Analyse des génomes

Fatena Bellahcene, Xavier Brochet,

Laetitia Régnier, Marie-Paule Lefranc

M2 Bioinformatique

Montpellier, 13,14 et 16 janvier 2009

http://www.imgt.org

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

http://www.imgt.org

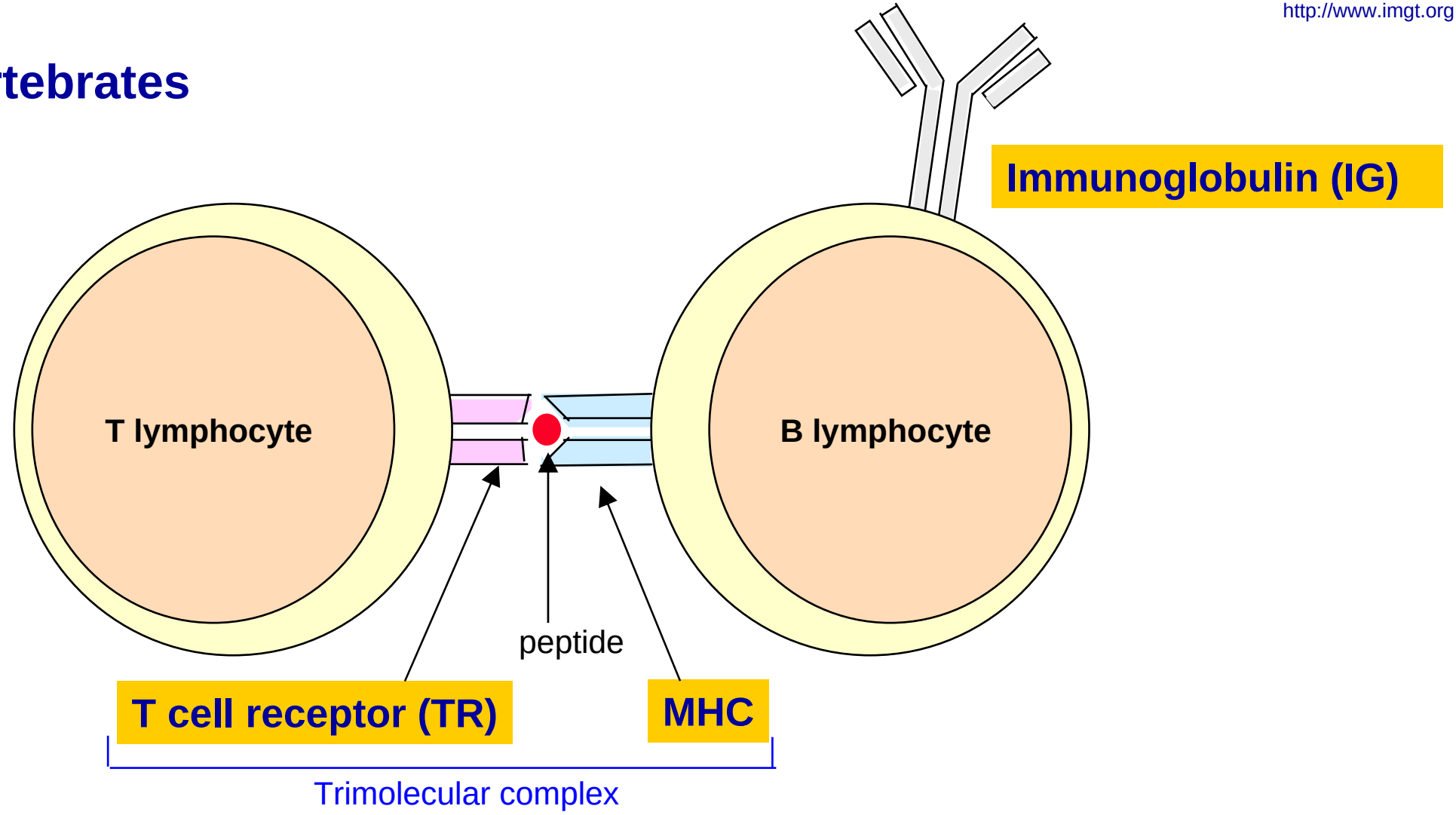
IMGT®, the international ImMunoGeneTics information system®

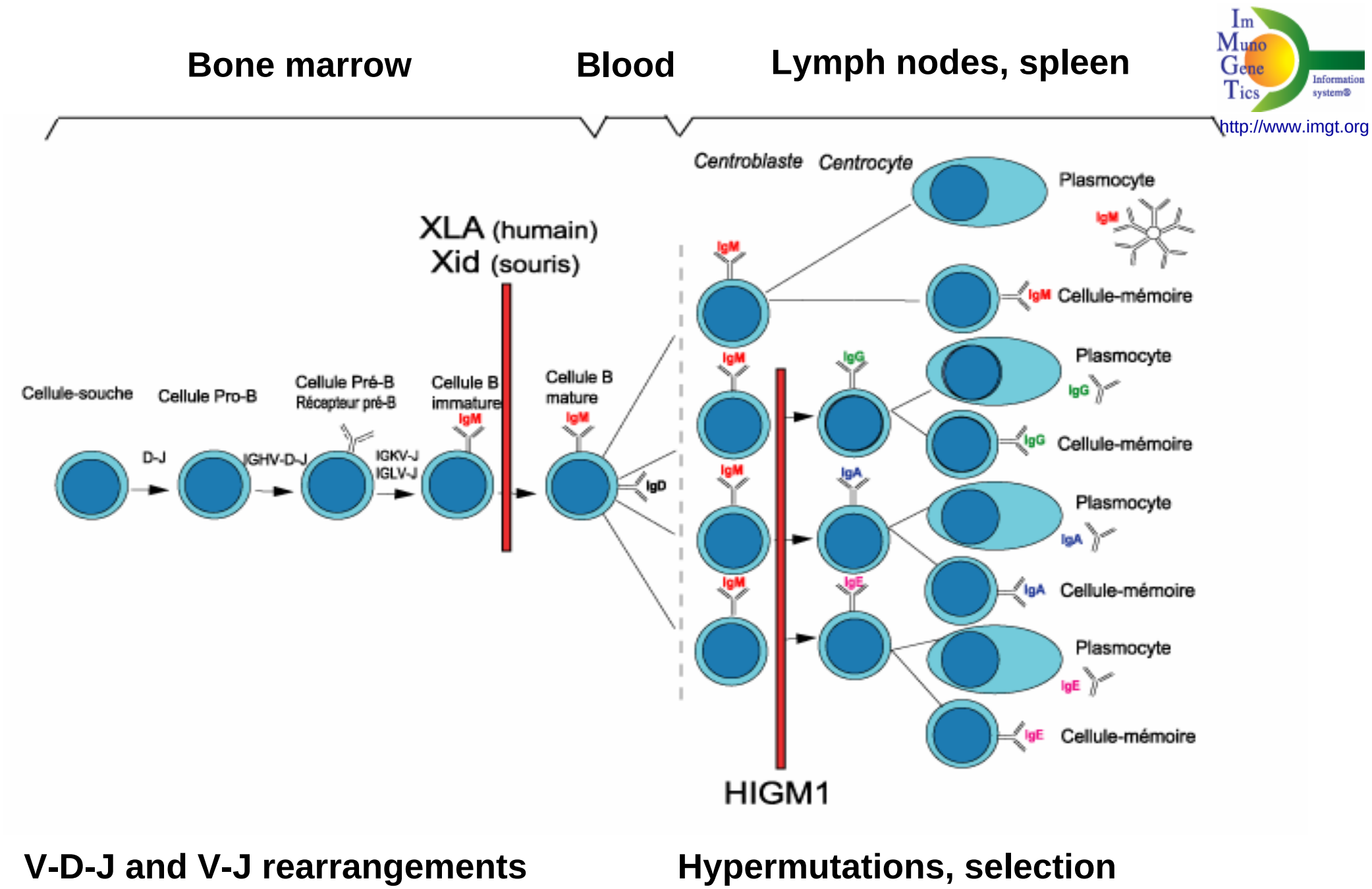
Pourquoi IMGT® a-t-il été créé?

