Analyse des génomes

Fatena Bellahcene, Odile Colas, Laetitia Régnier, Jérôme Lane, Marie-Paule Lefranc

M2 Bioinformatique

Montpellier, 11-14 mars 2008

Plan

- Analyse de génomes (en immunogénétique)
- Axiomes et concepts d'IMGT-ONTOLOGY
- Un outil que vous utiliserez: IMGT/V-QUEST

Notions requises et …à redécouvrir:
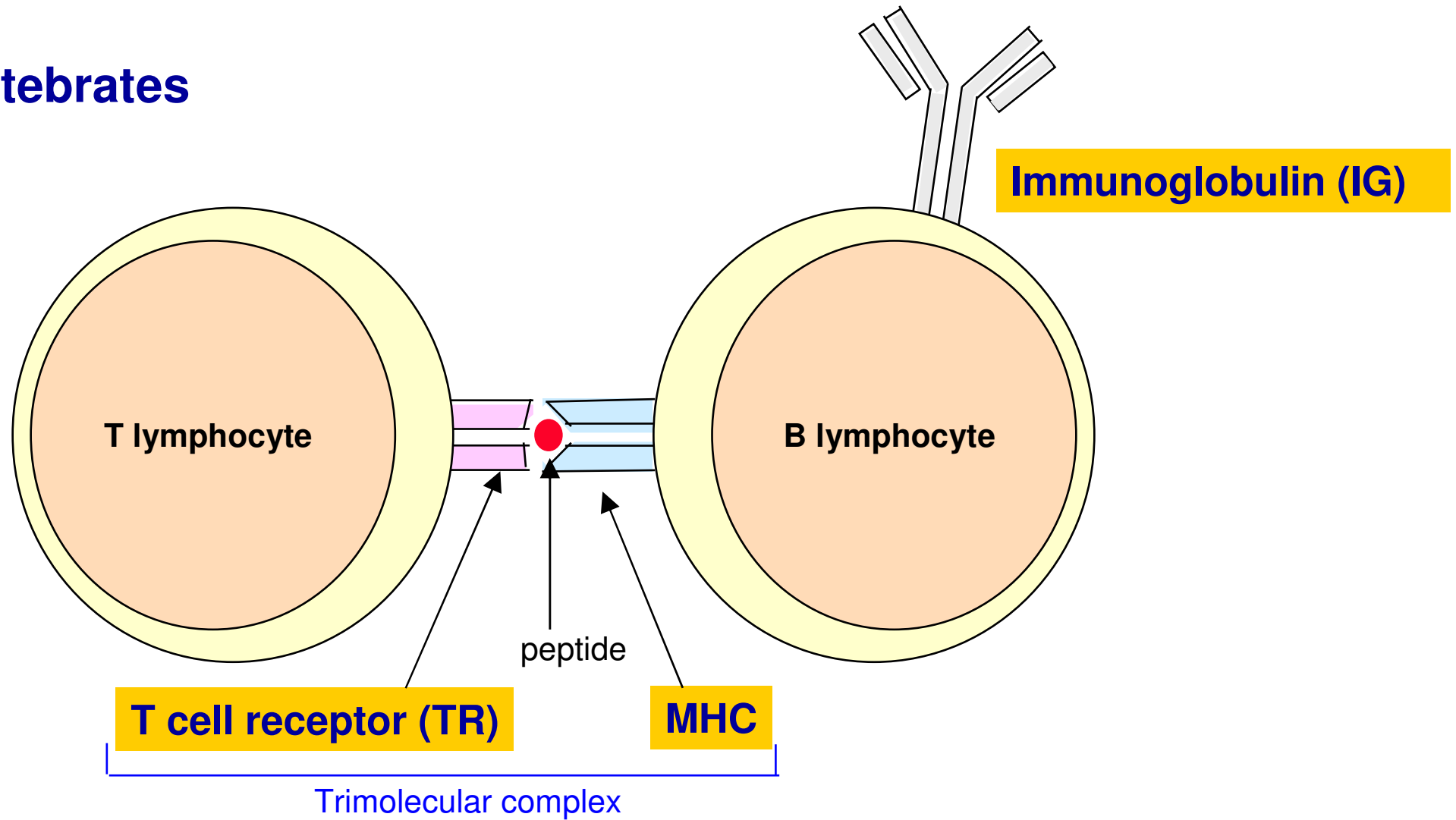
- Locus, gènes, allèles
- Orientation et localisation
- Description
- Classification
- Fonctionalité
- Relations séquences-structures-fonctions

IMGT®, the international ImMunoGeneTics information system®

Pourquoi IMGT® a-t-il été créé?

Bone marrow **Blood** **Lymph nodes, spleen**

http://imgt.cines.fr
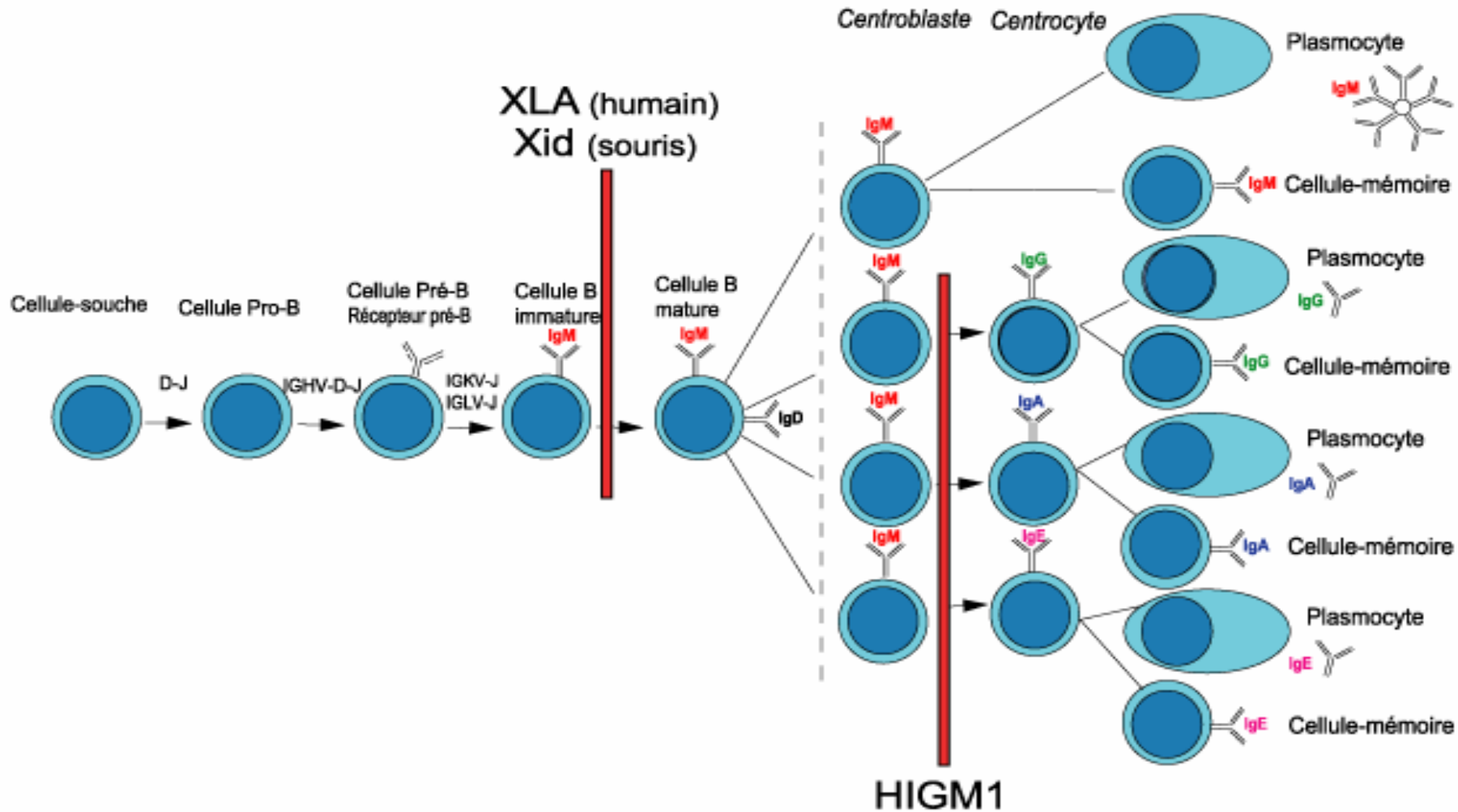
V-D-J and V-J rearrangements **Hypermutations, selection**

