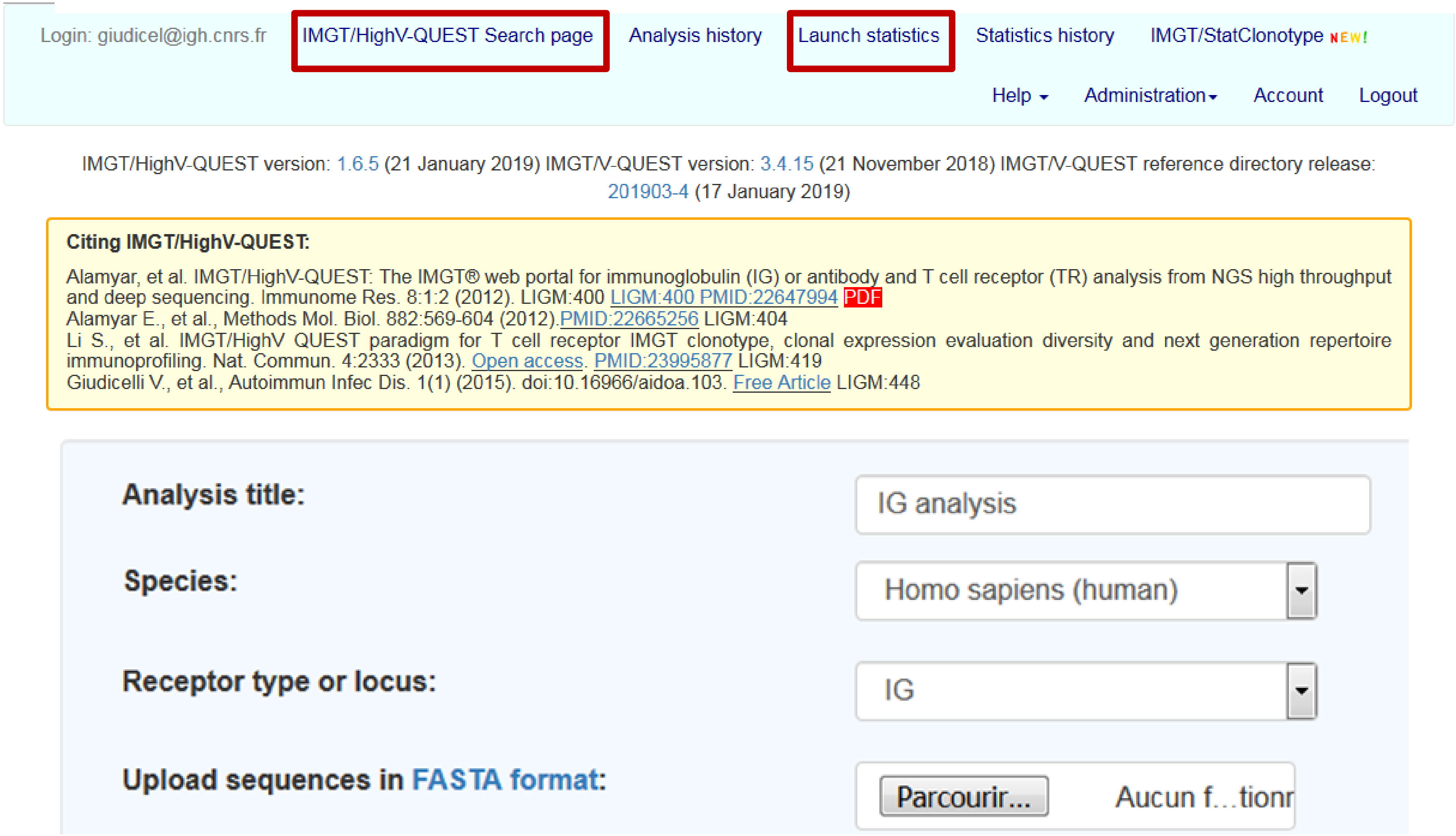
Email	@ Email
Password	*** Password

☐ I'm not a robot

Alamyar E. et al. Immunome Res. 8:26 (2012)

