

IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunoinformatics

Marie-Paule Lefranc
Université Montpellier, CNRS
Institut Universitaire de France

Singapore Immunoinformatics Symposium
March 1, 2004, Institute for Infocomm Research, Singapore



The international ImMunoGeneTics information system®
Coordinator: M.-P. Lefranc, Montpellier, France <http://imgt.cines.fr>



What is an ontology? and Why?

An **ontology** is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.

What is an ontology?

An **ontology** is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.

What is an ontology?

An **ontology** is the definition of the concepts...

Example: a gene

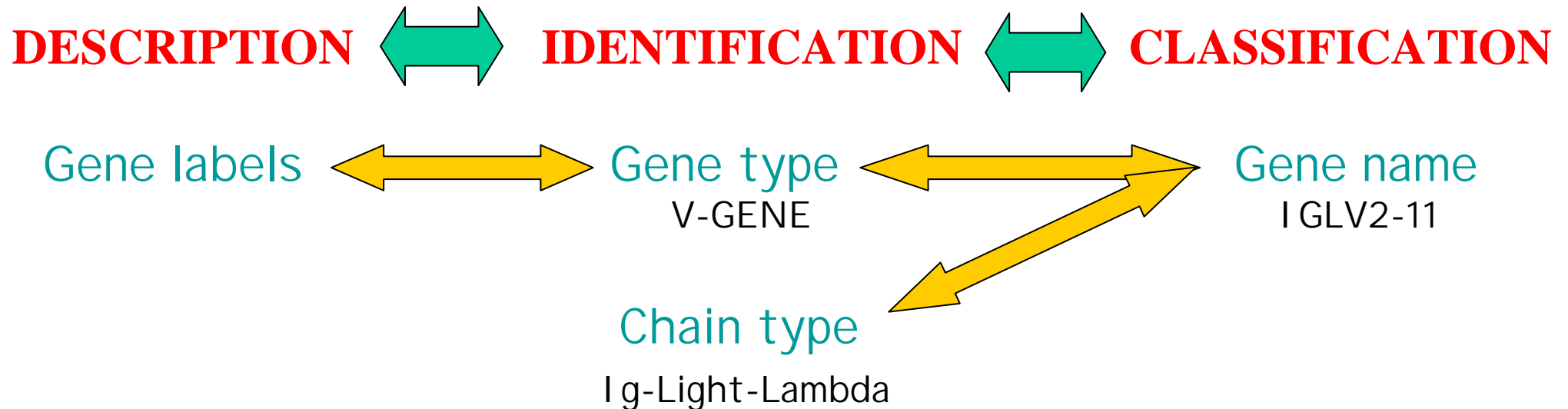
- gene type: concept of **IDENTIFICATION**
- gene name: concept of **CLASSIFICATION**
- gene labels: concept of **DESCRIPTION**



Controlled vocabulary

What is an ontology?

An **ontology** is the definition of the concepts and of their relations...



- In many ontologies, no distinction between « concepts » and « instances »
- In GO (GeneOntology), only 2 types of relations: « is a », « is part of »

Why an ontology?

An **ontology** is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.



Human beings



Information systems

Why an ontology?

An **ontology** is the definition of the concepts and of their relations, necessary to share, to reuse and to represent the knowledge, in a domain.



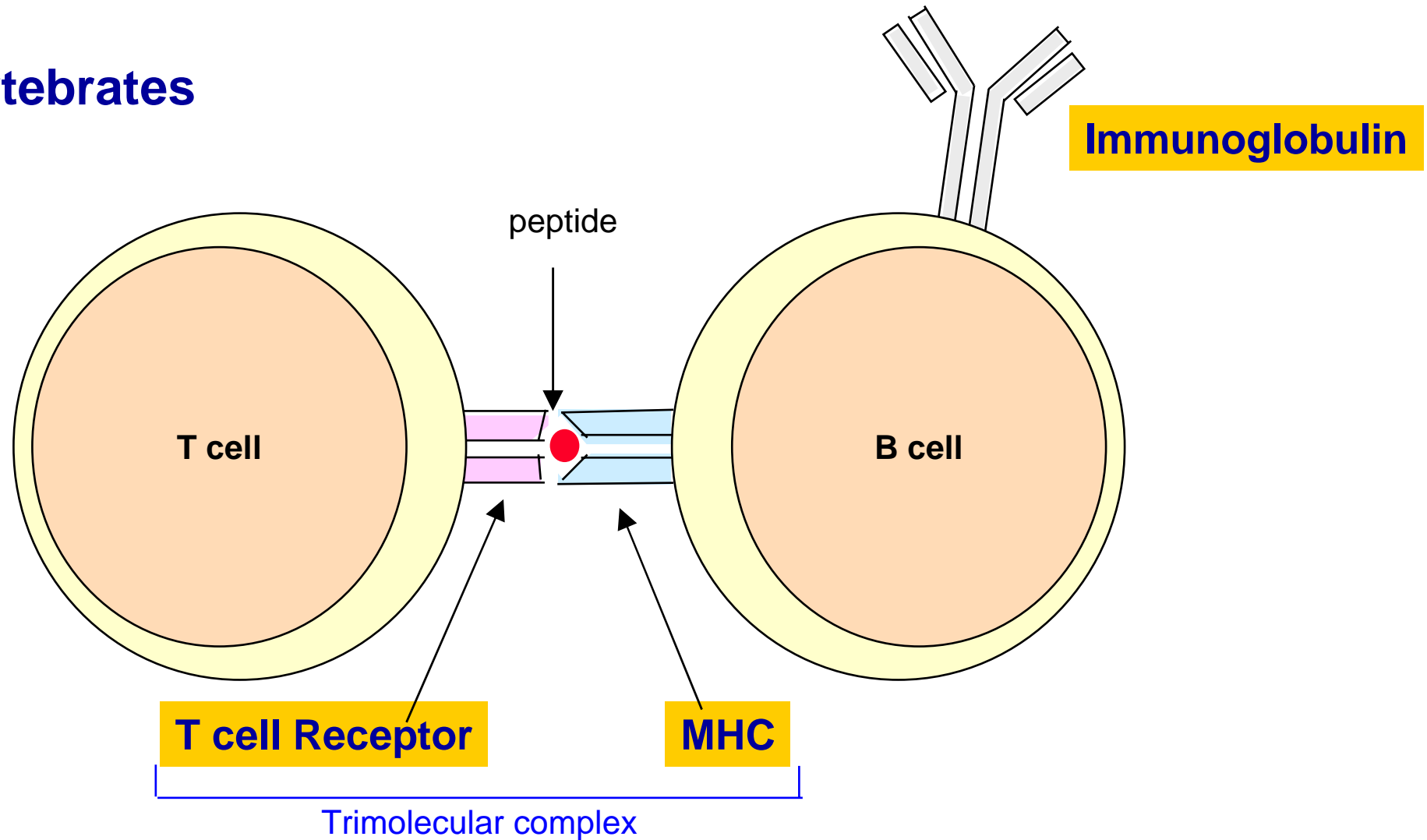
Immunology



Immunoinformatics

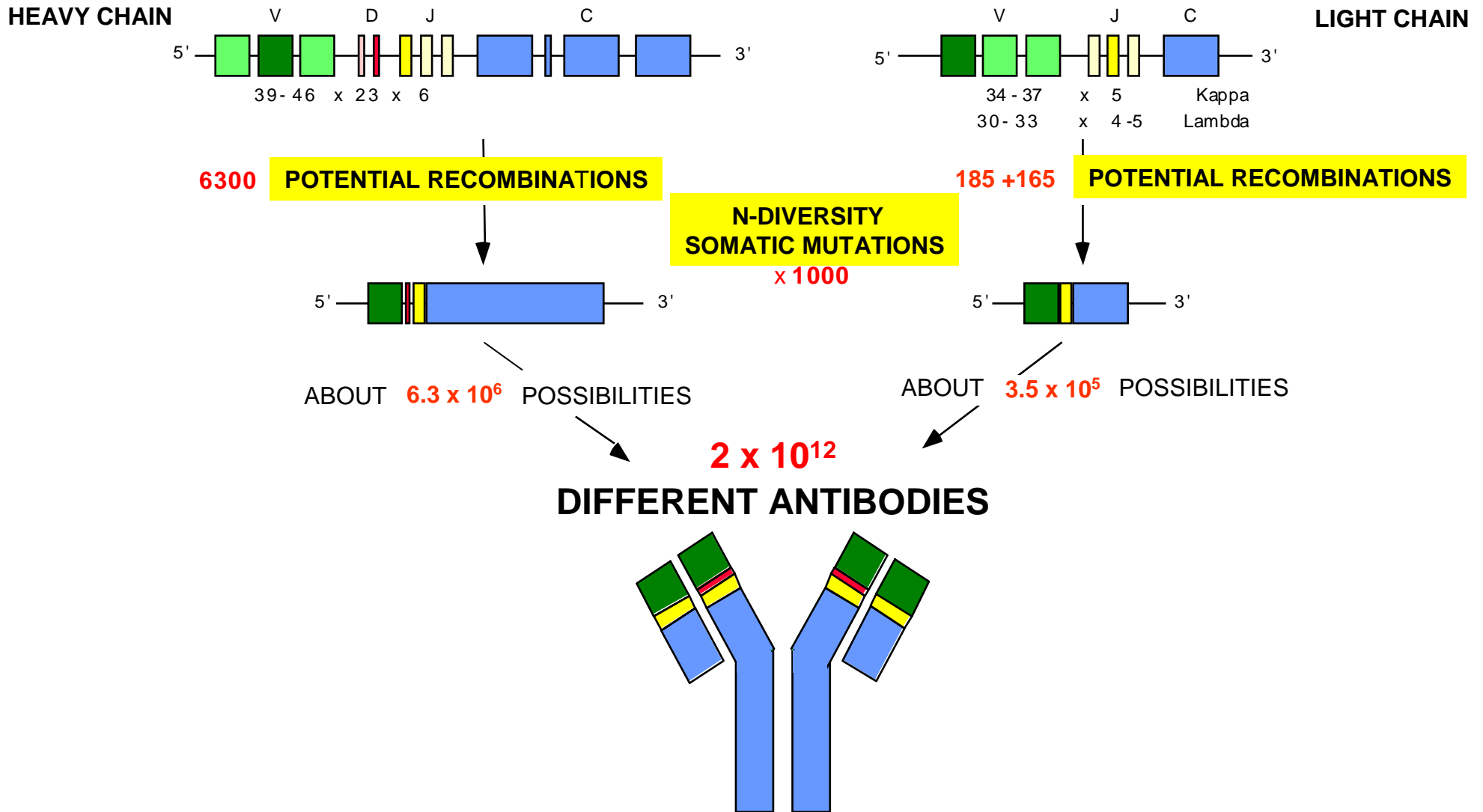
IMGT domain of research: the adaptive immune system