Spacefill 3D representation of an IgG

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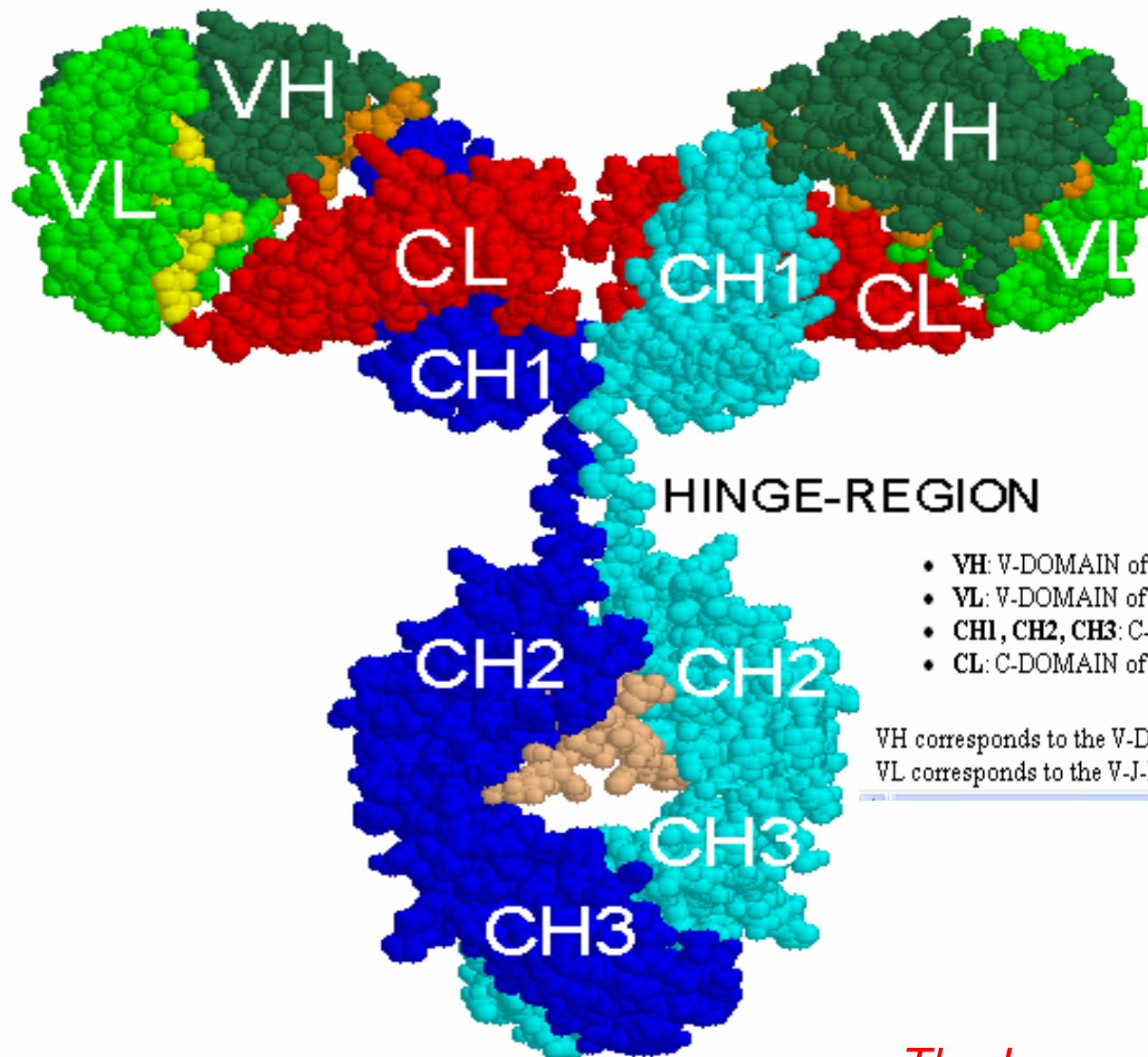
- **VH**: V-DOMAIN of the immunoglobulin heavy chain
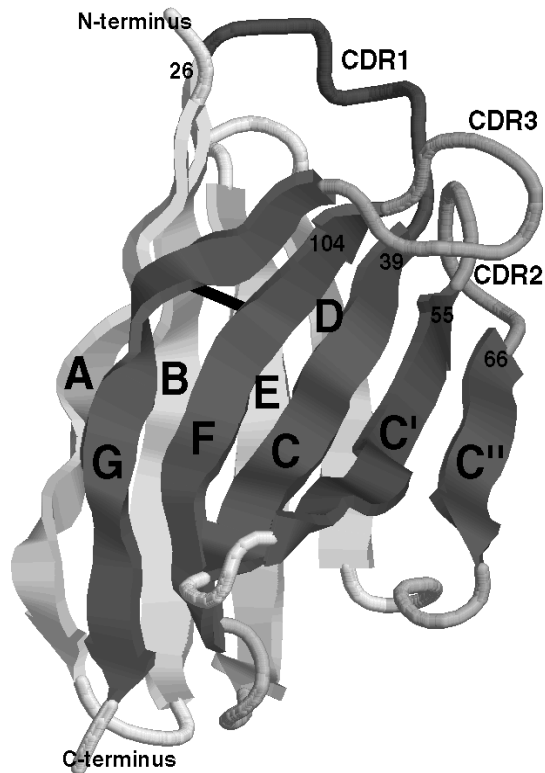- **VL**: V-DOMAIN of the immunoglobulin light chain
- **CH1, CH2, CH3**: C-DOMAIN of the immunoglobulin heavy chain
- **CL**: C-DOMAIN of the immunoglobulin light chain

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

Structural domains

IG and TR

MHC
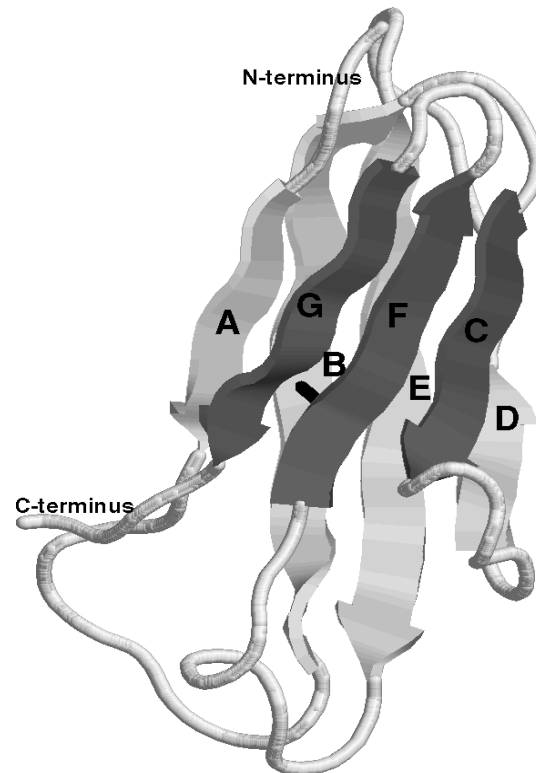
V-DOMAIN

C-DOMAIN

G-DOMAINs

Immunoglobulin (IG)   T cell receptor (TR)

