



IMGT-ONTOLOGY 2012

Véronique Giudicelli and Marie-Paule Lefranc*

IMGT®, the international ImMunoGenetics information system®, Université Montpellier 2, Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine, UPR CNRS, Montpellier, France

Edited by:

John Hancock, Medical Research Council, UK

Reviewed by:

Ruth Lovering, University College London, UK

Anna Maria Masci, Duke University, USA

***Correspondence:**

Marie-Paule Lefranc, IMGT®, the international ImMunoGenetics information system®, Université Montpellier 2, Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine, UPR CNRS 1142, 141 rue de la Cardonille, 34396 Montpellier cedex 5, France. e-mail: Marie-Paule.Lefranc@igh.cnrs.fr

Immunogenetics is the science that studies the genetics of the immune system and immune responses. Owing to the complexity and diversity of the immune repertoire, immunogenetics represents one of the greatest challenges for data interpretation: a large biological expertise, a considerable effort of standardization and the elaboration of an efficient system for the management of the related knowledge were required. IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>) has reached that goal through the building of a unique ontology, IMGT-ONTOLOGY, which represents the first ontology for the formal representation of knowledge in immunogenetics and immunoinformatics. IMGT-ONTOLOGY manages the immunogenetics knowledge through diverse facets that rely on the seven axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope: "IDENTIFICATION," "DESCRIPTION," "CLASSIFICATION," "NUMEROTATION," "LOCALIZATION," "ORIENTATION," and "OBTENTION." The concepts of identification, description, classification, and numerotation generated from the axioms led to the elaboration of the IMGT® standards that constitute the IMGT Scientific chart: IMGT® standardized keywords (concepts of identification), IMGT® standardized labels (concepts of description), IMGT® standardized gene and allele nomenclature (concepts of classification) and IMGT unique numbering and IMGT Collier de Perles (concepts of numerotation). IMGT-ONTOLOGY has become the global reference in immunogenetics and immunoinformatics for the knowledge representation of immunoglobulins (IG) or antibodies, T cell receptors (TR), and major histocompatibility (MH) proteins of humans and other vertebrates, proteins of the immunoglobulin superfamily (IgSF) and MH superfamily (MhSF), related proteins of the immune system (RPI) of vertebrates and invertebrates, therapeutic monoclonal antibodies (mAbs), fusion proteins for immune applications (FPIA), and composite proteins for clinical applications (CPCA).

Keywords: IMGT, immunogenetics, immunoinformatics, IMGT-ONTOLOGY, immunoglobulin, antibody, T cell receptor, immune repertoire

INTRODUCTION

Immunogenetics is the science that studies the genetics of the immune system and immune responses. Among them, the adaptive immune response, acquired by vertebrates with jaws or *gnathostomata*, is characterized by an extreme diversity of the specific antigen receptors that comprise the immunoglobulins (IG) or antibodies and the T cell receptors (TR). The potential repertoire of each individual is estimated to comprise about 2×10^{12} different IG and TR, and the limiting factor is only the number of B and T cells that an organism is genetically programmed to produce. This huge diversity results from the complex and unique molecular synthesis and genetics of the antigen receptor chains that include DNA molecular rearrangements (combinatorial diversity) in multiple loci (three for IG and four for TR in humans) located on different chromosomes (four in humans), nucleotide deletions and insertions at the rearrangement junctions (or N-diversity) and, for the IG, somatic hypermutations (for review see Lefranc and Lefranc, 2001a,b).

Owing to the complexity and diversity of the immune repertoires and their implications in fundamental and medical research, immunogenetics represents one of the greatest challenges for data

interpretation: a large biological expertise, a considerable effort of standardization and the elaboration of an efficient system for the management of the related knowledge were required. To answer that challenge, IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>) was created in 1989 by the Laboratoire d'ImmunoGénétique Moléculaire LIGM (Université Montpellier 2 and CNRS) at Montpellier, France (Lefranc et al., 2009; Lefranc, 2011a). IMGT® has become the global reference in immunogenetics and immunoinformatics. IMGT® is a high-quality integrated knowledge resource that provides a common access to standardized data from genome, proteome, genetics, two-dimensional (2D) and three-dimensional (3D) structures. It comprises 7 databases (sequence, gene, structure and specialist databases), 17 online tools and more than 15,000 pages of web resources (Lefranc et al., 2009).

IMGT® has reached that goal through the building of a unique ontology, IMGT-ONTOLOGY started in 1989 and, since then, in constant evolution and extension (Giudicelli and Lefranc, 1999; Lefranc et al., 2004, 2005a, 2008; Duroux et al., 2008; Lefranc, 2011b,c,d,e,f, 2013). IMGT-ONTOLOGY represents the first ontology for the formal representation of knowledge in

immunogenetics and immunoinformatics. IMGT-ONTOLOGY manages the immunogenetics knowledge through diverse facets that rely on the seven axioms of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope: “IDENTIFICATION,” “DESCRIPTION,” “CLASSIFICATION,” “NUMEROTATION,” “LOCALIZATION,” “ORIENTATION,” and “OBTENTION” (Duroux et al., 2008). These axioms postulate that any object, any process and any relation has to be identified, described, classified, numbered, localized, and orientated, and that the way it is obtained can be characterized. The IMGT-ONTOLOGY concepts were generated from these axioms. The concepts of identification, description, classification, and numerotation led to the elaboration of the IMGT® standards that constitute the IMGT Scientific chart: IMGT® standardized keywords (concepts of identification), IMGT® standardized labels (concepts of description), IMGT® standardized IG and TR gene and allele nomenclature (concepts of classification) and IMGT unique numbering and IMGT Collier-de-Perles (concepts of numerotation). One major feature of IMGT-ONTOLOGY is the formalization of the specific relations that link, on a semantic point of view, the different concepts and capture the immunogenetics complexity. These relations are fundamental for data consistency and biological interpretation.

MATERIALS AND METHODS

An ontology is defined as “an explicit specification of a conceptualization” (Gruber, 1993; Guarino and Giaretta, 1995; Guarino, 1997). The building of IMGT-ONTOLOGY has consisted in the conceptualization and in the formalization of the related knowledge in immunogenetics, and in the definition of the relations between concepts. The first concepts were defined as “relevant and fundamental criteria which are needed to characterize IG and TR data” (Giudicelli and Lefranc, 1999). Since then, the IMGT-ONTOLOGY concepts have been largely extended to molecular components other than IG and TR, that include major histocompatibility (MH) proteins of humans and other vertebrates, proteins of the immunoglobulin superfamily (IgSF), and MH superfamily (MhSF), related proteins of the immune system (RPI) of vertebrates and invertebrates, therapeutic monoclonal antibodies (mAbs), fusion proteins for immune applications (FPiA), and composite proteins for clinical applications (CPCA).

Concepts are characterized by their properties which may be simple attributes or relations between concepts. The relation of subsumption (*is_a*) allows to structure the IMGT-ONTOLOGY concepts, and to represent them as nodes of the graph with their level of granularity. The concepts that correspond to the finest level of granularity (and the highest level of precision) in branches of the graph are designated as “leafconcept.” Concepts from which a hierarchy is generated with several levels before reaching the leafconcepts are designated as “highconcept.”

IMGT-ONTOLOGY is being formalized in OWL-DL¹ language using the Protégé editor² (Noy et al., 2003). The formalized concepts of identification are available for downloading or browsing on the National Center for Biomedical Ontology (NCBO)

BioPortal³ (Noy et al., 2009; Musen et al., 2012) and on the IMGT® web site (<http://www.imgt.org>; Lefranc, 2011a,b,c,d,e,f).

The semantic relations (other than subsumption) are formalized as OWL object properties (see OWL 2 Web Ontology Language <http://www.w3.org/TR/owl-primer/>): Object properties allow to link specifically two concepts through the statement “Subject > Property > Object” where “Subject” is the concept being characterized by the object property, “Property” the name of a given property defined in the ontology and “Object” the name of the concept that is linked. These properties are restricted using in particular universal quantification (all connected individuals by the property must be instances of a given class), existential quantification (all individuals of the class for which the property is defined are connected to at least one individuals of the class mentioned in the restriction) and cardinality restrictions (quantification of the number of connected individual with the property). These relations can be displayed on NCBO BioPortal in “IMGT-ONTOLOGY > Terms > Details” page. They are indicated in the “Equivalent Class” section if they are necessary and sufficient to define the concept, or in the “Sub Class Of” section if they are necessary only (for instance, the relations “*is_defined_by*” and “*_has_*” of the “D-gene” (which is a “Molecule_EntityType” leaf-concept, see below “Molecule_EntityType” Concept), are examples of relations in “Equivalent Class” and “Sub Class Of” sections, respectively). The formalization of these relations highlights and focuses on the dependencies between the terms that are closely interconnected at the level of immunogenetics knowledge and set up the constraints that must be respected in the IMGT® databases and tools and in immunoinformatics.

RESULTS

IMGT-ONTOLOGY IDENTIFICATION AXIOM

The IDENTIFICATION axiom of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope (Duroux et al., 2008) postulates that, for molecular components, any molecule and its relations have to be identified (Lefranc, 2011b). IMGT-ONTOLOGY concepts of identification generated from the IDENTIFICATION axiom led to the IMGT® standardized keywords for molecular components (IG, TR, MH, RPI, FPiA, and CPCA) in IMGT® databases and tools.

IMGT-ONTOLOGY concepts of identification

“*Molecule_EntityType*” concept. The objective of IMGT-ONTOLOGY was to identify the type of any molecular entity at each step of its synthesis. An insight of the knowledge related to the synthesis of an IG is schematized in **Figure 1**. It illustrates the concept of “Molecule_EntityType” and the other related concepts of identification and the relations that link them.

