Ontologies
Langage OWL

IMGT-ONTOLOGY

Application à la formalisation des concepts de description d'IMGT-ONTOLOGY avec l'éditeur Protégé

Véronique Giudicelli
LIGM, IGH, UPR CNRS 1142
141 rue de la Cardonille
34296 Montpellier Cedex 5
tel: 04.34.35.99.28 fax: 04.34.35.99.01
e-mail: Veronique.Giudicelli@igh.cnrs.fr

23 mai 2013

Plan

1- Cours
- **Ontologies**
- **le Web sémantique**
- **Le langage OWL**
- **Exemples d'ontologies en biologie.**

2- Domaine d'application en immunogénétique:
IMGT, The international ImMunoGenetics information system
IMGT-ONTOLOGY

3- Application
- **Formalisation des concepts de description**

En philosophie, l'ontologie est l'étude de l'être en tant qu'être, c'est-à-dire l'étude des propriétés générales de ce qui existe.

Une ontologie est un ensemble structuré de connaissances dans un domaine particulier (comme par exemple celui de l'immunogénétique).

Une ontologie cherche à en représenter le sens (sémantique).

Nécessité de définir un <u>vocabulaire standardisé</u> utilisé dans le domaine de connaissance

Quelle est la signification des termes dans le contexte?

=> Terminologie + définition pour chaque terme.

Comment ces termes sont–ils organisés? Comment sont-il reliés?

=> Propriétés: elles expriment les relations entre les connaissances

Ontologies

Une **ontologie** définit **formellement** les **termes** employés pour **décrire et représenter** un **domaine de connaissance**.

Les ontologies sont destinées à être utilisées par:

- des personnes
- des bases de données
- des applications

ayant besoin de partager des informations

Ontologies

Au sein d'une ontologie, les termes sont regroupés sous forme de concepts (ou classes) sémantiques: ils définissent un groupe d'individus (instances) possédant des propriétés similaires.

Les ontologies regroupent les concepts élémentaires, leur définition (informatiquement exploitable) et leurs relations.

Les ontologies doivent permettre le partage et la réutilisation des connaissances.

Bases de connaissances:

Une ontologie ainsi que l'ensemble des instances individuelles des concepts constituent une base de connaissances. Une frontière subtile marque la fin d'une ontologie et le début d'une base de connaissances.

World Wide Web
(Créé par Tim Berners Lee en 1990)

=> Pour partager des informations/connaissances et les relier

IMGT Index

Here you are: IMGT Web resources > IMGT Index

Ontology (IMGT-ONTOLOGY)

- Introduction
- IMGT-ONTOLOGY axioms and concepts

An ontology is a concise and non ambiguous description of the more significant and relevant concepts in a application domain.
IMGT-ONTOLOGY [1], is the first ontology which allows the management of the immunogenetics knowledge for all vertebrate species.

Citing IMGT-ONTOLOGY:
- Giudicelli, V. and Lefranc, M.-P. "Ontology for Immunogenetics: IMGT-ONTOLOGY", Bioinformatics, 15, 1047-1054 (1999) PMID: 10745995, LIGM:221, PDF
- Lefranc, M.-P. et al. "IMGT-ONTOLOGY for Immunogenetics and Immunoinformatics, http://www.imgt.org", *In Silico* Biology, 2004, 4, 17-29. Epub 2003, 4, 0004, LIGM:278, PMID: 15089751
- Lefranc, M.-P. et al. "IMGT-Choreography for Immunogenetics and Immunoinformatics", *In Silico* Biology, 2005, 5, 45-60, Epub 2005, 5, 0006, 24 Dec 2004, LIGM:294, PMID: 15972004
- Duroux, P. et al., "IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm", Biochimie, 90, 570-583 (2008). Epub 2007 Sep11. PMID: 17949886

Introduction

The molecular synthesis and genetics of the immunoglobulin (IG) and T cell receptor (TR) chains and the polymorphism of the MHC are particularly complex, and therefore one of the first tasks of IMGT-ONTOLOGY comprises a formal specification of the terms to be used in the domain of immunogenetics and bioinformatics [2-8].

IMGT-ONTOLOGY includes a controlled vocabulary and annotation rules which are indispensable to ensure accuracy, consistency and coherence in IMGT® [5]. IMGT-ONTOLOGY allows scientists and clinicians to use identical terms with the same meaning. This provides a semantic repository to be included in more general molecular biology ontologies, and will be therefore of a great help to increase interoperability between specialist and generalist databases.

IMGT-ONTOLOGY axioms and concepts

Seven IMGT-ONTOLOGY axioms have been defined [1,6-8]: '**IDENTIFICATION**', '**DESCRIPTION**', '**CLASSIFICATION**', '**NUMEROTATION**', '**LOCALIZATION**', '**ORIENTATION**', and '**OBTENTION**'. They constitute the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope [9,10].

The IMGT-ONTOLOGY concepts of identification, description, classification, numerotation, localization, orientation and obtention were generated from these axioms and described in ref. [1,2,6-8].

Concepts of interaction that are necessary to define interactions between entities were defined based on these concepts.

World Wide Web

- **Les ressources sous forme de page HTML (Langage HTML)**
- **Les relations entre ses pages qui se traduisent par des liens hypertextes**

World Wide Web

Limites

• **Balises HTML: la connaissance est encapsulée dans une couche de présentation**

• **Les liens hypertexte ne permettent pas de qualifier les relations entre les informations**

• **La qualité des pages est très hétérogène**

$\Rightarrow$ **Ces informations ne sont pas exploitables automatiquement**

World Wide Web Consortium (W3C, http://www.w3.org)
(créé en 1994)

- Développement de technologies (spécifications, guides, logiciels et outils)
- Proposition des standards, pour la croissance et l'exploitation du Web

En particulier, en 1998, le langage de balisage extensible XML (Extensible Markup Language) devient une recommandation du W3C

- XML devrait pouvoir être utilisé sans difficulté sur Internet
- Les documents XML devraient être lisibles par l'homme et raisonnablement clairs
- Il devrait être facile de créer des documents XML
- Il devrait être facile d'écrire des programmes traitant les documents XML

Séquence EMBL AM947564

Fichier   Édition   Affichage   Historique   Marque-pages   Outils   ?

HighVquest admin page | http://www.ebi.ac.u...47564&display=text | +

http://www.ebi.ac.uk/ena/data/view/AM947564&display=text | ☆ ▾ C | 🔍 ▾ ebi embl

Les plus visités | Débuter avec Firefox | À la une

```
ID   AM947564; SV 1; linear; genomic DNA; STD; HUM; 287 BP.
XX
AC   AM947564;
XX
DT   01-APR-2008 (Rel. 95, Created)
DT   01-APR-2008 (Rel. 95, Last updated, Version 1)
XX
DE   Homo sapiens partial IGHV gene for immunoglobulin heavy chain variable
DE   region, patient 4
XX
KW   .
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae;
OC   Homo.
XX
RN   [1]
RP   1-287
RA   Capello D.;
RT   ;
RL   Submitted (13-MAR-2008) to the EMBL/GenBank/DDBJ databases.
RL   Capello D., Clinical & Experimental Medicine, University of Eastern
RL   Piedmont, Via Solaroli, 17 Novara, 28100, ITALY.
XX
RN   [2]
RA   Capello D., Martini M., Gloghini A., Cerri M., Rasi S., Deambrogi C.,
RA   Rossi D., Spina M., Tirelli U., Larocca L., Carbone A., Gaidano G.;
RT   "Molecular analysis of immunoglobulin variable genes in HIV-related non
RT   Hodgkin lymphoma reveals implications for disease pathogenesis and
RT   histogenesis";
RL   Unpublished.
XX
DR   IMGT/LIGM; AM947564.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..287
FT                   /organism="Homo sapiens"
FT                   /isolate="patient 4"
FT                   /mol_type="genomic DNA"
FT                   /cell_type="B lymphocyte-derived lymphoma"
```

zotero

Séquence AM947564 en format XML

RDF : Resource Description Framework

**Créé en 1999 pour décrire les métadonnées
dans l'objectif de traiter l'information automatiquement,
de favoriser l'interopérabilité des connaissances et de structurer
les informations.**

**Métadonnées: information permettant d'en décrire une autre
• Les métadonnées sont habituellement comprises comme des données à propos des données.
• Un catalogue de bibliothèque contient de l'information (métadonnées) à propos de publications (données)
• Un système de fichier informatique définit des droits de lecture, écriture, etc. (métadonnées) à propos des fichiers (données).**

RDF

- **Standard du W3C**
- **Basé sur XML**
- **Un début vers un Web structuré**
- **RDF emploie les URI comme schéma de nommage (pour éviter la confusion entre termes identiques mais différents sémantiquement selon le contexte)**
- **On peut tout décrire**
- **Grande souplesse quand à l'extensibilité**

URI: Uniform Resource Identifier

RDF

RDF se base sur une description des connaissances à l'aide de phrases simples :

C'est un moyen d'exprimer des relations.

Ces relations sont décrites sous forme de graphe.

Chaque nœud du graphe est une ressource ou une valeur.

Et chaque nœud est relié à un autre par un arc "nommé"

Sujet
(source, ressource)

Prédicat
(propriété)

Objet
(cible, valeur)

urn:users:456

http://xulfr.org/users#name

laurent

nœuds du graphe

ressource — prédicat → literal

prédicat → Autre ressource

Shémas issus de http://xulfr.org/wiki/FormatRdf/Introduction
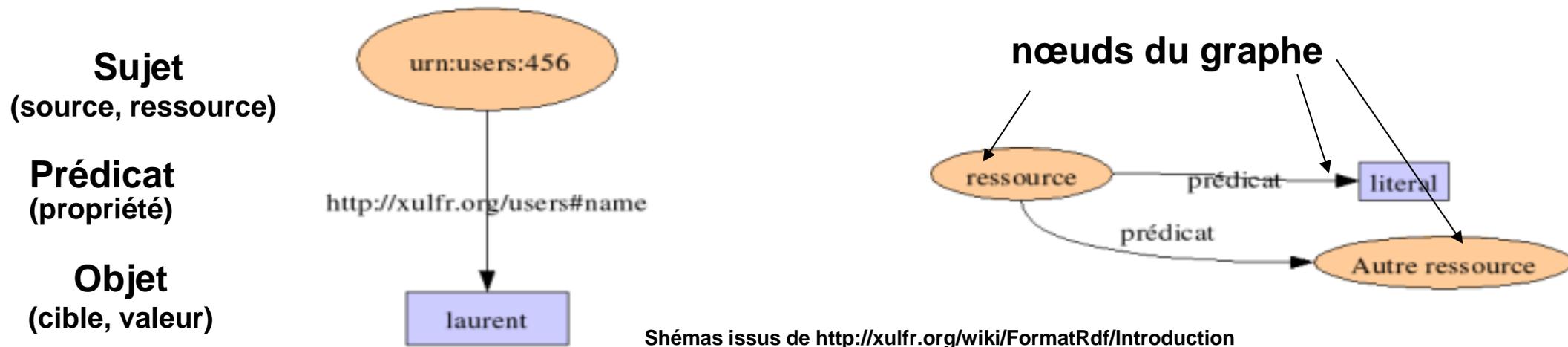
**=> Ceci correspond à un lien qui comporte un sens :
c'est du Web sémantique**

Les objectifs du Web sémantique sont de partager les connaissances et de pouvoir les manipuler automatiquement.

Pour ceci, il faut :
- **qualifier le savoir (à l'aide de métadonnées),**
- **le formaliser en utilisant une syntaxe extensible (par exemple avec XML) et**
- **le structurer pour éviter les duplications.**

ONTOLOGIE

Sur le plan informatique, les ressources sont définies les unes par rapport aux autres selon un graphe. Cette structure permet une automatisation de la manipulation des données.

