

IMGT® biocuration of IG and TR in IMGT/LIGM-DB and IMGT/GENE-DB

Géraldine Folch*, Joumana Michaloud*, Safa Aouinti, Mélissa Cambon, Imène Chentli, Saida Hadi-Saljoqi, Karthik Kalyan, Anjana Kushwaha, Arthur Lavoie, Claudio Lorenzi, Perrine Pégorier, Laurène Picandet, Patrice Duroux, Véronique Giudicelli, Sofia Kossida and Marie-Paule Lefranc

*Equal contribution

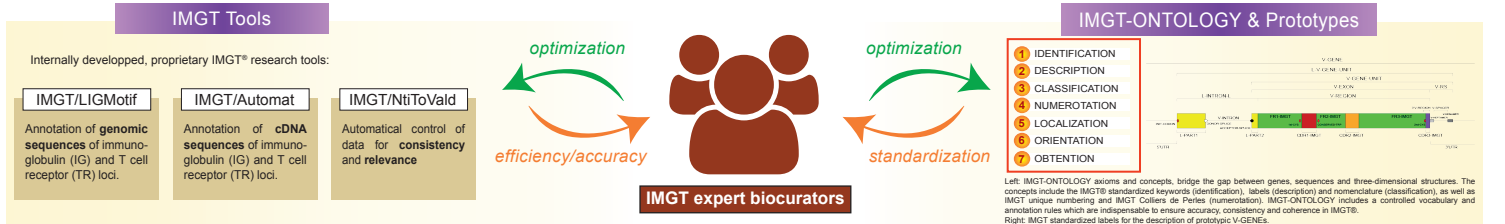
IMGT®, the international ImMunoGeneTics information system®
 Université de Montpellier (UM), Laboratoire d'Immunogénétique Moléculaire (LIGM),
 Institut de Génétique Humaine (IGH), UMR 9002 CNRS-UM, Montpellier (France)



IMGT®, the international ImMunoGeneTics information system®, <http://www.imgt.org>, is the global reference in immunogenetics and immunoinformatics [1]. By managing the extreme diversity and complexity of the antigen receptors of the adaptive immune response, the immunoglobulins (IG) or antibodies and the T cell receptors (TR) [2,3] (2.10¹² different specificities per individual), IMGT® is at the origin of immunoinformatics, a science at the interface between immunogenetics and bioinformatics [4]. IMGT® is based on the concepts of IMGT-ONTOLOGY [5] and these concepts are used for expert annotation and standardized knowledge in IMGT/LIGM-DB, the IMGT® database of IG and TR nucleotide sequences from human and other vertebrate species and in IMGT/GENE-DB, the IMGT® gene and allele database. The IMGT/LIGM-DB biocuration pipeline of IG and TR sequences includes IMGT/LIGMOTif, for the analysis of large genomic DNA sequences, and IMGT/Automat, for the automatic annotation of rearranged cDNA sequences. Analysis results are checked for consistency, both manually and by using IMGT® tools (IMGT/NtiToVald, IMGT/V-QUEST, IMGT/BLAST, etc.). The annotated sequences are integrated in IMGT/LIGM-DB and include the sequence identification (IMGT® keywords), the gene and allele classification (IMGT® nomenclature), the constitutive and specific motif description (IMGT® labels in capital letters, no plural), the translation of the coding regions (IMGT® unique numbering) [4,5]. For genomic IMGT/LIGM-DB sequences containing either an IG or TR variable (V), diversity (D) or joining (J) gene in germline configuration or a constant (C) gene, the gene and allele information is entered in IMGT/GENE-DB. In parallel, the IMGT® Repertoire is updated (Locus representations, Gene tables and Protein displays (for new genes), Alignments of alleles (for new and/or confirmatory alleles) and the IMGT® reference directory [1,4] is completed (sequences used for gene and allele comparison and assignment in IMGT® tools (IMGT/V-QUEST, IMGT/HighV-QUEST for next generation sequencing (NGS), IMGT/DomainCapAlign) and databases (IMGT/2Dstructure-DB, IMGT/3Dstructure-DB). An IMGT/GENE-DB entry also provides information on the rearranged cDNA and gDNA entries (with links to IMGT/LIGM-DB) and on the three-dimensional structures (with links to IMGT/3Dstructure-DB). IMGT/GENE-DB is the official repository of IG and TR genes and alleles. IMGT® gene names were approved by HGNC and endorsed by WHO-IUIS, the World Health Organization (WHO)-International Union of Immunological Societies (IUIS) Nomenclature Subcommittee for IG and TR. Reciprocal links exist between IMGT/GENE-DB and HGNC, NCBI and Vega. The definition of antibodies published by the WHO International Nonproprietary Name (INN) Programme is based on the IMGT® concepts [6], and allows easy retrieval via IMGT/mAb-DB query [1,4]. The IMGT® standardized annotation has allowed to bridge the gaps for IG or antibodies and TR between fundamental and medical research, veterinary research, repertoire analysis, biotechnology related to antibody engineering, diagnostics and therapeutical approaches.

