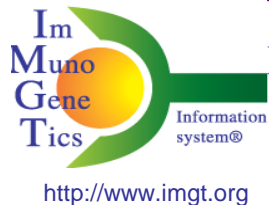


Utilisation d'IMGT pour l'humanisation et l'ingénierie des anticorps thérapeutiques

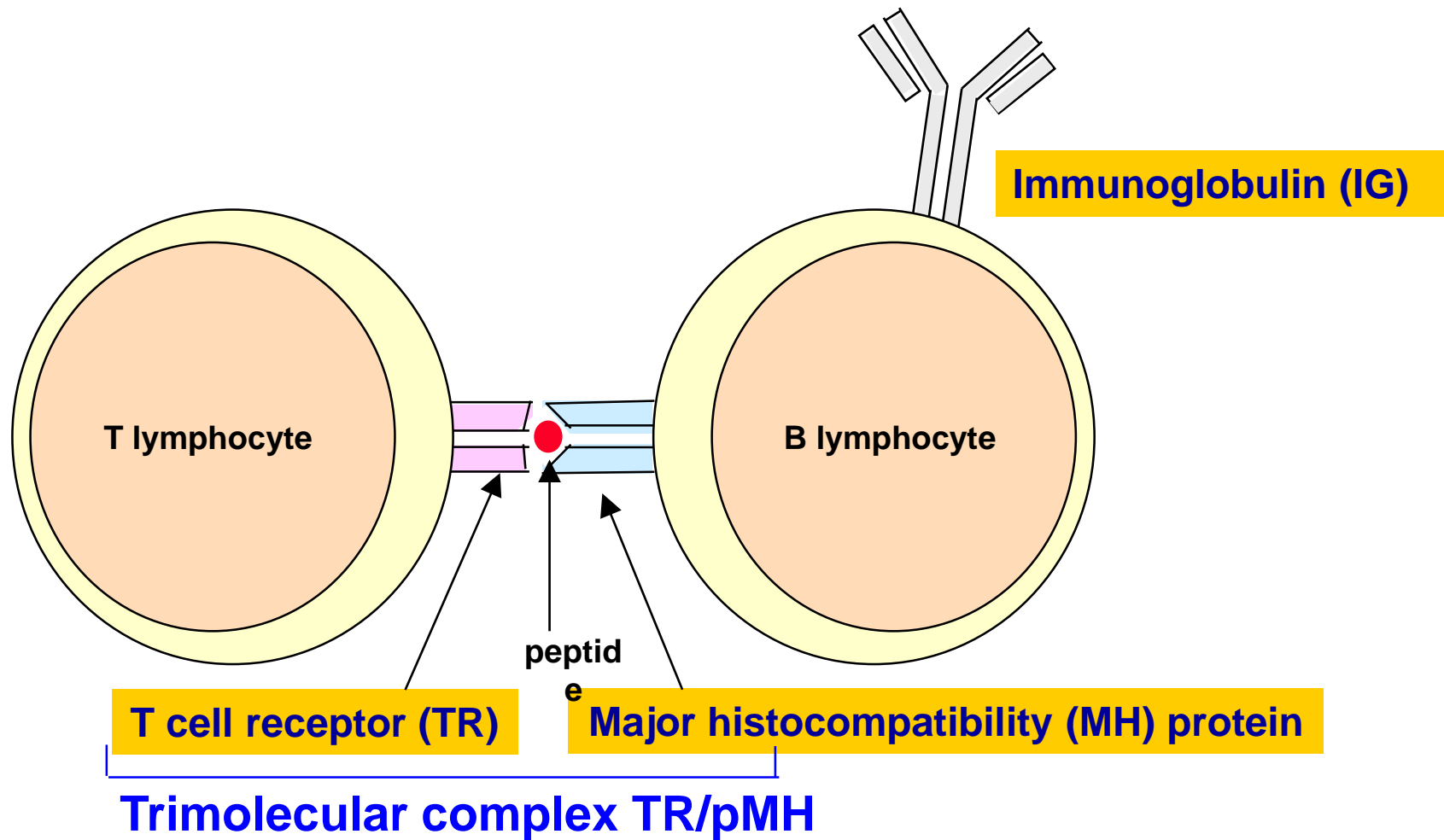
Marie-Paule Lefranc

IMGT®, the international ImMunoGeneTics information system®
Université de Montpellier, CNRS Institut de Génétique Humaine IGH,
Montpellier, France



Recherche, développement et production d'anticorps monoclonaux thérapeutiques
Villejuif, 4-6 avril 2018

IMGT® domain: the adaptive immune response



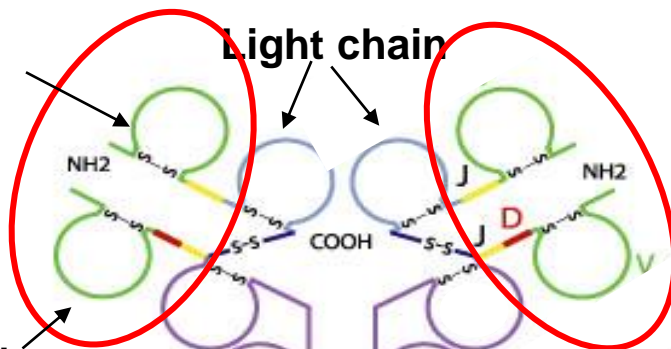
Vertebrates *gnathostomata* (with jaws): specificity, memory

Immunoglobulin (IG)

T cell receptor (TR)

V-DOMAIN

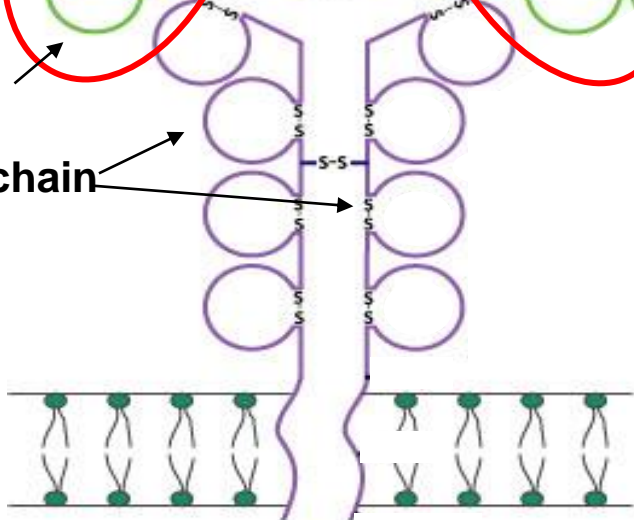
V-J-REGION



V-DOMAIN

V-D-J-REGION

Heavy chain



Membrane IgM

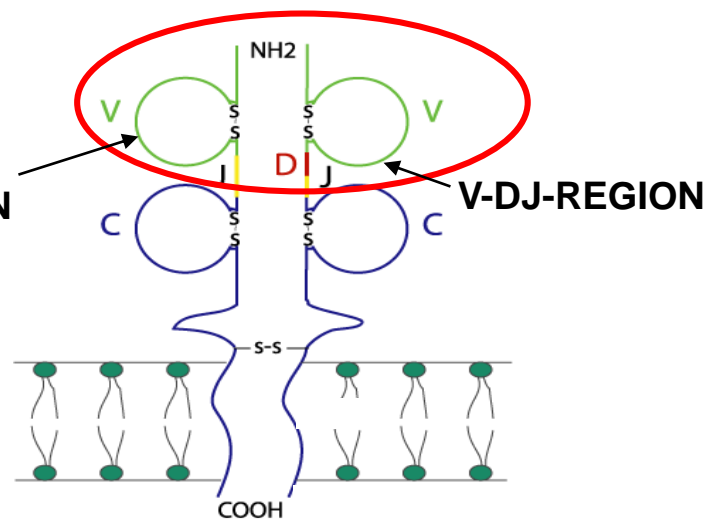
Contribution of the
2 V-DOMAINS
to the antigen binding
site

Alpha

Beta

Gamma

Delta

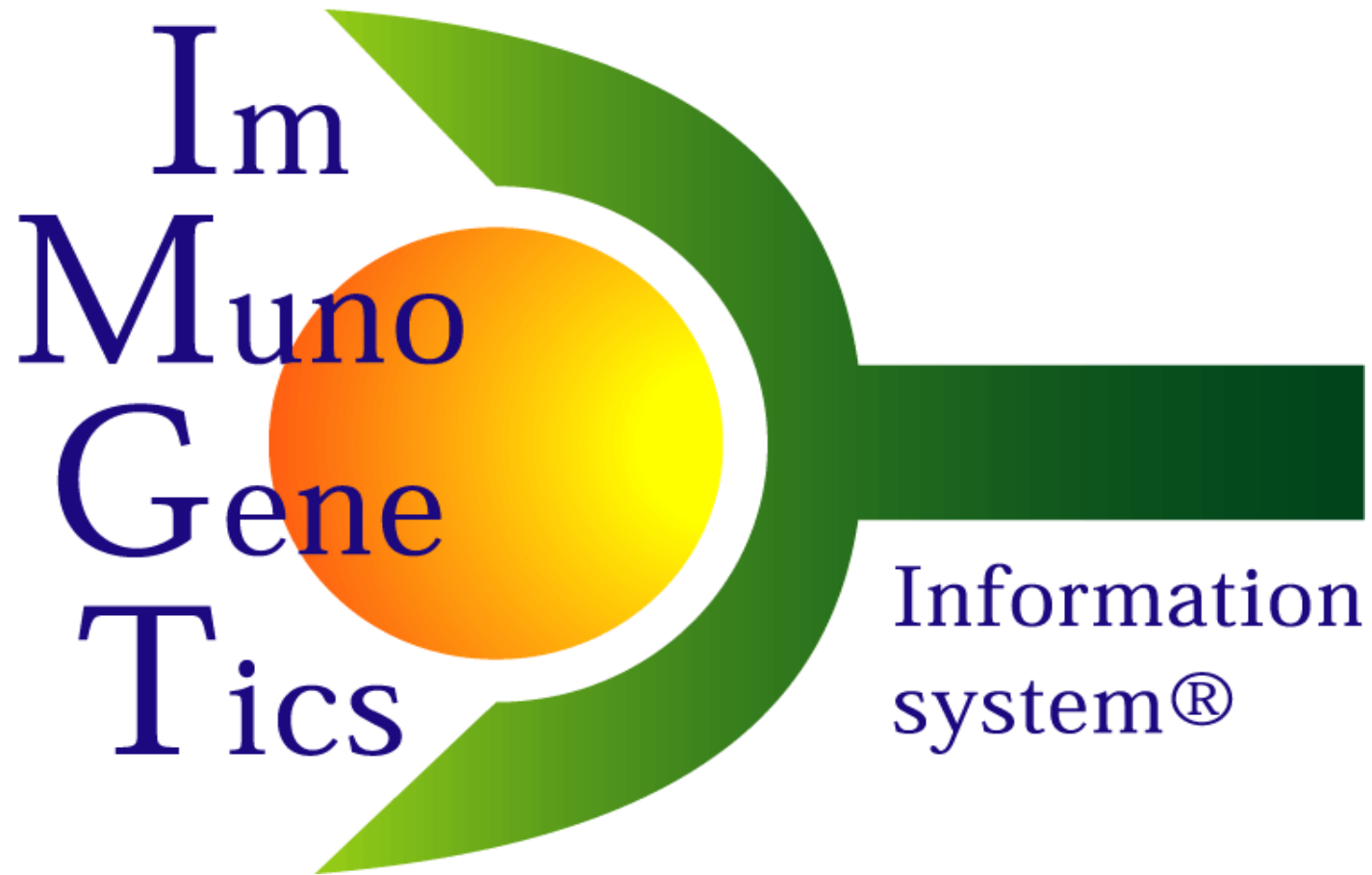


V-J-REGION

V-DJ-REGION

T cell receptor

IMGT[®], the international ImMunoGeneTics information system[®]
<http://www.imgt.org>

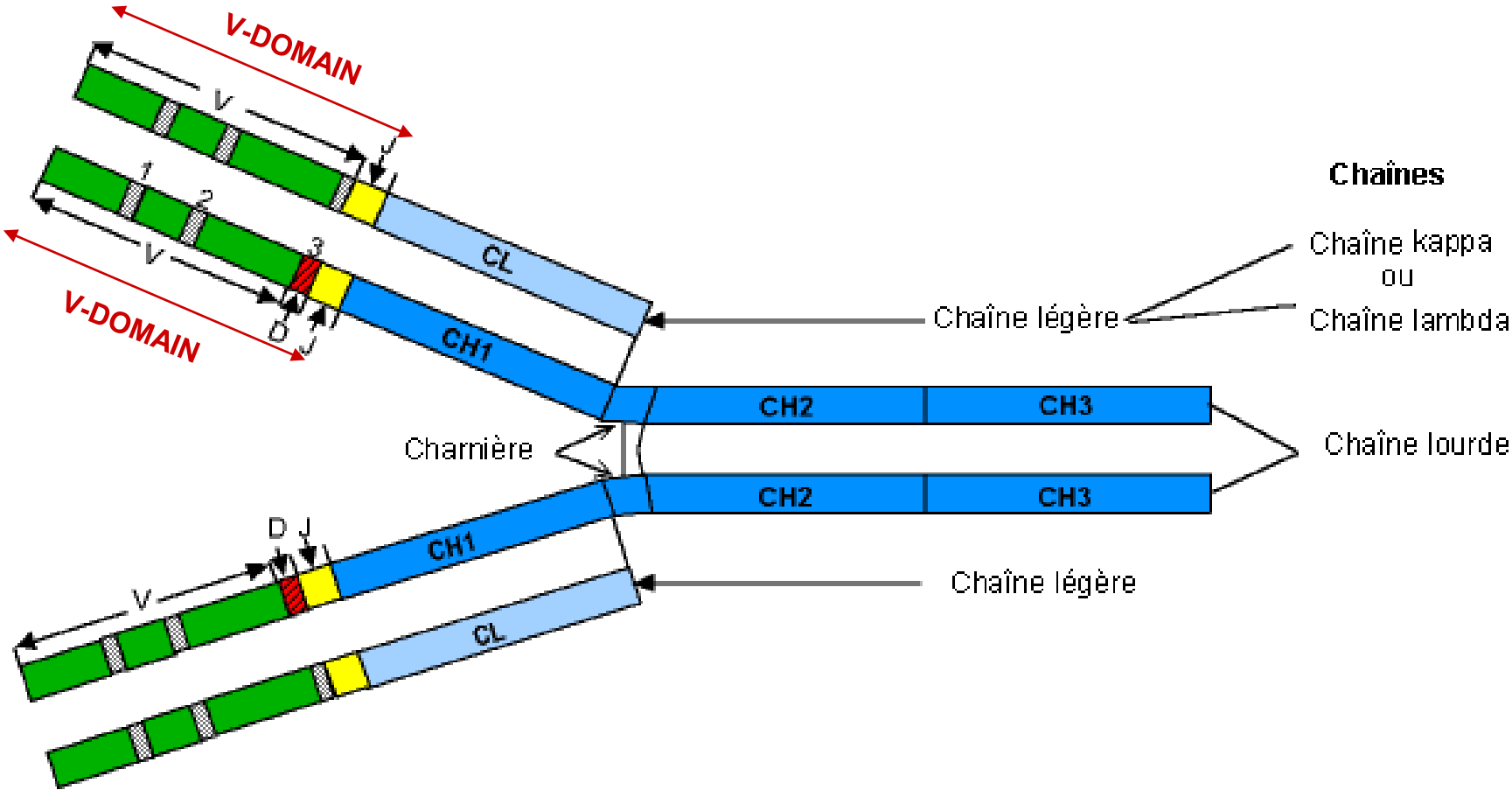


Why was IMGT® necessary?

