Immunoinformatique

Marie-Paule Lefranc

Laboratoire d'ImmunoGénétique Moléculaire
Université Montpellier 2, UPR CNRS 1142, IGH
Institut Universitaire de France

FMBS312, Master recherche 2 Bio-Med, TC3,
Universités Montpellier 1 et 2, ENSCM
30 septembre 2010

IMGT®, the international ImMunoGeneTics information system®

http://www.imgt.org

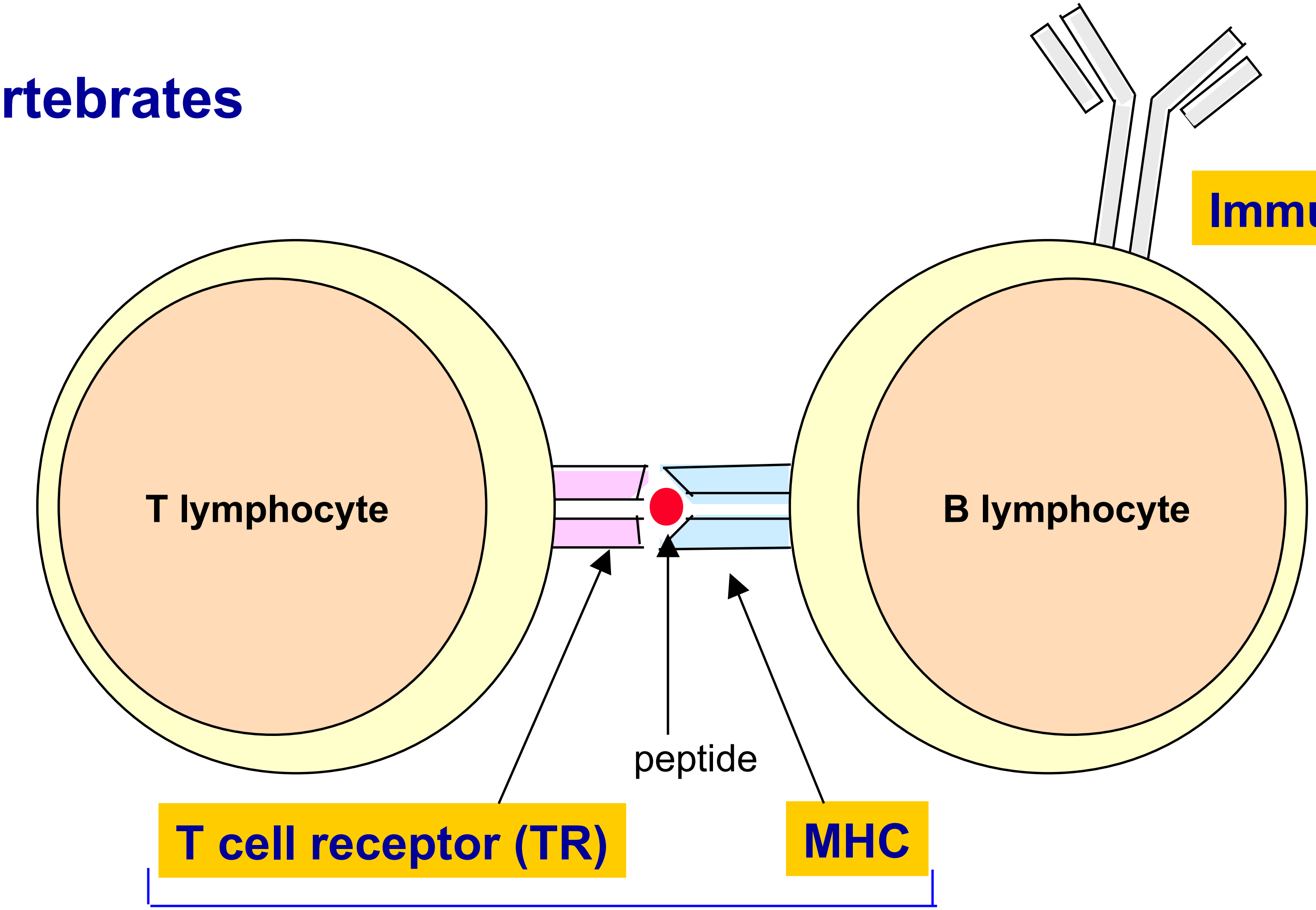
Im
Muno
Gene
Tics

Information
system®

IMGT® domain: the adaptive immune response

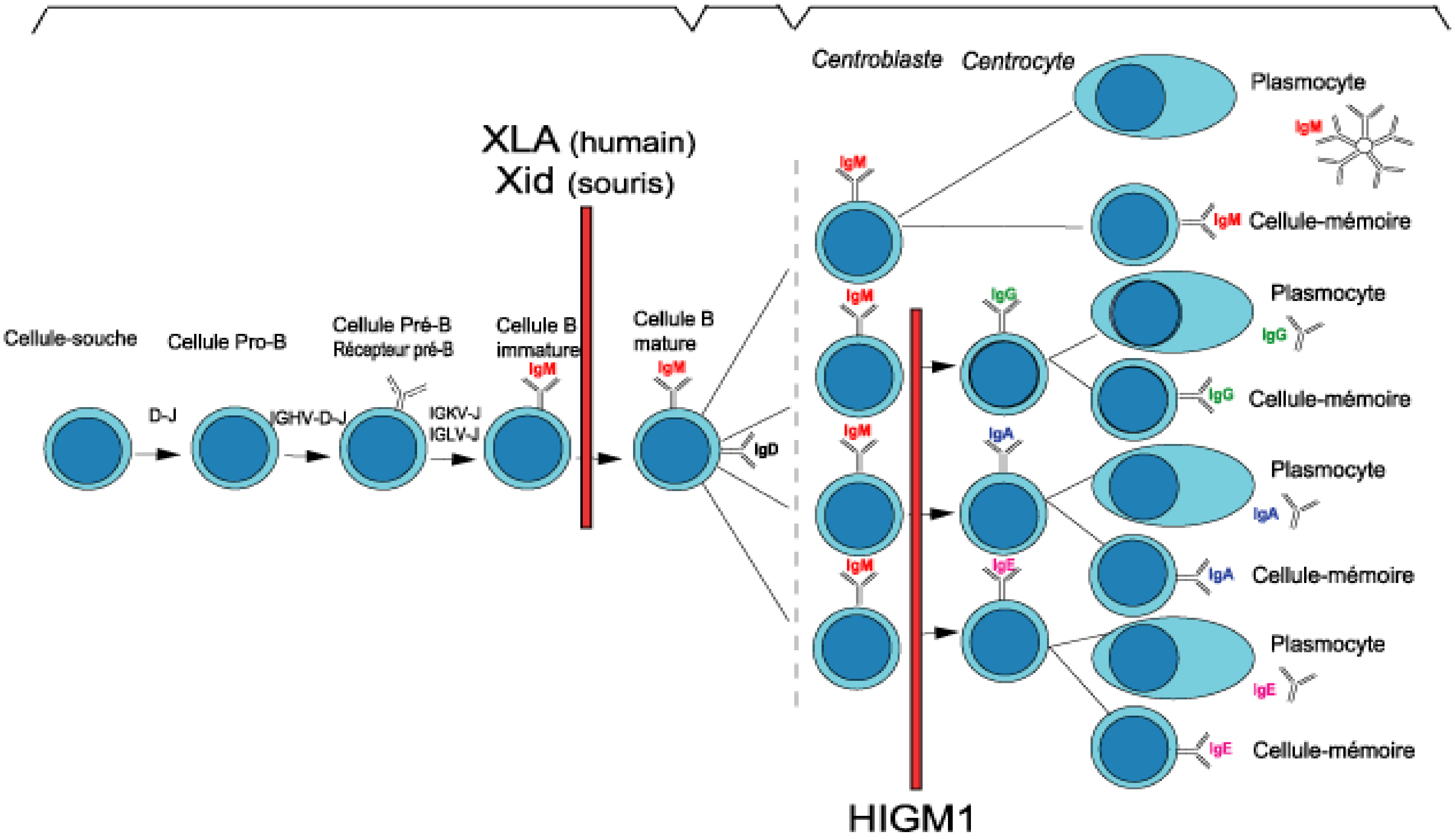
http://www.imgt.org

Vertebrates



Bone marrow Blood Lymph nodes, spleen

http://www.imgt.org



V-D-J and V-J rearrangements

Hypermutations, selection

When and why was IMGT created?

IMGT birth and objectives

IMGT was created in June 1989, following the Human Genome Meeting HGM, at New Haven, USA.

- ***Thomas Shows and Ken Kidd, meeting organizers***
- ***Lap Chee Tsui, chromosome 7 responsable***
- ***Claude Boucheix (CD nomenclature)***

1. At New haven, entry of the 16 human T cell receptor gamma genes (TRG) in the genome database, Genome Database (GDB).

2. IMGT was created to answer the need to manage the T cell receptor (TR) and immunoglobulin (IG) genes and to enter them in the generalist databases.

Why was IMGT necessary?

1. Unique structure of the immunoglobulin (IG) and T cell receptor (TR) genes.

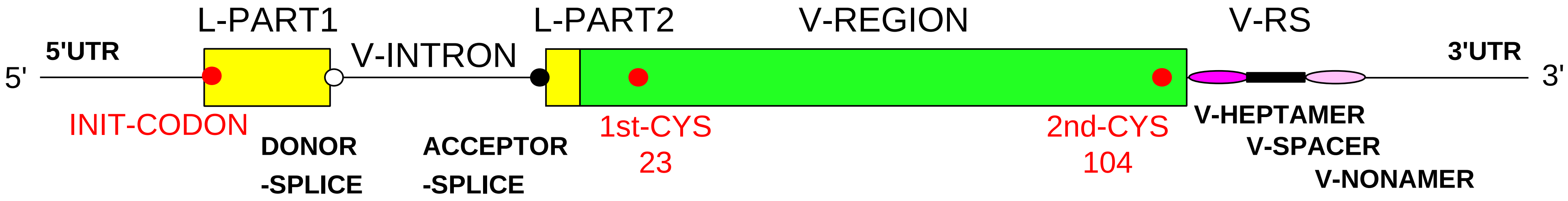
Genomic DNA in germline configuration

http://www.imgt.org

V-GENE

>X62106.0|HSVI2|*Homo sapiens* VI-2 gene for immunoglobulin heavy chain

```
tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag    60
ccacaggtaa gaggctccct agtcccagtg atgagaaaga gattgagtcc agtccaggga   120
gatctcatcc acttctgtgt tctctccaca ggagcccact cccaggtgca gctggtgcag   180
tctggggctg aggtgaagaa gcctggggcc tcagtgaagg tctcctgcaa ggcttctgga   240
tacaccttca ccggctacta tatgcactgg gtgcgacagg cccctggaca agggcttgag   300
tggatgggat ggatcaaccc taacagtggt ggcacaaact atgcacagaa gtttcagggc   360
agggtcacca tgaccaggga cacgtccatc agcacagcct acatggagct gagcaggctg   420
agatctgacg acacggccgt gtattactgt gcgagagaca cagtgtgaaa acccacatcc   480
tgagggtgtc agaaacccaa gggaggaggc ag
```

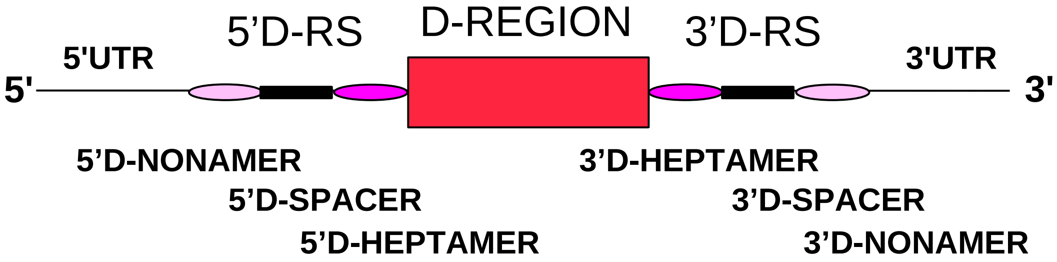


Genomic DNA in germline configuration

http://www.imgt.org

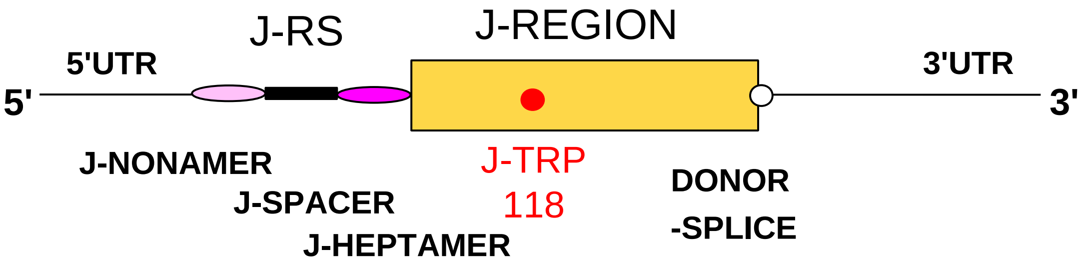
D-GENE

ggtttttgcc tgagctgaga accactgtgc taactgggga cacagtgatt
ggcagctcta caaaaacc



J-GENE

ggtttct gtgcccctgg ctcagggctg actcaccgtg gctgaatact
tccagcactg gggccagggc accctggtca ccgtctcctc ag

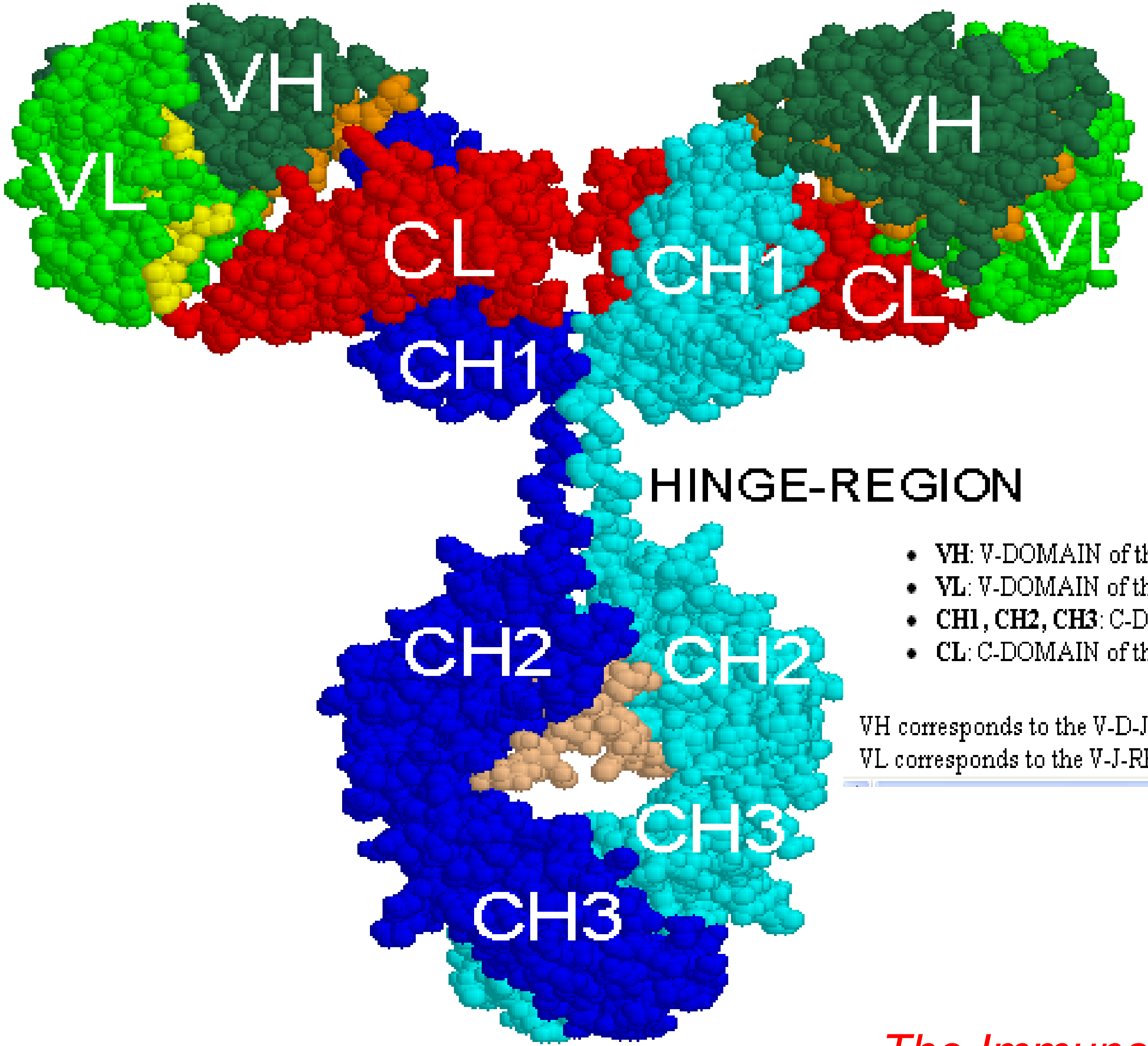


Why was IMGT necessary?

2. Huge diversity of the immunoglobulin and T cell receptor repertoires.

Spacefill 3D representation of an IgG

http://www.imgt.org



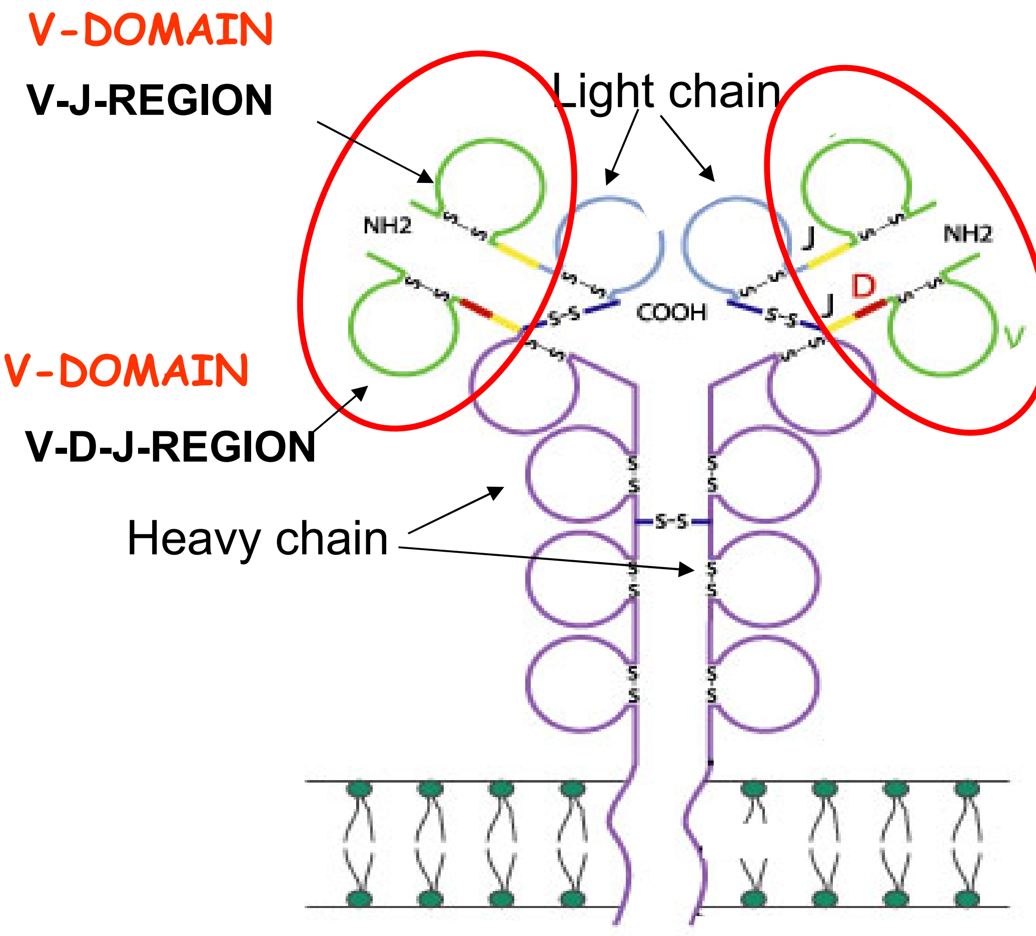
- **VH**: V-DOMAIN of the immunoglobulin heavy chain
- **VL**: V-DOMAIN of the immunoglobulin light chain
- **CH1, CH2, CH3**: C-DOMAIN of the immunoglobulin heavy chain
- **CL**: C-DOMAIN of the immunoglobulin light chain

VH corresponds to the V-D-J-REGION (in green (V), orange (DJ)) of the heavy chain.
VL corresponds to the V-J-REGION (in green (V) and yellow (J)) of the light chain.

The Immunoglobulin FactsBook, 2001

Immunoglobulin (IG) T cell receptor (TR)

http://www.imgt.org



V-DOMAIN

V-J-REGION

Light chain

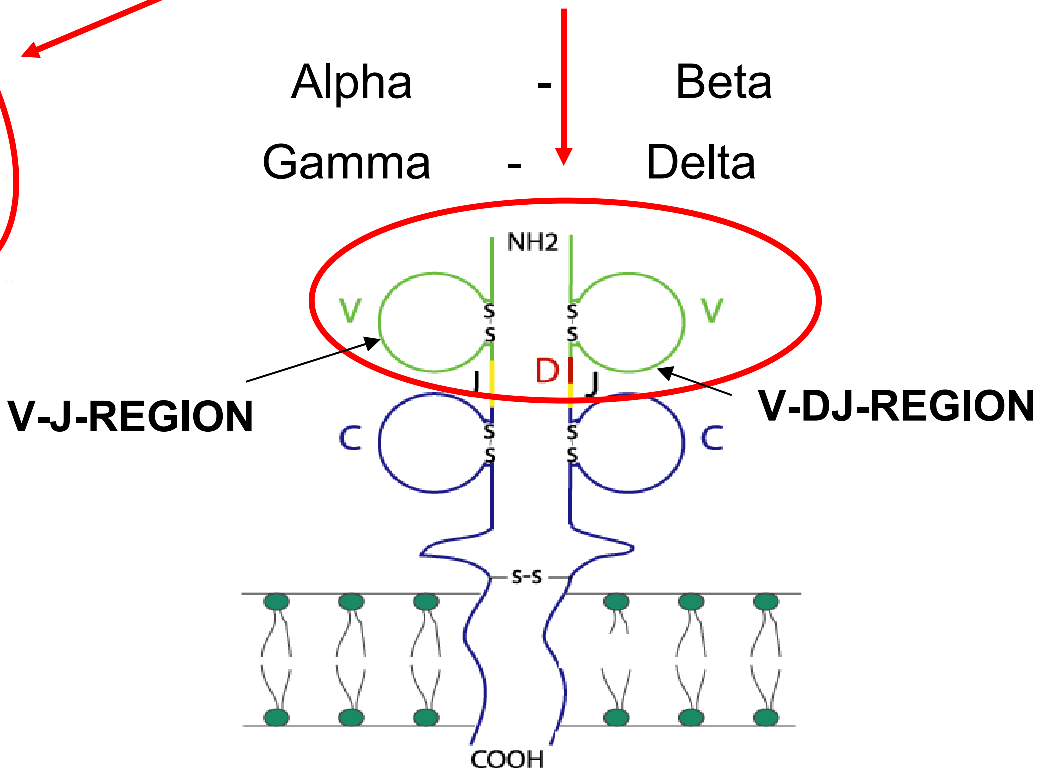
V-DOMAIN

V-D-J-REGION

Heavy chain

Membrane IgM

Contribution of the
2 V-DOMAINs
to the antigen binding site



Alpha - Beta

Gamma - Delta

V-J-REGION

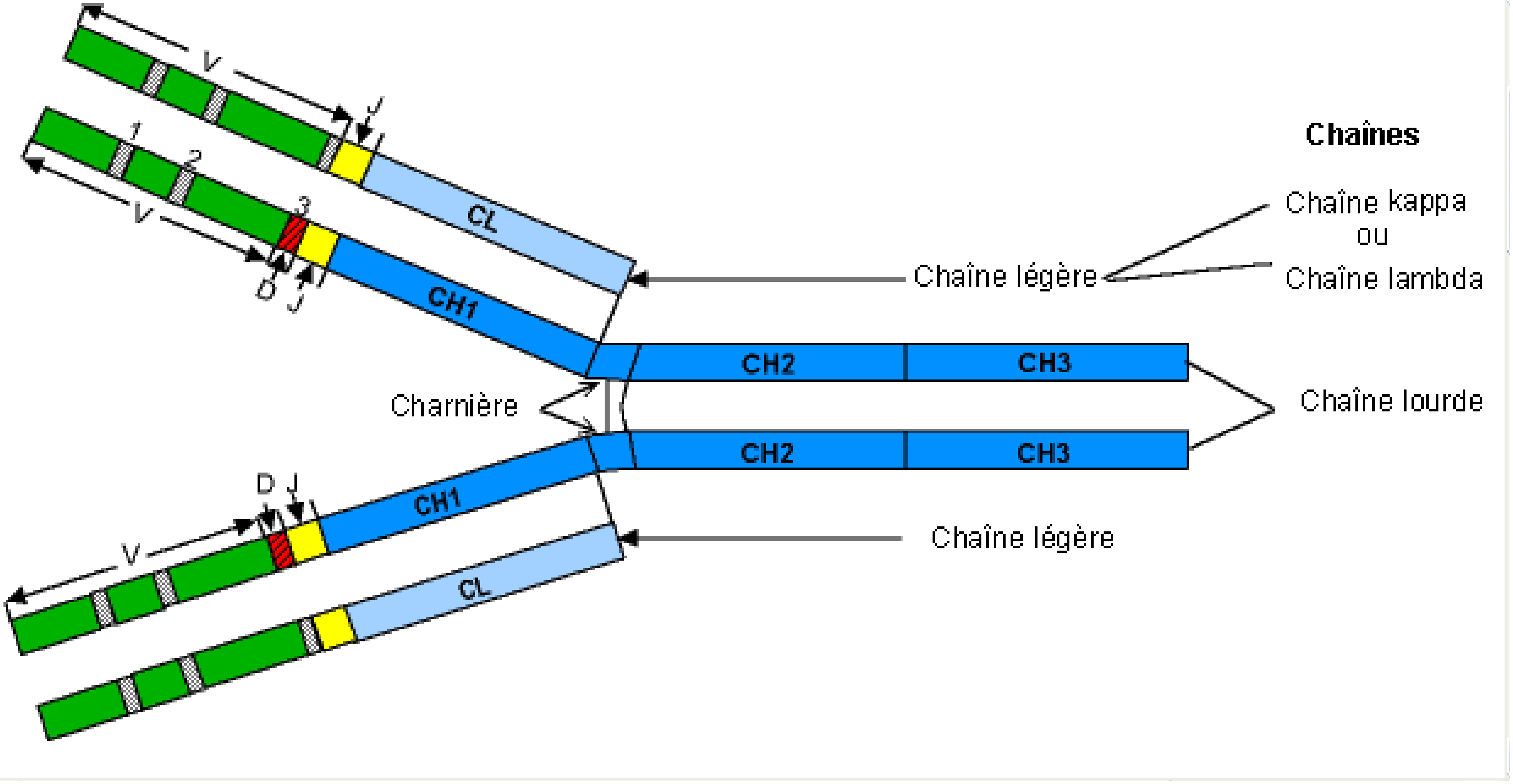
V-DJ-REGION

T cell receptor

The Immunoglobulin FactsBook, 2001

Immunoglobulin IgG

http://www.imgt.org



Chaînes

Chaîne kappa
ou
Chaîne lambda

Chaîne légère

Chaîne lourde

Charnière

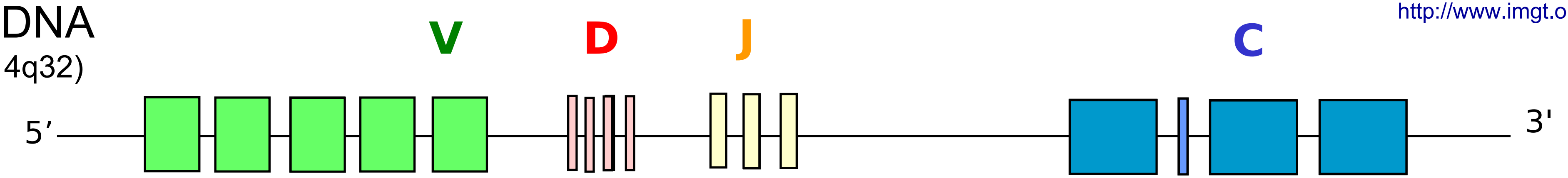
Chaîne légère

IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

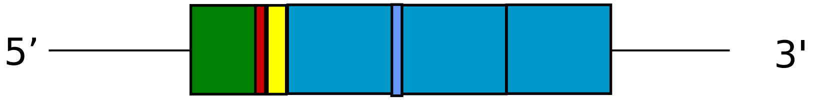
http://www.imgt.org

genomic DNA
(IGH Locus 14q32)