Vertebrates



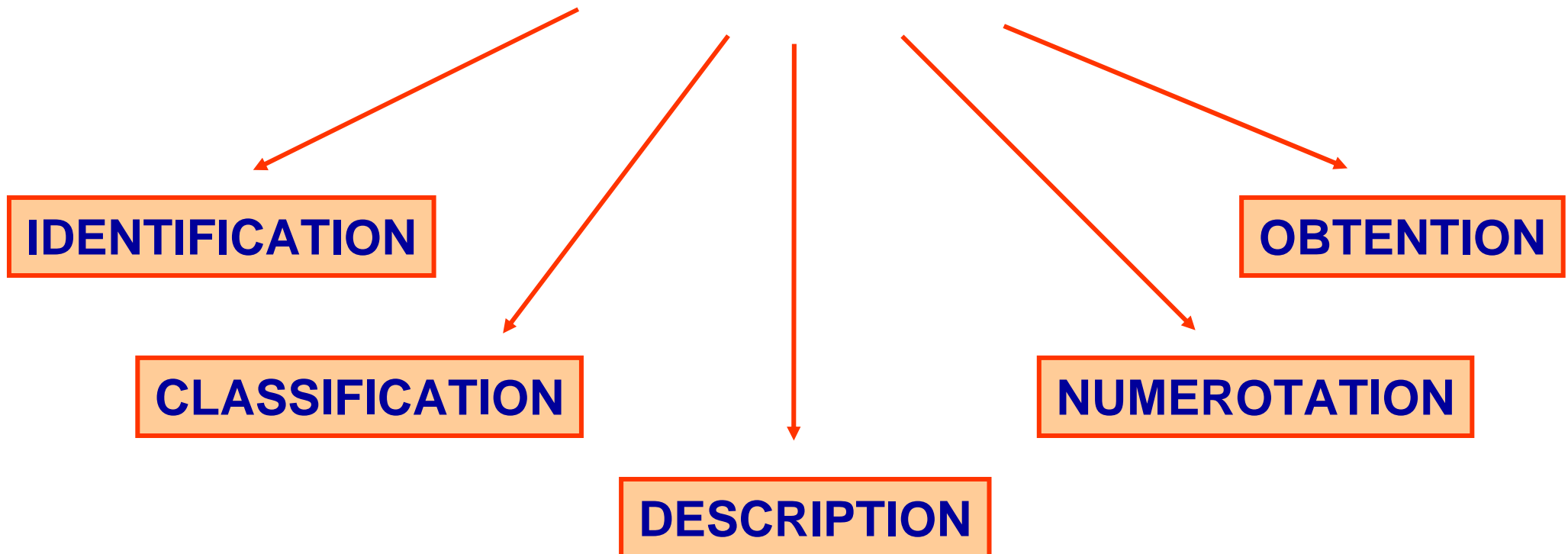
Immunoglobulin (IG) and T cell receptor (TR) synthesis

150 FUNCTIONAL IG GENES

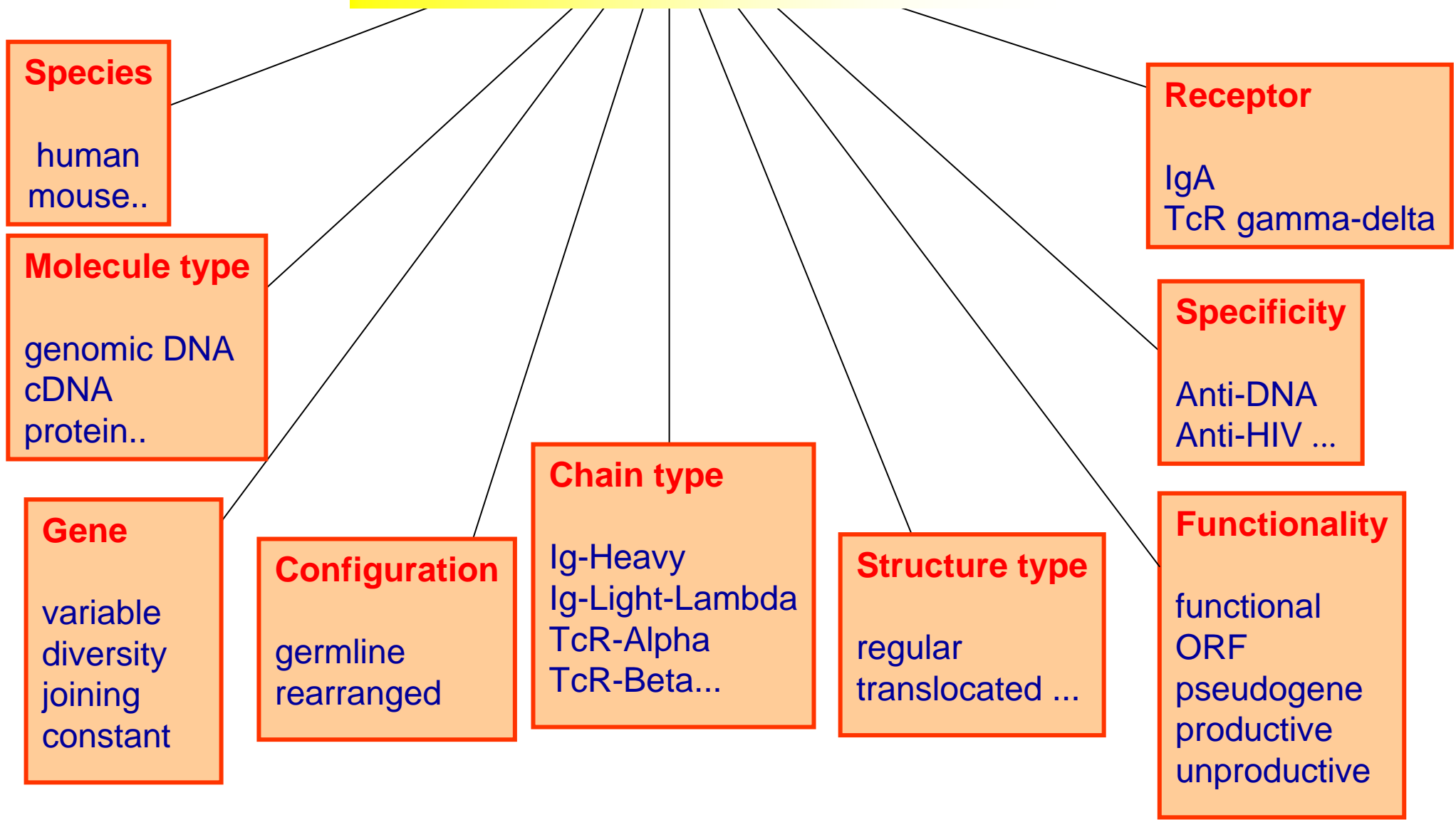


IMGT-ONTOLOGY five main concepts

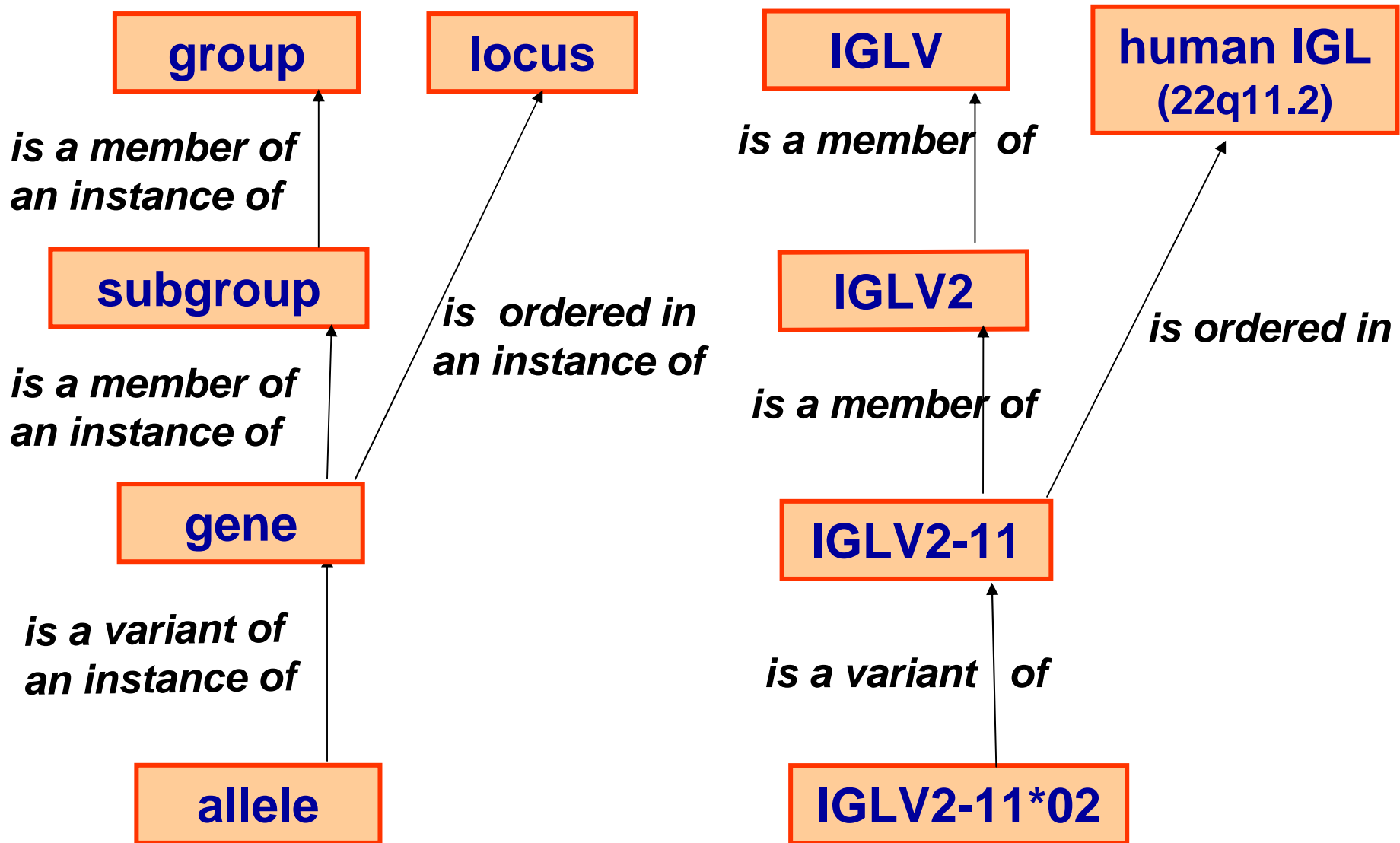
to share, reuse and represent knowledge
in immunogenetics