V-DOMAIN

V-J-REGION

Light chain

V-DOMAIN

V-D-J-REGION

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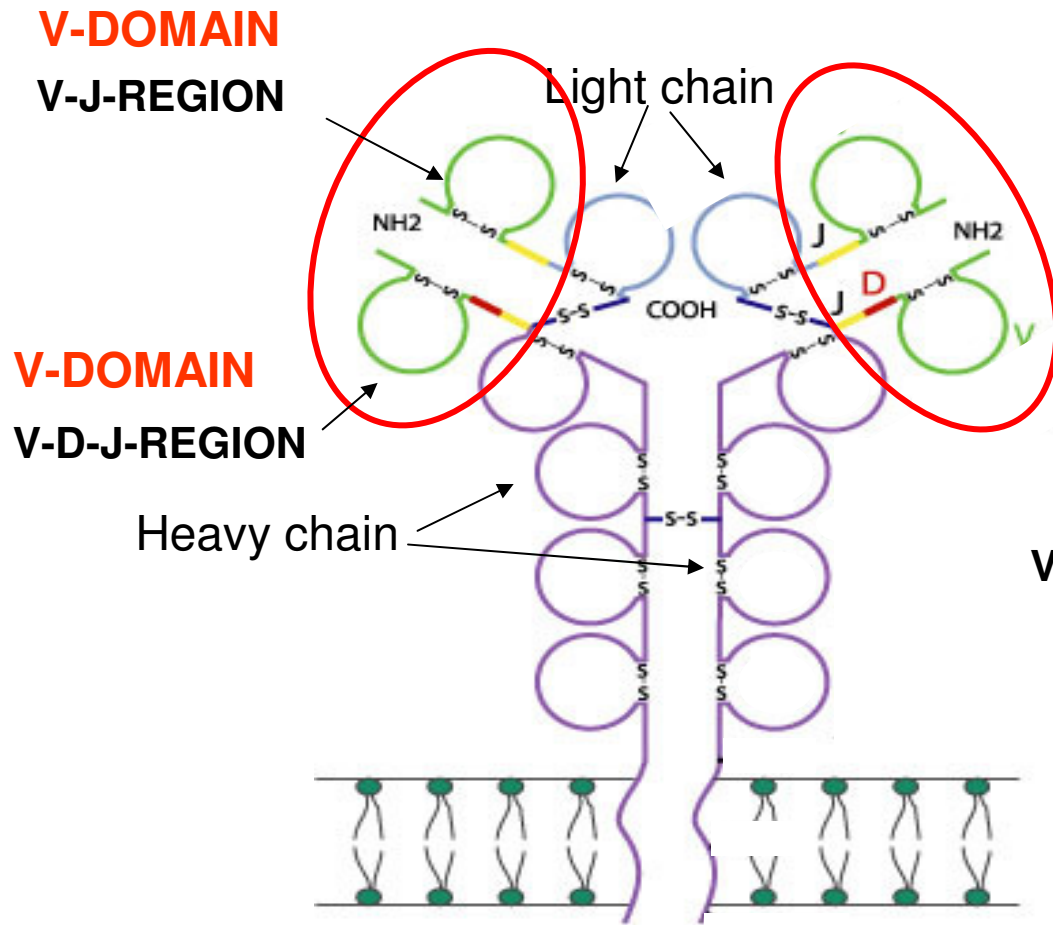
Contribution of the
2 V-DOMAINs
to the antigen binding site

Alpha	-	Beta
Gamma	-	Delta

V-J-REGION   **V-DJ-REGION**

Membrane IgM   **T cell receptor**

The Immunoglobulin FactsBook, 2001

Immunoglobulin IgG

IMGT Repertoire, http://imgt.cines.fr

Immunoglobulin (IG) synthesis

http://imgt.cines.fr
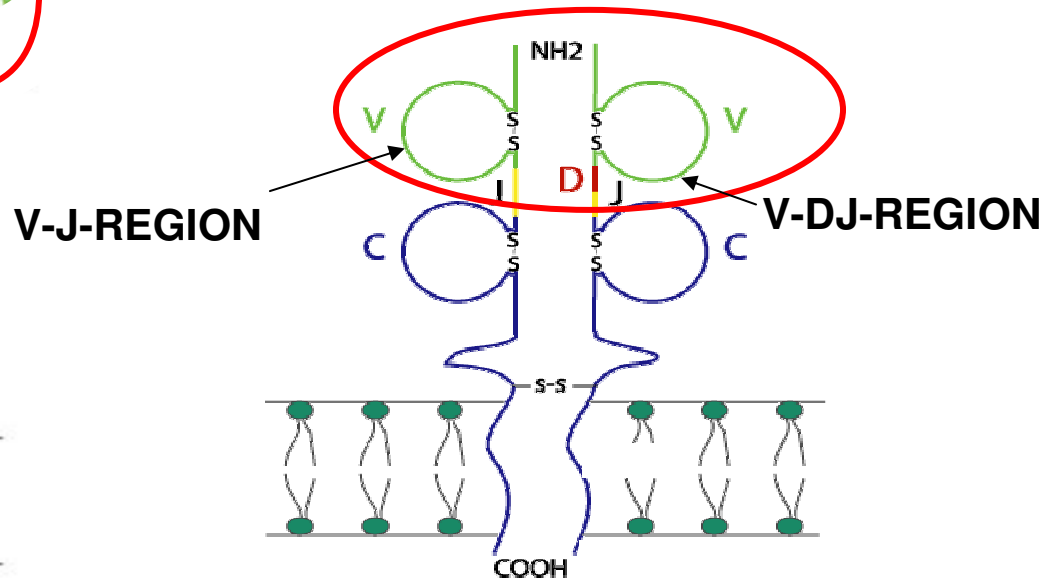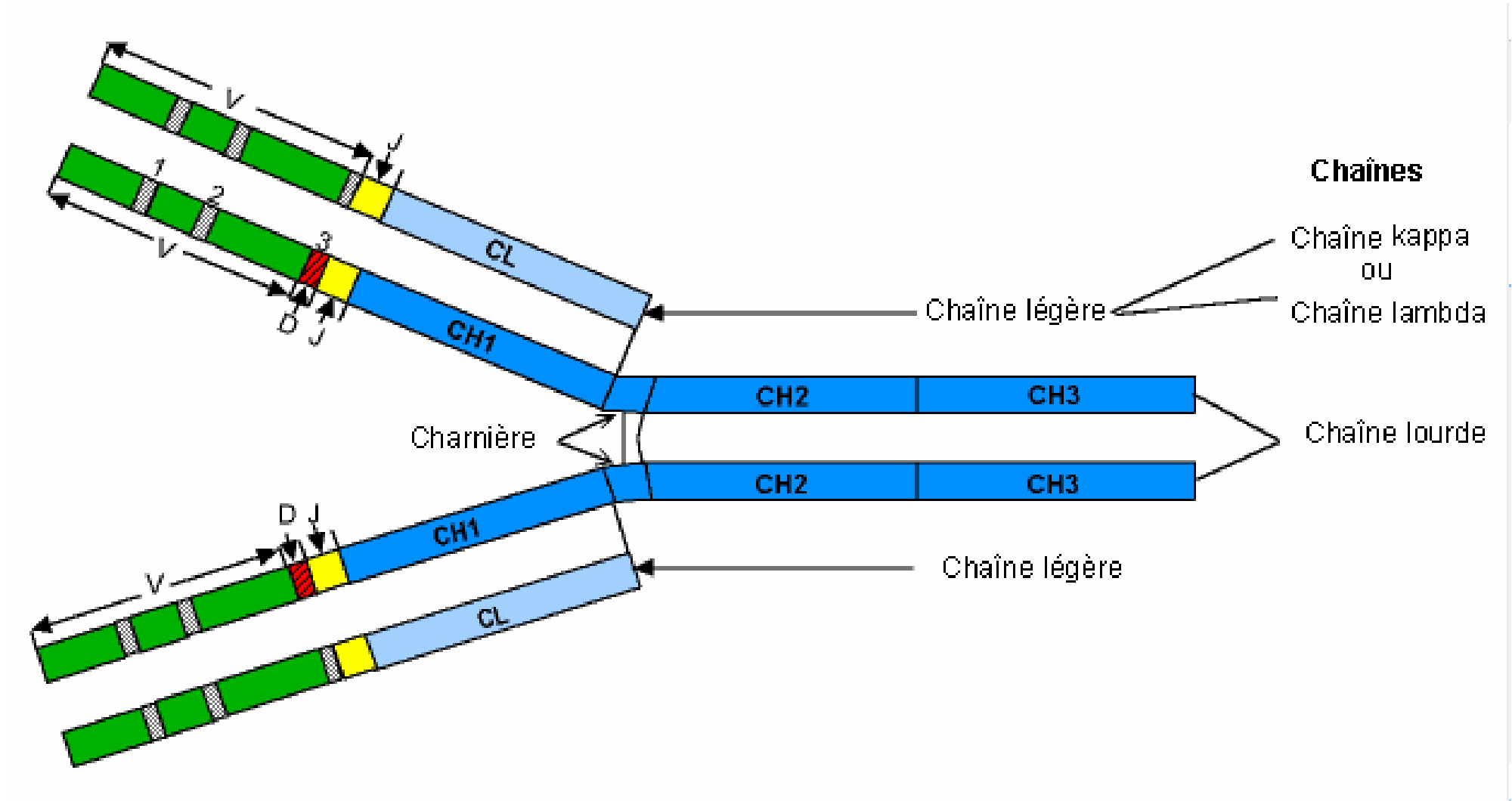
genomic DNA
(IGH Locus 14q32)