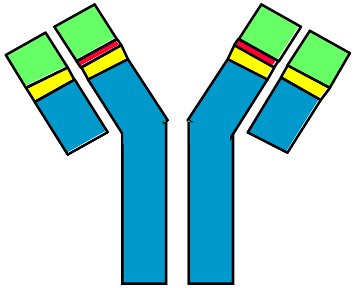


rearranged
DNA

mRNA



2 x 10^{12} different IG
per individual

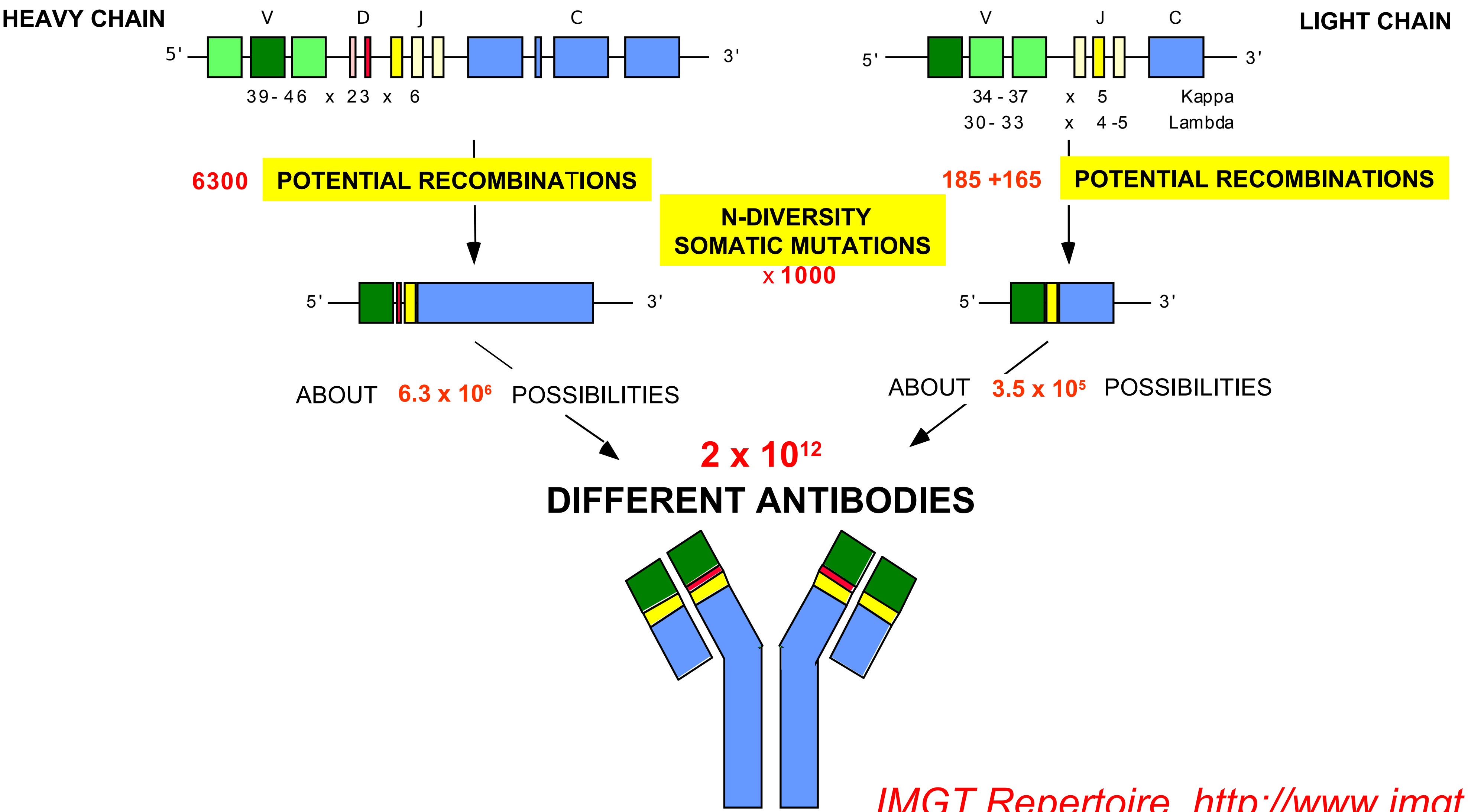


IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

http://www.imgt.org

150
FUNCTIONAL IG GENES



HEAVY CHAIN

39-46 x 23 x 6

LIGHT CHAIN

34 - 37 x 5 Kappa

30 - 33 x 4 -5 Lambda

6300 POTENTIAL RECOMBINATIONS

185 +165 POTENTIAL RECOMBINATIONS

N-DIVERSITY
SOMATIC MUTATIONS
x 1000

ABOUT **6.3×10^6** POSSIBILITIES

ABOUT **3.5×10^5** POSSIBILITIES

2×10^{12}
DIFFERENT ANTIBODIES

IMGT Repertoire, http://www.imgt.org

What was the first goal?

CONTENT

1. Immunoglobulins AND T cell receptors
2. Human AND other vertebrate species

DATA

1. Genes AND alleles
2. Sequences (germline AND rearranged)
3. Polymorphisms: genomic AND somatic hypermutations
4. 3D structures

IMGT® Web resources

http://www.imgt.org

IMGT Repertoire 10,000 pages HTML

- Chromosomal localizations
- Locus representations
- Genes tables
- Potential repertoire
- Alignments of alleles
- Protein displays
- IMGT Colliers de Perles...

IMGT Index

IMGT Bloc-notes

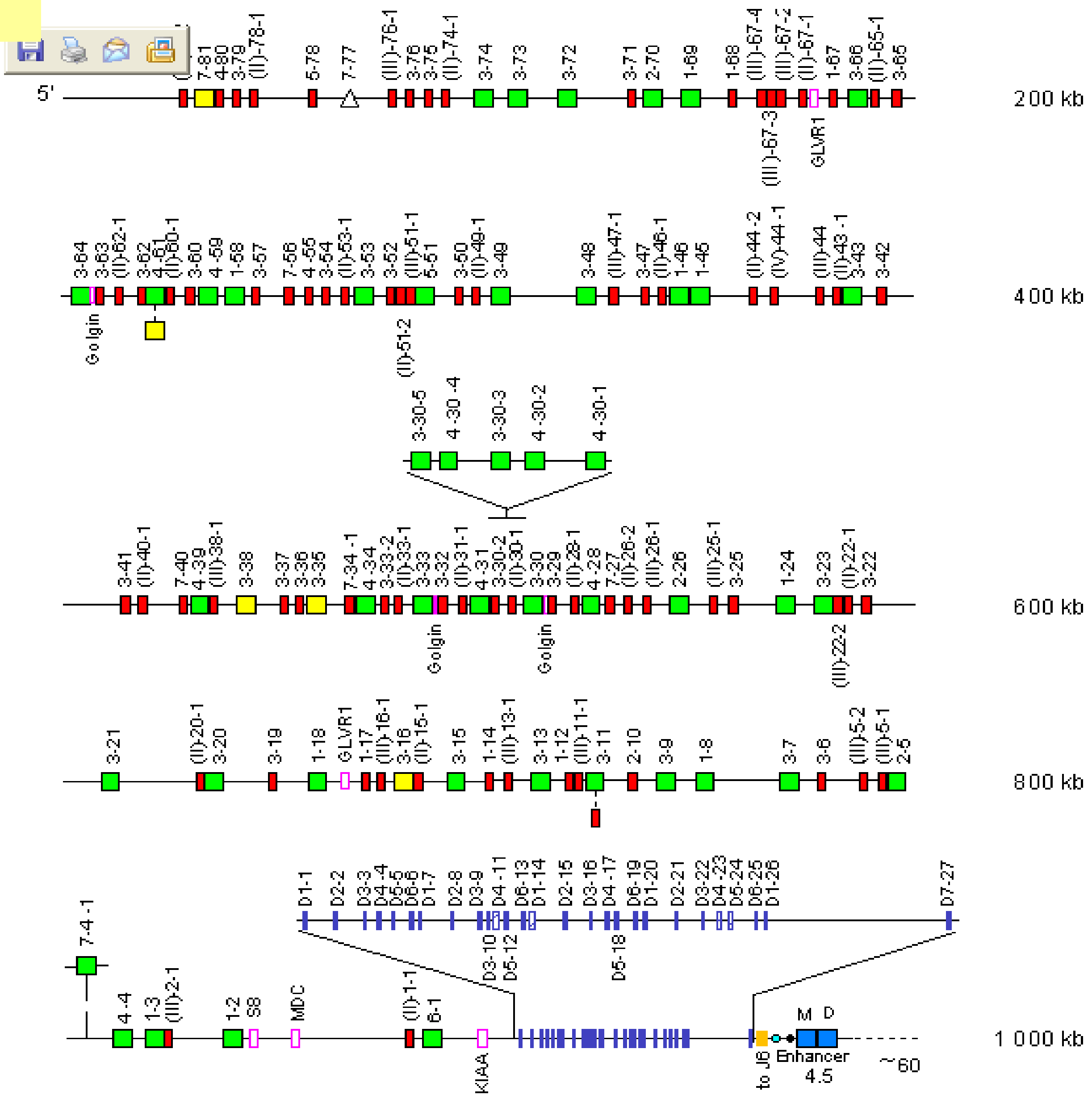
- Interesting links, PubMed,
- Meeting announcements
- IMGT Immunoinformatics page,...

IMGT Education

- Tutorials, IMGT Lexique, ...

Human IGH locus

Chromosome
14q32.33



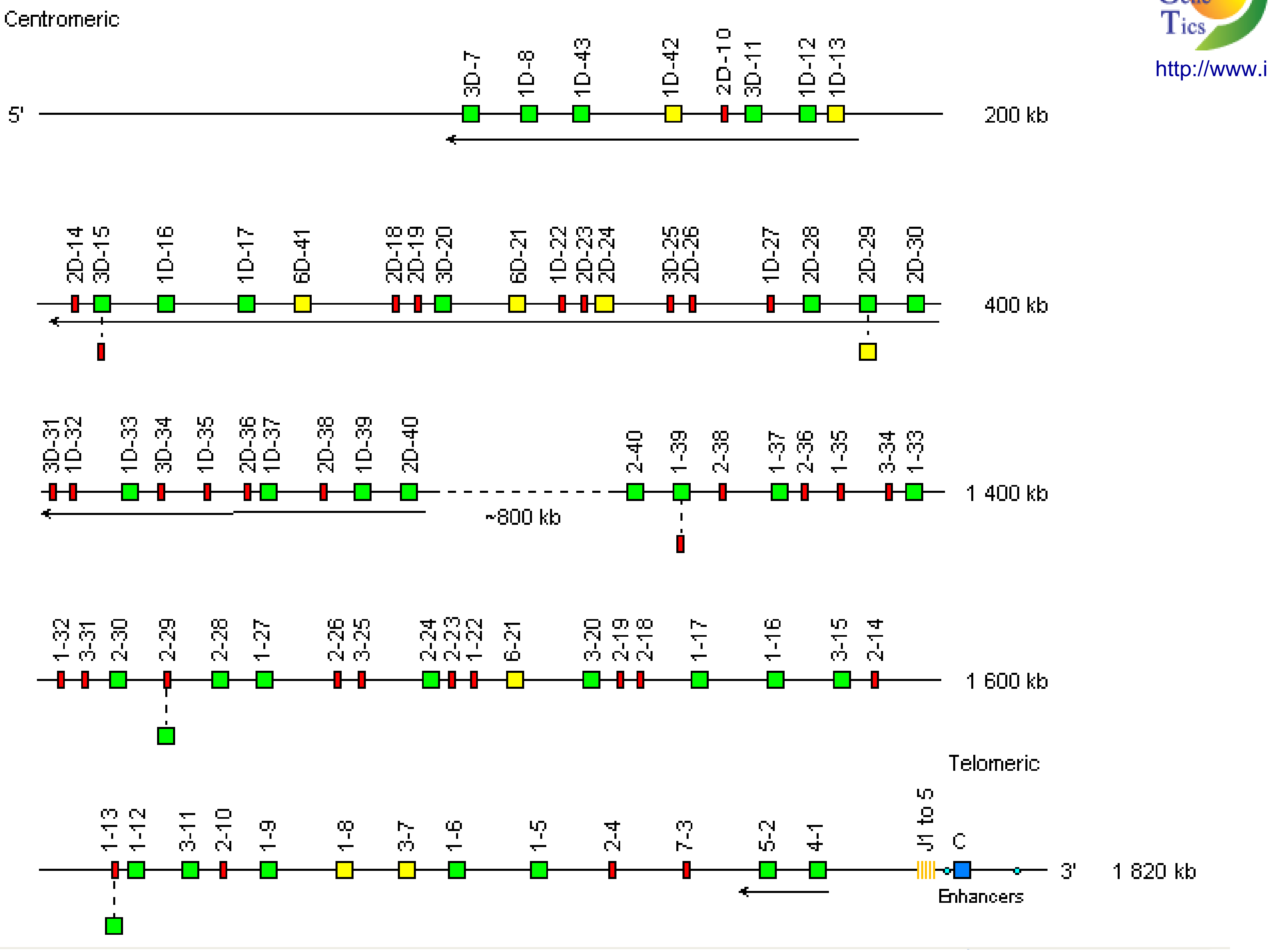
IMGT Repertoire, http://www.imgt.org

http://www.imgt.org

Human IGK locus

Chromosome
2p11.2

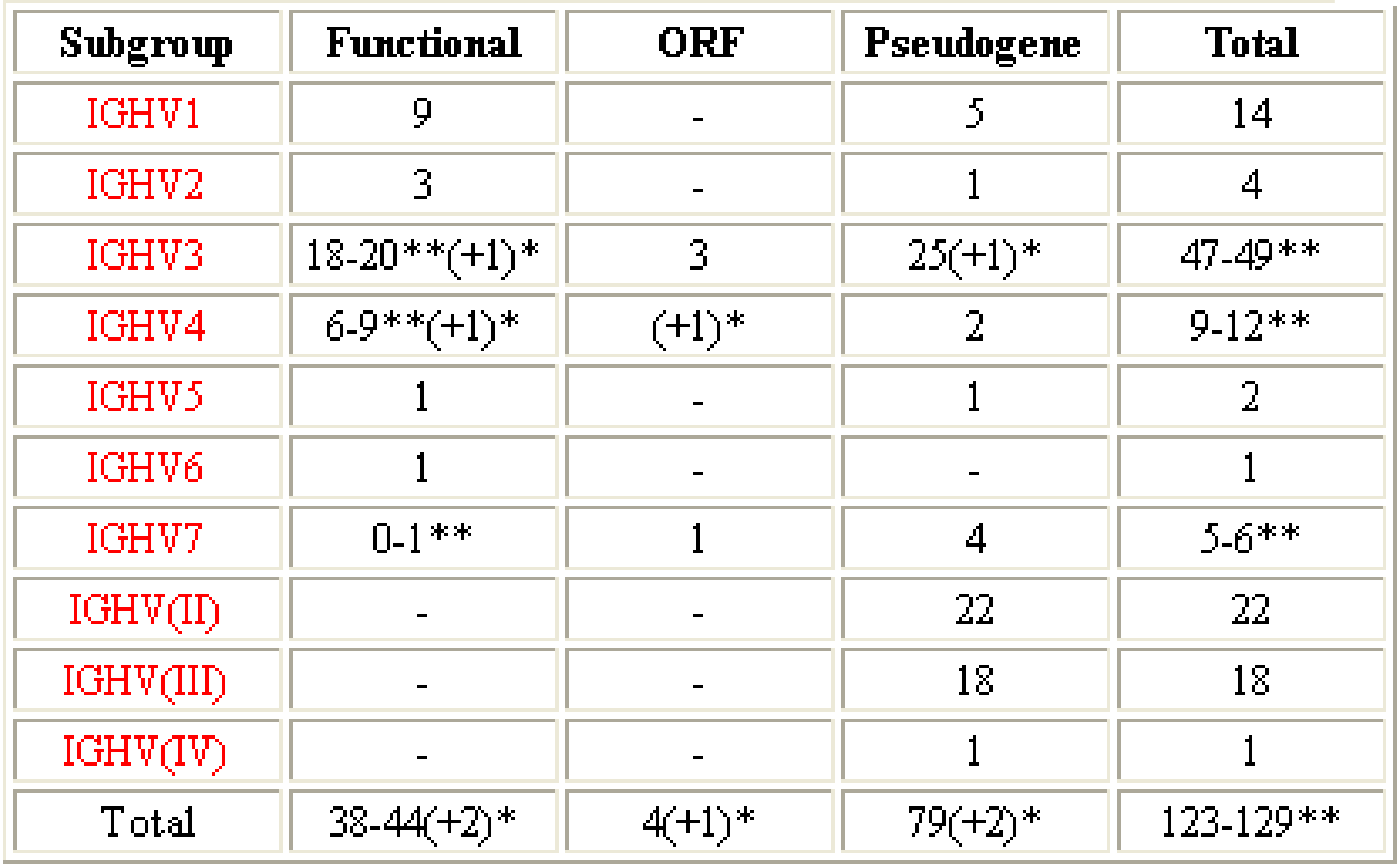
http://www.imgt.org



IMGT Repertoire, http://www.imgt.org

Human IGHV genes on chromosome 14 at 14q32.33

http://www.imgt.org

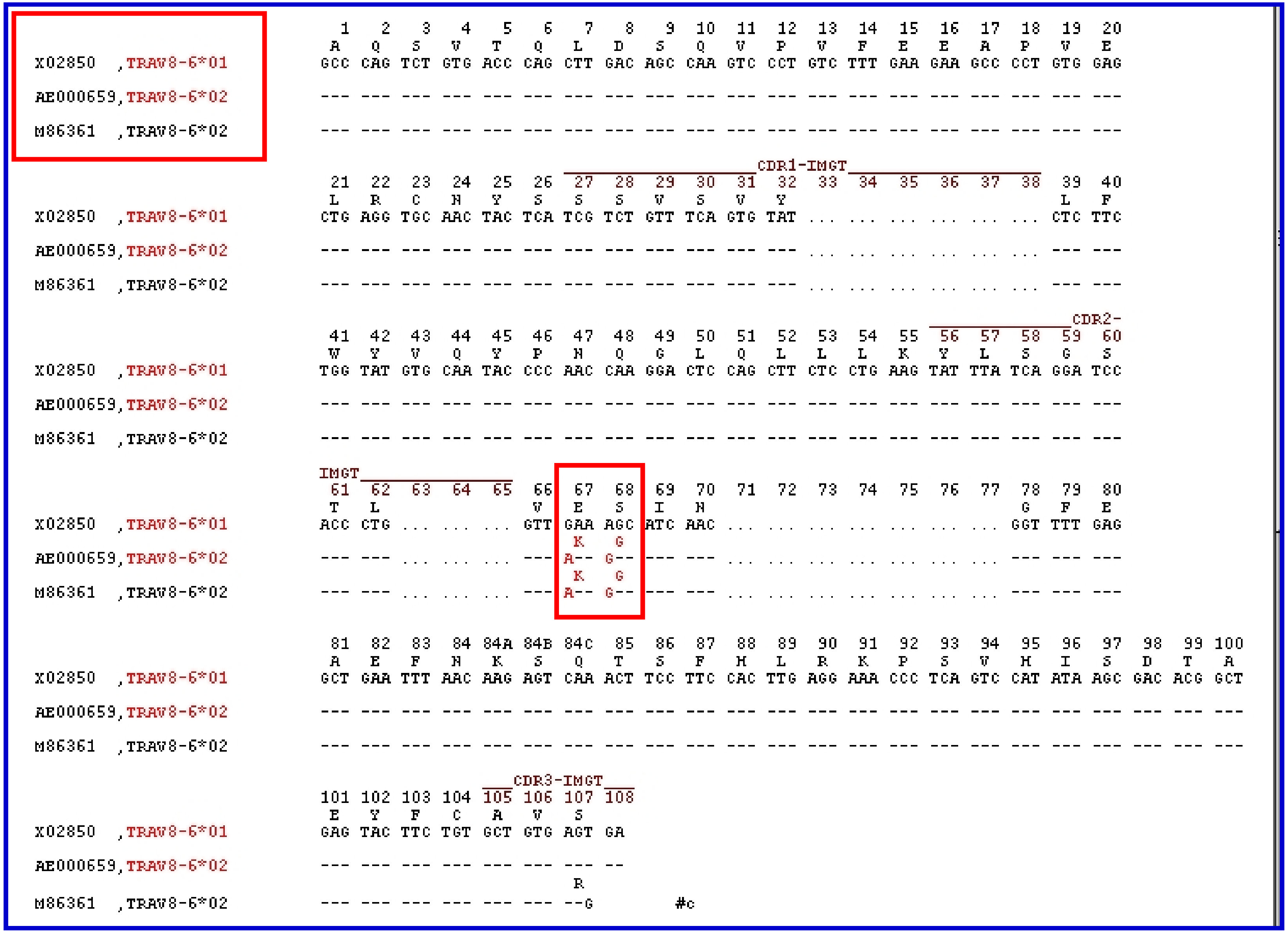


Subgroup	Functional	ORF	Pseudogene	Total
IGHV1	9	-	5	14
IGHV2	3	-	1	4
IGHV3	18-20**(+1)*	3	25(+1)*	47-49**
IGHV4	6-9**(+1)*	(+1)*	2	9-12**
IGHV5	1	-	1	2
IGHV6	1	-	-	1
IGHV7	0-1**	1	4	5-6**
IGHV(II)	-	-	22	22
IGHV(III)	-	-	18	18
IGHV(IV)	-	-	1	1
Total	38-44(+2)*	4(+1)*	79(+2)*	123-129**

IMGT Repertoire, http://www.imgt.org

IMGT Alignment of alleles

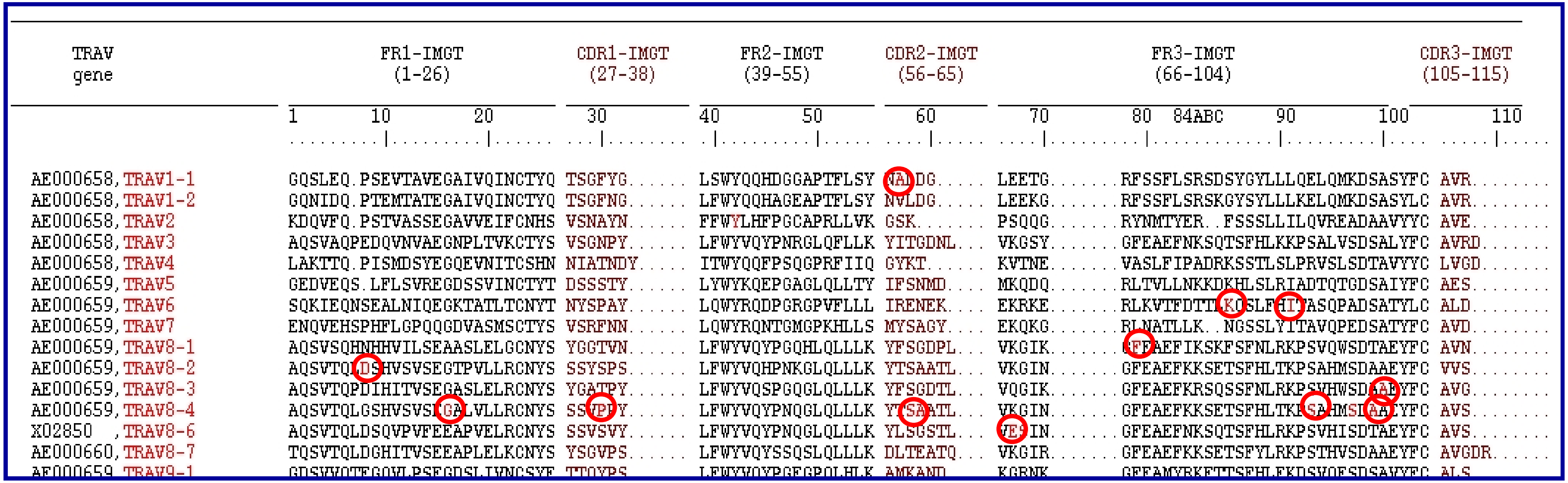
http://www.imgt.org



```
                          1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
                          A   Q   S   V   T   Q   L   D   S   Q   V   P   V   F   E   E   A   P   V   E
X02850    ,TRAV8-6*01    GCC CAG TCT GTG ACC CAG CTT GAC AGC CAA GTC CCT GTC TTT GAA GAA GCC CCT GTG GAG

AE000659,TRAV8-6*02      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361    ,TRAV8-6*02    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                 ____________________CDR1-IMGT___________________
                         21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
                          L   R   C   N   Y   S   S   S   V   S   V   Y                           L   F
X02850    ,TRAV8-6*01    CTG AGG TGC AAC TAC TCA TCG TCT GTT TCA GTG TAT ... ... ... ... ... ... CTC TTC

AE000659,TRAV8-6*02      --- --- --- --- --- --- --- --- --- --- --- --- ... ... ... ... ... ... --- ---

M86361    ,TRAV8-6*02    --- --- --- --- --- --- --- --- --- --- --- --- ... ... ... ... ... ... --- ---

                                                                             ____________CDR2-
                         41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
                          W   Y   V   Q   Y   P   N   Q   G   L   Q   L   L   L   K   Y   L   S   G   S
X02850    ,TRAV8-6*01    TGG TAT GTG CAA TAC CCC AAC CAA GGA CTC CAG CTT CTC CTG AAG TAT TTA TCA GGA TCC

AE000659,TRAV8-6*02      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361    ,TRAV8-6*02    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                         IMGT_______________
                         61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
                          T   L               V   E   S   I   N                               G   F   E
X02850    ,TRAV8-6*01    ACC CTG ... ... ... GTT GAA AGC ATC AAC ... ... ... ... ... ... ... GGT TTT GAG
                                                  K   G
AE000659,TRAV8-6*02      --- --- ... ... ... --- A-- G-- --- --- ... ... ... ... ... ... ... --- --- ---
                                                  K   G
M86361    ,TRAV8-6*02    --- --- ... ... ... --- A-- G-- --- --- ... ... ... ... ... ... ... --- --- ---

                         81  82  83  84 84A 84B 84C  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100
                          A   E   F   N   K   S   Q   T   S   F   H   L   R   K   P   S   V   H   I   S   D   T   A
X02850    ,TRAV8-6*01    GCT GAA TTT AAC AAG AGT CAA ACT TCC TTC CAC TTG AGG AAA CCC TCA GTC CAT ATA AGC GAC ACG GCT

AE000659,TRAV8-6*02      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

M86361    ,TRAV8-6*02    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                         ___CDR3-IMGT___
                         101 102 103 104 105 106 107 108
                          E   Y   F   C   A   V   S
X02850    ,TRAV8-6*01    GAG TAC TTC TGT GCT GTG AGT GA

AE000659,TRAV8-6*02      --- --- --- --- --- --- --- --
                                                  R
M86361    ,TRAV8-6*02    --- --- --- --- --- --- --G        #c
```