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

IMGT® is at the origin of Immunoinformatics

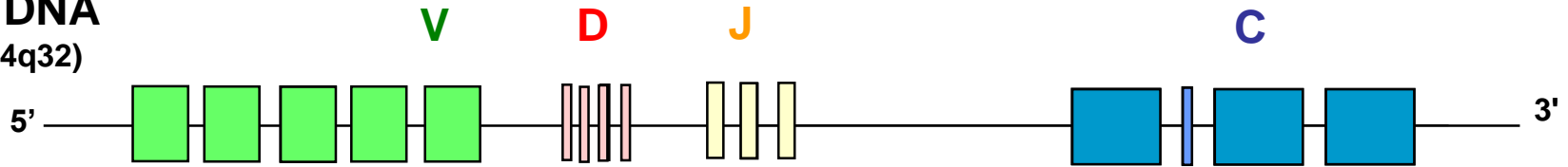
Structure of an immunoglobulin (IG) or antibody



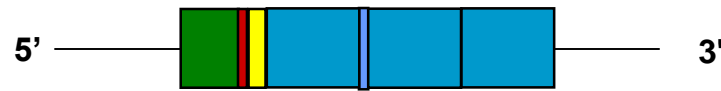
Potentially 2×10^{12} different IG (or antibody) specificities

Immunoglobulin (IG) synthesis

genomic DNA
(IGH Locus 14q32)

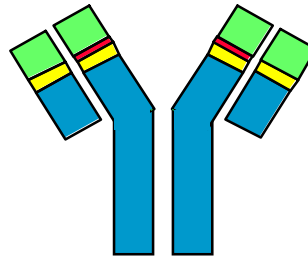


**rearranged
DNA**



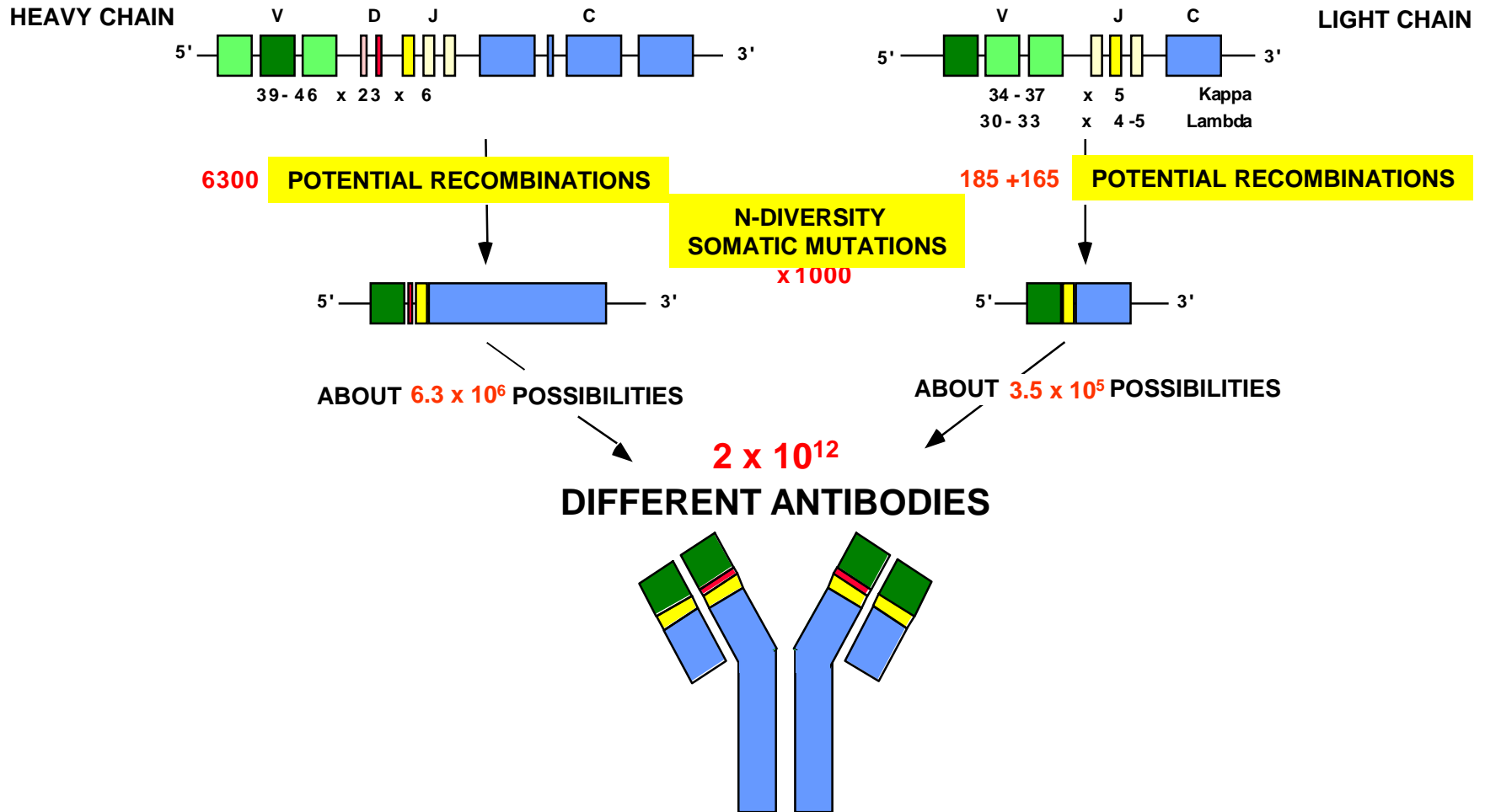
mRNA

2×10^{12} different IG
per individual



Immunoglobulin (IG) synthesis

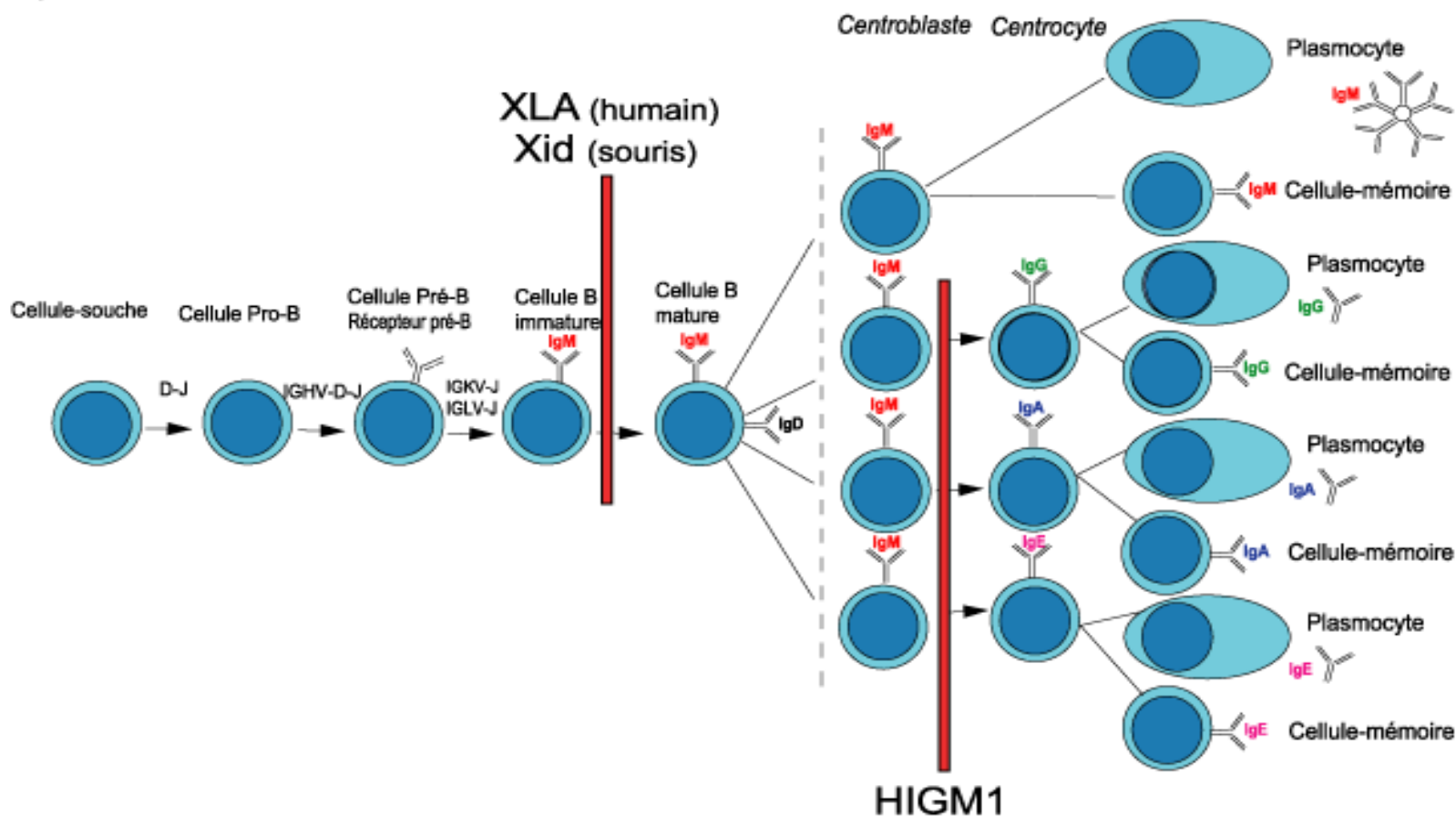
150 FUNCTIONAL IG GENES



Bone marrow

Blood

Lymph nodes, spleen



V-D-J and V-J rearrangements

Hypermutations, selection

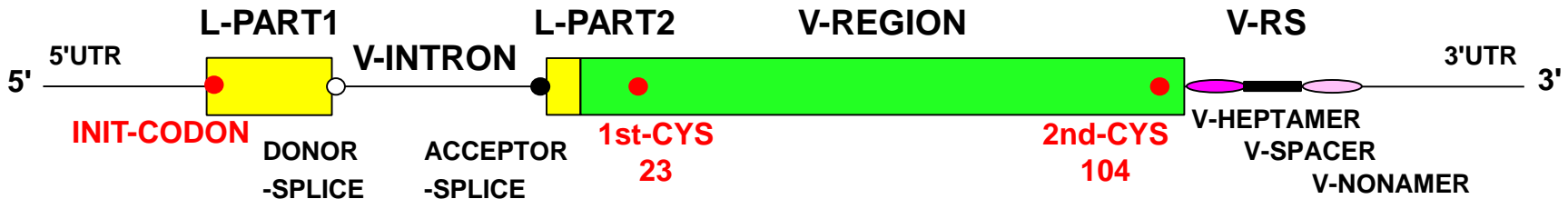
Genomic DNA in germline configuration

V-GENE

>X62106.0|HSV12|*Homo sapiens* VI-2 gene for immunoglobulin heavy chain

```

tgagagctcc gttcctcacc atggactgga cctggaggat cctcttcttg gtggcagcag      60
ccacagggttaa gaggctcct agtcccagtg atgagaaaga gattgagtcc agtccaggga      120
gatctcatcc acttctgtgt tctctcca ca ggagcccact ccaggtgca gctgggtgcag      180
tctggggctg aggtgaagaa gcttggggcc tcagtgaagg tctcctgcaa ggcttctgga      240
tacaccttca ccggctacta tatgcaactg gtgacacagg ccctggaca agggcttgag      300
tggatgggat ggatcaacc taacagtggg ggcacaaact atgcacagaa gtttcagggc      360
agggtcacca tgaccaggga cacgtccatc agcacagcct acatggagct gagcaggctg      420
agatctgacg acacggcogt gtattactgt gcgagagaca cagtggtgaaa acccacatcc      480
tgagggtgtc agaaacccaa gggaggagggc ag
    
```



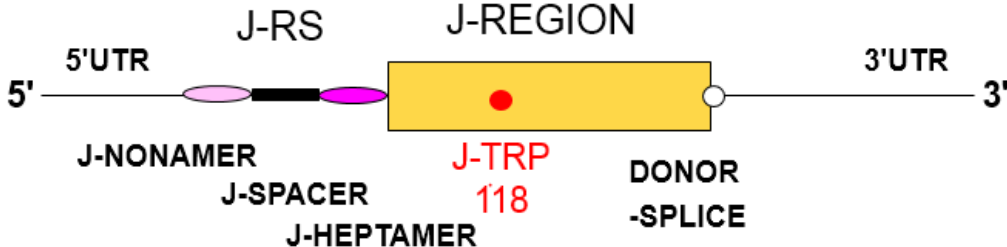
Genomic DNA in germline configuration

J-GENE

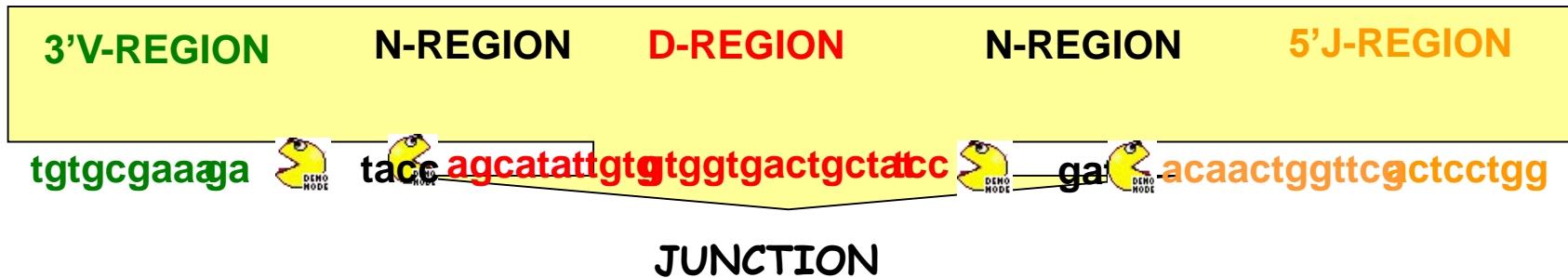
>J00256|IGHJ1*01|Homo sapiens J-GENE

```

acccgggct gtgggtttct gtgcccctgg ctcagggctg actcaccgtg gctgaatact    60
tccagcactg gggccagggc accctgggtca cgtctcctc aggtgagtct gctgtactgg    120
ggatagcggg gagccatgtg tactgggcca agcaagggct ttggcttcag                170
    
```