IMGT® domain: the adaptive immune response

http://www.imgt.org

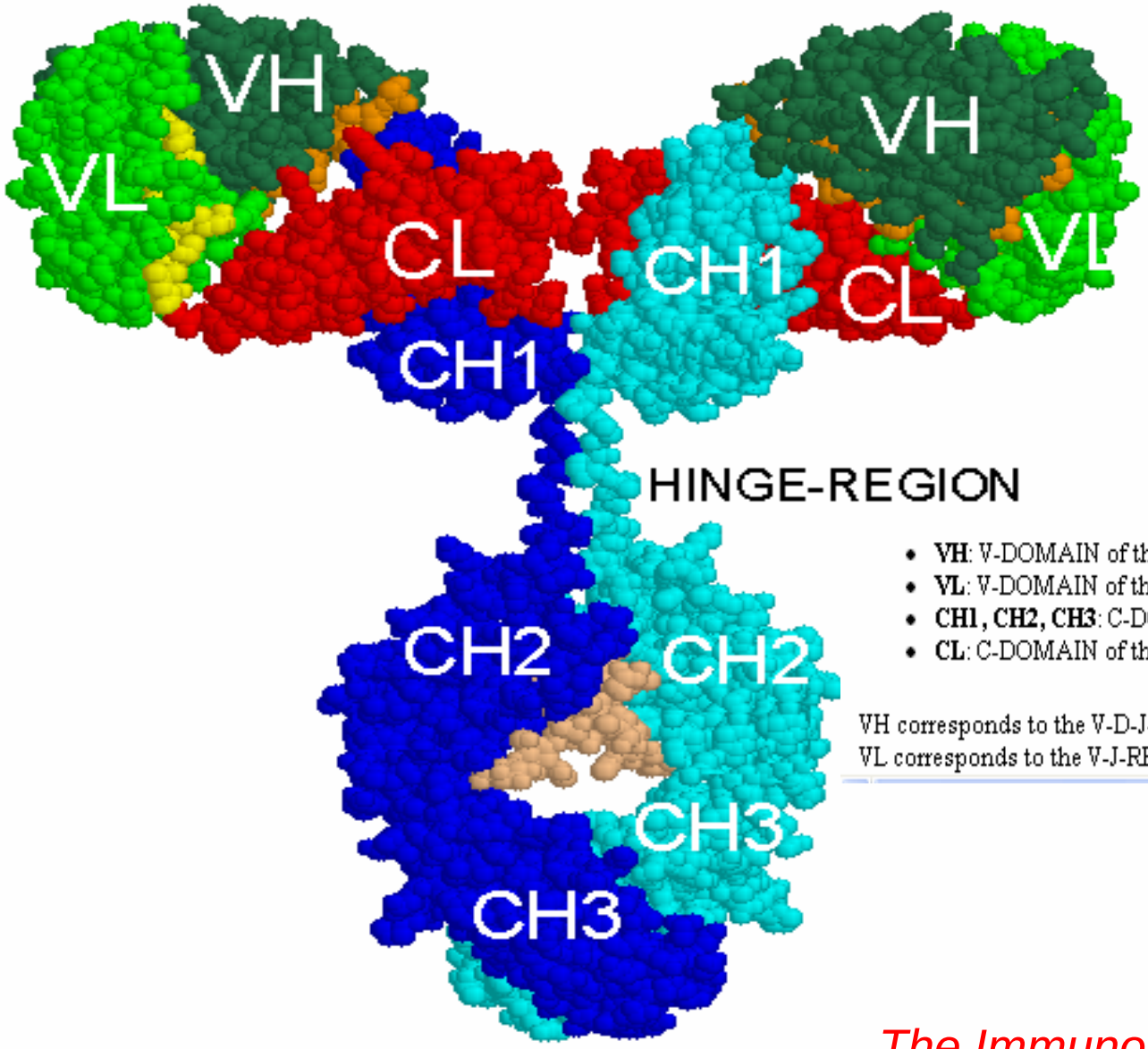
Vertebrates




Bone marrow
Blood
Lymph nodes, spleen
http://www.imgt.org
Centroblaste
Centrocyte
Plasmocyte
IgM
XLA (humain)
Xid (souris)
IgM
IgM Cellule-mémoire
Cellule-souche
Cellule Pro-B
Cellule Pré-B
Récepteur pré-B
Cellule B
immature
IgM
Cellule B
mature
IgM
D-J
IGHV-D-J
IGKV-J
IGLV-J
IgD
IgM
IgG
Plasmocyte
IgG
IgG Cellule-mémoire
IgM
IgA
Plasmocyte
IgA
IgM
IgE
IgA Cellule-mémoire
Plasmocyte
IgE
IgE Cellule-mémoire
HIGM1
V-D-J and V-J rearrangements
Hypermutations, selection