IMGT Repertoire, http://www.imgt.org

IMGT Protein display

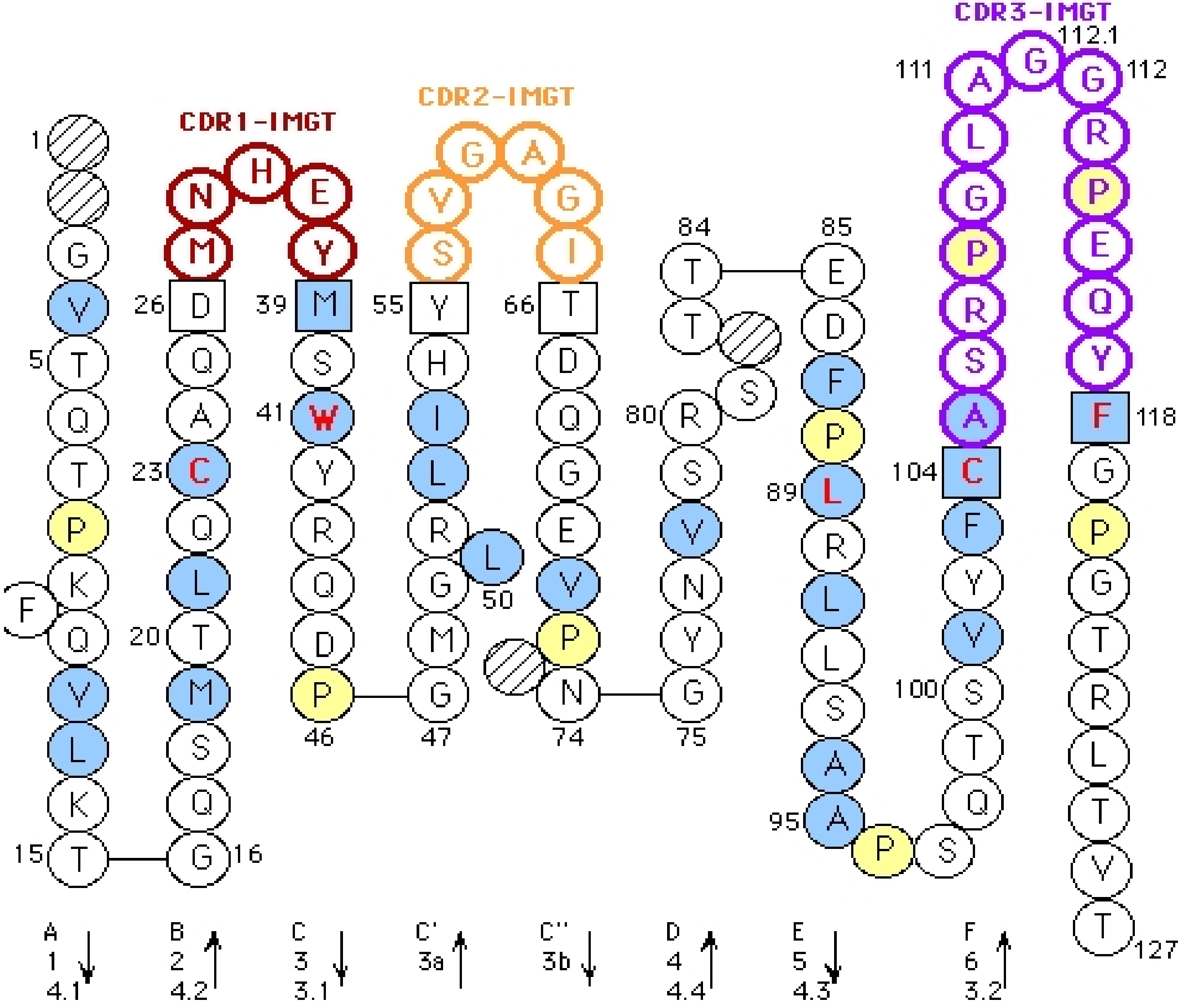


TRAV gene	FR1-IMGT (1-26)	CDR1-IMGT (27-38)	FR2-IMGT (39-55)	CDR2-IMGT (56-65)	FR3-IMGT (66-104)	CDR3-IMGT (105-115)
	1 10 20	30	40 50	60	70 80 84ABC 90 100	110
AE000658,TRAV1-1	GQSLEQ.PSEVTAVEGAIVQINCTYQ	TSGFYG......	LSWYQQHDGGAPTFLSY	NALDG.....	LEETG.......RFSSFLSRSDSYGYLLLQELQMKDSASYFC	AVR........
AE000658,TRAV1-2	GQNIDQ.PTEMTATEGAIVQINCTYQ	TSGFNG......	LFWYQQHAGEAPTFLSY	NVLDG.....	LEEKG.......RFSSFLSRSKGYSYLLLKELQMKDSASYLC	AVR........
AE000658,TRAV2	KDQVFQ.PSTVASSEGAVVEIFCNHS	VSNAYN......	FFWYLHFPGCAPRLLVK	GSK.......	PSQQG.......RYNMTYER..FSSSLLILQVREADAAVYYC	AVE........
AE000658,TRAV3	AQSVAQPEDQVNVAEGNPLTVKCTYS	VSGNPY......	LFWYVQYPNRGLQFLLK	YITGDNL...	VKGSY.......GFEAEFNKSQTSFHLKKPSALVSDSALYFC	AVRD.......
AE000658,TRAV4	LAKTTQ.PISMDSYEGQEVNITCSHN	NIATNDY.....	ITWYQQFPSQGPRFIIQ	GYKT......	KVTNE.......VASLFIPADRKSSTLSLPRVSLSDTAVYYC	LVGD.......
AE000659,TRAV5	GEDVEQS.LFLSVREGDSSVINCTYT	DSSSTY......	LYWYKQEPGAGLQLLTY	IFSNMD....	MKQDQ.......RLTVLLNKKDKHLSLRIADTQTGDSAIYFC	AES........
AE000659,TRAV6	SQKIEQNSEALNIQEGKTATLTCNYT	NYSPAY......	LQWYRQDPGRGPVFLLL	IRENEK....	EKRKE.......RLKVTFDTTLKGSLFHITASQPADSATYLC	ALD........
AE000659,TRAV7	ENQVEHSPHFLGPQQGDVASMSCTYS	VSRFNN......	LQWYRQNTGMGPKHLLS	MYSAGY....	EKQKG.......RLNATLLK..NGSSLYITAVQPEDSATYFC	AVD........
AE000659,TRAV8-1	AQSVSQHNHHVILSEAASLELGCNYS	YGGTVN......	LFWYVQYPGQHLQLLLK	YFSGDPL...	VKGIK.......GFEAEFIKSKFSFNLRKPSVQWSDTAEYFC	AVN........
AE000659,TRAV8-2	AQSVTQLDSHVSVSEGTPVLLRCNYS	SSYSPS......	LFWYVQHPNKGLQLLLK	YTSAATL...	VKGIN.......GFEAEFKKSETSFHLTKPSAHMSDAAEYFC	VVS........
AE000659,TRAV8-3	AQSVTQPDIHITVSEGASLELRCNYS	YGATPY......	LFWYVQSPGQGLQLLLK	YFSGDTL...	VQGIK.......GFEAEFKRSQSSFNLRKPSVHWSDAAEYFC	AVG........
AE000659,TRAV8-4	AQSVTQLGSHVSVSEGALVLLRCNYS	SSVPPY......	LFWYVQYPNQGLQLLLK	YTSAATL...	VKGIN.......GFEAEFKKSETSFHLTKPSAHMSDAAEYFC	AVS........
X02850 ,TRAV8-6	AQSVTQLDSQVPVFEEAPVELRCNYS	SSVSVY......	LFWYVQYPNQGLQLLLK	YLSGSTL...	VESIN.......GFEAEFNKSQTSFHLRKPSVHISDTAEYFC	AVS........
AE000660,TRAV8-7	TQSVTQLDGHITVSEEAPLELKCNYS	YSGVPS......	LFWYVQYSSQSLQLLLK	DLTEATQ...	VKGIR.......GFEAEFKKSETSFYLRKPSTHVSDAAEYFC	AVGDR......
AE000659,TRAV9-1	GDSVVQTEGQVLPSEGDSLIVNCSYE	TTQYPS	LFWYVQYPGEGPQLHLK	AMKAND	KGRNK	GFEAMYRKETTSFHLEKDSVQESDSAVYFC	ALS

IMGT Repertoire, http://www.imgt.org

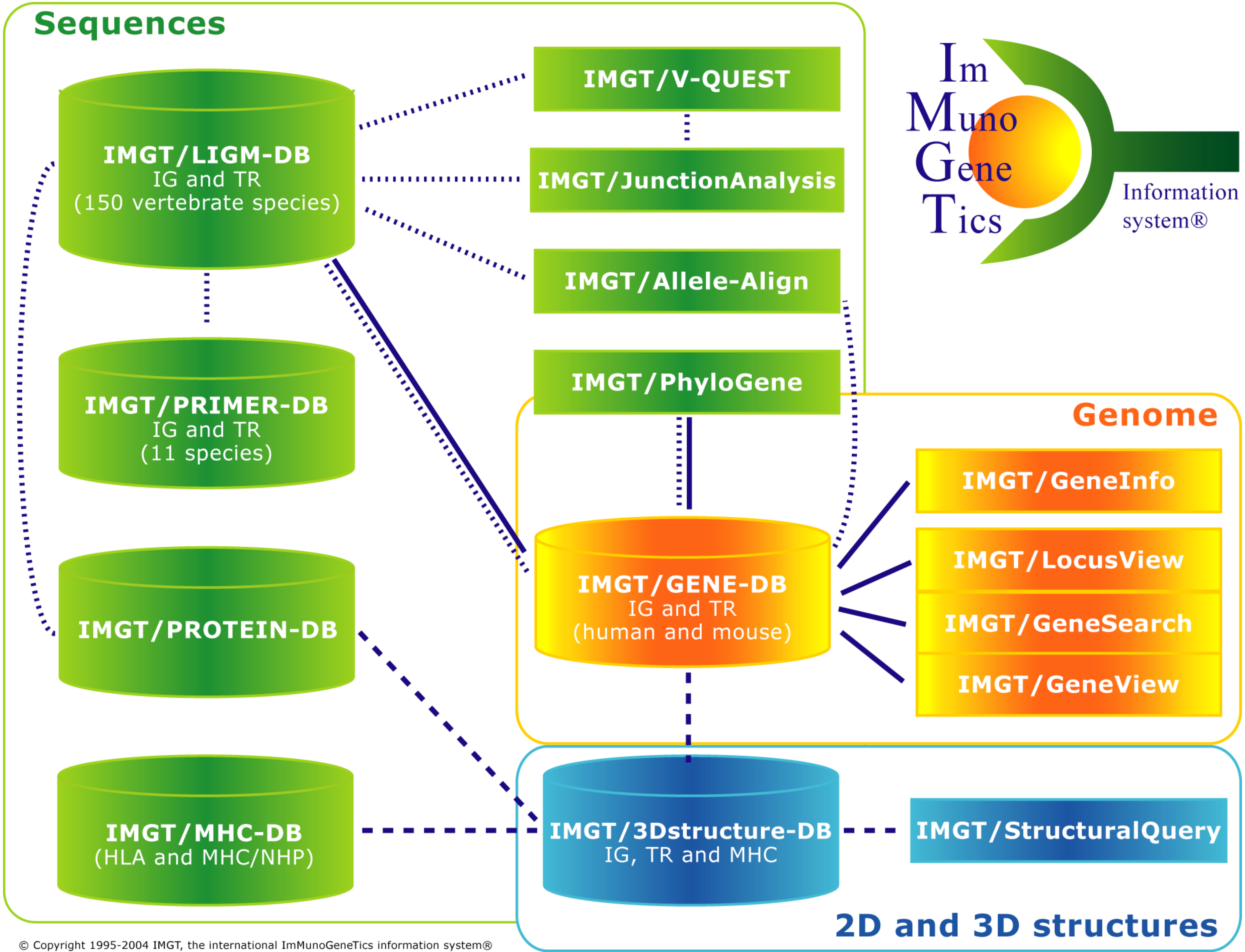
IMGT Collier de Perles

http://www.imgt.org



IMGT Repertoire, http://www.imgt.org

IMGT databases and tools



Im
Muno
Gene
Tics
Information system®

Sequences

IMGT/LIGM-DB
IG and TR
(150 vertebrate species)

IMGT/PRIMER-DB
IG and TR
(11 species)

IMGT/PROTEIN-DB

IMGT/MHC-DB
(HLA and MHC/NHP)

IMGT/V-QUEST

IMGT/JunctionAnalysis

IMGT/Allele-Align

IMGT/PhyloGene

Genome

IMGT/GENE-DB
IG and TR
(human and mouse)

IMGT/GeneInfo

IMGT/LocusView

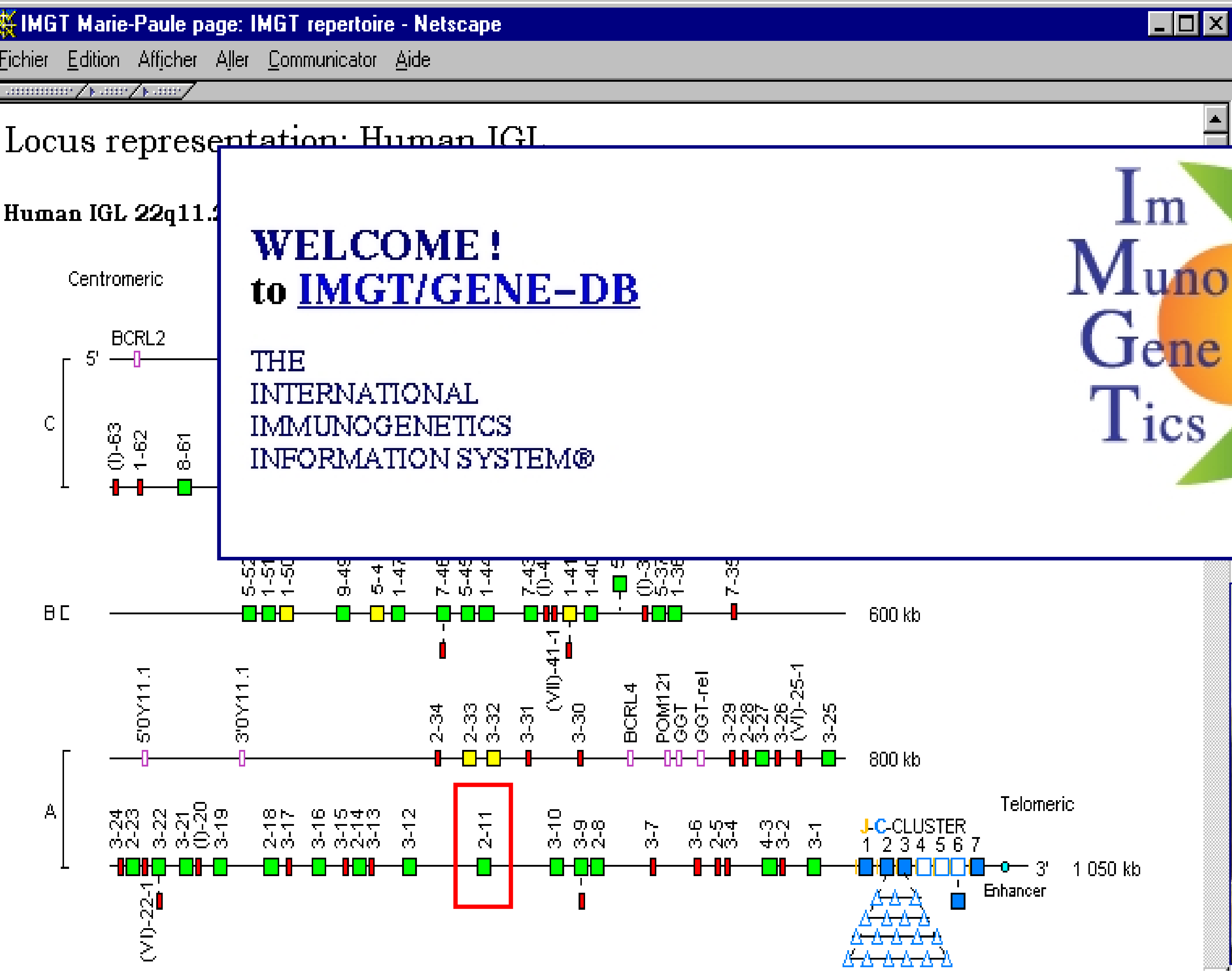
IMGT/GeneSearch

IMGT/GeneView

2D and 3D structures

IMGT/3Dstructure-DB
IG, TR and MHC

IMGT/StructuralQuery



Locus representation: Human IGL

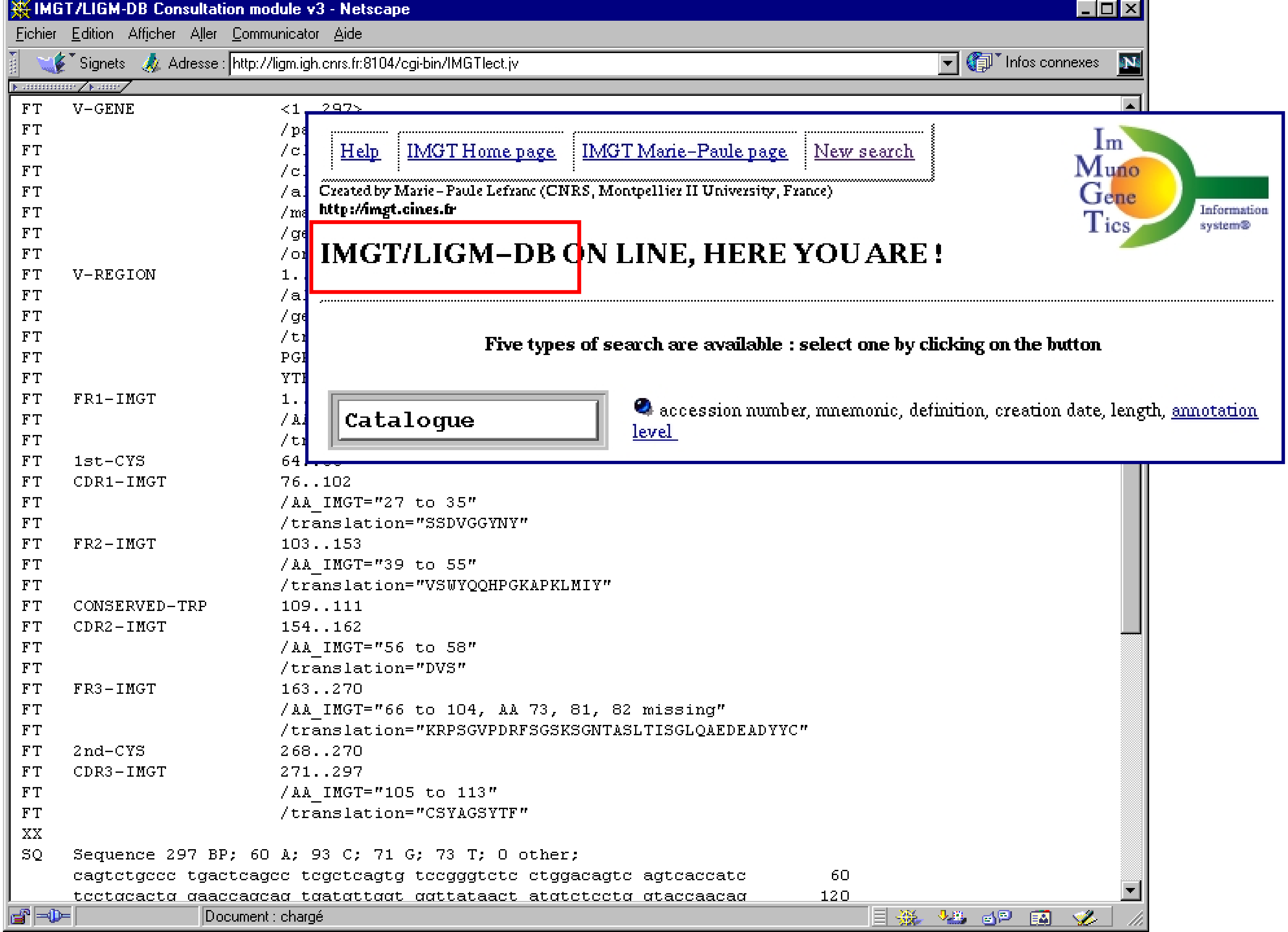
Human IGL 22q11.

WELCOME !
to IMGT/GENE-DB

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

IMmunoGeneTics
Information
system®

Giudicelli V. et al. Nucleic Acids Res. 33, D256-D261 (2005)



```
FT   V-GENE          <1  297>
FT                   /p
FT                   /c
FT                   /c
FT                   /a
FT                   /m
FT                   /g
FT                   /o
FT   V-REGION        1.
FT                   /a
FT                   /g
FT                   /t
FT                   PG
FT                   YT
FT   FR1-IMGT        1.
FT                   /A
FT                   /t
FT   1st-CYS         64
FT   CDR1-IMGT       76..102
FT                   /AA_IMGT="27 to 35"
FT                   /translation="SSDVGGYNY"
FT   FR2-IMGT        103..153
FT                   /AA_IMGT="39 to 55"
FT                   /translation="VSWYQQHPGKAPKLMIY"
FT   CONSERVED-TRP   109..111
FT   CDR2-IMGT       154..162
FT                   /AA_IMGT="56 to 58"
FT                   /translation="DVS"
FT   FR3-IMGT        163..270
FT                   /AA_IMGT="66 to 104, AA 73, 81, 82 missing"
FT                   /translation="KRPSGVPDRFSGSKSGNTASLTISGLQAEDEADYYC"
FT   2nd-CYS         268..270
FT   CDR3-IMGT       271..297
FT                   /AA_IMGT="105 to 113"
FT                   /translation="CSYAGSYTF"
XX
SQ   Sequence 297 BP; 60 A; 93 C; 71 G; 73 T; 0 other;
     cagtctgccc tgactcagcc tcgctcagtg tccgggtctc ctggacagtc agtcaccatc        60
     tcctgcactg gaaccagcag tgatgttggt ggttataact atgtctcctg gtaccaacag       120
```

Created by Marie-Paule Lefranc (CNRS, Montpellier II University, France)
http://imgt.cines.fr

IMGT/LIGM-DB ON LINE, HERE YOU ARE !