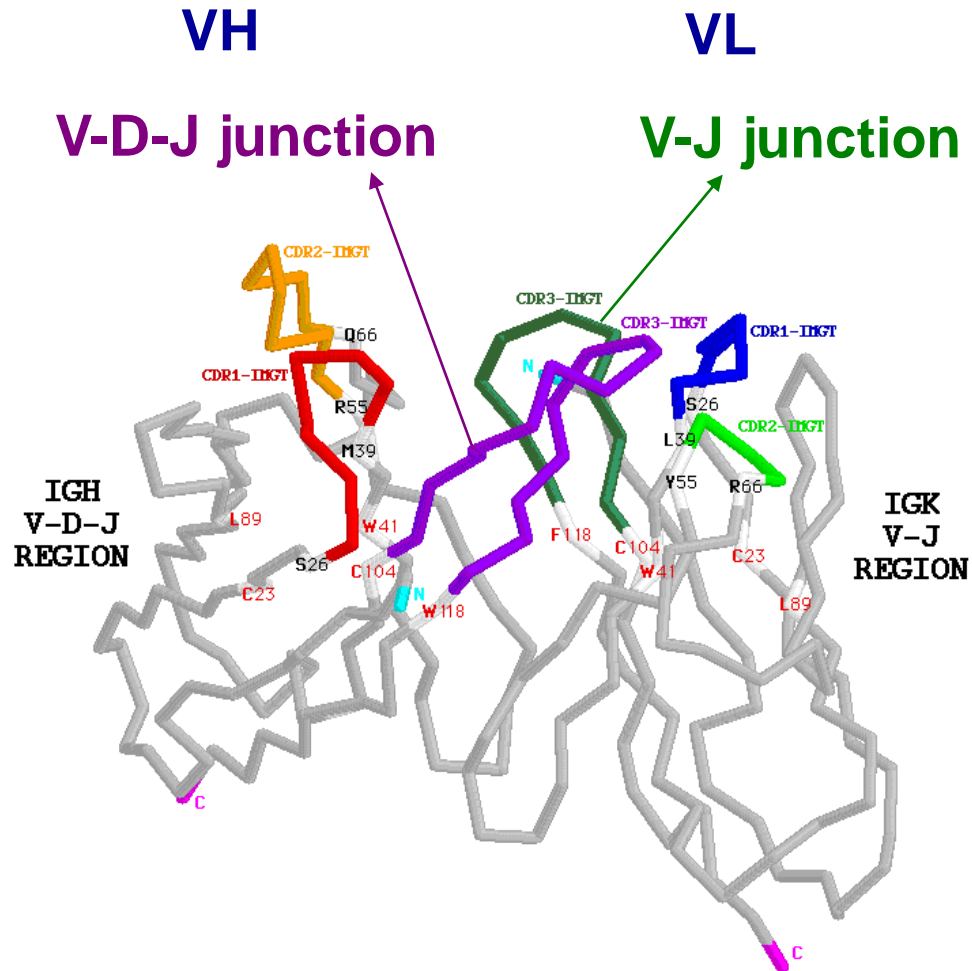


Immunoglobulin V-D-J generation of sequence diversity



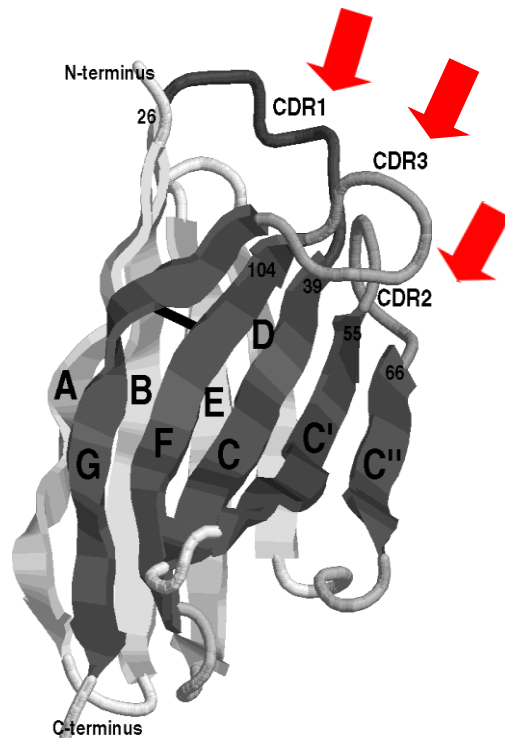
C A P Y R G D T Y D Y S W
 tgt gcg cca tac cgg ggt gac act tat gat tac tcc
 tgg

V domains of an IG (or antibody)

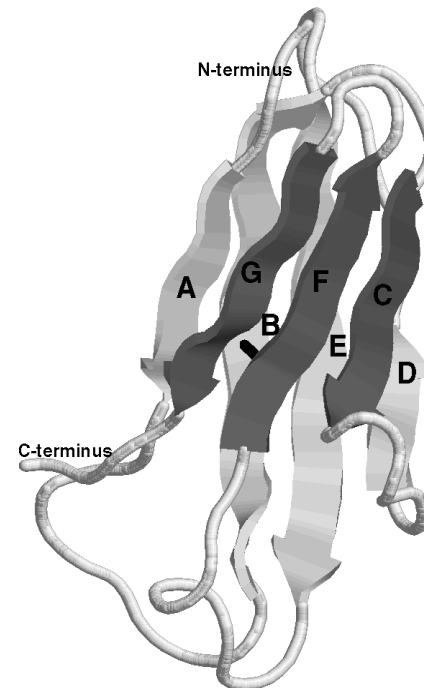


Structural domains of the IG and TR

V-DOMAIN



C-DOMAIN



CDR: complementarity determining region

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

JUNCTION alignments

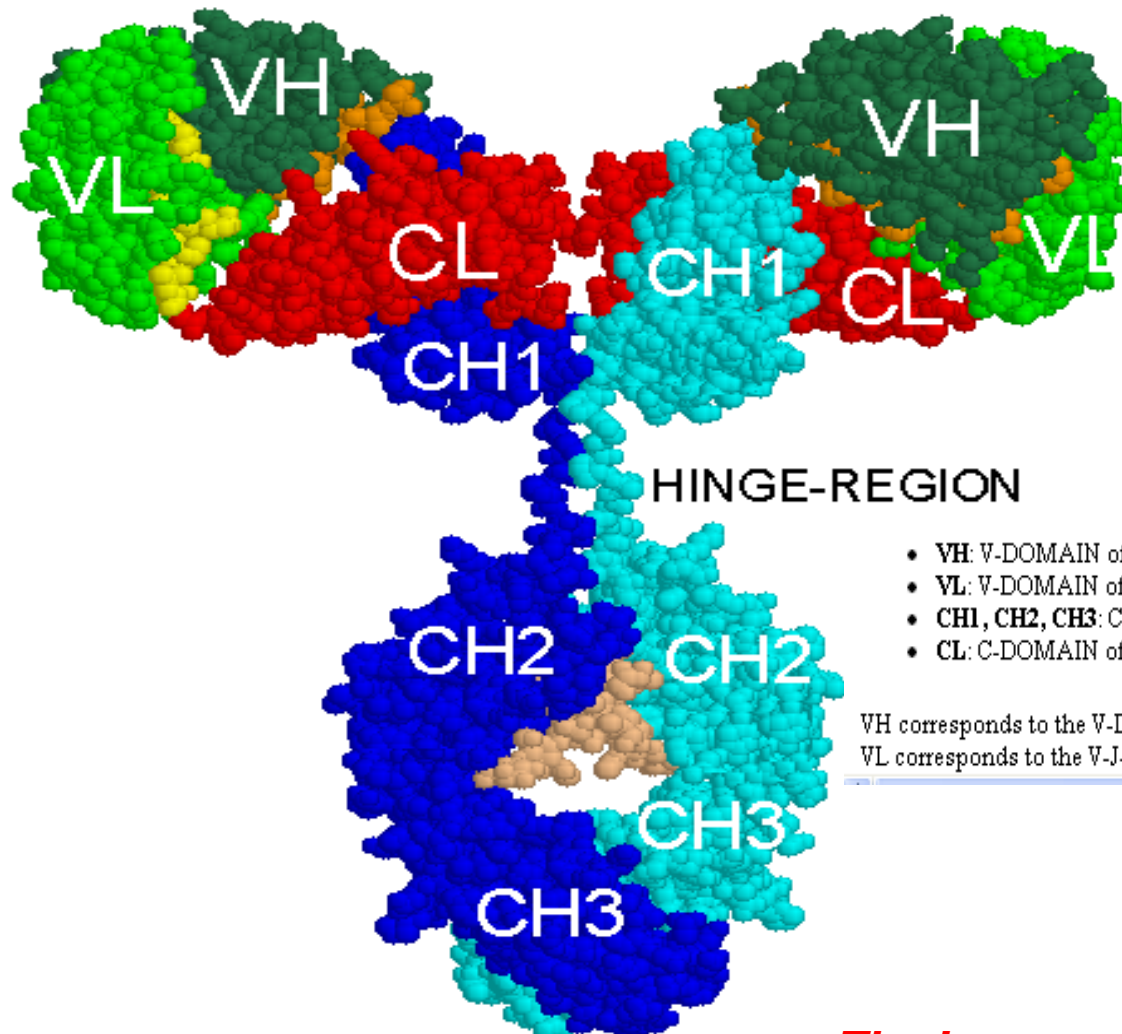
 Click on mutated (underscored) amino acid to see the original one:

	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	
#1 Z70256	C	<u>V</u>	R	V	V	Q	<u>R</u>	L	V								P	K	Y	H	F	D	<u>H</u>	W	+	15	2,438.76	
#2 Z70257	C	A	R	D	G	S	<u>S</u>	Y	A							R	P	Y	W	Y	F	D	L	W	+	16	2,256.49	
#3 Z70606	C	A	R	A	T	T	H												Y	A	F	D	V	W	+	11	1,604.77	
#4 Z70608	C	A	R	V	T	I	F	G	V	V						I	P	R	G	N	A	F	D	I	W	+	18	2,426.78
#5 Z70610	C	A	R	V	G	S	D	F	W	S	G					Y	S	R	H	D	A	F	D	I	W	+	19	2,539.73
#6 Z70611	C	A	R	H	G	<u>N</u>	Y	N	A								G	V	D	W	F	D	P	W	+	15	2,072.21	
#7 Z70613	C	A	R	A	A	A	G												T	S	L	F	D	Y	W	+	12	1,531.71
#8 Z70614	C	A	R	H	Y	N	<u>S</u>	G											T	Y	P	L	D	Y	W	+	13	1,846.01
#9 Z70615	C	A	R	G	L	V	<u>K</u>	<u>R</u>	V							S	E	Y	W	Y	F	D	L	W	+	16	2,419.75	
#10 Z70616	C	A	R	A	G	L	G												S	H	W	F	D	P	W	+	12	1,602.8
#11 Z70620	C	A	R	D	R	G	G	M									V	R	D	A	F	D	I	W	+	14	1,932.17	
#12 Z70621	C	A	R	H	H	D	L	W	F								G	E	F	D	P	L	D	Y	W	+	16	2,307.53
#13 Z70622	C	A	R	D	C	P	A	P	A	K	M	Y	Y	Y	G	S	G	I	C	T	F	D	Y	W	+	16	2,307.53	

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

'Volume' classes		'Hydropathy' classes						
	in Å ³	Hydrophobic		Neutral		Hydrophilic		
Very large	189-228	F			Y			
Large	162-174	I	L					
Medium	138-154	V				H		
Small	108-117							
Very small	60-90	A						
		Aliphatic		Sulfur	Hydroxyl	Basic	Acidic	Amide
		Nonpolar		Uncharged	Charged		Uncharged	
				Polar				

Spacefill 3D representation of an IgG



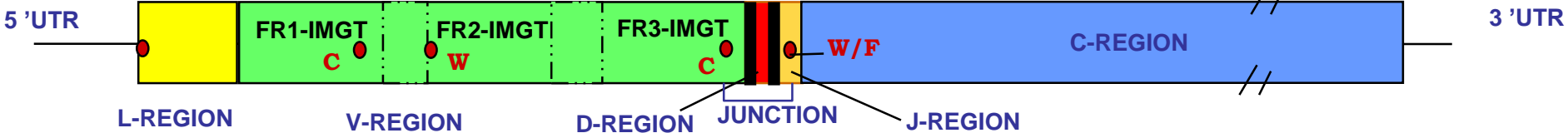
- VH: V-DOMAIN of the immunoglobulin heavy chain
- VL: V-DOMAIN of the immunoglobulin light chain
- CH1, CH2, CH3: C-DOMAIN of the immunoglobulin heavy chain
- CL: C-DOMAIN of the immunoglobulin light chain

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.

cdNA

```

.....gagga ttcaccatgg aactggggct ccgctggggt ttccttggtg cttttttaga 120
aggtgtccag tgtgaggtgc aactggtgga gtctggggga ggctgtgca agccgggggg 180
gtccctgaga ctctctgtg cagcctctg attagcttc agtacctatg ccatgaactg 240
ggtccgccag gctccagga aggggctgga atgggtctca agtattagta gtagaagtga 300
ttacata tac tatagagact cagtgaagg cggattcacc atctccagag acaacgcaa 360
gaattcactg tatctgcaa tgaatagcct gagagtcgac gacacggctg tctattactg 420
tgcgagaga ttgtaatg gtgctatatg ttatggttc ccctggg gccagggaac 480
cctggtcacc gtctcctcag catccccgac cagccaggtag gtcttcccgc tgagcctctg 540
cagcaccag ccagatggga acgtggtcat cgctgctg gtccagggct tcttccccca 600
ggagccactc agtgtgacct ggagcgaag cggacagggc gtgaccgcca gaaacttccc 660
accagccag gatgcctccg gggacctgta caccacgagc agccagctga ccctgccggc 720
cacacagtgc ctagecggca agtccgtgac atgccacgtg aagcactaca cgaatcccag 780
ccaggatgtg actgtgcct gccagttcc ctcaactcca cctaccccat ctccctcaac 840
tccacctacc ccattctcct catgctgcc cccccgactg tcaactgcacc gaccggccct 900
cgaggacctg ctcttaggtt cagaagcga cctcacgtgc aactgaccg gcctgagaga 960
tgctcgggt gtcacctca cctggacgcc ctcaagtggg aagagcgtg ttcaaggacc 1020
acctgagcgt gacctctgtg gctgctacag cgtgtccagt gtctgcccgg gctgtgccga 1080
gccatggaac catgggaaga ccttcaactg cactgctgcc taccccagat ccaagacccc 1140
gctaaccgcc accctctcaa aatccgaaa cacattccgg cccgaggtec acctgctgcc 1200
gccgccgtcg gaggagctgg ccctgaacga gctggtgacg ctgacgtgcc tggcacgtgg 1260
cttcagccc aaggacgtgc tggttcgtg gctgcagggg tcacaggagc tgccccgcga 1320
gaagtacctg acttgggcat cccggcagga gccagccag ggcaccacca ccttcgctgt 1380
gaccagcata ctgctcgtgg cagccgagga ctggaagaag ggggacacct tctcctgcat 1440
gggtgggccac gaggcctgc cgctggcctt cacacagaag accatcgacc gcttggcggg 1500
taaaccacc catgtcaatg tgtctgttgt catggcggag gtggacggca cctgctactga 1560 ....
    
```



IMGT birth and objectives: 1989

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 responsible*
- * *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.

Lefranc M-P.

Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics.

Front Immunol. 2014 Feb 05;5:22.

doi: 10.3389/fimmu.2014.00022.

[Open access.](#) [PMID: 24600447](#)

Lefranc M-P et al.

IMGT®, the international ImMunoGeneTics information system® 25 years on.

Nucleic Acids Res. 2015 Jan;43(Database issue):D413-22.

doi: 10.1093/nar/gku1056. Epub 2014 Nov 5

[Free Article.](#) [PMID: 25378316](#)

Part 2.