Différence entre une thésaurus et une ontologie :

- Un thésaurus relie des concepts entre eux selon des relations précises : synonyme, homonyme, hiérarchie, terme associé.
- Une ontologie ajoute des règles et des outils de comparaison sur et entre les termes, groupes de termes et relations : équivalence, symétrie, contraire, cardinalité, transitivité.

=> Dans cet objectif RDF est trop limité en particulier, il ne permet pas de donner la nature des relations (transitive inverse, ..), possède une logique trop limitée pour faire du raisonnement

OWL (Web Ontology Language)

En 2004, OWL devient une recommandation du W3C

OWL découle de RDF et RDFS, possède des connecteurs logiques, d'exprimer des cardinalités sur les propriétés et d'en spécifier la nature.

Une ontologie formalisée en OWL comprend
- Un espace de nom, :
- L'entête <owl :Ontology> pour décrire l'ontologie
- La définition des classes, des propriétés et des instances

Classes, sous-classes, héritage et spécialisation

Exemple des les formes géométriques: représentation hiérarchique de concepts avec la relation « is_a »

(OWL, Web Ontology Language)

Notion de Classe :

   Définit un groupe d'individus possédant des propriétés similaires.
   Thing est la classe mère.

spécialisation

owl:Thing

Is_a

Quadrilatere

Def : <u>polygone</u> à 4 côtés

Is_a

Parallelogramme

Def: <u>quadrilatère</u> (convexe) dont les côtés sont <u>parallèles</u> deux à deux

Is_a

Rectangle

Def: Parallélogramme à angles droits

Is_a

Carre

Def: rectangle dont les cotés sont égaux
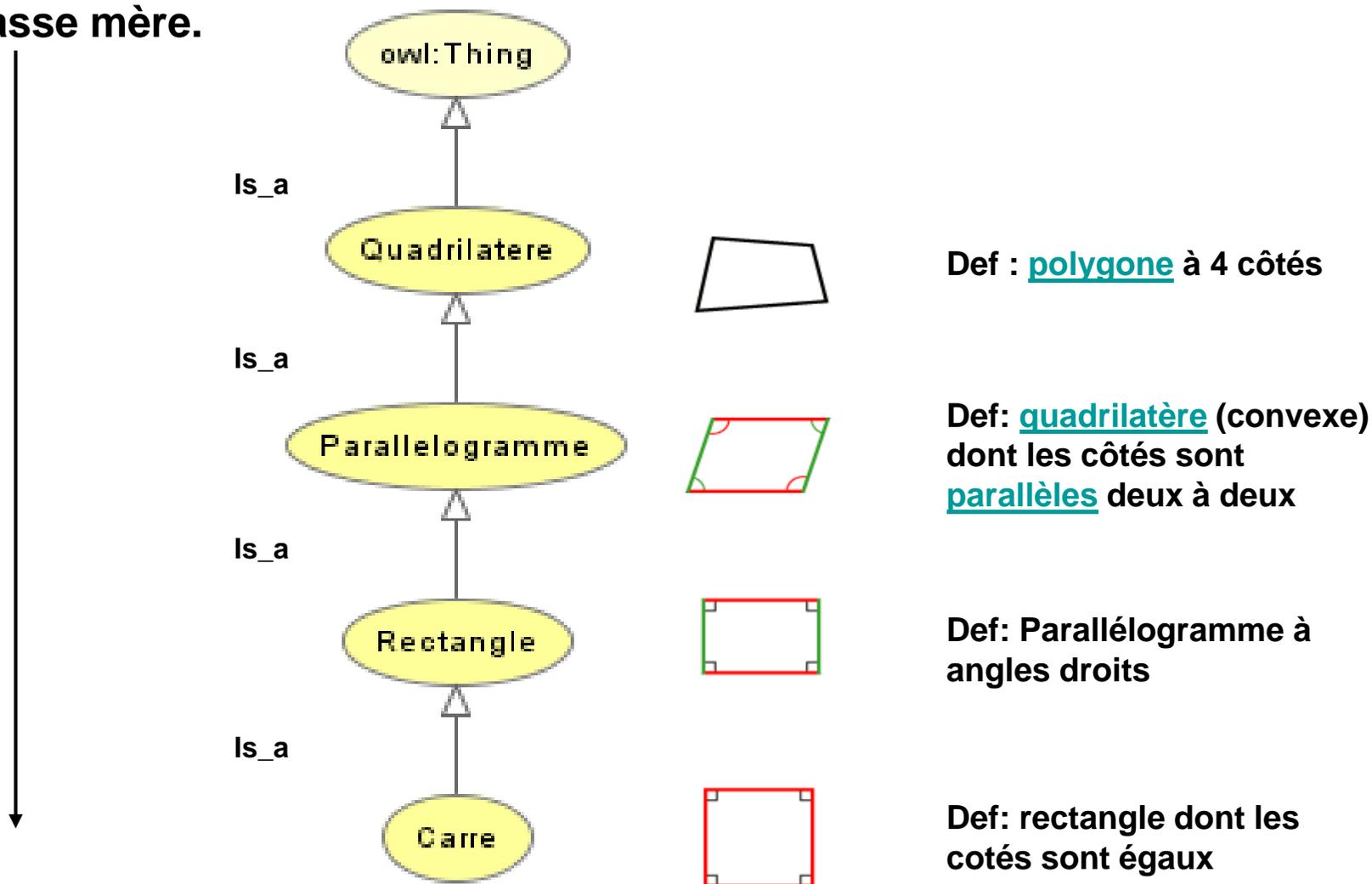
# Taxonomy : représentation hiérarchique de concepts avec la relation « is_a »

Instances

Ce sont les « individus » qui peuplent les classes
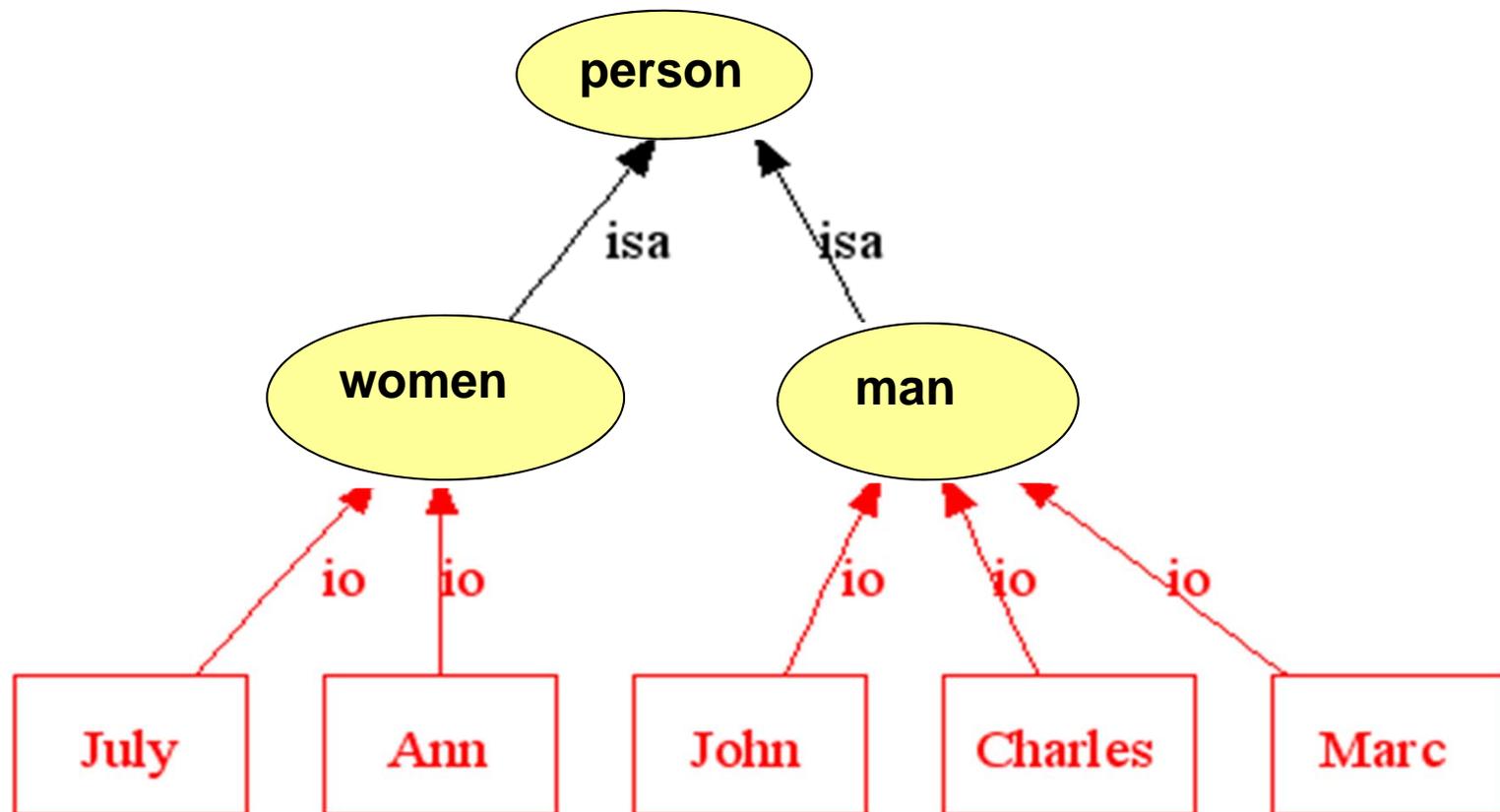
Propriétés/Relations
(OWL, Web Ontology Language)
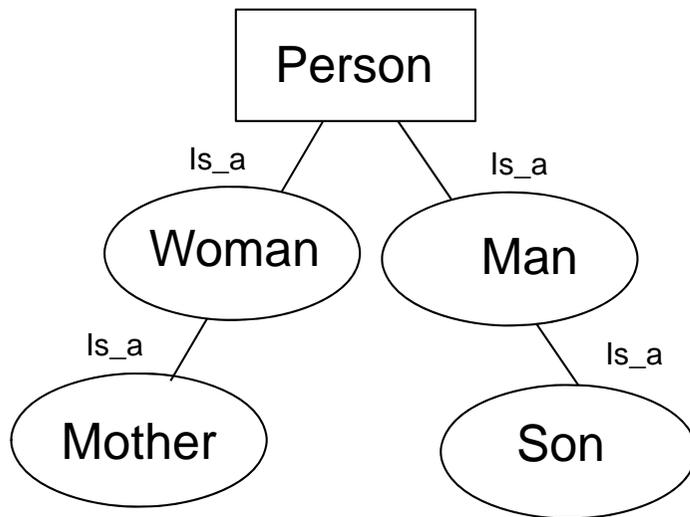
Existe-t-il des relations particulières
Autres que « is_a » :

- fondamentales sur le plan sémantique
- caractéristiques des concepts/classes
- qui lient les concepts/classes entre eux?

Propriétés/Relations
(OWL, Web Ontology Language)

-Propriété : de type de données:

    classe =>ensemble de valeurs (numériques, alphabétique,…)

      - date de naissance

      - numéro de sécurité sociale

      …..