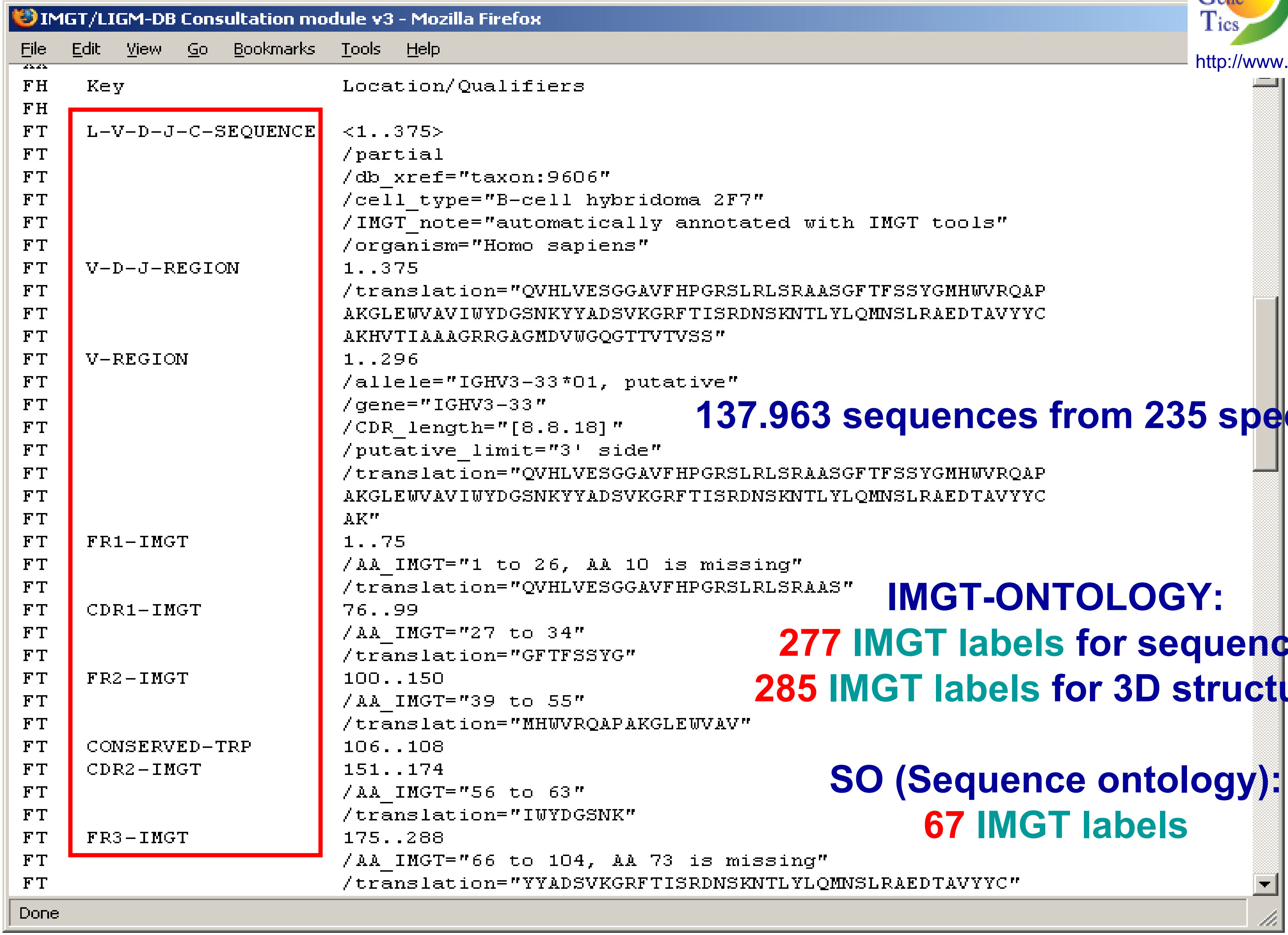
Five types of search are available : select one by clicking on the button

accession number, mnemonic, definition, creation date, length, annotation level

IMGT/LIGM-DB

http://www.imgt.org

D
E
S
C
R
I
P
T
I
O
N



```
FH   Key                  Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE   <1..375>
FT                        /partial
FT                        /db_xref="taxon:9606"
FT                        /cell_type="B-cell hybridoma 2F7"
FT                        /IMGT_note="automatically annotated with IMGT tools"
FT                        /organism="Homo sapiens"
FT   V-D-J-REGION         1..375
FT                        /translation="QVHLVESGGAVFHPGRSLRLSRAASGFTFSSYGMHWVRQAP
FT                        AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC
FT                        AKHVTIAAAGRRGAGMDVWGQGTTVTVSS"
FT   V-REGION             1..296
FT                        /allele="IGHV3-33*01, putative"
FT                        /gene="IGHV3-33"
FT                        /CDR_length="[8.8.18]"
FT                        /putative_limit="3' side"
FT                        /translation="QVHLVESGGAVFHPGRSLRLSRAASGFTFSSYGMHWVRQAP
FT                        AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC
FT                        AK"
FT   FR1-IMGT             1..75
FT                        /AA_IMGT="1 to 26, AA 10 is missing"
FT                        /translation="QVHLVESGGAVFHPGRSLRLSRAAS"
FT   CDR1-IMGT            76..99
FT                        /AA_IMGT="27 to 34"
FT                        /translation="GFTFSSYG"
FT   FR2-IMGT             100..150
FT                        /AA_IMGT="39 to 55"
FT                        /translation="MHWVRQAPAKGLEWVAV"
FT   CONSERVED-TRP        106..108
FT   CDR2-IMGT            151..174
FT                        /AA_IMGT="56 to 63"
FT                        /translation="IWYDGSNK"
FT   FR3-IMGT             175..288
FT                        /AA_IMGT="66 to 104, AA 73 is missing"
FT                        /translation="YYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC"
```

137.963 sequences from 235 species

IMGT-ONTOLOGY:

277 IMGT labels for sequences

285 IMGT labels for 3D structures

SO (Sequence ontology):

67 IMGT labels

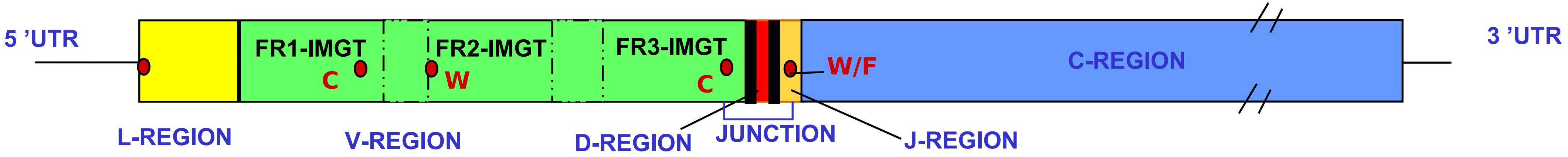
cDNA (in databases: mRNA!)

```
……….gagga ttcaccatgg aactggggct ccgctgggtt ttccttgttg ctttttaga 120
aggtgtccag tgtgaggtgc aactggtgga gtctggggga ggcctggtca agccgggggg 180
gtccctgaga ctctcctgtg cagcctctgg attaagcttc agtacctatg ccatgaactg 240
ggtccgccag gctccaggga aggggctgga atgggtctca agtattagta gtagaagtga 300
ttacatatac tatagagact cagtgaaggg ccgattcacc atctccagag acaacgccaa 360
gaattcactg tatctgcaaa tgaatagcct gagagtcgac gacacggctg tctattactg 420
tgcgagagat tcttgtaatg gtgctatatg ttatggtttc agtccctggg gccagggaac 480
cctggtcacc gtctcctcag catccccgac cagccccaag gtcttcccgc tgagcctctg 540
cagcacccag ccagatggga acgtggtcat cgcctgcctg gtccagggct tcttccccca 600
ggagccactc agtgtgacct ggagcgaaag cggacagggc gtgaccgcca gaaacttccc 660
acccagccag gatgcctccg gggacctgta caccacgagc agccagctga ccctgccggc 720
cacacagtgc ctagccggca agtccgtgac atgccacgtg aagcactaca cgaatcccag 780
ccaggatgtg actgtgccct gcccagttcc ctcaactcca cctaccccat ctccctcaac 840
tccacctacc ccatctccct catgctgcca cccccgactg tcactgcacc gaccggccct 900
cgaggacctg ctcttaggtt cagaagcgaa cctcacgtgc acactgaccg gcctgagaga 960
tgcctcgggt gtcaccttca cctggacgcc ctcaagtggg aagagcgctg ttcaaggacc 1020
acctgagcgt gacctctgtg gctgctacag cgtgtccagt gtcctgccgg gctgtgccga 1080
gccatggaac catgggaaga ccttcacttg cactgctgcc taccccgagt ccaagacccc 1140
gctaaccgcc accctctcaa aatccggaaa cacattccgg cccgaggtcc acctgctgcc 1200
gccgccgtcg gaggagctgg ccctgaacga gctggtgacg ctgacgtgcc tggcacgtgg 1260
cttcagcccc aaggacgtgc tggttcgctg gctgcagggg tcacaggagc tgccccgcga 1320
gaagtacctg acttgggcat cccggcagga gcccagccag ggcaccacca ccttcgctgt 1380
gaccagcata ctgcgcgtgg cagccgagga ctggaagaag gggacacct tctcctgcat 1440
ggtgggccac gaggccctgc cgctggcctt cacacagaag accatcgacc gcttggcggg 1500
taaacccacc catgtcaatg tgtctgttgt catggcggag gtggacggca cctgctactga 1560 …..
```

cDNA

http://www.imgt.org

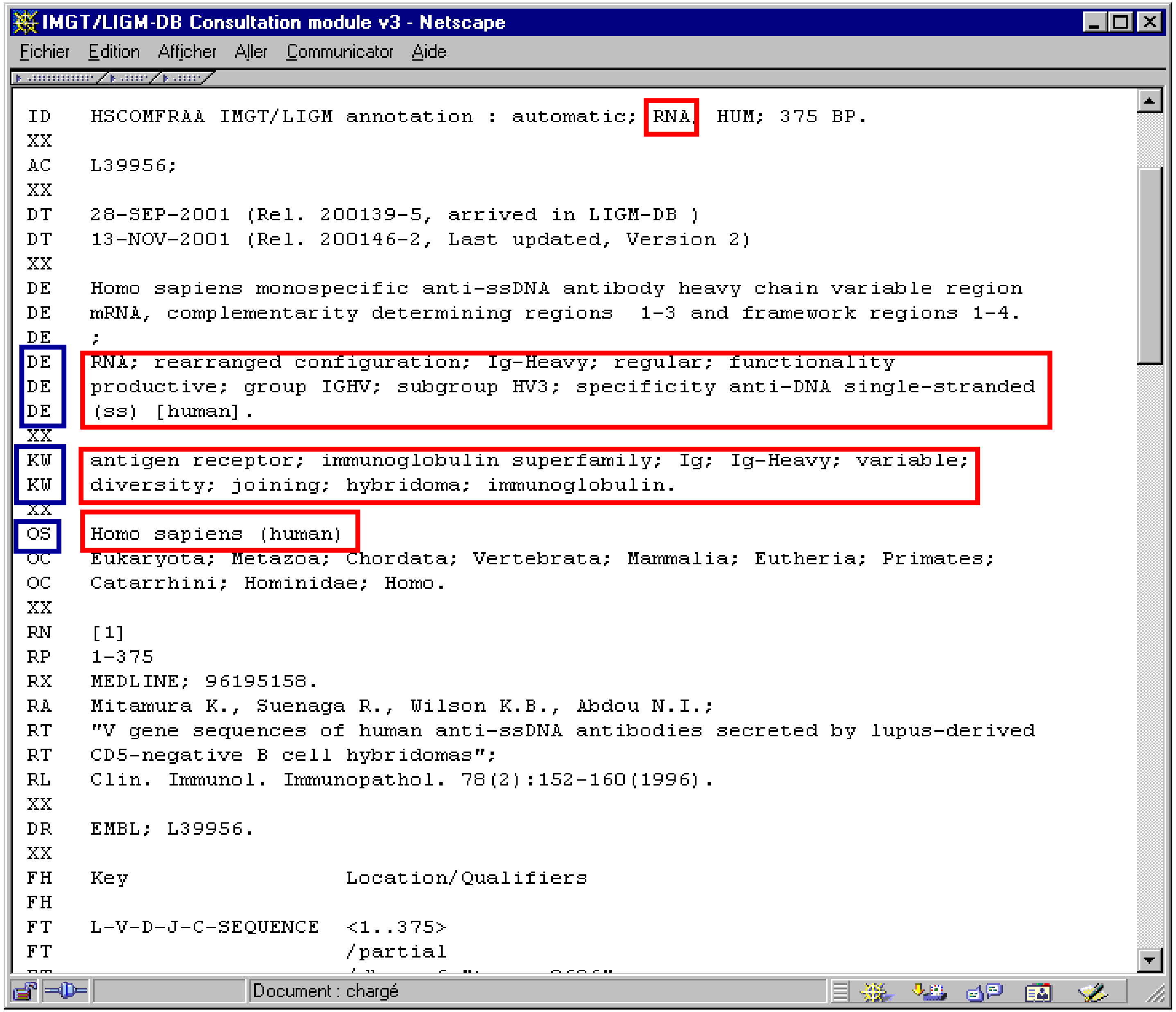
```
………..gagga ttcaccatgg aactggggct ccgctgggtt ttccttgttg cttttttaga 120
aggtgtccag tgtgaggtgc aactggtgga gtctggggga ggcctggtca agccgggggg 180
gtccctgaga ctctcctgtg cagcctctgg attaagcttc agtacctatg ccatgaactg 240
ggtccgccag gctccaggga aggggctgga atgggtctca agtattagta gtagaagtga 300
ttacatatac tatagagact cagtgaaggg ccgattcacc atctccagag acaacgccaa 360
gaattcactg tatctgcaaa tgaatagcct gagagtcgac gacacggctg tctattactg 420
tgcgagagat tcttgtaatg gtgctatatg ttatggtttc agtcctggg gccagggaac 480
cctggtcacc gtctcctcag catccccgac cagccccaag gtcttcccgc tgagcctctg 540
cagcacccag ccagatggga acgtggtcat cgcctgcctg gtccagggct tcttccccca 600
ggagccactc agtgtgacct ggagcgaaag cggacagggc gtgaccgcca gaaacttccc 660
acccagccag gatgcctccg gggacctgta caccacgagc agccagctga ccctgccggc 720
cacacagtgc ctagccggca agtccgtgac atgccacgtg aagcactaca cgaatcccag 780
ccaggatgtg actgtgccct gcccagttcc ctcaactcca cctaccccat ctccctcaac 840
tccacctacc ccatctccct catgctgcca cccccgactg tcactgcacc gaccggccct 900
cgaggacctg ctcttaggtt cagaagcgaa cctcacgtgc acactgaccg gcctgagaga 960
tgcctcgggt gtcaccttca cctggacgcc ctcaagtggg aagagcgctg ttcaaggacc 1020
acctgagcgt gacctctgtg gctgctacag cgtgtccagt gtcctgccgg gctgtgccga 1080
gccatggaac catgggaaga ccttcacttg cactgctgcc taccccgagt ccaagacccc 1140
gctaaccgcc accctctcaa aatccggaaa cacattccgg cccgaggtcc acctgctgcc 1200
gccgccgtcg gaggagctgg ccctgaacga gctggtgacg ctgacgtgcc tggcacgtgg 1260
cttcagcccc aaggacgtgc tggttcgctg gctgcagggg tcacaggagc tgccccgcga 1320
gaagtacctg acttgggcat cccggcagga gcccagccag ggcaccacca ccttcgctgt 1380
gaccagcata ctgcgcgtgg cagccgagga ctggaagaag ggggacacct tctcctgcat 1440
ggtgggccac gaggccctgc cgctggcctt cacacagaag accatcgacc gcttggcggg 1500
taaacccacc catgtcaatg tgtctgttgt catggcggag gtggacggca cctgctactga 1560 …..
```



The IMGT-ONTOLOGY concepts

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

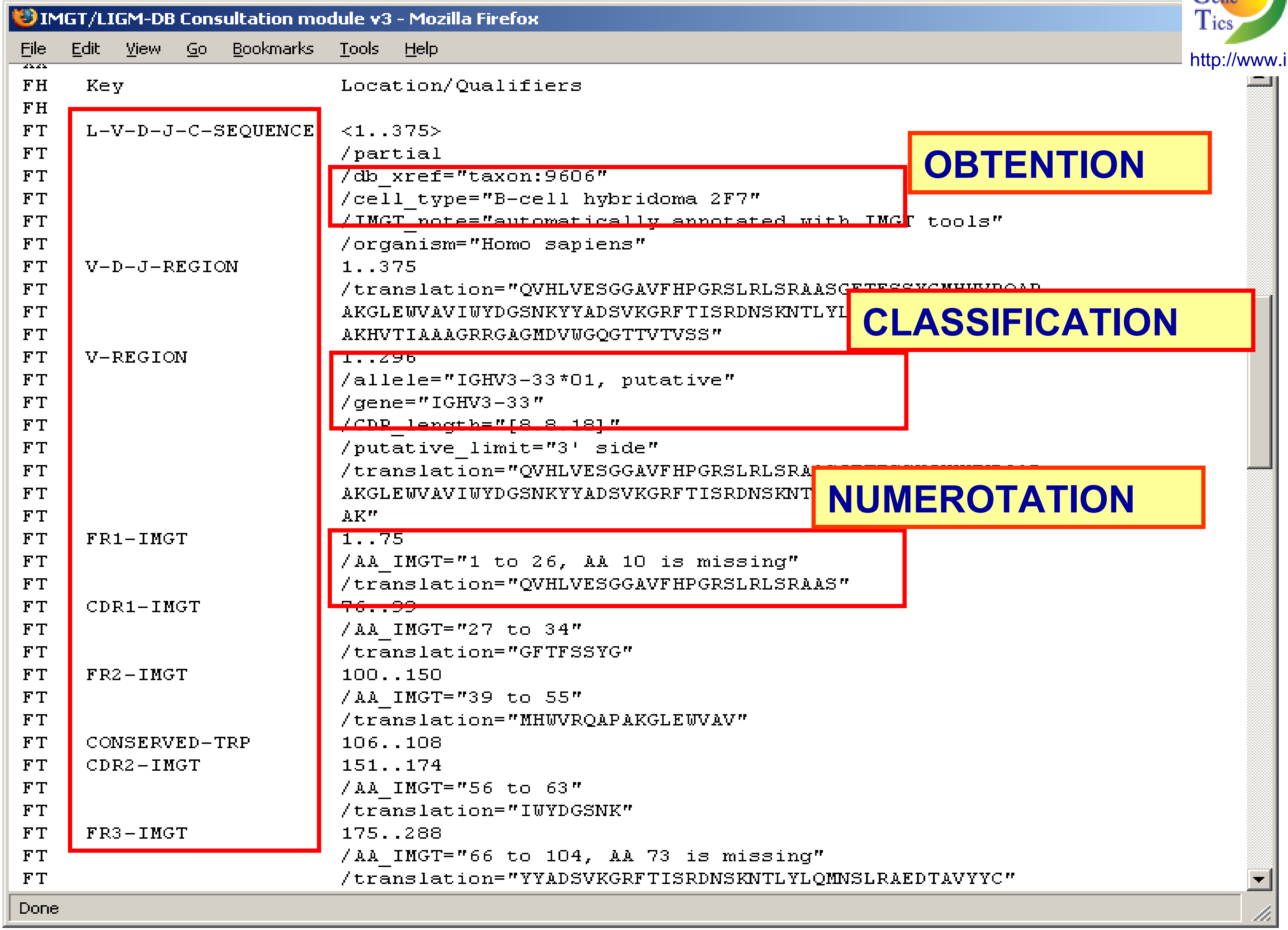
IDENTIFICATION



```
ID   HSCOMFRAA IMGT/LIGM annotation : automatic; RNA; HUM; 375 BP.
XX
AC   L39956;
XX
DT   28-SEP-2001 (Rel. 200139-5, arrived in LIGM-DB )
DT   13-NOV-2001 (Rel. 200146-2, Last updated, Version 2)
XX
DE   Homo sapiens monospecific anti-ssDNA antibody heavy chain variable region
DE   mRNA, complementarity determining regions  1-3 and framework regions 1-4.
DE   ;
DE   RNA; rearranged configuration; Ig-Heavy; regular; functionality
DE   productive; group IGHV; subgroup HV3; specificity anti-DNA single-stranded
DE   (ss) [human].
XX
KW   antigen receptor; immunoglobulin superfamily; Ig; Ig-Heavy; variable;
KW   diversity; joining; hybridoma; immunoglobulin.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-375
RX   MEDLINE; 96195158.
RA   Mitamura K., Suenaga R., Wilson K.B., Abdou N.I.;
RT   "V gene sequences of human anti-ssDNA antibodies secreted by lupus-derived
RT   CD5-negative B cell hybridomas";
RL   Clin. Immunol. Immunopathol. 78(2):152-160(1996).
XX
DR   EMBL; L39956.
XX
FH   Key                 Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE  <1..375>
FT                       /partial
```

The IMGT-ONTOLOGY concepts

http://www.imgt.org



```
FH   Key                 Location/Qualifiers
FH
FT   L-V-D-J-C-SEQUENCE  <1..375>
FT                       /partial
FT                       /db_xref="taxon:9606"
FT                       /cell_type="B-cell hybridoma 2F7"
FT                       /IMGT_note="automatically annotated with IMGT tools"
FT                       /organism="Homo sapiens"
FT   V-D-J-REGION        1..375
FT                       /translation="QVHLVESGGAVFHPGRSLRLSRAASG
FT                       AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDNSKNTLYL
FT                       AKHVTIAAAGRRGAGMDVWGQGTTVTVSS"
FT   V-REGION            1..296
FT                       /allele="IGHV3-33*01, putative"
FT                       /gene="IGHV3-33"
FT                       /CDR_length="[8.8.18]"
FT                       /putative_limit="3' side"
FT                       /translation="QVHLVESGGAVFHPGRSLRLSRA
FT                       AKGLEWVAVIWYDGSNKYYADSVKGRFTISRDNSKNT
FT                       AK"
FT   FR1-IMGT            1..75
FT                       /AA_IMGT="1 to 26, AA 10 is missing"
FT                       /translation="QVHLVESGGAVFHPGRSLRLSRAAS"
FT   CDR1-IMGT           76..99
FT                       /AA_IMGT="27 to 34"
FT                       /translation="GFTFSSYG"
FT   FR2-IMGT            100..150
FT                       /AA_IMGT="39 to 55"
FT                       /translation="MHWVRQAPAKGLEWVAV"
FT   CONSERVED-TRP       106..108
FT   CDR2-IMGT           151..174
FT                       /AA_IMGT="56 to 63"
FT                       /translation="IWYDGSNK"
FT   FR3-IMGT            175..288
FT                       /AA_IMGT="66 to 104, AA 73 is missing"
FT                       /translation="YYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC"
```

DESCRIPTION

OBTENTION

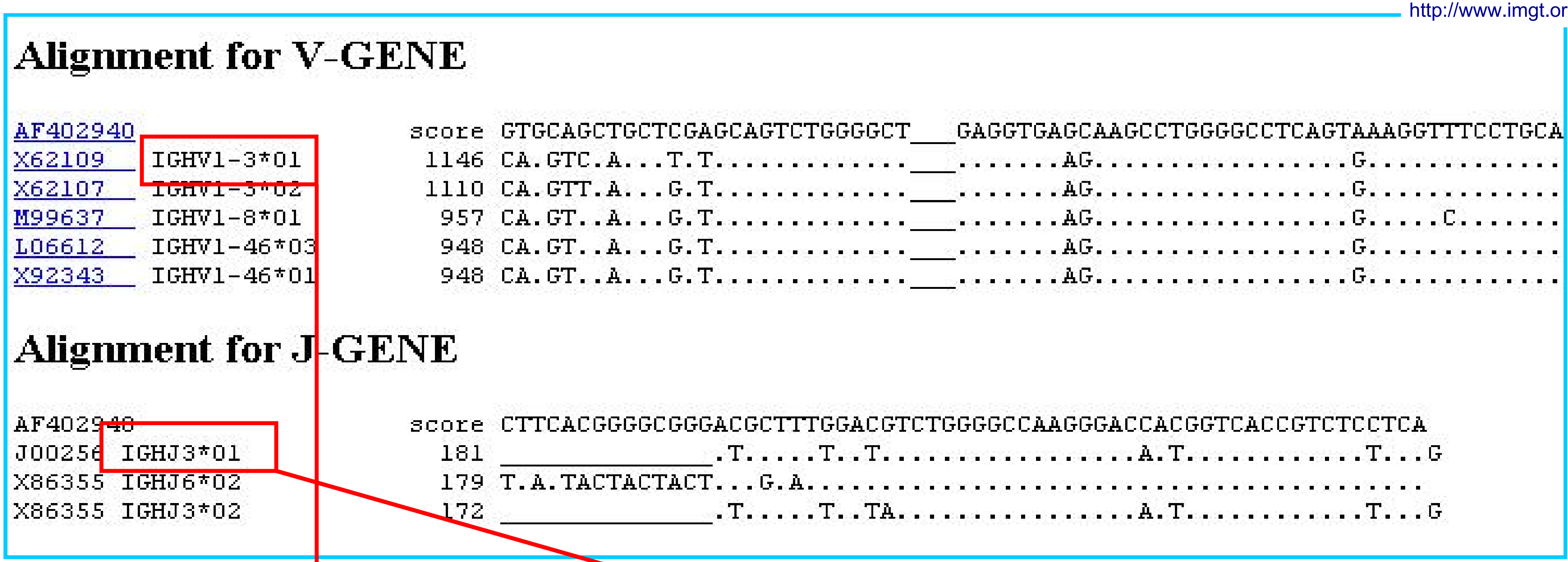
CLASSIFICATION

NUMEROTATION

IMGT/V-QUEST

IMGT/V-QUEST

http://www.imgt.org



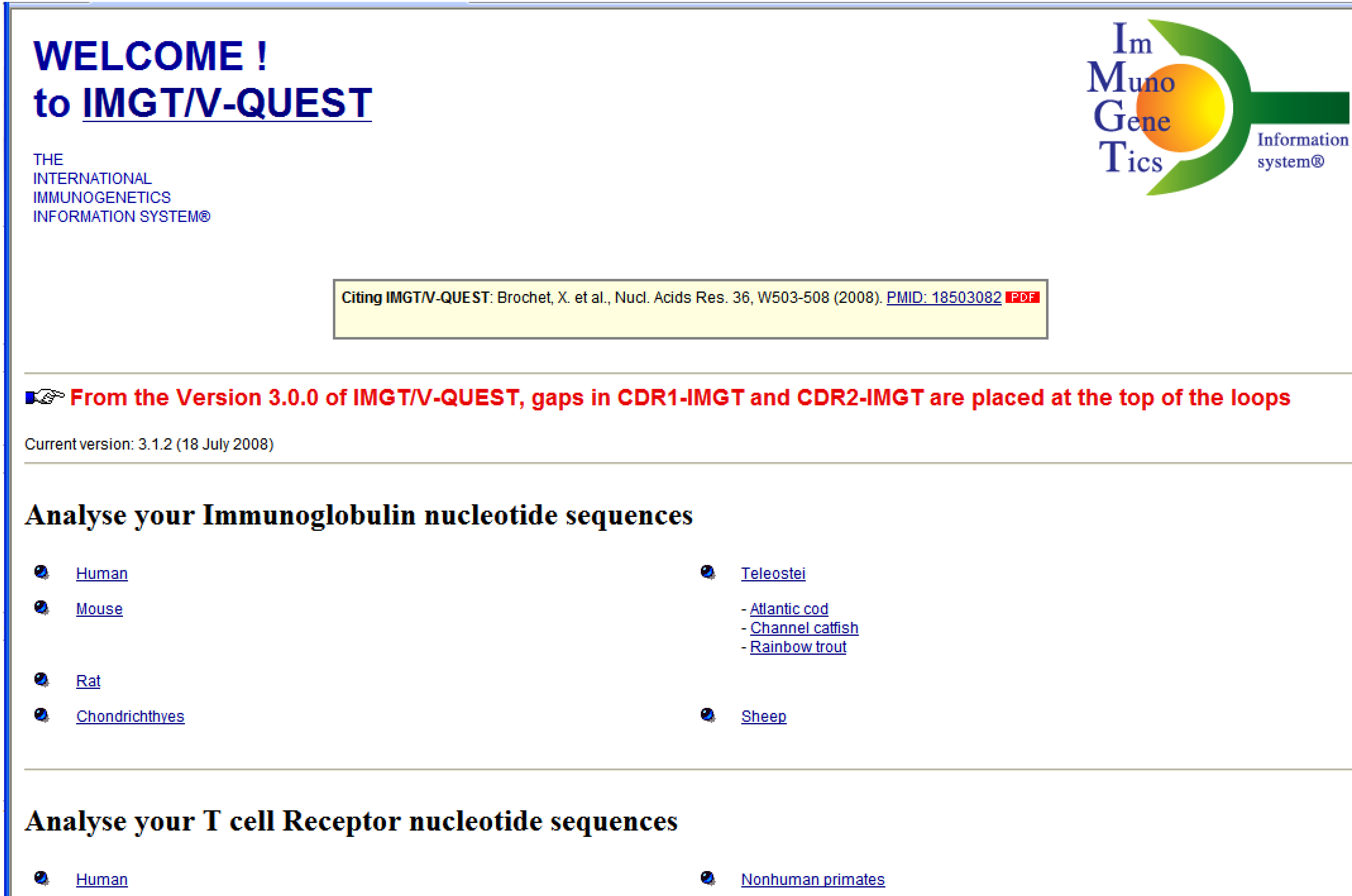
Alignment for V-GENE

```
AF402940                    score GTGCAGCTGCTCGAGCAGTCTGGGGCT___GAGGTGAGCAAGCCTGGGGCCTCAGTAAAGGTTTCCTGCA
X62109   IGHV1-3*01          1146 CA.GTC.A...T.T.............___.......AG.................G.............
X62107   IGHV1-3*02          1110 CA.GTT.A...G.T.............___.......AG.................G.............
M99637   IGHV1-8*01           957 CA.GT..A...G.T.............___.......AG.................G.....C.......
L06612   IGHV1-46*03          948 CA.GT..A...G.T.............___.......AG.................G.............
X92343   IGHV1-46*01          948 CA.GT..A...G.T.............___.......AG.................G.............
```

Alignment for J-GENE

```
AF402940                score CTTCACGGGGCGGACGCTTTGGACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCA
J00256 IGHJ3*01           181 ______________.T.....T..T................A.T............T...G
X86355 IGHJ6*02           179 T.A.TACTACTACT...G.A........................................
X86355 IGHJ3*02           172 ______________.T.....T..TA...............A.T............T...G
```

Giudicelli et al. Nucleic Acids Res. 36, W503-W508 (2008)



WELCOME !
to IMGT/V-QUEST

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

ImMunoGeneTics Information system®

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

☞ **From the Version 3.0.0 of IMGT/V-QUEST, gaps in CDR1-IMGT and CDR2-IMGT are placed at the top of the loops**

