IMGT®, the international ImMunoGeneTics information system®
http://www.imgt.org/
Laboratoire d'ImmunoGénétique Moléculaire, LIGM
Institut de Génétique Humaine, IGH, UPR CNRS 1142
Université Montpellier 2

Joumana Jabado-Michaloud

« Analyse de génome »

1

Analyse de génome

1- Présentation IMGT®, the international ImMunoGeneTics information system®

2- Rappel système immunitaire

3- Synthèse des immunoglobulines

- Origine de la diversité des domaines variables des anticorps

4- IMGT-ONTOLOGY

5- Analyse de génome

Les immunoglobulines:

- Chromosomes et locus
- Structure des gènes : V-GENE, D-GENE, J-GENE et C-GENE

6- Quelques Bases de données généralistes et serveurs

2

WELCOME !
to the IMGT Home page

THE INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

Im
MunO
GenE
Tics   Information
                       system®

http://www.imgt.org

IMGT®, the international ImMunoGeneTics information system® http://www.imgt.org, is the global reference in immunogenetics and immunoinformatics, created in 1989 by Marie-Paule Lefranc (Université Montpellier 2 and CNRS). IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility (MH) of human and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MhSF) and related proteins of the immune system (RPI) of vertebrates and invertebrates. IMGT® provides a common access to sequence, genome and structure Immunogenetics data, based on the concepts of IMGT-ONTOLOGY and on the IMGT Scientific chart rules. IMGT® works in close collaboration with EBI (Europe), DDBJ (Japan) and NCBI (USA). IMGT® consists of sequence databases, genome database, structure database, and monoclonal antibodies database, **Web resources** and **interactive tools**.

IMGT founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr), Université Montpellier 2, CNRS, LIGM, IGH, SFR, Montpellier (France)

IMGT® Site Map
Information on IMGT®, IMGT creations and updates, IMGT references, FAQ, Citing IMGT, Acknowledgments and Funding support
IMGT Livre d'or
The 2010 IMGT® Customer Satisfaction Survey

The Quality Management System of IMGT® Montpellier France has been approved by Lloyd's Register Quality Assurance France SAS to the following Quality Management System Standard: ISO 9001:2008

IMGT databases

- **IMGT/LIGM-DB** (doc) LIGM, Montpellier, France
  Nucleotide sequences of IG and TR from 331 species (**168 866 entries**)
- **IMGT/MH-DB** ANRI, BPRC, hosted at EBI
  Sequences of the human MH (HLA)
- **IMGT/PRIMER-DB** (doc) LIGM, Montpellier, France
  Oligonucleotides (primers) of IG and TR from 11 species (**1 864 entries**)
- **IMGT/CLL-DB** (bylaws) LIGM, Montpellier, France
  IG sequences from CLL, an initiative of the IMGT/CLL-DB group

- **IMGT/GENE-DB** (doc) LIGM, Montpellier, France
  International nomenclature for IG and TR genes from human, mouse, rat and rabbit (**2 956 genes, 4 471 alleles**)

- **IMGT/3Dstructure-DB and IMGT/2Dstructure-DB** (doc) LIGM, Montpellier, France
  3D structures (IMGT Colliers de Perles) of IG antibodies, TR, MH and RPI (**2 763 entries**)
  Source: PDB, INN, Kabat

- **IMGT/mAb-DB** (doc) LIGM, Montpellier, France
  Monoclonal antibodies (IG, mAb) and fusion proteins for immune applications (FPIA) (**426 entries**)

IMGT Web resources

- **IMGT Repertoire** (IG and TR, MH and RPI)
- **IMGT Scientific chart** (Sequence description, Numbering, Nomenclature, Representation rules)

- **IMGT Index** (FactsBook)
- **IMGT Bloc-notes** (Interesting links, PubMed, Meeting announcements, Postdoctoral positions and jobs, Messages, Search engines...)
- **IMGT Education** (IMGT Lexique, Aide-mémoire, Tutorials, Questions and answers, Enseignements...)
- **IMGT Posters and diaporama**

- The IMGT Medical page
- The IMGT Veterinary page
- The IMGT Biotechnology page
- The IMGT Immunoinformatics page

IMGT tools

- **IMGT/V-QUEST** (doc) (sequence alignment software for IG and TR)
- **IMGT/HighV-QUEST** (doc) (NGS High-Throughput analysis of IG and TR)

IMGT other accesses

- IMGT Other accesses (ARSA, SRS, MRS)
- Compare your sequence against IMGT (BLAST, FASTA)

# IMGT®, the international ImMunoGeneTics information system®

Created in 1989
A unique integrated knowledge resource for :
• immunoglobulins (IG)
• T cell receptors (TR)
• MH
• IG superfamily (IgSF)
• MH superfamily (MhSF)
• related proteins of the immune system (RPI)

7 Databases
17 online tools

Based on IMGT-ONTOLOGY, the first ontology for immunogenetics, and on the IMGT Scientific chart rules

4
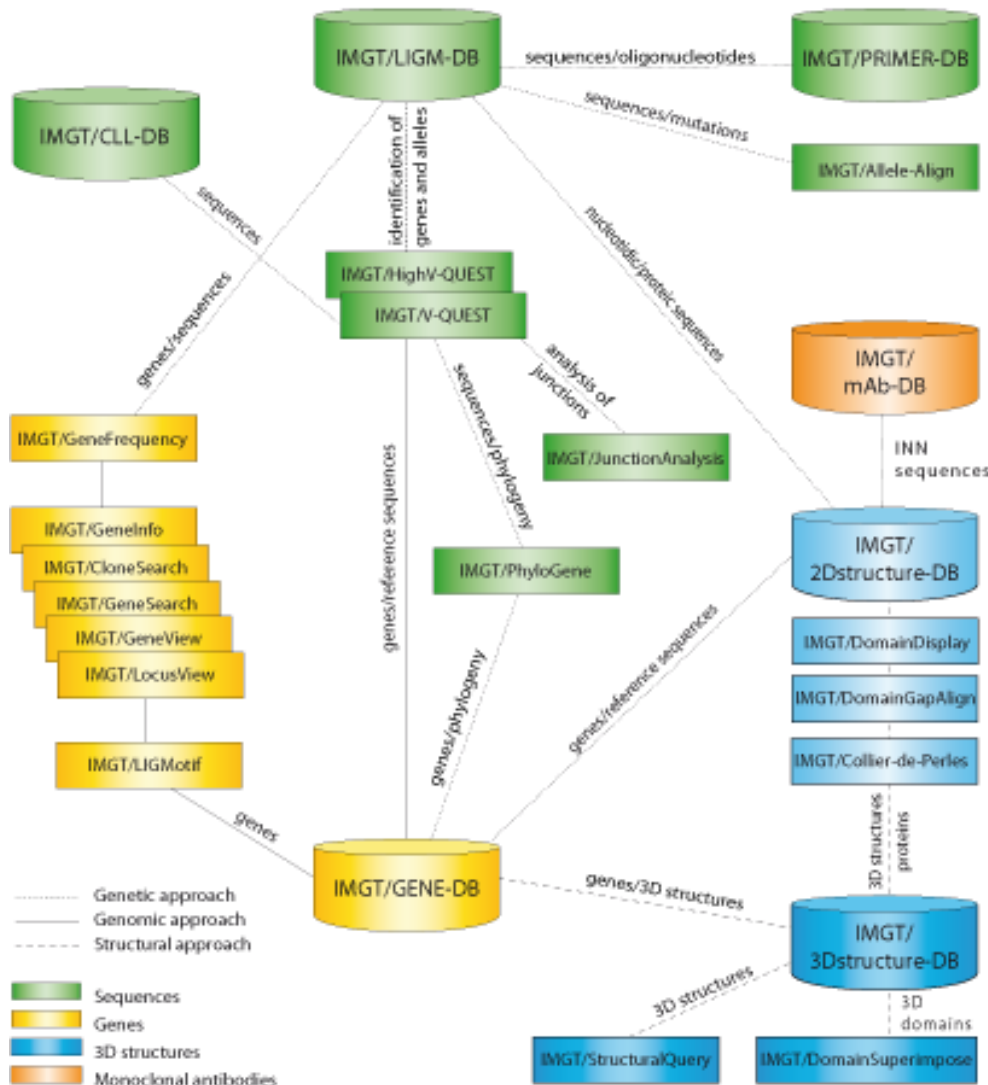
Rappel système immunitaire:

Le système immunitaire a pour rôle de maintenir l'intégrité biologique de l'organisme:
-reconnaissance et tolérance du soi
-reconnaissance et élimination du non soi

Deux types de réponses immunitaires rentrent en jeux :

I. La réponse immunitaire innée:
les barrières anatomiques, physiologiques, phagocytaires et inflammatoires. Elle constitue la première ligne de défense de l'organisme

