

**2010- 56<sup>th</sup> IUIS Council Meeting, August 21, 2010**, Kobe, Japan  
at the 14th International Congress of Immunology (ICI), Kobe, Japan

## WHO-IUIS Nomenclature Subcommittee for Immunoglobulins (IG) and T cell receptors (TR) report

### **1. The WHO-IUIS Nomenclature Subcommittee**

*Members of the WHO-IUIS Nomenclature Subcommittee/IMGT Nomenclature (IMGT-NC) Subcommittee for immunoglobulins and T cell receptors: Donald Capra (USA), Max Cooper (USA), Tasuku Honjo (Japan), Leroy Hood (USA), Gérard Lefranc (France), Marie-Paule Lefranc (France), Fumihiko Matsuda (Japan), Hans Zachau (Germany).*

Experts 2008-2010: Cynthia L. Baldwin (USA), Eva Bengtén (USA), Pierre Boudinot (France), Salvatrice Ciccarese (Italy), Deborah Dunn-Walters (UK), Jean-Pol Fripiat (France), Evelyne Jouvin-Marche (France), Azad K. Kaushik (Canada), Serge Muyldermans (Belgium), Bettina Wagner (USA).

*Standing Committee for Immunoglobulins and T cell receptors: Marie-Paule Lefranc, IMGT, Laboratoire d'ImmunoGénétique Moléculaire, Université Montpellier 2 and UPR CNRS, Institut de Génétique Humaine, Montpellier, France, Marie-Paule.Lefranc@igh.cnrs.fr.*

Since the creation of IMGT®, the international ImMunoGeneTics information system®, <http://www.imgt.org>, in 1989, at New Haven during the 10th Human Genome Mapping Workshop (HGM10), the standardized classification and nomenclature of the immunoglobulins and T cell receptors of human and other vertebrate species have been under the responsibility of the IMGT Nomenclature Committee (IMGT-NC). In 1995, following the first demonstration online of the nucleotide database IMGT/LIGM-DB at the 9th International Congress of Immunology in San Francisco, the IMGT-NC Subcommittee for immunoglobulins and T cell receptors has become the WHO-IUIS Nomenclature Subcommittee.

The WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors works in close collaboration with the IMGT-NC Subcommittee for the immunoglobulin superfamily (IgSF) and the major histocompatibility complex superfamily (MhcSF), the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC), the Mouse Genomic Nomenclature Committee (MGNC), the Nomenclature Committees of newly sequenced genomes, the national and international Immunology, Immunogenetics and Genetics Societies, the editors and publishers for recommendations to Authors.

### **21st Anniversary of IMGT® and 11th Anniversary of the approval of the IMGT immunoglobulin (IG) and T cell receptor (TR) gene names**

In May 2010, a presentation was given by Marie-Paule Lefranc and Véronique Giudicelli, at the 14th Human Genome Meeting HGM 2010, Next Generation Genomics and Medicine, Montpellier (France), organized by the Human Genome Organisation (HUGO), 18-21 May 2010. This presentation highlighted the 21st Anniversary of IMGT®, the international

ImMunoGeneTics information system® (<http://www.imgt.org>) and the 11th Anniversary of the approval of the IMGT immunoglobulin (IG) and T cell receptor (TR) gene names as official international nomenclature by the HUGO nomenclature committee (HGNC).