"IDENTIFICATION" concept



"CLASSIFICATION" concept



« Concepts »

« Instances »



Locus representation: Human IGL

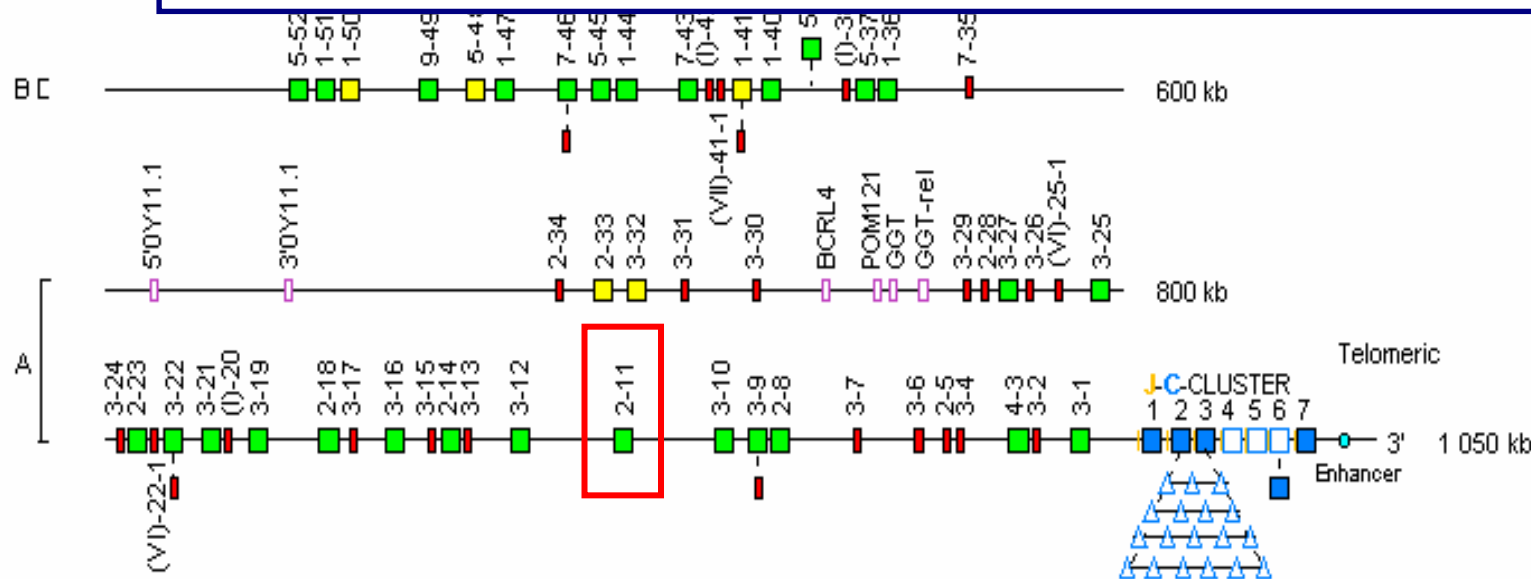
Human IGL 2

WELCOME! to IMGT/GENE-DB

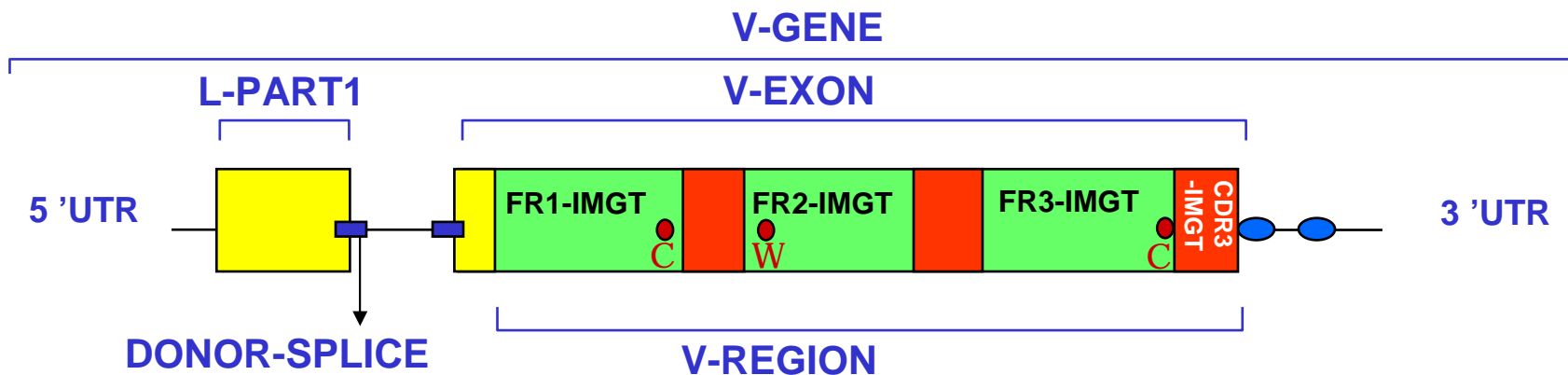
THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://imgt.cines.fr>



"DESCRIPTION" concept



Label 1

Label 2

V-GENE

V-EXON

FR3-IMGT

CDR3-IMGT

L-PART1

DONOR-SPLICE

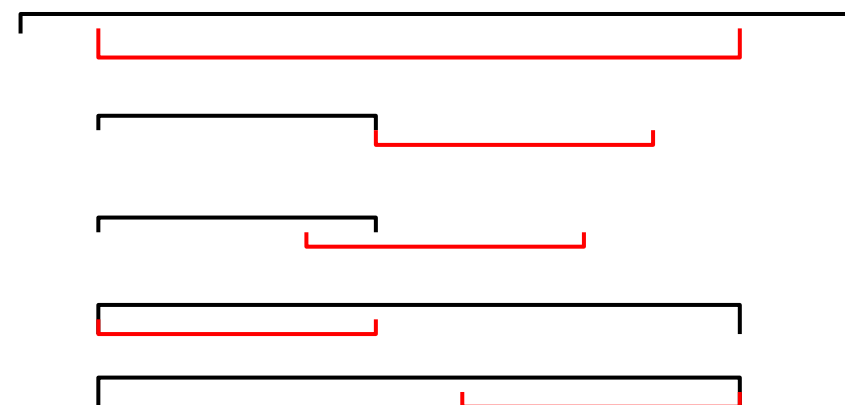
V-REGION

FR1-IMGT

V-REGION

CDR3-IMGT

Relations entre Labels



IMGT/LIGM-DB Consultation module v3 - Netscape
 Fichier Edition Afficher Aller Communicator Aide
 Signets Adresse : <http://ligm.igh.cnrs.fr:8104/cgi-bin/IMGTlect.jv> Infos connexes

FT	V-GENE	<1..297>
FT		(partial)
FT		
FT		
FT		
FT		
FT		
FT		
FT	V-REGION	
FT		
FT		
FT		
FT		
FT	FR1-IMGT	
FT		
FT		
FT	1st-CYS	
FT	CDR1-IMGT	
FT		
FT		
FT	FR2-IMGT	103..153
FT		/AA_IMGT="39 to 55"
FT		/translation="VSWYQQHPGKAPKLMIIY"
FT	CONSERVED-TRP	109..111
FT	CDR2-IMGT	154..162
FT		/AA_IMGT="56 to 58"
FT		/translation="DVS"
FT	FR3-IMGT	163..270
FT		/AA_IMGT="66 to 104, AA 73, 81, 82 missing"
FT		/translation="KRPSGVPPDRFSGSKSGNTASLTISGLQAEDEADYYC"
FT	2nd-CYS	268..270
FT	CDR3-IMGT	271..297
FT		/AA_IMGT="105 to 113"
FT		/translation="CSYAGSYTF"
XX		
SQ	Sequence 297 BP; 60 A; 93 C; 71 G; 73 T; 0 other;	
	cagtctgccc tgactcagcc tcgctcagtg tccgggtctc ctggacagtc agtcaccatc	60
	tctctactca gaaccacacg tcatattgat gattataact atatctccta ataccaacac	120

[Help](#) [IMGT Home page](#) [IMGT Marie-Paule page](#) [New search](#)


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

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


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

Catalogue

● accession number, mnemonic, definition, creation date, length, [annotation level](#)



<http://imgt.cines.fr>



Document : chargé

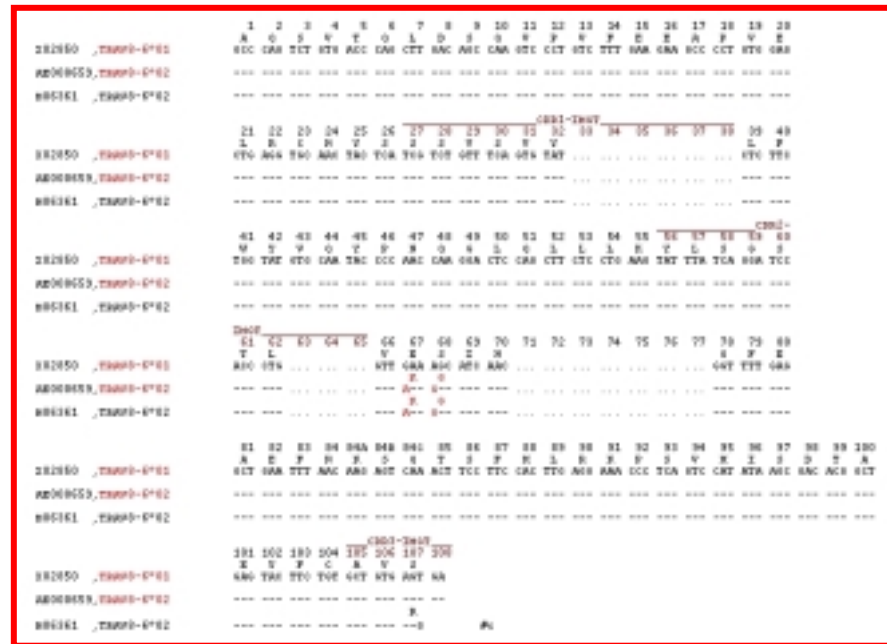
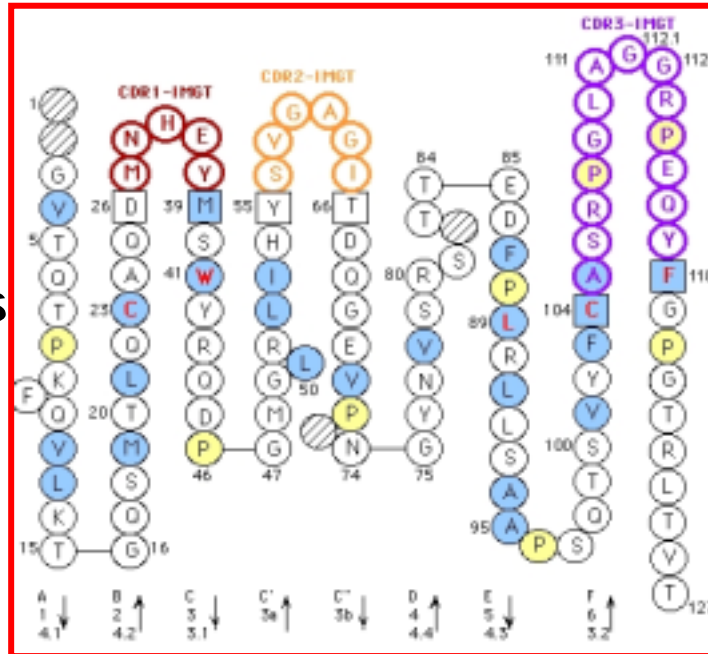