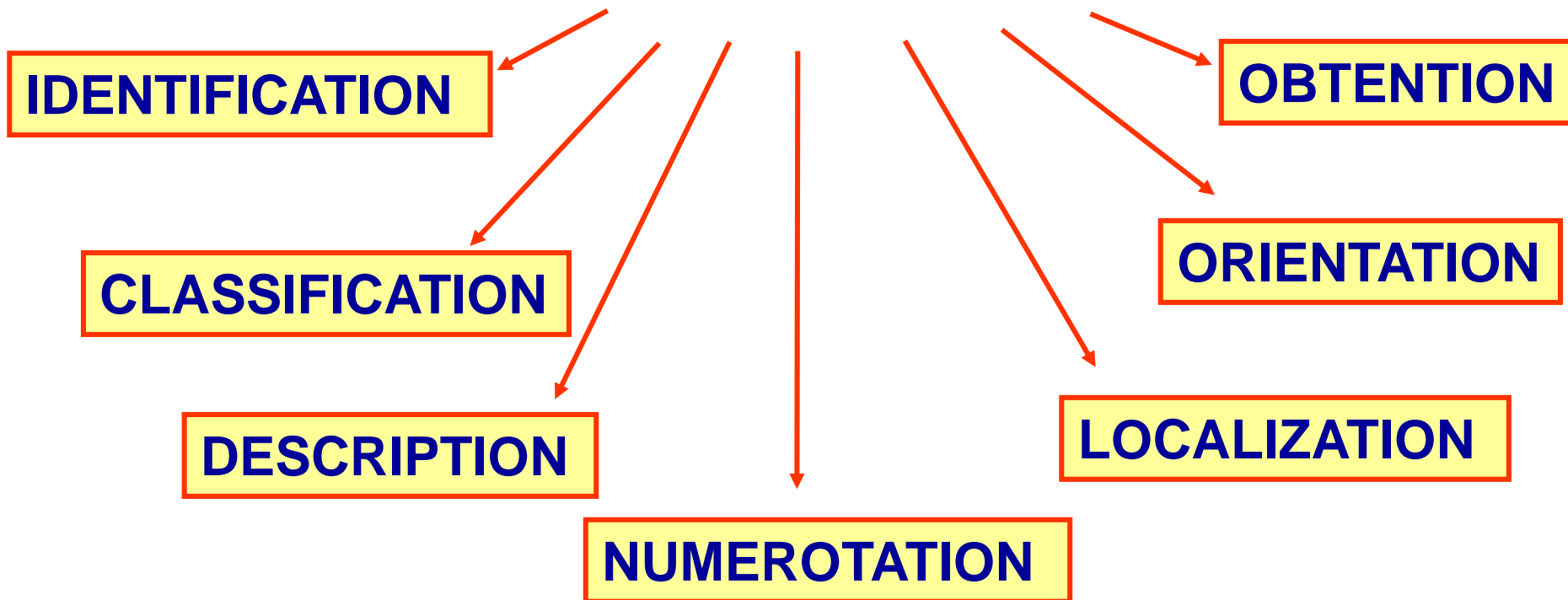
1. IMGT-ONTOLOGY

2. A new science: Immunoinformatics

3. An information system for IG, TR and MH.

IMGT-ONTOLOGY seven axioms:

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

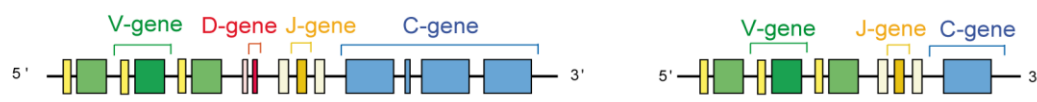


Concepts of IDENTIFICATION

IMGT® standardized keywords

IG-Heavy IGH **IG-Light-Kappa IGK**
IG-Light-Lambda IGL

Genome



1 DNA rearrangements



2 Transcription



Transcriptome

3 Translation



Proteome



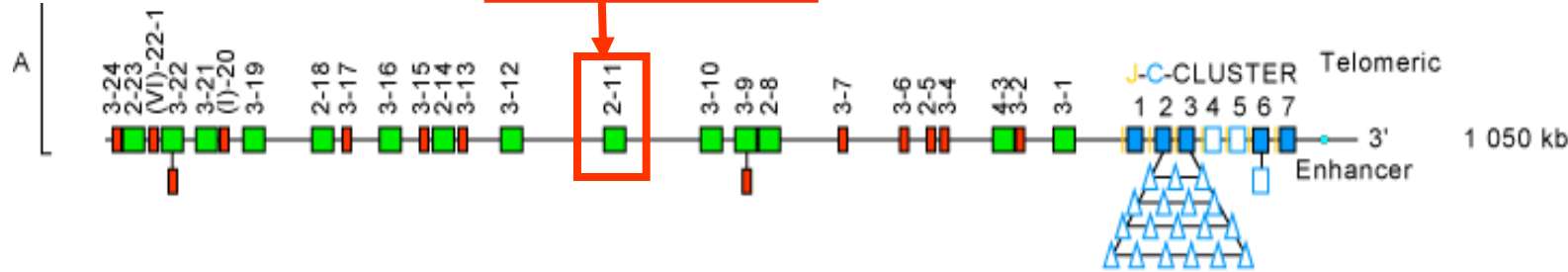
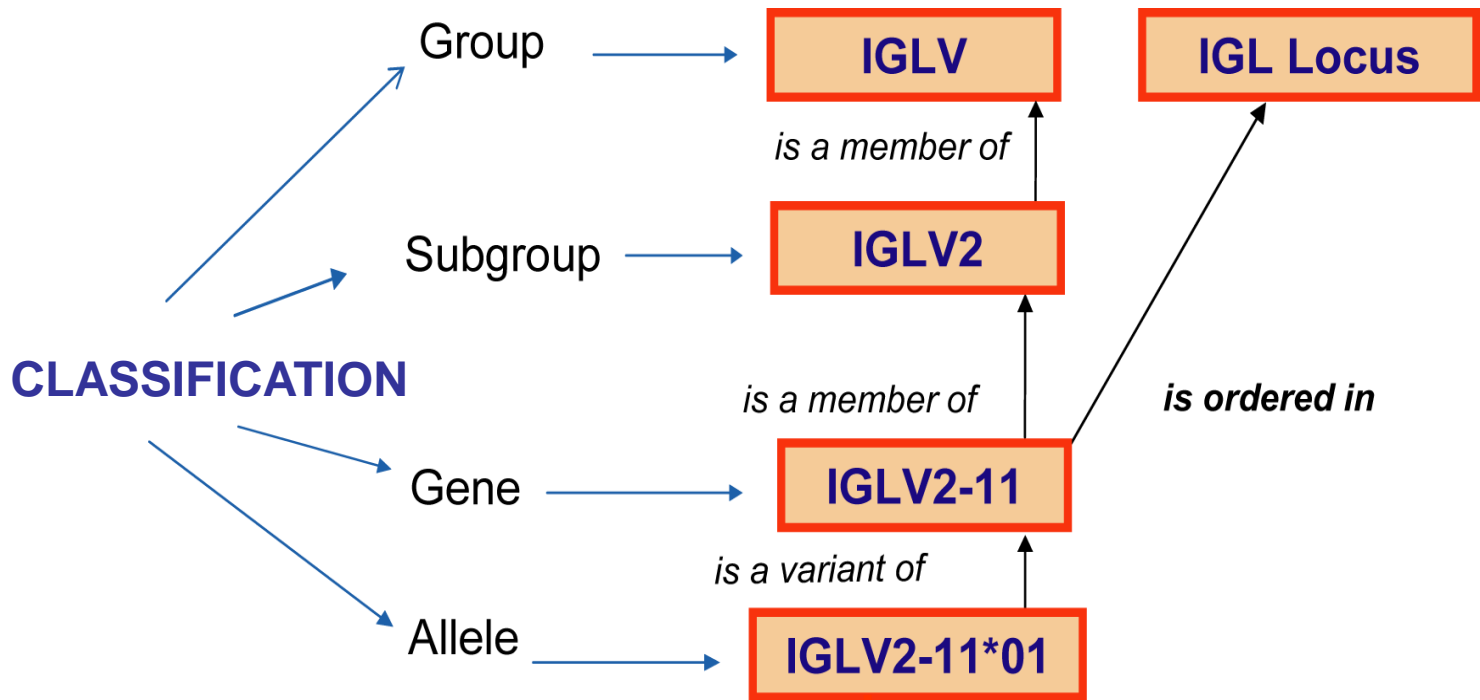
IG (or antibody)

"MoleculeType" concept	"ConfigurationType" concept
gDNA	germline
gDNA	rearranged
mRNA (or <i>in vitro</i> cDNA)	rearranged
protein	rearranged

Concepts of CLASSIFICATION

Gene name: *Homo sapiens* IGLV2-11

Definition: *Homo sapiens* immunoglobulin lambda variable 2-11



LOCALIZATION Human IGL locus (22q11.2)

1. Immunoglobulin (IG) and T cell receptor (TR) genes
2. Nomenclature for the variable (V), diversity (D), joining (J) and constant (C) genes
3. IMGT concepts of 'group', 'subgroup', 'gene' and 'allele'
4. All IMGT human gene names were approved by HUGO Gene Nomenclature Committee (HGNC) in 1999, and endorsed by NCBI in 2000
5. IMGT alleles validated by WHO-IUIS/IMGT-NC since 1989

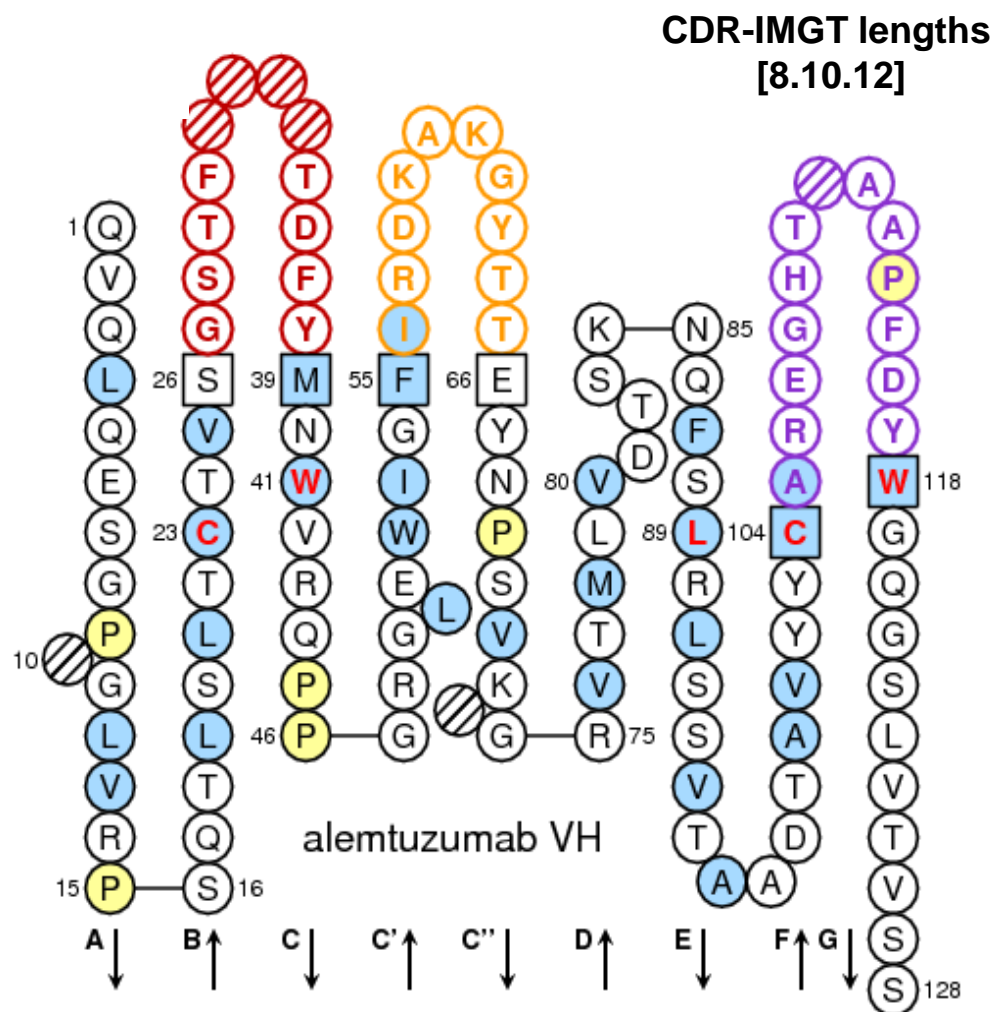
IMGT/GENE-DB: international reference database for IG and TR genes (links from NCBI Gene to IMGT/GENE-DB) and alleles

Concepts of NUMEROTATION

1. IMGT unique numbering

2. IMGT Collier de Perles

(first one in **1997**)



Concepts of NUMEROTATION

1. IMGT unique numbering

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

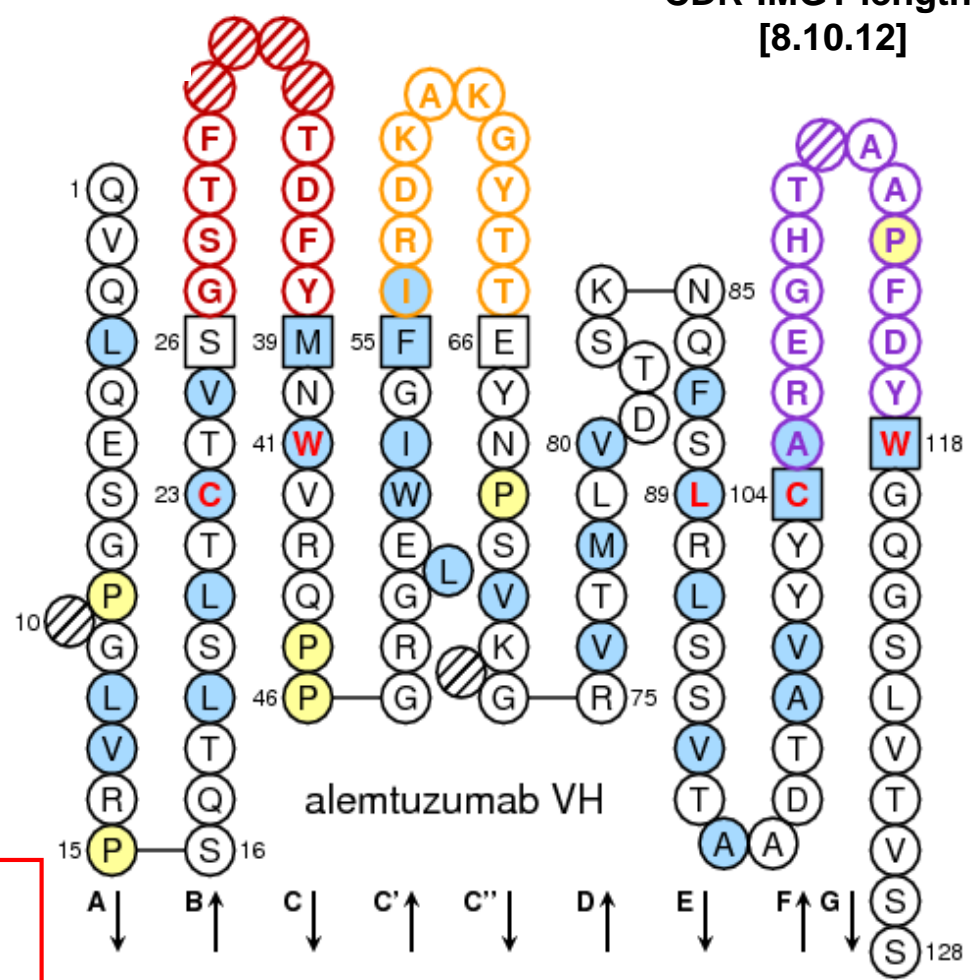
- conserved AA (and codons) always at the **same positions**:

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

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

CDR-IMGT lengths are crucial information

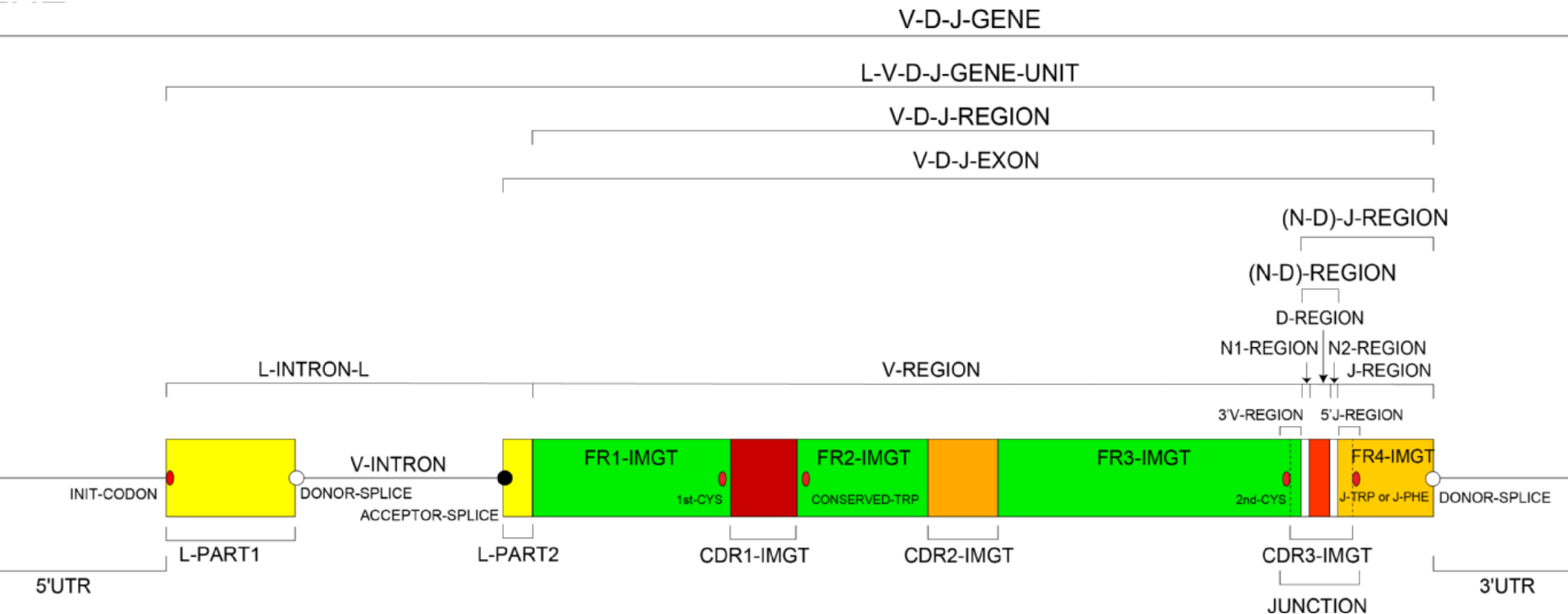
CDR-IMGT lengths
[8.10.12]

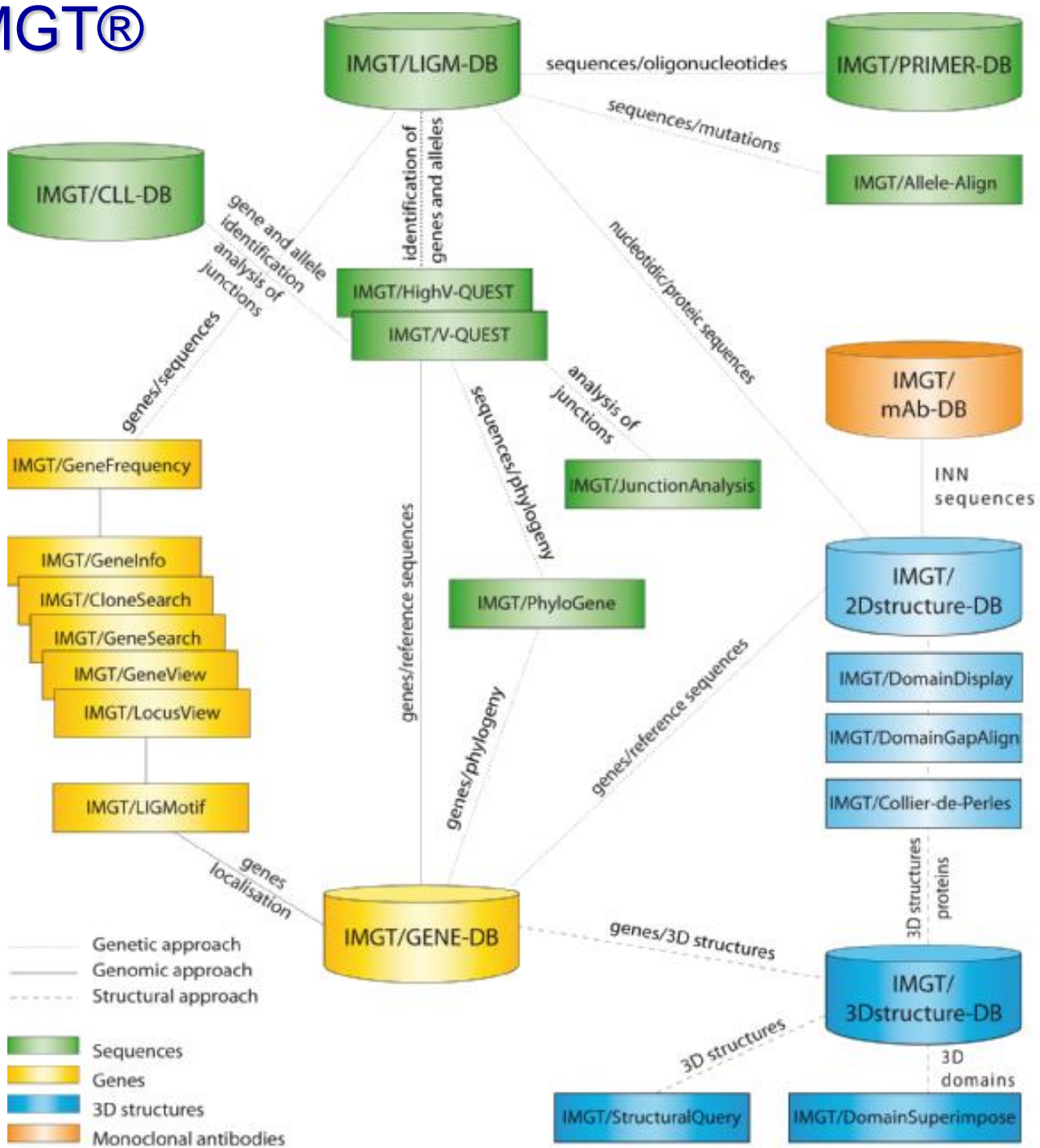


Concepts of DESCRIPTION

IMGT labels are in capital letters

Example: Prototype and IMGT Labels for a rearranged V-D-J-GENE in gDNA

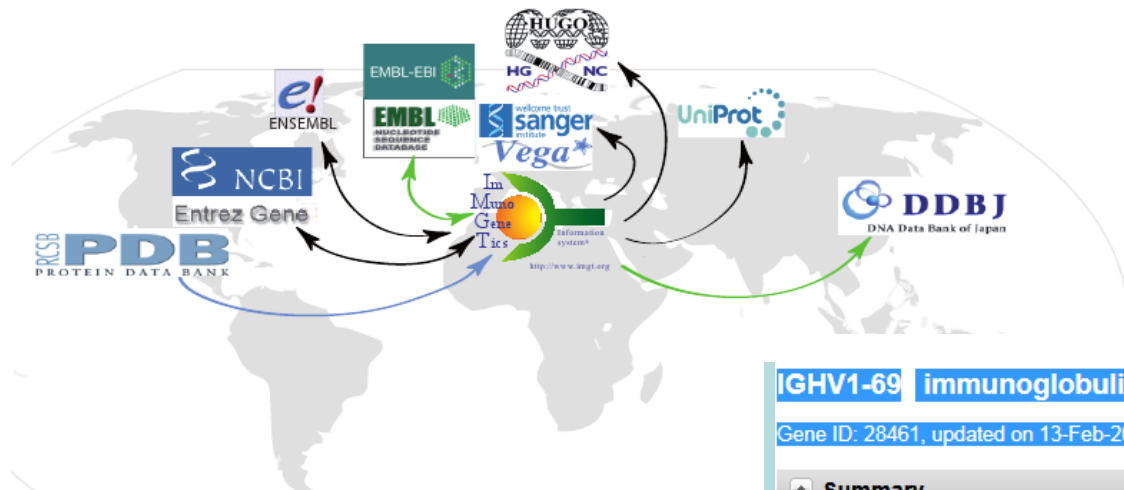




created in 1989
by Marie-Paule Lefranc
Montpellier University,
CNRS, France

- 7 databases
- 17 online tools
- 20,000 web pages
- Sequences
- Genes
- Structures
- Immunoglobulins (IG) (or antibodies)
- T cell receptors (TR)
- MH
- IgSF and MhSF

IMGT® source and reference for IG and TR genes



NCBI

IGHV1-69 immunoglobulin heavy variable 1-69 [*Homo sapiens*]

Gene ID: 28461, updated on 13-Feb-2016

Summary

- Official Symbol** IGHV1-69 provided by [HGNC](#)
- Official Full Name** immunoglobulin heavy variable 1-69 provided by [HGNC](#)
- Primary source** [HGNC:HGNC:5558](#)
- See related** [IMGT/GENE-DB:IGHV1-69](#)
- Gene type** other
- RefSeq status** VALIDATED
- Organism** [Homo sapiens](#)

HGNC

Symbol Report: IGHV1-69

APPROVED SYMBOL	IGHV1-69
APPROVED NAME	immunoglobulin heavy variable 1-69
HGNC ID	HGNC:5558
PREVIOUS SYMBOLS & NAMES	IGHV1-E
LOCUS TYPE	immunoglobulin gene
CHROMOSOMAL LOCATION	14q32.33
HCOP	GENE FAMILY Immunoglobulin heavy locus at 14q32.33
	Orthology Predictions for IGHV1-69

External links

SPECIALIST DATABASES	IMGT/GENE-DB
GENE RESOURCES	Entrez Gene: 28461 C Ensembl: ENSG00000211973 C Region in detail Sequence Vega: OTTHUMG00000151862 C Region in detail Sequence

THANK YOU for using IMGT/GENE-DB

THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®

<http://www.imgt.org>

Citing IMGT/GENE-DB : Giudicelli, V. et al. Nucleic Acids Res., 33: D256 - D261 (2005). PMID: 15608191



IMGT/GENE-DB program version: 3.1.11 (8 February 2016)
[IMGT/GENE-DB data upgrades](#)

IMGT/GENE-DB DETAILED RESULTS

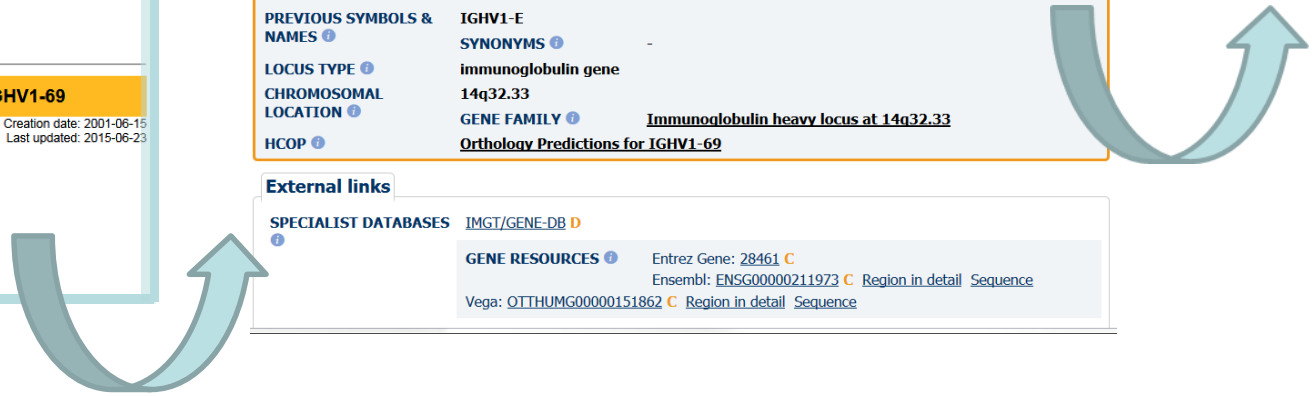
Selected genes: [Homo sapiens IGHV1-69](#)

IMGT/GENE-DB entry for *Homo sapiens* IGHV1-69

Creation date: 2001-06-15
 Last updated: 2015-06-23

IMGT gene name and definition

IMGT gene name: *Homo sapiens* IGHV1-69
IMGT gene definition: *Homo sapiens* immunoglobulin heavy variable 1-69



Chain details of b12 neutralizing mAb, anti-gp120 [HIV-1], IG, FAB-GAMMA-1_KAPPA *Homo sapiens* (human) [1n0x_H,1n0x_L]

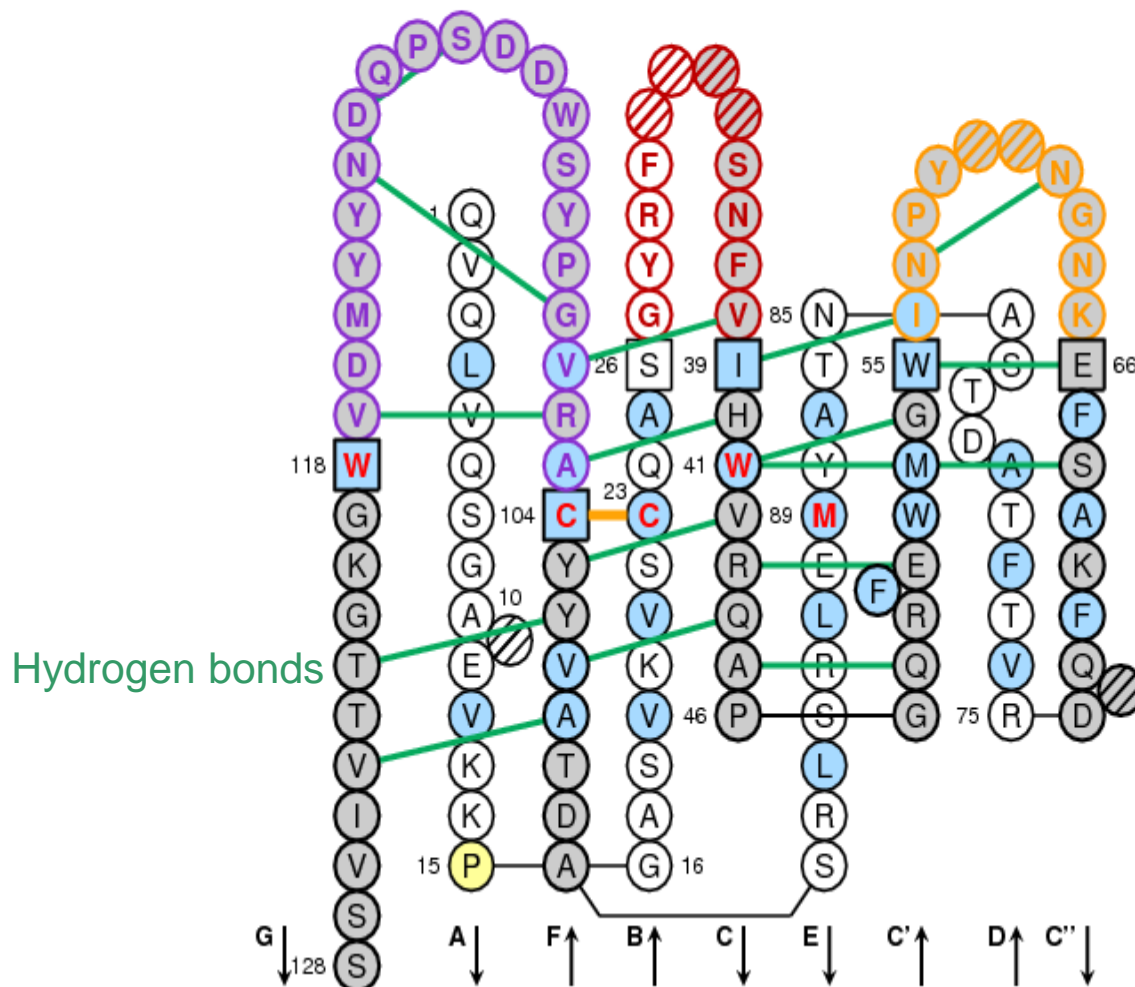
Chain ID	1n0x_H	DESCRIPTION
Chain length	230	
IMGT chain description	VH-CH1 = VH (1-127) [D1] + CH1 (128-219) [D2]	
Chain sequence	<p>[VH (1-127) [D1]] [CH1 (128-219)]</p> <p>QVQLVQSGAEVKKPGASVKVSCQASGYRFSNFIHWVRQAPGQRFEWMGWINPYNGNKEFSAKFQDRVFTADTSANTAYMELRSLRSAD TAVYYCARVGPYSWDDSPQDNYYMDVWGKGTIVVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF [D2]</p> <p>PAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHNKPSNTKVDKKVEPKSC</p> <p>Sequence in FASTA format Sequence in IMGT format</p>	
V-DOMAIN	IMGT domain description	VH (1-127) [D1] CLASSIFICATION
	IMGT gene and allele name	IGHV1-3*01 (79.60%)(human) Alignment details
	IMGT gene and allele name	IGHJ6*03 (93.80%)(human) Alignment details
	2D representation	IMGT Collier de Perles or IMGT Collier de Perles on 2 layers
	Contact analysis	Domain contacts (overview)
	CDR-IMGT lengths	[8.8.20]
	Sheet composition	[A' B D E] [A" C C' C" F G] NUMEROTATION
	<p>[CDR1] [CDR2]</p> <p>QVQLVQSGA.EVKKPGASVKVSCQASGYR...SNFIHWVRQAPGQRFEWMGWINPY..NGNKEFSAKFQ.DRVFTADTSANTAYME [CDR3]</p> <p>LRSLRSADTAVYYCARVGPYSWDDSPQDNYYMDVWGKGTIVVSS</p> <p>IMGT/DomainGapAlign results</p>	
IMGT domain	CH1 (128-219) [D2]	