Current version: 3.1.2 (18 July 2008)

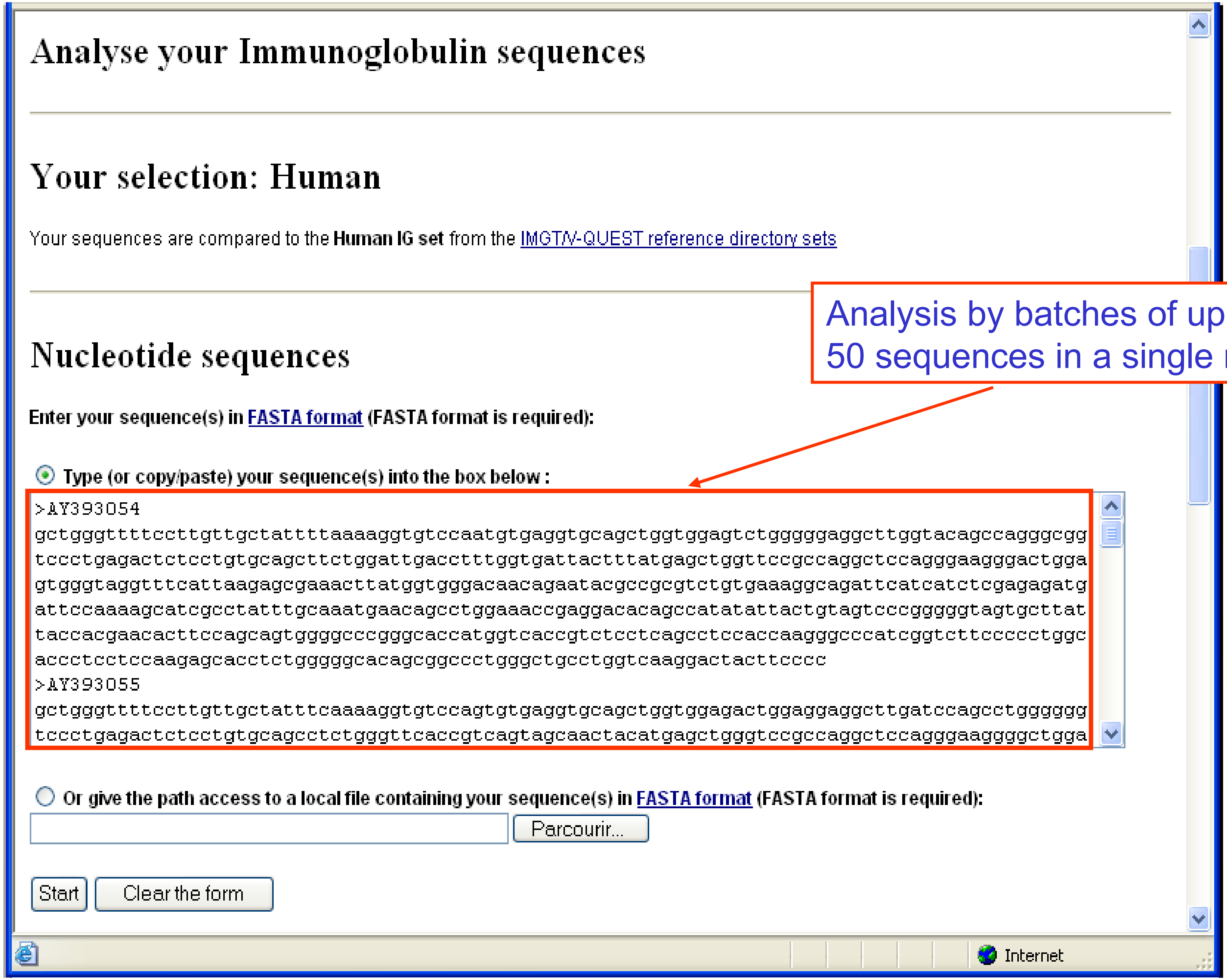
Analyse your Immunoglobulin nucleotide sequences

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

Analyse your T cell Receptor nucleotide sequences

- Human
- Nonhuman primates

http://www.imgt.org



Analyse your Immunoglobulin sequences

Your selection: Human

Your sequences are compared to the **Human IG set** from the IMGT/V-QUEST reference directory sets

Nucleotide sequences

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

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

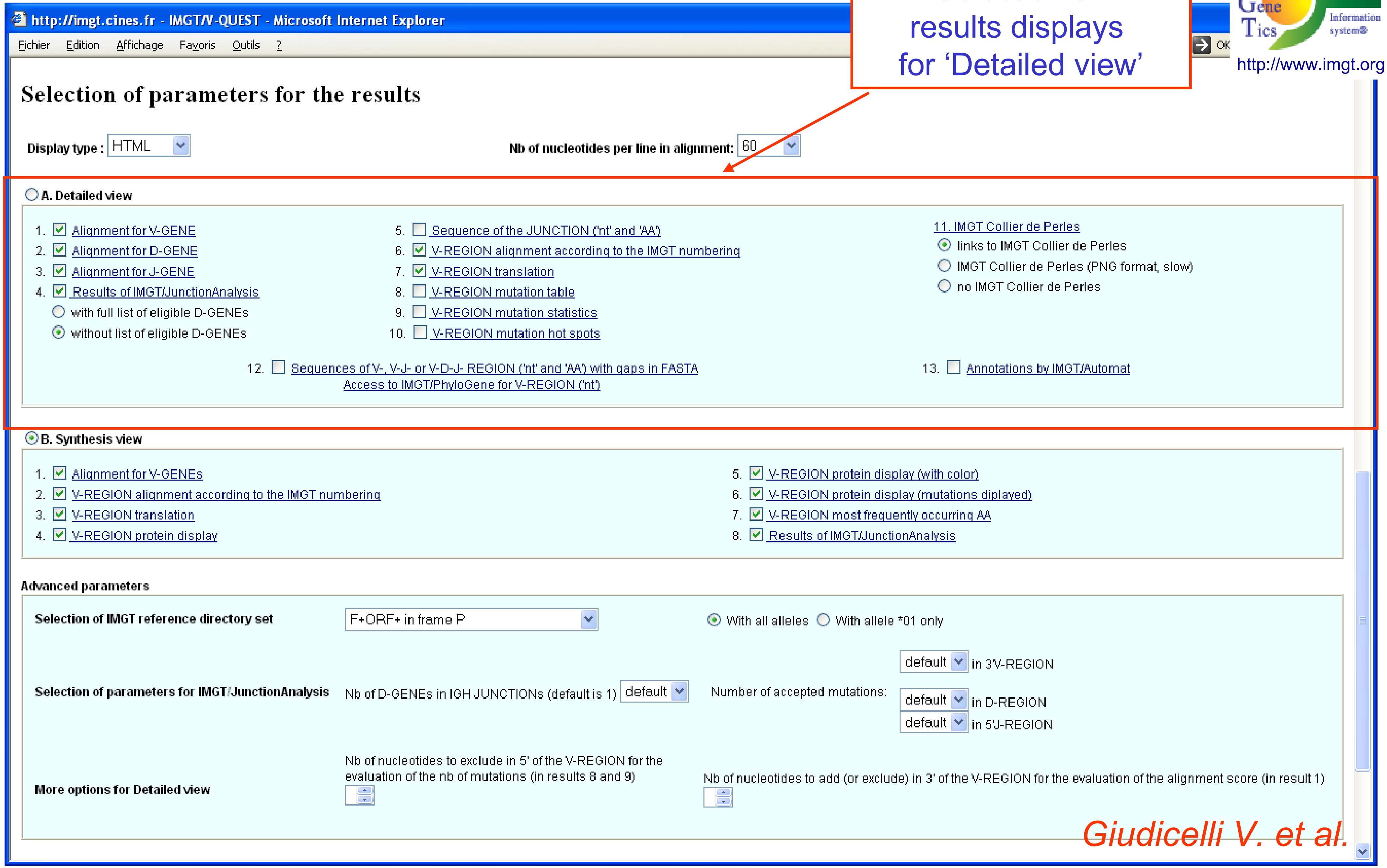
```
>AY393054
gctgggttttccttgttgctattttaaaaggtgtccaatgtgaggtgcagctggtggagtctgggggaggcttggtacagccagggcgg
tccctgagactctcctgtgcagcttctggattgacctttggtgattactttatgagctggttccgccaggctccagggaagggactgga
gtgggtaggtttcattaagagcgaaacttatggtgggacaacagaatacgccgcgtctgtgaaaggcagattcatcatctcgagagatg
attccaaaagcatcgcctatttgcaaatgaacagcctggaaaccgaggacacagccatatattactgtagtcccgggggtagtgcttat
taccacgaacacttccagcagtggggcccgggcaccatggtcaccgtctcctcagcctccaccaagggcccatcggtcttccccctggc
accctcctccaagagcacctctgggggcacagcggccctgggctgcctggtcaaggactacttcccc
>AY393055
gctgggttttccttgttgctatttcaaaaggtgtccagtgtgaggtgcagctggtggagactggaggaggcttgatccagcctgggggg
tccctgagactctcctgtgcagcctctgggttcaccgtcagtagcaactacatgagctgggtccgccaggctccagggaaggggctgga
```

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

Analysis by batches of up to 50 sequences in a single run

Selection of results displays for ‘Detailed view’

http://www.imgt.org



Selection of parameters for the results

Display type : HTML

Nb of nucleotides per line in alignment: 60

○ A. Detailed view

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-GENEs
   - ● without list of eligible D-GENEs
5. ☐ Sequence of the JUNCTION ('nt' and 'AA')
6. ☑ V-REGION alignment according to the IMGT numbering
7. ☑ V-REGION translation
8. ☐ V-REGION mutation table
9. ☐ V-REGION mutation statistics
10. ☐ V-REGION mutation hot spots
11. IMGT Collier de Perles
   - ● links to IMGT Collier de Perles
   - ○ IMGT Collier de Perles (PNG format, slow)
   - ○ no IMGT Collier de Perles
12. ☐ 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')
13. ☐ Annotations by IMGT/Automat

● B. Synthesis view

1. ☑ Alignment for V-GENEs
2. ☑ V-REGION alignment according to the IMGT numbering
3. ☑ V-REGION translation
4. ☑ V-REGION protein display
5. ☑ V-REGION protein display (with color)
6. ☑ V-REGION protein display (mutations diplayed)
7. ☑ V-REGION most frequently occurring AA
8. ☑ Results of IMGT/JunctionAnalysis

Advanced parameters

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

Selection of parameters for IMGT/JunctionAnalysis: Nb of D-GENEs in IGH JUNCTIONs (default is 1) default

Number of accepted mutations: default in 3'V-REGION; default in D-REGION; default in 5'J-REGION

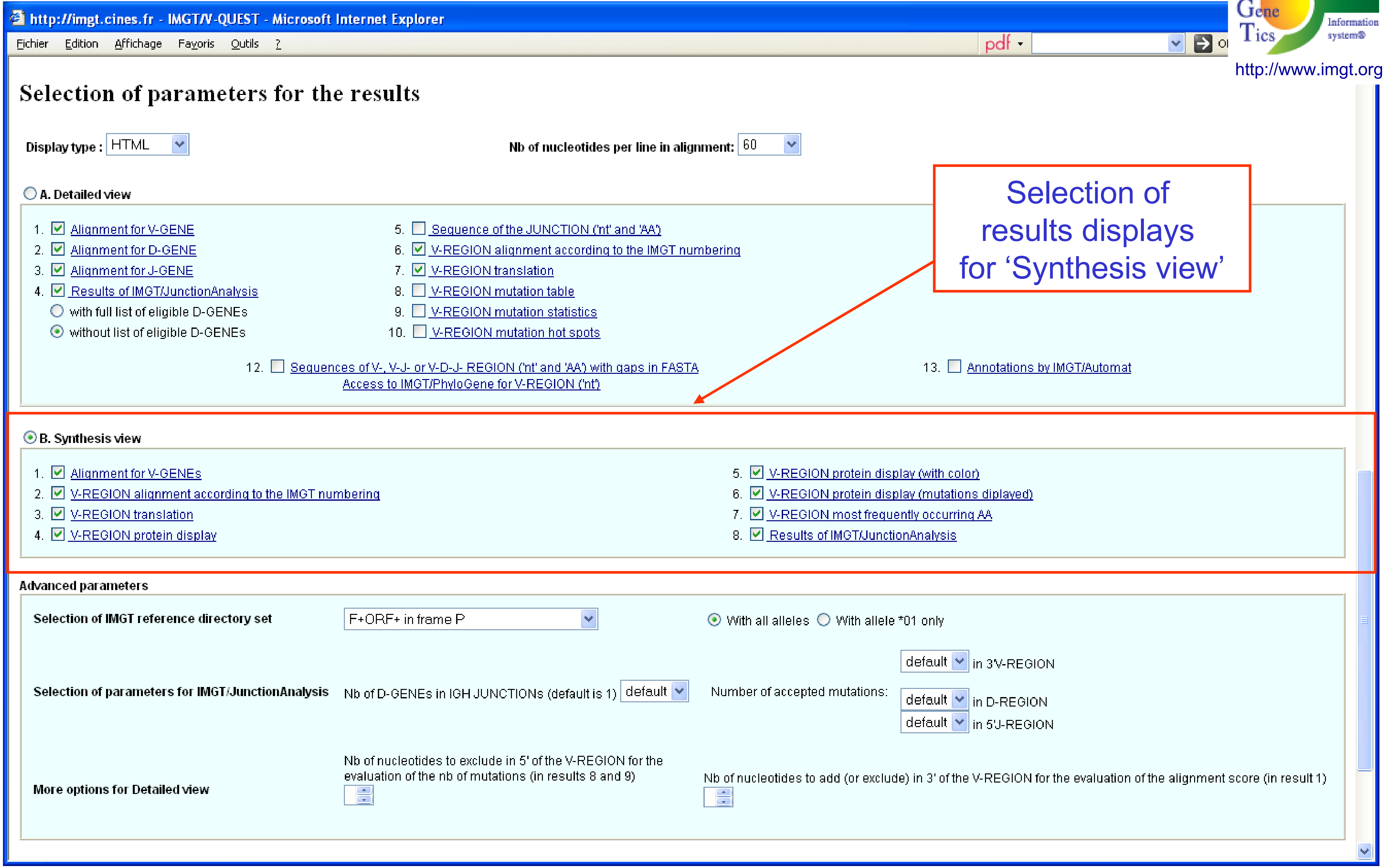
More options for Detailed view:

Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 8 and 9)

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

Giudicelli V. et al.

http://www.imgt.org



Selection of parameters for the results

Display type : HTML

Nb of nucleotides per line in alignment: 60

○ **A. Detailed view**

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-GENEs
   - ● without list of eligible D-GENEs
5. ☐ Sequence of the JUNCTION ('nt' and 'AA')
6. ☑ V-REGION alignment according to the IMGT numbering
7. ☑ V-REGION translation
8. ☐ V-REGION mutation table
9. ☐ V-REGION mutation statistics
10. ☐ V-REGION mutation hot spots

12. ☐ 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')

13. ☐ Annotations by IMGT/Automat

● **B. Synthesis view**

1. ☑ Alignment for V-GENEs
2. ☑ V-REGION alignment according to the IMGT numbering
3. ☑ V-REGION translation
4. ☑ V-REGION protein display
5. ☑ V-REGION protein display (with color)
6. ☑ V-REGION protein display (mutations diplayed)
7. ☑ V-REGION most frequently occurring AA
8. ☑ Results of IMGT/JunctionAnalysis

Advanced parameters

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

Selection of parameters for IMGT/JunctionAnalysis Nb of D-GENEs in IGH JUNCTIONs (default is 1) default

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

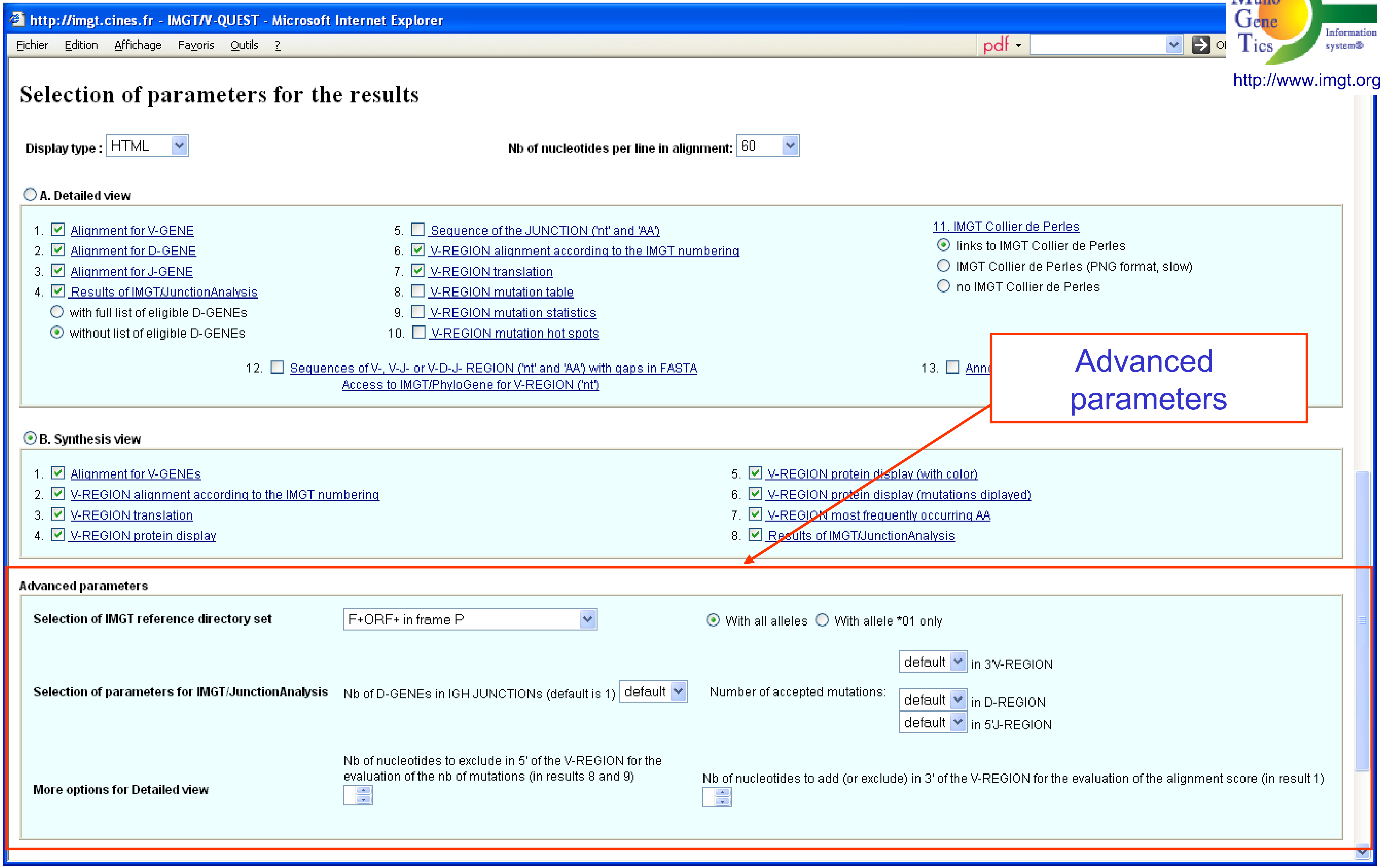
More options for Detailed view

Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 8 and 9)

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

Selection of results displays for 'Synthesis view'

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



Selection of parameters for the results

Display type : HTML

Nb of nucleotides per line in alignment: 60

○ **A. Detailed view**

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-GENEs
   - ● without list of eligible D-GENEs
5. ☐ Sequence of the JUNCTION ('nt' and 'AA')
6. ☑ V-REGION alignment according to the IMGT numbering
7. ☑ V-REGION translation
8. ☐ V-REGION mutation table
9. ☐ V-REGION mutation statistics
10. ☐ V-REGION mutation hot spots
11. IMGT Collier de Perles
   - ● links to IMGT Collier de Perles
   - ○ IMGT Collier de Perles (PNG format, slow)
   - ○ no IMGT Collier de Perles
12. ☐ 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')
13. ☐ Ann

● **B. Synthesis view**

1. ☑ Alignment for V-GENEs
2. ☑ V-REGION alignment according to the IMGT numbering
3. ☑ V-REGION translation
4. ☑ V-REGION protein display
5. ☑ V-REGION protein display (with color)
6. ☑ V-REGION protein display (mutations diplayed)
7. ☑ V-REGION most frequently occurring AA
8. ☑ Results of IMGT/JunctionAnalysis

Advanced parameters

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

Selection of parameters for IMGT/JunctionAnalysis Nb of D-GENEs in IGH JUNCTIONs (default is 1) default

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

More options for Detailed view

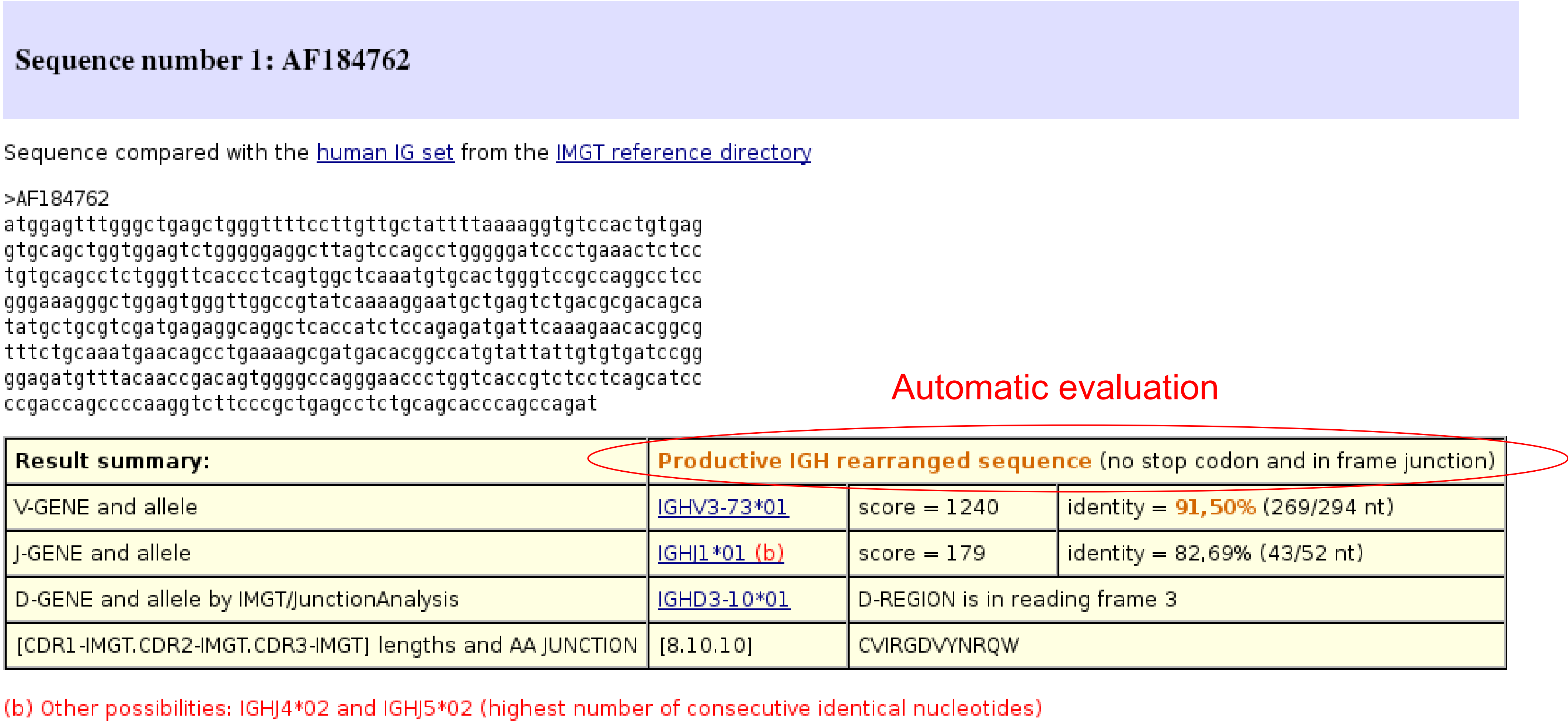
Nb of nucleotides to exclude in 5' of the V-REGION for the evaluation of the nb of mutations (in results 8 and 9)

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

Advanced parameters

A. ‘Detailed view’: Result summary

http://www.imgt.org



Sequence number 1: AF184762

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

```
>AF184762
atggagtttgggctgagctgggttttccttgttgctattttaaaaggtgtccactgtgag
gtgcagctggtggagtctgggggaggcttagtccagcctggggggatccctgaaactctcc
tgtgcagcctctgggttcaccctcagtggctcaaatgtgcactgggtccgccaggcctcc
gggaaagggctggagtgggttggccgtatcaaaaggaatgctgagtctgacgcgacagca
tatgctgcgtcgatgagaggcaggctcaccatctccagagatgattcaaagaacacggcg
tttctgcaaatgaacagcctgaaaagcgatgacacggccatgtattattgtgtgatccgg
ggagatgtttacaaccgacagtggggccagggaaccctggtcaccgtctcctcagcatcc
ccgaccagccccaaggtcttcccgctgagcctctgcagcacccagccagat
```

Automatic evaluation

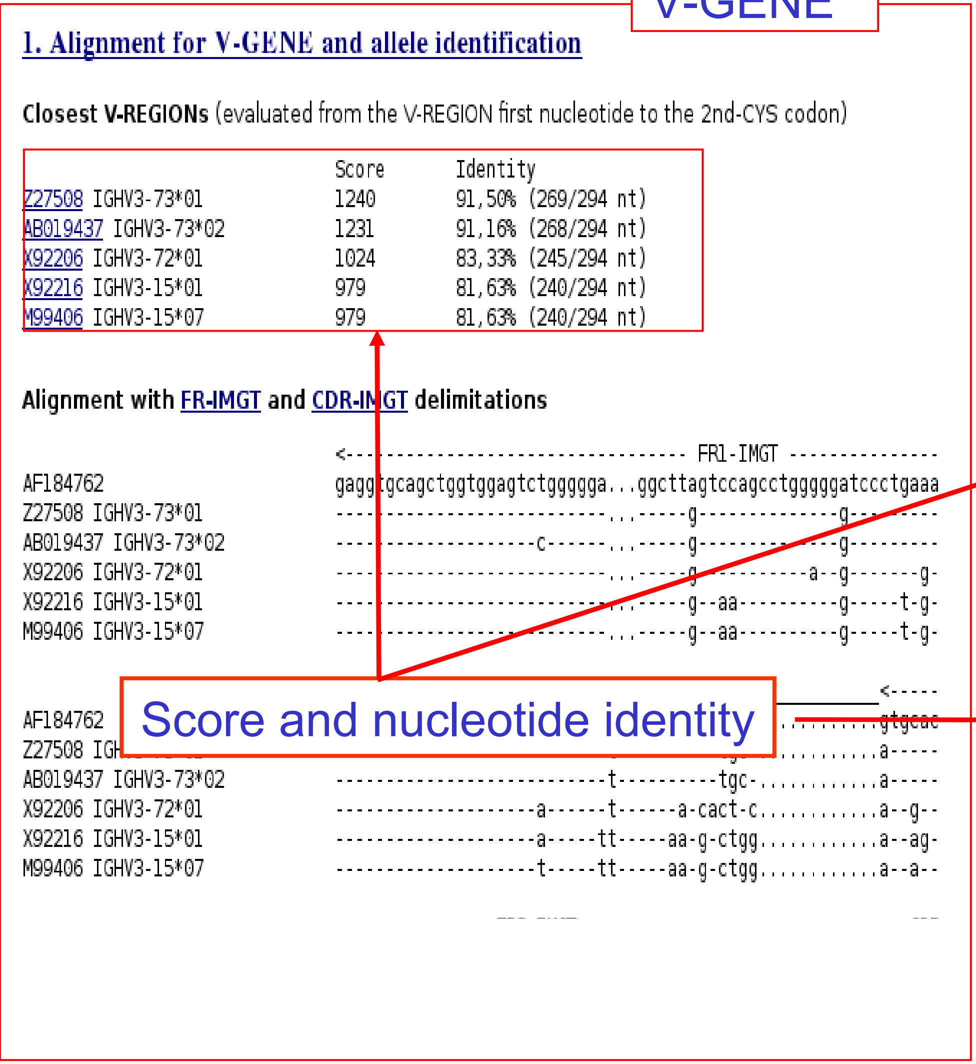
Result summary:	**Productive IGH rearranged sequence** (no stop codon and in frame junction)		
V-GENE and allele	IGHV3-73*01	score = 1240	identity = **91,50%** (269/294 nt)
J-GENE and allele	IGHJ1*01 (b)	score = 179	identity = 82,69% (43/52 nt)
D-GENE and allele by IMGT/JunctionAnalysis	IGHD3-10*01	D-REGION is in reading frame 3	
[CDR1-IMGT.CDR2-IMGT.CDR3-IMGT] lengths and AA JUNCTION	[8.10.10]	CVIRGDVYNRQW	