IMGT/HighV-QUEST submission

http://www.imgt.org



IMGT/HighV-QUEST version: 1.6.5 (21 January 2019) IMGT/V-QUEST version: 3.4.15 (21 November 2018) IMGT/V-QUEST reference directory release: 201903-4 (17 January 2019)

Citing IMGT/HighV-QUEST:

Alamyar, et al. IMGT/HighV-QUEST: The IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Res. 8:1:2 (2012). LIGM:400 LIGM:400 PMID:22647994 PDF
Alamyar E., et al., Methods Mol. Biol. 882:569-604 (2012).PMID:22665256 LIGM:404
Li S., et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype, clonal expression evaluation diversity and next generation repertoire immunoprofiling. Nat. Commun. 4:2333 (2013). Open access. PMID:23995877 LIGM:419
Giudicelli V., et al., Autoimmun Infec Dis. 1(1) (2015). doi:10.16966/aidoa.103. Free Article LIGM:448

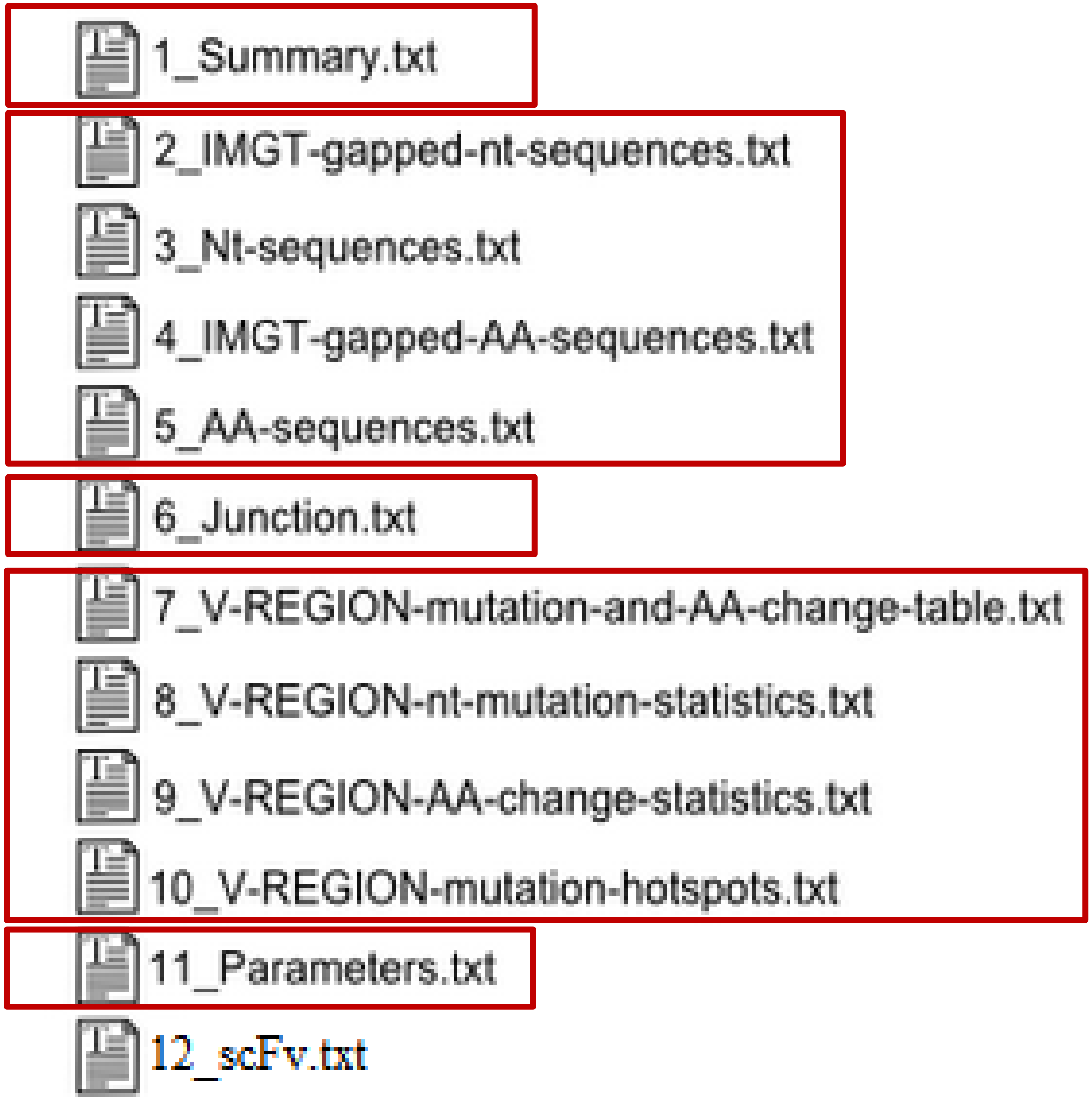
Analysis title: IG analysis

Species: Homo sapiens (human)

Receptor type or locus: IG

Upload sequences in FASTA format:

IMGT/HighV-QUEST results in 11 CSV files



Result summary

1_Summary.txt

Sequences of the
V-DOMAIN in nucleotides (nt)
and in amino acids (AA),
with and without IMGT gaps

2_IMGT-gapped-nt-sequences.txt
3_Nt-sequences.txt
4_IMGT-gapped-AA-sequences.txt
5_AA-sequences.txt

Analysis of the JUNCTION

6_Junction.txt

Description of V-REGION
nt mutations, AA changes,
mutation hotspots

7_V-REGION-mutation-and-AA-change-table.txt
8_V-REGION-nt-mutation-statistics.txt
9_V-REGION-AA-change-statistics.txt
10_V-REGION-mutation-hotspots.txt