II. La réponse immunitaire adaptative ou spécifique:
elle est apparue au cours de l'évolution il y a 450 millions d'années et elle est observée uniquement chez les vertébrés **gnathostomes** (à mâchoires). Elle est caractérisée par la **reconnaissance spécifique** d'antigène (AG) et la création **d'une mémoire immunitaire** ; cette dernière permet de lutter plus efficacement et plus rapidement lors d'une rencontre ultérieure du même antigène

Nature biologique et types de reconnaissance

Nature Reconnaissance	Immunité innée	Immunité adaptative
Cellules	Granulocytes Lymphocytes NK	**Lymphocytes T** **Lymphocytes B**
Molécules	Lectines Complement	**Immunoglobulines (IG)** **Recepteurs T (TR)** Complexes **CMH**-peptides

Lymphocytes T responsables de la réponse cellulaire
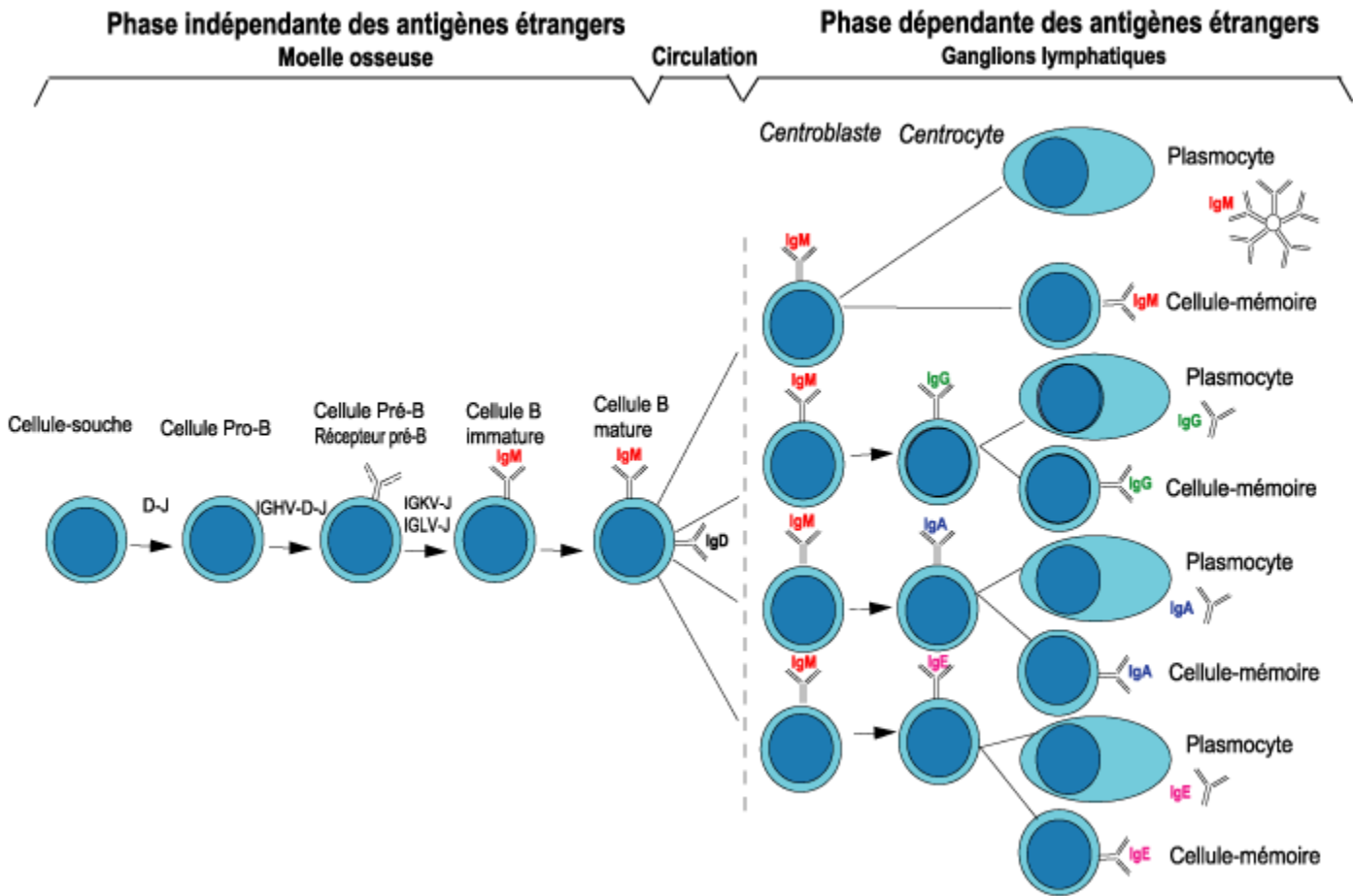Lymphocytes B responsables de la réponse Humorale

The main actors of adaptive immune response

Vertebrates

Immunoglobulin (IG)