V   D   J   C

5'   3'
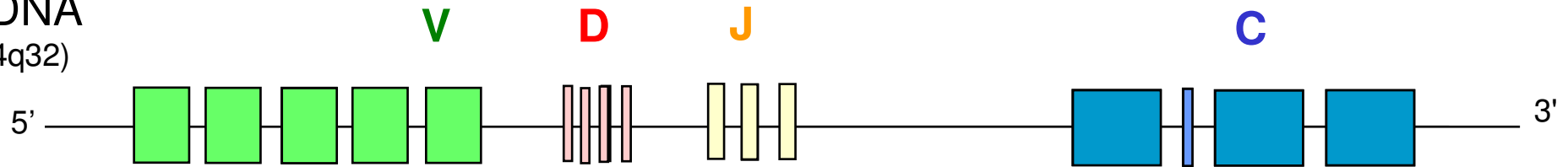
rearranged DNA

mRNA

5'   3'

2×10^{12} different IG
per individual

IMGT Repertoire, http://imgt.cines.fr

Immunoglobulin (IG) synthesis

http://imgt.cines.fr
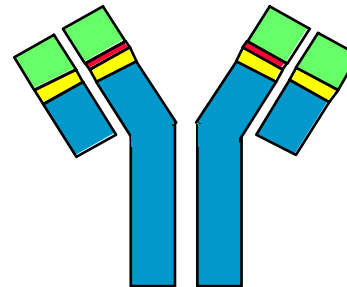
150
FUNCTIONAL IG GENES

HEAVY CHAIN

V   D   J   C

5'   3'

38 - 46   x   23   x   6

LIGHT CHAIN

V   J   C

5'   3'

30 - 35   x   5   Kappa
29 - 33   x   4 -5   Lambda
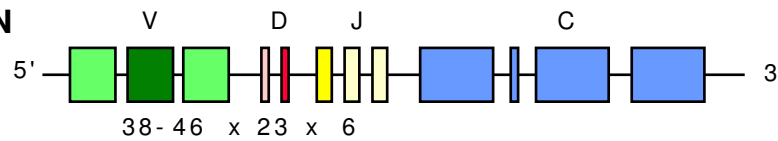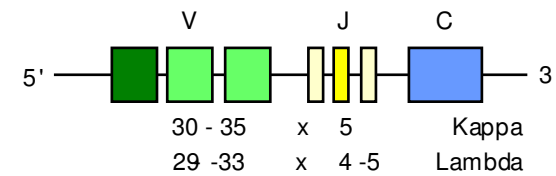
6300   **POTENTIAL RECOMBINATIONS**

185 +165   **POTENTIAL RECOMBINATIONS**

**N-DIVERSITY
SOMATIC MUTATIONS**
x 1000

5'   3'

5'   3'

ABOUT   **6.3 x 10^6**   POSSIBILITIES

ABOUT   **3.5 x 10^5**   POSSIBILITIES
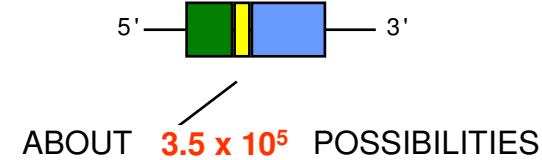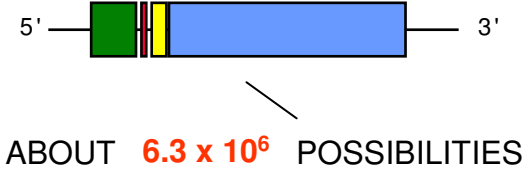
2 x 10^{12}

DIFFERENT ANTIBODIES

IMGT Repertoire, http://imgt.cines.fr
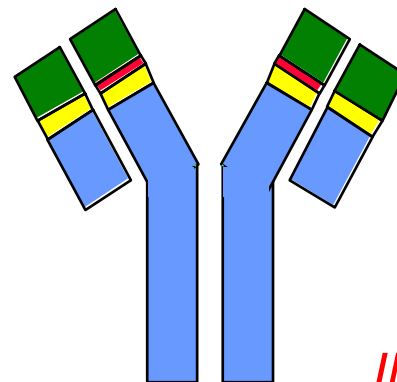
IMGT®

- 6 Bases de données
- 15 outils
- 10.000 pages de ressources Web
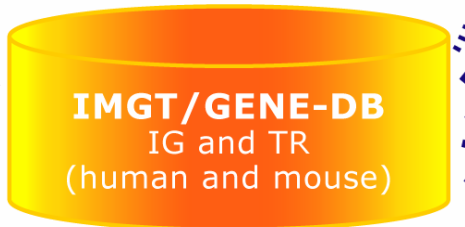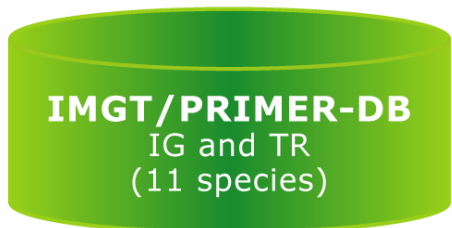
Axiomes et concepts d'IMGT-ONTOLOGY

IMGT-ONTOLOGY

http://imgt.cines.fr

IMGT-ONTOLOGY seven axioms:

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

IDENTIFICATION

OBTENTION

CLASSIFICATION

ORIENTATION

DESCRIPTION

LOCALIZATION

NUMEROTATION

Giudicelli and Lefranc, Bioinformatics 1999
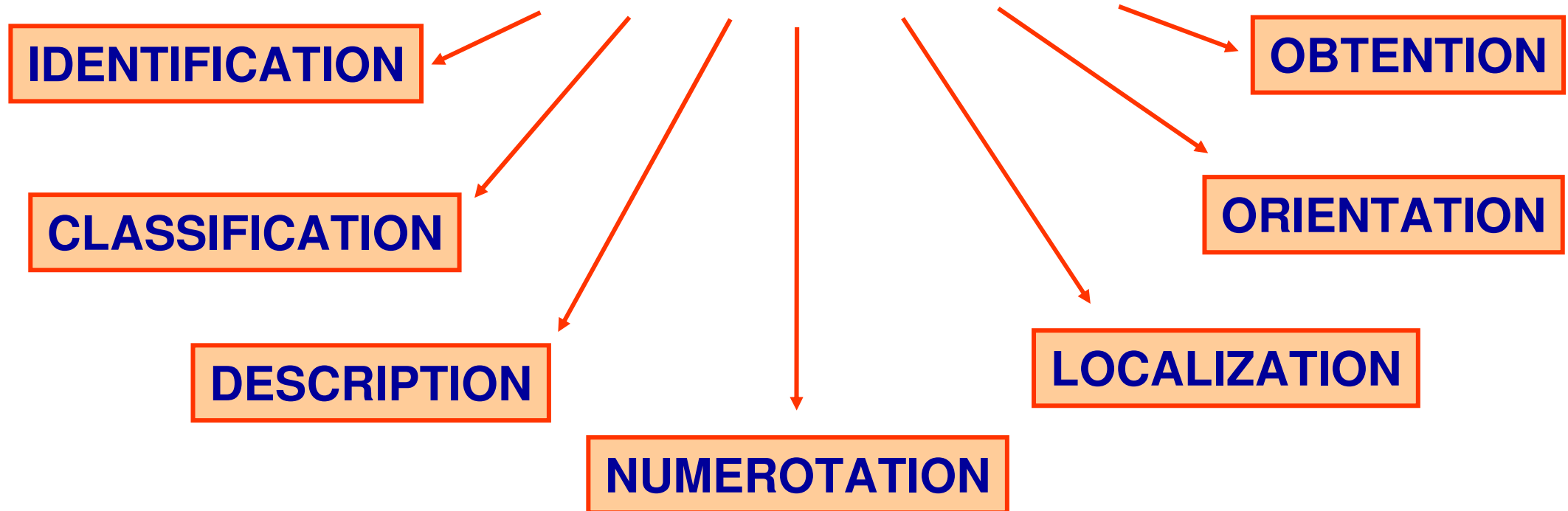
CLASSIFICATION axiom

Concepts

group

locus

is a member of
an instance of

subgroup

is a member of
an instance of

is ordered in
an instance of

gene

is a variant of
an instance of

allele

« Concepts »

Instances

IGLV

human IGL
(22q11.2)

is a member of

IGLV2

is ordered in

is a member of

IGLV2-11

is a variant of

IGLV2-11*02
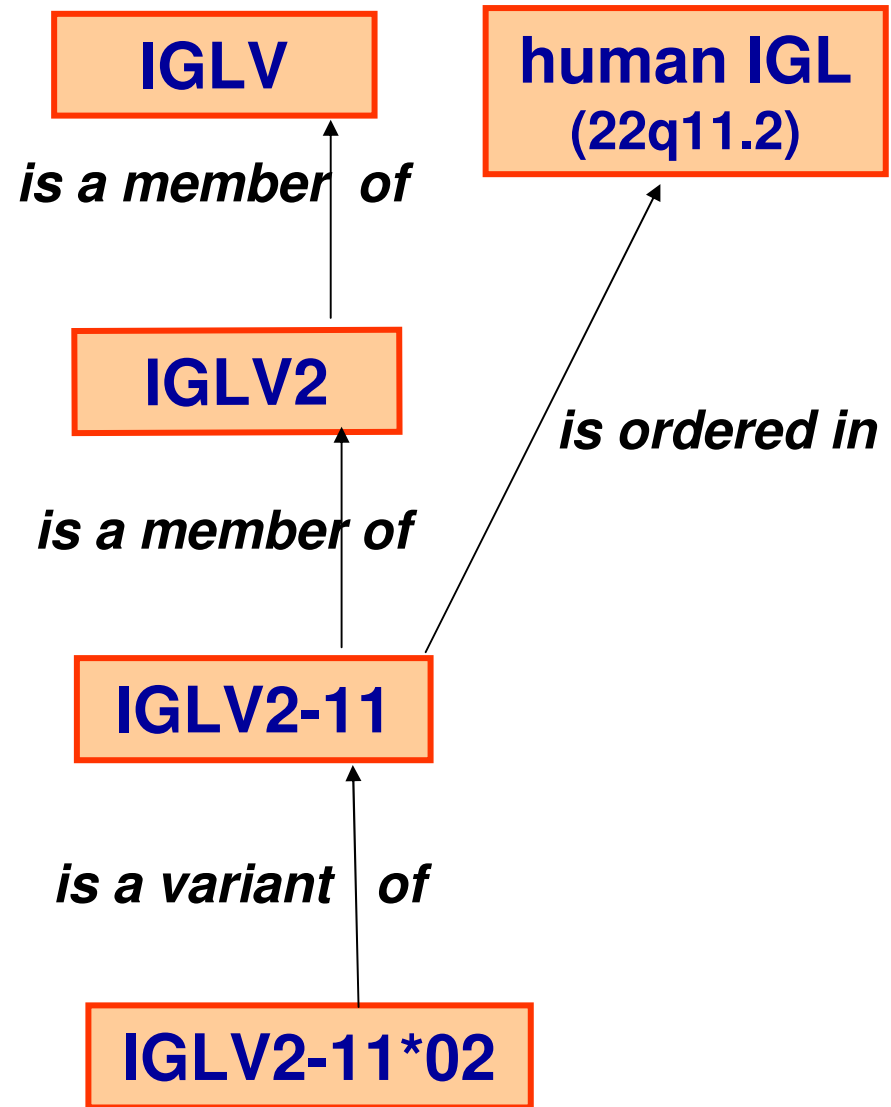
« Instances »