IMGT/3Dstructure-DB

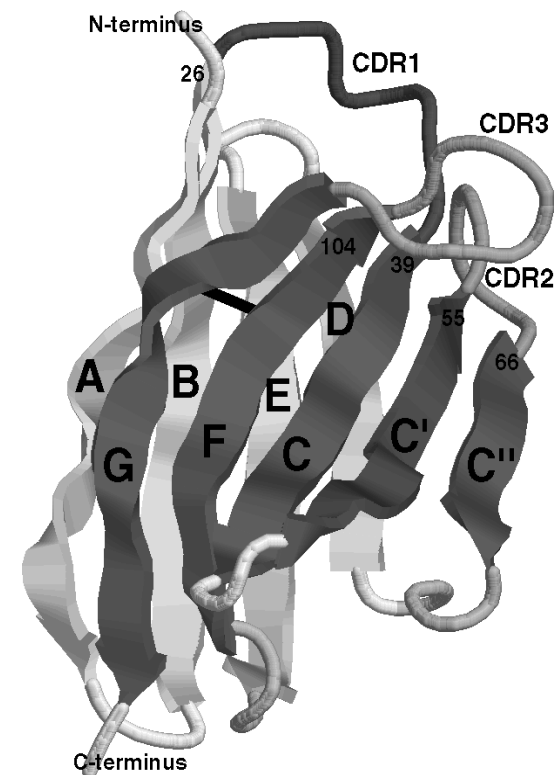
IMGT Collier de Perles : *Homo sapiens* (Human) IGHV V-DOMAIN from **b12** (1hzh_H)

<http://www.imgt.org>

CDR-IMGT lengths [8.8.20]



V-DOMAIN



IMGT/3Dstructure-DB: Contacts IG/Ag (Ligand)

IMGT/3Dstructure-DB Domain pair contacts (overview) of 1n0x

Click 'DomPair' for IMGT/3Dstructure-DB Domain pair contacts (list of Residue@Position pair contacts)

	Unit 1		Unit 2		Residue pair contacts	Number of residues			Atom pair contact types			
	Domain	Chain	Domain	Chain		Total	From 1	From 2	Total	Polar	Hydrogen	Nonpolar
DomPair	[D1] VH	1n0x_H	[D2] CH1	1n0x_H	10	11	5	6	56	11	1	45
DomPair			[D1] V-KAPPA	1n0x_L	59	46	21	25	596	61	7	535
DomPair			(Ligand)	1n0x_P	17	15	7	8	106	10	1	96
DomPair			(Ligand)	1n0x_R	5	6	4	2	34	12	1	22
DomPair	[D2] CH1	1n0x_H	[D1] VH	1n0x_H	10	11	6	5	56	11	1	45
DomPair			[D2] C-KAPPA	1n0x_L	56	47	25	22	413	28	4	385
DomPair	[D1] V-KAPPA	1n0x_L	[D1] VH	1n0x_H	59	46	25	21	596	61	7	535
DomPair			[D2] C-KAPPA	1n0x_L	15	14	6	8	126	22	5	104
DomPair			[D1] V-KAPPA	1n0x_M	4	5	2	3	52	11	1	41
DomPair			(Ligand)	1n0x_P	24	19	12	7	195	32	3	163
DomPair	[D2] C-KAPPA	1n0x_L	[D2] CH1	1n0x_H	56	47	22	25	413	28	4	385
DomPair			[D1] V-KAPPA	1n0x_L	15	14	8	6	126	22	5	104

B IMGT/3Dstructure-DB Domain pair contacts

Contacts of

Domain Chain
[D1] VH 1n0x_H

with

Domain Chain
(Ligand) 1n0x_P

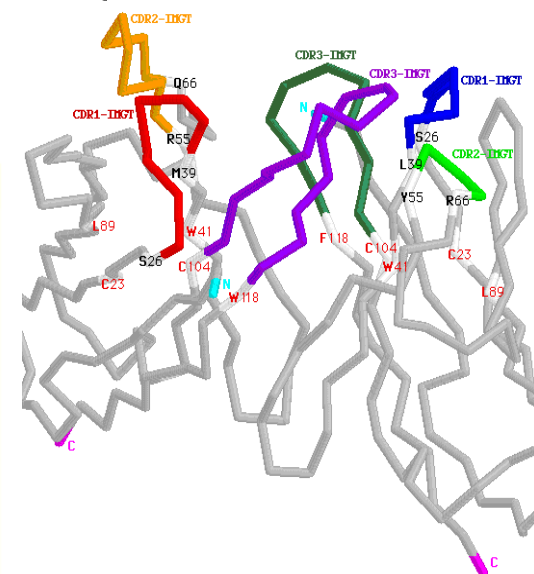
Summary:

Residue pair contacts	Number of residues			Atom pair contact types			
	Total	From 1	From 2	Total	Polar	Hydrogen	Nonpolar
17	15	7	8	106	10	1	96

List of the Residue@Position pair contacts:

Click 'R@P' for IMGT Residue@Position cards

Order				Order				Atom pair contact types			
IMGT Num	Residue	Domain	Chain	IMGT Num	Residue	Domain	Chain	Total	Polar	Hydrogen	Nonpolar
R@P	55	TRP W	[D1] VH 1n0x_H	R@P	19	ORN (Ligand)	1n0x_P	6	0	0	6
R@P	55	TRP W	[D1] VH 1n0x_H	R@P	20	LYS K (Ligand)	1n0x_P	8	0	0	8
R@P	62	ASN N	[D1] VH 1n0x_H	R@P	19	ORN (Ligand)	1n0x_P	1	0	0	1
R@P	64	ASN N	[D1] VH 1n0x_H	R@P	18	GLU E (Ligand)	1n0x_P	1	1	0	0
R@P	64	ASN N	[D1] VH 1n0x_H	R@P	19	ORN (Ligand)	1n0x_P	15	2	1	13
R@P	64	ASN N	[D1] VH 1n0x_H	R@P	20	LYS K (Ligand)	1n0x_P	8	1	0	7
R@P	65	LYS K	[D1] VH 1n0x_H	R@P	19	ORN (Ligand)	1n0x_P	1	1	0	0
R@P	65	LYS K	[D1] VH 1n0x_H	R@P	20	LYS K (Ligand)	1n0x_P	8	1	0	7
R@P	66	GLU E	[D1] VH 1n0x_H	R@P	20	LYS K (Ligand)	1n0x_P	9	2	0	7
R@P	66	GLU E	[D1] VH 1n0x_H	R@P	21	LYS K (Ligand)	1n0x_P	2	0	0	2
R@P	112.3	PRO P	[D1] VH 1n0x_H	R@P	4	SER S (Ligand)	1n0x_P	5	0	0	5
R@P	112.3	PRO P	[D1] VH 1n0x_H	R@P	5	TYR Y (Ligand)	1n0x_P	9	1	0	8
R@P	112.3	PRO P	[D1] VH 1n0x_H	R@P	6	MET M (Ligand)	1n0x_P	19	0	0	19
R@P	112.3	PRO P	[D1] VH 1n0x_H	R@P	17	ALA A (Ligand)	1n0x_P	2	0	0	2
R@P	112.2	GLN Q	[D1] VH 1n0x_H	R@P	6	MET M (Ligand)	1n0x_P	10	0	0	10
R@P	112.2	GLN Q	[D1] VH 1n0x_H	R@P	17	ALA A (Ligand)	1n0x_P	1	0	0	1
R@P	112.2	GLN Q	[D1] VH 1n0x_H	R@P	21	LYS K (Ligand)	1n0x_P	1	1	0	0



Two IMGT tools for interpretation of IG and TR sequences and NGS repertoires:

- IMGT/HighV-QUEST
- IMGT/StatClonotype

IMGT/HighV-QUEST: High throughput version of IMGT/V-QUEST

WELCOME!
to [IMGT/HighV-QUEST](#)

THE INTERNATIONAL IMMUNOGENETICS INFORMATION SYSTEM®



IMGT/HighV-QUEST version: [1.5.5](#) (9 June 2017) IMGT/V-QUEST version: [3.4.7](#) (8 June 2017) IMGT/V-QUEST reference directory release: [201723-4](#) (8 June 2017)

Citing IMGT/HighV-QUEST:

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

Please login here

E-mail:*

Password

Je ne suis pas un robot  reCAPTCHA
Confidentialité - Conditions

[New user](#) [Forgot password?](#)

- on the Web since 2010, freely available for academics (user registration required)
- deals with NGS sequences from Roche 454, Illumina, Ion Torrent, PacBio
- 2 modules: 1. Analyses up to 500,000 IG or TR rearranged sequences per batch
2. Statistics analysis (including IMGT clonotype (AA)) between batches (up to 1 million IG or TR sequence outputs from module 1).

IMG/V-QUEST results in 11 CSV files (1 line per analysed sequence, ~500 columns per line)

IMG/HighV-QUEST_main_folder (named according

1_Summary.txt

2_IMG-gapped-nt-sequences.txt

3_Nt-sequences.txt

4_IMG-gapped-AA-sequences.txt

5_AA-sequences.txt

6_Junction.txt

7_V-REGION-mutation-and-AA-change-table.txt

8_V-REGION-nt-mutation-statistics.txt

9_V-REGION-AA-change-statistics.txt

10_V-REGION-mutation-hotspots.txt

11_Parameters.txt

1. Identifies the germline V, D and J genes and alleles, score, %...

2-5. Provides sequences (nt and AA, with or without IMG gaps) with description of the V-DOMAIN

6. Analyses the V-(D)-J junction (IMG/JunctionAnalysis)

7-10. Characterizes:
nt mutations
AA changes
localisation of the hotspots

11. Parameters

IMGT/HighV-QUEST sequence analysis

1_Summary (CSV file): equivalent to Results summary of online IMGT/V-QUEST

Sequence number 1: seq1

Sequence compared with the [human IG set](#) from the [IMGT reference directory](#)

```
>seq1
caggtgcagctggtgcagctctggggctgaggtgaagaagcctggggcctcagtgaaggctc
tctgcaaggcttctggttacaccttaccagttatggtatcagctgggtgcgacaggcc
cctggacaagggcttgagtggatgggatggatcaacccttacaatggtaacacaaattat
gcacagaagctccagggcagagtcacatgaccacagacacatccacgagcacagcctac
atggagctgaggagcctgagatttgacgacacggccgtatattactgtgagagatttc
tggagcagtggtgatatctggggccaagggacaatggtcaccgtctcgagc
```

Result summary:	Productive IGH rearranged sequence (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV1-18*01 F	score = 1363	identity = 97.22% (280/288 nt)
J-GENE and allele	Homsap IGHJ3*02 F	score = 137	identity = 75.51% (37/49 nt)
D-GENE and allele by IMGT/JunctionAnalysis	Homsap IGHD6-19*01 F	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.11]	[8.8.10]	CARDFWSSGDIW

IMGT/HighV-QUEST sequence analysis

6_Junction (CSV file)

equivalent to IMGT/JunctionAnalysis detailed analysis

Input	V name	3'V-REGION	P	N	D-REGION	5'J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
seq1	Homsap IGHV1-18*01	tgtgcgagaga	t	ttctggagcagtg.....tgatatctgg	Homsap IGHJ3*02	Homsap IGHD6-19*01	0	0	0	3/6

Translation of the JUNCTION

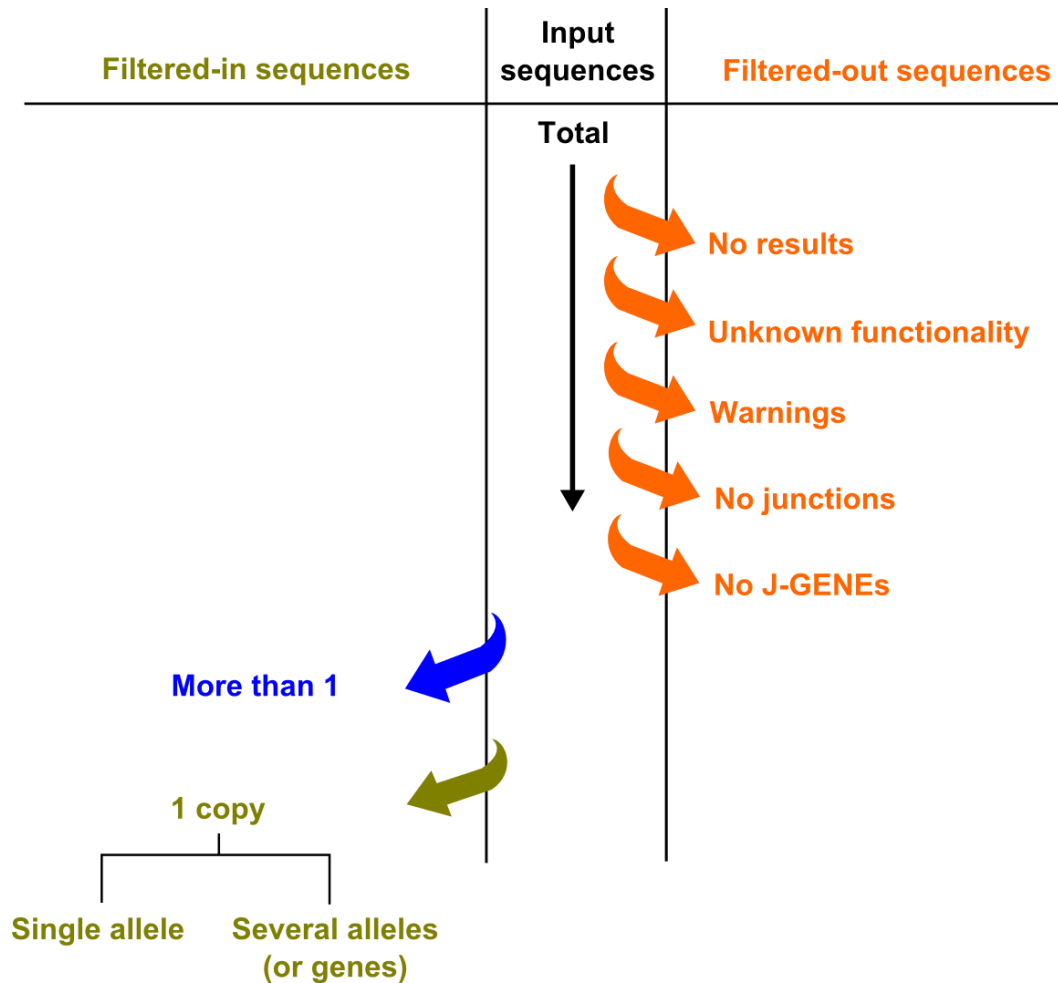
Click on mutated (underlined) amino acid to see the original one:

	104	105	106	107	108	109	113	114	115	116	117	118	Frame	CDR3-IMGT length	Molecular mass	pI	PhysicoChemical Descriptor (by BRFAA)
	C	A	<u>R</u>	D	F	W	S	S	G	D	I	W					
seq1	tgt	gcg	aga	gat	ttc	tgg	agc	agt	ggt	gat	atc	tgg	+	10	1,442.57	4.44	CARDFWSSGDIW

Be aware that some allele reference sequences may be incomplete or from cDNAs. In those cases, IMGT/JunctionAnalysis uses automatically the allele *01 for the analysis of the JUNCTION.