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

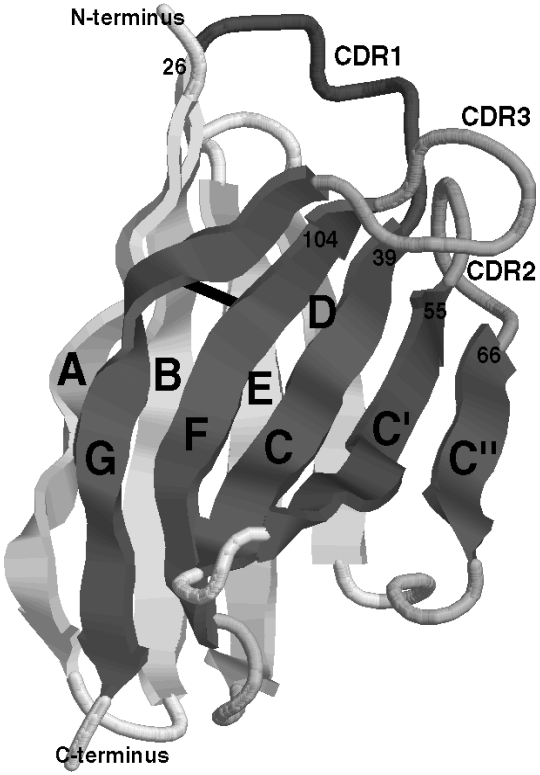
Structural domains

http://www.imgt.org

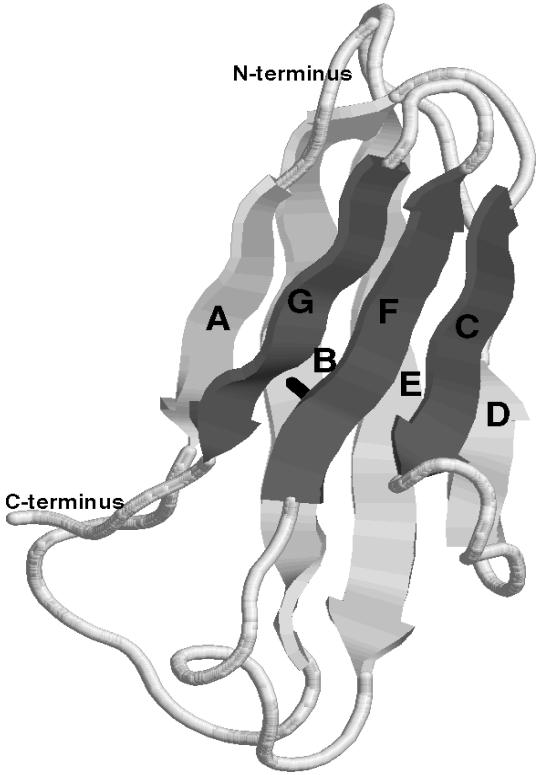
IG and TR

MHC

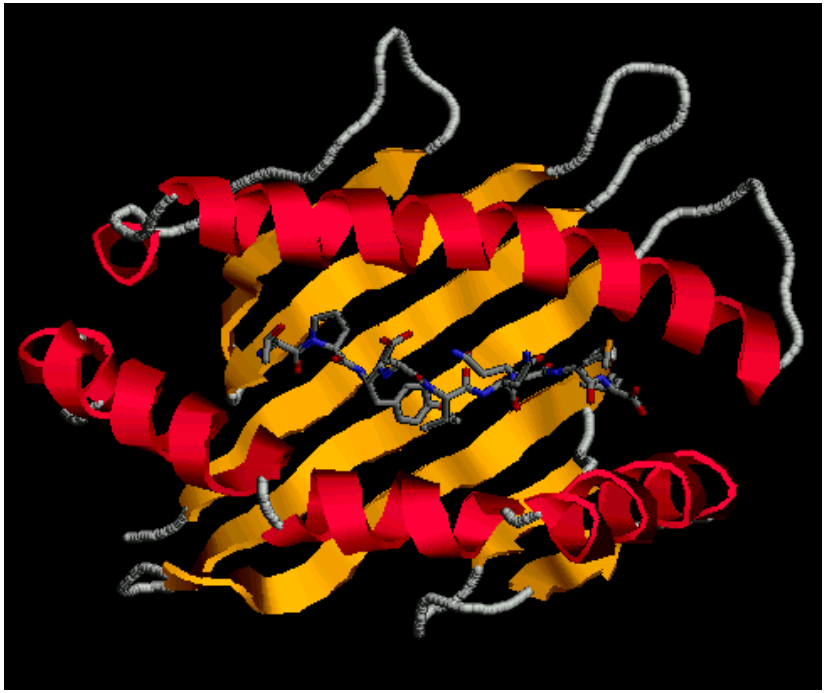
V-DOMAIN



C-DOMAIN



G-DOMAINs



Immunoglobulin (IG) | T cell receptor (TR)