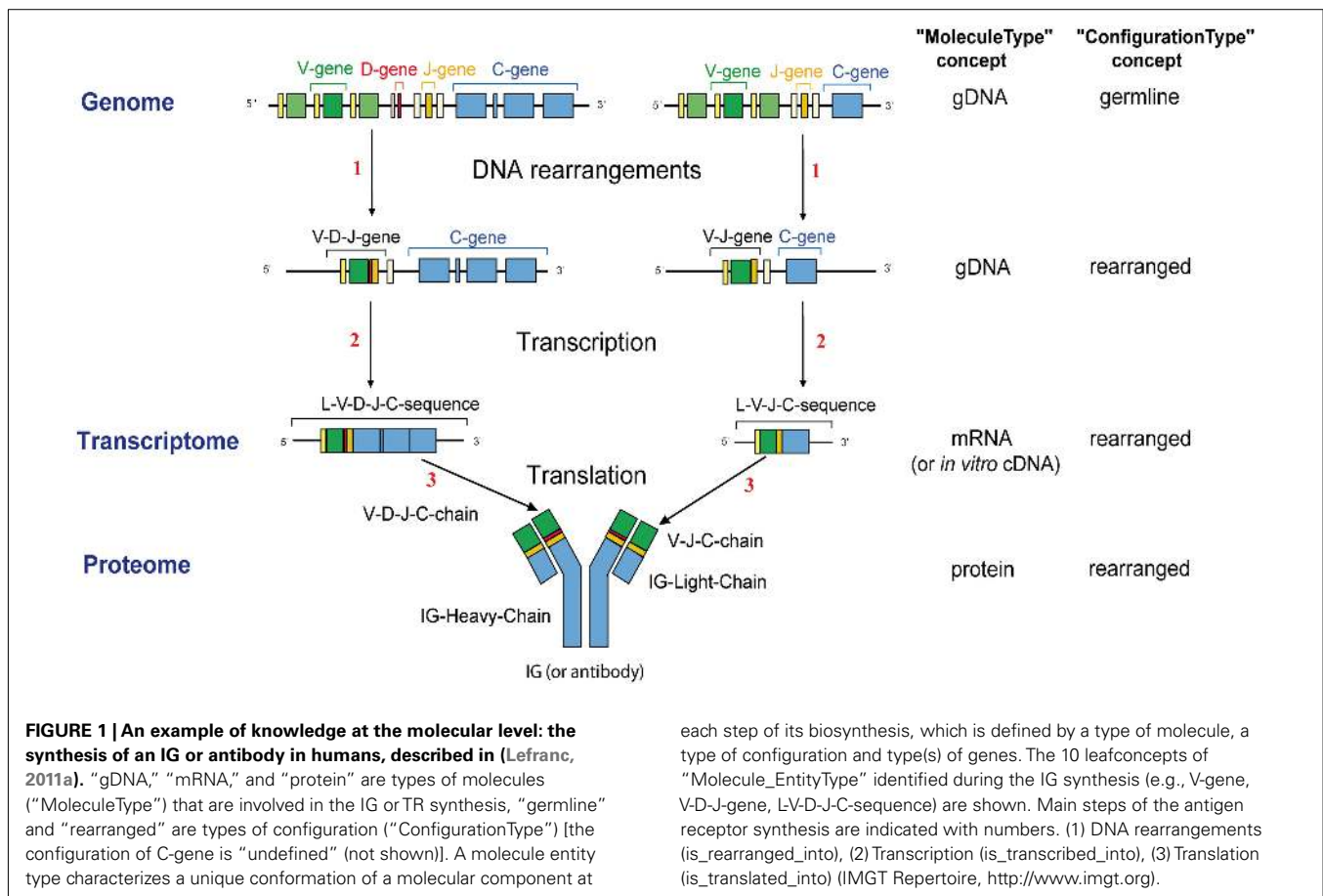
The “Molecule_EntityType” concept is fully defined by the concepts of “MoleculeType,” “GeneType,” and “ConfigurationType” (**Figure 2**).

- “MoleculeType” allows one to identify the type of molecule, based on the type of the constitutive elements and on the concepts of obtention. The “MoleculeType” concept comprises four major leafconcepts: “gDNA,” “mRNA,” “cDNA,” “protein.”

¹<http://www.w3.org/TR/owl2-overview/>

²<http://protege.stanford.edu>

³<http://biportal.bioontology.org/ontologies/1491>



- “GeneType” allows one to identify the type of gene. The “GeneType” concept comprises six leafconcepts: “variable” (V), “diversity” (D), “joining” (J), and “constant” (C) are the four gene types specific of IG and TR, and “conventional-with-leader,” “conventional-without-leader” are the two gene types of conventional genes.
- “ConfigurationType” allows one to identify the type of configuration of a gene, and by extension, the type of configuration of the Molecule_EntityType leafconcepts that contain it. The “ConfigurationType” concept comprises four leafconcepts: “germline,” “partially-rearranged” and “rearranged” for V, D, and J genes and the molecule entities that contain them, and “undefined” for C and conventional genes and related entities.

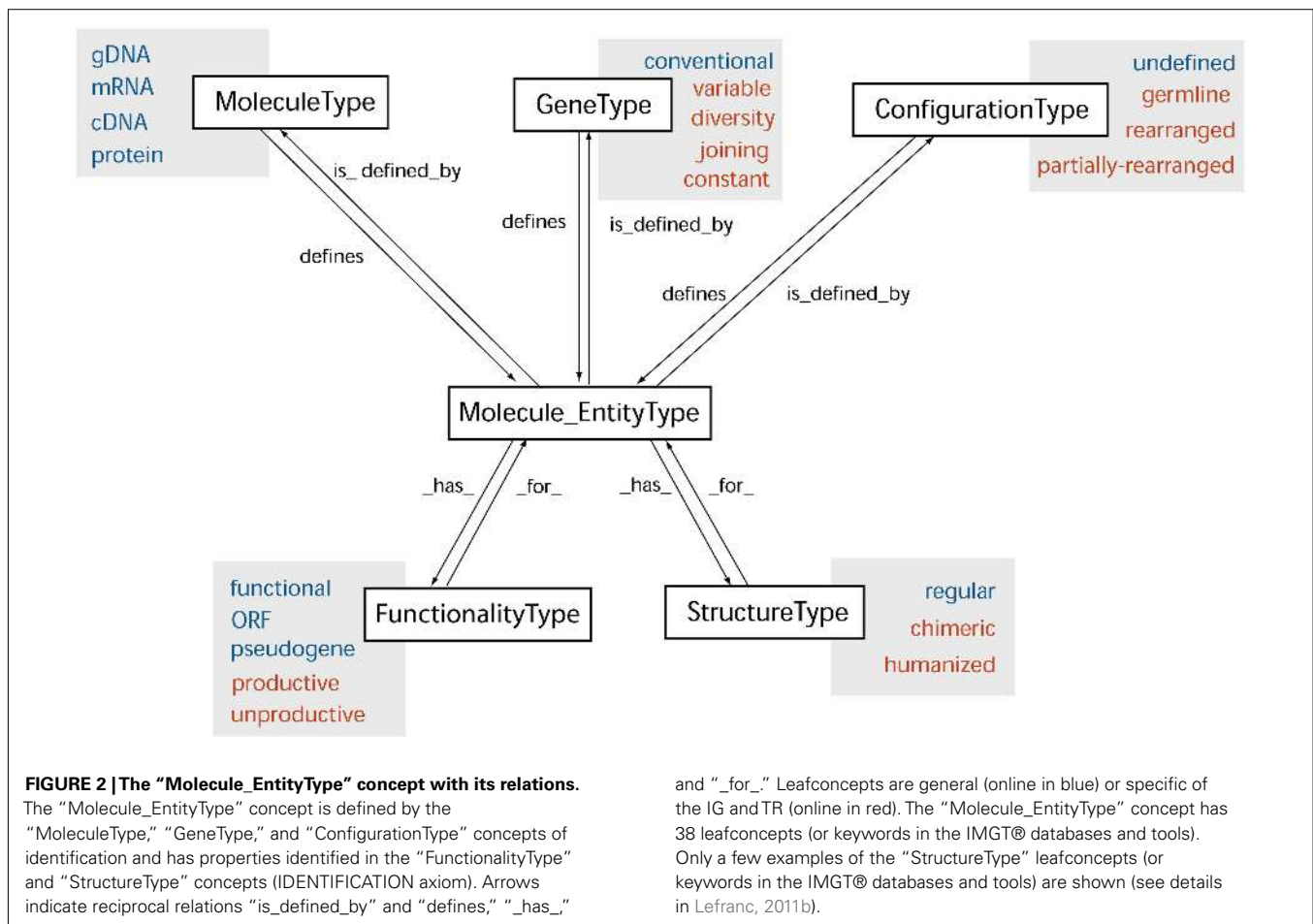
The “Molecule_EntityType” concept comprises 38 leafconcepts (Table 1). For examples, “V-gene” identifies, for gDNA, molecule entities with a germline V gene, “V-D-J-gene” identifies, for gDNA, molecule entities with rearranged V, D, and J genes, and “L-V-D-J-C-sequence” identifies, for cDNA, molecule entities with rearranged V, D, and J genes spliced to a C gene. The four “MoleculeUnit” leafconcepts that are “gene” (10), “transcript” (11), “sequence” (11), and “chain” (6) identify the type of entities based on the “MoleculeType” only, as indicated by the suffix (Table 1).

In addition to the relation “is_defined_by,” a “Molecule_EntityType” “has” properties identified in the “FunctionalityType” and “StructureType” concepts (Figure 2).

- “FunctionalityType” is a concept of identification that allows one to identify, whatever the molecule type (gDNA, cDNA, mRNA, or protein), the type of functionality of a Molecule_EntityType leafconcept. The “FunctionalityType” concept comprises five leafconcepts: “functional,” “ORF” (open reading frame) and “pseudogene” identify the functionality of Molecule_EntityType leafconcepts in undefined configuration (conventional genes and IG and TR C genes), or in germline configuration (IG and TR V, D, and J genes before DNA rearrangements); “productive” and “unproductive” identify the functionality of Molecule_EntityType leafconcepts in rearranged or partially-rearranged configuration (IG and TR entities after DNA rearrangements, and by extension fusion entities resulting from translocations, and hybrid entities obtained by biotechnology molecular engineering).
- “StructureType” is a concept of identification that allows one to identify, whatever the molecule type (gDNA, cDNA, mRNA, protein), the type of structure of Molecule_EntityType leafconcepts.

The semantic relations of “Molecule_EntityType” are formalized as properties (in OWL).

“ChainType,” “DomainType,” and “ReceptorType” concepts. One of the goals of IMGT-ONTOLOGY has been to represent knowledge in order to manage molecular components from



sequences to 3D structures in IMGT® databases and tools. The three concepts “ChainType,” “DomainType,” and “ReceptorType” have been fundamental in that knowledge representation.

“ChainType” is a concept of identification that allows one to identify the type of chain. “ChainType” is a “highconcept” that comprises four levels (Figure 3): “MolecularComponentLevelChainType,” “ReceptorLevelChainType,” “ClassLevelChainType,” and “GeneLevelChainType.” The concepts are organized in an acyclic graph based on the subsumption relation, the depth of which depends on the precision that needs to (or that can be) reported for the data identification. The finest level of granularity, the “GeneLevelChainType” concept, identifies the type of chain by reference to the gene(s) which code(s) the chain. It represents the main concept for a very precise identification because it establishes a relationship with “Gene” (concept of classification) (the reciprocal relations are: “is_coded_by” and “codes”). The number of “ChainType” leafconcepts of the “GeneLevelChainType” depends on the number of functional genes and ORF (“FunctionalityType”) per haploid genome, in a given species (in the case of the IG and TR genes, it is the number of functional and ORF C genes which is taken into account).