- Propriété d'objet : classe => classe

Mother

mother_of*son_of*

Son

brother_of*
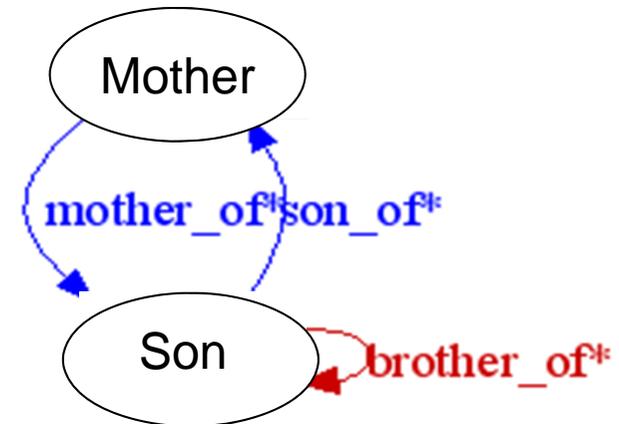
- Propriétés d'annotations: non héritées

Définir et caractériser les Propriétés/Relations (OWL, Web Ontology Language)

domain : classes pour lesquelles est définie la propriété,
range : classes reliées par la propriété au domain.

domain	relation	range
Son	brother_of	Son
Mother	mother_of	Son
Son	son_of	Mother

Caractéristiques des Propriétés/relations: symétriques, transitives, inverses et fonctionnelles
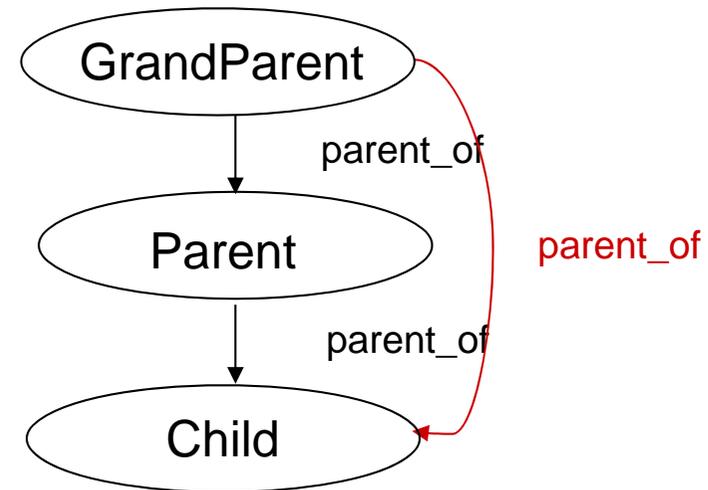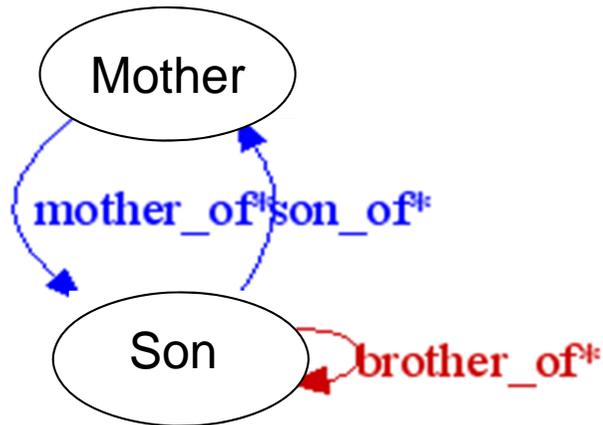
(familles de 3 générations avec de jeunes enfants)

Symétrique: brother_of
Transitive: parent_of
Inverse: son_of / mother_of
Fonctionnelle : son_of

Restriction sur les Propriétés/Relations : cardinalités

Mother

mother_of

child_of

Child
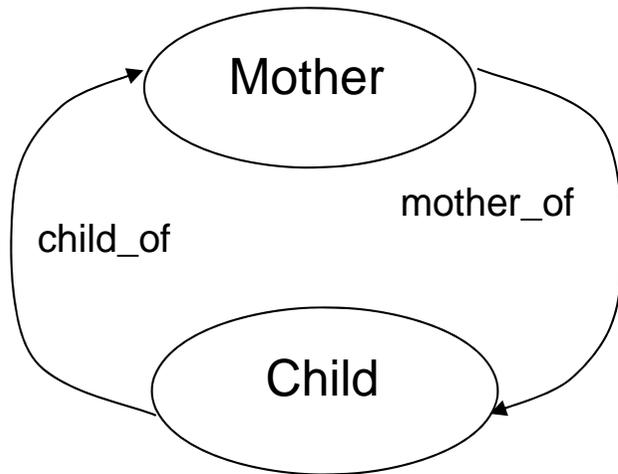
Ex de familles nombreuses de moins de 10 enfants

- Une mère a au moins 3 enfants
- Une mère a au plus 10 enfants
- Un enfant a une et une seule mère

minCardinality: toute instance de la classe sera relié par la propriété à au moins x individus Exemple: propriété mother_of de la classe Mother : minCardinality=3

maxCardinality: toute instance de la classe sera relié par la propriété à au plus x individus Exemple: propriété mother_of de la classe Mother : maxCardinality=10

cardinality: toute instance de la classe sera relié par la propriété à exactement x individus Exemple: propriété son_of de la classe Mother : cardinality=1

Les ontologies en Biologie

- On s'intéresse aux ontologies qui sont du domaine public.
- Leur nombre augmente régulièrement
  (besoin de définir, de partager).
- Elles couvrent des sujets et domaines différents.

NCBO Bioportal (http://bioportal.bioontology.org/)

OBO (Open Biological Ontologies) (http://www.obofoundry.org/)

recensent les ontologies en biologie.

NCBO Bioportal (http://bioportal.bioontology.org/)

NCBO Bioportal

BioPortal   Browse   Search   Mappings   Recommender   Annotator   Resource Index   Projects   Recently Viewed   Sign In   Help   Feedback

NCBI organismal classification   Terms ▾

Jump To: [   ]

| Details | Visualization | Notes (0) | Term Mappings (9) | Term Resources |

- root
  - cellular organisms
    - Archaea
    - Bacteria
    - **Eukaryota**
  - other sequences
  - unclassified sequences
  - Viroids
  - Viruses
- taxonomic_rank
  - class
  - family
  - forma
  - genus
  - infraclass
  - infraorder
  - kingdom
  - order
  - parvorder
  - phylum
  - species
  - species group
  - species subgroup
  - subclass
  - subfamily
  - subgenus
  - subkingdom
  - suborder
  - subphylum
  - subspecies
  - subtribe
  - superclass
  - superfamily
  - superkingdom
  - superorder

Preferred Name	Eukaryota
Synonyms	"eukaryotes"
	"eucaryotes"
ID	NCBITaxon:2759
Full Id	http://purl.org/obo/owl/NCBITaxon#NCBITaxon_2759
Database References	GC_ID
Exact Synonym	"eukaryotes"
	"eucaryotes"
Related Synonym	"eukaryotes"
	"Eucarya"
	"Eukarya"
	"Eucaryotae"
	"Eukaryotae"
Is A	cellular organisms

OBO (http://www.obofoundry.org/)

OBO (http://www.obofoundry.org/)

Mosquito insecticide resistance	environment	MIRO	mosquito_insecticide_resistance.obo	2010/08/29
Mouse adult gross anatomy	anatomy	MA	adult_mouse_anatomy.obo	2010/04/04
Mouse gross anatomy and development	anatomy	EMAP	EMAP.obo	2007/06/17
Mouse pathology	health	MPATH	mouse_pathology.obo	2010/01/14
NCBI organismal classification	taxonomy	NCBITaxon	taxonomy.dat	
NCI Thesaurus	health	ncithesaurus	EVS/	
NIF Cell	neuroscience	NIF_Cell	NIF-Cell.owl	
NIF Dysfunction	neuroscience	NIF_Dysfunction	NIF-Dysfunction.owl	
NIF Gross Anatomy	neuroscience	NIF_GrossAnatomy	NIF-GrossAnatomy.owl	
NMR-instrument specific component of metabolomics investigations	experiments	NMR	NMR.owl	
OBO relationship types	all	OBO_REL	ro.obo	2010/06/07
Ontology for biomedical investigations	experiments	OBI	obi.owl	
Ontology for General Medical Science	medicine	OGMS	ogms.owl	
Pathogen transmission	health	TRANS	transmission_process.obo	2009/05/28
Pathway ontology	biological process	PW	pathway.obo	
Physico-chemical methods and properties		FIX	fix.obo	
Physico-chemical process		REX	rex.obo	2007/03/08
Plant environmental conditions	environment	EO	environment_ontology.obo	2009/01/09
Plant growth and developmental stage	anatomy	PO	po_temporal.obo?view=co	
Plant structure	anatomy	PO	po_anatomy.obo?view=co	
Platynereis stage ontology	anatomy	PD_ST	Pdu_Stages.obo	
Protein covalent bond	proteins	RESID	RESIDUES.XML	
Protein Domains	proteins	IPR	InterPro FTP directory	
Protein modification	proteins	MOD	PSI-MOD.obo	
Protein-protein interaction	experiments	MI	psi-mi25.obo	
Proteomics data and process provenance	proteins	ProPreO	ProPreO-060506.owl	
RNA ontology	molecular structure	RNAO	rnao.obo	
Sample processing and separation techniques	provenance	SEP	sep.obo	
Sequence types and features	biological sequence	SO	so.obo	2010/08/27
Software ontology		SWO	softwareontology.owl	
Spatial Ontology	anatomy	BSPO	spatial.obo	2010/05/18

Gene Ontology.

- GO a été créée en 1998. GO résulte d'une collaboration entre plusieurs bases de données: FlyBase (drosophile), the Saccharomyces Genome Database, et des bases de données de génomes (homme et souris), etc.

- GO comprend 3 parties axées sur :

    – **la fonction moléculaire** (fonction des gènes exprimés ex: ATPase activity.

    – **le processus biologique** (rôles biologique généraux de fonctions moléculaire complexes ex: la mitose).

    – **les composants cellulaires** (structures subcellulaires, localisation des complexes macromoleculaires ex: le noyau, le télomère).

Gene Ontology (http://www.geneontology.org/)

the Gene Ontology

AmiGO

Search   Browse   BLAST   Homolog Annotations   Tools & Resources   Help

Search GO mitosis   ● terms ○ genes or proteins ☐ exact match Envoyer

Term Search Results

57 results for **mitosis** in terms fields **term accession, term name and synonyms**

▼ **Filter search results** ❓

Ontology   Set filters
All
biological process   Remove all filters
cellular component
molecular function

Results are sorted by **relevance**. To change the sort order, click on the column headers.

1 2 View all results

Select all   Clear all   Perform an action with this page's selected terms... ▾   Go!

rel ↓		Accession , Term	Ontology
☐	GO:0007067 : **mitosis** [show def]	3835 gene products view in tree	biological process
☐	GO:0051337 : a**mitosis** [show def]	0 gene products view in tree	biological process
☐	GO:0010458 : exit from **mitosis** [show def]	338 gene products view in tree	biological process
☐	GO:0007088 : regulation of **mitosis** [show def]	1075 gene products view in tree	biological process
☐	GO:0055047 : generative cell **mitosis**	2 gene products biological process

A AmiGO: Term Details for GO:0007067   +

amigo.geneontology.org/cgi-bin/amigo/term_details?term=GO:0007067&session_id=3811amigo1369227609

8 ▾ ncbo bioportal

Les plus visités   Débuter avec Firefox   À la une

Term Information

Accession	GO:0007067
Ontology	**Biological Process**
Synonyms	None
Definition	A cell cycle process comprising the steps by which the nucleus of a eukaryotic cell divides; the process involves condensation of chromosomal DNA into a highly compacted form. Canonically, mitosis produces two daughter nuclei whose chromosome complement is identical to that of the mother cell. *Source:* GOC:dph, GOC:ma, GOC:mah, ISBN:0198547684
Comment	None
Subset	Generic GO slim
Community	Add usage comments for this term on the GONUTS wiki.

Back to top

Term Neighborhood for mitosis (GO:0007067)

Filter lineage gene product counts ?