Human IGH locus

Chromosome 14q32.33
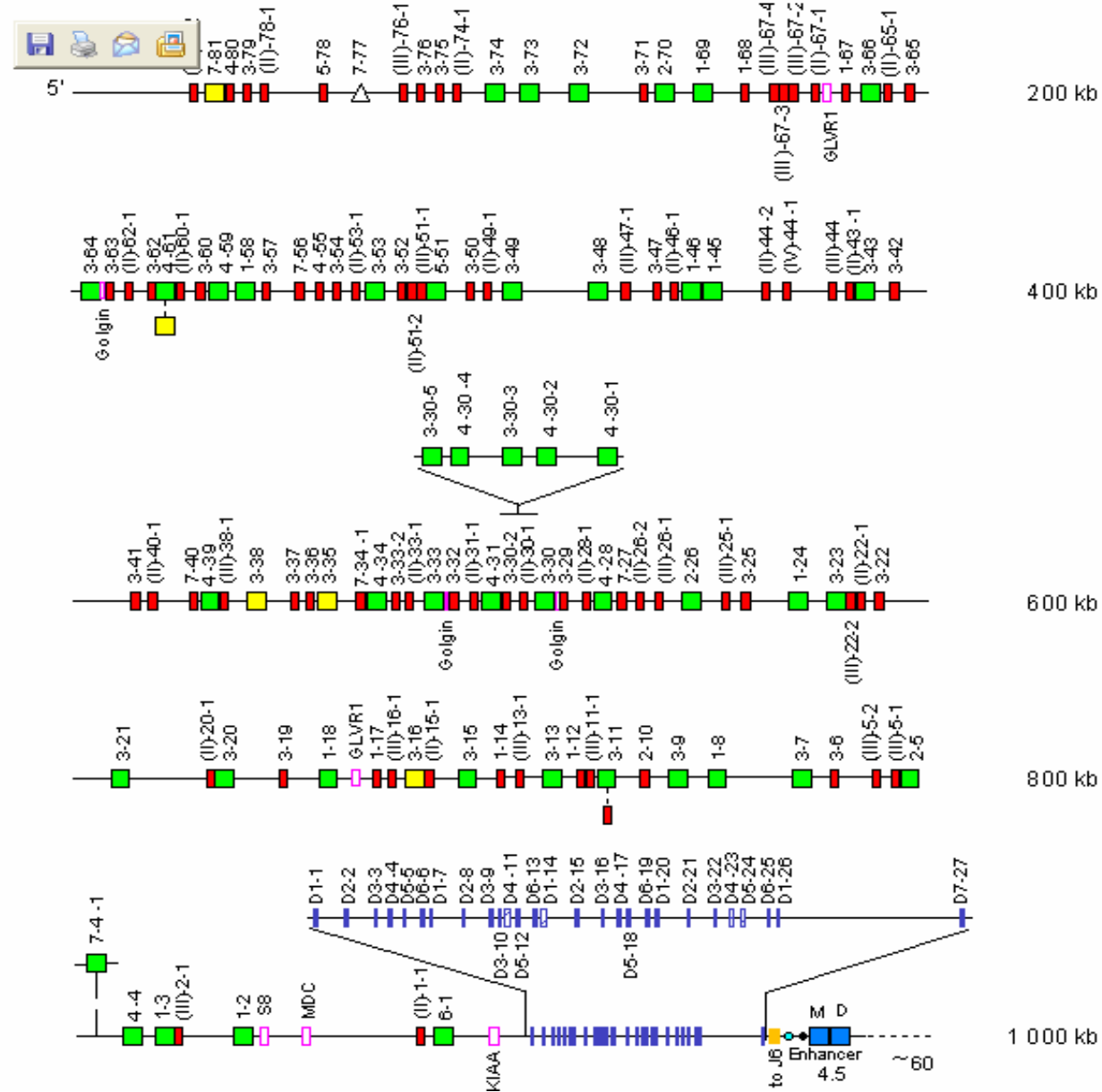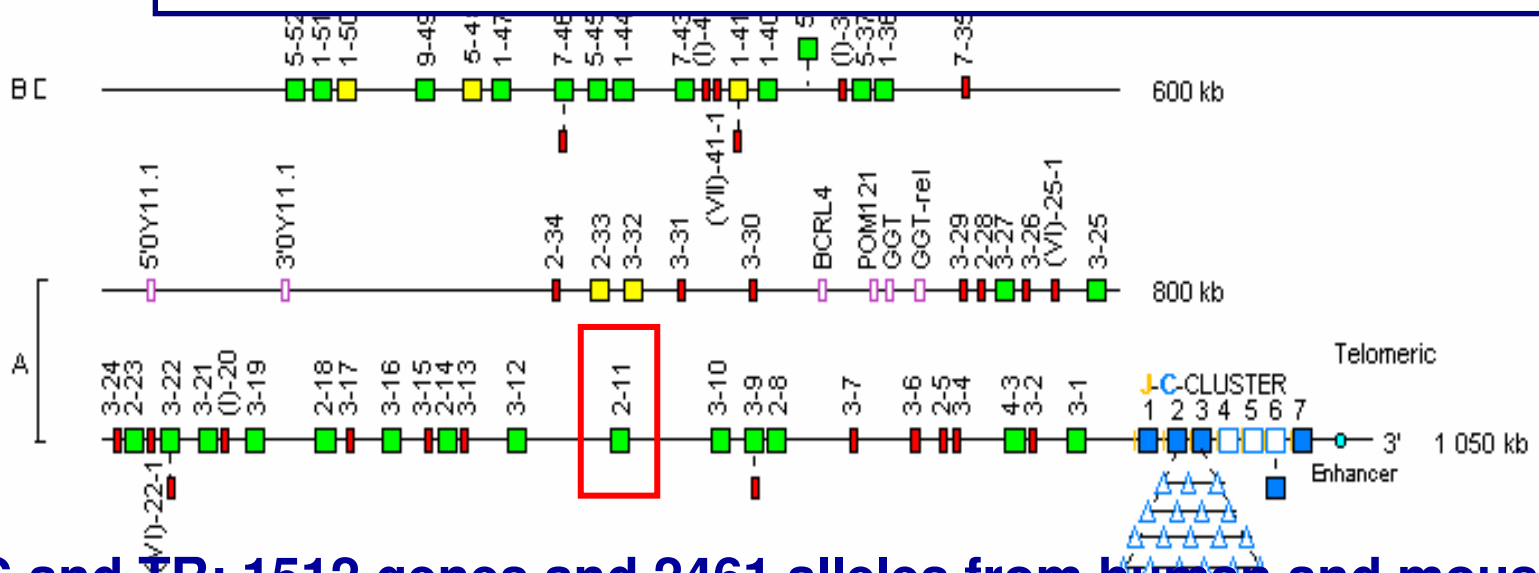
IG and TR: 1512 genes and 2461 alleles from human and mouse

CLASSIFICATION axiom

- The IMGT-ONTOLOGY main concepts of classification include 'group', 'subgroup', 'gene', 'allele'.

- They allowed to set up the nomenclature for IG and TR genes (V, D, J, C genes).

- IMGT gene names were approved by HGNC in 1999 and entered in GDB, LocusLink and Entrez Gene (NCBI).

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

- WHO-IUIS/IMGT 2007 report (*Dev. Comp. Immunol., Immunogenetics*).

DESCRIPTION axiom

PROTOTYPE for a V-GENE

http://imgt.cines.fr

V-GENE

L-PART1

V-EXON

5'UTR

FR1-IMGT
C

FR2-IMGT
W

FR3-IMGT
C

CDR3-IMGT

3'UTR

DONOR-SPLICE

V-REGION
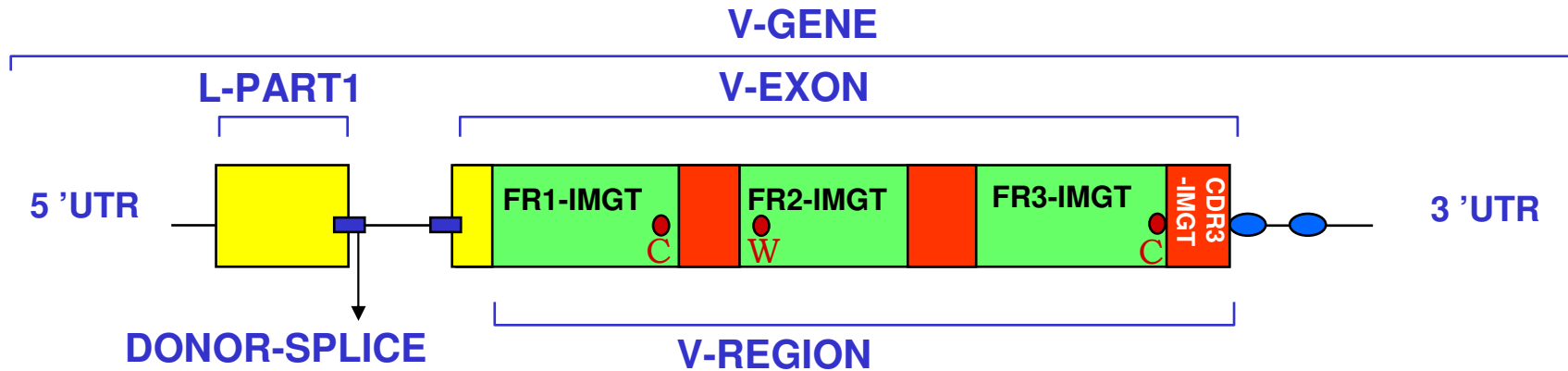
Label 1	Label 2	Relations entre Labels
V-GENE	V-EXON	
FR3-IMGT	CDR3-IMGT	
L-PART1	DONOR-SPLICE	
V-REGION	FR1-IMGT	
V-REGION	CDR3-IMGT	
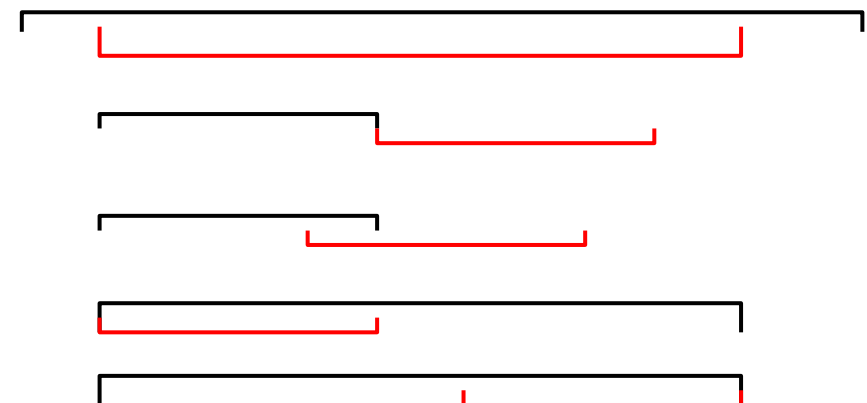
An example of V-GENE