IMGT/HighV-QUEST Module 2: Statistics for sequence analysis interpretation

- Up to 1,000,000 sequence results
- Performed on filtered-in sequences (reliable set)



Characterization of IMGT clonotypes (AA)

An IMGT clonotype (AA) is defined by:

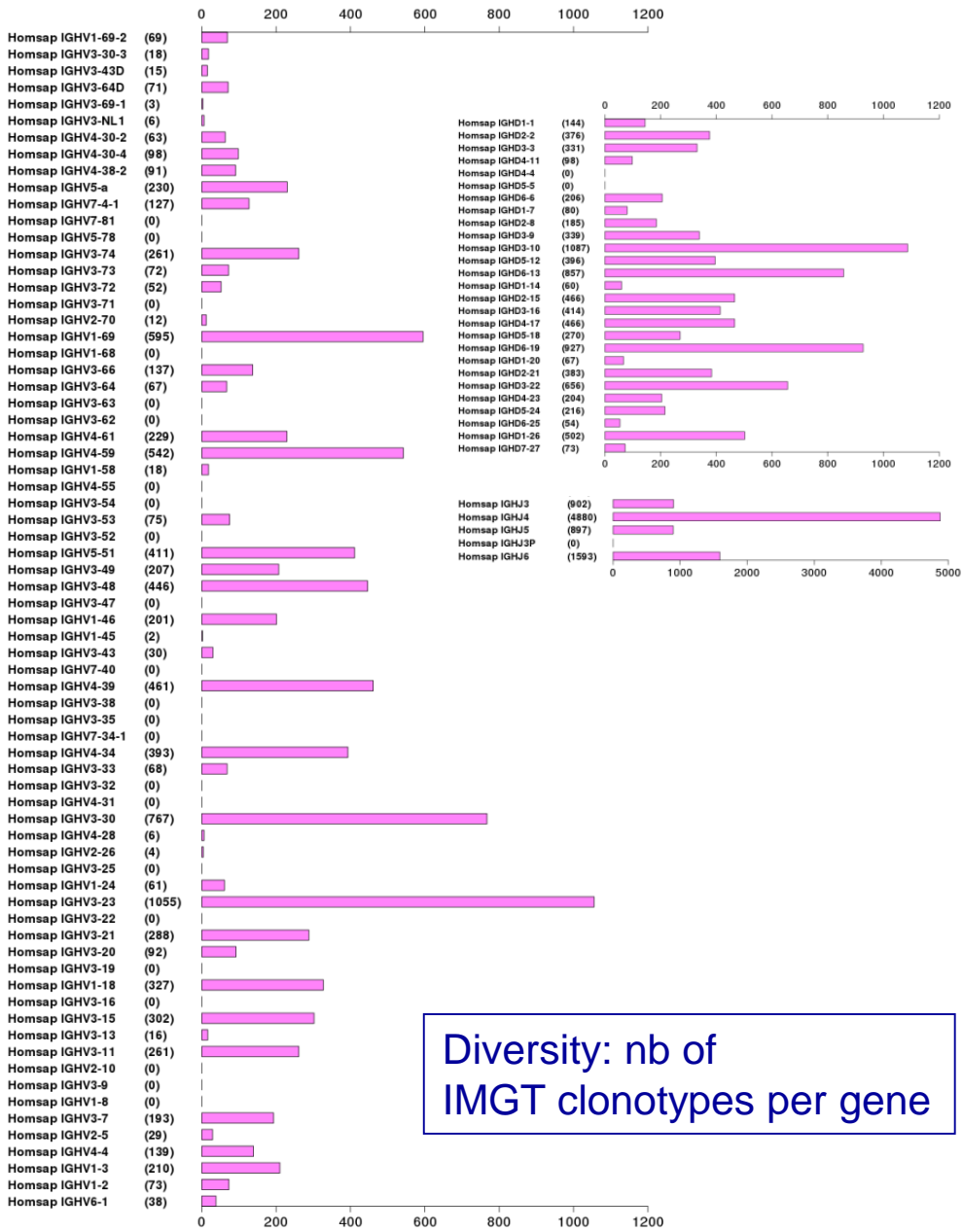
- a **unique** V-(D)-J-rearrangement (V and J genes and alleles) (nt)
- a **unique** CDR3 (AA)
- Conserved anchors 104, 118 (C104, W or F118)

An example of table for TRB IMGT clonotypes (AA)

#	Exp. ID	Nb			IMGT clonotype (AA) definition					IMGT clonotype (AA) representative sequence				IMGT clonotypes (nt)	
		Total nb of '1 copy'	Total nb of 'More than 1'	Total	V gene and allele	D gene and allele	J gene and allele	CDR3-IMGT length (AA)	CDR3-IMGT sequence (AA)	Anchors 104,118	V %	Sequence length	Functionality		Sequence ID
1	761-MID4	123	12	135	Homsap TRBV6-1*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-2*01 F	12 AA	ASLIGLDTGELF	C,F	100	493	productive	GQMC0HM04IGRCE length=493	Sequences file
2	1260-MID4	48	2	50	Homsap TRBV29-1*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-7*01 F	11 AA	SASQGSPYEYQ	C,F	100	464	productive	GQMC0HM04JAQ96 length=464	Sequences file
3	1061-MID4	48	0	48	Homsap TRBV6-1*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-2*01 F	11 AA	ASAGGDTGELF	C,F	100	493	productive	GQMC0HM04IPY4R length=493	Sequences file
4	839-MID4	47	1	48	Homsap TRBV5-1*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-7*01 F	12 AA	ASSLGQGEYEYQ	C,F	100	495	productive	GQMC0HM04JLVOJ length=495	Sequences file
5	167-MID4	39	3	42	Homsap TRBV29-1*01 F	Homsap TRBD1*01 F	Homsap TRBJ1-3*01 F	15 AA	SVEKDRGPHSGNTIY	C,F	100	463	productive	GQMC0HM04IWSJK length=463	Sequences file
6	391-MID4	38	4	42	Homsap TRBV20-1*01 F	Homsap TRBD2*01 F	Homsap TRBJ2-1*01 F	14 AA	SARSPLAGDYNEQF	C,F	100	479	productive	GQMC0HM04JHQGB length=479	Sequences file

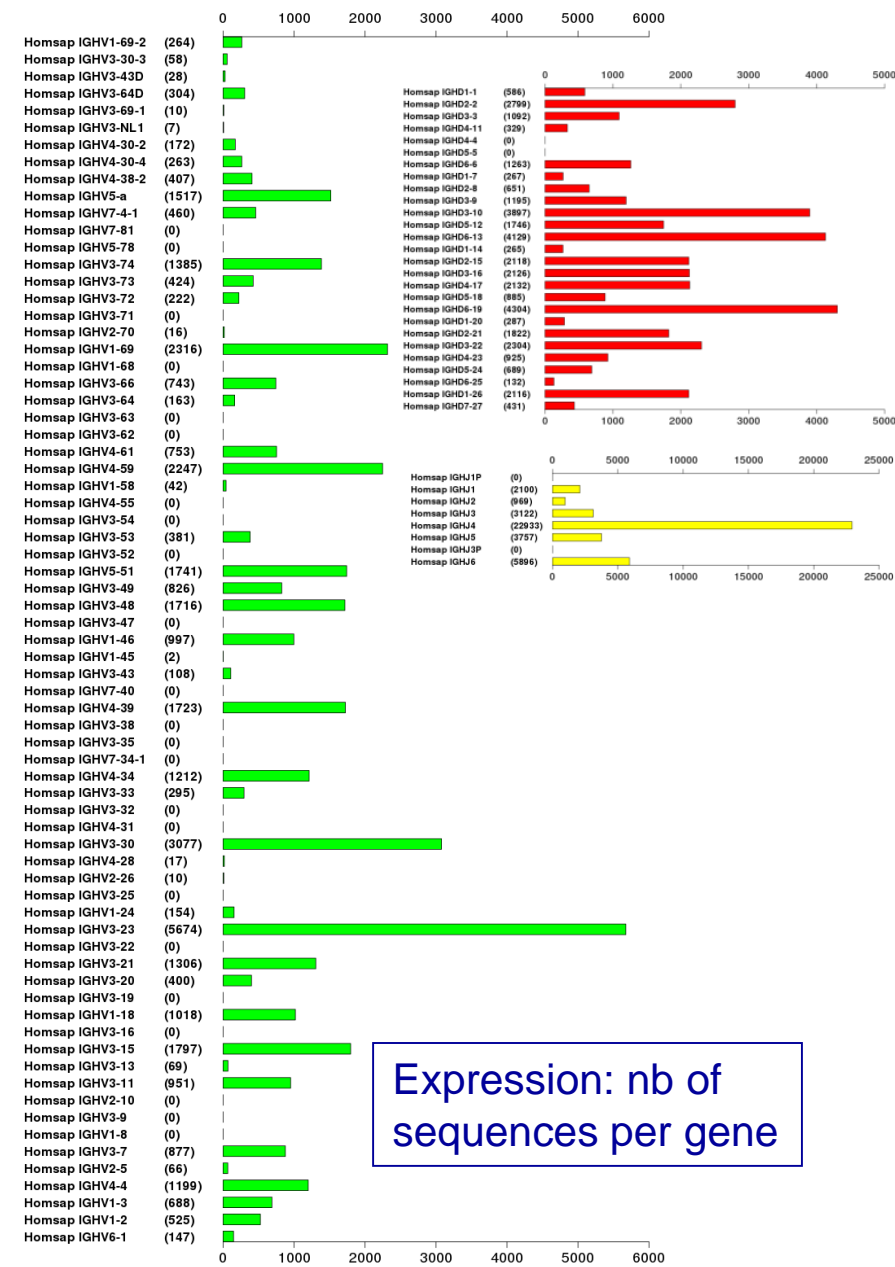
Evaluation of the clonotype diversity and expression per gene

IMGT clonotype (AA) diversity



Diversity: nb of IMGT clonotypes per gene

IMGT clonotype (AA) expression



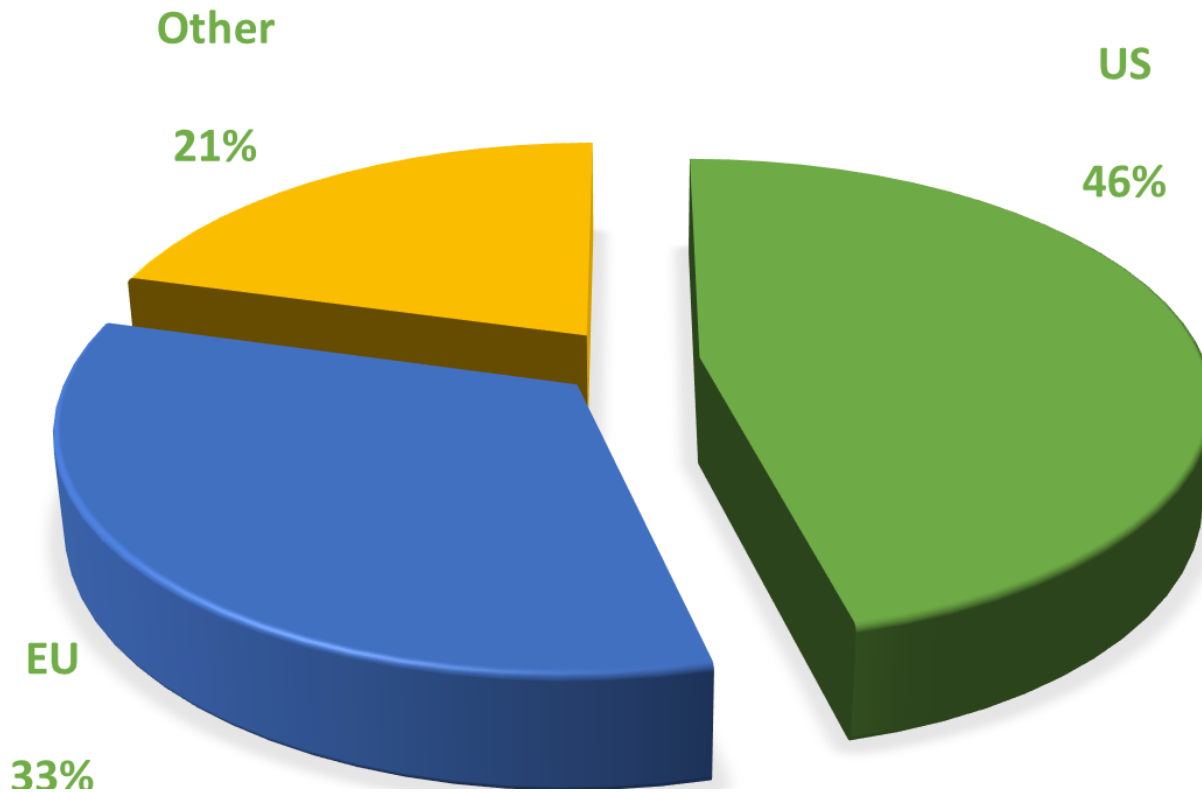
Expression: nb of sequences per gene

Identification of 'common' IMGT clonotypes (AA) in different sets

ID		IMGT clonotype (AA) definition						IMGT clonotype (AA) represe		
#	Exp. ID	CDR3-IMGT length (AA)	CDR3-IMGT sequence (AA)	V gene and allele	D gene and allele	J gene and allele	Anchors 104,118	V %	Sequence length	Functionality
ASSTRAGGPETDTQY										
1	204-Treg	16 AA	ASSTRAGGPETDTQY	Homsap TRBV7-6*01 F	Homsap TRBD2*02 F	Homsap TRBJ2-3*01 F	C104,F118	100	495	productive
	245-CD4n	16 AA	ASSTRAGGPETDTQY	Homsap TRBV7-6*01 F	Homsap TRBD2*02 F	Homsap TRBJ2-3*01 F	C104,F118	100	501	productive
ASSLGLAGLTDQY										
2	828-Treg	14 AA	ASSLGLAGLTDQY	Homsap TRBV5-1*01 F	Homsap TRBD2*01 F	Homsap TRBJ2-3*01 F	C104,F118	100	487	productive
	1074-CD4n	14 AA	ASSLGLAGLTDQY	Homsap TRBV7-2*01 F	Homsap TRBD2*01 F	Homsap TRBJ2-3*01 F	C104,F118	98.19	411	productive
ASSLGLAGPTDQY										
3	832-Treg	14 AA	ASSLGLAGPTDQY	Homsap TRBV5-1*01 F	Homsap TRBD2*01 F	Homsap TRBJ2-3*01 F	C104,F118	99.63	460	productive
	1075-CD4n	14 AA	ASSLGLAGPTDQY	Homsap TRBV5-1*01 F	Homsap TRBD2*01 F	Homsap TRBJ2-3*01 F	C104,F118	100	434	productive
ASSLLGGQGANEQF										
4	864-Treg	14 AA	ASSLLGGQGANEQF	Homsap TRBV27*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-1*01 F	C104,F118	100	505	productive
	1102-CD4n	14 AA	ASSLLGGQGANEQF	Homsap TRBV27*01 F	Homsap TRBD1*01 F	Homsap TRBJ2-1*01 F	C104,F118	100	481	productive