Data source	Species
No filter	No filter
ASAP	A. fumigatus
AspGD	A. niger
CGD	A. oryzae

Ancestors and Children	Inferred Tree View	Graph View	Other Views	Downloads	Mappings

Ancestors of mitosis (GO:0007067)

subject ⇕	relation ⇕	object ⇕	annotations ⇕
mitosis	Ⓟ part_of (inferred)	biological_process (GO:0008150)	501703
mitosis	Ⓘ is_a (inferred)	cellular component organization or biogenesis (GO:0071840)	61308
mitosis	Ⓘ is_a (inferred)	cellular component organization (GO:0016043)	54002
mitosis	Ⓟ part_of (inferred)	cellular process (GO:0009987)	287339
mitosis	Ⓟ part_of (inferred)	single-organism process (GO:0044699)	172947
mitosis	Ⓘ is_a (inferred)	organelle organization (GO:0006996)	29739
mitosis	Ⓟ part_of (inferred)	single-organism cellular process (GO:0044763)	142265

zotero

biological process

cellular component organization or biogenesis

cellular process

single-organism process

cellular component organization

single-organism cellular process

organelle organization

cell cycle

organelle fission

nuclear division

cell cycle process

mitosis

Is a   A → B

Part of   A → B

Has part   A ⇠ B

Regulates   A → B

Positively regulates   A → B

Negatively regulates   A → B

Occurs in   A → B

goslim_metagenomics

goslim_candida

goslim_plant

goslim_pir

goslim_aspergillus

goslim_generic

goslim_yeast

goslim_pombe

QuickGO - http://www.ebi.ac.uk/QuickGO

Sequence Ontology

(http://www.sequenceontology.org/)

- **Vocabulaire contrôlé pour l'annotation des séquences et l'annotation des génomes**

- **proposer une représentation structurée des annotations**

- **vocabulaire pour la description des mutations**

Sequence Ontology:
une ontologie pour décrire les séquences biologiques

Fichier   Édition   Affichage   Historique   Marque-pages   Outils   ?

The MISO Sequence Ontology Browser   +

http://www.sequenceontology.org/browser/obob.cgi

sequence ontology

MISO The Sequence Ontology Browser

Home   **Browser**   **Wiki**   **GFF3**   **Resources**   **About**   **Request A Term**   **Site Map**

Search for Terms: immmunoglobulin   Search   Select Release: CURRENT RELEASE

⊞ sequence_attribute
⊞ sequence_collection
⊞ sequence_feature
⊞ sequence_variant

⊞ Obsolete Terms

⊞ Relationship

Welcome to MISO, the Sequence Ontology Browser

▷ Search for a SO term by entering a SO term name or synonym in the query box above;

▷ Explore the structure of SO and browse for SO terms using the expandable, cascading tree on the left;

▷ Go to the detail page for a term where you can:

▷ Get details about a term, it's definition and relationships;

▷ See graphical views of a term's place in the ontology and link to it's neighbors;

▷ Export details about a term in a variety of formats;

▷ And access and contribute detailed documentation about a term and it's history by linking through to the SO wiki

Le système d'information
IMGT, the international ImMunoGeneTics information system®
http://www.imgt.org

Immunoglobulin (IG)   T cell receptor (TR)

[publication info top right]

V-DOMAIN

V-J-REGION

V-DOMAIN

V-D-J-REGION

Light chain

Heavy chain

NH2

COOH

S-S
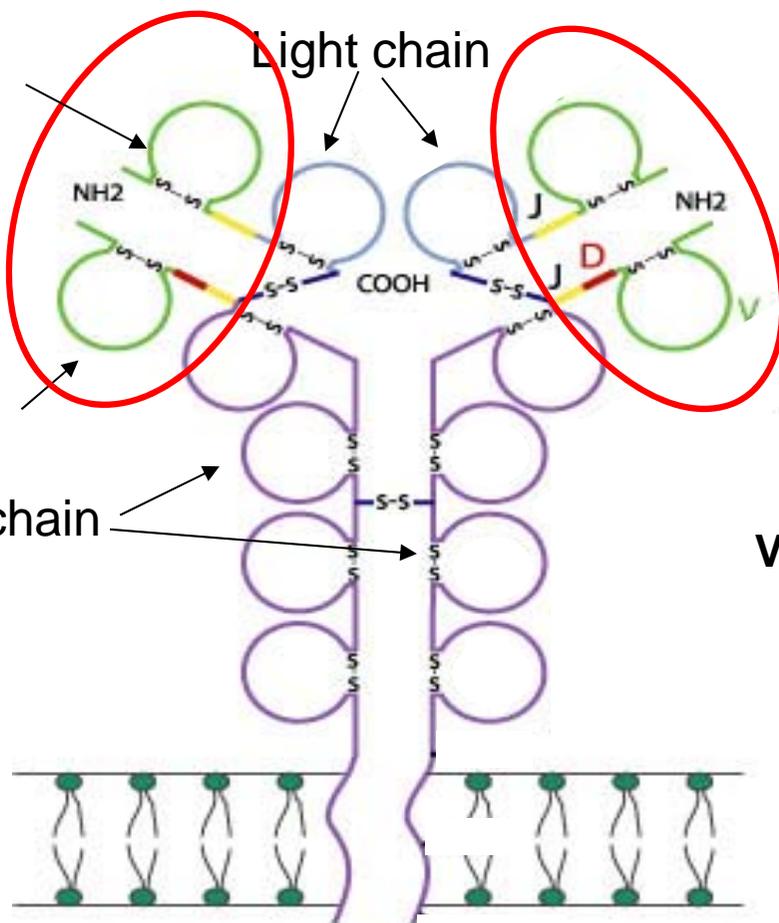
Membrane IgM

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

| Alpha | - | Beta |
| Gamma | - | Delta |

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

T cell receptor

The Immunoglobulin FactsBook, 2001
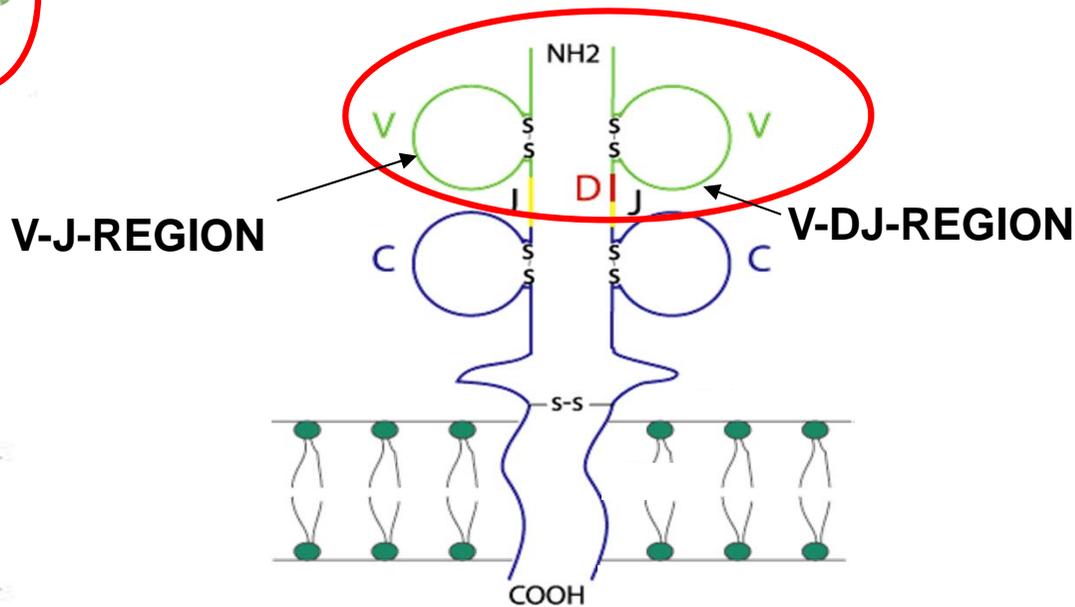
Part of Human *(Homo sapiens)* IGH locus on chromosome 14 (14q32.33)
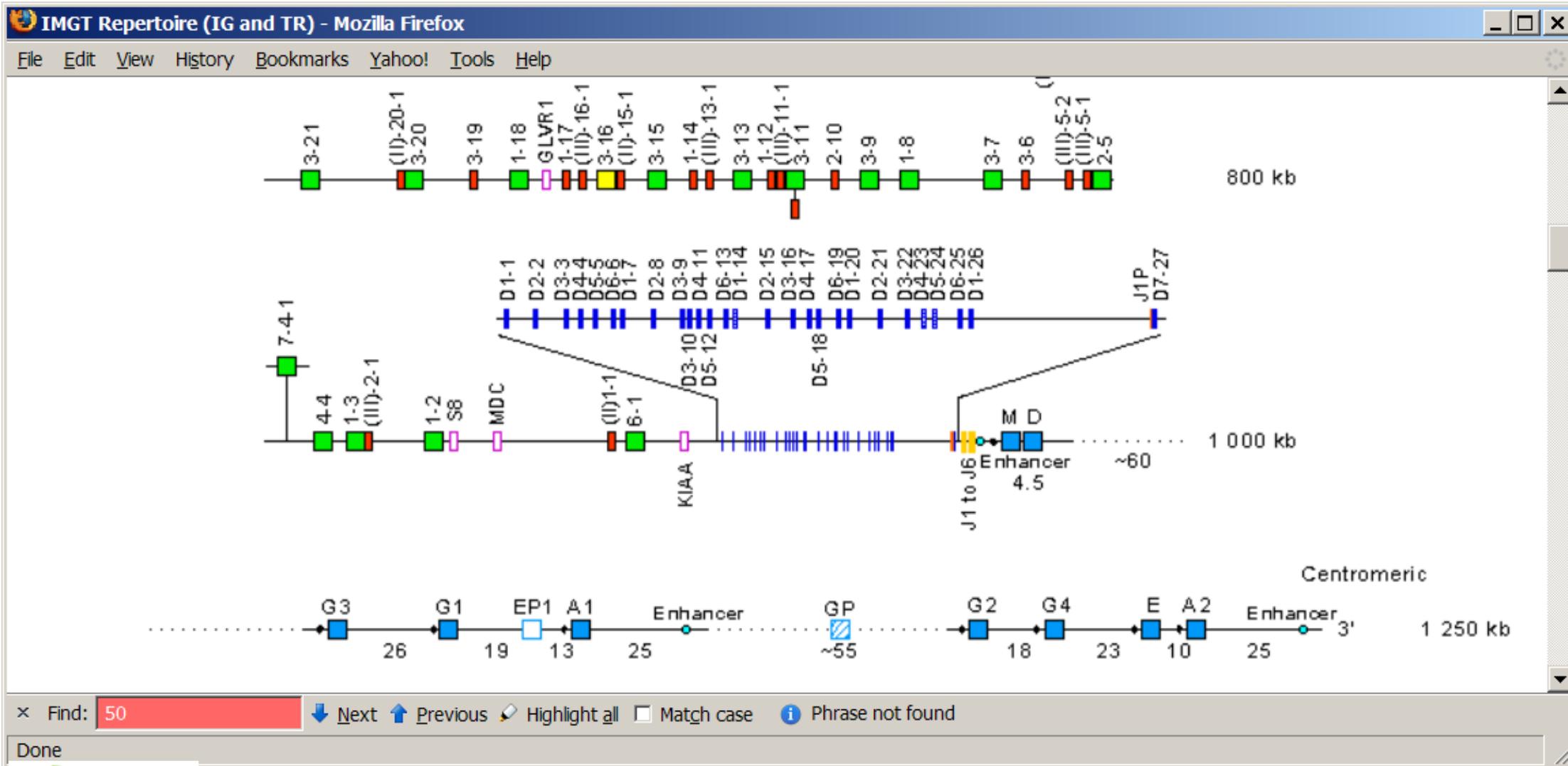
Synthèse des immunoglobulines
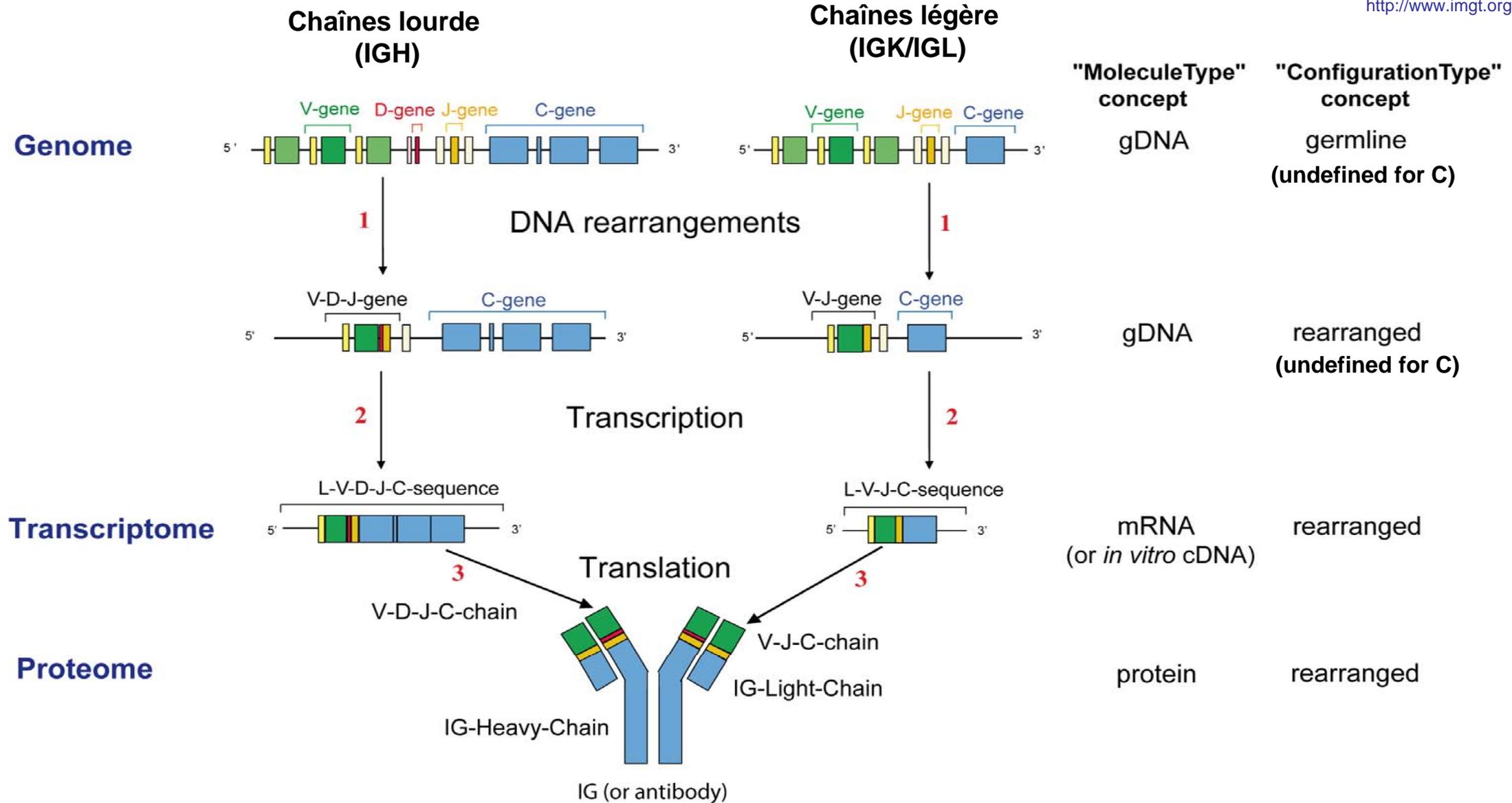
Particularités des séquences IG et TR