<http://imgt.cines.fr>



"NUMEROTATION" concept

Collier de Perles

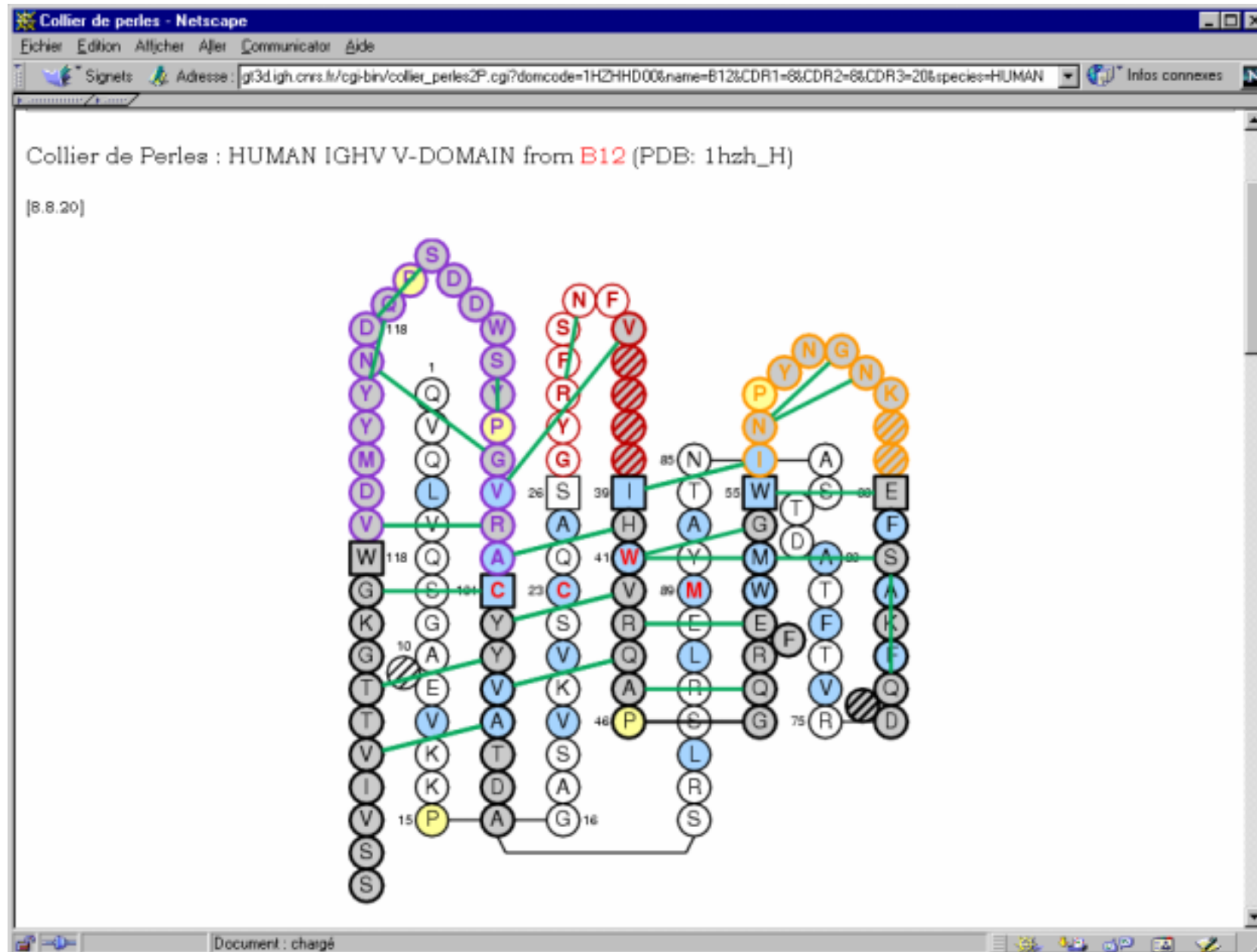


Alignment of alleles

Protein Display

TRAV gene	FR1-IMGT (1-26)	CDR1-IMGT (27-38)	FR2-IMGT (39-55)	CDR2-IMGT (56-65)	FR3-IMGT (66-104)	CDR3-IMGT (105-115)
AE000658, TRAV1-1	GQSLEQ PSEVTAVEGAIVQINCTYQ	TSGFYG.....	LSWYQQHDGGAPTFLSY	NALDGL.....	LEETG.....	RFSSFLSRSDSYGYLLQLQMKDSASYFC AVR.....
AE000658, TRAV1-2	GQNIIDQ PTEMTEGAIVQINCTYQ	TSGFNG.....	LFWYQQHAGEAPTFLSY	NVLDGL.....	LEELG.....	RFSSFLSRSKGYSYLLKELQMKDSASYLC AVR.....
AE000658, TRAV2	KDQVFQ PSTVASSEGAVVEIFCNHS	VSNAYN.....	FFWYLFHPGCCAPRLLVK	GSK.....	PSQQG.....	RYNMTYER.FSSLLILQVREADAAVYYC AVE.....
AE000658, TRAV3	AQSVAQPEDQVNVAEGNPLTVKCTYS	VSGNPY.....	LFWYVQYPNRLGQLLLK	YITGDNL.....	VKGSY.....	GFEAEFNKSTSPHLKPSALVSDSALYFC AVRD.....
AE000658, TRAV4	LAKITQ PISMDSYEGQEVNITCSHN	NIATNDY.....	ITWYQQFPSPQGRFIIQ	GYKT.....	KVINE.....	VASLFI PADRKSSITLSLPRVLSDTAVYYC LVGD.....
AE000659, TRAV5	GEDVEQS LFLSVREGDSSVINGTYT	DSSSTY.....	LYWYKQEPGAGLQLLTY	IFSNMD.....	MKQDQ.....	RLTVLLNKKDKHLRLRIADTQTGDSAIYFC AES.....
AE000659, TRAV6	SQKIEQNSEALNIQEGKTATLTCNYT	NYSPAY.....	LQWYRQDPGRGPVFLLL	IRENEK.....	EKRKE.....	RLKVTFDITLTKQSLFHITASQPADSATYLC ALD.....
AE000659, TRAV7	ENQVEHSPHFLGPQQGDVASMSTYS	VSRFNN.....	LQWYRQNTGMGPKHLLS	MYSAGY.....	EKQKG.....	RLNATLLK.NGSSLYITAVQPEDSATYFC AVD.....
AE000659, TRAV8-1	AQSVSQNHVILSEAALELGCNYS	YGQTVN.....	LFWYVQYPGQLQLLLK	YFSGDPL.....	VKGIK.....	GFEAEFIKSKFSPNLRKPSVQWSDAEYFC AVN.....
AE000659, TRAV8-2	AQSVTQLDHSVSVSEGTPVLLRCNYS	SSYSPS.....	LFWYVQHPNKGQLLLK	YTSATL.....	VKGIN.....	GFEAEFKKSETSPHLTKPSAHMSDAAEYFC VVS.....
AE000659, TRAV8-3	AQSVTQPDIIHITVSEGASLELRCNYS	YGATPY.....	LFWYVQSPGQLQLLLK	YFSGDTL.....	VQGIK.....	GFEAEFKRSQSSPNLRKPSVHWSDAEYFC AVG.....
AE000659, TRAV8-4	AQSVTQLGSHVSVSEGALVLLRCNYS	SSVPPY.....	LFWYVQYPNQGLQLLLK	YTSATL.....	VKGIN.....	GFEAEFKKSETSPHLTKPSAHMSDAAEYFC AVS.....
XO2850, TRAV8-6	AQSVTQLDSQVPVFEAPVELRCNYS	SSVSVY.....	LFWYVQYPNQGLQLLLK	YLSGSL.....	VESIN.....	GFEAEFNKSTSPHLRKPSTVHSDAEYFC AVS.....
AE000660, TRAV8-7	TQSVTQLDGHITVSEEAPELRCNYS	YSGVPS.....	LFWYVQYSSQSLQLLLK	DLTEATQ.....	VKGIR.....	GFEAEFKKSETSPYLRKPSVHWSDAEYFC AVGDR.....
AE000659, TRAV9-1	QDSVWQTEGQVLRSEGRSLIVNCSYF	TTQVPS.....	LFWYVQVPGEGPOLHLK	AMKAMD.....	KGRMK.....	GFEAMVRYKETTSEHLEKDSVQVSDSAVYFC ALS.....