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

1-3. ‘Detailed view’: Alignments for V-, D- and J-GENE

http://www.imgt.org

V-GENE



1. Alignment for V-GENE and allele identification

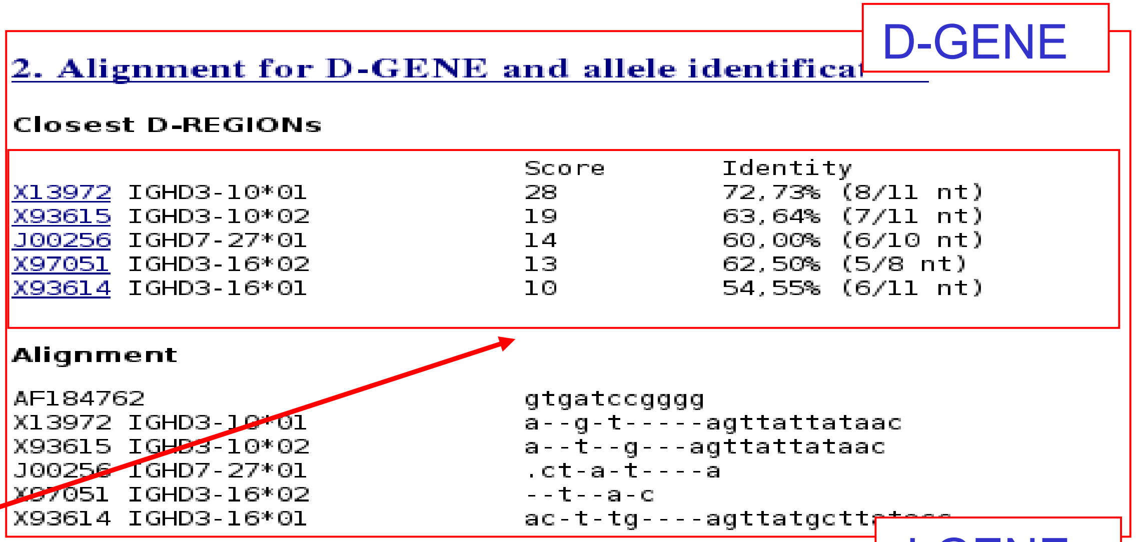
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

```
                              Score       Identity
Z27508 IGHV3-73*01            1240        91,50% (269/294 nt)
AB019437 IGHV3-73*02          1231        91,16% (268/294 nt)
X92206 IGHV3-72*01            1024        83,33% (245/294 nt)
X92216 IGHV3-15*01            979         81,63% (240/294 nt)
M99406 IGHV3-15*07            979         81,63% (240/294 nt)
```

Alignment with FR-IMGT and CDR-IMGT delimitations

```
                              <-------------------------------- FR1-IMGT --------------
AF184762                      gaggtgcagctggtggagtctgggggа...ggcttagtccagcctggggggatccctgaaa
Z27508 IGHV3-73*01            --------------------------...-----g-------------g---------
AB019437 IGHV3-73*02          -------------------c------...-----g-------------g---------
X92206 IGHV3-72*01            --------------------------...-----g----------a--g-------g-
X92216 IGHV3-15*01            --------------------------...-----g--aa---------g-----t-g-
M99406 IGHV3-15*07            --------------------------...-----g--aa---------g-----t-g-

                                                                               <-----
AF184762                                                              ..........gtgcac
Z27508 IGH                                                            ..........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 identificat

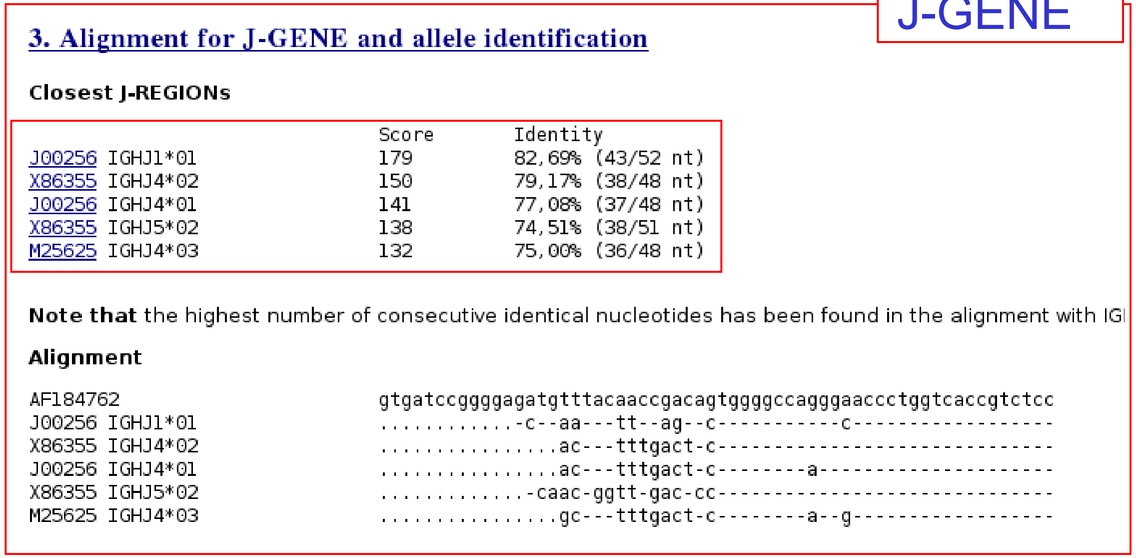
Closest D-REGIONs

```
                                          Score      Identity
X13972 IGHD3-10*01                        28         72,73% (8/11 nt)
X93615 IGHD3-10*02                        19         63,64% (7/11 nt)
J00256 IGHD7-27*01                        14         60,00% (6/10 nt)
X97051 IGHD3-16*02                        13         62,50% (5/8 nt)
X93614 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
```

J-GENE



3. Alignment for J-GENE and allele identification

Closest J-REGIONs

```
                              Score      Identity
J00256 IGHJ1*01               179        82,69% (43/52 nt)
X86355 IGHJ4*02               150        79,17% (38/48 nt)
J00256 IGHJ4*01               141        77,08% (37/48 nt)
X86355 IGHJ5*02               138        74,51% (38/51 nt)
M25625 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                      gtgatccggggagatgtttacaaccgacagtggggccagggaaccctggtcaccgtctcc
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-ggtt-gac-cc-----------------------------
M25625 IGHJ4*03                ................gc---tttgact-c--------a--g--------------------
```

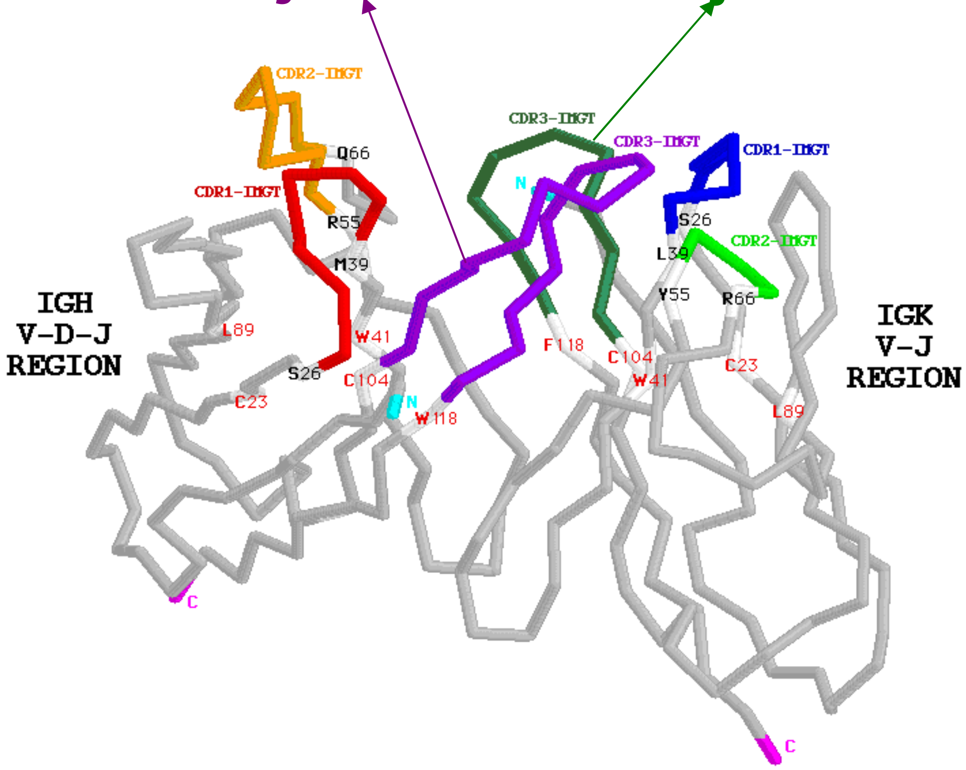
Score and nucleotide identity

Junctions of the V-DOMAINs

http://www.imgt.org

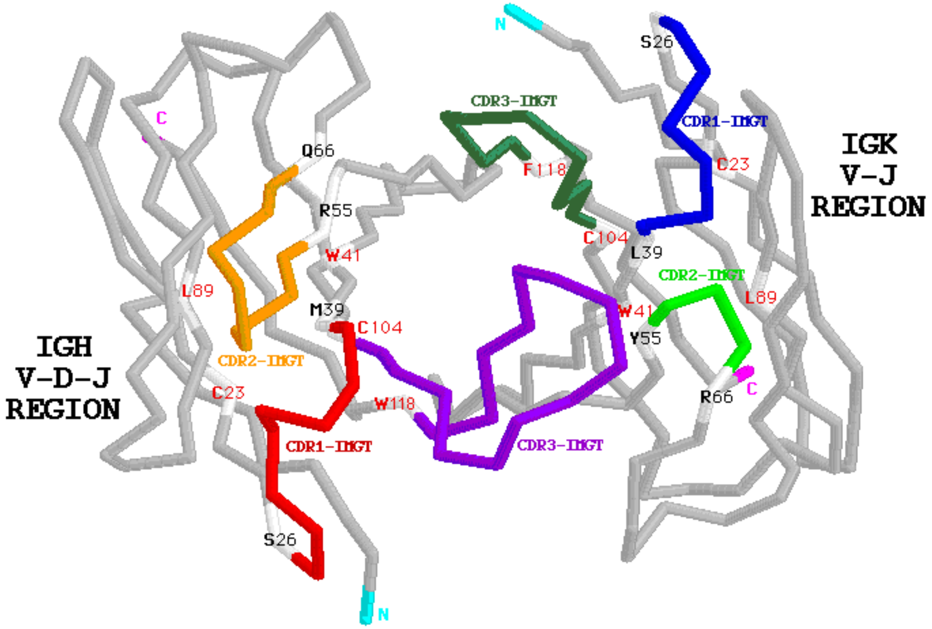
VH **V-KAPPA**

V-D-J junction **V-J junction**



Side view

VH **V-KAPPA**



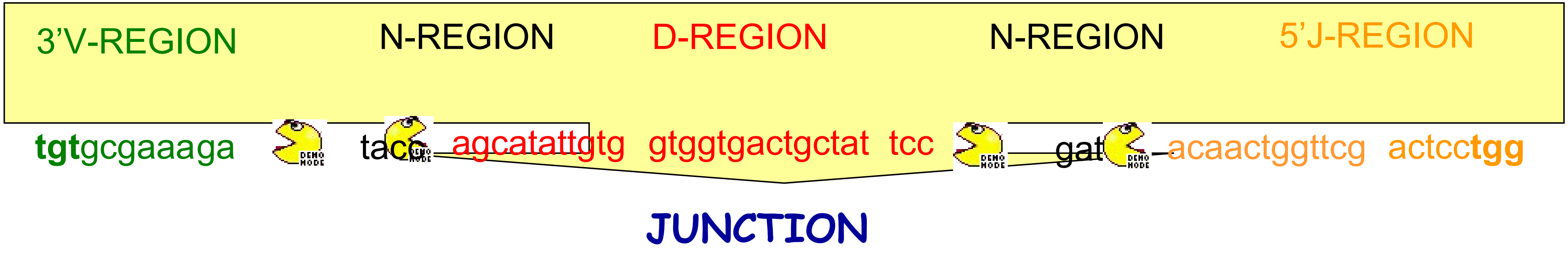
View from above

Mouse(Mus musculus) E5.2Fv

CDR3-IMGT= Complementarity determining region (105-117)
V-J junction (104-118)
V-D-J junction (104-118)

Generation of the JUNCTION diversity

http://www.imgt.org

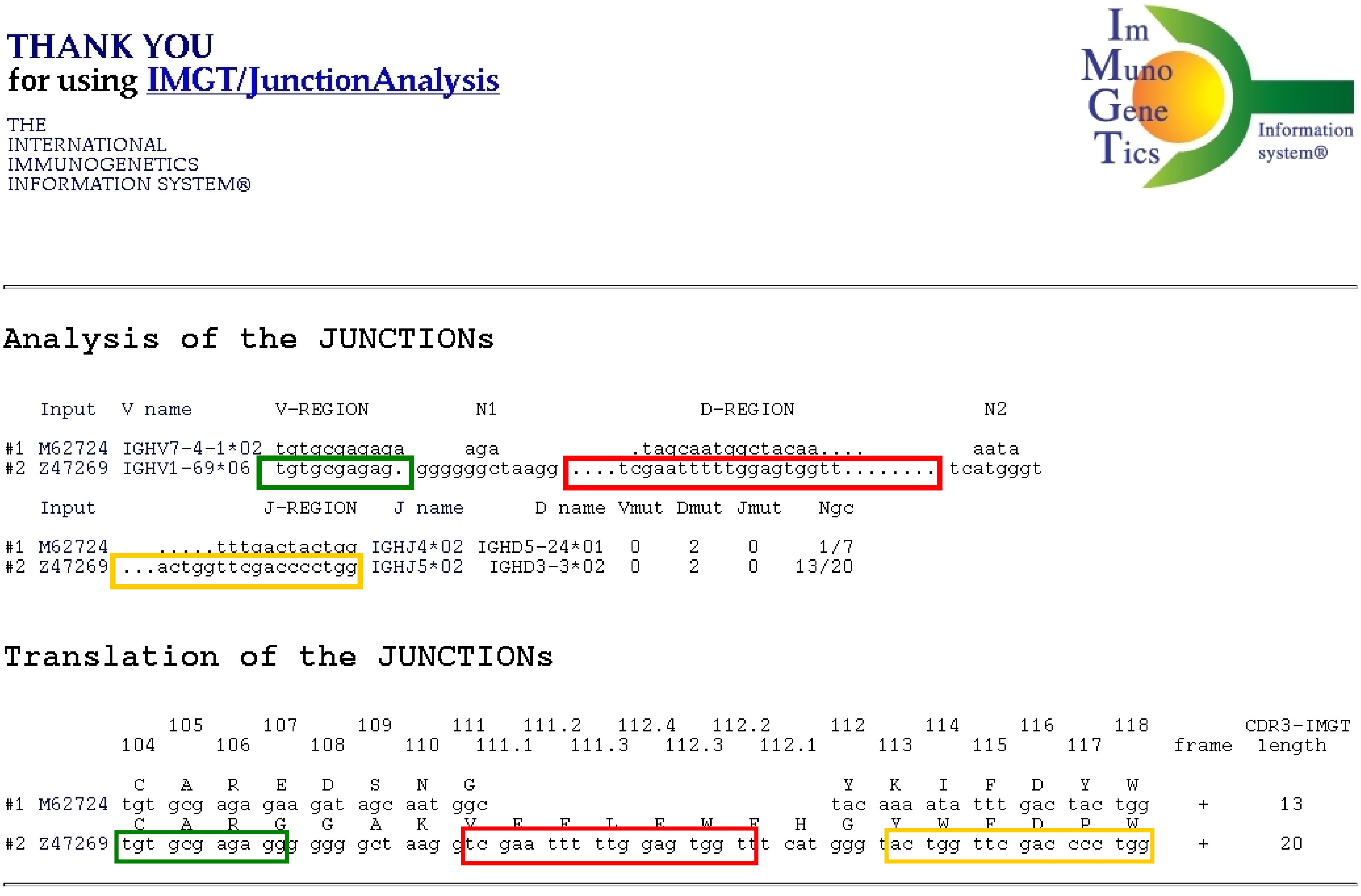


3’V-REGION N-REGION D-REGION N-REGION 5’J-REGION

tgtgcgaaaga tacc agcatattgtg gtggtgactgctat tcc gat acaactggttcg actcc**tgg**

JUNCTION

C	**A**	**P**	**Y**	**R**	**G**	**D**	**T**	**Y**	**D**	**Y**	**S**	**W**
tgt	gcg	cca	tac	cgg	ggt	gac	act	tat	gat	t**ac**	**tcc**	**tgg**



THANK YOU
for using IMGT/JunctionAnalysis

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Analysis of the JUNCTIONs

```
   Input   V name        V-REGION         N1                D-REGION              N2

#1 M62724 IGHV7-4-1*02 tgtgcgagaga      aga          .tagcaatggctacaa....         aata
#2 Z47269 IGHV1-69*06   tgtgcgagag. gggggctaagg ....tcgaatttttggagtggtt........ tcatgggt
   Input               J-REGION    J name      D name Vmut Dmut Jmut    Ngc

#1 M62724      .....tttgactactgg IGHJ4*02 IGHD5-24*01  0    2    0     1/7
#2 Z47269   ...actggttcgaccctgg IGHJ5*02  IGHD3-3*02  0    2    0   13/20
```

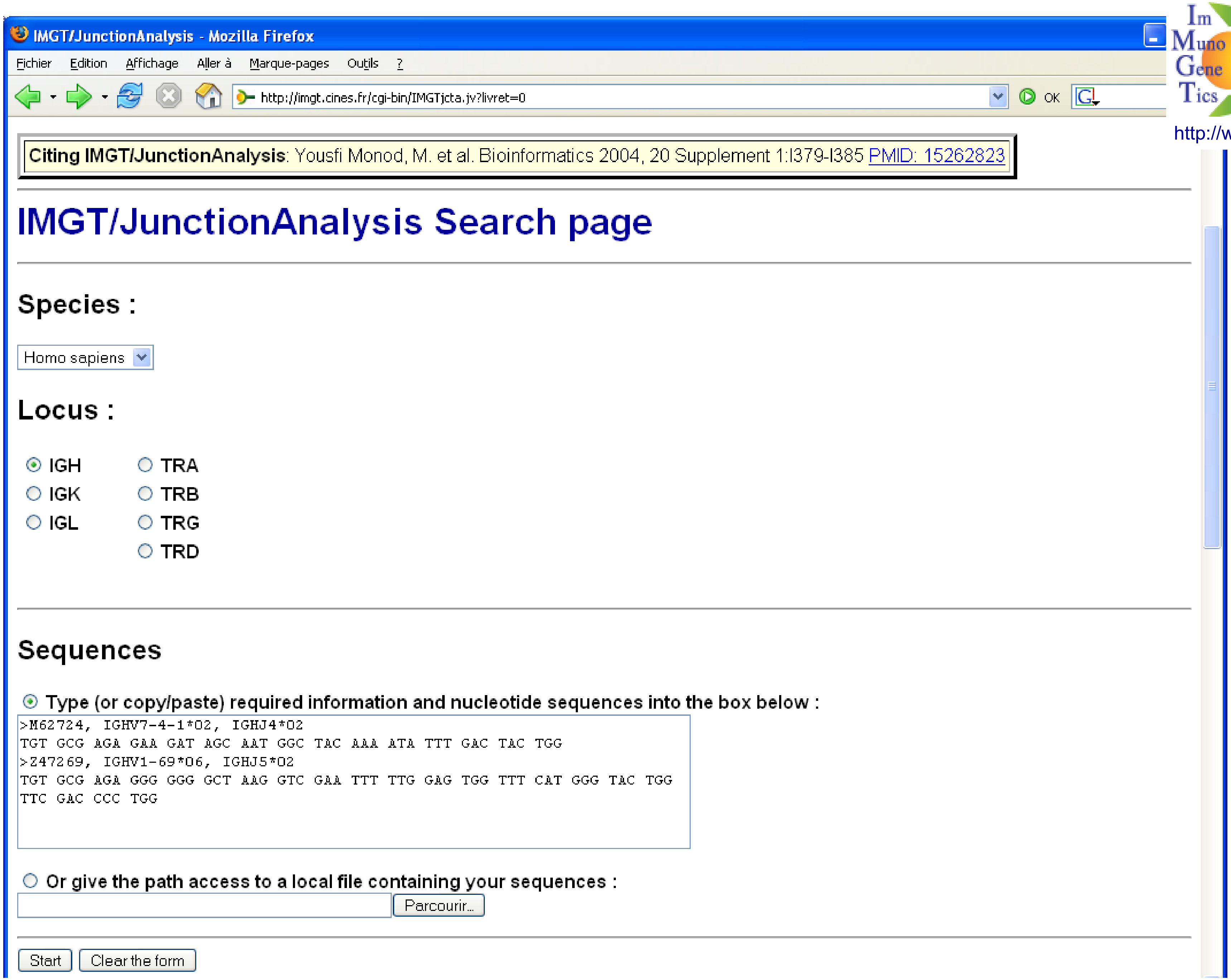
Translation of the JUNCTIONs

```
              105     107     109     111   111.2   112.4   112.2     112     114     116     118          CDR3-IMGT
          104     106     108     110   111.1   111.3   112.3   112.1     113     115     117      frame   length

           C   A   R   E   D   S   N   G                                   Y   K   I   F   D   Y   W
#1 M62724 tgt gcg aga gaa gat agc aat ggc                                 tac aaa ata ttt gac tac tgg    +      13
           C   A   R   G   G   A   K   V   E   F   L   E   W   F   H   G   Y   W   F   D   P   W
#2 Z47269 tgt gcg aga ggg ggg gct aag gtc gaa ttt ttg gag tgg ttt cat ggg tac tgg ttc gac ccc tgg    +      20
```

-> IMGT/JunctionAnalysis Search page
-> IMGT/JunctionAnalysis Documentation

Yousfi Monod et al. Bioinformatics, 20, I379-I385 (2004)

http://www.imgt.org



Citing IMGT/JunctionAnalysis: Yousfi Monod, M. et al. Bioinformatics 2004, 20 Supplement 1:I379-I385 PMID: 15262823

IMGT/JunctionAnalysis Search page

Species :

Homo sapiens

Locus :

● IGH	○ TRA
○ IGK	○ TRB
○ IGL	○ TRG
	○ TRD

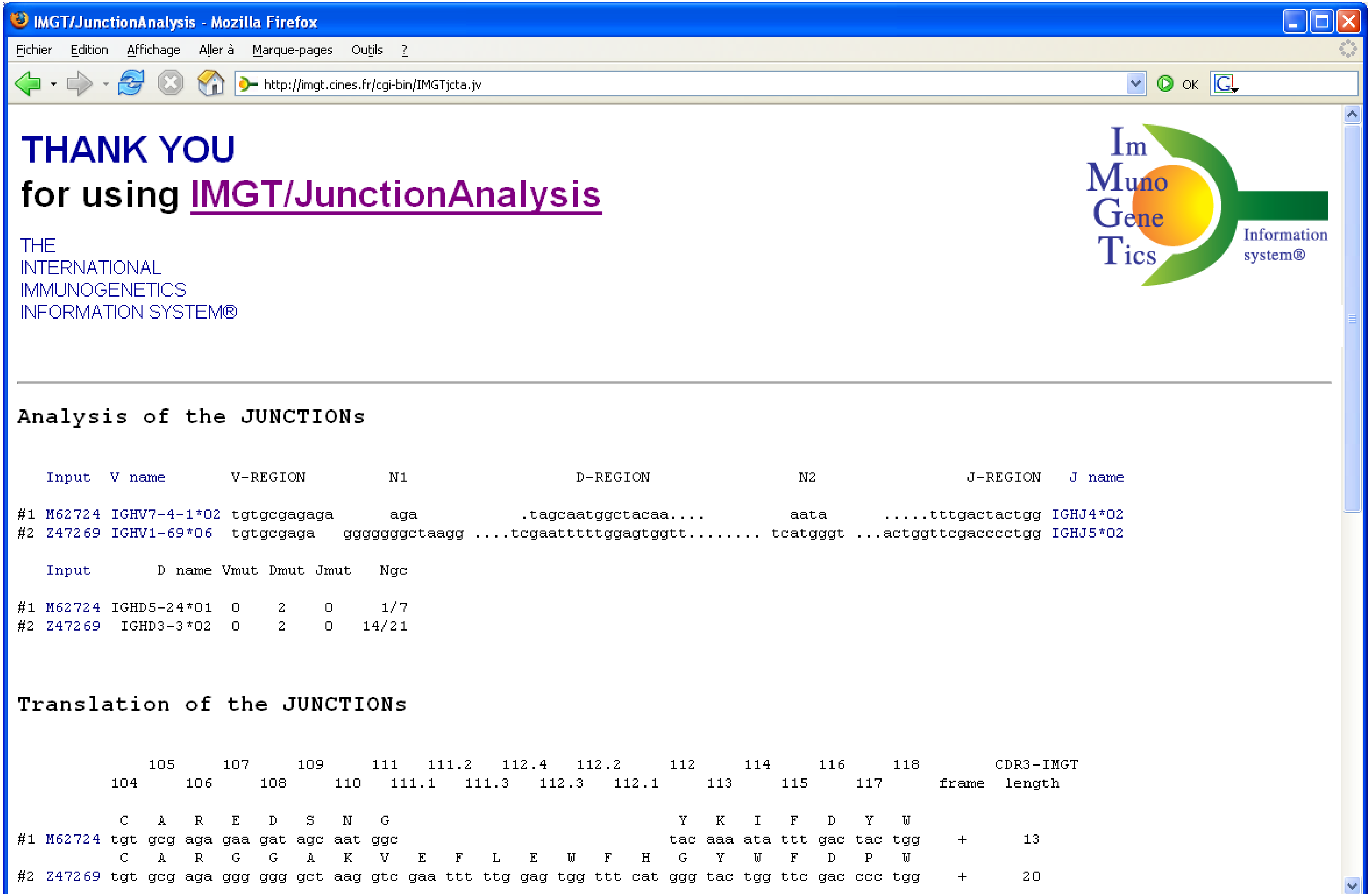
Sequences

● Type (or copy/paste) required information and nucleotide sequences into the box below :