The “ChainType” concept is defined by the “Molecule_EntityType” and the “DomainType” concepts of identification, and also defined by concepts of classification (see IMGT-ONTOLOGY CLASSIFICATION Axiom) as the type of chain

depends on the taxon (Figure 4). “DomainType” allows one to identify the type of domain. A domain is a chain subunit characterized by its three-dimensional (3D) structure, and by extension its amino acid sequence and the nucleotide sequence which encodes it.

The “ChainType” concept represents a key concept that allows to link the “Molecule_EntityType” (sequences in databases) to the concept of “ReceptorType” (3D structures in databases; Figure 4). “ReceptorType” allows one to identify the type of receptor. “ReceptorType” is defined by the “ChainType” leafconcept(s) that identify the associated chains of a receptor. “ReceptorType” is a “highconcept” with a hierarchy of four levels of granularity (depending on the “ChainType” hierarchy). The “ReceptorType” concept has properties identified in the “FormatType,” “SpecificityType,” and “FunctionType” concepts (Figure 4; Lefranc, 2011b).

IMGT® standardized keywords in databases and tools

The leafconcepts of identification are IMGT® standardized keywords in the IMGT® databases and tools (Lefranc, 2005). The list of IMGT® standardized keywords is available from the IMGT/LIGM-DB database (Giudicelli et al., 2006) query page (IMGT® Home page; <http://www.imgt.org>) and in the IMGT Scientific chart at <http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGT3Dkeywords.html>. More than 325 IMGT® standardized keywords (189 for sequences and 137 for

Table 1 | “Molecule_EntityType” leafconcepts and related concepts. The 38 “Molecule_EntityType” leafconcepts are shown with the leafconcepts of “GeneType,” “ConfigurationType,” and “MoleculeType” that define them. The four leafconcepts of “MoleculeUnit” are based on “MoleculeType” only.

MoleculeUnit leafconcepts	Molecule_EntityType leafconcepts	GeneType leafconcepts	ConfigurationType leafconcepts	MoleculeType leafconcepts		
gene	V-gene*	V	germline	gDNA		
	D-gene*	D				
	J-gene*	J				
	J-C-gene	J, C				
	C-gene*	C	undefined			
	conventional-gene	conventional				
	V-D-gene	V, D	partially-rearranged			
	D-J-gene	D, J				
	V-J-gene*	V, J	rearranged			
	V-D-J-gene*	V, D, J				
transcript	L-V-transcript	V	germline	mRNA		
	D-transcript	D				
	J-transcript	J				
	J-C-transcript	J, C				
	C-transcript	C	undefined			
	Lnt-transcript	conventional				
	nt-transcript	conventional				
	L-V-D-transcript	V, D	partially-rearranged			
	D-J-C-transcript	D, J, C				
	L-V-J-C-transcript	V, J, C	rearranged			
	L-V-D-J-C-transcript	V, D, J, C				
	sequence	L-V-sequence	V		germline	cDNA
		D-sequence	D			
		J-sequence	J			
J-C-sequence		J, C				
C-sequence		C	undefined			
Lnt-sequence		conventional				
nt-sequence		conventional				
L-V-D-sequence		V, D	partially-rearranged			
D-J-C-sequence		D, J, C				
L-V-J-C-sequence*		V, J, C	rearranged			
L-V-D-J-C-sequence*		V, D, J, C				
chain		L-AA-chain	conventional	undefined	Protein	
		AA-chain	conventional			
		L-V-J-C-chain	V, J, C	rearranged		
	L-V-D-J-C-chain	V, D, J, C				
	V-J-C-chain*	V, J, C				
	V-D-J-C-chain*	V, D, J, C				

*Indicates the 10 leafconcepts that are the most classical representatives of IG and TR identification. These leafconcepts are illustrated in **Figure 1**.

3D structures) were precisely defined. They represent the controlled vocabulary assigned during the annotation process and allow standardized search criteria for querying the IMGT® databases and for the extraction of sequences and 3D structures. IMGT/HighV-QUEST, the IMGT® tool for analysis of IG and TR nucleotide sequences obtained from next generation sequencing (NGS; Alamyar et al., 2012), provides an evaluation of the configuration (“ConfigurationType”) and, accordingly, of the sequence functionality (“FunctionalityType”): such precision and standardization in the NGS results are of the utmost importance for the

reuse of data for the statistical analyses required for the comparison of immune repertoires (Prabakaran et al., 2012) and for data interpretation.

IMGT-ONTOLOGY DESCRIPTION AXIOM

The DESCRIPTION axiom of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope (Duroux et al., 2008) postulates that, for molecular components, any molecule and its relations have to be described (Lefranc, 2011c). IMGT-ONTOLOGY concepts of description generated from the DESCRIPTION axiom led

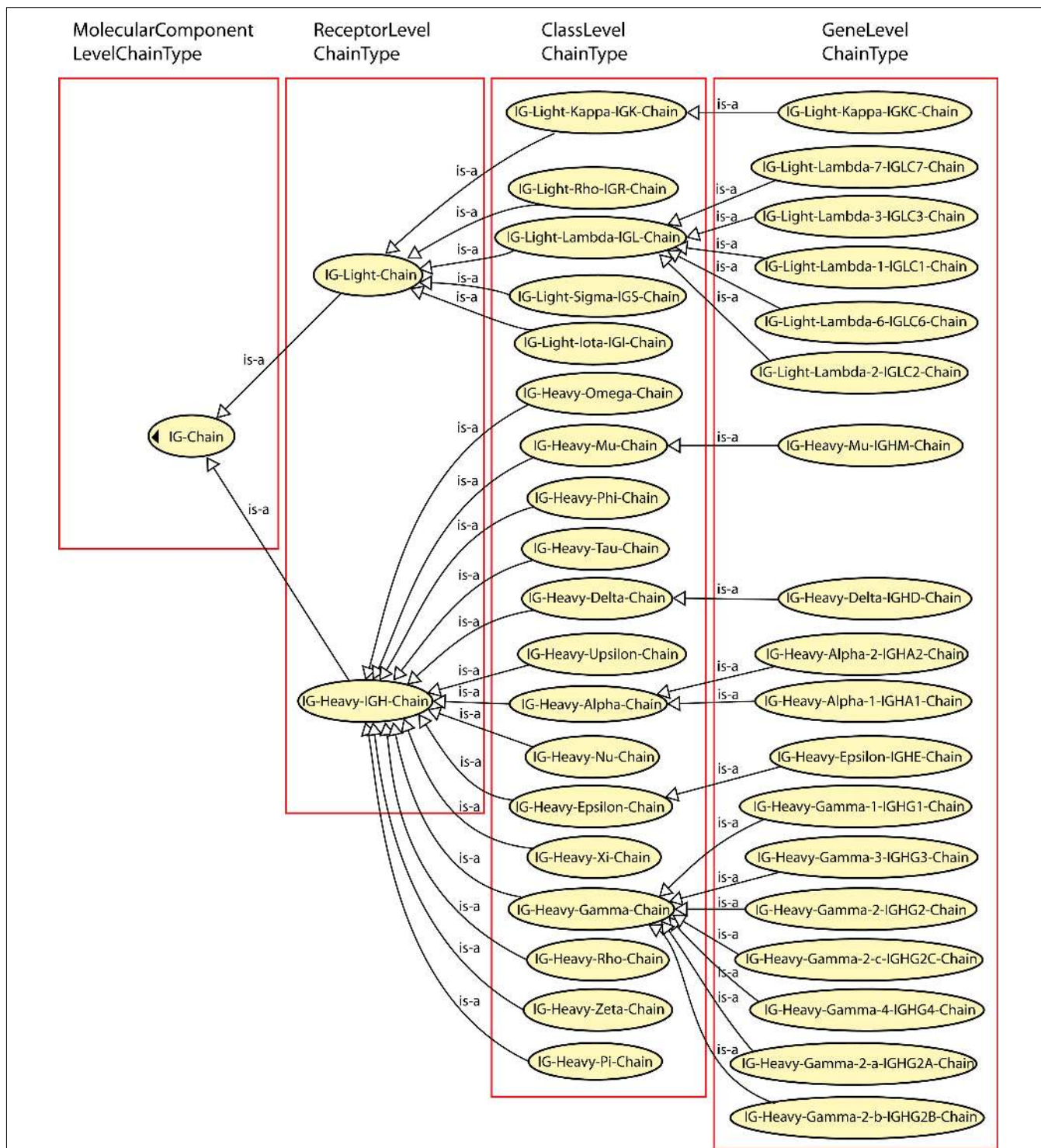


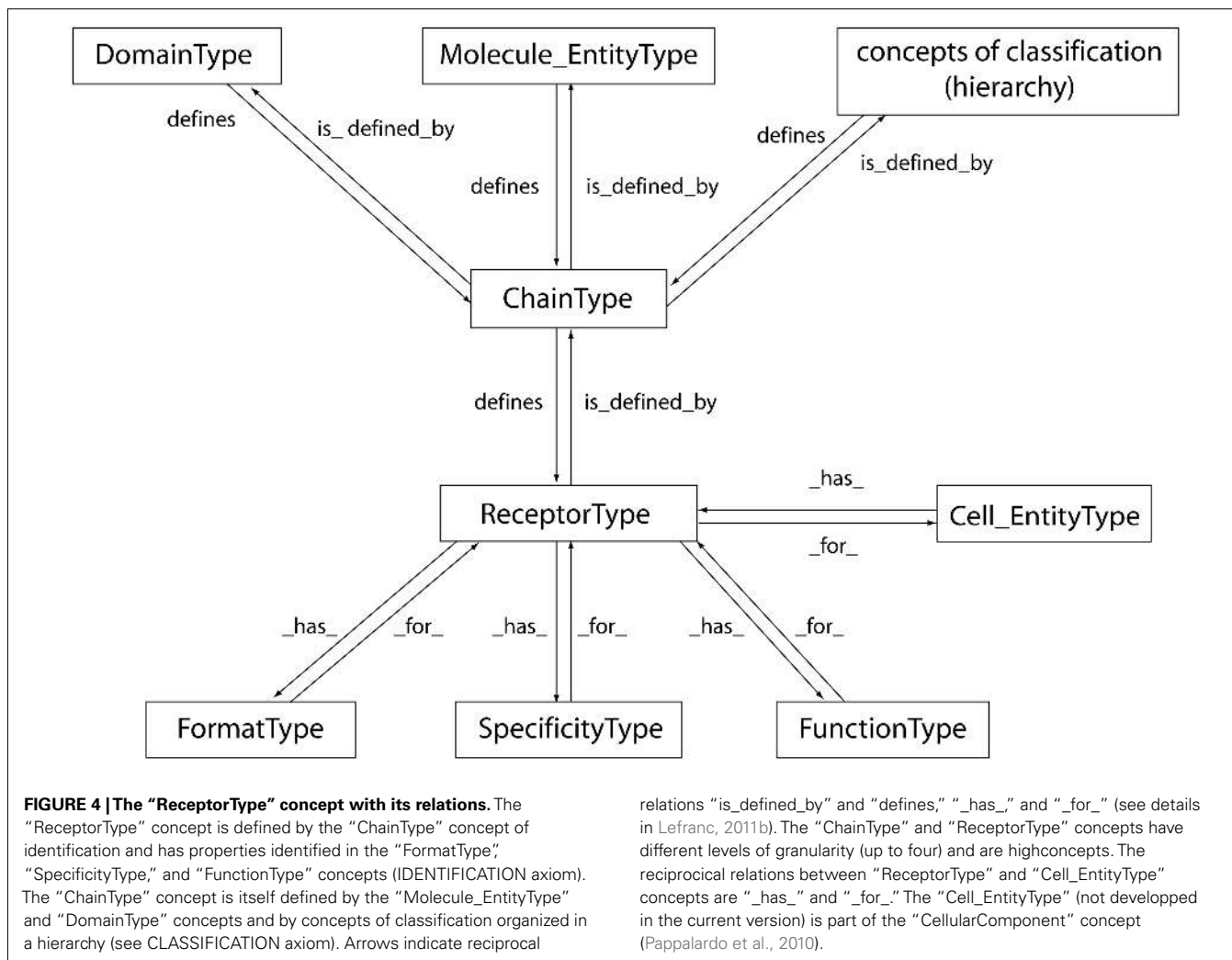
FIGURE 3 | "ChainType" "highconcept." The hierarchy of "ChainType" for the identification of IG chains comprises four levels of granularity which are associated with an increasing level of precision:

the "MolecularComponentLevelChainType," the "ReceptorLevelChainType," the "ClassLevelChainType," and the "GeneLevelChainType."

to the IMGT® standardized labels for molecular components (IG, TR, MH, RPI, FPIA, and CPCA) in IMGT® databases and tools.

IMGT-ONTOLOGY concepts of description

Concepts of description have been progressively elaborated in order to take into account the entities of the different steps of



the molecular synthesis of the antigen receptors (IG and TR) and, more generally, of all molecular components and to describe all motifs of biological interest of sequences and 2D and 3D structures in databases and tools.

“Molecule_EntityPrototype” concept. The “Molecule_EntityPrototype” is a concept, generated from the DESCRIPTION axiom, that provides the description of the “Molecule_EntityType” concept (IDENTIFICATION axiom). There are as many leafconcepts in the “Molecule_EntityPrototype” as there are leafconcepts in the “Molecule_EntityType.” Thus the “Molecule_EntityPrototype” comprises 38 leafconcepts that describe the organization of each entity with its constitutive motifs and relations. Each “Molecule_EntityPrototype” leafconcept is linked to a “Molecule_EntityType” leafconcept by the reciprocal relations “describes” and “is_described_by.” For example, a “V-gene” is described by “V-GENE,” and a “V-D-J-gene” by “V-D-J-GENE.” Leafconcepts of description (labels in the IMGT® databases and tools) are written in capital letters.

Prototypes and relations between concepts of description. In order to visualize the organization of each entity, prototypes were

defined. A prototype is a graphical representation of a “Molecule_EntityPrototype” leafconcept. Two prototypes of “V-GENE” and “V-D-J-GENE” are shown in **Figure 5** as examples of a germline entity and of a rearranged entity, respectively. Twenty-seven labels for “V-GENE” and 33 labels for “V-D-J-GENE” (20 of them being shared by the two prototypes), on a total of 277 different labels for sequences in IMGT/LIGM-DB, are necessary and sufficient for a complete description of these prototypes. The organization of a prototype is based on the relations that order two labels.

IMGT-ONTOLOGY formalizes the topological relations that define the relative position of two labels. A set of twelve relations are necessary and sufficient to describe the relations between labels in a prototype (Duroux et al., 2008; Lane et al., 2010; **Table 2**). The reciprocal relations “is_in_5_prime_of” and “is_in_3_prime_of” describe the relative position of labels on a 5′–3′ DNA strand when there is no intersection or contiguity between labels (Lane et al., 2010).

IMGT® standardized labels in databases and tools

The leafconcepts of description are IMGT® standardized labels in the databases and tools (Lefranc, 2005). The IMGT®

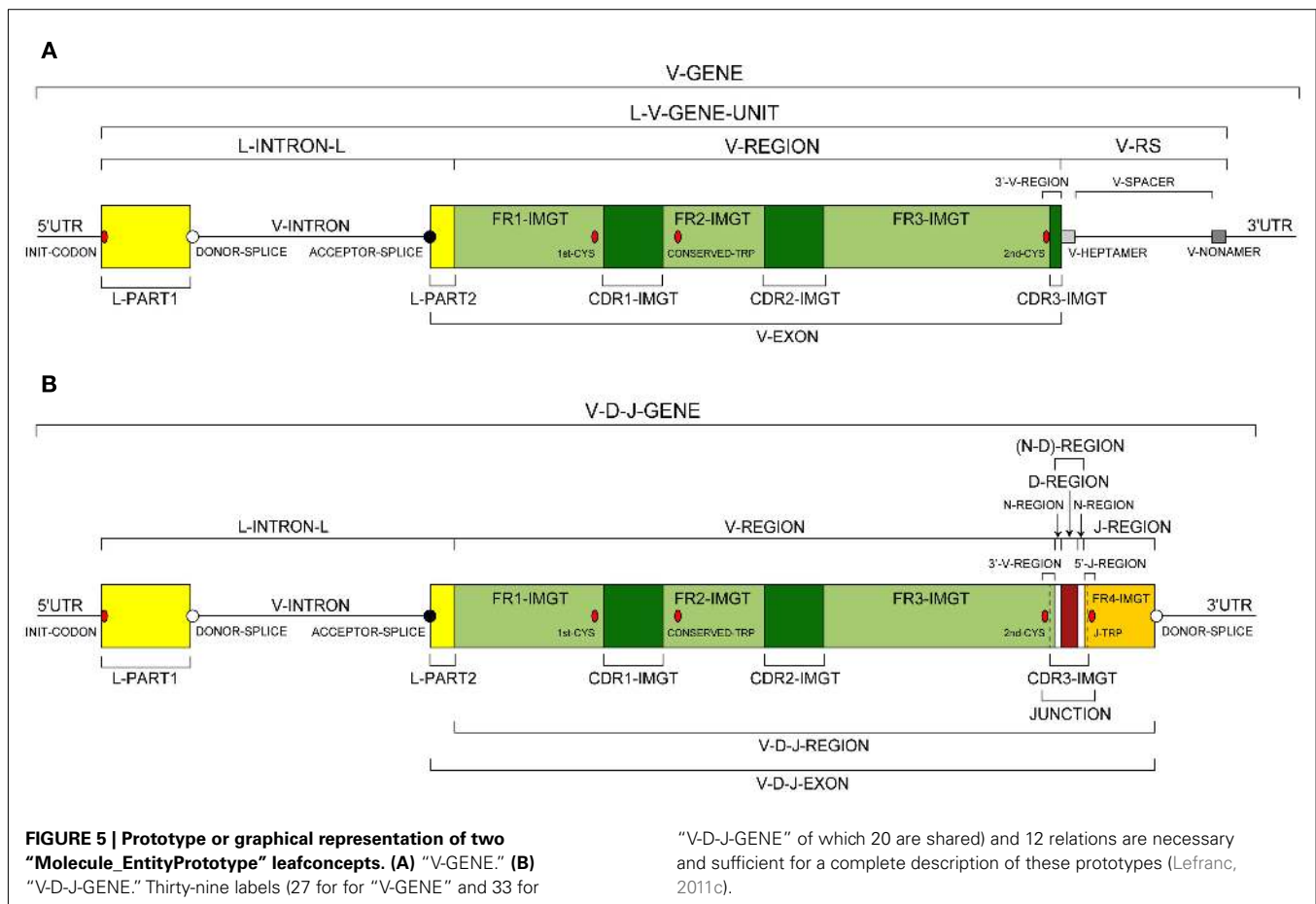


Table 2 | IMGT-ONTOLOGY relations between labels used for the description of prototypes.

Relation	Reciprocal relation
"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"
"overlaps_at_its_5_prime_with"	"overlaps_at_its_3_prime_with"
"included_in"	"includes"
"is_in_5_prime_of"	"is_in_3_prime_of"

standardized labels are available from the IMGT/LIGM-DB database (Giudicelli et al., 2006) query page (IMGT® Home page; <http://www.imgt.org>) and in the IMGT Scientific chart at: <http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGT3Dkeywords.html> (definitions of these labels are available at: <http://www.imgt.org/IMGTScientificChart/SequenceDescription/IMGT3Dlabeldef.html>). More than 560 IMGT® standardized labels (277 for sequences and 285 for 3D structures) were precisely defined.

IMGT/Automat, the IMGT® tool for the annotation of rearranged cDNA (Giudicelli et al., 2005a) implements corresponding labels and prototypes. IMGT® standardized labels and the organization of "Molecule_EntityPrototype" have recently

been implemented in IMGT/LIGMotif for the automation of the annotation of large genomic sequences (Lane et al., 2010). A set of specific labels was defined to describe the different organizations of IG and TR genes in clusters at the scale of the locus or of the chromosome.

been implemented in IMGT/LIGMotif for the automation of the annotation of large genomic sequences (Lane et al., 2010). A set of specific labels was defined to describe the different organizations of IG and TR genes in clusters at the scale of the locus or of the chromosome.

IMGT-ONTOLOGY CLASSIFICATION AXIOM

The CLASSIFICATION axiom of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope (Duroux et al., 2008) postulates that, for molecular components, any molecule and its relations have to be classified (Lefranc, 2011d). IMGT-ONTOLOGY concepts of classification generated from the CLASSIFICATION axiom led to the IMGT® standardized IG and TR gene and allele nomenclature.

IMGT-ONTOLOGY concepts of classification

The IMGT® standardized gene and allele nomenclature is based on the concepts of classification, generated from the CLASSIFICATION axiom, which defines the principles for the nomenclature of highly polymorphic multigene loci and families. In particular, the concepts of classification have allowed to classify the genes whatever the antigen receptor (IG or TR), whatever the locus (e.g., for mammals, immunoglobulin heavy IGH, immunoglobulin kappa IGK, immunoglobulin lambda IGL, T cell receptor alpha TRA, T cell receptor beta TRB, T cell receptor gamma TRG, and T cell receptor delta TRD), whatever the gene configuration (germline,

undefined, or rearranged), and whatever the species, from fish to human. Among the concepts of classification, the “Group,” “Subgroup,” “Gene,” and “Allele” concepts are essential for the IMGT® gene nomenclature (Giudicelli and Lefranc, 1999). They are shown with their semantic relations in **Figure 6** that are used for the V gene designation.