T lymphocyte

B lymphocyte

peptide

T cell receptor (TR)
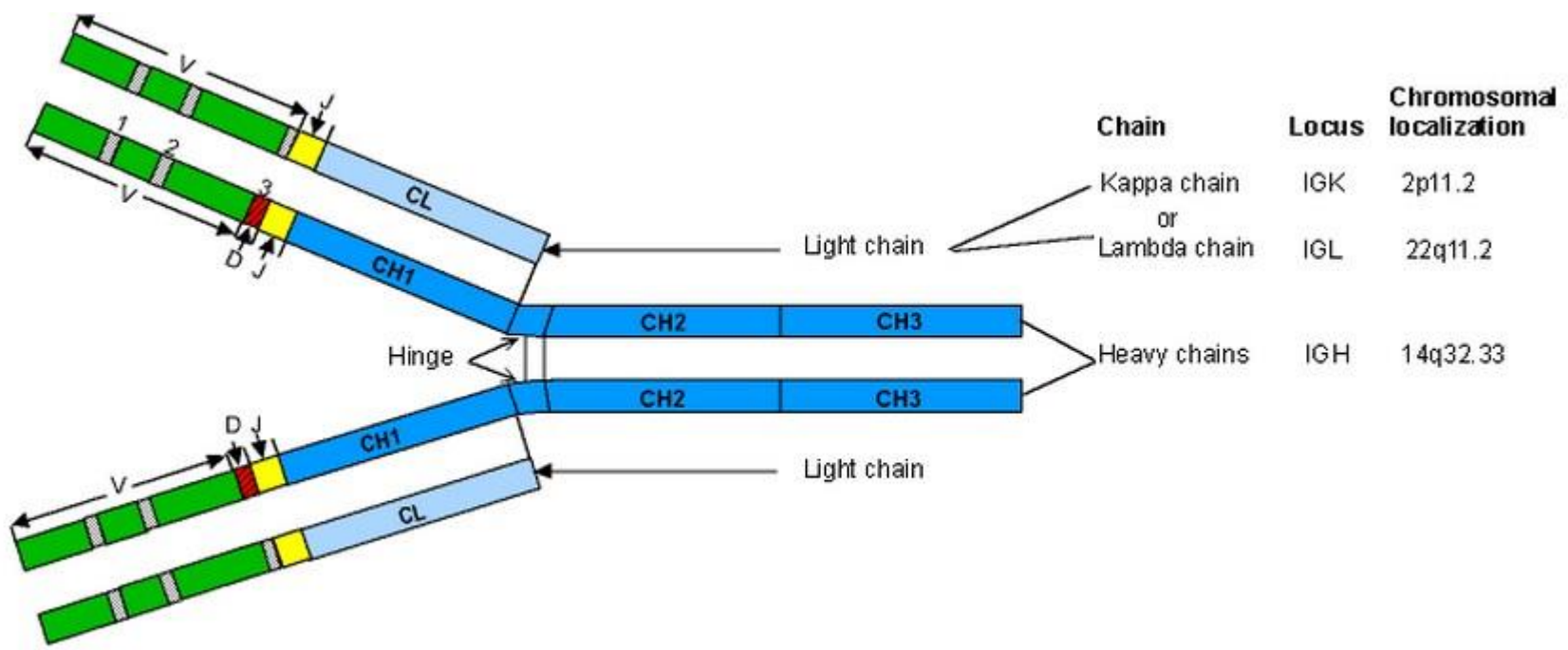
MHC

Trimolecular complex
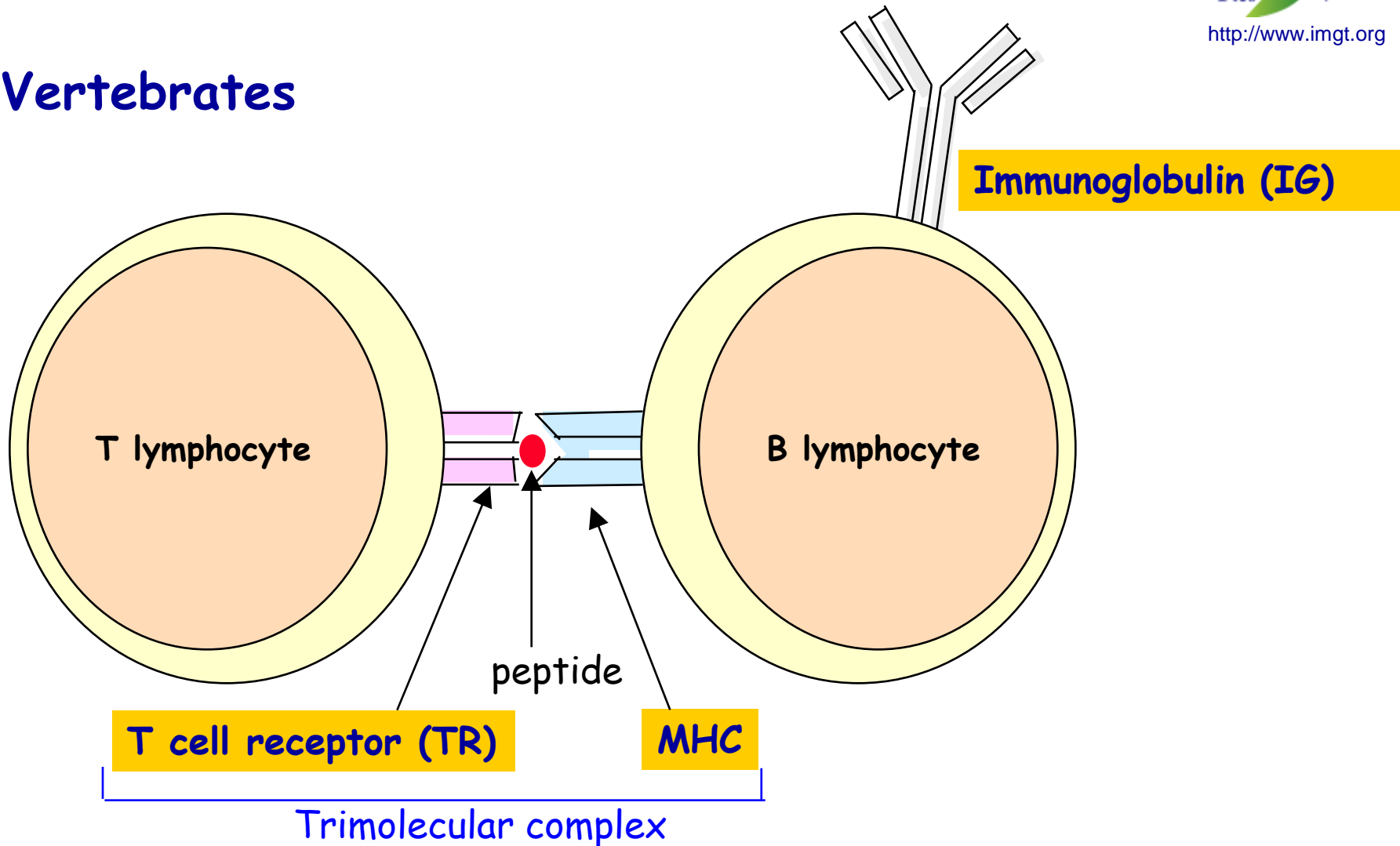
Presentation of peptides by the MHC to the T cell receptors (TR) at the surface of T cells.
➔ characterization of the TR/peptide/MHC trimolecular complexes (TR/pMHC) is crucial

7

Différenciation des lymphocytes B

Master STIC pour la santé spécialité BCD   5 Nov 2013

http://www.imgt.org

Chain	Locus	Chromosomal localization
Kappa chain or Lambda chain	IGK	2p11.2
	IGL	22q11.2
Heavy chains	IGH	14q32.33

Immunoglobuline membranaire (IGM)

receptor T (TR)

V-DOMAIN

V-J-REGION

Chaîne légère

Contribution des 2
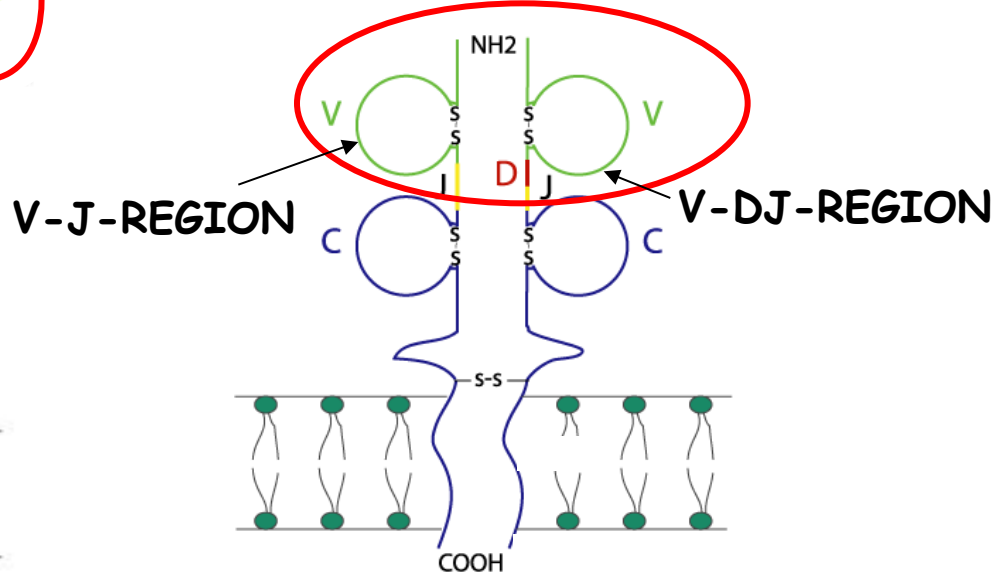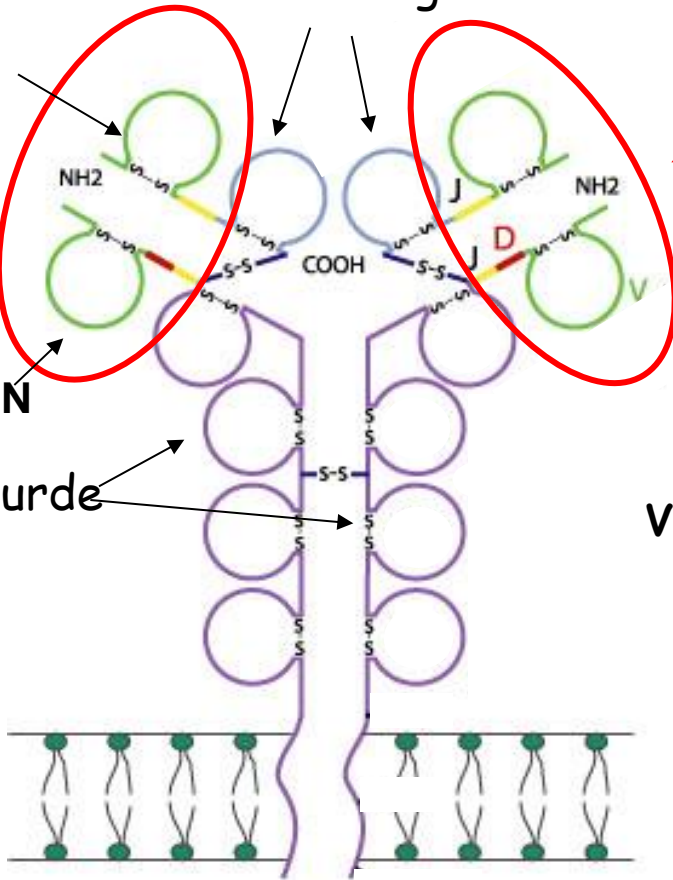2 V-DOMAINs
Au site de fixation de l'antigène

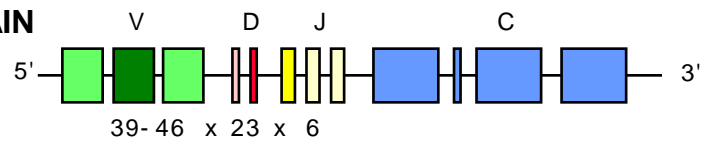**V-DOMAIN**

V-D-J-REGION

Chaîne lourde

Alpha	-	Beta
Gamma	-	Delta

NH2

V   V

D   J

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

C   C

COOH

The Immunoglobulin FactsBook, 2001

Synthèse des Immunoglobulines

Immunoglobulin (IG) synthesis

150
FUNCTIONAL IG GENES

HEAVY CHAIN

V   D   J   C

5' ——— 3'