- [1] Lefranc M-P et al. Nucleic Acids Res 43:413-422 (2015) PMID: 25378316.
- [3] Lefranc M-P, Lefranc G. The T cell receptor FactsBook (2001).
- [5] Giudicelli V, Lefranc, M-P. Front Genet 3:79 (2012) PMID: 22654892.
- [2] Lefranc M-P, Lefranc G. The Immunoglobulin FactsBook (2001).
- [4] Lefranc M-P. Front Immunol 5:22 (2014) PMID: 24600447.
- [6] Lefranc M-P. mAbs 3(1)-2 (2011) PMID: 21093347.

IMGT® Expert Biocuration Pipeline



IMGT/LIGM-DB

IMGT/LIGM-DB includes all germline (non-rearranged) and rearranged IG and TR genomic DNA and complementary DNA sequences published in generalist databases. IMGT/LIGM-DB allows searches from the Web interface according to biological and immunogenetic criteria. For a given entry, nine types of display are available including the IMGT flat file, the translation of the coding regions and the analysis by the IMGT-V-QUEST tool. The annotations hugely enhance the quality and the accuracy of the distributed detailed information.

177 115 sequences
351 species

IMGT/GENE-DB

IMGT/GENE-DB Query Page allows the search of IG/TR genes according to IMGT-ONTOLOGY's seven axioms. IMGT/GENE-DB entry displays accurate gene data related to genome (gene localization), allelic polymorphisms (number of alleles, IMGT reference sequences, functionality, etc.) gene expression (known cDNAs) and proteins structures (IMGT Collers de Perles, IMGT/3Dstructure-DB). It provides internal links to the IMGT sequence databases and the IMGT Web Resources as well as external links to genome and generalist sequence databases.

3 927 genes
5630 alleles
24 species

Web Resources

1 IMGT flat file

2 Gene table

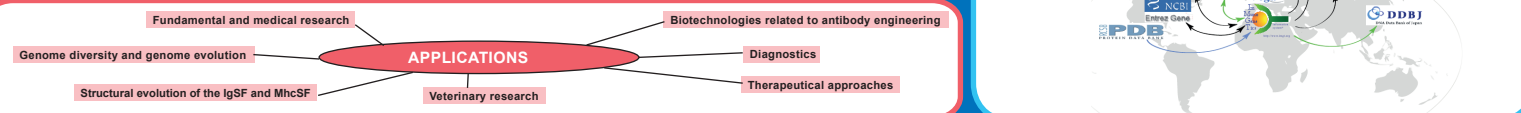
3 Locus representation

4 Chromosomal localization

INTEROPERABILITY

WEB INTERFACE

WEB RESOURCES



IMGT® founder and executive director emeritus: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)
 IMGT® director: Sofia Kossida (Sofia.Kossida@igh.cnrs.fr)
 Bioinformatics manager: Véronique Giudicelli (Veronique.Guindicelli@igh.cnrs.fr)
 Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)