IMGT Collier de Perles



V-DOMAIN 3D representation (TR A6, 1ao7)

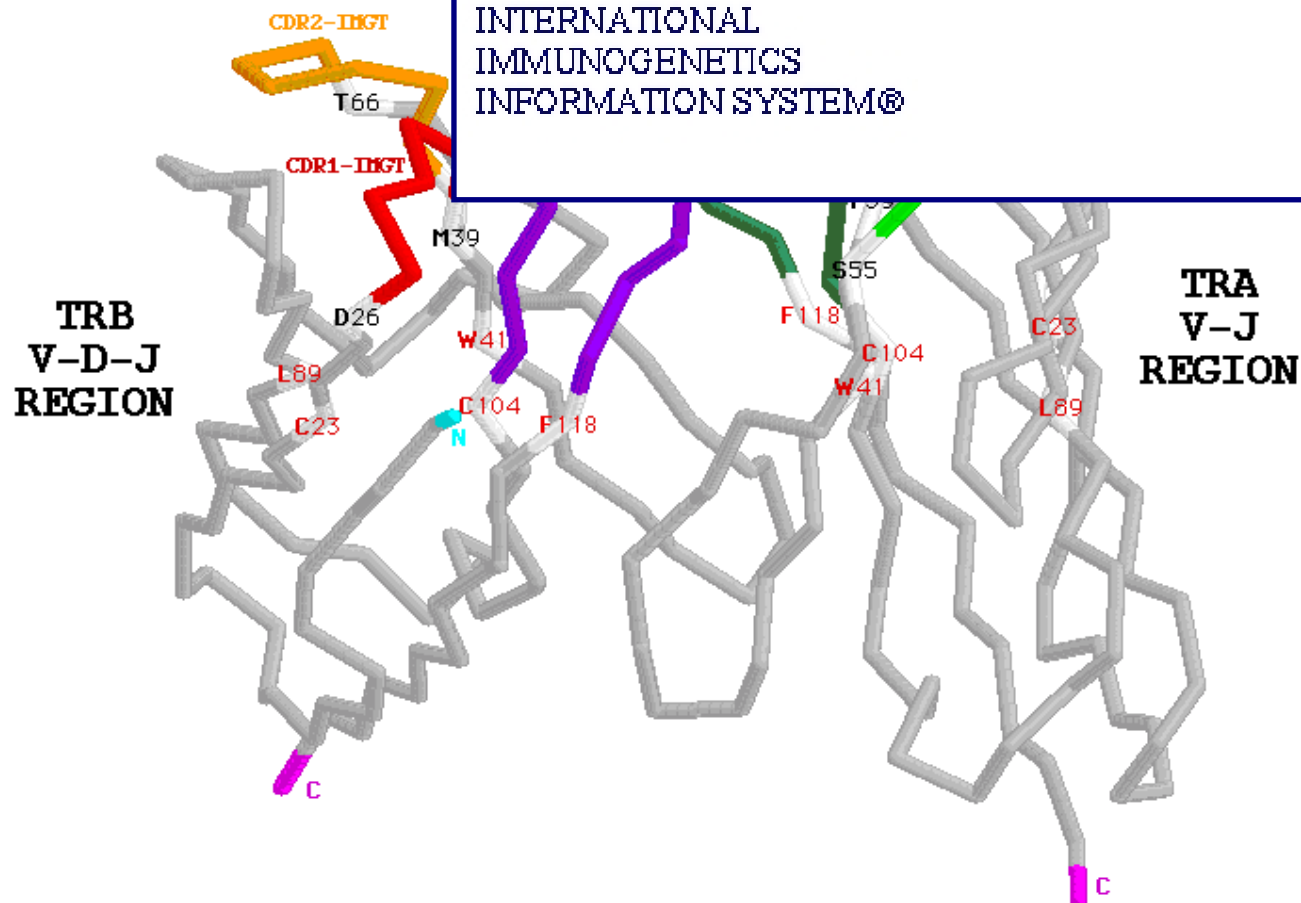
WELCOME !