IMGT® standardized IG and TR gene and allele nomenclature

In the context of the gene and allele classification, ontological principles defined in IMGT-ONTOLOGY have preceded the IMGT® standardized gene and allele nomenclature. This has been true for the human genes, and all IMGT® IG and TR gene names (Lefranc, 2000a,b; Lefranc and Lefranc, 2001a,b) were defined before the complete human genome sequencing (Lander et al., 2001; Venter et al., 2001). This is still the case for newly sequenced genomes and the denomination of IG and TR genes from a newly sequenced species is considerably facilitated by the preexisting nomenclature principles and rules. Full IMGT® standardized gene name comprises the latin names of the genus and species (e.g., *Homo sapiens* IGHV1-2). Gene names used in natural language and in publications may include abbreviation if needed for tables or figures (6-letter code for genus and species, 9-letter code for genus, species, and subspecies).

Interoperability between IMGT, HGNC, and NCBI

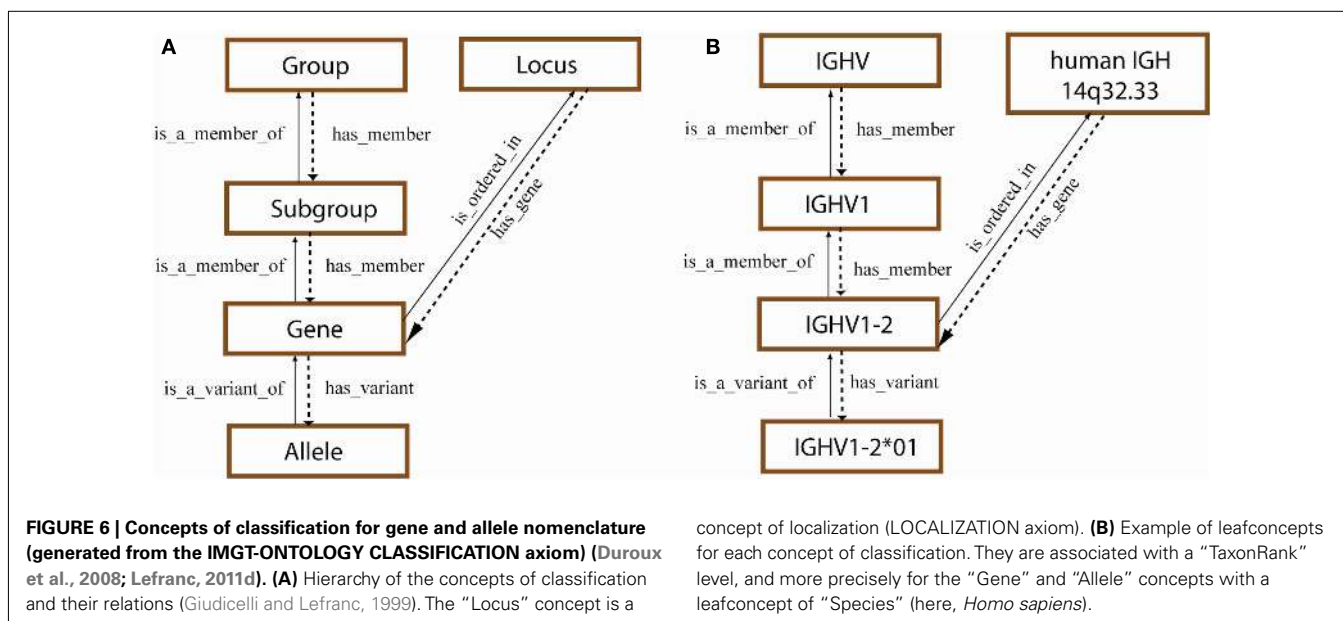
Since the creation of IMGT®, the international ImMunoGeneTics information system® in 1989, at New Haven during the 10th Human Genome Mapping Workshop (HGM10), the standardized classification and nomenclature of the IG and TR of humans and other vertebrate species have been under the responsibility of the IMGT Nomenclature Committee (IMGT-NC). The IMGT® gene nomenclature for human IG and TR genes (Lefranc, 2000a,b; Lefranc and Lefranc, 2001a,b) was approved by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) in 1999 (Wain et al., 2002) and endorsed by the World Health Organization-International Union

of Immunological Societies (WHO-IUIS; Lefranc, 2007, 2008). IMGT® IG and TR gene names are the official international reference and have been entered in IMGT/GENE-DB, the IMGT® gene database (Giudicelli et al., 2005b), in the Human Genome Database (GDB; Letovsky et al., 1998), in LocusLink at the National Center for Biotechnology Information (NCBI) in 1999–2000 (Maglott et al., 2000), in NCBI Entrez Gene when this gene database superseded LocusLink (Maglott et al., 2007), in NCBI Gene and in NCBI MapViewer, in Ensembl at the European Bioinformatics Institute (EBI) in 2006 (Hubbard et al., 2002), and in the Vega Genome Browser at the Wellcome Trust Sanger Institute (Ashurst et al., 2005). Amino acid sequences of human IG and TR C genes were provided to UniProt in 2008 (Bairoch et al., 2009). Close collaborations have been developed to maintain interoperability between the databases, with HGNC (Wain et al., 2004; Bruford et al., 2008), NCBI Gene (Maglott et al., 2011), Ensembl, Vega (Wilming et al., 2008), the Mouse Genomic Nomenclature Committee (MGNC), the Nomenclature Committees of newly sequenced genomes, for example, ZFIN for the zebrafish *Danio rerio* (Bradford et al., 2011) or external team contribution, for example, TRB locus of the rhesus macaque *Macaca mulatta* (Greenaway et al., 2009). IG and TR genes are also integrated in the HUGO ontology and NCI Metathesaurus available on the NCBO BioPortal⁴. Mapping between the HUGO ontology and IMGT-ONTOLOGY will be developed with the formalization of the concepts of classification in OWL.

IMGT-ONTOLOGY NUMEROTATION AXIOM

The NUMEROTATION axiom of the Formal IMGT-ONTOLOGY or IMGT-Kaleidoscope (Duroux et al., 2008) postulates that, for molecular components, any molecule and its relations have to be numbered (Lefranc, 2011e,f). Two major IMGT-ONTOLOGY concepts of numerotation generated from the NUMEROTATION

⁴<http://biportal.bioontology.org/ontologies/>



axiom comprises the “IMGT_unique_numbering” and “IMGT_Collier_de_Perles” (IMGT unique numbering and IMGT Colliers de Perles in IMGT® databases and tools).

“IMGT_unique_numbering”

The “IMGT_unique_numbering” concept (Lefranc, 2011e) defines a systematic and coherent numbering (amino acids and codons) for the description of “DomainType” leafconcepts. The “IMGT_unique_numbering” was originally defined for the IG and TR V-DOMAIN (Lefranc, 1997). It provides a standardized delimitation of the framework regions (FR-IMGT) and complementarity determining regions (CDR-IMGT), and therefore allows to correlate each position (amino acid or codon) with the structure (beta strand, loop, beta turn) and the function (antigen binding) of the V-DOMAIN. FR-IMGT and CDR-IMGT lengths became a major property of the IG and TR V-DOMAIN. The “IMGT_unique_numbering” concept has been extended to the V-LIKE-DOMAIN of the IgSF other than IG and TR (Lefranc, 1999; Lefranc et al., 2003), to the C domain (C-DOMAIN of IG and TR and C-LIKE-DOMAIN of IgSF other than IG and TR; Lefranc et al., 2005b) and to the G domain (G-DOMAIN of MH and G-LIKE-DOMAIN of MhSF other than MH) (Lefranc et al., 2005c). Thus, the “IMGT_unique_numbering” concept allows to number domain types that are characteristic of protein superfamilies, whatever the species, the molecule type or the chain type. Three leafconcepts have been defined for the variable (V) domain, the constant (C) domain, and the groove (G) domain: “IMGT_unique_numbering_for_V_domain” (Lefranc, 1997, 1999; Lefranc et al., 2003) and “IMGT_unique_numbering_for_C_domain” (Lefranc et al., 2005b) of the IG, TR and IgSF, and “IMGT_unique_numbering_for_G_domain” (Lefranc et al., 2005c) of the MH and MhSF.

“IMGT_Collier_de_Perles”

The “IMGT_Collier_de_Perles” concept (Lefranc, 2011f) corresponds to the graphical 2D representation of domains based on the set of rules defined by the “IMGT_unique_numbering.” This original and unique approach allows one to bridge the gap between sequences and 2D and 3D structures and greatly facilitates the domain comparison, position per position. Three leafconcepts are defined: “IMGT_Collier_de_Perles_for_V_domain” (Lefranc, 1999; Lefranc et al., 2003), “IMGT_Collier_de_Perles_for_C_domain” (Lefranc et al., 2005b) and “IMGT_Collier_de_Perles_for_G_domain” (Lefranc et al., 2005c).

Figure 7 shows graphical representations of “IMGT_Collier_de_Perles_for_V_domain” (Lefranc et al., 2003). The five highly conserved amino acids found in IG and TR V domains, whatever the species and molecule type, are highlighted (online in red letters): at position 23 (1st-CYS, or first conserved cysteine C), 41 (CONSERVED-TRP, or conserved tryptophan W), 89 (hydrophobic amino acid, here methionine M), 104 (2nd-CYS, or second conserved cysteine C), and 118 (here J-PHE, or J-REGION tryptophan W). This leafconcept allows, for the first time, to compare domains of IG and TR (V-DOMAIN) and of IgSF proteins other than IG or TR (V-LIKE-DOMAIN), on one layer (facilitating comparison with sequences) or on two layers (bridging comparison with 3D structures).