39- 46   x 23   x   6
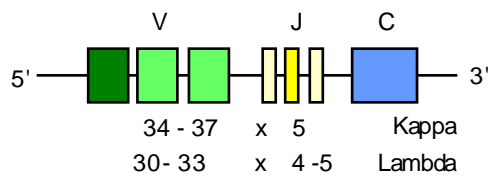
6300 **POTENTIAL RECOMBINATIONS**

5' ——— 3'

ABOUT **6.3 x 10^6** POSSIBILITIES

LIGHT CHAIN

V   J   C

5' ——— 3'

34 - 37   x   5   Kappa
30- 33   x   4 -5   Lambda

185 +165 **POTENTIAL RECOMBINATIONS**

**N-DIVERSITY
SOMATIC MUTATIONS**
x **1000**

5' ——— 3'

ABOUT **3.5 x 10^5** POSSIBILITIES

2 x 10^12

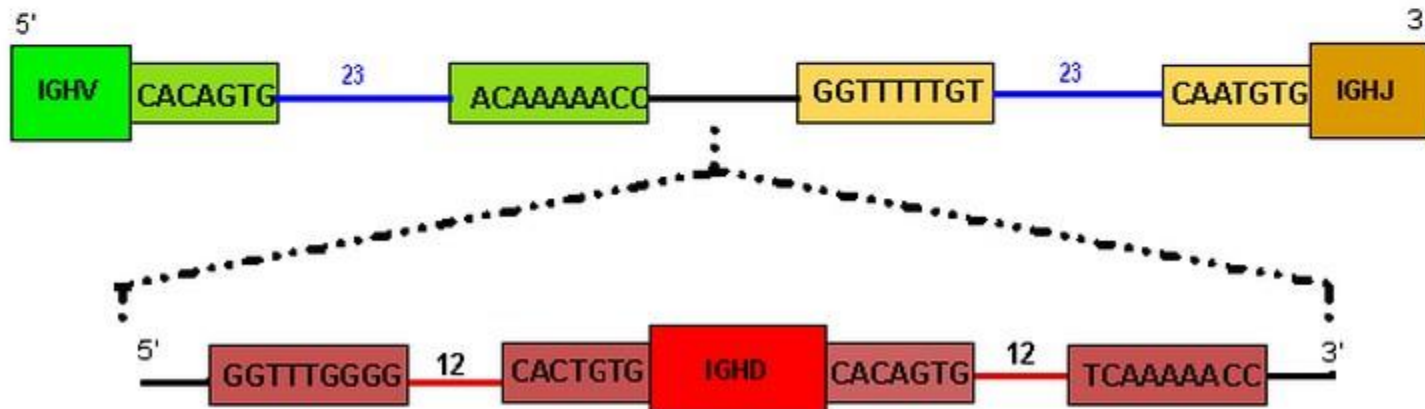## DIFFERENT ANTIBODIES
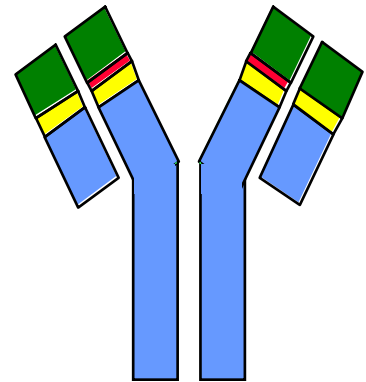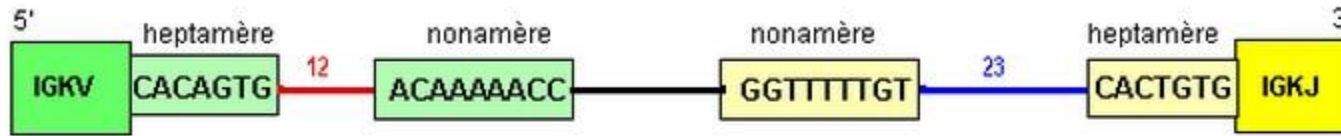
Chromosome 2p11.2

Chromosome 22q11.2
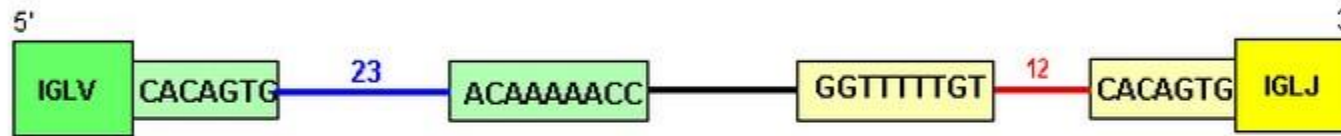
Chromosome 14q32.33

Generation of the JUNCTION diversity

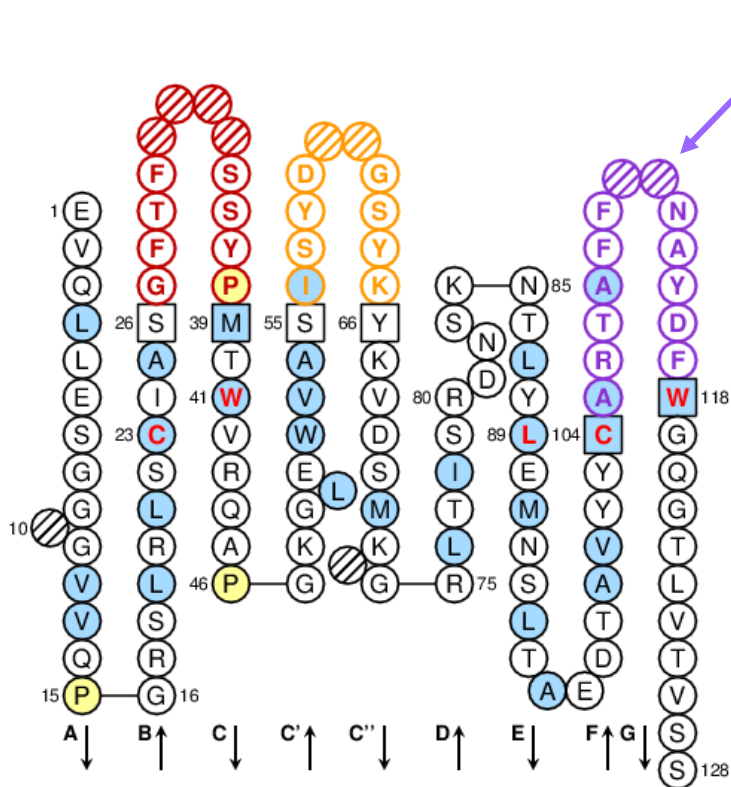| 3'V-REGION	N-REGION	D-REGION	N-REGION	5'J-REGION

tgtgcgaaga   tacc agcatattgtggtggtgactgctattcc gat acaactggttcgactcc**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 tac tcc tgg