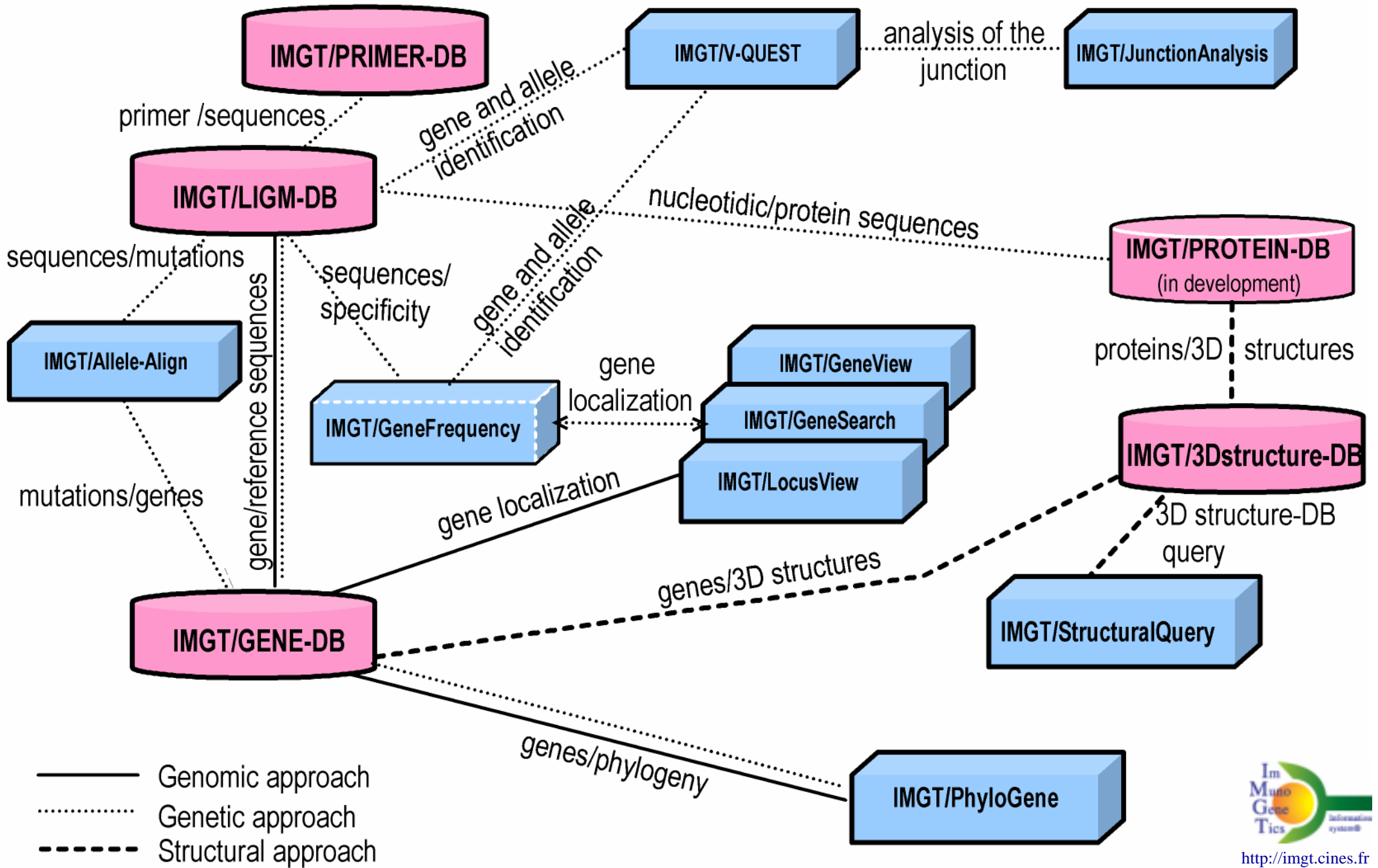
to IMGT/3Dstructure-DB

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®

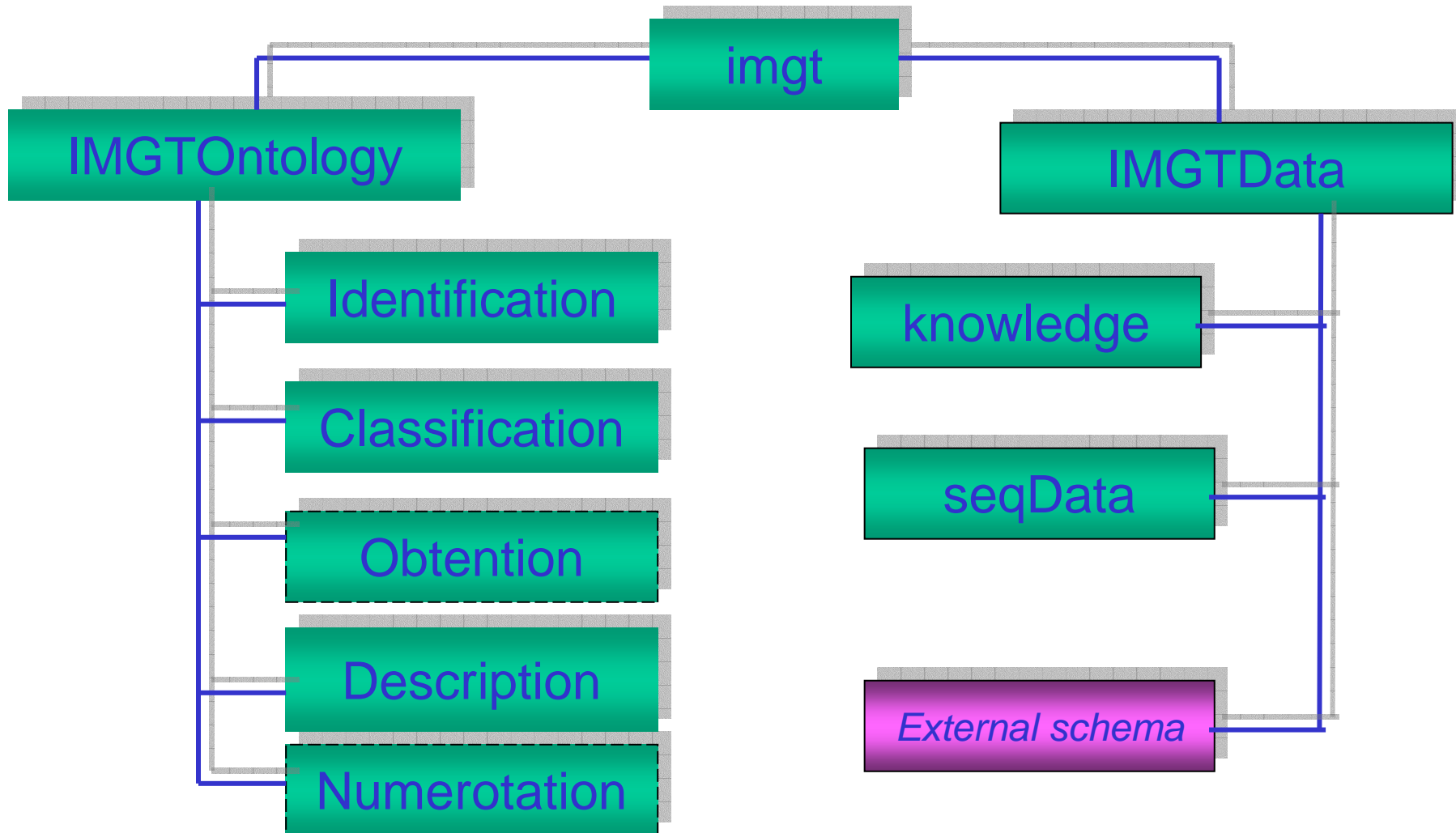


<http://imgt.cines.fr>

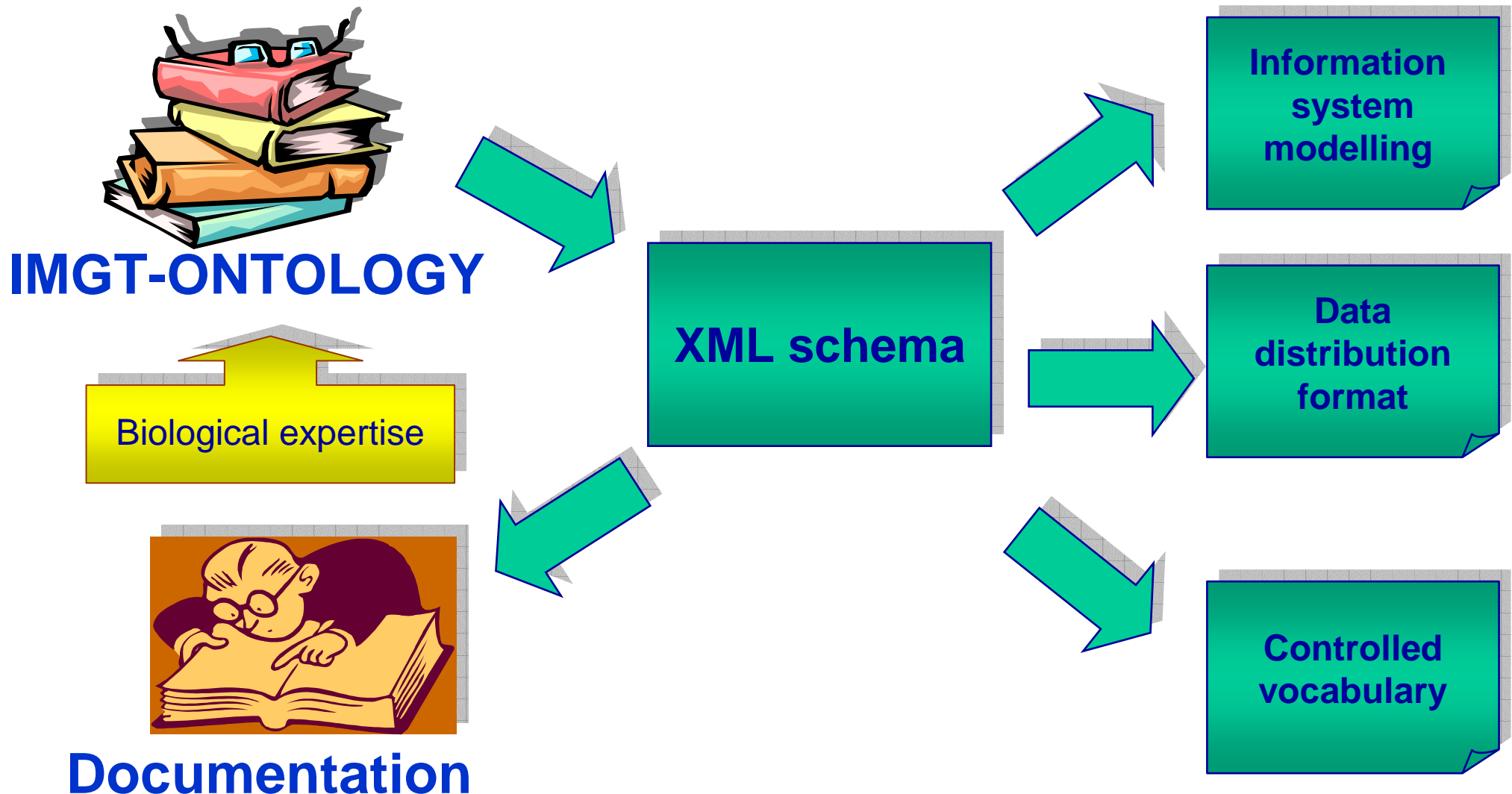




IMGT-ML schema



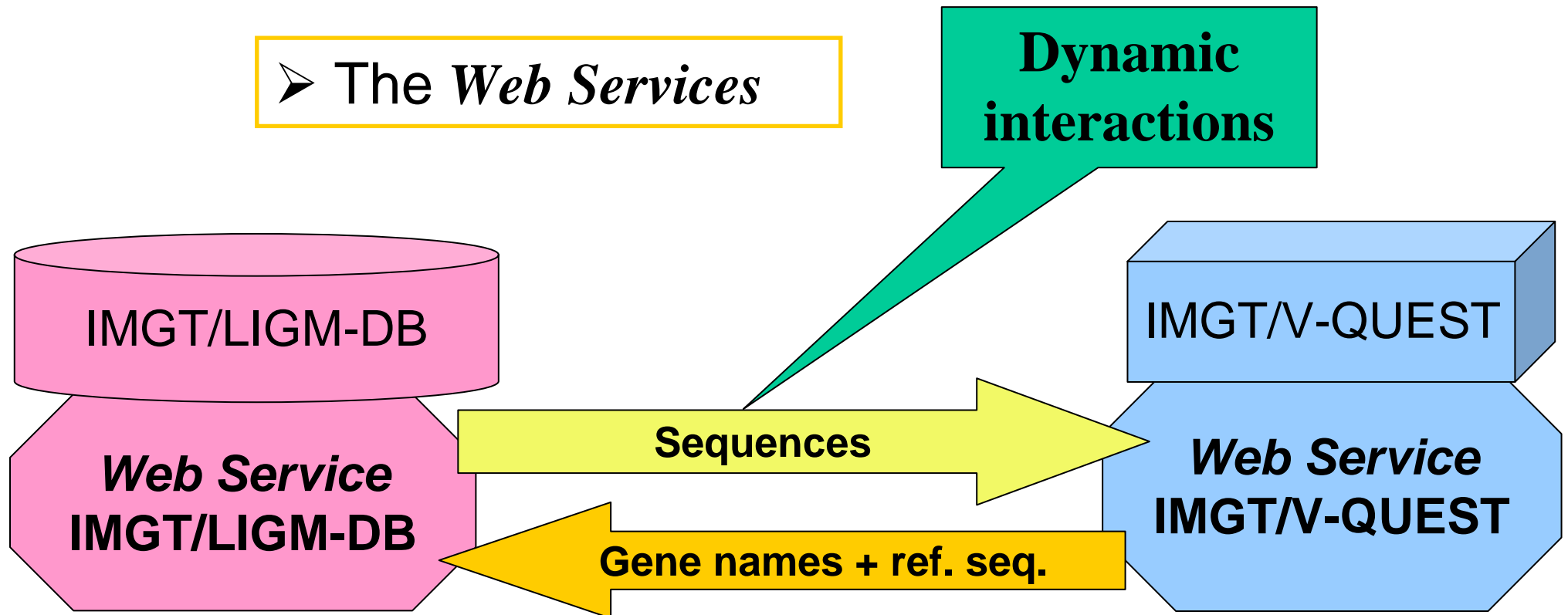
IMG-T-ML architecture



Informatic answers to the biological problems

- ❑ Use IMGT-ONTOLOGY (and IMGT-ML)
- ❑ Allow IMGT components to dynamically interact

➤ The *Web Services*



Example of IMGTV-QUEST results

Alignment for V-GENE

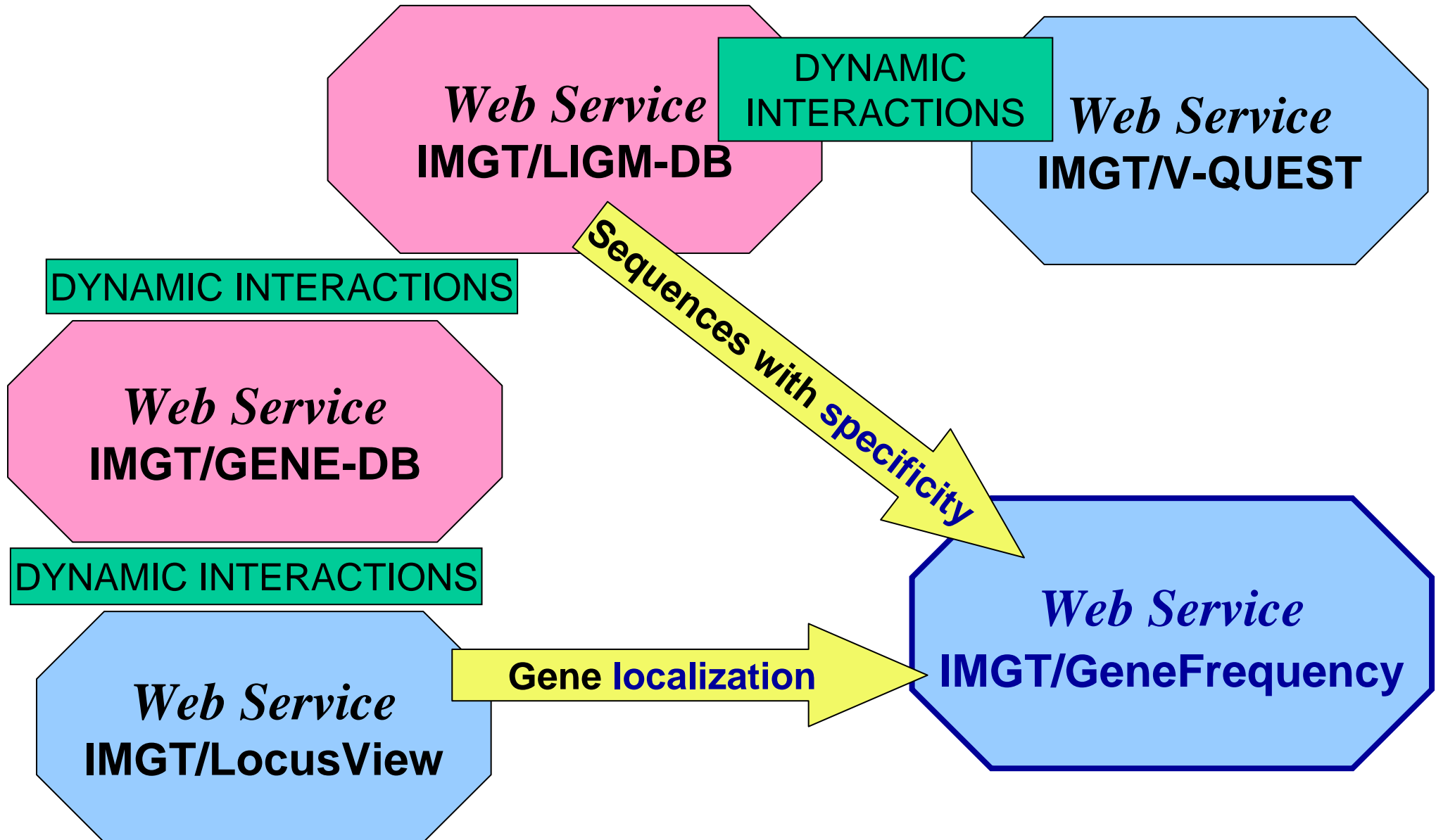
AF402940		score	GTGCAGCTGCTCGAGCAGTCTGGGGCT__GAGGTGAGCAAGCCTGGGGCCTCAGTAAAGGTTTCCTGCA
X62109	IGHV1-3*01	1146	CA.GTC.A...T.T.....AG.....G.....
X62107	IGHV1-3*02	1110	CA.GTT.A...G.T.....AG.....G.....
M99637	IGHV1-8*01	957	CA.GT..A...G.T.....AG.....G...C.....
L06612	IGHV1-46*03	948	CA.GT..A...G.T.....AG.....G.....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