IMGT/V-QUEST version and
data release, selected parameters

11_Parameters.txt

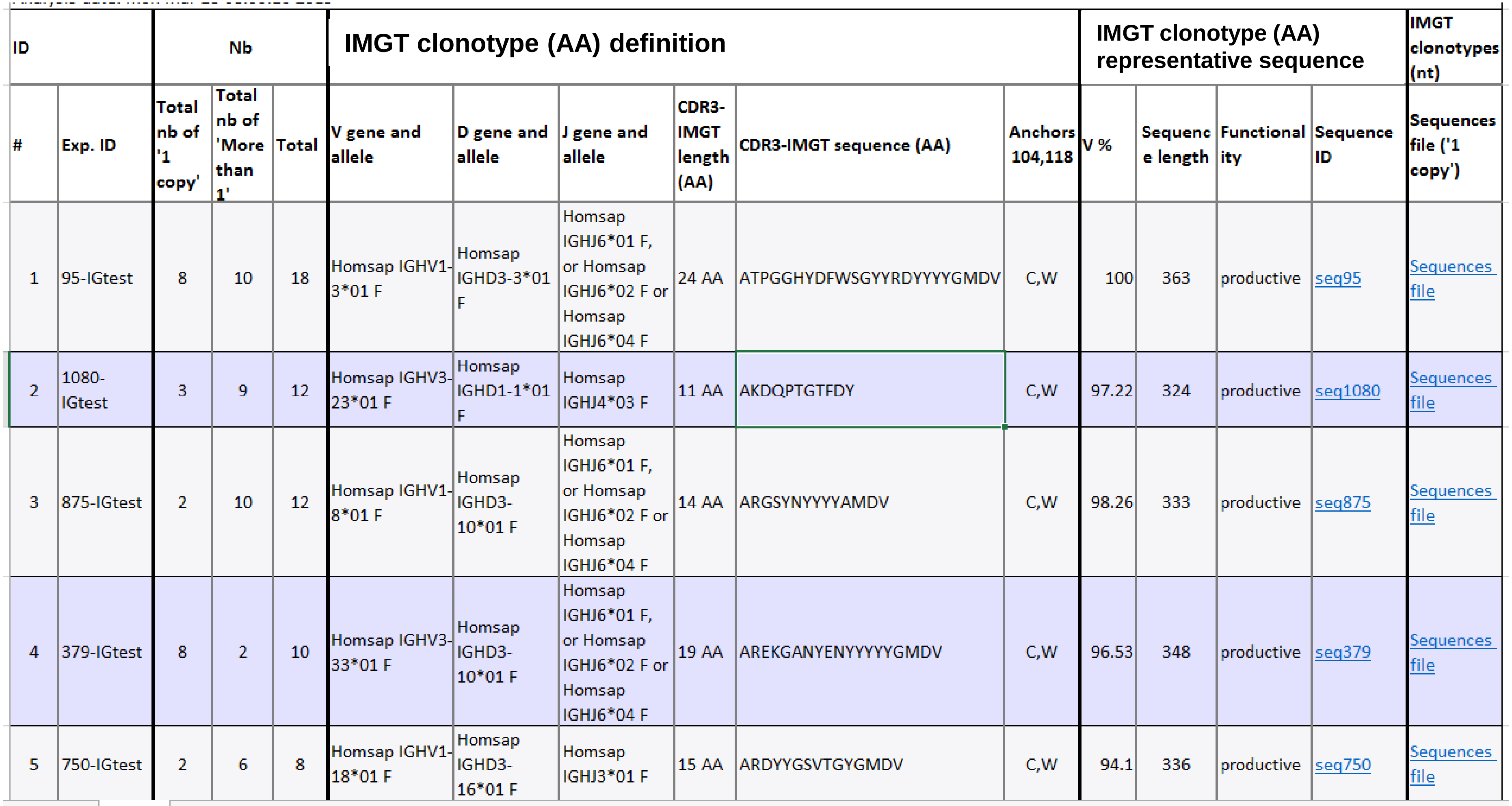
12_scFv.txt

IMGT/HighV-QUEST results of IMGT clonotypes (AA)

http://www.imgt.org

An IMGT clonotype (AA) is defined by:

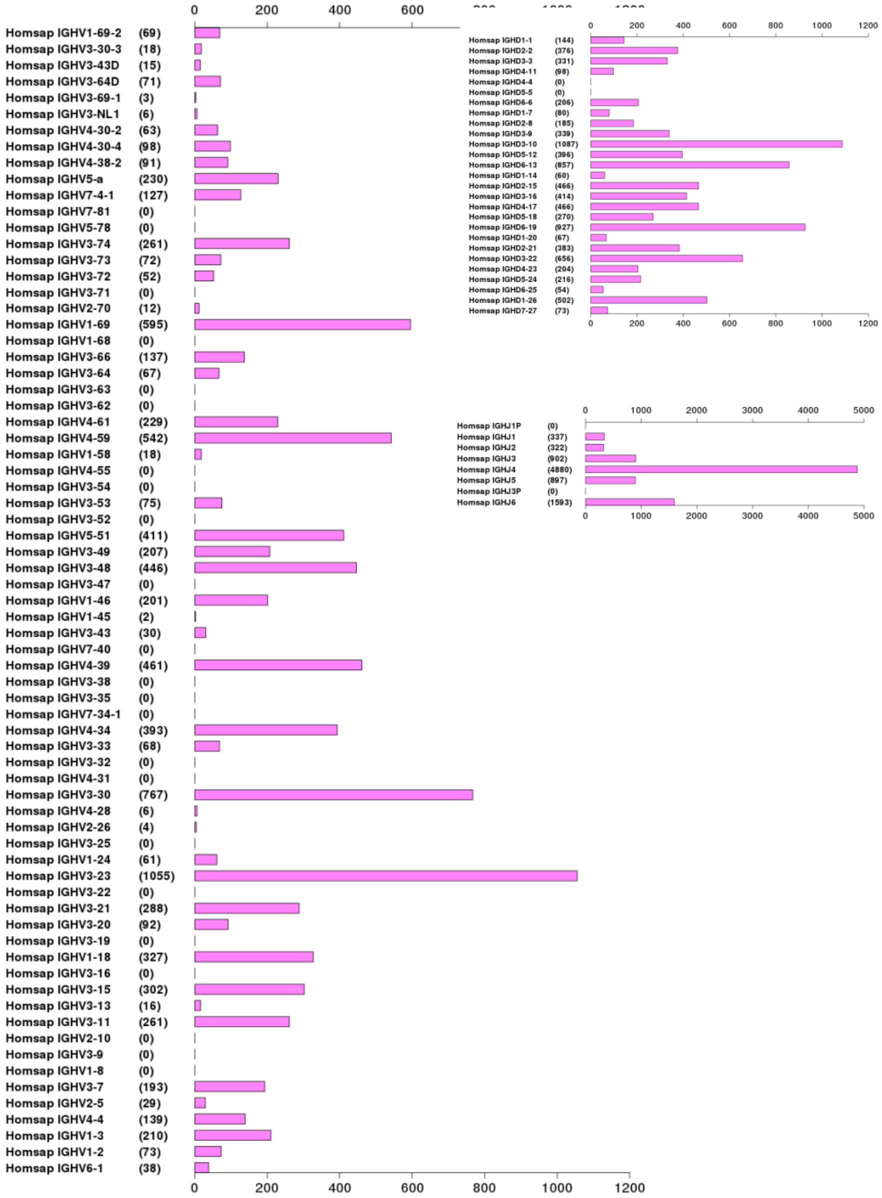
- a unique V-(D)-J-rearrangement (V and J genes and alleles) (nt)
- a unique CDR3 (AA)
- Conserved anchors C104 W/F118