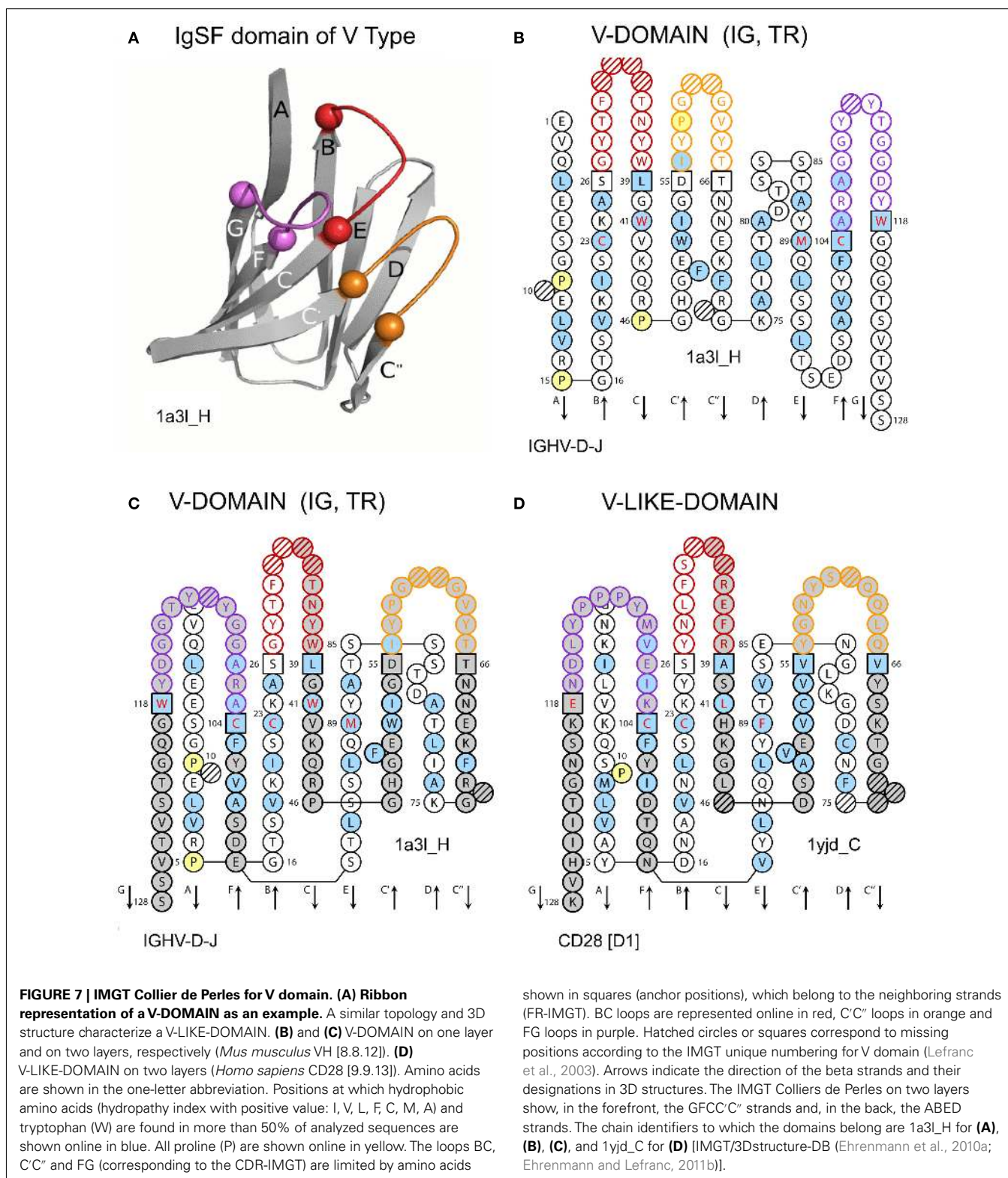
Figure 8 shows graphical representations of “IMGT_Collier_de_Perles_for_G_domain” (Lefranc et al., 2005c). This leafconcept allows, for the first time, to compare domains of the same chain (G-ALPHA1 and G-ALPHA2 of MH1), domains of different chains of the same receptor (G-ALPHA and G-BETA of MH2), or domains of MhSF proteins other than MH (G-ALPHA1-LIKE and G-ALPHA2-LIKE of RPI-MH1Like).

IMGT unique numbering and IMGT Collier de Perles in databases and tools

The IMGT unique numbering and the IMGT Colliers de Perles are used for the numbering of both the codons (in nucleotide sequences) and the amino acids (in protein sequences and structures; Ruiz and Lefranc, 2002; Garapati and Lefranc, 2007; Kaas and Lefranc, 2007; Kaas et al., 2007). By facilitating the comparison of residues between sequences, the IMGT unique numbering and the IMGT Colliers de Perles have been the basis for the description of the IG and TR gene allelic polymorphism and for the studies of IG somatic hypermutations in V-DOMAIN. They represent a major breakthrough for the analysis and the comparison of the huge repertoires of antigen receptors (potentially 2×10^{12} per individual). Indeed, the IMGT unique numbering and the IMGT Colliers de Perles represent a key component in immunogenetics studies by creating a strong and reliable interoperability between the IMGT® databases, tools, and web resources (Lefranc et al., 2009).

Rules for the IMGT unique numbering are implemented in IMGT® online tools: for the analysis of IG and TR rearranged cDNA sequences by IMGT/V-QUEST (Brochet et al., 2008; Giudicelli et al., 2011) and IMGT/JunctionAnalysis (Yousfi Monod et al., 2004; Bleakley et al., 2006; Giudicelli and Lefranc, 2011), for the analysis of cDNA sequences from high-throughput NGS sequencing by IMGT/HighV-QUEST (Alamyar et al., 2012) and for the analysis of amino acid sequences and 2D structures by IMGT/DomainGapAlign (Ehrenmann and Lefranc, 2011a), IMGT/DomainDisplay and IMGT/Collier-de-Perles (Ehrenmann et al., 2011). They are also implemented in IMGT® databases, and particularly in IMGT/3Dstructure-DB (Ehrenmann et al., 2010a; Ehrenmann and Lefranc, 2011b) where they have been fundamental in the setting up of the standardized definition of contact analysis (Kaas and Lefranc, 2005; Kaas et al., 2008; Ehrenmann et al., 2010a) and of paratope and epitope in crystal structures (Lefranc, 2009; Ehrenmann et al., 2010b).

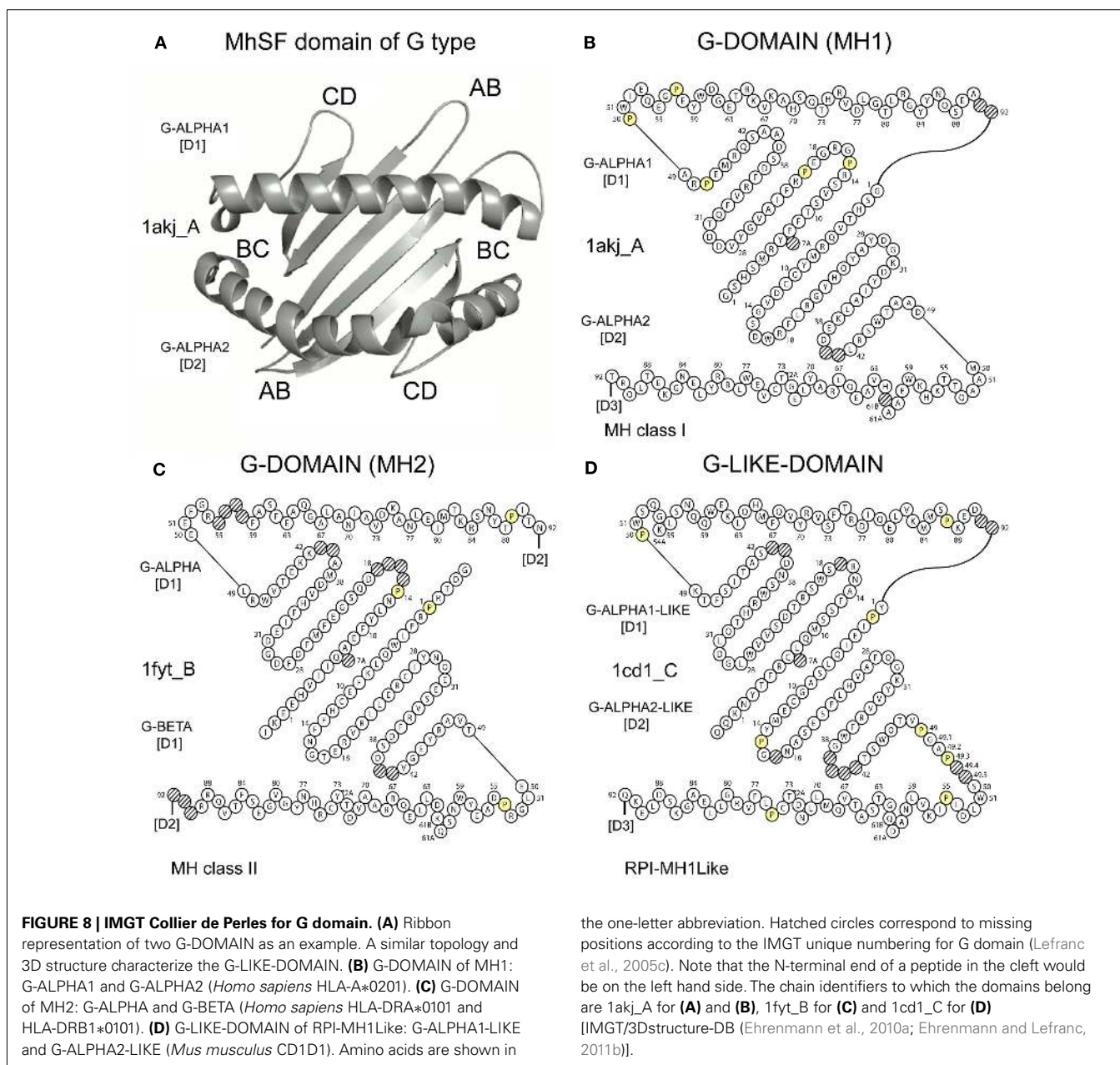
The IMGT Colliers de Perles are particularly useful in molecular engineering and antibody humanization design based on CDR grafting. Indeed they allow to precisely define the CDR-IMGT and to easily compare the amino acid sequences of FR-IMGT and CDR-IMGT between the mouse (or other species) and the closest human V-DOMAIN (Lefranc, 2009; Ehrenmann et al., 2010b). Analyses performed on humanized therapeutic antibodies underline the importance of a correct delimitation of the CDR regions to be grafted (Magdelaine-Beuzelin et al., 2007). The IMGT Colliers de Perles also allow a comparison to the IMGT Colliers de Perles statistical profiles for the human expressed IGHV, IGKV, and IGLV repertoires. These statistical profiles are based on the definition of 11 IMGT amino acid physicochemical characteristics classes which take into account the hydropathy, volume, and chemical characteristics of the 20 common



amino acids (Pommié et al., 2004). This comparison is useful to identify potential immunogenic residues at given positions in chimeric or humanized antibodies or to evaluate immunogenicity of therapeutic antibodies.

DISCUSSION

The standardization, the consistency and the reliability of the immunogenetics data in IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>) rely on



IMGT-ONTOLOGY, elaborated since 1989 in order to manage, to share and to represent the immunogenetics knowledge (Giudicelli and Lefranc, 1999; Lefranc et al., 2004, 2005a; 2008; Duroux et al., 2008; Lefranc, 2011a,b,c,d,e,f, 2013).