- Elles sont nombreuses (2.10^{12} IG par individu)

- Une forte similitude

- De petites différences très significatives

- Nécessité de standardiser la nomenclature des gènes

IMGT-ONTOLOGY

Pour standardiser, partager, réutiliser et représenter les connaissances en immunogénétique

IMGT-ONTOLOGY repose sur 7 axiomes:

IDENTIFICATION

OBTENTION

CLASSIFICATION

DESCRIPTION

ORIENTATION

LOCALIZATION

NUMEROTATION

http://www.imgt.org

IMGT-ONTOLOGY

http://www.imgt.org
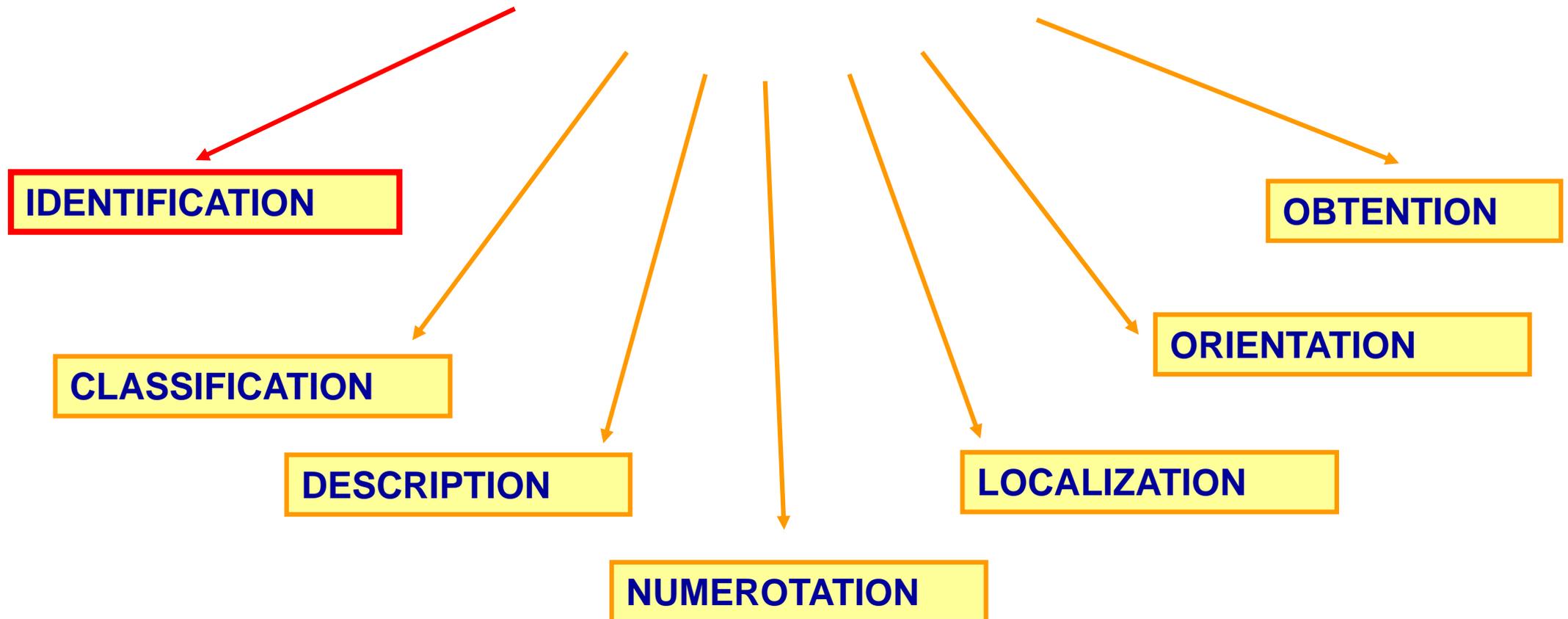
IMGT Scientific chart :

IMGT-ONTOLOGY en *langage naturel* pour les biologistes

http://imgt.cines.fr/textes/IMGTScientificChart/

En cours : édition en langage OWL
Version v1.0.0 sur le site BioPortal

(http://bioportal.bioontology.org/visualize/42685)

Relationships between Molecule_EntityType and the other IDENTIFICATION concepts
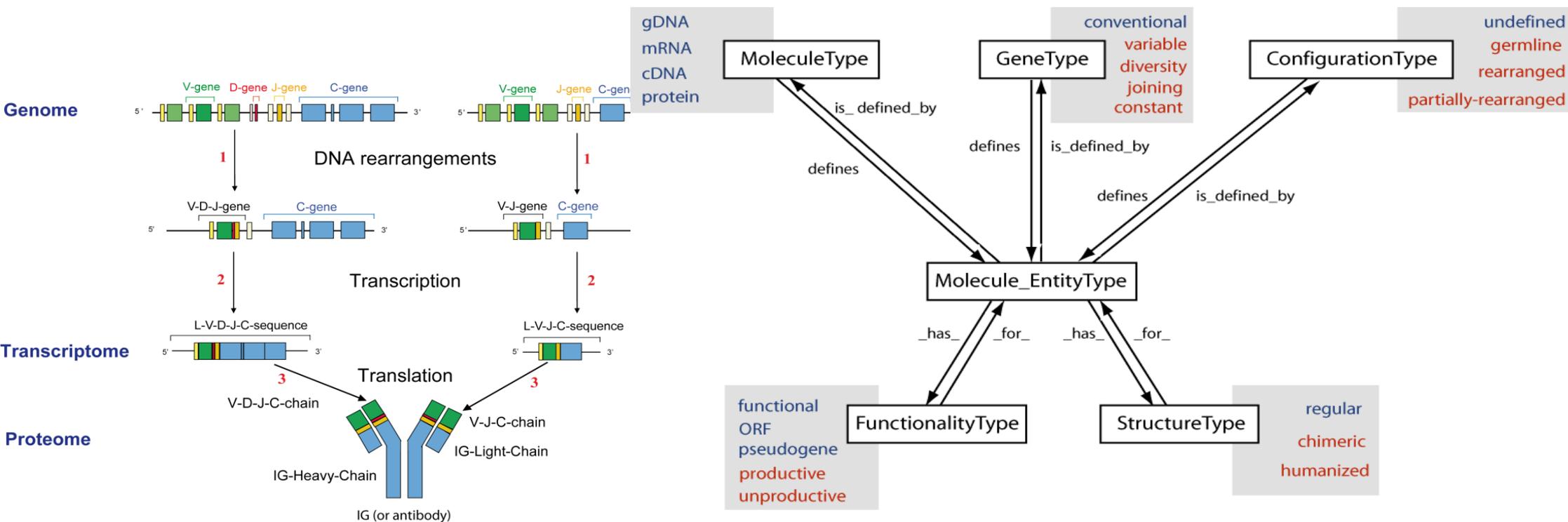
Relationships between the L-V-D-J-C-sequence concepts and the other IDENTIFICATION concepts
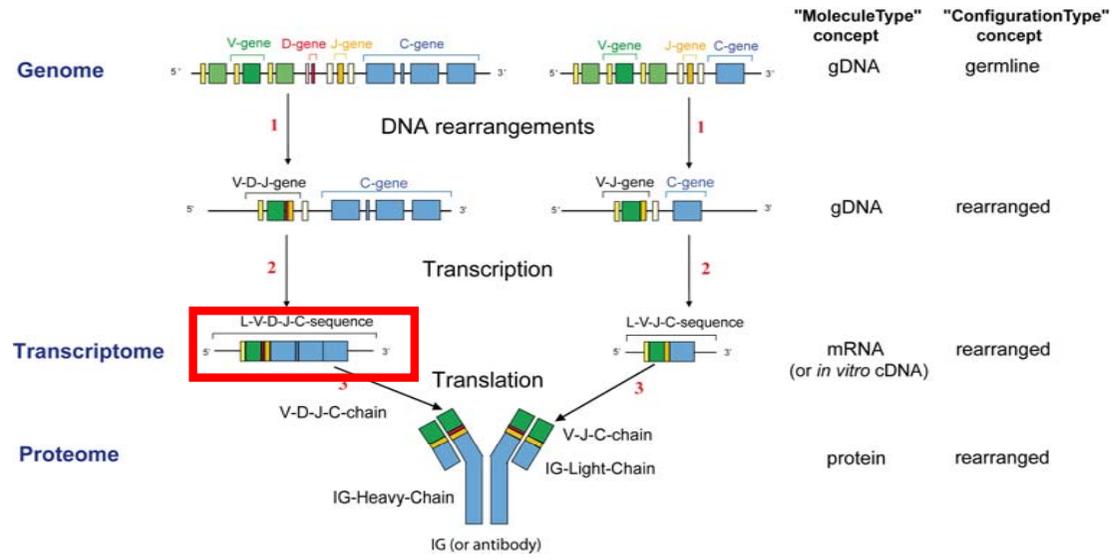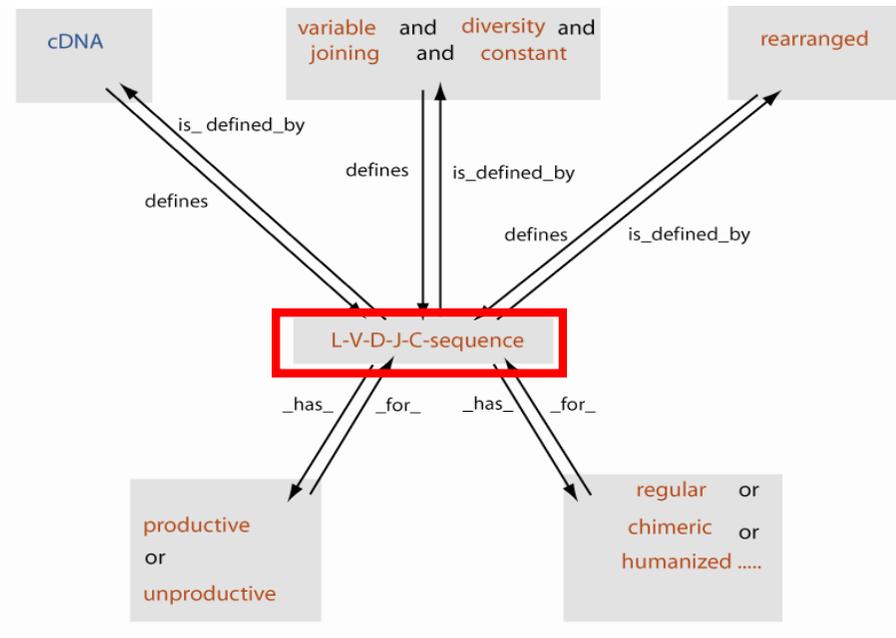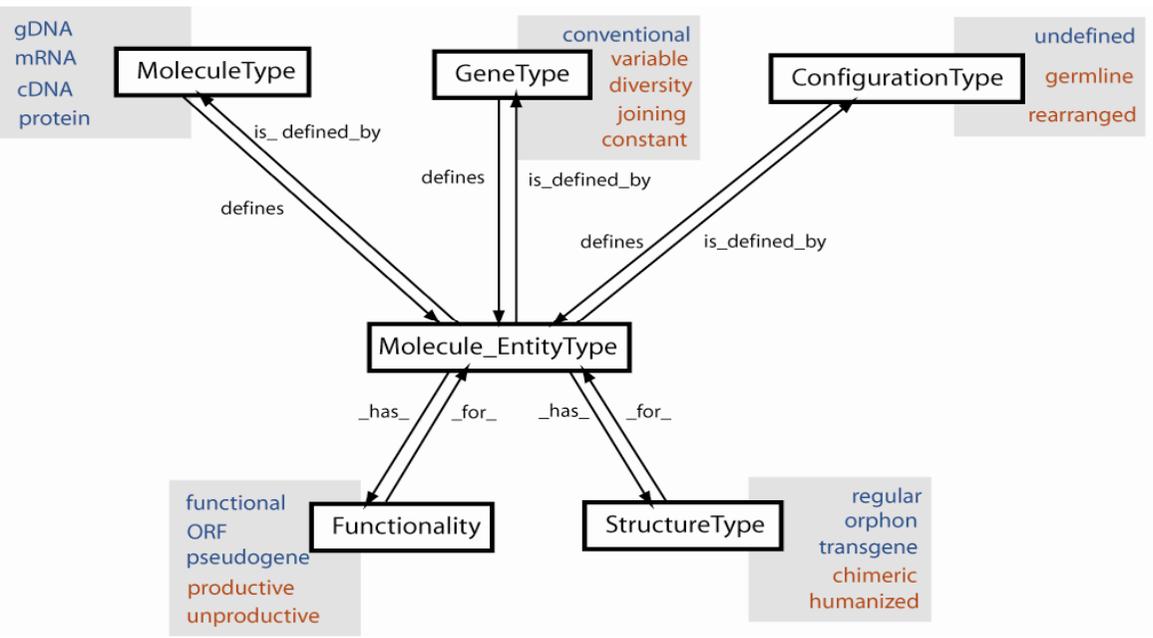
WELCOME !
to the IMGT Home page