http://www.imgt.org

V-DOMAIN

V-J-REGION

Light chain

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

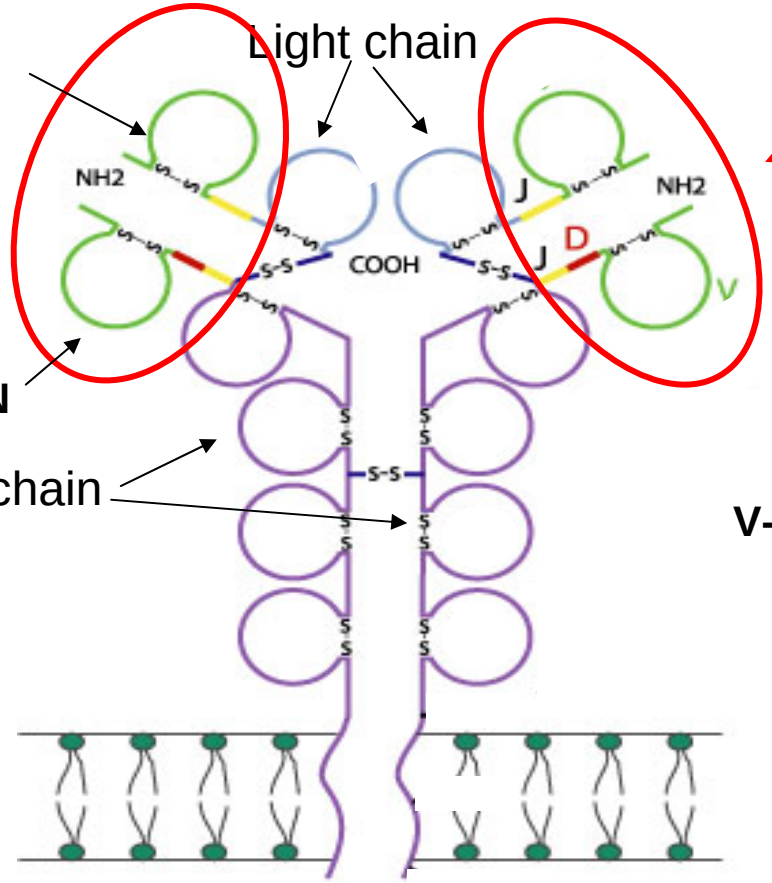
Alpha - Beta

Gamma - Delta

V-DOMAIN

V-D-J-REGION

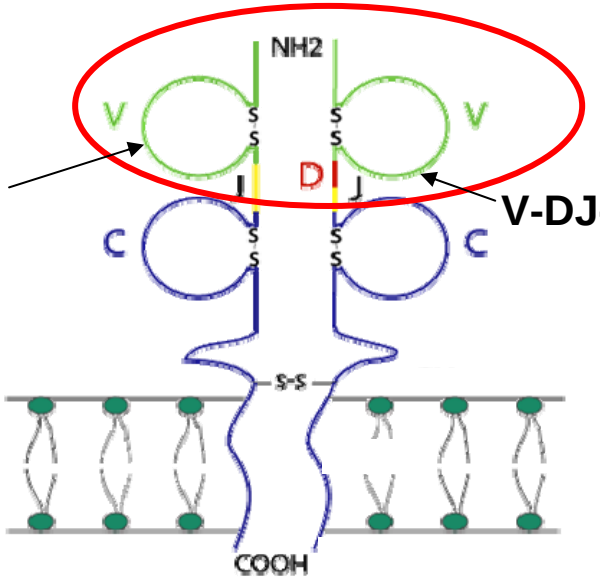
Heavy chain



Membrane IgM

V-J-REGION

V-DJ-REGION

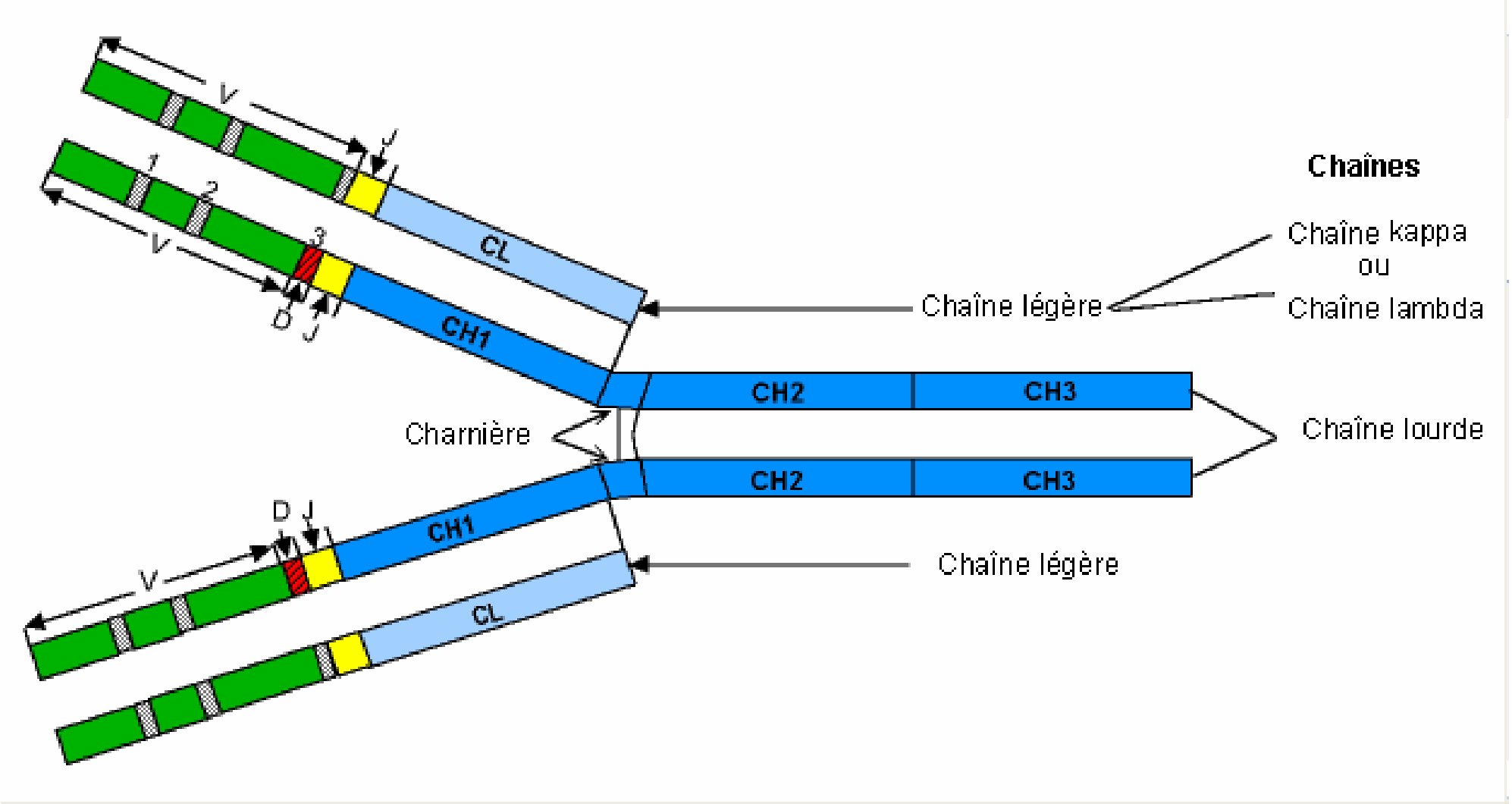


T cell receptor

The Immunoglobulin FactsBook, 2001

Immunoglobulin IgG

http://www.imgt.org

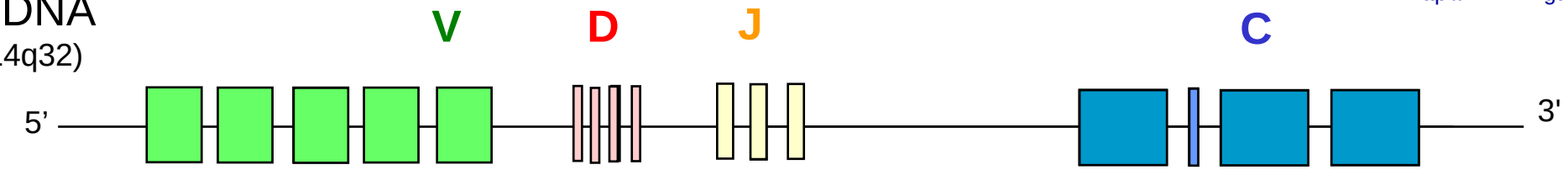


IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

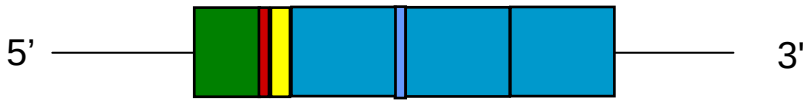
http://www.imgt.org

genomic DNA
(IGH Locus 14q32)