IMG/HighV-QUEST (web portal for NGS)

15.8 billions of IG and TR sequences analysed



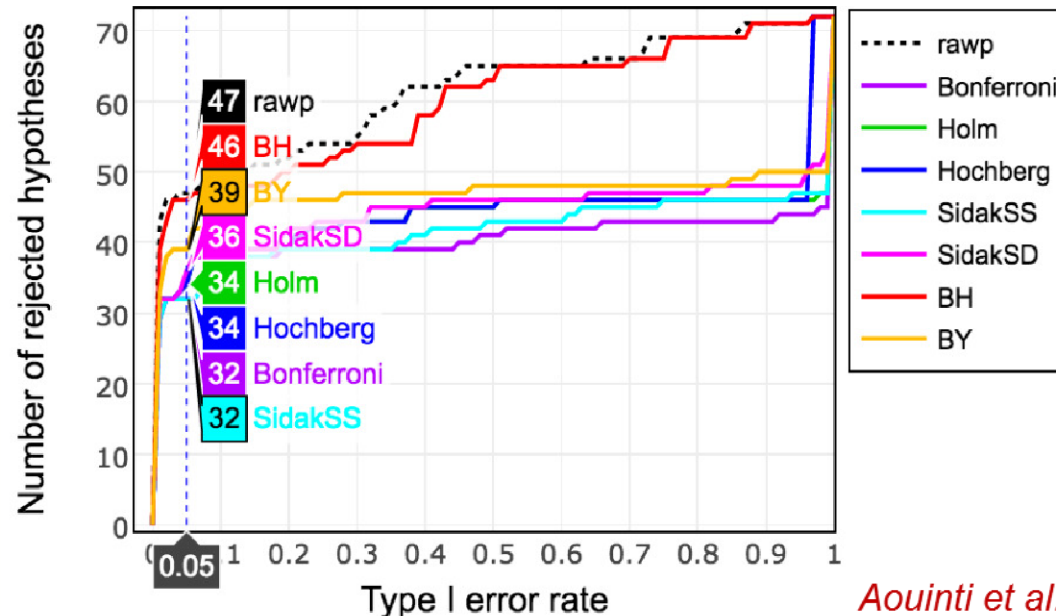
2419 users from 46 countries, March 2018

IMGT/StatClonotype: significance of differences in proportions

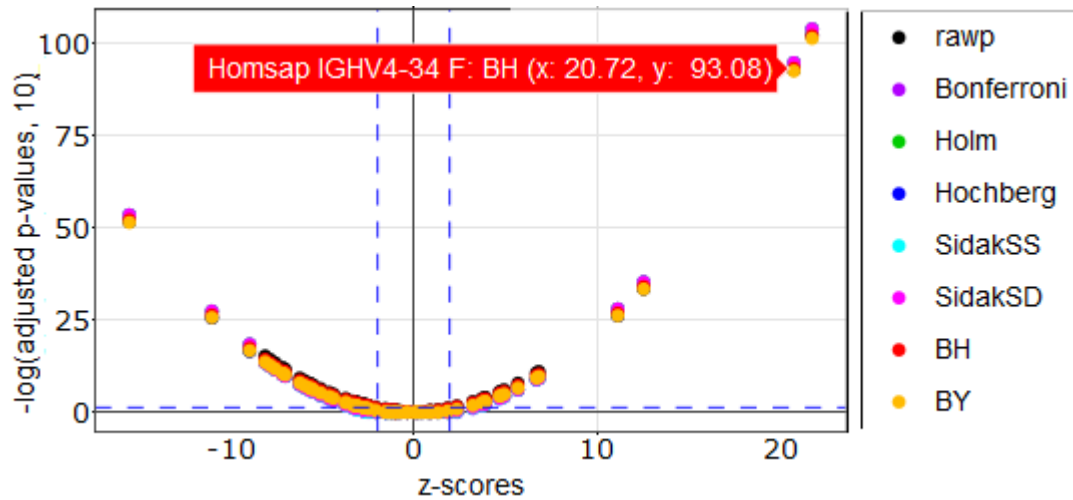
- Statistic test : z-score (exact Fisher's test for low or null occurrences)
- Adjustment of the p -values with 7 multiple testing procedures

Procedures	Type of control	Algorithm structure	dependence of p-values under H_0	Properties
Bonferroni	FWER	Single-step	Ignorance	The most conservative
Šidák (SS)	FWER	Single-step	Independence	Less conservative than Bonferroni
Holm	FWER	Step-down	Ignorance	Less conservative than Bonferroni
Šidák (SD)	FWER	Step-down	Dependence	Similar to Holm
Hochberg	FWER	Step-up	Independence	Step-up of Holm
Benjamini & Hochberg (BH)	FDR	Step-up	Independence	The least conservative
Benjamini & Yekutieli (BY)	FDR	Step-up	Ignorance	More conservative than BH

! We assume that the two compared sets are independent and the individual tests are independent of each other (i.e., multiple hypotheses are independent).



IMGT/StatClonotype: Multiple testing procedures plots for genes (also available for alleles)

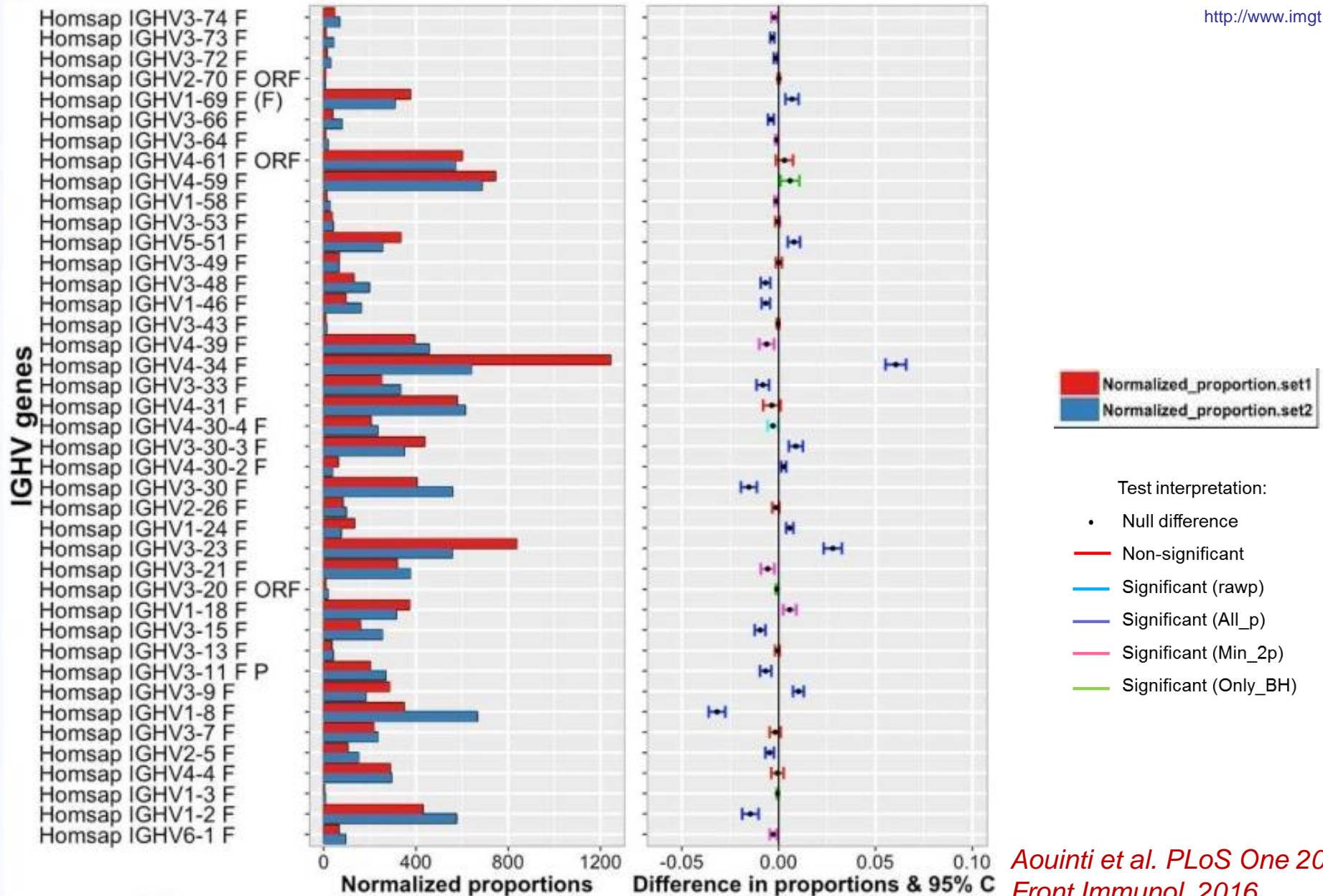


Scatter plot

Gene_Name	z (Test statistic)	-log ₁₀ (rawp)	-log ₁₀ (Bonferroni)	-log ₁₀ (Holm)	-log ₁₀ (Hochberg)	-log ₁₀ (SidakSS)	-log ₁₀ (SidakSD)	-log ₁₀ (BH)	-log ₁₀ (BY)
All	All	All	All	All	All	All	All	All	All
16 Homsap IGHV3-43 F	-1.01	0.5051566523996588	0	0	0.0140828634659759	8.34914720542639e-13	0.000831390732409686	0.411252021128257	0
17 Homsap IGHV4-39 F	-3.2	2.86192607708363	1.00459358065236	1.29372435301663	1.29372435301663	1.02561862489211	1.30442805347286	2.56089608141964	1.87418742205409
18 Homsap IGHV4-34 F	20.72	94.6407615059182	92.7834290094869	92.7895031571991	92.7895031571991	94.6407615059182	94.6407615059182	93.0844590051509	92.3977503457853
19 Homsap IGHV3-33 F	-5.08	6.42315798117749	4.56582548474622	4.71558780507955	4.71558780507955	4.56583130376861	4.71559190297203	5.90824816556843	5.22153950620287
20 Homsap IGHV4-31 F	-1.53	0.89957179774879	0	0	0.0140828634659759	0.0000266699546555017	0.0200607816043128	0.741209305653541	0.0545006462879823

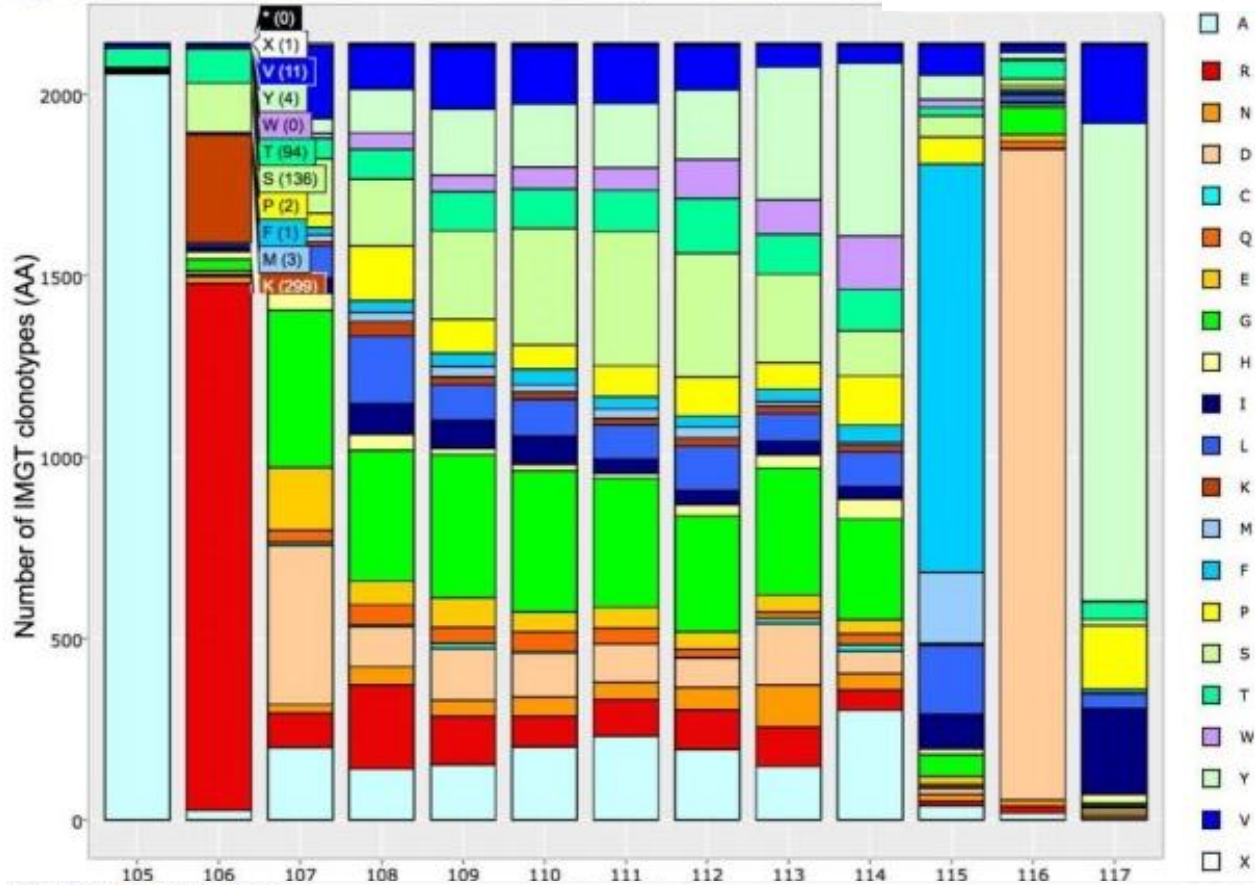
IMGT/StatClonotype: Statistical significance

Results: Test interpretation is displayed in Synthesis graphs



IMGT/StatClonotype: CDR-IMGT AA Properties

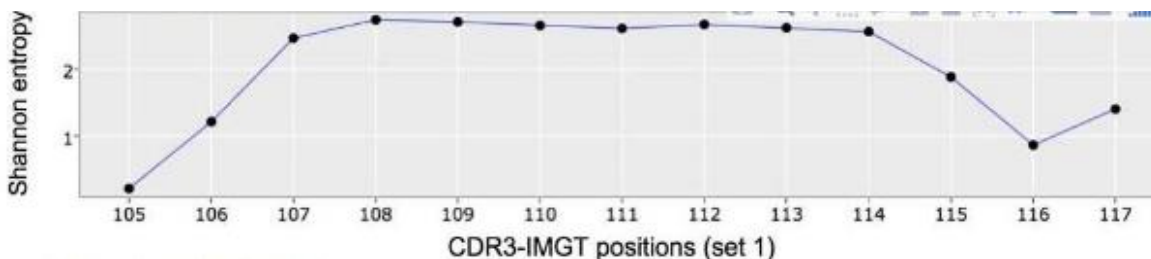
Bar graph of the 20 amino acids at CDR3-IMGT positions (set 1) (here CDR3-IMGT length=13 AA)



Displays:

- . 20 amino acids
- . Physicochemical
- . Hydropathy
- . Volume
- . Chemical
- . Charge
- . Hydrogen donor or acceptor atoms
- . Polarity

Variability plot



Variability indexes :

- . Shannon entropy
- . Wu-Kabat variability
- . Simpson index

Im
Muno
Gene
Tics



Information
system®

<http://www.imgt.org>