**THE INTERNATIONAL
IMMUNOGENETICS
INFORMATION SYSTEM®**

Im
Muno
Gene
Tics   Information
                    system®

http://www.imgt.org
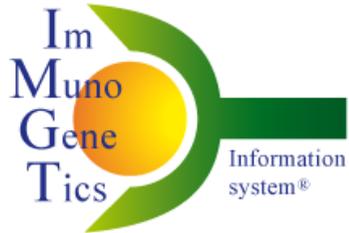
IMGT®, the international ImMunoGeneTics information system® http://www.imgt.org, is the global reference in immunogenetics and immunoinformatics, created in 1989 by Marie-Paule Lefranc (Université Montpellier 2 and CNRS). IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (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 335 species (**170 685 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 (**3 077 genes, 4 669 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 (**2 802 entries**)
  Source: PDB, INN, Kabat

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

Les objectifs de IMGT/LIGM-DB

- Contenir toutes les séquences IG et TR des bases EMBL/GenBank/DDBJ
  + de 170.000 séquences de 335 espèces

- Expertiser les séquences en fonction des connaissances en immunogénétique
  - identification des gènes,
    - description des motifs (régions codantes, …)

- Suivi et intégration des nouvelles connaissances dans le système d'information IMGT®

A l'arrivée dans IMGT®

EMBL FLAT-FILE

```
ID    AABR03051069; SV 1; linear; genomic DNA; WGS; ROD; 16176 BP.
XX
AC    AABR03051069; AABR03000000;
XX
DT    13-AUG-2003 (Rel. 76, Created)
DT    13-AUG-2003 (Rel. 76, Last updated, Version 1)
XX
DE    Rattus norvegicus chromosome 6 clone CH230-392J6; CH230-207C13;
DE    CH230-361I3; CH230-68G3; CH230-83C24; CH230-408M5; CH230-30N12;
DE    CH230-110O20; CH230-321B4; CH230-83D21; CH230-232G6; CH230-170A13;
DE    CH230-375C4; CH230-331M14; CH230-186O12; CH230-49E8; CH230-62J1;
DE    CH230-447G6 strain BN/SsNHsdMCW RNOR03303698, whole genome shotgun
DE    sequence.
XX
KW    WGS.
XX
OS    Rattus norvegicus (Norway rat)
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC    Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea;
OC    Muridae; Murinae; Rattus.
```

Après expertise dans IMGT®

IMGT/LIGM-DB - Mozilla Firefox

File   Edit   View   History   Bookmarks   Tools   Help

IMGT FLAT-FILE

```
ID    AABR03051069 IMGT/LIGM annotation : by annotators; genomic DNA; ROD; 16176 BP.
XX
AC    AABR03051069; AABR03000000;
XX
DT    17-OCT-2006 (Rel. 200643-2, arrived in LIGM-DB )
DT    17-JUL-2007 (Rel. 200729-2, Last updated, Version 3)
XX
DE    Rattus norvegicus chromosome 6 clone CH230-392J6; CH230-207C13;
DE    CH230-361I3; CH230-68G3; CH230-83C24; CH230-408M5; CH230-30N12;
DE    CH230-110O20; CH230-321B4; CH230-83D21; CH230-232G6; CH230-170A13;
DE    CH230-375C4; CH230-331M14; CH230-186O12; CH230-49E8; CH230-62J1;
DE    CH230-447G6 strain BN/SsNHsdMCW RNOR03303698, whole genome shotgun
DE    sequence. ;
DE    genomic DNA; germline configuration; Ig-Heavy; regular; group IGHV.
XX
KW    antigen receptor; immunoglobulin superfamily; IG; IG-Heavy; variable;
KW    pseudogene; IMGT reference sequence; immunoglobulin; gDNA; germline;
KW    V-gene.
XX
OS    Rattus norvegicus (Norway rat)
OC    cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
OC    Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata;
OC    Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda;
OC    Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Glires;
OC    Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
XX
```

Identification de la séquence

Done

Result: the formalization of the relationships between concepts of IDENTIFICATION allows to represent the rules of the dependency between keywords used in IMGT® databases
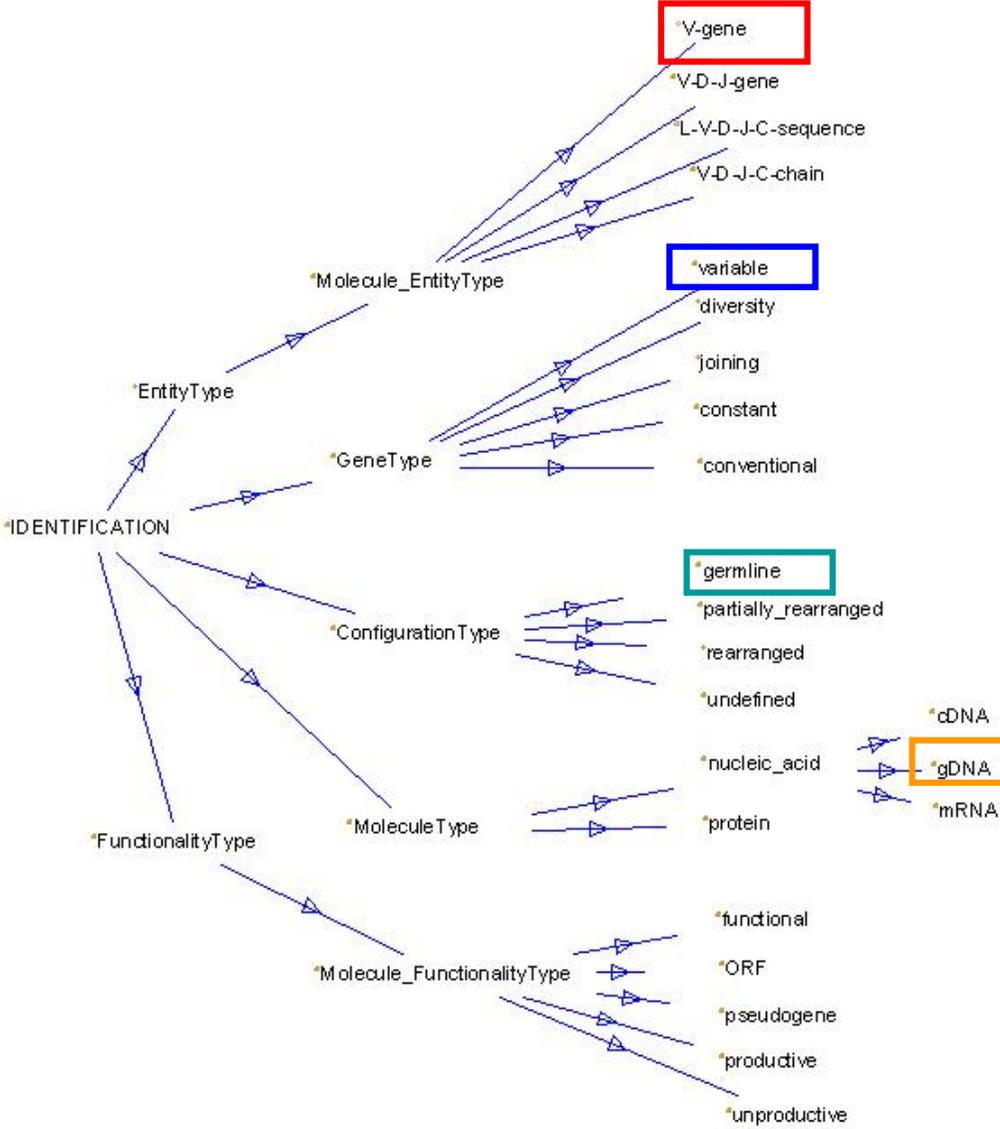
IMGT/LIGM-DB flatfile

```
ID    X07448 IMGT/LIGM annotation : by annotators; genomic DNA; HUM; 618 BP.
XX
AC    X07448;
XX
DT    15-MAY-1995 (Rel. 2, arrived in LIGM-DB )
DT    20-OCT-2008 (Rel. 200843-1, Last updated, Version 10)
XX
DE    Human V35 gene for Ig heavy chain  ;
DE    genomic DNA; germline configuration; Ig-Heavy; regular; functionality
DE    functional; group IGHV; subgroup HV1.
XX
KW    antigen receptor; Immunoglobulin superfamily (IgSF);
KW    immunoglobulin (IG); IG-Heavy; variable; IMGT reference sequence; gDNA;
KW    germline; functional; V-gene.
XX
OS    Homo sapiens (human)
OC    cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa;
OC    Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata;
OC    Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda;
OC    Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates;
OC    Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae;
OC    Homo/Pan/Gorilla group; Homo.
XX
RN    [1]
RP    1-618
RX    PUBMED; 2841108.
RA    Matsuda F., Lee K.H., Nakai S., Sato T., Kodaira M., Zong S.Q., Ohno H.,
RA    Fukuhara S., Honjo T.;
RT    "Dispersed localization of D segments in the human immunoglobulin
RT    heavy-chain locus";
RL    EMBO J. 7(4):1047-1051(1988).
XX
```