IMGT-ONTOLOGY has been developed to be used by any scientific domain which deals with immunogenetics. This includes fundamental, medical, veterinary, clinical, pharmaceutical and biotechnological research. Closely related terms have been integrated in some other biological ontologies (Table 3). Chain types have been included in NCI Thesaurus, Logical Observation Identifier Names and Codes (LOINC), Molecule role (INOH Protein name/family name ontology) (IMR), National Drug File Reference Terminology (NDRFT). IMGT® standardized labels that describe

specifically IG and TR sequences and 3D structures and 64 of the IMGT® standardized labels, in particular those for genomic sequences, have been included in Sequence Ontology (SO; Eilbeck et al., 2005) and in SNP-Ontology. IG and TR gene names were entered in HUGO and NCI Metathesaurus (Table 3). These ontologies are available on the NCBO BioPortal (Noy et al., 2009), opening opportunities of mapping with them.

IMGT® standards derived from IMGT-ONTOLOGY concepts allow interoperability between external databases and tools. Interoperability between IMGT®, HGNC, NCBI, Ensembl, and Vega for the concepts of classification has been described (see Interoperability between IMGT, HGNC, and NCBI). The IMGT numbering is integrated in external Web resources: it is proposed, for

Table 3 | Formal IMGT-ONTOLOGY axioms, IMGT-ONTOLOGY concepts, IMGT® standards, and external resources.

Formal IMGT-ONTOLOGY axioms	IDENTIFICATION	DESCRIPTION	CLASSIFICATION	NUMEROTATION
IMGT-ONTOLOGY concepts ^a	Concepts of identification ^b (/1491)	Concepts of description ^c	Concepts of classification ^d	Concepts of numerotation ^{e,f}
IMGT® standards	IMGT® standardized keywords	IMGT® standardized labels	IMGT® standardized IG and TR gene names	IMGT unique numbering IMGT Colliers de Perles IgBlast
External resources (ontologies, databases, and tools)	NCI Thesaurus (/1032)	Sequence types and features (SO) (/1109)	HUGO (/1528)	
	Logical Observation Identifier Names and Codes (LOINC) (/1350)	SNP-Ontology (/1058)	NCI Metathesaurus (/1499)	
	Molecule role (INOH Protein name/family name ontology) (IMR) (/1029)		HGNC (Bruford et al., 2008)	
	National Drug File Reference Terminology (NDRFT) (/1352)		NCBI gene (Maglott et al., 2011) Ensembl (Hubbard et al., 2002) Vega (Wilming et al., 2008)	

(/number) indicates, for ontologies at NCBO BioPortal, the identifier to be added to <http://bioportal.bioontology.org/ontologies>.

^aGiudicelli and Lefranc (1999), Lefranc et al. (2004, 2005a, 2008), Duroux et al. (2008), Lefranc (2013).

^{b-f}Lefranc (2011b,c,d,e,f).

example, as domain system numbering in the sequence analysis tool IgBlast⁵.

The IMGT® standards generated from IMGT-ONTOLOGY are extensively reused by scientists in very diverse domains for the interpretation of immunogenetics data. The first example is the acknowledgment of the IMGT® gene names as the official nomenclature for IG and TR genes (Wain et al., 2002; Lefranc, 2007, 2008), referenced and recorded in genome sites (NCBI Gene; Maglott et al., 2011). The second example concerns the medical and clinical research which requires a high level of standardization for the results of data analysis in order to take therapeutical decisions: the European Research Initiative on chronic lymphocytic leukemia (CLL) (ERIC) includes 130 laboratories in 26 countries. ERIC has recommended the use of IMGT/V-QUEST (Brochet et al., 2008; Giudicelli et al., 2011), the IMGT® tool for the analysis of IG and TR rearranged sequences, as a reference for determining the rate of IGHV gene mutations, an important prognostic factor for CLL patients (Ghia et al., 2007; Giudicelli and Lefranc, 2008; Langerak et al., 2011). Results provided with the IMGT® standards are integrated in clinical reports (Rosenquist, 2008). The third example is the definition of monoclonal antibodies (mAb, suffix -mab) and fusion proteins for immune applications (FPIA, suffix -cept) of the World Health Organization/International Nonproprietary Name (WHO/INN) programme that are based on the IMGT-ONTOLOGY concepts (Lefranc, 2011g). INN mAb and FPIA have been entered in IMGT/mAb-DB and IMGT/2Dstructure-DB, allowing queries of sequences, 2D structures (or IMGT Collier de Perles) and, if available, 3D structures. The fourth example of great interest

for pharmaceutical companies involved in antibody engineering and humanization for therapeutical use is the characterization of the three hypervariable loops (or CDR-IMGT) of an IG or TR variable domain using the IMGT/DomainGapAlign and IMGT/Collier-de-Perles tools. The objective of antibody humanization is to graft the CDR-IMGT of an antibody, usually murine, and of a given specificity onto a human domain framework, thus preserving the original murine antibody specificity while decreasing its immunogenicity (Lefranc, 2009; Ehrenmann et al., 2010b).

IMGT-ONTOLOGY and IMGT® standards ensure the coherency of the IMGT® information system whose data permanently evolve with the most recent advances in science and methodologies. They form a unique and necessary whole for the modeling, the representation and the sharing of the immunogenetics knowledge by both humans and automated resources.

ACKNOWLEDGMENTS

We are grateful to Gérard Lefranc for helpful comments. We thank the IMGT® team, and all the previous collaborators and biocurators for the expertise and constant motivation. IMGT® is an Institutional Academic Member of the International Medical Informatics Association (IMIA). IMGT® is a registered mark of the Centre National de la Recherche Scientifique (CNRS). IMGT® is certified ISO 9001:2008 and has received the National (CNRS, INSERM, CEA, INRA) Bioinformatics Platform labels: RIO in 2001 and IBiSA in 2007. IMGT® is Bioinformatics Platform of ELIXIR, ReNaBi, GDRACCITH, Cancéropôle GSO, GPTR Sud de France and SFR Biocampus. IMGT® was funded in part by the BIOMED1 (BIOCT930038), Biotechnology BIOTECH2 (BIO4CT960037), 5th PCRDT Quality of Life and Management of Living Resources programmes (QLG2-2000-01287) and 6th

⁵<http://www.ncbi.nlm.nih.gov/igblast/>

PCRD Information Society Technology (ImmunoGrid, IST-2004-028069) programmes of the European Union (EU). IMGT® was granted access by GENCI to the CINES HPC resources (2010-036029). IMGT® is currently supported by the Ministère

de l'Enseignement Supérieur et de la Recherche (MESR), CNRS, Université Montpellier 2, Région Languedoc-Roussillon, Agence Nationale de la Recherche ANR (BIOSYS-06-135457, FLAVORES), and the Labex MabImprove (2011–2020).

REFERENCES

- Alamyar, E., Giudicelli, V., Shuo, L., Duroux, P., and Lefranc, M.-P. (2012). 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.
- Ashurst, J. L., Chen, C.-K., Gilbert, J. G. R., Jekosch, K., Keenan, S., Meidl, P., Searle, S. M., Stalker, J., Storey, R., Trevanion, S., Wilming, L., and Hubbard, T. (2005). The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* 33, D459–465.
- Bairoch, A., Bougueleret, L., and UniProt Consortium. (2009). The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* 37, D169–D174.
- Bleakley, K., Giudicelli, V., Wu, Y., Lefranc, M.-P., and Biau, G. (2006). IMGT standardization for statistical analyses of T cell receptor junctions: the TRAV-TRAJ example. *In silico Biol. (Gedrukt)* 6, 573–588.
- Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D. G., Knight, J., Mani, P., Martin, R., Moxon, S. A. T., Paddock, H., Pich, C., Ramachandran, S., Ruef, B. J., Ruzicka, L., Bauer Schaper, H., Schaper, K., Shao, X., Singer, A., Sprague, J., Sprunger, B., Van Slyke, C., and Westerfield, M. (2011). ZFIN: enhancements and updates to the zebrafish model organism database. *Nucleic Acids Res.* 39, D822–D829.
- Brochet, X., Lefranc, M.-P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* 36, W503–W508.
- Bruford, E. A., Lush, M. J., Wright, M. W., Sneddon, T. P., Povey, S., and Birney, E. (2008). The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.* 36, D445–D448.
- Duroux, P., Kaas, Q., Brochet, X., Lane, J., Ginestoux, C., Lefranc, M.-P., and Giudicelli, V. (2008). IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. *Biochimie* 90, 570–583.
- Ehrenmann, F., Giudicelli, V., Duroux, P., and Lefranc, M.-P. (2011). IMGT/Collier de Perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). *Cold Spring Harb. Protoc.* 2011, 726–736.
- Ehrenmann, F., Kaas, Q., and Lefranc, M.-P. (2010a). IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhSF. *Nucleic Acids Res.* 38, D301–D307.
- Ehrenmann, F., Duroux, P., Giudicelli, V., and Lefranc, M.-P. (2010b). “Standardized sequence and structure analysis of antibody using IMGT®,” in *Antibody Engineering*, Vol. 2, eds R. Kontermann and S. Dübel (Berlin: Springer-Verlag), 11–31.
- Ehrenmann, F., and Lefranc, M.-P. (2011a). IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF). *Cold Spring Harb. Protoc.* 2011, 737–749.
- Ehrenmann, F., and Lefranc, M.-P. (2011b). IMGT/3Dstructure-DB: querying the IMGT database for 3D structures in immunology and immunoinformatics (IG or antibodies, TR, MH, RPI, and FPIA). *Cold Spring Harb. Protoc.* 2011, 750–761.
- Elibeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., Durbin, R., and Ashburner, M. (2005). The sequence ontology: a tool for the unification of genome annotations. *Genome Biol.* 6, R44.
- Garapati, V. P., and Lefranc, M.-P. (2007). IMGT Colliers de Perles and IgSF domain standardization for T cell costimulatory activatory (CD28, ICOS) and inhibitory (CTLA4, PDCD1 and BTLA) receptors. *Dev. Comp. Immunol.* 31, 1050–1072.
- Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C., Stiglenbauer, S., Stevenson, F., Davi, F., and Rosenquist, R. (2007). ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia* 21, 1–3.
- Giudicelli, V., Brochet, X., and Lefranc, M.-P. (2011). IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harb. Protoc.* 2011, 695–715.
- Giudicelli, V., Chaume, D., Jabado-Michaloud, J., and Lefranc, M.-P. (2005a). Immunogenetics sequence annotation: the strategy of IMGT based on IMGT-ONTOLOGY. *Stud. Health Technol. Inform.* 116, 3–8.
- Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2005b). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* 33, D256–D261.
- Giudicelli, V., Duroux, P., Ginestoux, C., Folch, G., Jabado-Michaloud, J., Chaume, D., and Lefranc, M.-P. (2006). IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res.* 34, D781–D784.
- Giudicelli, V., and Lefranc, M.-P. (1999). Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* 15, 1047–1054.
- Giudicelli, V., and Lefranc, M.-P. (2008). “IMGT® standardized analysis of immunoglobulin rearranged sequences,” in *Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia*, eds P. Ghia, R. Rosenquist, and F. Davi (Milan: Wolters Kluwer Health), 33–52.
- Giudicelli, V., and Lefranc, M.-P. (2011). IMGT/JunctionAnalysis: IMGT standardized analysis of the V-J and V-D-J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb. Protoc.* 2011, 716–725.
- Greenaway, H. Y., Kurniawan, M., Price, D. A., Douek, D. C., Davenport, M. P., and Venturi, V. (2009). Extraction and characterization of the rhesus macaque T-cell receptor beta-chain genes. *Immunol. Cell Biol.* 87, 546–553.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquis.* 5, 199–220.
- Guarino, N. (1997). Understanding, building and using ontologies. *Int. J. Hum. Comput. Stud.* 46, 293–310.
- Guarino, N., and Giarretta, P. (1995). “Ontologies and knowledge bases: towards a terminological clarification,” in *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, ed. N. Mars (Amsterdam, NL: IOS Press), 25–32.
- Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyraas, E., Gilbert, J., Hammond, M., Humniecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Res.* 30, 38–41.
- Kaas, Q., Duprat, E., Tourneur, G., and Lefranc, M.-P. (2008). “IMGT standardization for molecular characterization of the T cell receptor/peptide/MHC complexes,” in *Immunoinformatics, Immunomics Reviews, Series of Springer Science and Business Media LLC*, eds C. Schoenbach, S. Ranganathan, and V. Brusci (New York: Springer), 19–49.
- Kaas, Q., Ehrenmann, F., and Lefranc, M.-P. (2007). IG, TR and IgSF, MHC and MhSF: what do we learn from the IMGT Colliers de Perles? *Brief. Funct. Genomic. Proteomic.* 6, 253–264.
- Kaas, Q., and Lefranc, M.-P. (2005). T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In silico Biol. (Gedrukt)* 5, 505–528.
- Kaas, Q., and Lefranc, M.-P. (2007). IMGT Colliers de Perles: standardized sequence-structure representations of the IgSF and MhSF superfamily domains. *Curr. Bioinform.* 2, 21–30.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, J., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grahham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer,