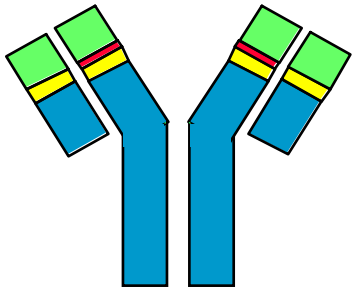


rearranged
DNA

mRNA



2 x 10^{12} different IG
per individual



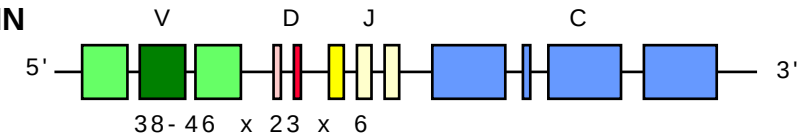
IMGT Repertoire, http://www.imgt.org

Immunoglobulin (IG) synthesis

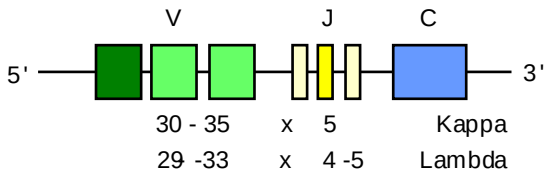
http://www.imgt.org

150
FUNCTIONAL IG GENES

HEAVY CHAIN



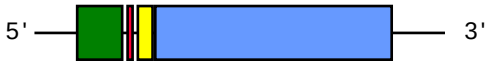
LIGHT CHAIN



6300 **POTENTIAL RECOMBINATIONS**

185 +165 **POTENTIAL RECOMBINATIONS**

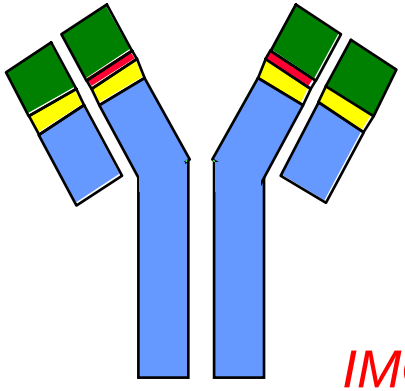
N-DIVERSITY
SOMATIC MUTATIONS
x 1000



ABOUT **6.3×10^6** POSSIBILITIES

ABOUT **3.5×10^5** POSSIBILITIES

2×10^{12}
DIFFERENT ANTIBODIES

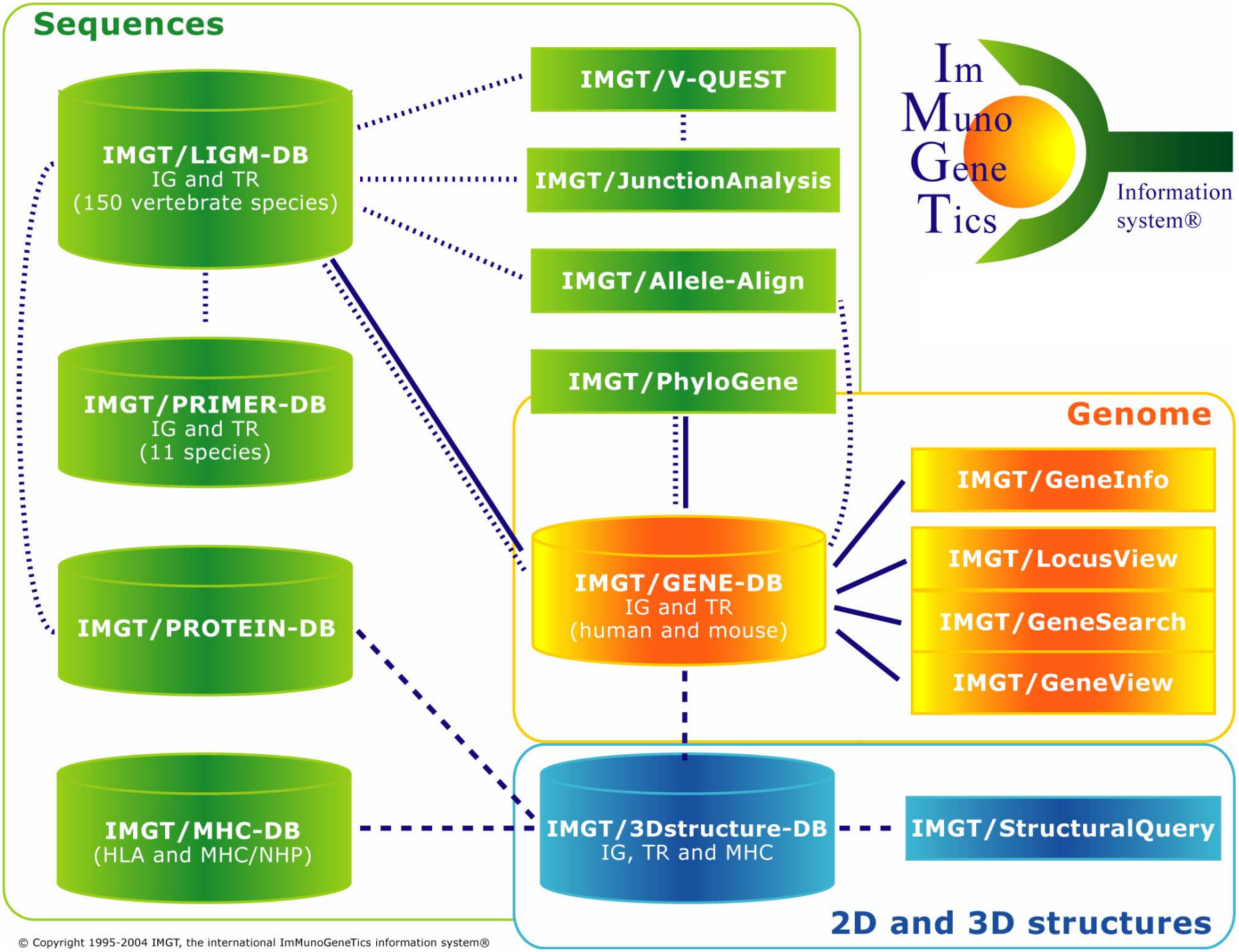


IMGT Repertoire, http://www.imgt.org

http://www.imgt.org

IMGT®

- 6 Bases de données
- 15 outils
- 10.000 pages de ressources Web



IMGT® Ressources Web

http://www.imgt.org

IMGT Repertoire 10.000 pages HTML

- Chromosomal localizations
- Locus representations
- Gene tables
- Alignments of alleles
- Protein displays...

IMGT Bloc-notes

- Interesting links…

The IMGT Immunoinformatics page

- Resources…

The IMGT Biotechnology page

- Clinical monoclonal antibodies…

IMGT Education

- Tutorials, IMGT Lexique…

http://www.imgt.org

Axiomes et concepts d'IMGT-ONTOLOGY

http://www.imgt.org

IMGT-ONTOLOGY seven axioms:

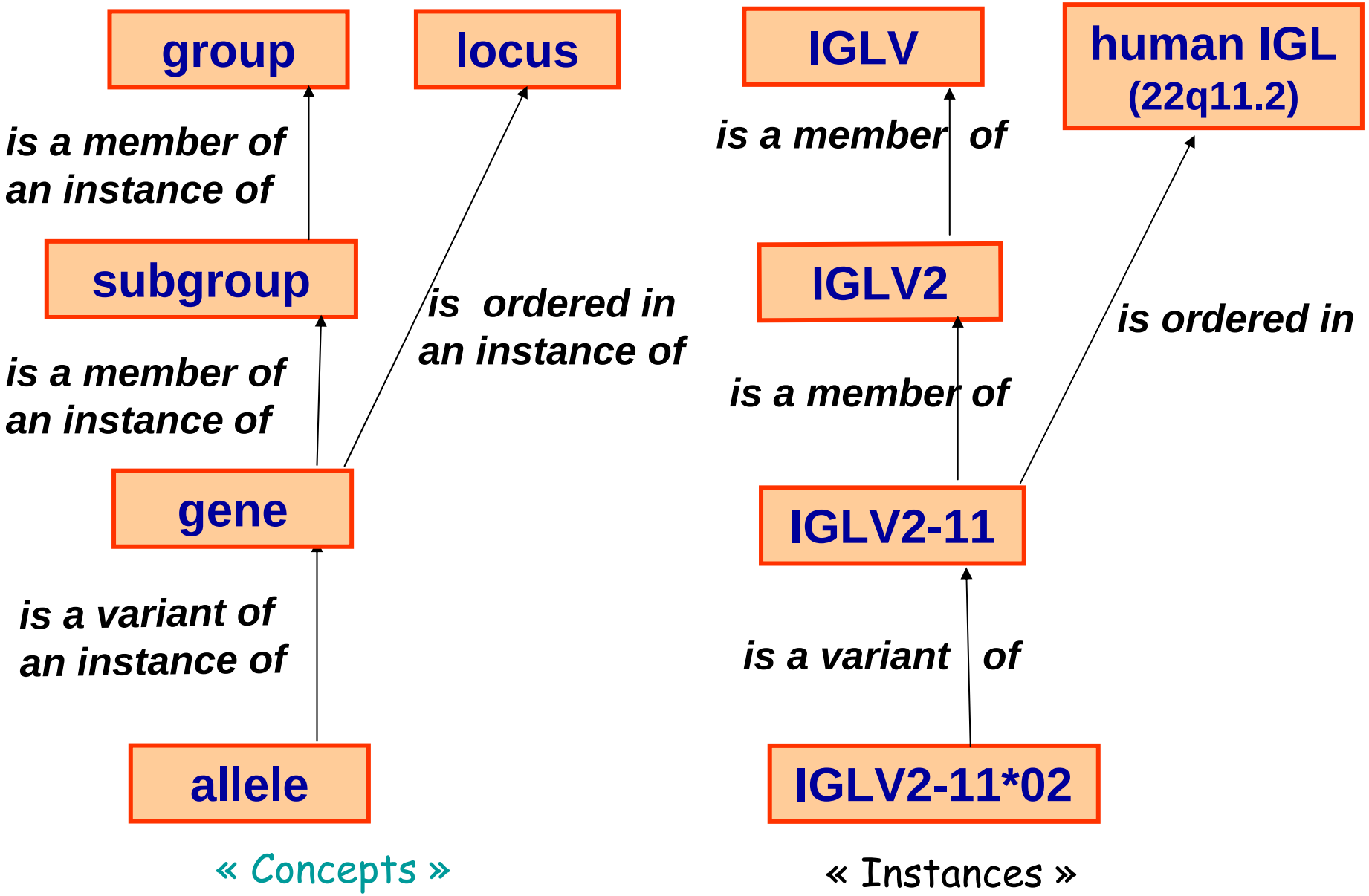
To share, reuse and represent knowledge in Immunogenetics and Life Sciences