IMGT-ONTOLOGY

Pour standardiser, partager, réutiliser et représenter les connaissances en immunogénétique

IMGT-ONTOLOGY repose sur 7 axiomes:

IDENTIFICATION

OBTENTION

CLASSIFICATION

ORIENTATION

DESCRIPTION

LOCALIZATION

NUMEROTATION
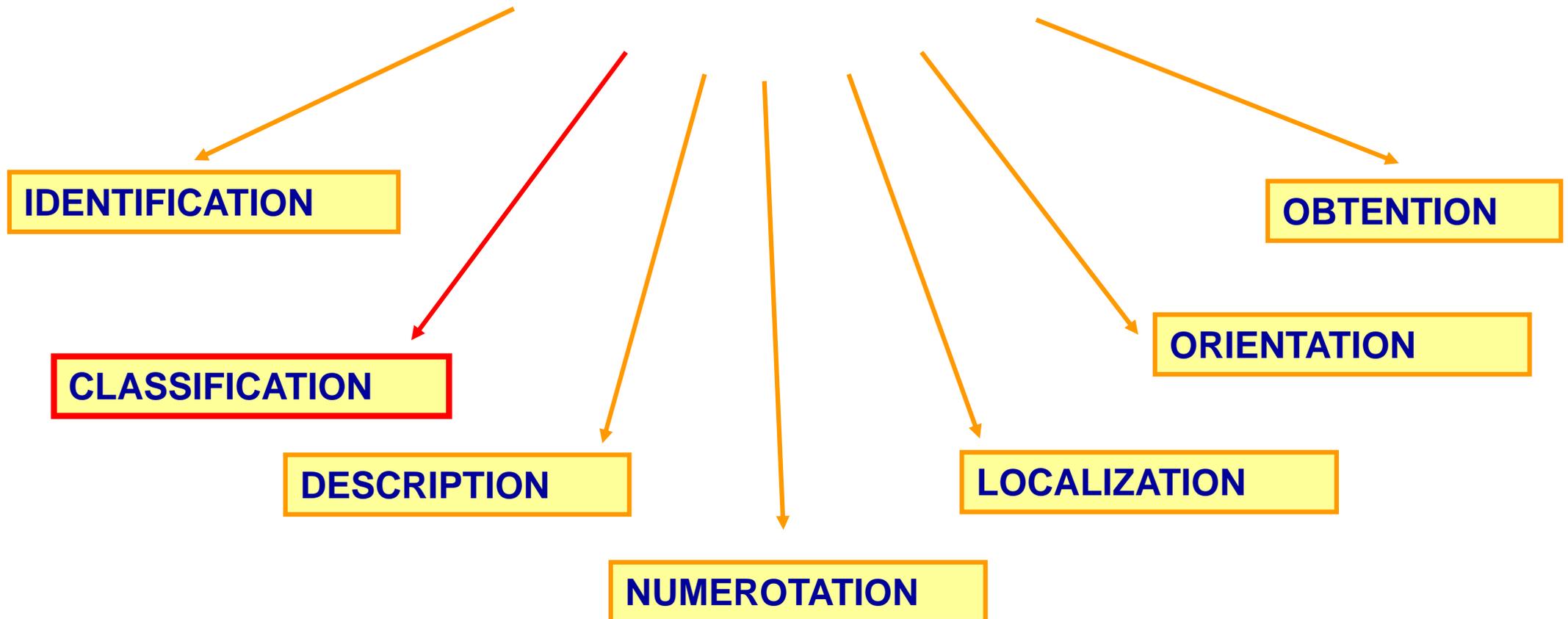
2 – Concepts de classification d'IMGT-ONTOLOGY

Group: A group is a set of genes which share the same 'gene type' and participate potentially to the synthesis of a polypeptide of the same 'chain type'.
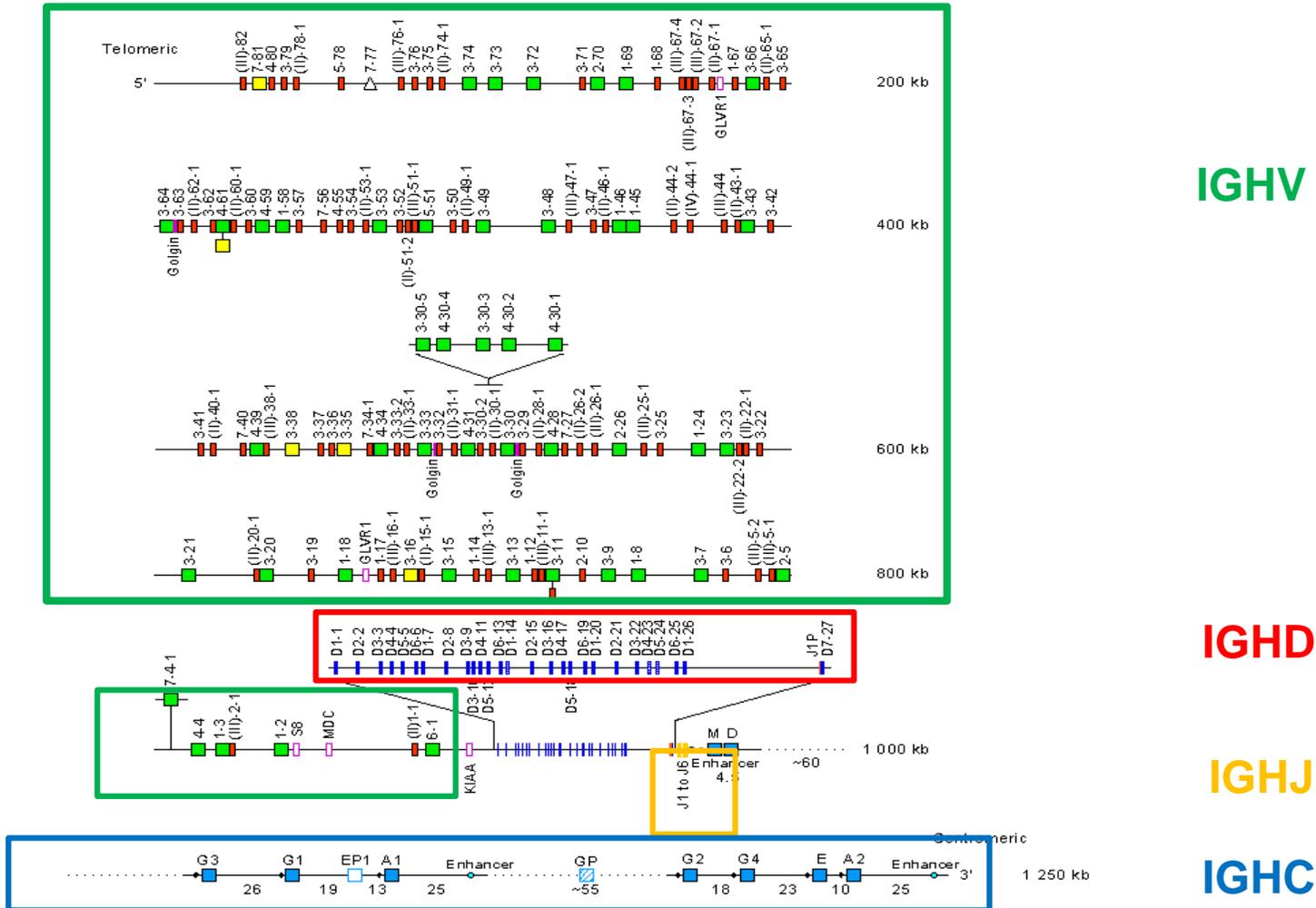
Human IGH locus on chromosome 14 (14q32.33)

2 – Concepts de classification d'IMGT-ONTOLOGY

Subgroup:

A subgroup is a set of IG or TR genes (C-GENE, V-GENE, D-GENE , J-GENE)
which belong to the same group, in a given species, and which share at least
75% identity at the nucleotide level (in the germline configuration for V, D, and J).