>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
```


L-PART1   L-PART2   V-REGION   V-RS

5'UTR   V-INTRON   3'UTR

DONOR-SPLICE   ACCEPTOR-SPLICE   1st-CYS 23   2nd-CYS 104   V-HEPTAMER V-SPACER V-NONAMER
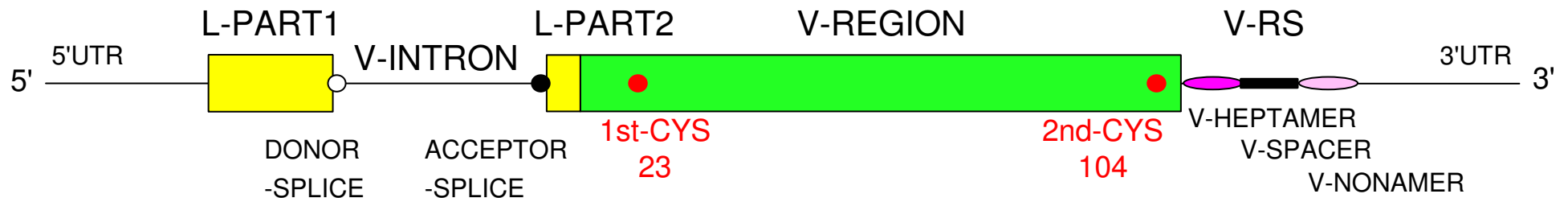
An example of D-GENE

```
>J00256|IGHD7-27*01|Homo sapiens D-GENE

ccagccgcag ggtttttggc tgagctgaga accactgtgc taactggggga cacagtgatt        60
ggcagctcta caaaaaccat gctcccccgg g
```
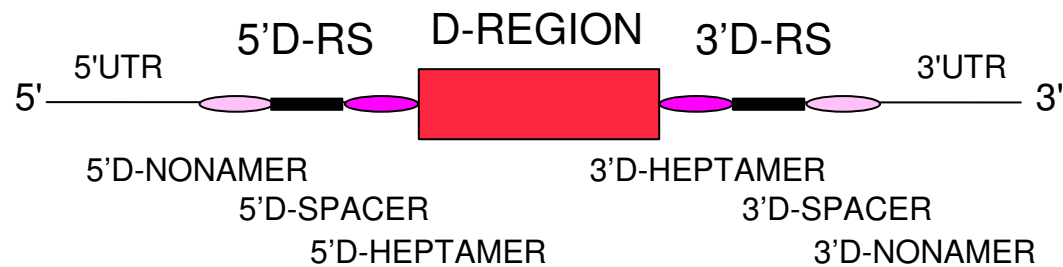
An example of J-GENE

>J00256|IGHJ1*01|*Homo sapiens* J-GENE

```
acccccgggct gtgggtttct gtgcccctgg ctcagggctg actcaccgtg gctgaatact       60
tccagcactg gggccagggc accctggtca ccgtctcctc aggtgagtct gctgtactgg      120
ggatagcggg gagccatgtg tactgggcca agcaagggct ttggcttcag                 170
```

J-RS   J-REGION

5'UTR   3'UTR

5' ——————◯▬◖—[●   ]—◯—————— 3'

J-NONAMER   J-TRP   DONOR
   J-SPACER   118   -SPLICE
      J-HEPTAMER
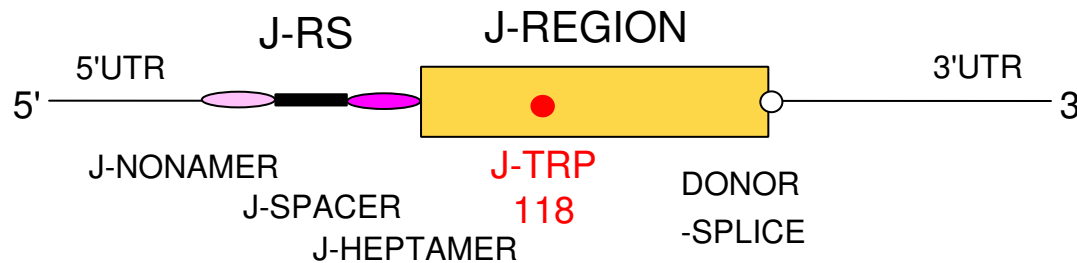
Heavy chain   **WGXG** (J-TRP)
Light chain   **FGXG** (J-PHE)

IMGT/LIGM-DB

IMGT/LIGM-DB Consultation module v3 - Mozilla Firefox

http://imgt.cines.fr

```
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="IGHV
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"
```

Done

D E S C R I P T I O N

Avril 2007: 105 188 sequences from 150 species
Mars 2008: 121 983 sequences from 221 species

IMGT-ONTOLOGY:
277 IMGT labels for sequences
285 IMGT labels for 3D structures

SO (Sequence ontology):
67 IMGT labels

DESCRIPTION axiom

- The IMGT-ONTOLOGY concepts of description comprise the standardized IMGT labels and relations.

- They have allowed to describe the IG, TR and MHC sequences and 3D structures, whatever the receptor type, the chain type, or the species.

- They are particularly useful to describe IG, TR, and MHC and their complexes (IG/antigen, TR/pMHC).

- It is possible to query the IMGT® databases (IMGT/LIGM-DB for sequences, IMGT/3Dstructure-DB for 3D structures) with IMGT labels.

- Sequence Ontology (SO) includes IMGT labels.

IMGT/V-QUEST

WELCOME !
to the IMGT/V-QUEST Search page

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Im
Muno
Gene
Tics

Information
system®

http://imgt.cines.fr

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

- Chondrichthyes

- Teleostei
  - Atlantic cod
  - Channel catfish
  - Rainbow trout

- Sheep

Analyse your T cell Receptor nucleotide sequences

- Human

- Mouse

- Non-human primates

Selection of ouput parameters

http://imgt.cines.fr

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

More options

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)

Nb of D-GENEs in IGH JUNCTIONs (default is 1)   default

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

http://imgt.cines.fr

Sequences aligned with IGHV5-51*03

1. Alignement for V-GENE

```
                                  <---------------------------------- FR1-IMGT ---------------
X56368 IGHV5-51*03                gaggtgcagctggtgcagtctggagca...gaggtgaaaaagccggggggagtctctgaag
AY393084              1327        -------gc-------------------...----------------------------
AY393088              1309        -------gc-------------------...---------------------------a
AY393089              1255        ----g-------a---------------...--a--------------------a---
AY393091              1237        ----------------------------...----------------------------
AY393092              1255        ----g-------a---------------...--a--------------------a---
AY393094              1300        -------gc--t----------------...----------------------------
                                  ---------------->_____      CDR1-IMGT _____<-----
X56368 IGHV5-51*03                atctcctgtaagggttctggatacagctttaccagctactgg............atcggc
AY393084              1327        -----t-------c---------------------ta-----...........------
AY393088              1309        ----------g--c---------------------tc-----...........------
AY393089              1255        ------------c---------------c---g---------...........------
AY393091              1237        ------------------------t-tt--c----aa-----...........------
AY393092              1255        ------------c---------------c---g---------...........------
AY393094              1300        ------------c---------------t--------ta-----..........------
                                  --------------- FR2-IMGT ------------------->_____ CDR
X56368 IGHV5-51*03                tgggtgcgccagatgcccgggaaaggcctggagtggatgggggatcatctatcctggtgac
AY393084              1327        -------------------------------------------------t----------
AY393088              1309        -------------------------------------------------t----------
AY393089              1255        -------------------------------------------g----a---g-g-----------
AY393091              1237        -------------a----c------------------g--c-------a------a---t
AY393092              1255        -------------------------------------g----a---g-g-----------
AY393094              1300        -------------------------------------------------t--------t----------
```

6. Protein display (mutations displayed)