14

N-diversity: unicity of the CDR3

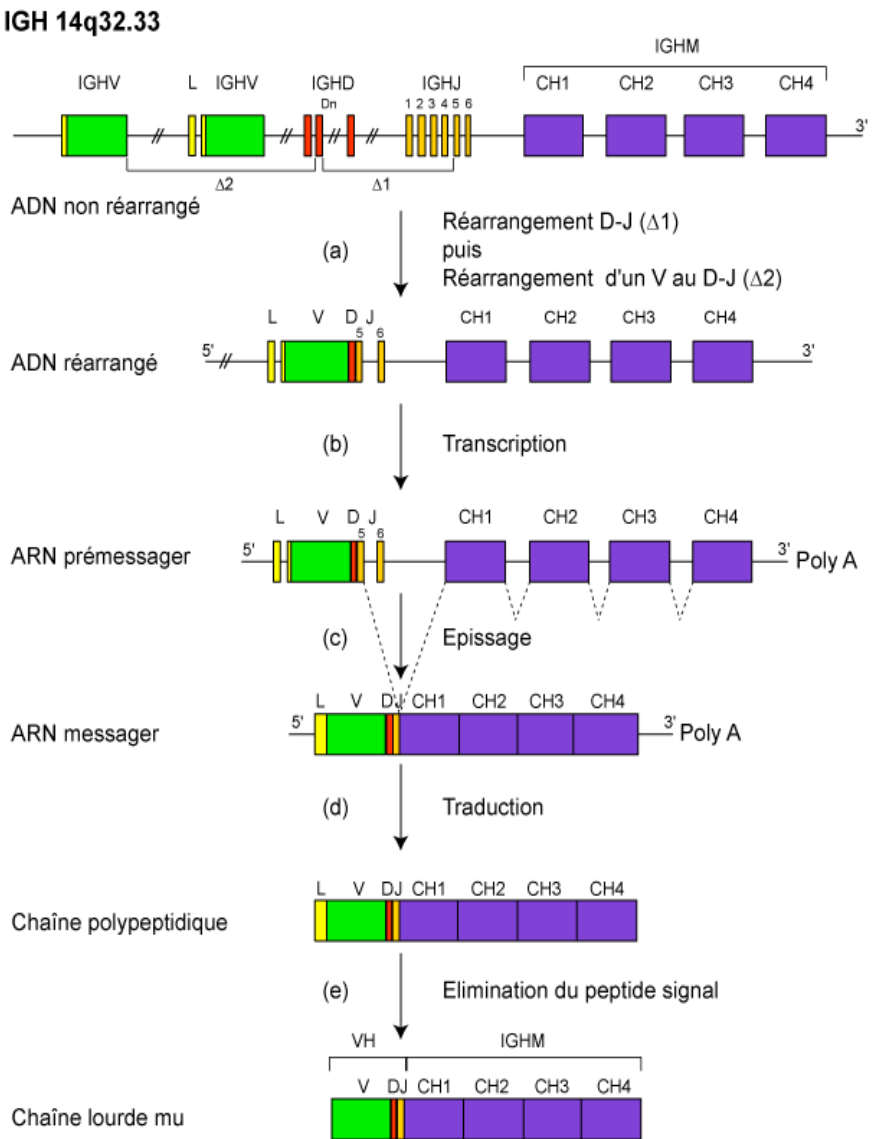| tgt	gcg	agg	aca	gct	ttc	ttt	aac	gcc	tat	gac	ttc	tgg
C	A	R	T	A	F	F	N	A	Y	D	F	W

⇒ CDR3 (complementarity determining region 3): third loop of the V-DOMAIN

⇒ Characterized by its length, its unique sequence:
- useful for diagnosis/prognosis of leukemia/lymphomas
- used as marker of clonal population

2 D representation of a V-DOMAIN
IMGT collier de Perles

Synthèse d'une chaîne lourde mu d'immunoglobuline

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

IMGT-ONTOLOGIE

Master STIC pour la santé spécialité BCD   5 Nov 2013
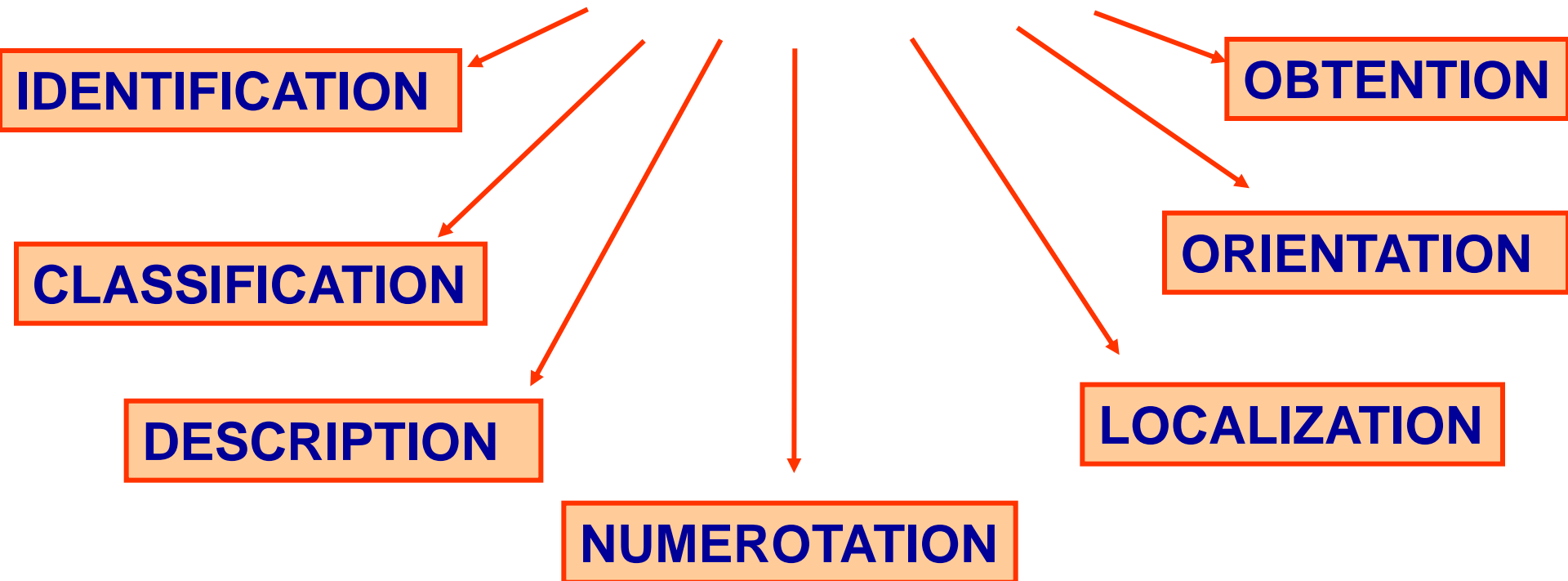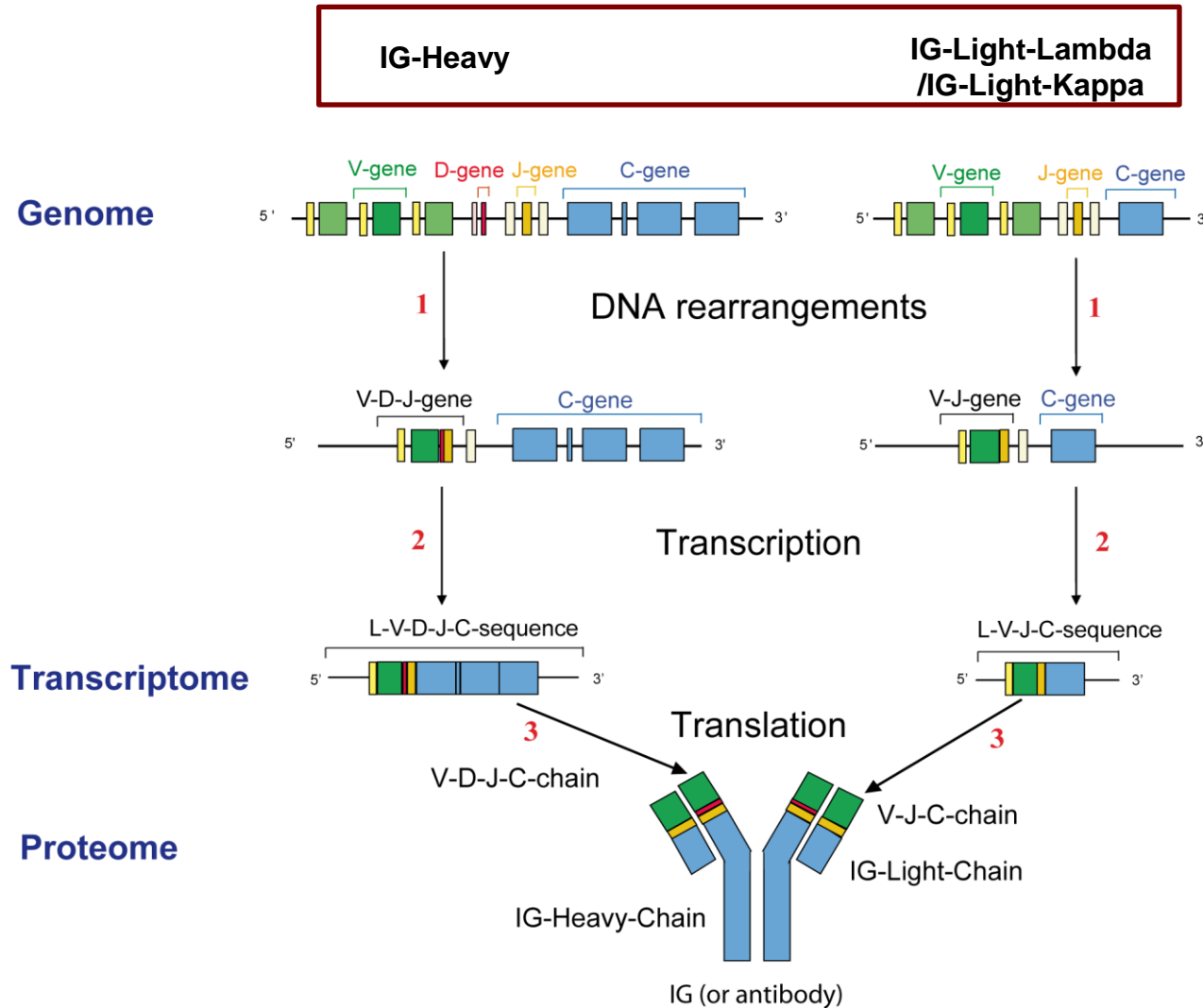
Concepts of CLASSIFICATION :
the IG and TR gene nomenclature