- S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weisenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramsay, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nord-siek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Hausler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrino, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lane, J., Duroux, P., and Lefranc, M.-P. (2010). From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences. *BMC Bioinformatics* 11, 223. doi:10.1186/1471-2105-11-223
- Langerak, A. W., Davi, F., Ghia, P., Hadzidimitriou, A., Murray, F., Potter, K. N., Rosenquist, R., Stamatopoulos, K., and Belessi, C. (2011). Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia* 25, 979–984.
- Lefranc, M. P. (1997). Unique database numbering system for immunogenetic analysis. *Immunol. Today* 18, 509.
- Lefranc, M. P. (1999). The IMGT unique numbering for Immunoglobulins, T cell receptors and Ig-like domains. *Immunologist* 7, 132–136.
- Lefranc, M.-P. (2000a). “Nomenclature of the human immunoglobulin genes,” in *Current Protocols in Immunology*, eds J. E. Coligan, B. E. Bierer, D. E. Margulies, E. M. Shevach, and W. Strober (Hoboken, NJ: John Wiley and Sons, Inc), A.1P.1–A.1P.37.
- Lefranc, M.-P. (2000b). “Nomenclature of the human T cell receptor genes,” in *Current Protocols in Immunology*, eds J. E. Coligan, B. E. Bierer, D. E. Margulies, E. M. Shevach, and W. Strober (Hoboken, NJ: John Wiley and Sons, Inc), A.1O.1–A.1O.23.
- Lefranc, M.-P. (2005). IMGT, the international ImMunoGeneTics information system: a standardized approach for immunogenetics and immunoinformatics. *Immunome Res.* 1, 3.
- Lefranc, M.-P. (2007). WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report. *Immunogenetics* 59, 899–902.
- Lefranc, M.-P. (2008). WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev. Comp. Immunol.* 32, 461–463.
- Lefranc, M.-P. (2009). “Antibody databases and tools: The IMGT® experience,” in *Therapeutic Monoclonal Antibodies: From Bench to Clinic*, ed. Z. An (Hoboken, NJ: John Wiley and Sons, Inc), 91–114.
- Lefranc, M.-P. (2011a). IMGT, the international ImMunoGeneTics information system. *Cold Spring Harb. Protoc.* 2011, 595–603.
- Lefranc, M.-P. (2011b). From IMGT-ONTOLOGY IDENTIFICATION axiom to IMGT standardized keywords: for immunoglobulins (IG), T cell receptors (TR), and conventional genes. *Cold Spring Harb. Protoc.* 2011, 604–613.
- Lefranc, M.-P. (2011c). From IMGT-ONTOLOGY DESCRIPTION axiom to IMGT standardized labels: for immunoglobulin (IG) and T cell receptor (TR) sequences and structures. *Cold Spring Harb. Protoc.* 2011, 614–626.
- Lefranc, M.-P. (2011d). From IMGT-ONTOLOGY CLASSIFICATION Axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harb. Protoc.* 2011, 627–632.
- Lefranc, M.-P. (2011e). IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of, IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb. Protoc.* 2011, 633–642.
- Lefranc, M.-P. (2011f). IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harb. Protoc.* 2011, 643–651.
- Lefranc, M.-P. (2011g). Antibody nomenclature: from IMGT-ONTOLOGY to INN definition. *MAbs* 3, 1–2.
- Lefranc, M.-P. (2013). “IMGT-ONTOLOGY,” in *Encyclopedia of Systems Biology*, eds W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota (New York: Springer) (in press).
- Lefranc, M.-P., Clement, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D., and Lefranc, G. (2005a). IMGT-Choreography for immunogenetics and immunoinformatics. *In silico Biol. (Gedrukt)* 5, 45–60.
- Lefranc, M.-P., Pommier, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piédade, I., Rouard, M., Foulquier, E., Thouvenin, V., and Lefranc, G. (2005b). IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* 29, 185–203.
- Lefranc, M.-P., Duprat, E., Kaas, Q., Tranne, M., Thiriot, A., and Lefranc, G. (2005c). IMGT unique numbering for MHC groove G-DOMAIN and MHC superfamily (MhcSF) G-LIKE-DOMAIN. *Dev. Comp. Immunol.* 29, 917–938.
- Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Bosc, N., Folch, G., Guiraudou, D., Jabado-Michaloud, J., Magris, S., Scaviner, D., Thouvenin, V., Combres, K., Girod, D., Jeanjean, S., Prétat, C., Yousfi-Monod, M., Duprat, E., Kaas, Q., Pommier, C., Chaume, D., and Lefranc, G. (2004). IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In silico Biol. (Gedrukt)* 4, 17–29.
- Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G., and Duroux, P. (2009). IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res.* 37, D1006–D1012.
- Lefranc, M.-P., Giudicelli, V., Regnier, L., and Duroux, P. (2008). IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. *Brief. Bioinform.* 9, 263–275.
- Lefranc, M.-P., and Lefranc, G. (2001a). *The Immunoglobulin FactsBook*. London: Academic Press.
- Lefranc, M.-P., and Lefranc, G. (2001b). *The T Cell Receptor FactsBook*. London: Academic Press.
- Lefranc, M.-P., Pommier, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., and Lefranc, G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* 27, 55–77.
- Letovsky, S. I., Cottingham, R. W., Porter, C. J., and Li, P. W. (1998). GDB: The Human Genome Database. *Nucleic Acids Res.* 26, 94–99.
- Magdelaine-Beuzelin, C., Kaas, Q., Wehbi, V., Ohresser, M., Jefferis, R., Lefranc, M.-P., and Watier, H. (2007). Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Crit. Rev. Oncol. Hematol.* 64, 210–225.
- Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2007). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 35, D26–D31.
- Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2011). Entrez Gene:

- gene-centered information at NCBI. *Nucleic Acids Res.* 39, D52–D57.
- Maglott, D. R., Katz, K. S., Sicotte, H., and Pruitt, K. D. (2000). NCBI's LocusLink and RefSeq. *Nucleic Acids Res.* 28, 126–128.
- Musen, M. A., Noy, N. F., Shah, N. H., Whetzel, P. L., Chute, C. G., Story, M.-A., Smith, B., and the NCBO team. (2012). The National Center for Biomedical Ontology. *J. Am. Med. Inform. Assoc.* 19, 190–195.
- Noy, N. F., Crubezy, M., Ferguson, R. W., Knublauch, H., Tu, S. W., Vendetti, J., and Musen, M. A. (2003). Protégé-2000: an open-source ontology-development and knowledge-acquisition environment. *AMIA Annu. Symp. Proc.* 953.
- Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D. L., Storey, M.-A., Chute, C. G., and Musen, M. A. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.* 37, W170–W173.
- Pappalardo, F., Lefranc, M.-P., Lollini, P.-L., and Motta, S. (2010). A novel paradigm for cell and molecule interaction ontology: from the CMM model to IMGT-ONTOLOGY. *Immunome Res.* 6, 1.
- Pommié, C., Levadoux, S., Sabatier, R., Lefranc, G., and Lefranc, M.-P. (2004). IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. *J. Mol. Recognit.* 17, 17–32.
- Prabakaran, P., Chen, W., Singarayan, M. G., Stewart, C. C., Streaker, E., Feng, Y., and Dimitrov, D. S. (2012). Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics* 64, 337–350.
- Rosenquist, R. (2008). “How to report IG sequence data in clinical routine: cases difficult to categorize,” in *Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia*, eds P. Ghia, R. Rosenquist, and F. Davi (Milan: Wolters Kluwer Health), 113–124.
- Ruiz, M., and Lefranc, M.-P. (2002). IMGT gene identification and Colliers de Perles of human immunoglobulins with known 3D structures. *Immunogenetics* 53, 857–883.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huseon, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Bid-dick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabriellian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferreira, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science* 291, 1304–1351.
- Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W., and Povey, S. (2002). Guidelines for human gene nomenclature. *Genomics* 79, 464–470.
- Wain, H. M., Lush, M. J., Ducluzeau, F., Khodiyar, V. K., and Povey, S. (2004). Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Res.* 32, D255–D257.
- Wilming, L. G., Gilbert, J. G. R., Howe, K., Trevanion, S., Hubbard, T., and Harrow, J. L. (2008). The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* 36, D753–D760.
- Yousfi Monod, M., Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2004). IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics* 20(Suppl. 1), i379–i385.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 February 2012; accepted: 24 April 2012; published online: 23 May 2012.

Citation: Giudicelli V and Lefranc M-P (2012) IMGT-ONTOLOGY 2012. *Front. Gene.* 3:79. doi: 10.3389/fgene.2012.00079

This article was submitted to *Frontiers in Bioinformatics and Computational Biology*, a specialty of *Frontiers in Genetics*. Copyright © 2012 Giudicelli and Lefranc. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.