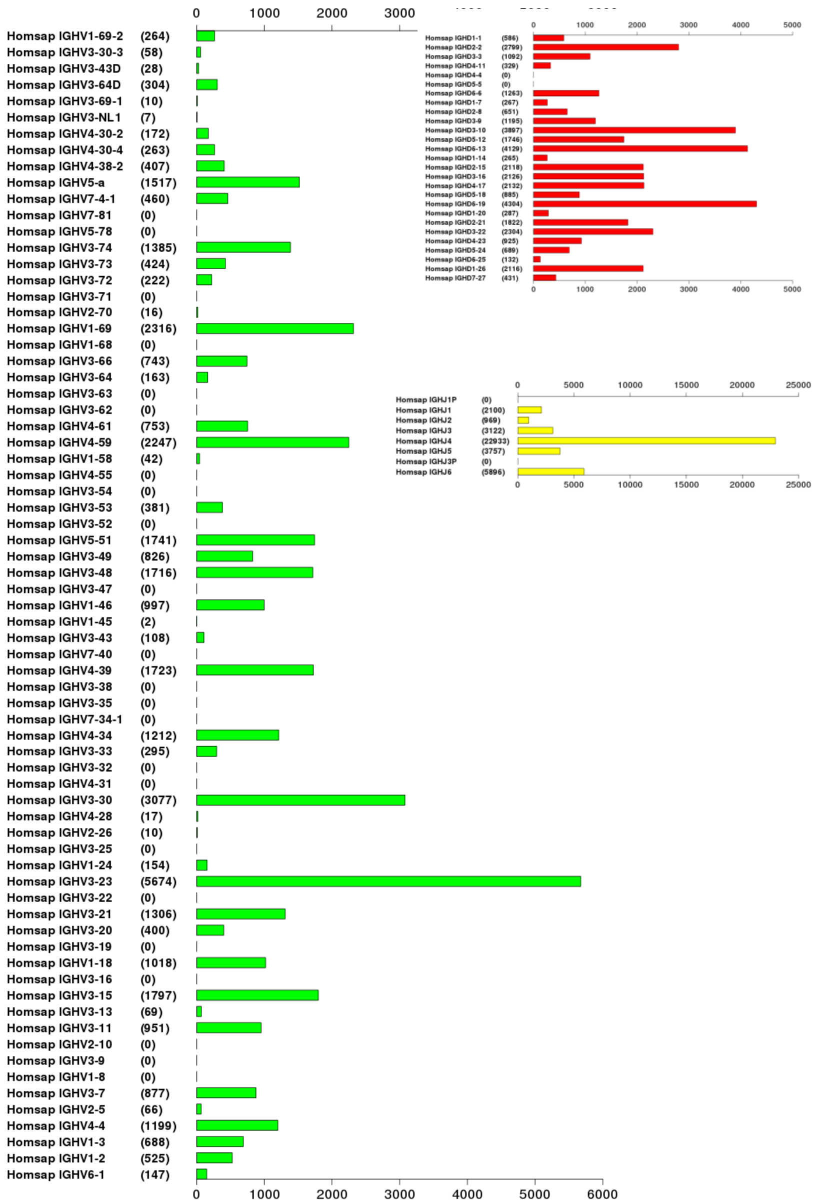
ID		Nb			IMGT clonotype (AA) definition						IMGT clonotype (AA) representative sequence				IMGT clonotypes (nt)
#	Exp. ID	Total nb of '1 copy'	Total nb of 'More than 1'	Total	V gene and allele	D gene and allele	J gene and allele	CDR3-IMGT length (AA)	CDR3-IMGT sequence (AA)	Anchors 104,118	V %	Sequence length	Functionality	Sequence ID	Sequences file ('1 copy')
1	95-IGtest	8	10	18	Homsap IGHV1-3*01 F	Homsap IGHD3-3*01 F	Homsap IGHJ6*01 F, or Homsap IGHJ6*02 F or Homsap IGHJ6*04 F	24 AA	ATPGGHYDFWSGYYRDYYYYGMDV	C,W	100	363	productive	seq95	Sequences file
2	1080-IGtest	3	9	12	Homsap IGHV3-23*01 F	Homsap IGHD1-1*01 F	Homsap IGHJ4*03 F	11 AA	AKDQPTGTFDY	C,W	97.22	324	productive	seq1080	Sequences file
3	875-IGtest	2	10	12	Homsap IGHV1-8*01 F	Homsap IGHD3-10*01 F	Homsap IGHJ6*01 F, or Homsap IGHJ6*02 F or Homsap IGHJ6*04 F	14 AA	ARGSYNYYYYAMDV	C,W	98.26	333	productive	seq875	Sequences file
4	379-IGtest	8	2	10	Homsap IGHV3-33*01 F	Homsap IGHD3-10*01 F	Homsap IGHJ6*01 F, or Homsap IGHJ6*02 F or Homsap IGHJ6*04 F	19 AA	AREKGANYENYYYYYGMDV	C,W	96.53	348	productive	seq379	Sequences file
5	750-IGtest	2	6	8	Homsap IGHV1-18*01 F	Homsap IGHD3-16*01 F	Homsap IGHJ3*01 F	15 AA	ARDYYGSVTGYGMDV	C,W	94.1	336	productive	seq750	Sequences file