http://imgt.cines.fr

- Approved by HUGO Nomenclature Committee (HGNC) in 1999

20
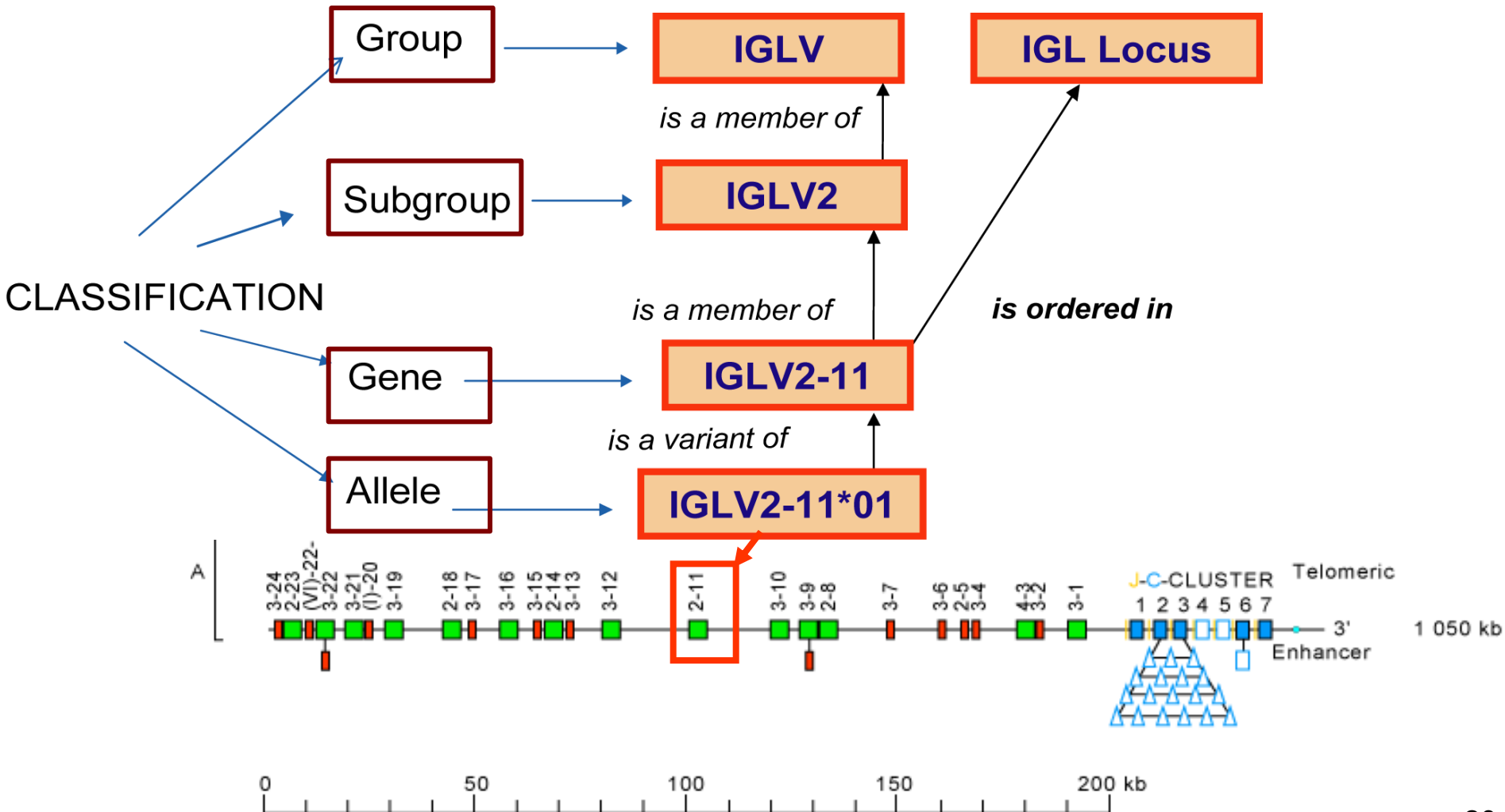
Concepts of DESCRIPTION: IMGT labels

Labels for a germline V-GENE in gDNA
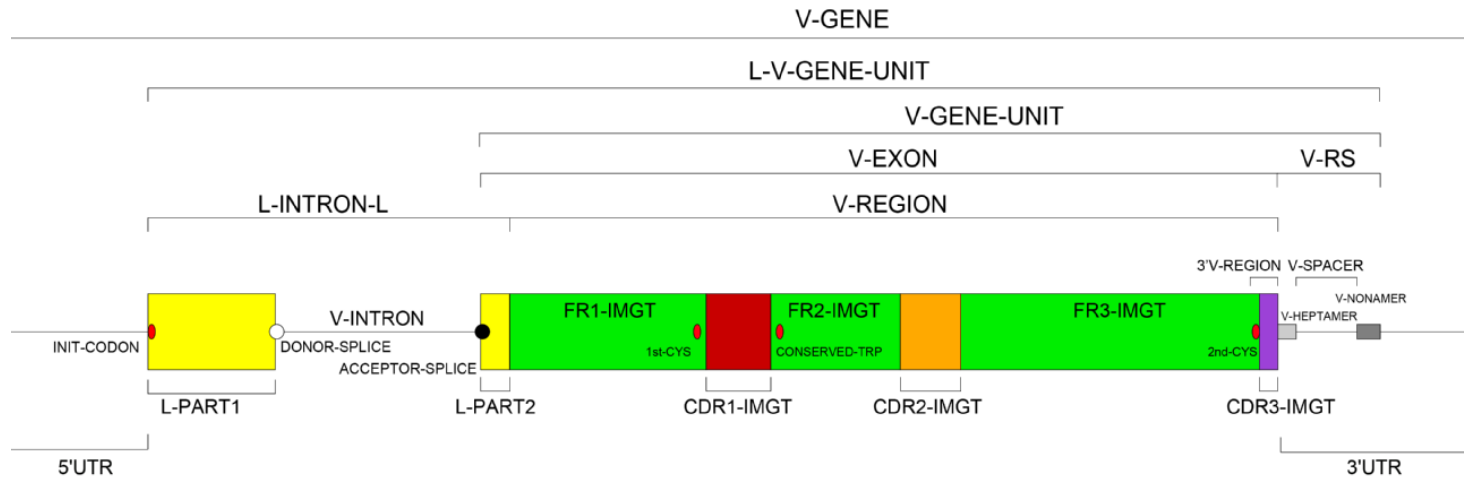
Labels for a rearranged V-D-J-GENE in gDNA
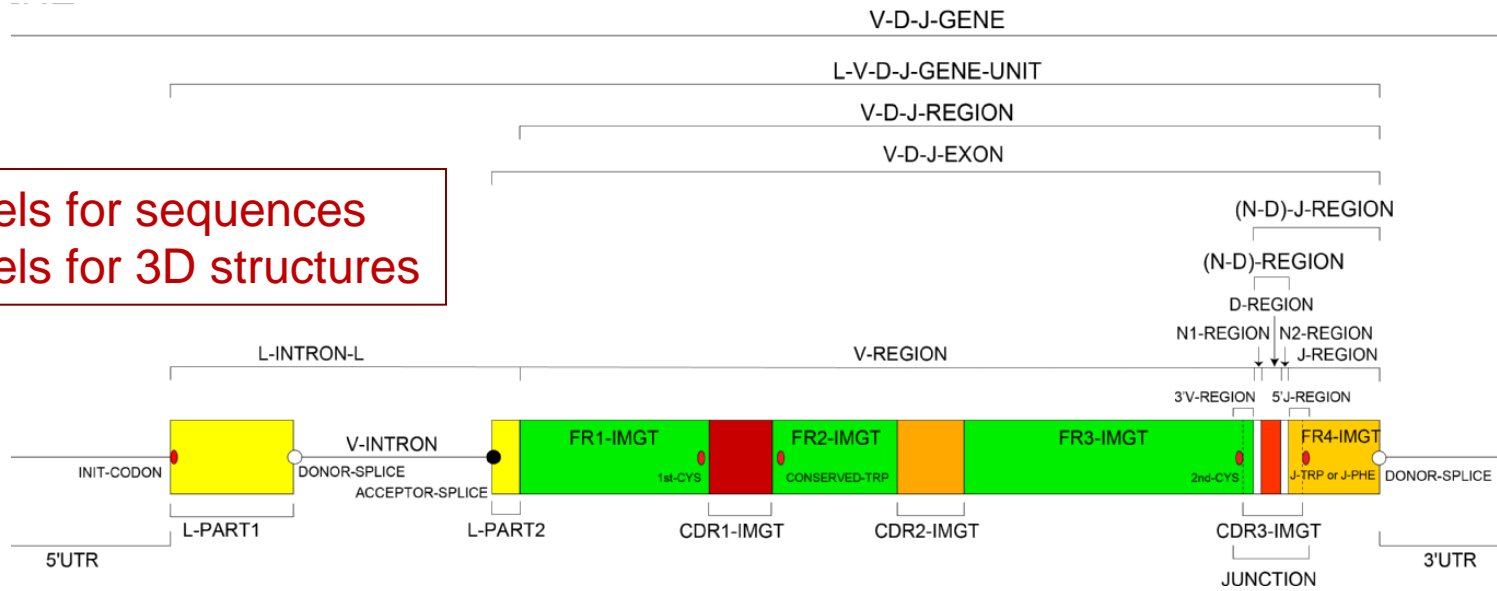
287 labels for sequences
370 labels for 3D structures