- IDENTIFICATION
- CLASSIFICATION
- DESCRIPTION
- NUMEROTATION
- LOCALIZATION
- ORIENTATION
- OBTENTION

Giudicelli and Lefranc, Bioinformatics 1999

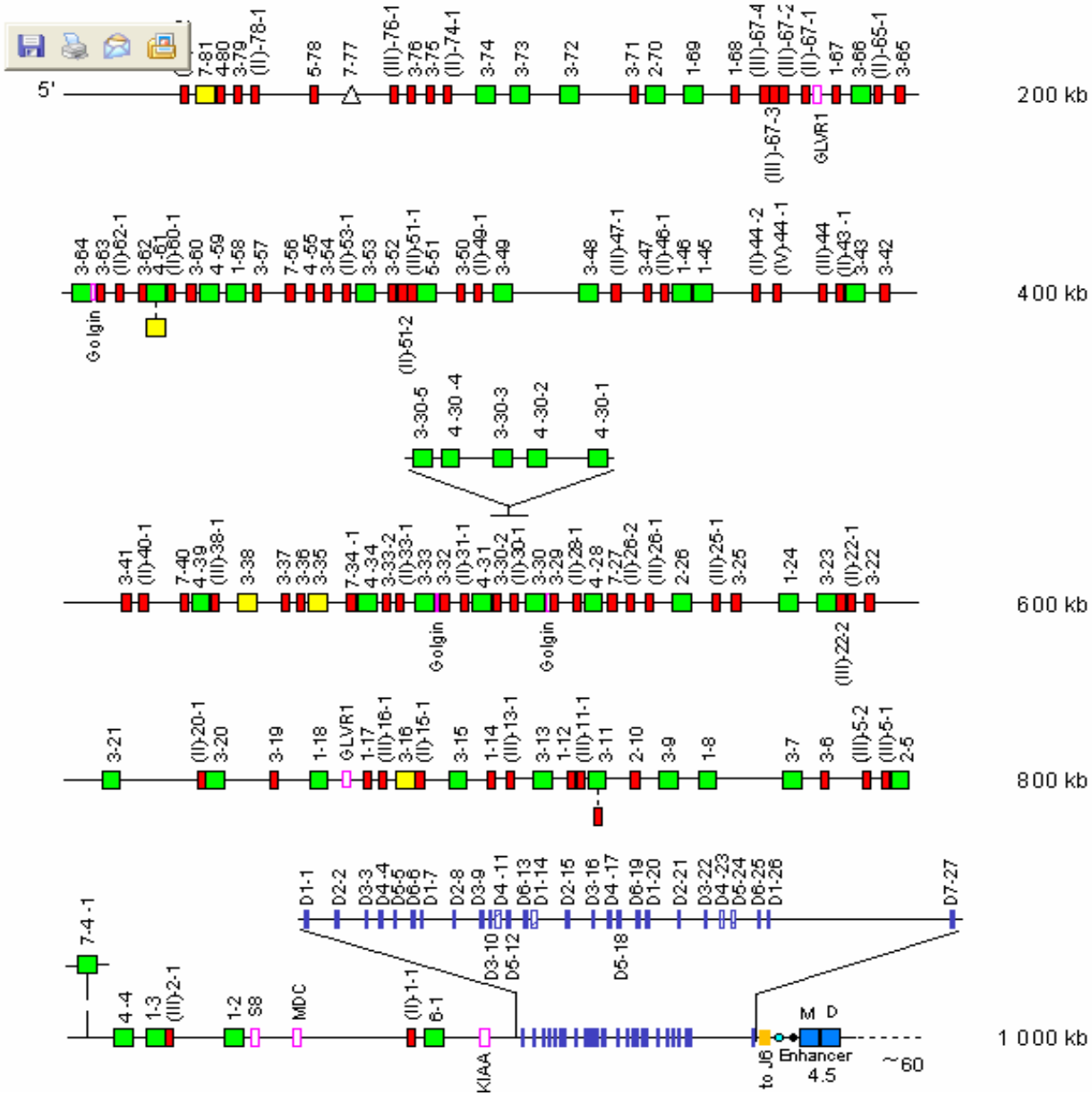
CLASSIFICATION axiom

http://www.imgt.org



Human IGH locus

Chromosome 14q32.33



http://www.imgt.org

IMGT Marie-Paule page: IMGT repertoire - Netscape

Fichier Edition Afficher Aller Communicator Aide

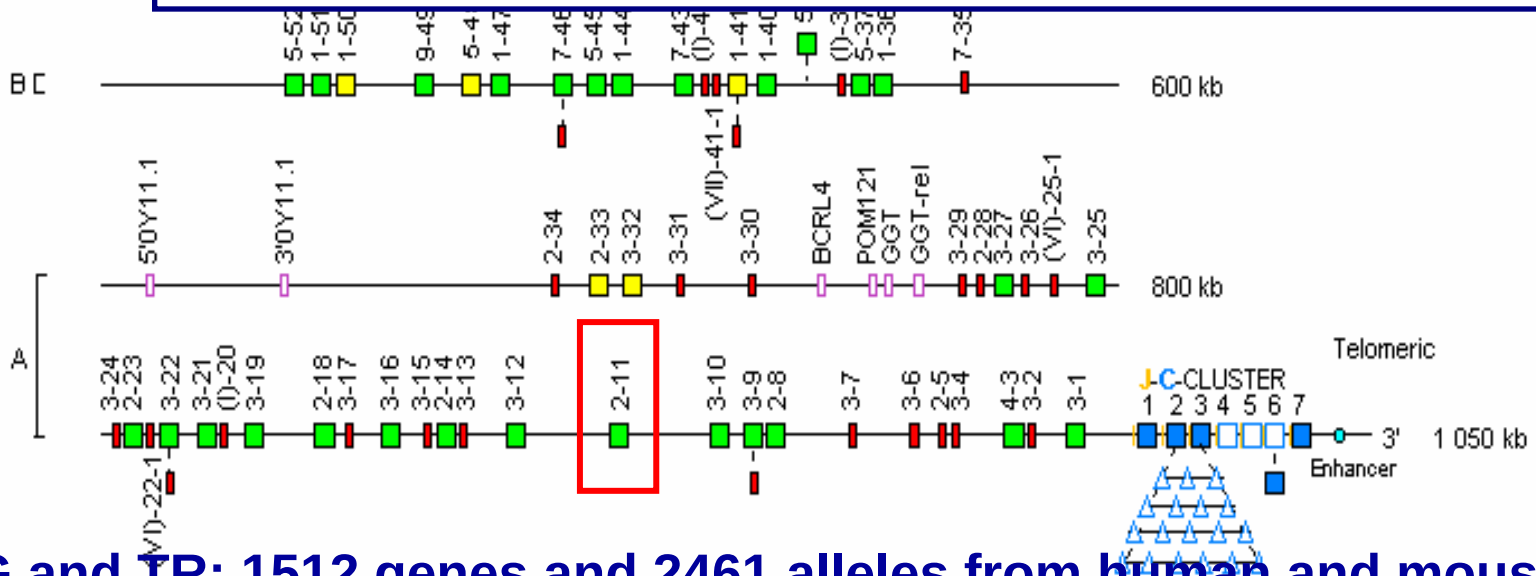
Locus representation: Human IGL

Human IGL 2

WELCOME !
to IMGT/GENE-DB

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Im Muno Gene Tics
Information system®



IG and TR: 1512 genes and 2461 alleles from human and mouse

Document : chargé

CLASSIFICATION axiom

http://www.imgt.org

- 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).

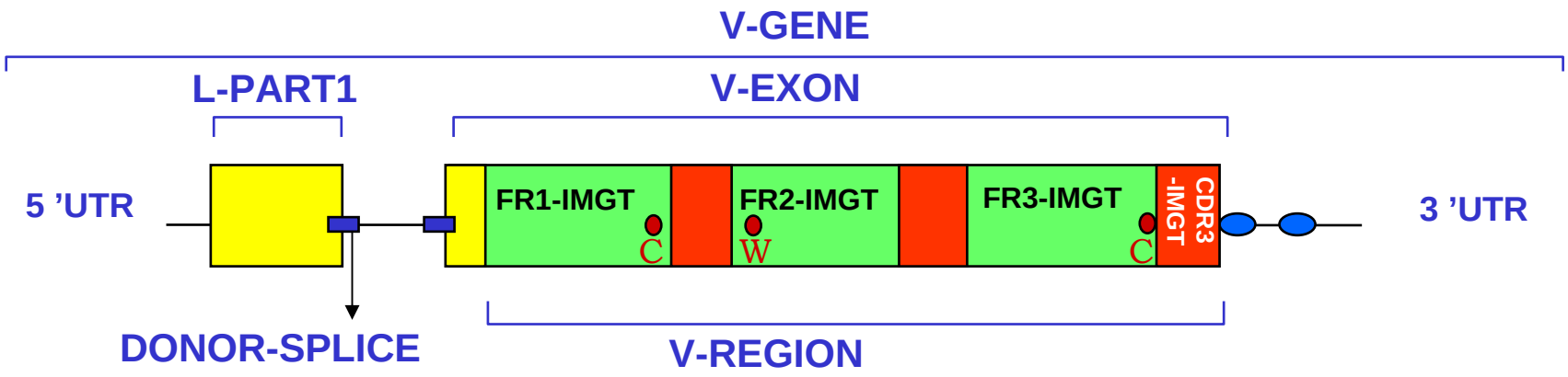
- IMGT gene names were approved by HGNC in 1999 and entered in GDB, LocusLink and Entrez Gene (NCBI).