Evaluation of the clonotype diversity and expression per gene

IMGT clonotype (AA) diversity



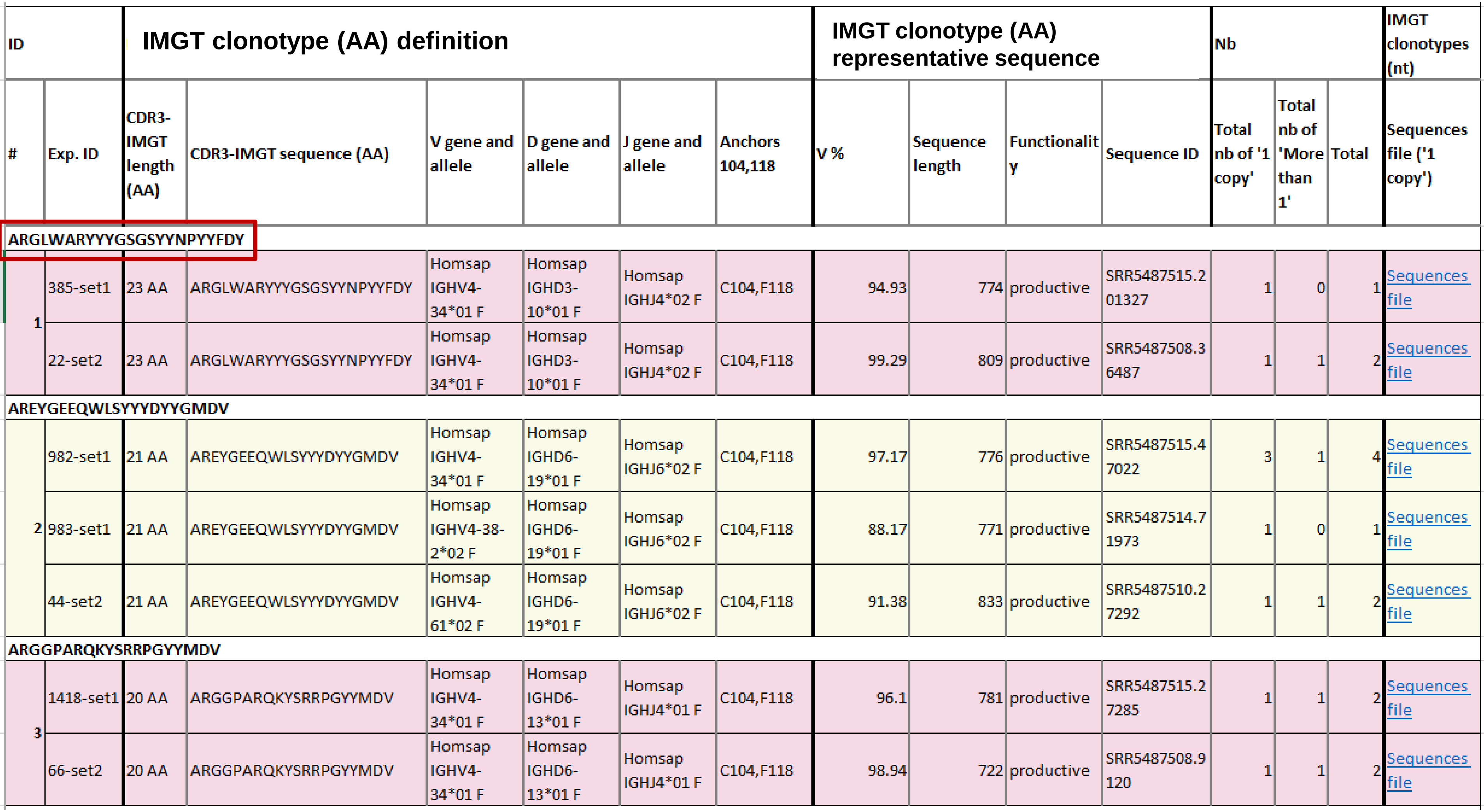
IMGT clonotype (AA) expression



Comparison several different sequences batches

http://www.imgt.org

identification of IMGT clonotypes (AA) with same CDR3-IMGT (AA)