ImMunoGeneTics
Information system®

http://www.imgt.org

IMGT Collier de Perles

Based on the IMGT unique numbering
(first one in 1997)

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

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

- delimitation of the FR-IMGT
and CDR-IMGT is standardized

- CDR-IMGT lengths are crucial
information

CDR-IMGT lengths
[8.10.12]

alemtuzumab VH
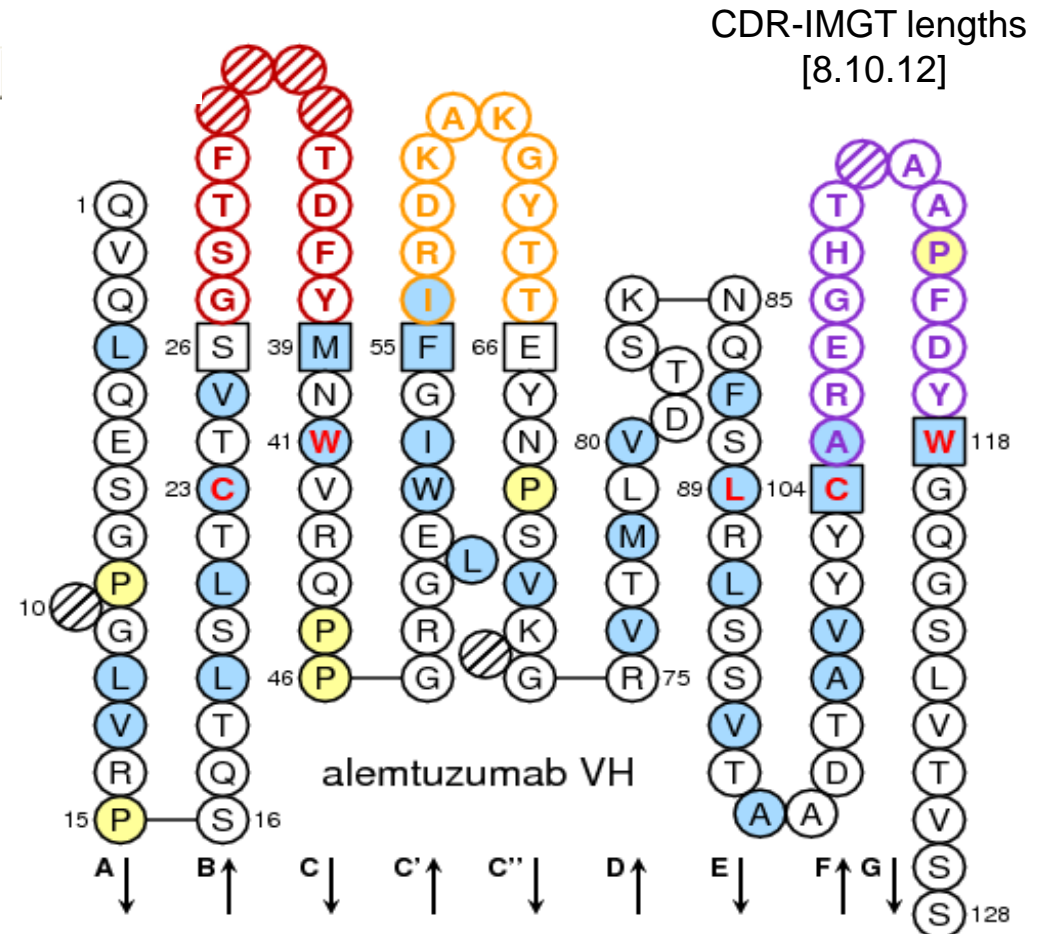
IMGT FLAT-FILE

```
ID    M18809 IMGT/LIGM annotation : by annotators; genomic DNA; HUM; 1194 BP.
AC    M18809;
DT    15-MAY-1995 (Rel. 2, arrived in LIGM-DB )
DT    27-FEB-2007 (Rel. 200709-2, Last updated, Version 7)
XX
DE    Human Ig rearranged H-chain gene (WS1) V-region (VDJ4).  ;
DE    genomic DNA; rearranged configuration; Ig-Heavy; regular; functionality
DE    productive  group IGHV; subgroup HV5.
XX
KW    antigen receptor; Immunoglobulin superfamily (IgSF);
KW    Immunoglobulin (IG); IG-Heavy; IG-Heavy-Mu; variable; diversity;
KW    joining; gDNA; rearranged; V-D-J-gene.
XX
OS    Homo sapiens (human)
OC    cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
.........
OC    Homo/Pan/Gorilla group; Homo.
XX
FH    Key              Location/Qualifiers
FH
FT    L-V-D-J-GENE     1..1194
FT                     /db_xref="taxon:9606"
FT                     /map="14q32.33"
FT                     /organism="Homo sapiens"
FT                     /mol_type="genomic DNA"
FT    5'UTR            1..198
FT    L-PART1          199..244
FT                     /protein_id="AAA36053.1"
FT                     /translation="MGSTAILALLLAVLQ"
FT    INIT-CODON       199..201
FT    DONOR-SPLICE     244..246
FT    V-INTRON         245..327
FT    ACCEPTOR-SPLICE  325..329
FT    V-D-J-EXON       328..717
FT                     /codon_start=3
FT                     /translation="VCAEVQLVQSGAEVKKPGESLKISCKGSGYSFSTYWIAWLR
FT                     QMPGKGLEWMGIIYAGDSETRYTPSFRGQVTISADKSTSTAYLQWSSLKASDTAM
FT                     YYCARLEGRGYTGYALPYYFDYWGQGPLVTVSS"
FT    V-EXON           328..632
FT                     /codon_start=3
FT                     /translation="VCAEVQLVQSGAEVKKPGESLKISCKGSGYSFSTYWIAWLR
FT                     QMPGKGLEWMGIIYAGDSETRYTPSFRGQVTISADKSTSTAYLQWSSLKASDTAM
FT                     YYCAR"
FT    L-PART2          328..338
FT                     /codon_start=3
FT                     /translation="VCA"
FT    V-D-J-REGION     339..717
FT                     /translation="EVQLVQSGAEVKKPGESLKISCKGSGYSFSTYWIAWLRQMP
FT                     GKGLEWMGIIYAGDSETRYTPSFRGQVTISADKSTSTAYLQWSSLKASDTAMYYC
FT                     ARLEGRGYTGYALPYYFDYWGQGPLVTVSS"
FT    V-D-REGION       339..671
FT                     /putative_limit="3' side"
FT                     /translation="EVQLVQSGAEVKKPGESLKISCKGSGYSFSTYWIAWLRQMP
FT                     GKGLEWMGIIYAGDSETRYTPSFRGQVTISADKSTSTAYLQWSSLKASDTAMYYC
FT                     ARLEGRGYTGYALPY"
FT    V-REGION         339..632
FT                     /allele="IGHV5-51*01"
FT                     /gene="IGHV5-51"
FT                     /CDR_length="[8.8.19]"
FT                     /putative_limit="3' side"
FT                     /translation="EVQLVQSGAEVKKPGESLKISCKGSGYSFSTYWIAWLRQMP
FT                     GKGLEWMGIIYAGDSETRYTPSFRGQVTISADKSTSTAYLQWSSLKASDTAMYYC
FT                     AR"
FT    FR1-IMGT         339..413
FT                     /AA_IMGT="AA 1 to 26,AA 10 is missing"
FT                     /translation="EVQLVQSGAEVKKPGESLKISCKGS"
FT    1st-CYS          402..404
```

1

IDENTIFICATION: Keywords

genomic DNA=MoleculeType
rearranged=ConfigurationType
regular=StructureType
productive=Functionality
Homo sapiens=Taxon
Ig-Heavy=ChainType
variable, diversity, joining=GeneType