- 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*).

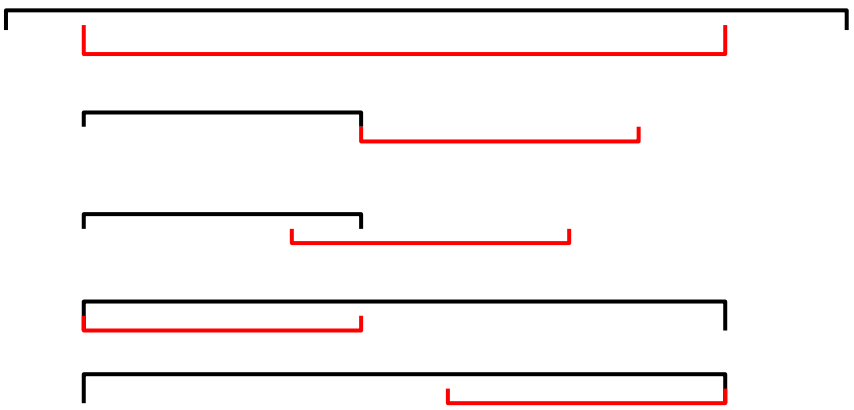
DESCRIPTION axiom

PROTOTYPE for a V-GENE



Label 1	Label 2
V-GENE	V-EXON
FR3-IMGT	CDR3-IMGT
L-PART1	DONOR-SPLICE
V-REGION	FR1-IMGT
V-REGION	CDR3-IMGT

Relations entre Labels

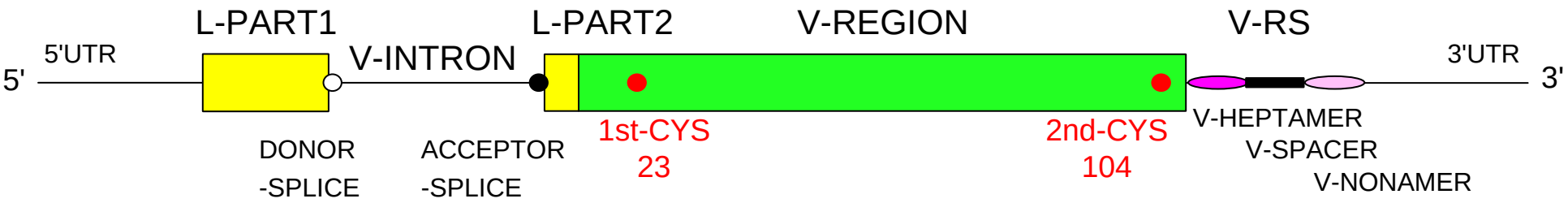


An example of V-GENE

http://www.imgt.org

>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
```

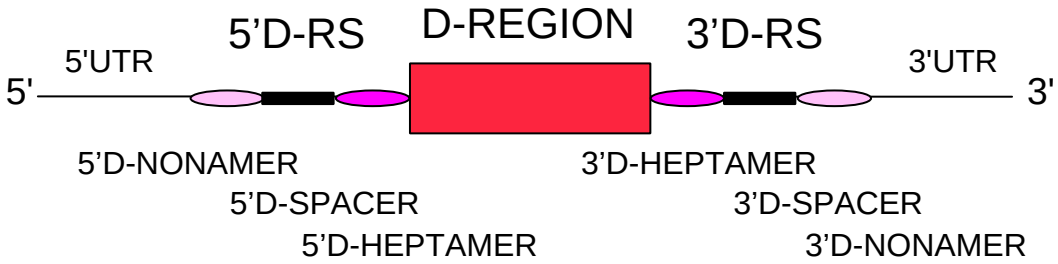


An example of D-GENE

http://www.imgt.org

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

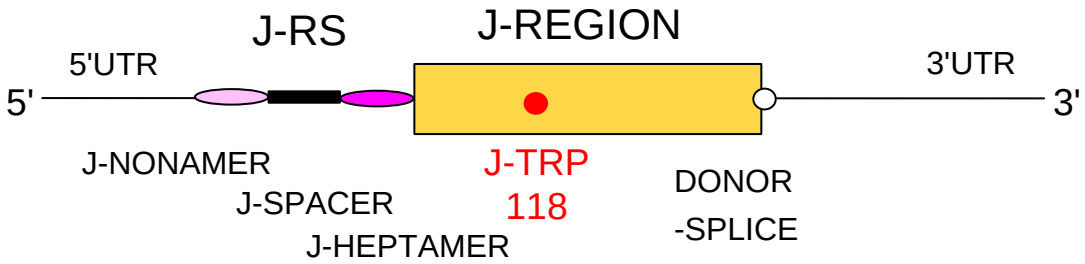
```
ccagccgcag ggtttttggc tgagctgaga accactgtgc taactgggga cacagtgatt      60
ggcagctcta caaaaaccat gctcccccgg g
```



An example of J-GENE

http://www.imgt.org

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

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



Heavy chain **WGXG** (J-TRP)
Light chain **FGXG** (J-PHE)

IMGT/LIGM-DB

http://www.imgt.org

IMGT/LIGM-DB Consultation module v3 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

```
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_length="3[illegible]"
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

DESCRIPTION

Avril 2007: 105 188 sequences from 150 species
Mars 2008: 121 983 sequences from 221 species
Janv. 2009: 129 644 sequences from 229 species

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

SO (Sequence ontology):
67 IMGT labels

DESCRIPTION axiom

http://imgt.cines.fr

- 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.

http://www.imgt.org

IMGT/V-QUEST

V-QUEST Search Page - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

WELCOME !
to the IMGT/V-QUEST Search page

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

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

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

Analyse your Immunoglobulin nucleotide sequences

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

Analyse your T cell Receptor nucleotide sequences

- Human
- Mouse
- Non-human 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