ID		IMGT clonotype (AA) definition						IMGT clonotype (AA) representative sequence				Nb			IMGT clonotypes (nt)
#	Exp. ID	CDR3-IMGT length (AA)	CDR3-IMGT sequence (AA)	V gene and allele	D gene and allele	J gene and allele	Anchors 104,118	V %	Sequence length	Functionality	Sequence ID	Total nb of '1 copy'	Total nb of 'More than 1'	Total	Sequences file ('1 copy')
ARGLWARYYYGSGSYYNPYYFDY															
1	385-set1	23 AA	ARGLWARYYYGSGSYYNPYYFDY	Homsap IGHV4-34*01 F	Homsap IGHD3-10*01 F	Homsap IGHJ4*02 F	C104,F118	94.93	774	productive	SRR5487515.201327	1	0	1	Sequences file
	22-set2	23 AA	ARGLWARYYYGSGSYYNPYYFDY	Homsap IGHV4-34*01 F	Homsap IGHD3-10*01 F	Homsap IGHJ4*02 F	C104,F118	99.29	809	productive	SRR5487508.36487	1	1	2	Sequences file
AREYGEEQWLSYYYDYYGMDV															
2	982-set1	21 AA	AREYGEEQWLSYYYDYYGMDV	Homsap IGHV4-34*01 F	Homsap IGHD6-19*01 F	Homsap IGHJ6*02 F	C104,F118	97.17	776	productive	SRR5487515.47022	3	1	4	Sequences file
	983-set1	21 AA	AREYGEEQWLSYYYDYYGMDV	Homsap IGHV4-38-2*02 F	Homsap IGHD6-19*01 F	Homsap IGHJ6*02 F	C104,F118	88.17	771	productive	SRR5487514.71973	1	0	1	Sequences file
	44-set2	21 AA	AREYGEEQWLSYYYDYYGMDV	Homsap IGHV4-61*02 F	Homsap IGHD6-19*01 F	Homsap IGHJ6*02 F	C104,F118	91.38	833	productive	SRR5487510.27292	1	1	2	Sequences file
ARGGPARQKYSRRPGYYMDV															
3	1418-set1	20 AA	ARGGPARQKYSRRPGYYMDV	Homsap IGHV4-34*01 F	Homsap IGHD6-13*01 F	Homsap IGHJ4*01 F	C104,F118	96.1	781	productive	SRR5487515.27285	1	1	2	Sequences file
	66-set2	20 AA	ARGGPARQKYSRRPGYYMDV	Homsap IGHV4-34*01 F	Homsap IGHD6-13*01 F	Homsap IGHJ4*01 F	C104,F118	98.94	722	productive	SRR5487508.9120	1	1	2	Sequences file

IMGT/V-QUEST and IMGT/HighV-QUEST

- **For extensive studies of the IG and TR repertoires:**

- V, D, J gene and allele usage
- CDR3-IMGT analysis (length distribution, AA composition)
- Mutation characterization
- Clonal diversity and clonal expression

⇒ repertoire immunoprofiling for comparison of B or T cell populations in normal and pathological situations

- **IMGT standards for sequence interpretation and comparison**

IMGT contributes to :

- the EuroClonality-NGS Consortium (21 diagnostic research groups)
- The Adaptive Immune Receptor Repertoire (AIRR) Community of The Antibody Society.

Acknowledgements

Information system®
http://www.imgt.org

IMGT/HighV-QUEST was granted access to the HPC resources of Centre Informatique National de l'Enseignement Supérieur (CINES) 2019-A0050306029 GENCI (Grand Equipement National de Calcul Intensif).

MAbImprove
LabEx Tours Montpellier

IBiSA
INFRASTRUCTURES BIOLOGIE SANTÉ ET AGRONOMIE

UNIVERSITÉ MONTPELLIER

Agence Nationale de la Recherche
ANR

la Région Languedoc Roussillon

IGH

cnrs

and the companies that support the IMGT efforts of standardization.