Alignment for J-GENE

AF402940		score	CTTCACGGGGCGGGACGCTTTGGACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCA
J00256	IGHJ3*01	181	_____T.....T..T.....A.T.....T...G
X86355	IGHJ6*02	179	T.A.TACTACTACT...G.A.....
X86355	IGHJ3*02	172	_____T.....T..TA.....A.T.....T...G



Diagram of collaboration: Analyse de repertoires



Example of IMGT/GeneFrequency results

Your Selection :

Human IGH, IGK and IGL Locus Specificity anti-thyroid peroxidase (TPO)

For the D and J genes, the number of genes is shown between parentheses when genes names could not be indicated for a click on the zoom for the D and J genes names.

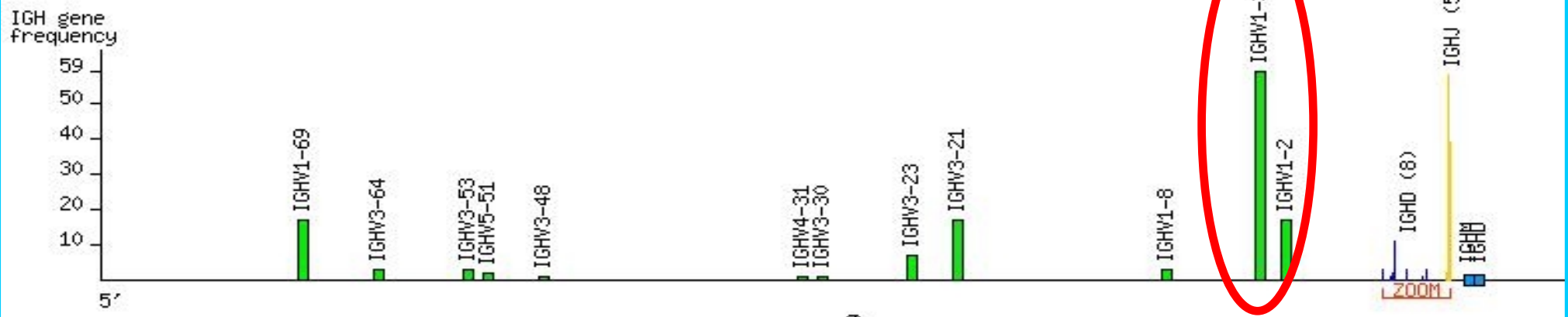
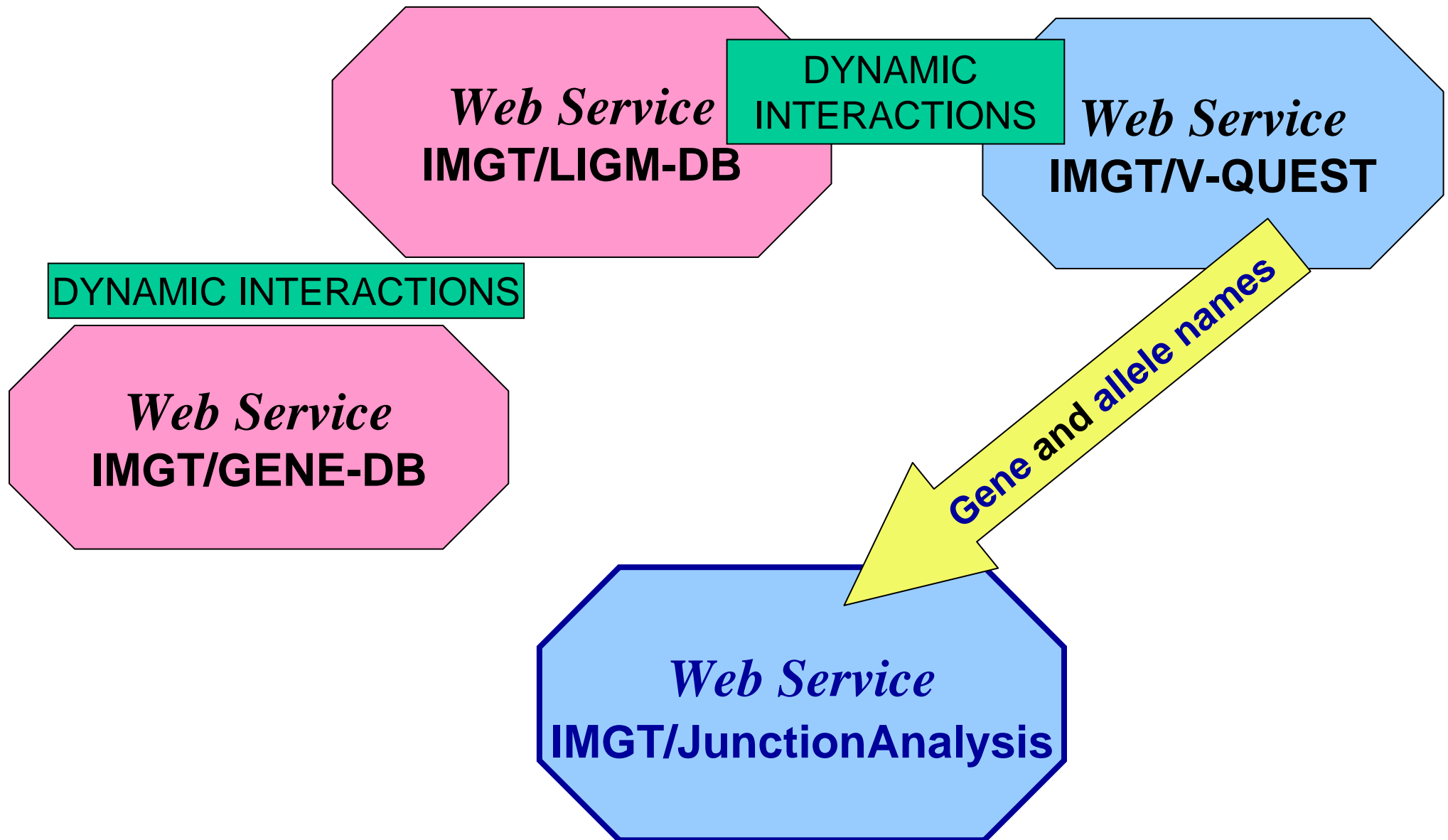


Diagram of collaboration: Analyse des jonctions



Example of IMGT/JunctionAnalysis results



<http://imgt.cines.fr>

Analysis of the JUNCTIONS

Input	V name	V-REGION	D-REGION	N2	J-REGION
#1 AF402940	IGHV1-3*01	tgtgcgagag.gcttcacgggg.....	cgggacgctttggacgtctgg

Input	J name	D name	Vmut	Dmut	Jmut	Ngc
#1 AF402940	IGHJ3*01	IGHD3-10*01	0	4	2	5/6

Translation of the JUNCTIONS

	105	107	109	112	114	116	118	CDR3-IMGT								
	104	106	108	110	113	115	117	frame length								
	C	A	R	G	F	T	G	R	D	A	L	D	V	W		
#1 AF402940	tgt	gcg	aga	ggc	ttc	acg	ggg	cgg	gac	gct	ttg	gac	gtc	tgg	+	12

THANK YOU for using IMGT/JunctionAnalysis

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://imgt.cines.fr>

Analysis of the JUNCTIONS

Input	V name	V-REGION	N1	D-REGION	N2
#1 M62724	IGHV7-4-1*02	tgttccgagaga	aga	.tagcaatggctacaa....	aata
#2 Z47269	IGHV1-69*06	tgtgcgagag.	gggggggctaaggtogaatttttggagtggtt.....	tcatgggt

Input	J-REGION	J name	D name	Vmut	Dmut	Jmut	Ngc
#1 M62724tttaactactag	IGHJ4*02	IGHD5-24*01	0	2	0	1/7
#2 Z47269	...actggttcgaccctgg	IGHJ5*02	IGHD3-3*02	0	2	0	13/20

Translation of the JUNCTIONS

	104	105	106	107	108	109	110	111	111.1	111.2	111.3	112.4	112.2	112.1	112	113	114	115	116	117	118	frame	CDR3-IMGT length	
#1 M62724	C	A	R	E	D	S	N	G							Y	K	I	F	D	Y	W	+	13	
#2 Z47269	tgt	gcg	aga	gaa	gat	agc	aat	ggc							tac	aaa	ata	ttt	gac	tac	tgg	+	20	
	C	A	R	G	G	A	K	V	F	F	L	F	W	F	H	G	V	W	F	D	P	W		
	tgt	gcg	aga	ggg	ggg	gct	aag	gtc	gaa	ttt	ttg	gag	tgg	ttt	cat	ggg	tac	tgg	ttc	gac	ccc	tgg	+	20

-> [IMGT/JunctionAnalysis Search page](#)
-> [IMGT/JunctionAnalysis Documentation](#)