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

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

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

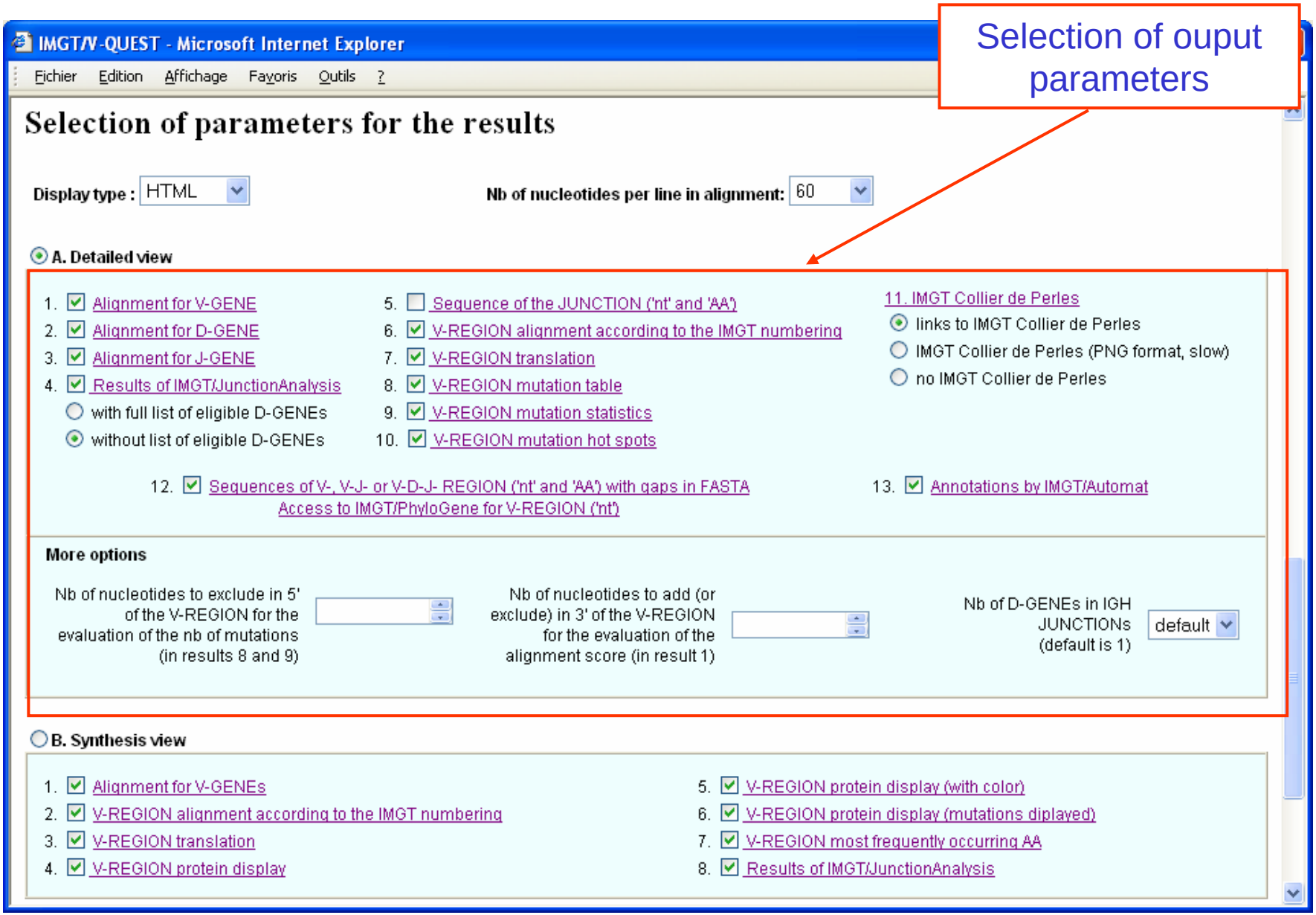
```
>AY393054
gctgggttttccttgttgctattttaaaaggtgtccaatgtgaggtgcagctggtggagtctgggggaggcttggtacagccagggcgg
tccctgagactctcctgtgcagcttctggattgacctttggtgattactttatgagctggttccgccaggctccagggaagggactgga
gtgggtaggtttcattaagagcgaaacttatggtgggacaacagaatacgccgcgtctgtgaaaggcagattcatcatctcgagagatg
attccaaaagcatcgcctatttgcaaatgaacagcctggaaaccgaggacacagccatatattactgtagtcccgggggtagtgcttat
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):

Parcourir...

Start | Clear the form

Internet



Selection of ouput parameters

http://www.imgt.org

IMGT/V-QUEST - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.imgt.org

Sequences aligned with IGHV5-51*03

1. Alignement for V-GENE

```
                                <---------------------------------- FR1-IMGT ---------------
X56368 IGHV5-51*03              gaggtgcagctggtgcagtctggagca...gaggtgaaaaagccgggggagtctctgaag
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              tgggtgcgccagatgcccgggaaaggcctggagtggatggggatcatctatcctggtgac
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----------
```

IMGT/V-QUEST - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

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://www.imgt.org

http://www.imgt.org

IMGT/V-QUEST - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

3. V-REGION translation

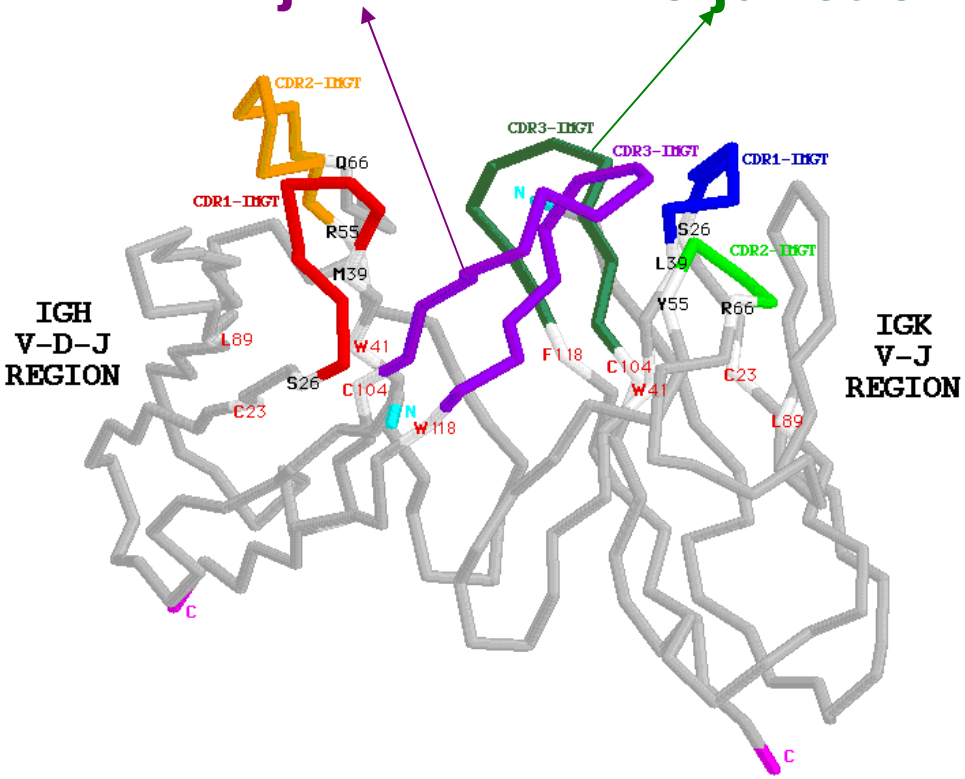
```
                                <-------------------------------------------  FR1 - IMGT
                                1               5                   10                  15
                                 E   V   Q   L   V   Q   S   G   A       E   V   K   K   P
X56368 IGHV5-51*03              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://www.imgt.org

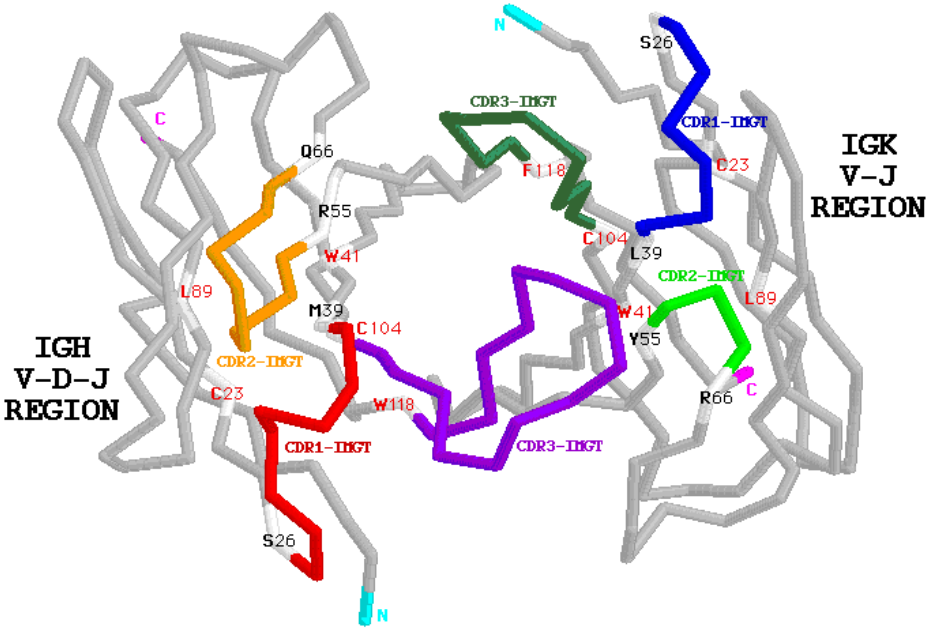
VH V-KAPPA

V-D-J junction V-J junction



Side view of the V-DOMAINs

VH V-KAPPA



View from above the CDRs

Mouse(Mus musculus) E5.2Fv

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