Concepts of CLASSIFICATION: IMGT® gene nomenclature

http://www.imgt.org

Human IGL locus (22q11.2)
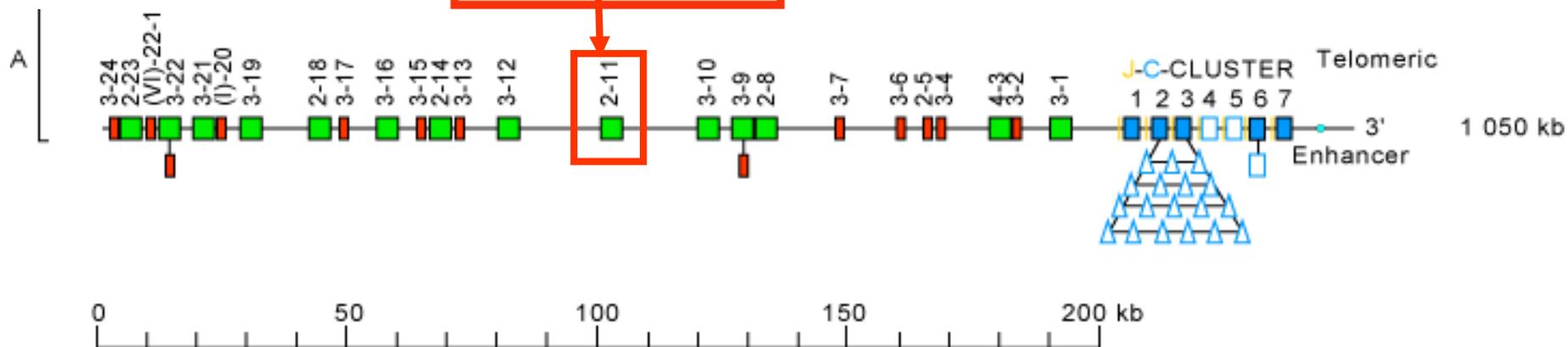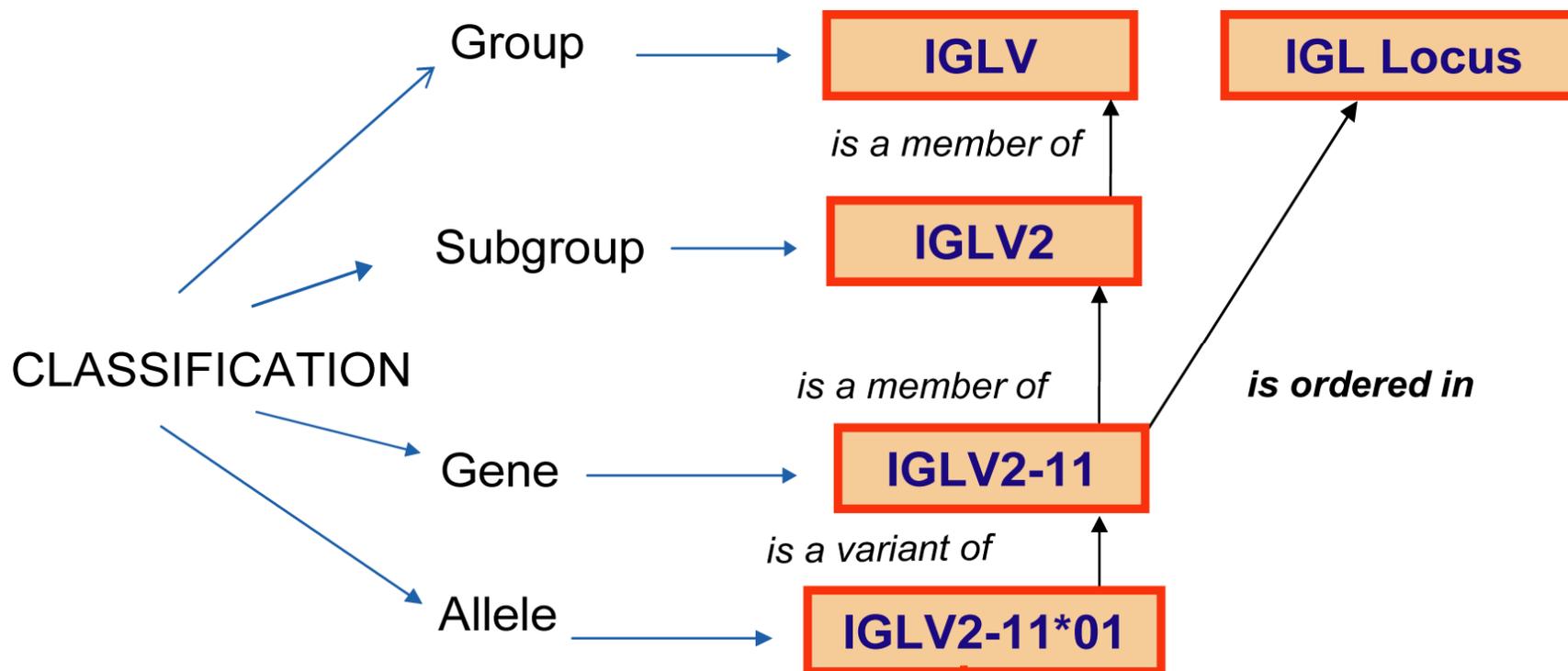
IMGT-ONTOLOGY

Pour standardiser, partager, réutiliser et représenter les connaissances en immunogénétique

IMGT-ONTOLOGY repose sur 7 axiomes:

IDENTIFICATION

OBTENTION

CLASSIFICATION

ORIENTATION

DESCRIPTION

LOCALIZATION

NUMEROTATION
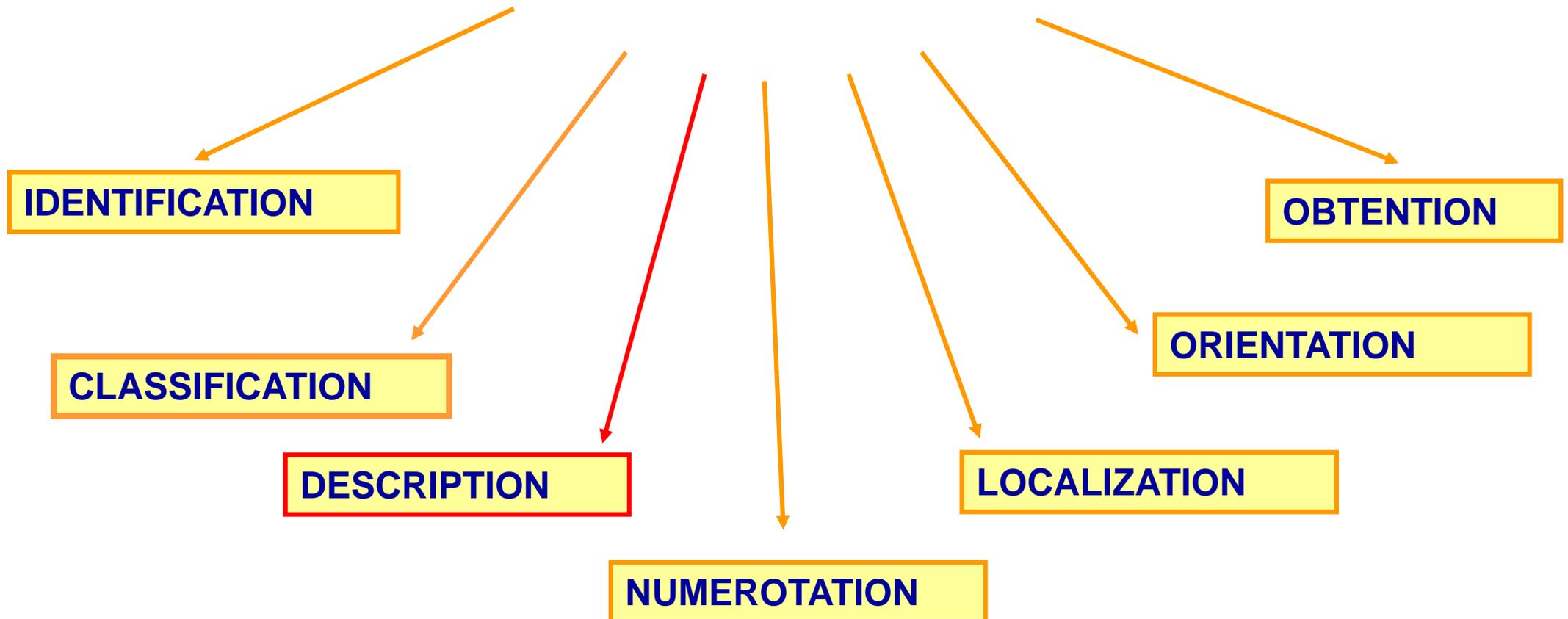
3- Concepts de description: décrire les motifs constitutifs

1 – Description d'une entité 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
aggtcacca tgaccaggga cacgtccatc agcacagcct acatggagct gagcaggctg      420
agatctgacg acacggccgt gtattactgt gcgagagaca cagtgtgaaa tgaaa acccacatcc
tgagggtg c agaaacccaa gggaggaggc ag
```


V-GENE

5'UTR — L-PART1 — V-INTRON — L-PART2 — V-REGION — V-RS — 3'UTR

INIT-CODON DONOR-SPLICE ACCEPTOR-SPLICE 1st-CYS 23 2nd-CYS 104 V-HEPTAMER V-SPACER V-NONAMER
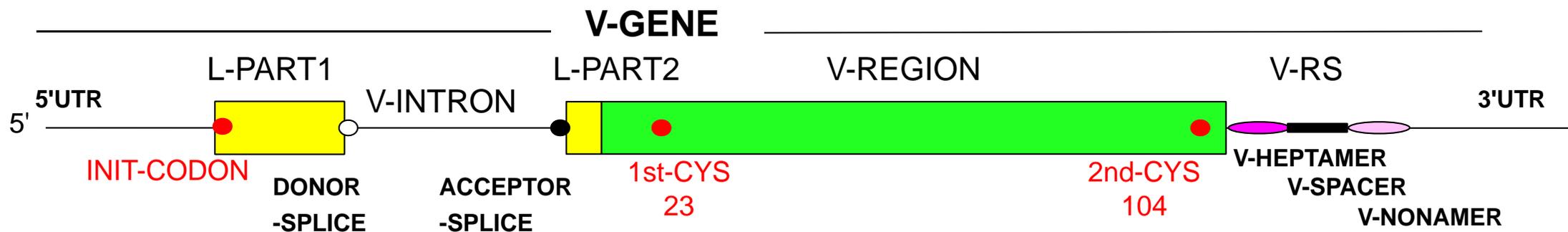
3- Concepts de description: les relations
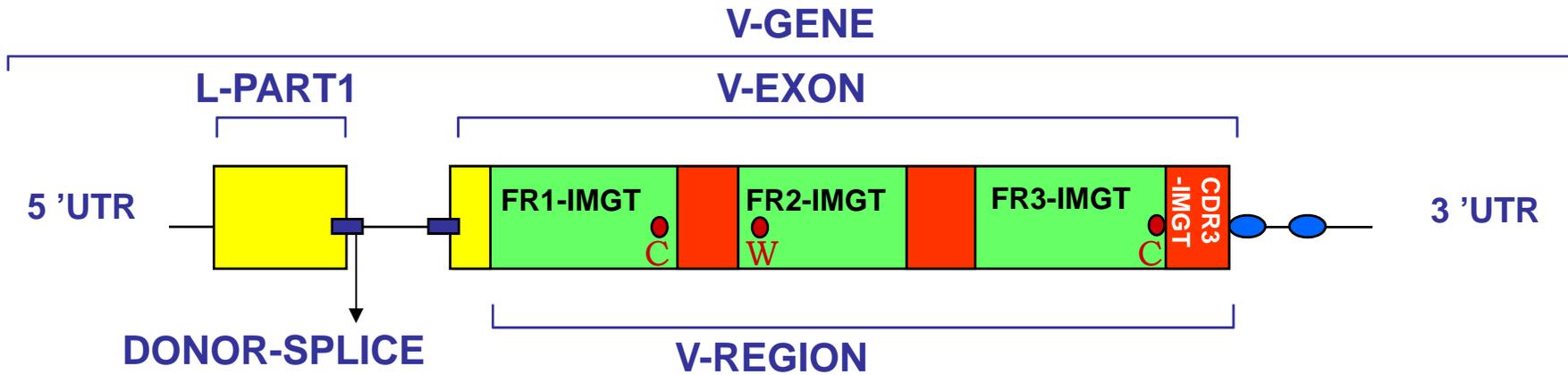
Label 1	Label 2	Label relations	
V-GENE	V-EXON		Includes/included_in
FR3-IMGT	CDR3-IMGT		Adjacent _at_its_5_prime_to/ Adjacent _at_its_3_prime_to
L-PART1	DONOR-SPLICE		Overlaps_in_5_prime/ Overlaps_in_3_prime
V-REGION	FR1-IMGT		Includes_with_same 5_prime/ included_with_same_5_prime_in
V-REGION	CDR3-IMGT		Includes_with_same 3_prime/ included_with_same_3_prime_in

Protégé
(http://protege.stanford.edu/)

Protégé est un logiciel gratuit (JAVA), plate-forme open-source qui fournit une une suite d'outils pour construire des bases de connaissance et des ontologies.

Protégé inclut de nombreux plugins pour la manipulation et la représentation d'ontologies dans différents formats.

L'éditeur Protégé-OWL permet aux utilisateurs de construire des ontologies pour le Web sémantique en OWL.

Application: formalisation dans Protégé des labels qui composent Le V-GENE et de leur relations
(Source d'information: les informations de l'article de Biochimie)

1- faire l'inventaire des termes (qui correspondent à autant de concepts)
2- les classer dans 2 catégories: codants ou non codants
3- créer l'arborescence des concepts
4- les visualiser avec OntoGraf

Relation	Reciprocal relation (inverse)
"adjacent_at_its_5_prime_to"	"adjacent_at_its_3_prime_to"
"included_with_same_5_prime_in",	"includes_with_same_5_prime",
"included_with_same_3_prime_in",	"includes_with_same_3_prime",
"overlap_at_its_5_prime_with"	"overlap_at_its_3_prime_with"
"included_in"	"includes"

5- créer les relations (et leurs caractéristiques)
6- définir les restrictions sur les relations
7- faire les tests de consistance

ImMunoGeneTics

Im

Muno

Gene

Tics

Information system®

http://www.imgt.org