```
>M62724, IGHV7-4-1*02, IGHJ4*02
TGT GCG AGA GAA GAT AGC AAT GGC TAC AAA ATA TTT GAC TAC TGG
>Z47269, IGHV1-69*06, IGHJ5*02
TGT GCG AGA GGG GGG GCT AAG GTC GAA TTT TTG GAG TGG TTT CAT GGG TAC TGG
TTC GAC CCC TGG
```

○ Or give the path access to a local file containing your sequences :

Yousfi Monod et al. Bioinformatics, 20, I379-I385 (2004)



THANK YOU
for using IMGT/JunctionAnalysis

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

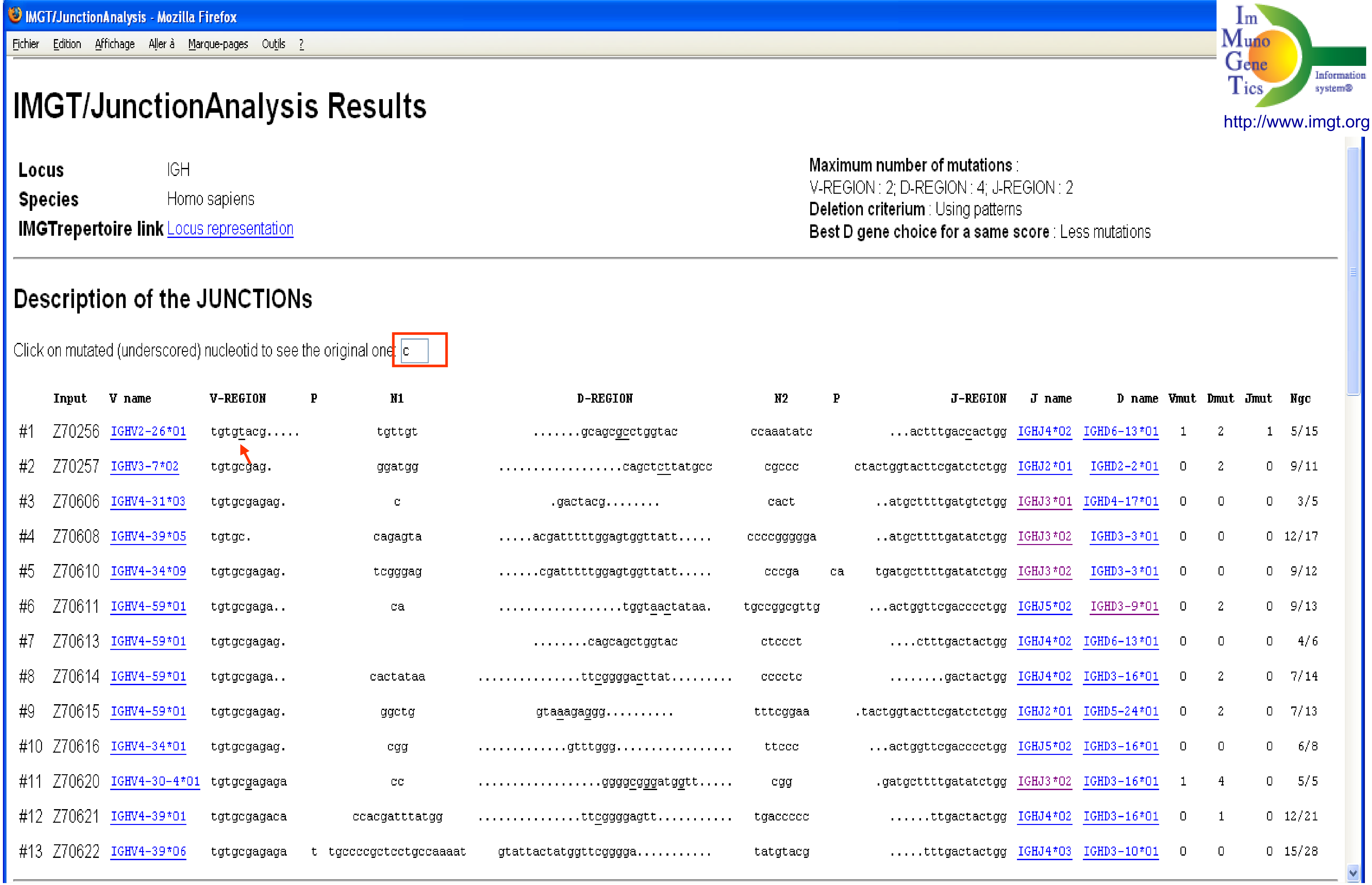
Analysis of the JUNCTIONs

```
   Input  V name        V-REGION         N1                  D-REGION               N2                J-REGION   J name

#1 M62724 IGHV7-4-1*02 tgtgcgagaga      aga           .tagcaatggctacaa....        aata      .....tttgactactgg IGHJ4*02
#2 Z47269 IGHV1-69*06  tgtgcgaga   gggggggctaagg ....tcgaattttggagtggtt........ tcatgggt ...actggttcgacccctgg IGHJ5*02

   Input       D name Vmut Dmut Jmut   Ngc

#1 M62724 IGHD5-24*01  0    2    0     1/7
#2 Z47269  IGHD3-3*02  0    2    0   14/21
```

Translation of the JUNCTIONs

```
              105     107     109     111   111.2   112.4   112.2     112     114     116     118        CDR3-IMGT
          104     106     108     110   111.1   111.3   112.3   112.1     113     115     117      frame  length

           C   A   R   E   D   S   N   G                               Y   K   I   F   D   Y   W
#1 M62724 tgt gcg aga gaa gat agc aat ggc                             tac aaa ata ttt gac tac tgg    +      13
           C   A   R   G   G   A   K   V   E   F   L   E   W   F   H   G   Y   W   F   D   P   W
#2 Z47269 tgt gcg aga ggg ggg gct aag gtc gaa ttt ttg gag tgg ttt cat ggg tac tgg ttc gac ccc tgg    +      20
```

Yousfi Monod et al. Bioinformatics, 20, I379-I385 (2004)



http://www.imgt.org

IMGT/JunctionAnalysis Results

Locus IGH
Species Homo sapiens
IMGTrepertoire link Locus representation

Maximum number of mutations :
V-REGION : 2; D-REGION : 4; J-REGION : 2
Deletion criterium : Using patterns
Best D gene choice for a same score : Less mutations

Description of the JUNCTIONs

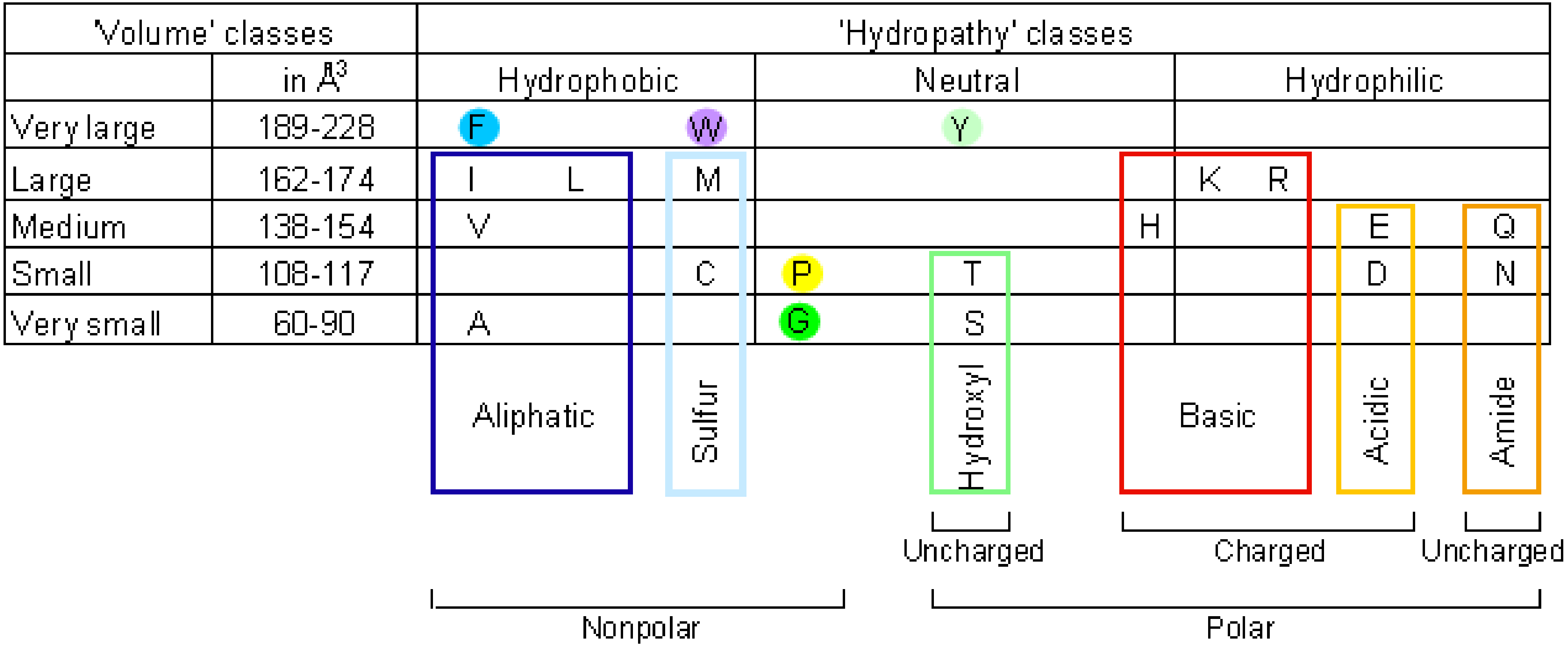
Click on mutated (underscored) nucleotid to see the original one: c

	Input	V name	V-REGION	P	N1	D-REGION	N2	P	J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
#1	Z70256	IGHV2-26*01	tgtgtacg.....		tgttgt	.......gcagcgcctggtac	ccaaatatc		...actttgacccactgg	IGHJ4*02	IGHD6-13*01	1	2	1	5/15
#2	Z70257	IGHV3-7*02	tgtgcgag.		ggatgg	..................cagctcttatgcc	cgccc		ctactggtacttcgatctctgg	IGHJ2*01	IGHD2-2*01	0	2	0	9/11
#3	Z70606	IGHV4-31*03	tgtgcgagag.		c	.gactacg........	cact		..atgcttttgatgtctgg	IGHJ3*01	IGHD4-17*01	0	0	0	3/5
#4	Z70608	IGHV4-39*05	tgtgc.		cagagta	.....acgattttggagtggttatt.....	ccccgggga		..atgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	12/17
#5	Z70610	IGHV4-34*09	tgtgcgagag.		tcgggag	......cgattttggagtggttatt.....	cccga	ca	tgatgcttttgatatctgg	IGHJ3*02	IGHD3-3*01	0	0	0	9/12
#6	Z70611	IGHV4-59*01	tgtgcgaga..		ca	..................tggtaactataa.	tgccggcgttg		...actggttcgacccctgg	IGHJ5*02	IGHD3-9*01	0	2	0	9/13
#7	Z70613	IGHV4-59*01	tgtgcgagag.			........cagcagctggtac	ctcccct		....ctttgactactgg	IGHJ4*02	IGHD6-13*01	0	0	0	4/6
#8	Z70614	IGHV4-59*01	tgtgcgaga..		cactataa	...............ttcggggacttat.........	cccctc		........gactactgg	IGHJ4*02	IGHD3-16*01	0	2	0	7/14
#9	Z70615	IGHV4-59*01	tgtgcgagag.		ggctg	gtaaagaggg..........	tttcggaa		.tactggtacttcgatctctgg	IGHJ2*01	IGHD5-24*01	0	2	0	7/13
#10	Z70616	IGHV4-34*01	tgtgcgagag.		cgg	.............gtttggg................	ttccc		...actggttcgacccctgg	IGHJ5*02	IGHD3-16*01	0	0	0	6/8
#11	Z70620	IGHV4-30-4*01	tgtgcgagaga		cc	..................ggggcgggatggtt.....	cgg		.gatgcttttgatatctgg	IGHJ3*02	IGHD3-16*01	1	4	0	5/5
#12	Z70621	IGHV4-39*01	tgtgcgagaca		ccacgatttatgg	...............ttcggggagtt...........	tgacccccc		......ttgactactgg	IGHJ4*02	IGHD3-16*01	0	1	0	12/21
#13	Z70622	IGHV4-39*06	tgtgcgagaga	t	tgccccgctcctgccaaaat	gtattactatggttcgggga..........	tatgtacg		.....tttgactactgg	IGHJ4*03	IGHD3-10*01	0	0	0	15/28

Yousfi Monod et al. Bioinformatics, 20, I379-I385 (2004)

The eleven IMGT amino acid classes according to the physico-chemical properties

http://www.imgt.org

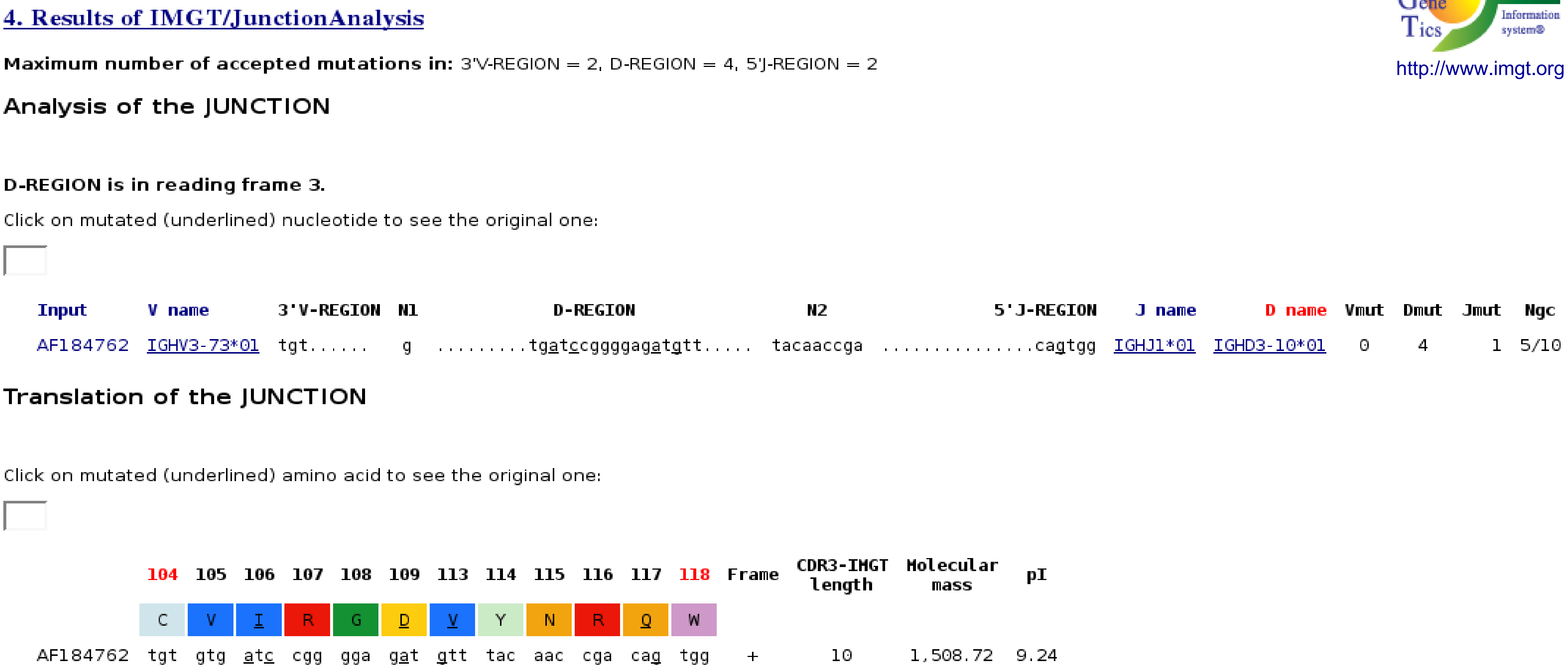


'Volume' classes		'Hydropathy' classes		
	in Å^3	Hydrophobic	Neutral	Hydrophilic
Very large	189-228	F W	Y	
Large	162-174	I L M		K R
Medium	138-154	V	H	E Q
Small	108-117	C	P T	D N
Very small	60-90	A	G S	

Pommié et al. J. Mol Recognit. ***17****, 17-32, 2004*

4. ‘Detailed view’: Results of IMGT/JunctionAnalysis

http://www.imgt.org



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:

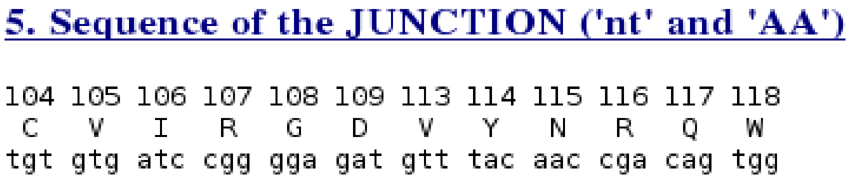
Input	V name	3'V-REGION	N1	D-REGION	N2	5'J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
AF184762	IGHV3-73*01	tgt......	g	.........tgatccggggagatgtt.....	tacaaccga	...............cagtgg	IGHJ1*01	IGHD3-10*01	0	4	1	5/10

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	I	R	G	D	V	Y	N	R	Q	W				
AF184762	tgt	gtg	atc	cgg	gga	gat	gtt	tac	aac	cga	cag	tgg	+	10	1,508.72	9.24

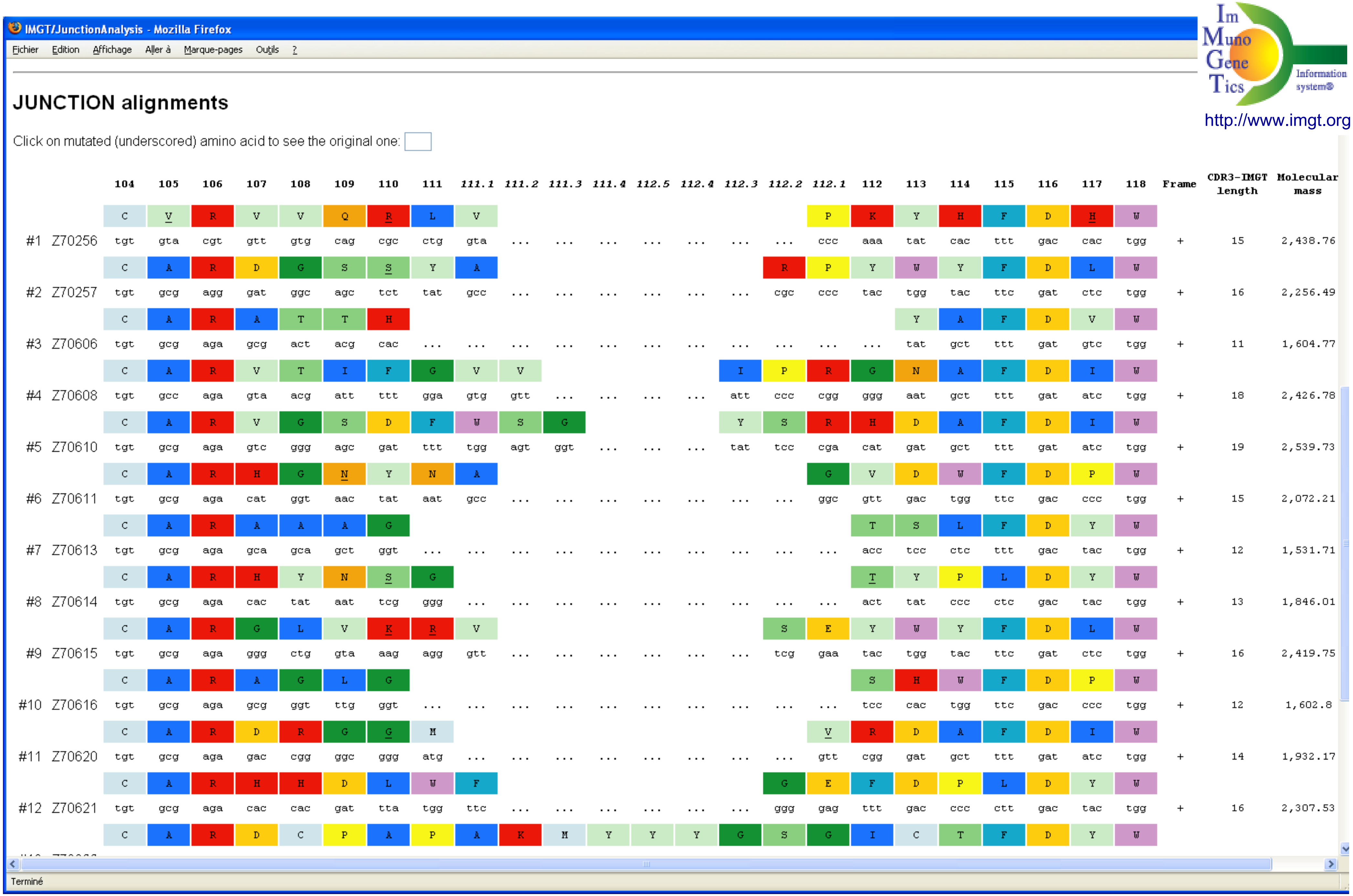
5. ‘Detailed view’: Sequence of the JUNCTION (‘nt’ and ‘AA’)



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 gtt tac aac cga cag tgg
```

http://www.imgt.org



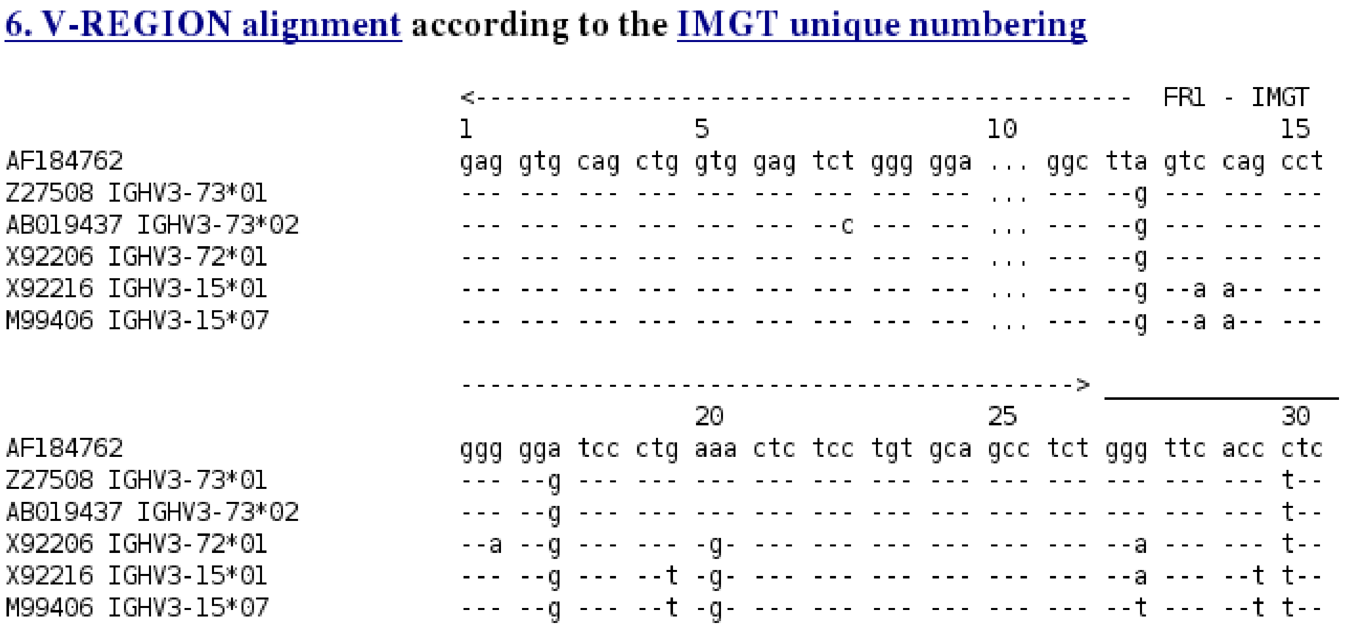
JUNCTION alignments

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

	104	105	106	107	108	109	110	111	111.1	111.2	111.3	111.4	112.5	112.4	112.3	112.2	112.1	112	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass
	C	V	R	V	V	Q	R	L	V								P	K	Y	H	F	D	H	W			
#1 Z70256	tgt	gta	cgt	gtt	gtg	cag	cgc	ctg	gta	...	...	...	...	...	...	...	ccc	aaa	tat	cac	ttt	gac	cac	tgg	+	15	2,438.76
	C	A	R	D	G	S	S	Y	A							R	P	Y	W	Y	F	D	L	W			
#2 Z70257	tgt	gcg	agg	gat	ggc	agc	tct	tat	gcc	...	...	...	...	...	...	cgc	ccc	tac	tgg	tac	ttc	gat	ctc	tgg	+	16	2,256.49
	C	A	R	A	T	T	H												Y	A	F	D	V	W			
#3 Z70606	tgt	gcg	aga	gcg	act	acg	cac	...	...	...	...	...	...	...	...	...	...	...	tat	gct	ttt	gat	gtc	tgg	+	11	1,604.77
	C	A	R	V	T	I	F	G	V	V					I	P	R	G	N	A	F	D	I	W			
#4 Z70608	tgt	gcc	aga	gta	acg	att	ttt	gga	gtg	gtt	...	...	...	...	att	ccc	cgg	ggg	aat	gct	ttt	gat	atc	tgg	+	18	2,426.78
	C	A	R	V	G	S	D	F	W	S	G				Y	S	R	H	D	A	F	D	I	W			
#5 Z70610	tgt	gcg	aga	gtc	ggg	agc	gat	ttt	tgg	agt	ggt	...	...	...	tat	tcc	cga	cat	gat	gct	ttt	gat	atc	tgg	+	19	2,539.73
	C	A	R	H	G	N	Y	N	A								G	V	D	W	F	D	P	W			
#6 Z70611	tgt	gcg	aga	cat	ggt	aac	tat	aat	gcc	...	...	...	...	...	...	...	ggc	gtt	gac	tgg	ttc	gac	ccc	tgg	+	15	2,072.21
	C	A	R	A	A	A	G											T	S	L	F	D	Y	W			
#7 Z70613	tgt	gcg	aga	gca	gca	gct	ggt	...	...	...	...	...	...	...	...	...	...	acc	tcc	ctc	ttt	gac	tac	tgg	+	12	1,531.71
	C	A	R	H	Y	N	S	G										T	Y	P	L	D	Y	W			
#8 Z70614	tgt	gcg	aga	cac	tat	aat	tcg	ggg	...	...	...	...	...	...	...	...	...	act	tat	ccc	ctc	gac	tac	tgg	+	13	1,846.01
	C	A	R	G	L	V	K	R	V							S	E	Y	W	Y	F	D	L	W			
#9 Z70615	tgt	gcg	aga	ggg	ctg	gta	aag	agg	gtt	...	...	...	...	...	...	tcg	gaa	tac	tgg	tac	ttc	gat	ctc	tgg	+	16	2,419.75
	C	A	R	A	G	L	G											S	H	W	F	D	P	W			
#10 Z70616	tgt	gcg	aga	gcg	ggt	ttg	ggt	...	...	...	...	...	...	...	...	...	...	tcc	cac	tgg	ttc	gac	ccc	tgg	+	12	1,602.8
	C	A	R	D	R	G	G	M									V	R	D	A	F	D	I	W			
#11 Z70620	tgt	gcg	aga	gac	cgg	ggc	ggg	atg	...	...	...	...	...	...	...	...	gtt	cgg	gat	gct	ttt	gat	atc	tgg	+	14	1,932.17
	C	A	R	H	H	D	L	W	F							G	E	F	D	P	L	D	Y	W			
#12 Z70621	tgt	gcg	aga	cac	cac	gat	tta	tgg	ttc	...	...	...	...	...	...	ggg	gag	ttt	gac	ccc	ctt	gac	tac	tgg	+	16	2,307.53
	C	A	R	D	C	P	A	P	A	K	M	Y	Y	Y	G	S	G	I	C	T	F	D	Y	W			