### **From HGM10 (1989) to HGM 2010: IG and TR gene concept and IMGT/GENE-DB**

In 1989, at HGM10 in New Haven (USA), the T cell receptor gamma genes were, for the first time, entered in the newly created Genome DataBase (GDB). This was a first major step as it acknowledged the concept of gene for the antigen receptors, the immunoglobulins (IG) and T cell receptors (TR), despite their unique particularities. Indeed, IG and TR chains are coded by genes belonging to four types, variable (V), diversity (D), joining (J) and constant (C), which show an unusual molecular organization due to the necessity of DNA rearrangements for the IG and TR chain synthesis, in B and T cells, respectively. These rearrangements contribute to the huge diversity and fine specificity of the variable domains of the IG and TR that bind specifically to the antigens, in the adaptive immune response. The potential expressed repertoire is estimated to  $2 \times 10^{12}$  different IG and TR per individual. Owing to these particularities, IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>), was created in 1989 by Marie-Paule Lefranc at LIGM, Montpellier, France. IMGT® is a high-quality integrated knowledge resource specialized in the IG, TR, major histocompatibility complex (MHC), immunoglobulin superfamily (IgSF), MHC superfamily (MhcSF) and related proteins of the immune system (RPI) of human and other vertebrate species. IMGT/GENE-DB (*Nucl. Acids Res.*, 33, D256-D261, 2005), the IMGT® genome database, was developed to standardize and classify the IG and TR gene data and to manage the related knowledge. The official nomenclature of human IG and TR genes and alleles, based on IMGT-ONTOLOGY, the first ontology for immunogenetics, was approved in 1999 by the HUGO Nomenclature Committee (HGNC) and acknowledged by the WHO-IUIS. Rules for the identification of gene and allele functionality were defined, with each IG and TR gene and allele being represented by an IMGT reference sequence. Another breakthrough is the IMGT unique numbering and its graphical representation, the IMGT Collier de Perles, which allow the standardization per domain type. In March 2010, IMGT/GENE-DB includes the 674 human IG and TR genes (1245 alleles). Five hundred ninety-six genes (1139 alleles) are organized in 7 loci on 4 chromosomes, spanning a total of 6 megabases: IGH (14q32.33), IGK (2p11.2), IGL (22q11.2), TRA (14q11.2), TRD (14q11.2), TRB (7q34) and TRG (7p14) (*The Immunoglobulin FactsBook*, 2001; *The T cell receptor FactsBook*, 2001). Seventy-eight orphans (106 alleles) are found outside the main loci. IMGT® gene data are provided on Ensembl Genome Browser (EBI) via a DAS server. IMGT/GENE-DB gene entries are cross-referenced by HGNC database, Genatlas, Entrez Gene (NCBI) and Vega (Wellcome Trust Sanger Institute). IMGT/GENE-DB reference sequences are crucial for the assignment of new alleles of IG and TR from different haplotypes (1,000 genomes project), for gene expression studies in normal and pathologic situations (cDNA high-throughput sequencing) and for biotechnology related to antibody engineering and antibody humanization.

### **2. Requirements for the assignment of new alleles for variable (V), diversity (D) and joining (J) genes of immunoglobulin (IG) and T cell receptor (TR) genes.**

The assignment of alleles for variable (V), diversity (D) and joining (J) genes of immunoglobulin (IG) and T cell receptor (TR) genes has been managed by the IMGT Nomenclature committee, since the creation of IMGT® in 1989 (*The Immunoglobulin*

*FactsBook 2001, The T cell receptor FactsBook, 2001*). The allele assignments are based on the IMGT Scientific chart rules and concepts of classification of IMGT-ONTOLOGY.

IMGT Index>Allele

<http://www.imgt.org/textes/IMGTindex/allele.html>

(created 03/12/1999)

IMGT allele nomenclature for sequence polymorphisms

<http://www.imgt.org/textes/IMGTScientificChart/Nomenclature/IMGTallelepolymorphism.html>

(created: 04/05/1998).

As the number of sequences increases, the WHO-IUIS Nomenclature Subcommittee for immunoglobulins (IG) and T cell receptors (TR) added an additional paragraph on the IMGT web site reminding the requirements for a new allele to be approved and entered in the IMGT® databases and tools.

The requirements are the following:

1. sequences of new alleles should be from **genomic DNA** and from V, D and J genes in **germline** configuration.
2. sequences of new alleles should include the **complete gene unit**, that is:
  - L-V-GENE-UNIT (or exceptionally, and depending on approval, V-GENE-UNIT)
  - D-GENE-UNIT
  - J-GENE-UNIT.

The definitions of the gene units are given in Table 1.

Table 1. Definition of the V, D and J gene units.

IMGT label name	Definition
D-GENE-UNIT	germline genomic DNA including 5'D-RS, D-REGION and 3'D-RS
J-GENE-UNIT	germline genomic DNA including 5'J-RS and J-REGION
L-V-GENE-UNIT	germline genomic DNA including L-PART1, V-INTRON, V-EXON and V-RS
V-GENE-UNIT	germline genomic DNA including V-EXON and V-RS

IMGT label names and their definition are available at: <http://www.imgt.org/cgi-bin/IMGTlect.jv?query=7>

3. sequences should be submitted to GenBank, EMBL-Bank or DDBJ.

Allele numbers are provided on a chronological basis, and as the IMGT nomenclature committee (IMGT-NC) becomes aware of new alleles.

Sequences are entered in IMGT/LIGM-DB only when sequences become publicly available. Therefore authors who would like to keep the anteriority of new alleles, should contact the IMGT nomenclature committee with the submitted sequences and accession numbers.

The IG and TR genes and alleles are managed in the IMGT/GENE-DB database which allows a search per gene name, based on the concept of classification of IMGT-ONTOLOGY, generated from the CLASSIFICATION axiom, and which includes links to IMGT/LIGM-DB, IMGT Repertoire and external databases. Links on individual entries between IMGT®, Entrez Gene and GeneCards have been established. In August 2008, IMGT/GENE-DB contains 1645 genes (673 from human and 972 from mouse) and 2598 alleles (1217 from human and 1381 from mouse). It comprises genes and alleles for the seven loci (IGH, IGK, IGL, TRA, TRB, TRG and TRD). The IG and/or TR genes from 50 other species are currently available in the IMGT Repertoire in “Gene tables” and in “Alignment of alleles”. For genes not yet localized in the loci or for incomplete loci, a provisional nomenclature has been assigned. With the completion of the genomes and the sequence annotation, these genes will be progressively entered in IMGT/GENE-DB.