2

CLASSIFICATION: Nomenclature

IGHV=Group
IGHV5=Subgroup
IGHV5-51=Gene
IGHV5-51*01=Allele

3

DESCRIPTION:Labels

L-V-D-J-SEQUENCE=Entity
V-D-J-REGION=ComposedRegion
V-REGION=CoreRegion
FR1-IMGT=SubRegion
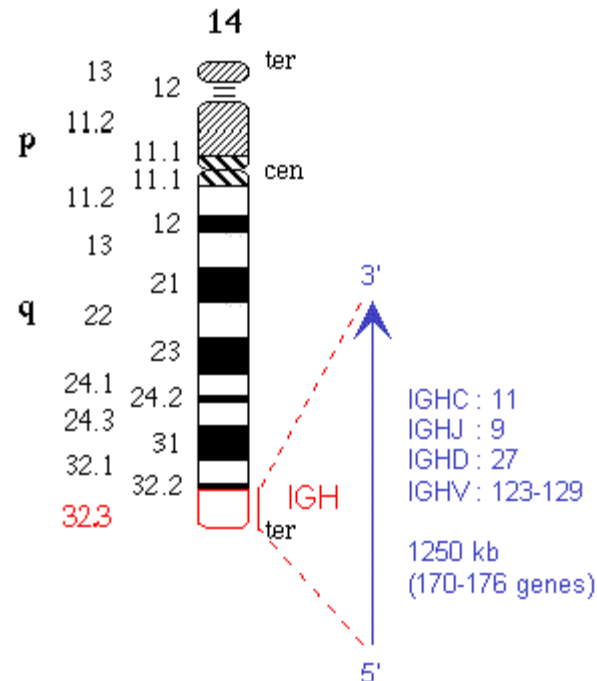1st-CYS=ConservedAminoAcid

4

NUMEROTATION

V [8.8.19]=V-REGION CDR lengths
1 to 26, AA 10 is missing=AA IMGT
numbering

ImMunoGeneTics
Information system®

23

Analyse de génome

Master STIC pour la santé spécialité BCD   5 Nov 2013

Chromosomal localization: human (Homo sapiens) IGH

14

IGHC : 11
IGHJ : 9
IGHD : 27
IGHV : 123-129

1250 kb
(170-176 genes)

The **human IGH locus** is located on the chromosome 14
at band 14q32.33, at the telomeric extremity of the long arm

Human *(Homo sapiens)* IGH locus on chromosome 14 (14q32.33)

The orientation of the human IGH locus on the chromosome is reverse (REV).

- Human IGH orphons on chromosome 15 (15q11.2)
- Human IGH orphons on chromosome 16 (16p11.2)

26

Centromeric

V-GENE

>X62106.0|HSVI2|*Homo sapiens* VI-2 gene for immunoglobulin heavy chain
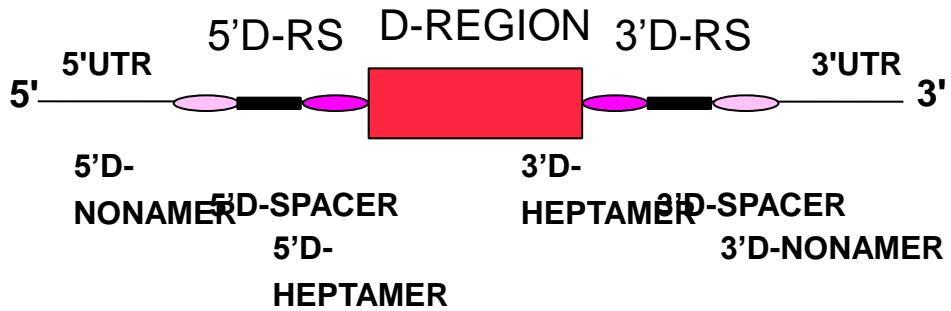
```
tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag   60
ccacaggtaa gaggctccct agtcccagtg atgagaaaga gattgagtcc agtccaggga  120
gatctcatcc acttctgtgt tctctccaca ggagcccact cccaggtgca gctggtgcag  180
tctggggctg aggtgaagaa gcctggggcc tcagtgaagg tctcctgcaa ggcttctgga  240
tacaccttca ccggctacta 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
```
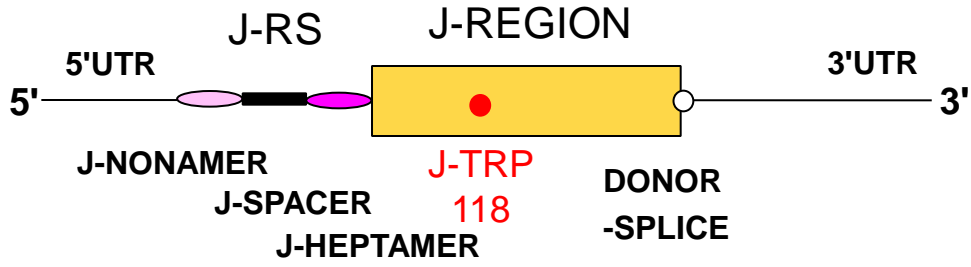
Genomic DNA in germline configuration

D-GENE

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

ccagccgcag ggtttttc tgagctgaga ac accactgtgc taactgggga attagtgatt
ggcagctct a caaaaaccat gctccccgg g

5'D-RS   D-REGION   3'D-RS

5'UTR   3'UTR

5'D-NONAMER   3'D-HEPTAMER
5'D-SPACER   3'D-SPACER
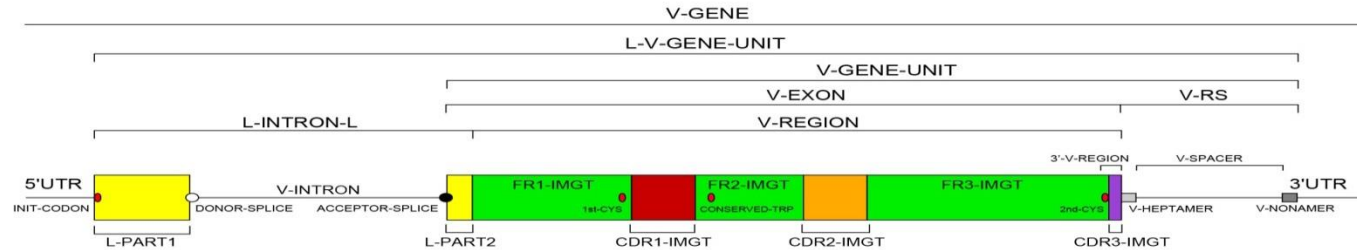5'D-HEPTAMER   3'D-NONAMER

J-GENE

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

accccgggct gtgggtttct cccctgg ctcagggctg act actcaccgtg gctgaatact
tccagca ctg gggccagggc accctggtca ccgtctcctc aggtgagtct gctgtactgg
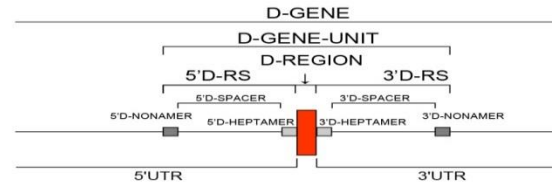ggatagcggg gagccatgtg tactgggcca agcaaggct ttggcttcag

J-RS   J-REGION

5'UTR   3'UTR

J-NONAMER
J-SPACER   J-TRP 118   DONOR-SPLICE
J-HEPTAMER

Séquences déscription ou prototypes

http://www.imgt.org/IMGTScientificChart/SequenceDescription/prototypes.html
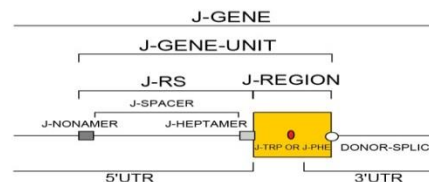
Quelques Bases de données et serveurs génomiques

Les bases de données généralistes:
- Génomique: Genbank, EMBL-Bank (ENA), DDBJ, IMGT/LIGM-DB (IG-TR)
- Protéines: UniProt
- Structurale: PDB, IMGT/structure-DB

Les grands serveurs de génomes:
NCBI aux USA, EBI en Angleterre et DDBJ au japon, Gold, Ensembl…

Master STIC pour la santé spécialité BCD   5 Nov 2013

Diagram of the "typical" eukaryotic protein-coding **gene**

Promoteur

Enhancers

DNA

ATG

Codon-stop

Transcription

Introns

Exon   Exon   Exon   Exon   Exon

Pre-mRNA

ATG

Codon-stop

Splicing

5'UTR   3'UTR

Translation

Protein