6. ‘Detailed view’: V-REGION alignment

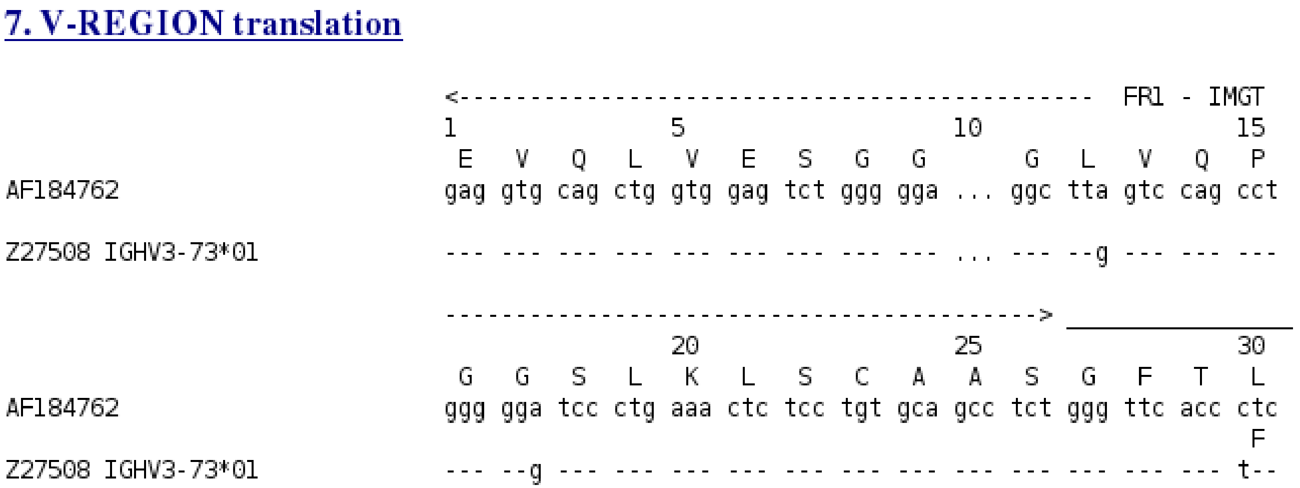
http://www.imgt.org



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

```
                         <---------------------------------------------  FR1 - IMGT
                         1               5                   10                  15
AF184762                 gag gtg cag ctg gtg gag tct ggg gga ... ggc tta gtc cag cct
Z27508 IGHV3-73*01       --- --- --- --- --- --- --- --- --- ... --- --g --- --- ---
AB019437 IGHV3-73*02     --- --- --- --- --- --- --c --- --- ... --- --g --- --- ---
X92206 IGHV3-72*01       --- --- --- --- --- --- --- --- --- ... --- --g --- --- ---
X92216 IGHV3-15*01       --- --- --- --- --- --- --- --- --- ... --- --g --a a-- ---
M99406 IGHV3-15*07       --- --- --- --- --- --- --- --- --- ... --- --g --a a-- ---

                         ------------------------------------------> ________________
                                         20                  25                  30
AF184762                 ggg gga tcc ctg aaa ctc tcc tgt gca gcc tct ggg ttc acc ctc
Z27508 IGHV3-73*01       --- --g --- --- --- --- --- --- --- --- --- --- --- --- t--
AB019437 IGHV3-73*02     --- --g --- --- --- --- --- --- --- --- --- --- --- --- t--
X92206 IGHV3-72*01       --a --g --- --- -g- --- --- --- --- --- --- --a --- --- t--
X92216 IGHV3-15*01       --- --g --- --t -g- --- --- --- --- --- --- --a --- --t t--
M99406 IGHV3-15*07       --- --g --- --t -g- --- --- --- --- --- --- --t --- --t t--
```

7. ‘Detailed view’: V-REGION translation



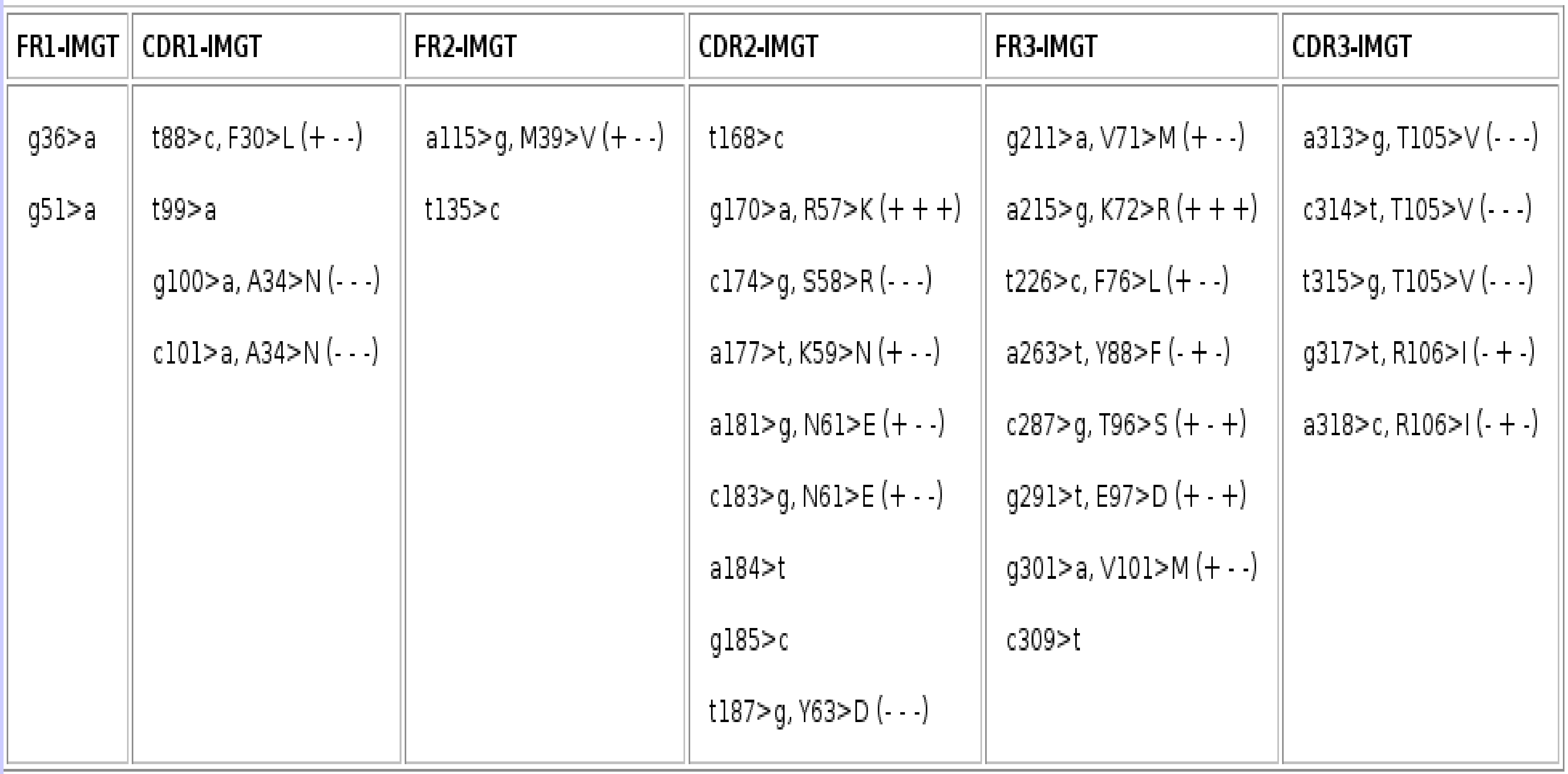
7. V-REGION translation

```
                         <---------------------------------------------  FR1 - IMGT
                         1               5                   10                  15
                          E   V   Q   L   V   E   S   G   G       G   L   V   Q   P
AF184762                 gag gtg cag ctg gtg gag tct ggg gga ... ggc tta gtc cag cct

Z27508 IGHV3-73*01       --- --- --- --- --- --- --- --- --- ... --- --g --- --- ---

                         ------------------------------------------> ________________
                                         20                  25                  30
                          G   G   S   L   K   L   S   C   A   A   S   G   F   T   L
AF184762                 ggg gga tcc ctg aaa ctc tcc tgt gca gcc tct ggg ttc acc ctc
                                                                                  F
Z27508 IGHV3-73*01       --- --g --- --- --- --- --- --- --- --- --- --- --- --- t--
```

8-10. ‘Detailed view’: mutation analysis

http://www.imgt.org

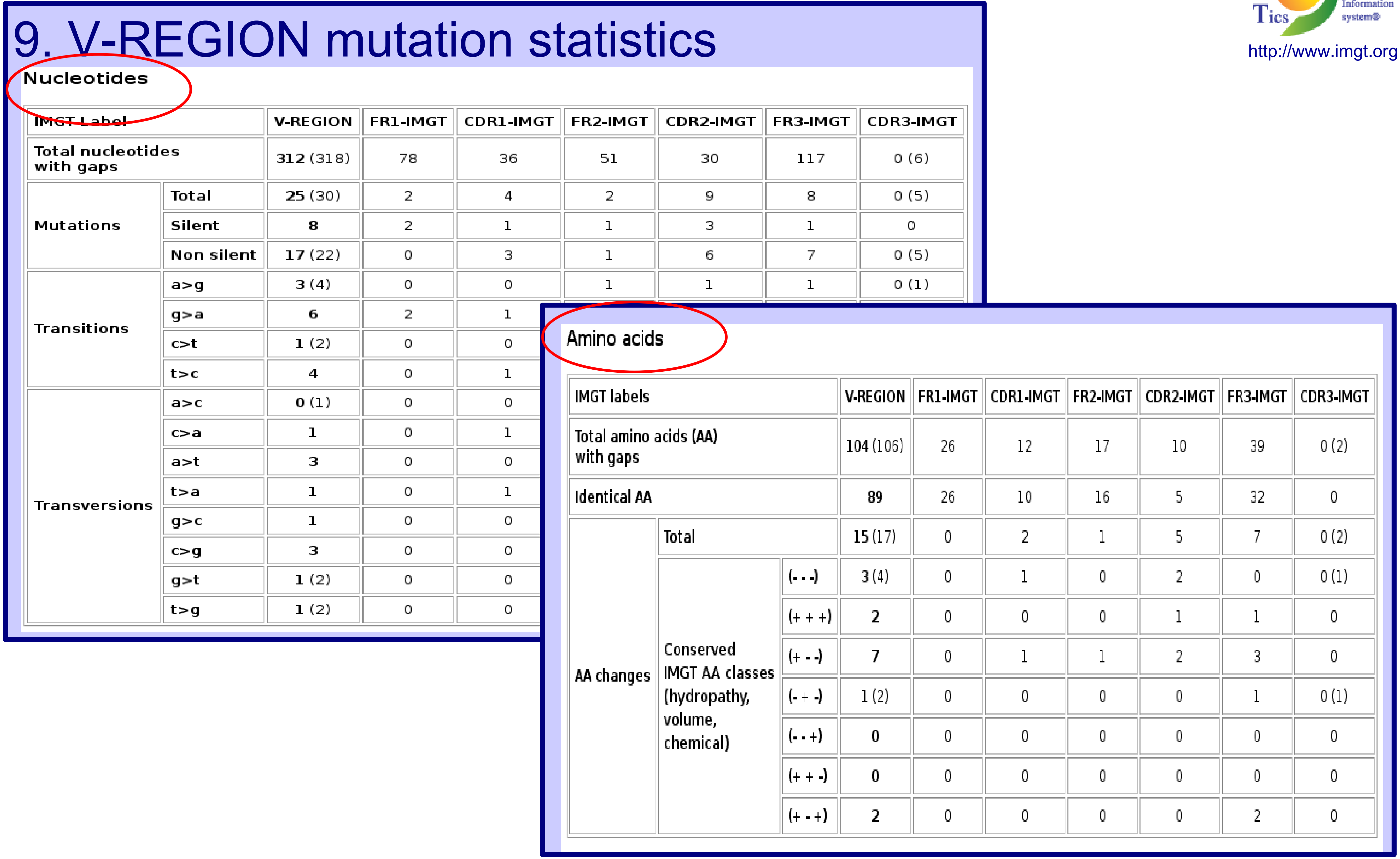
8. V-REGION mutation table



FR1-IMGT	CDR1-IMGT	FR2-IMGT	CDR2-IMGT	FR3-IMGT	CDR3-IMGT
g36>a g51>a	t88>c, F30>L (+ - -) t99>a g100>a, A34>N (- - -) c101>a, A34>N (- - -)	a115>g, M39>V (+ - -) t135>c	t168>c g170>a, R57>K (+ + +) c174>g, S58>R (- - -) a177>t, K59>N (+ - -) a181>g, N61>E (+ - -) c183>g, N61>E (+ - -) a184>t g185>c t187>g, Y63>D (- - -)	g211>a, V71>M (+ - -) a215>g, K72>R (+ + +) t226>c, F76>L (+ - -) a263>t, Y88>F (- + -) c287>g, T96>S (+ - +) g291>t, E97>D (+ - +) g301>a, V101>M (+ - -) c309>t	a313>g, T105>V (- - -) c314>t, T105>V (- - -) t315>g, T105>V (- - -) g317>t, R106>I (- + -) a318>c, R106>I (- + -)

8-10. ‘Detailed view’: mutation analysis

http://www.imgt.org



9. V-REGION mutation statistics

Nucleotides

IMGT Label		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
	Non silent	17 (22)	0	3	1	6	7	0 (5)
Transitions	a>g	3 (4)	0	0	1	1	1	0 (1)
	g>a	6	2	1				
	c>t	1 (2)	0	0				
	t>c	4	0	1				
Transversions	a>c	0 (1)	0	0				
	c>a	1	0	1				
	a>t	3	0	0				
	t>a	1	0	1				
	g>c	1	0	0				
	c>g	3	0	0				
	g>t	1 (2)	0	0				
	t>g	1 (2)	0	0				

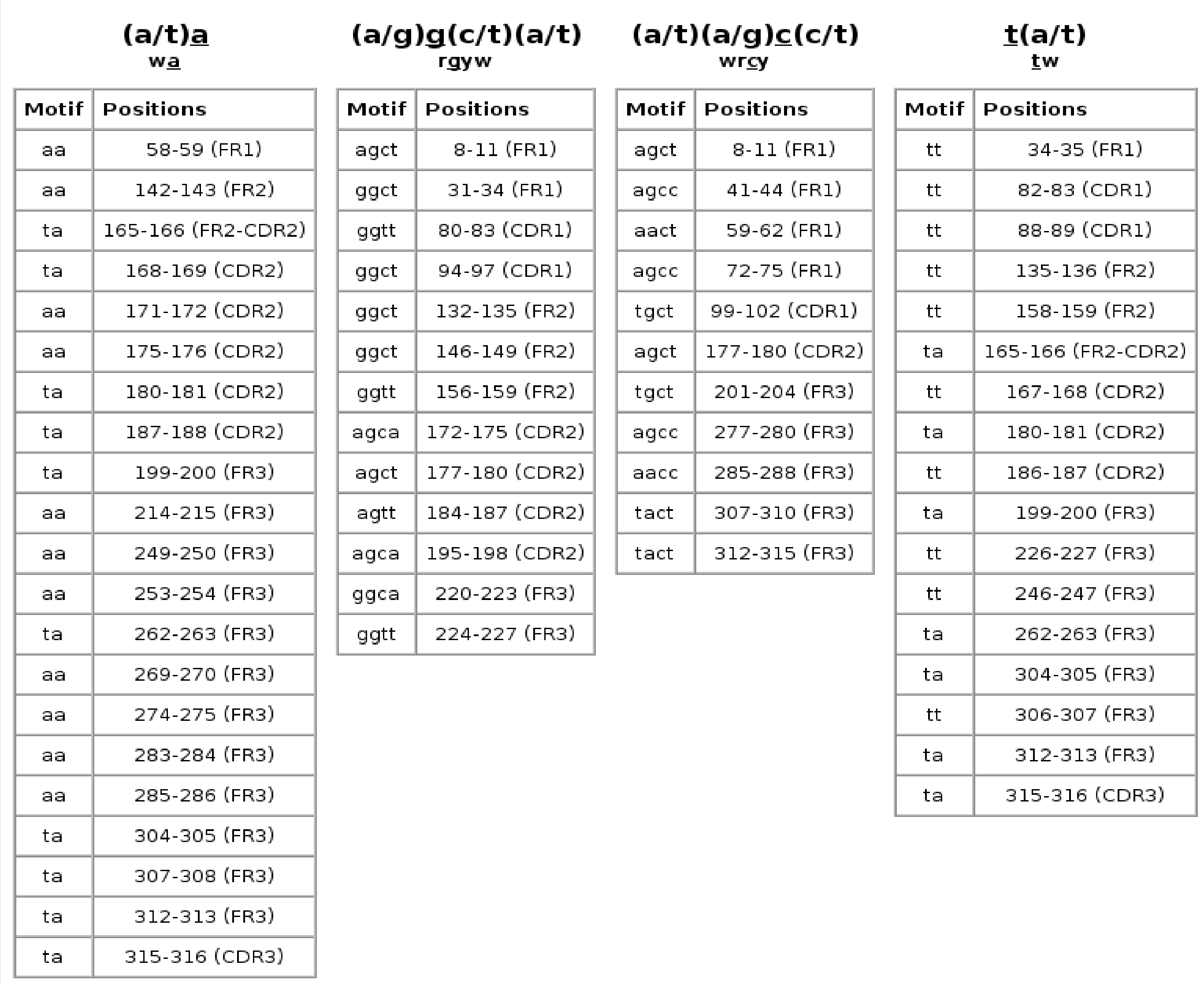
Amino acids

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

8-10. ‘Detailed view’: mutation analysis

http://www.imgt.org

10. Mutation hot spots (in germline V-REGION)



(a/t)a
wa

Motif	Positions
aa	58-59 (FR1)
aa	142-143 (FR2)
ta	165-166 (FR2-CDR2)
ta	168-169 (CDR2)
aa	171-172 (CDR2)
aa	175-176 (CDR2)
ta	180-181 (CDR2)
ta	187-188 (CDR2)
ta	199-200 (FR3)
aa	214-215 (FR3)
aa	249-250 (FR3)
aa	253-254 (FR3)
ta	262-263 (FR3)
aa	269-270 (FR3)
aa	274-275 (FR3)
aa	283-284 (FR3)
aa	285-286 (FR3)
ta	304-305 (FR3)
ta	307-308 (FR3)
ta	312-313 (FR3)
ta	315-316 (CDR3)

(a/g)g(c/t)(a/t)
rgyw

Motif	Positions
agct	8-11 (FR1)
ggct	31-34 (FR1)
ggtt	80-83 (CDR1)
ggct	94-97 (CDR1)
ggct	132-135 (FR2)
ggct	146-149 (FR2)
ggtt	156-159 (FR2)
agca	172-175 (CDR2)
agct	177-180 (CDR2)
agtt	184-187 (CDR2)
agca	195-198 (CDR2)
ggca	220-223 (FR3)
ggtt	224-227 (FR3)

(a/t)(a/g)c(c/t)
wrcy

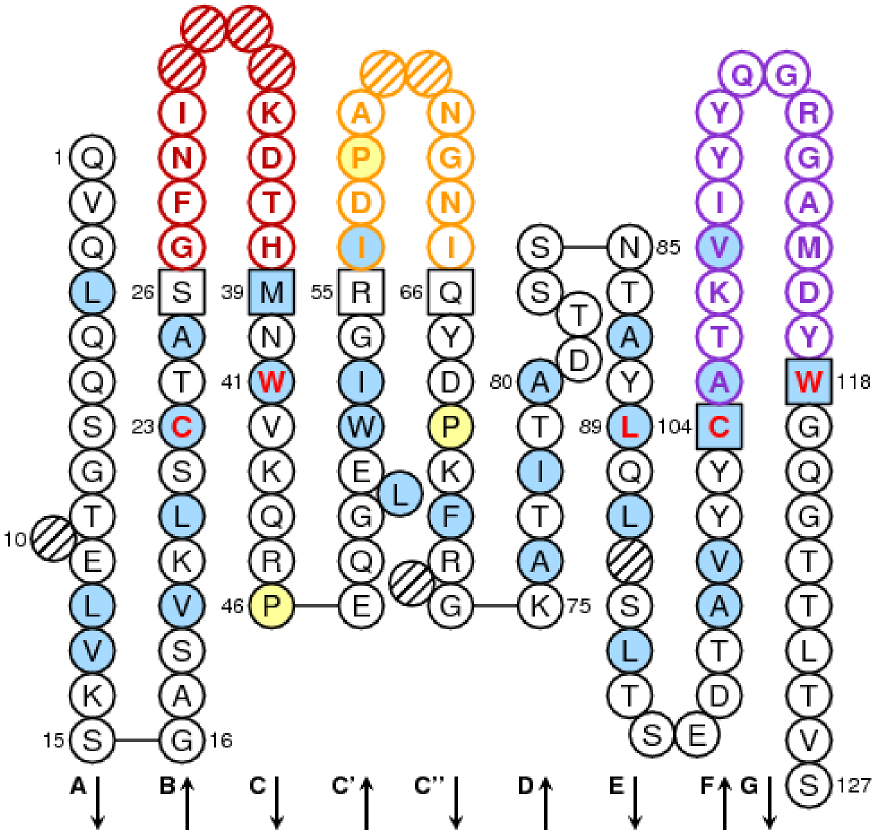
Motif	Positions
agct	8-11 (FR1)
agcc	41-44 (FR1)
aact	59-62 (FR1)
agcc	72-75 (FR1)
tgct	99-102 (CDR1)
agct	177-180 (CDR2)
tgct	201-204 (FR3)
agcc	277-280 (FR3)
aacc	285-288 (FR3)
tact	307-310 (FR3)
tact	312-315 (FR3)

t(a/t)
tw

Motif	Positions
tt	34-35 (FR1)
tt	82-83 (CDR1)
tt	88-89 (CDR1)
tt	135-136 (FR2)
tt	158-159 (FR2)
ta	165-166 (FR2-CDR2)
tt	167-168 (CDR2)
ta	180-181 (CDR2)
tt	186-187 (CDR2)
ta	199-200 (FR3)
tt	226-227 (FR3)
tt	246-247 (FR3)
ta	262-263 (FR3)
ta	304-305 (FR3)
tt	306-307 (FR3)
ta	312-313 (FR3)
ta	315-316 (CDR3)

11. IMGT Collier de Perles

http://www.imgt.org



IMGT unique numbering rules

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

CDR-IMGT lengths are in themselves a crucial information.

[8.8.15]

Standardized description of the V-DOMAIN, whatever the antigen receptor, the chain type and the species.

13. ‘Detailed view’: Annotation by IMGT/Automat

13. Annotations by IMGT/Automat

```
        Label                  Location/Qualifiers
V-D-J-REGION                   58..415
                               /Nucleotide sequence
                               gaggtgcagctggtggagtctgggggaggcttagtccagcctggggggatccctgaaactc
                               tcctgtgcagcctctgggttcaccctcagtggctcaaatgtgcactgggtccgccaggcc
                               tccgggaaagggctggagtgggttggccgtatcaaaaggaatgctgagtctgacgcgaca
                               gcatatgctgcgtcgatgagaggcaggctcaccatctccagagatgattcaaagaacacg
                               gcgtttctgcaaatgaacagcctgaaaagcgatgacacggccatgtattattgtgtgatc
                               cggggagatgtttacaaccgacagtggggccagggaaccctggtcaccgtctcctcag
                               /Translation
                               EVQLVESGGGLVQPGGSLKLSCAASGFTLSGSNVHWVRQASGKGLEWVGRIKRNAESDAT
                               AYAASMRGRLTISRDDSKNTAFLQMNSLKSDDTAMYYCVIRGDVYNRQWGQGTLVTVSS
V-REGION                       58..351
                               /allele="IGHV3-73*01"
                               /CDR_length="[8.10.10]"
                               /gene="IGHV3-73"
                               /Nucleotide sequence
                               gaggtgcagctggtggagtctgggggaggcttagtccagcctggggggatccctgaaactc
                               tcctgtgcagcctctgggttcaccctcagtggctcaaatgtgcactgggtccgccaggcc
                               tccgggaaagggctggagtgggttggccgtatcaaaaggaatgctgagtctgacgcgaca
                               gcatatgctgcgtcgatgagaggcaggctcaccatctccagagatgattcaaagaacacg
                               gcgtttctgcaaatgaacagcctgaaaagcgatgacacggccatgtattattgt
                               /Translation
                               EVQLVESGGGLVQPGGSLKLSCAASGFTLSGSNVHWVRQASGKGLEWVGRIKRNAESDAT
                               AYAASMRGRLTISRDDSKNTAFLQMNSLKSDDTAMYYC
FR1-IMGT                       58..132
                               /AA_IMGT="AA 1 to 26, AA 10 is missing"
                               /Nucleotide sequence
                               gaggtgcagctggtggagtctgggggaggcttagtccagcctggggggatccctgaaactc
                               tcctgtgcagcctct
                               /Translation
                               EVQLVESGGGLVQPGGSLKLSCAAS
1st-CYS                        121..123
                               /Nucleotide sequence
                               tgt
                               /Translation
                               C
CDR1-IMGT                      133..156
                               /AA_IMGT="AA 27 to 34"
                               /Nucleotide sequence
                               gggttcaccctcagtggctcaaat
                               /Translation
                               GFTLSGSN
FR2-IMGT                       157..207
                               /AA_IMGT="AA 39 to 55"
                               /Nucleotide sequence
                               gtgcactgggtccgccaggcctccgggaaagggctggagtgggttggccgt
                               /Translation
                               VHWVRQASGKGLEWVGR
CONSERVED-TRP                  163..165
                               /Nucleotide sequence
                               tgg
                               /Translation
                               W
```

IMGT/V-QUEST provides a full automatic annotation of the V-J-REGION or V-D-J-REGION.

Im
Muno
Gene
Tics

Information
system®

http://www.imgt.org