### ***IMGT/GENE-DB global statistics per species***

<b>Species</b>	<b>Number of IG and TR genes</b>	<b>Number of IG and TR alleles</b>
<i>Bos taurus</i>	135	135
<i>Canis lupus familiaris</i>	83	83
<i>Danio rerio</i>	58	58
<i>Homo sapiens</i>	684	1254
<i>Mus cookii</i>	1	1
<i>Mus minutoides</i>	2	2
<i>Mus musculus</i>	811	1240
<i>Mus pahari</i>	3	3
<i>Mus saxicola</i>	1	1
<i>Mus spretus</i>	14	17
<i>Oncorhynchus mykiss</i>	130	132
<i>Oryctolagus cuniculus</i>	185	224
<i>Rattus norvegicus</i>	560	564
<i>Rattus rattus</i>	1	1
<b>Total</b>	<b>2668</b>	<b>3715</b>

### ***IMGT/GENE-DB statistics for human***

#### **Human IG**

<b>Numbers</b>	<b>IGH</b>				<b>IGK</b>			<b>IGL</b>			<b>PREBV</b>	<b>Total</b>
	<b>IGHV</b>	<b>IGHD</b>	<b>IGHJ</b>	<b>IGHC</b>	<b>IGKV</b>	<b>IGKJ</b>	<b>IGKC</b>	<b>IGLV</b>	<b>IGLJ</b>	<b>IGLC</b>		
<b>Genes</b>	164	37	9	12	108	5	1	79	11	14	0	<b>440</b>

<b>Alleles</b>	397	44	17	52	135	9	5	129	14	28	0	<b>830</b>
----------------	-----	----	----	----	-----	---	---	-----	----	----	---	------------

<b>Numbers</b>	<b>TRA</b>			<b>TRB</b>				<b>TRG</b>			<b>TRD</b>				<b>Total</b>
	<b>TRAV</b>	<b>TRAJ</b>	<b>TRAC</b>	<b>TRBV</b>	<b>TRBD</b>	<b>TRBJ</b>	<b>TRBC</b>	<b>TRGV</b>	<b>TRGJ</b>	<b>TRGC</b>	<b>TRDV</b>	<b>TRDD</b>	<b>TRDJ</b>	<b>TRDC</b>	
<b>Genes</b>	54	61	1	76	2	14	2	15	5	2	3	3	4	1	<b>243</b>
<b>Alleles</b>	112	68	1	162	3	16	4	23	6	13	6	3	4	1	<b>422</b>

## Human TR

### *IMGT/GENE-DB statistics for mouse*

## Mouse IG

<b>Numbers</b>	<b>IGH</b>				<b>IGK</b>			<b>IGL</b>			<b>Total</b>
	<b>IGHV</b>	<b>IGHD</b>	<b>IGHJ</b>	<b>IGHC</b>	<b>IGKV</b>	<b>IGKJ</b>	<b>IGKC</b>	<b>IGLV</b>	<b>IGLJ</b>	<b>IGLC</b>	
<b>Genes</b>	332	30	4	9	177	5	6	12	7	7	<b>589</b>
<b>Alleles</b>	485	35	9	26	205	10	8	19	7	10	<b>814</b>

## Mouse TR

<b>Numbers</b>	<b>TRA</b>			<b>TRB</b>				<b>TRG</b>			<b>TRD</b>				<b>Total</b>
	<b>TRAV</b>	<b>TRAJ</b>	<b>TRAC</b>	<b>TRBV</b>	<b>TRBD</b>	<b>TRBJ</b>	<b>TRBC</b>	<b>TRGV</b>	<b>TRGJ</b>	<b>TRGC</b>	<b>TRDV</b>	<b>TRDD</b>	<b>TRDJ</b>	<b>TRDC</b>	
<b>Genes</b>	98	60	1	35	2	14	7	7	4	4	6	2	2	1	<b>243</b>
<b>Alleles</b>	234	70	2	61	2	19	9	28	4	5	10	2	3	1	<b>450</b>

### *IMGT/GENE-DB statistics for rat*

## Rat IG

<b>Numbers</b>	<b>IGH</b>				<b>IGK</b>			<b>IGL</b>			<b>Total</b>
	<b>IGHV</b>	<b>IGHD</b>	<b>IGHJ</b>	<b>IGHC</b>	<b>IGKV</b>	<b>IGKJ</b>	<b>IGKC</b>	<b>IGLV</b>	<b>IGLJ</b>	<b>IGLC</b>	
<b>Genes</b>	342	20	4	8	163	7	1	7	4	4	<b>560</b>
<b>Alleles</b>	342	20	4	11	163	7	2	7	4	4	<b>564</b>