IMGT-Choreography: Expressed IG and TR repertoires

THANK YOU
for using [IMGT/JunctionAnalysis](#)

THE
INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®



<http://imgt.cines.fr>

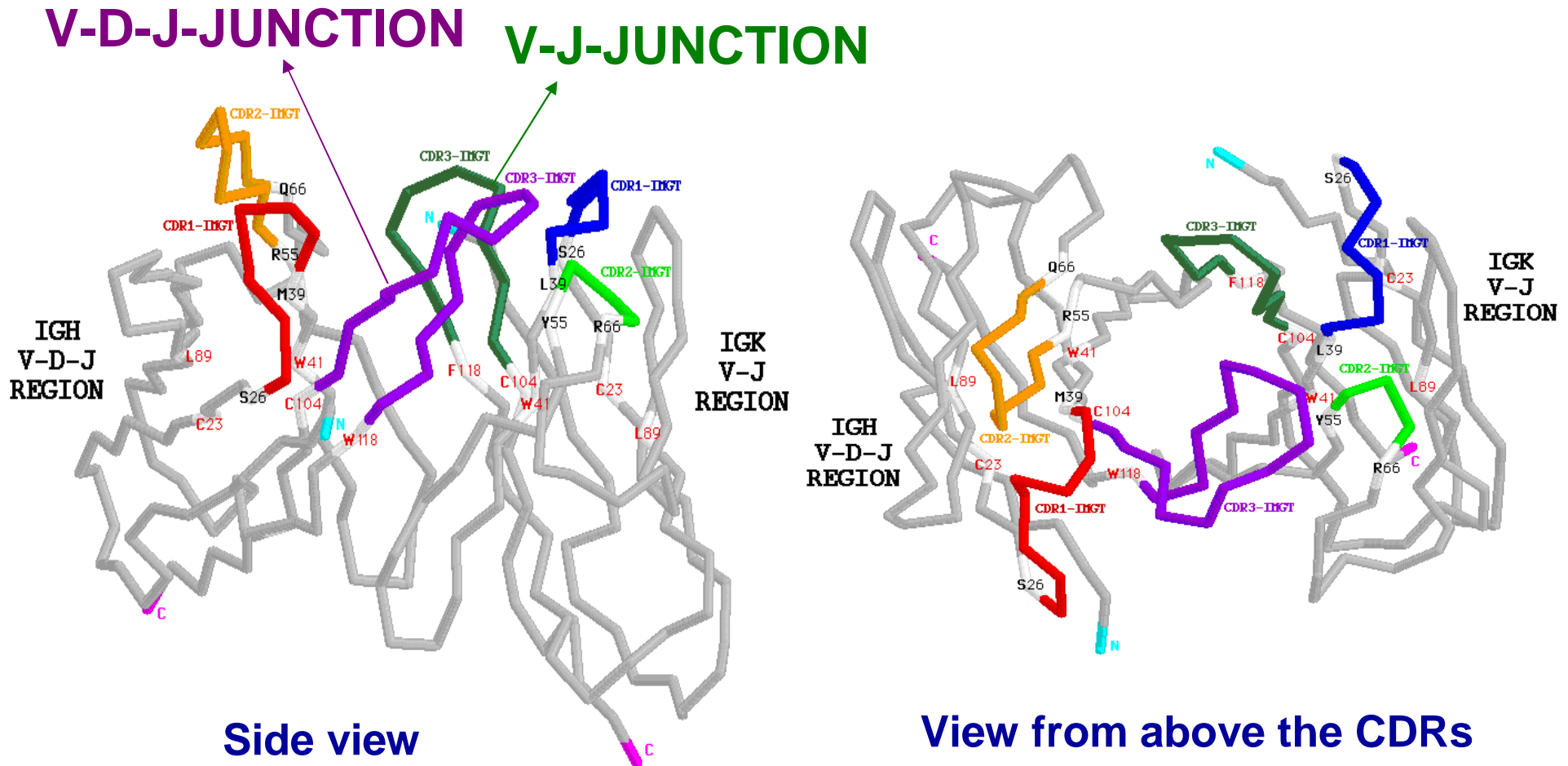
Analysis of the JUNCTIONS

	Input	V name	V-REGION	N	J-REGION	J name	Vmut	Jmut	Ngc
#1	AF490920	IGKV1-33*01	tgtcaacactatgatgatttccc...		attcactttc	IGKJ3*01	3	0	0/0
#2	AF490935	IGKV4-1*01	tgtcagcaatattatagtactctc.		..tcactttc	IGKJ4*01	0	0	0/0
#3	AF490937	IGKV4-1*01	tgtcagcaatattatagtggctctcc		.gtacactttt	IGKJ2*01	2	0	0/0
#4	AF490932	IGKV3-15*01	tgtcagcactataataactggcctcc	cc	tgtacactttt	IGKJ2*01	1	0	2/2

Translation of the JUNCTIONS

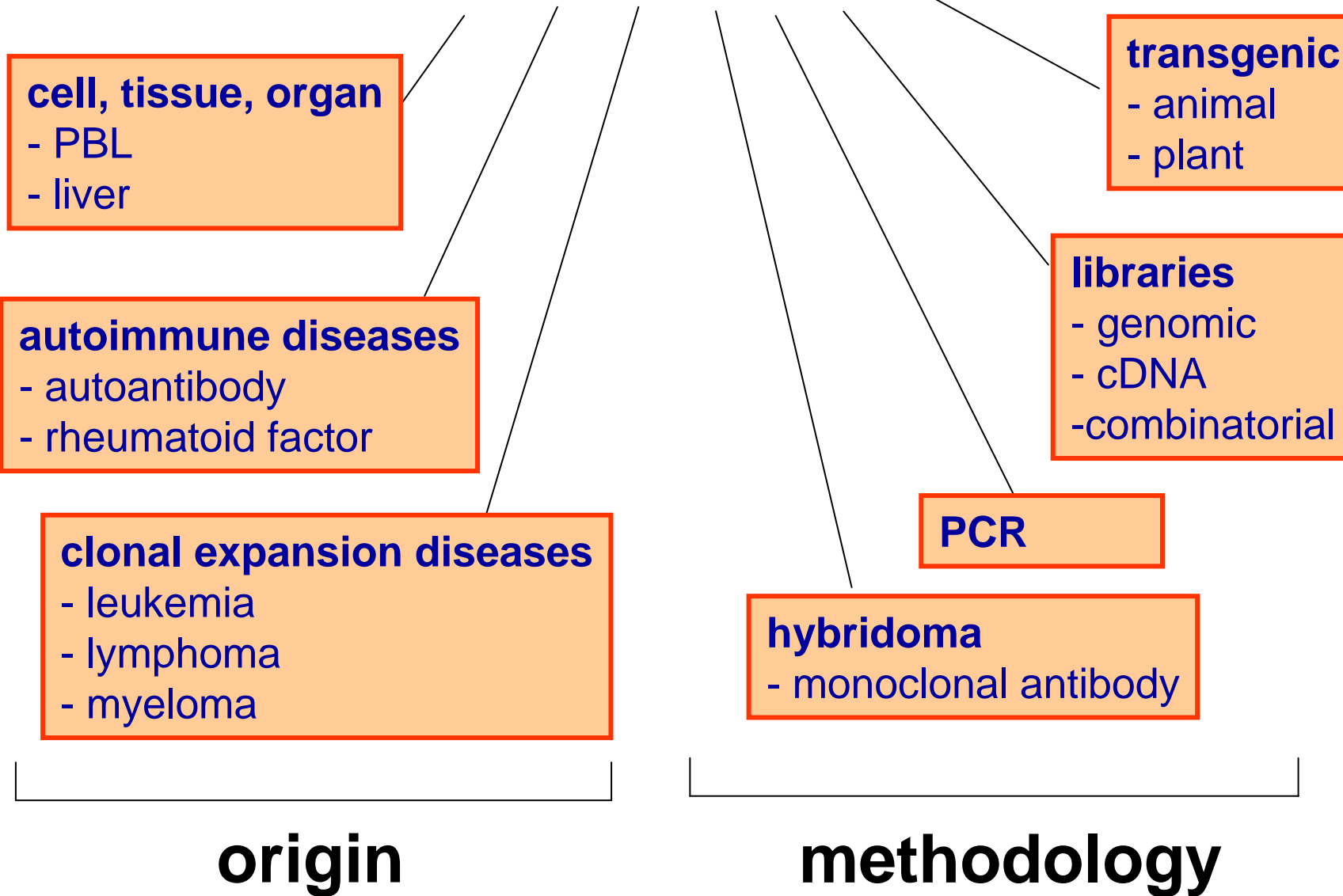
		104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	frame	CDR3-IMGT length
#1	AF490920	tgt	caa	cac	tat	gat	gat					ttc	cca	ttc	act	ttc	+	9
		C	Q	Q	Y	Y	S					T	P	L	T	F		
#2	AF490935	tgt	cag	caa	tat	tat	agt					act	cct	ctc	act	ttc	+	9
		C	Q	Q	Y	Y	S				G	P	P	Y	T	F		
#3	AF490937	tgt	cag	caa	tat	tat	agt				ggt	cct	ccg	tac	act	ttt	+	10
		C	Q	H	Y	N	N	W			P	P	L	Y	T	F		
#4	AF490932	tgt	cag	cac	tat	aat	aac	tgg	cct	ccc	ctg	tac	act	ttt			+	11

IMGT-Choreography: 3D structures/specificities



V-DOMAINS (*Mus musculus* E5.2 Fv)

"OBTENTION" concept



Immunoinformatics



<http://imgt.cines.fr>

Data integration specific to Immunology

- *interactions host-pathogens
- *vaccinology
- *immunomodulation...

Gene

Transcript

Microarrays

Protein

3D

Organelle

Bioinformatics,
databases and tools

Cell

Gene regulation
Pathways
Networks

Population

Organism

Organ

Tissue

Mathematical and
computational models

Collection of
clinical data

Who is using IMGT?

Medical research:

repertoire in autoimmune diseases, AIDS, leukemias, lymphomas, myelomas, translocations, detection of residual diseases

Therapeutic approaches:

immunotherapy, grafts, immunomodulation, immunosuppression

Veterinary research:

I G and TR repertoire of domestic and farm species

Biotechnology related to antibody engineering:

chimeric, humanized, human antibodies, scFv, combinatorial libraries, intrabodies

Genome diversity:

comparative and developmental immunology, evolution of the adaptive immune system

IMGT, the international ImMunoGeneTics information system®

<http://imgt.cines.fr>



The IMGT team at Montpellier, France



<http://imgt.cines.fr>