```
                              FR1-IMGT          CDR1-IMGT       FR2-IMGT          CD
                              (1-26)            (27-38)         (39-55)           (

                              1         10        20      30      40        50
                              .........|.........|.....   ...|........   .|.........|.....  ..
X56368 IGHV5-51*03            EVQLVQSGA.EVKKPGESLKISCKGS   GYSFTSYW....   IGWVRQMPGKGLEWMGI  IY
AY393084               1327   --R------.---------------A-   ------N-....   ----------------  -F
AY393088               1309   --R------.--------------RA-   ------H-....   ----------------  -F
AY393089               1255   -G--E----.---------------A-   --T-A---....   --------------V--  V-
AY393091               1237   ---------.----------------    --I--K--....   ------I-------VA-  -N
AY393092               1255   -G--E----.---------------A-   --T-A---....   --------------V--  V-
AY393094               1300   --R------.---------------A-   ------N-....   --------------I--  -F


                              R2-IMGT            FR3-IMGT
                              56-65)             (66-104)

                               60       70        80        90        100
                              ..|.....  ....|.........|.........|.........|....
X56368 IGHV5-51*03            PGDSDT..  RYSPSFQ.GQVTISADKSISTAYLQWSSLKASDTAMYYC  AR
AY393084               1327   ------..  I-------.------V--------------------I---  --QNPPEYSGA
AY393088               1309   ------..  I-----E.----------------------T-S-----  --QNPPEYSGA
AY393089               1255   ---F--..  K------.-H----V-R------------------F---  --EMLYGSGGY
AY393091               1237   -D----..  --NT---.------V----------------T-------  --QNPPEYSGA
AY393092               1255   ---F--..  K------.-H------R-T---------------VI---  --EMLYGSGGY
AY393094               1300   ------..  I------.------V----N------------T-------  --QNPPEYSGA
```

http://imgt.cines.fr

IMGT/V-QUEST - Mozilla Firefox

File   Edit   View   Go   Bookmarks   Tools   Help

3. V-REGION translation

```
                        <------------------------------------------------ FR1 - IMGT
                        1                  5                 10                 15
                        E   V   Q   L   V   Q   S   G   A       E   V   K   K   P
X56368 IGHV5-51*03      gag gtg cag ctg gtg cag tct gga gca ... gag gtg aaa aag ccg
                                    R
AY393084        1327    --- --- -gc --- --- --- --- --- --- ... --- --- --- --- ---
                                    R
AY393088        1309    --- --- -gc --- --- --- --- --- --- ... --- --- --- --- ---
                                G           E
AY393089        1255    --- -g- --- --- -a- --- --- --- --- ... --a --- --- --- ---

AY393091        1237    --- --- --- --- --- --- --- --- --- ... --- --- --- --- ---
                                G           E
AY393092        1255    --- -g- --- --- -a- --- --- --- --- ... --a --- --- --- ---
                                    R
AY393094        1300    --- --- -gc --t --- --- --- --- --- ... --- --- --- --- ---


                        ---------------------------------------------->  _____
                                           20                25                 30
                        G   E   S   L   K   I   S   C   K   G   S   G   Y   S   F
X56368 IGHV5-51*03      ggg gag tct ctg aag atc tcc tgt aag ggt tct gga tac agc ttt
                                                                A
AY393084        1327    --- --- --- --- --- --- --t --- --- -c- --- --- --- --- ---
                                                            R   A
AY393088        1309    --- --- --- --- --a --- --- --- --- -g- -c- --- --- --- ---
                                                                A                   T
AY393089        1255    --- --- --- --a --- --- --- --- --- -c- --- --- --- -c- ---
                                                                                    I
AY393091        1237    --- --- --- --- --- --- --- --- --- --- --- --- --t -tt --c
                                                                A                   T
AY393092        1255    --- --- --- --a --- --- --- --- --- -c- --- --- --- -c- ---
                                                                A
AY393094        1300    --- --- --- --- --- --- --- --- --- -c- --- --- --- --t ---
```

V-DOMAINs: VH and V-KAPPA

http://imgt.cines.fr

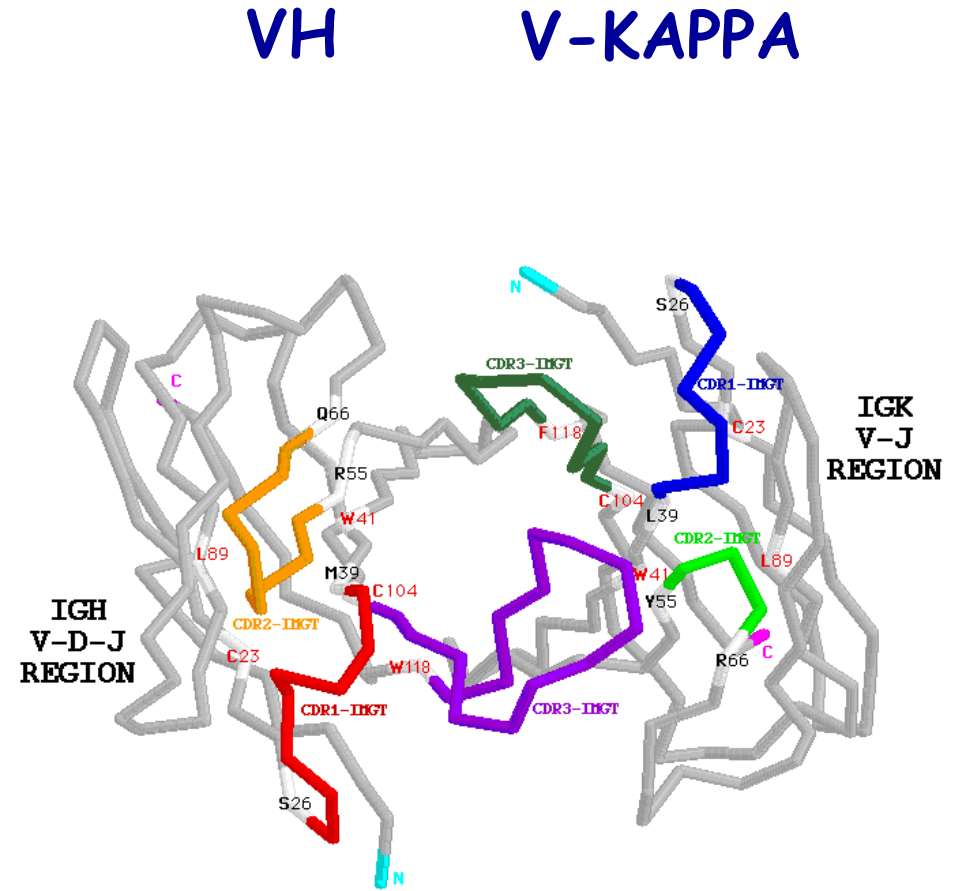**VH** **V-KAPPA**

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

VH **V-KAPPA**

Side view of the V-DOMAINs

View from above the CDRs

Mouse *(Mus musculus)* **E5.2Fv**

CDR-IMGT= Complementarity determining region (en couleurs)
FR-IMGT= Framework region (en gris)