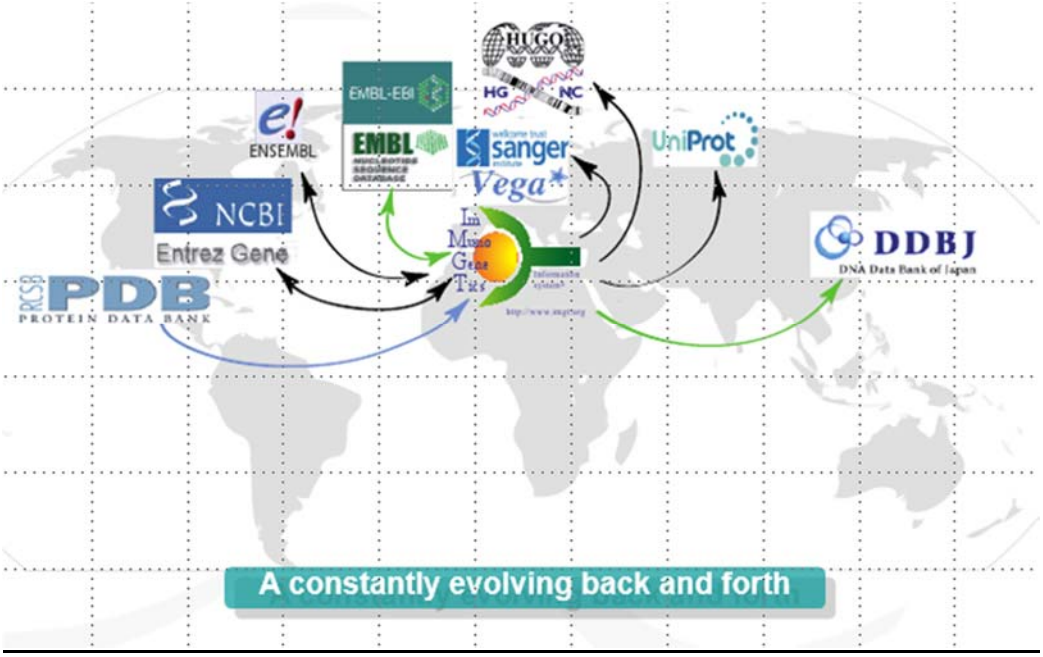
**IMGT®, the international ImMunoGeneTics information system®**
http://www.imgt.org

## 1. Description

**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 de Montpellier 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). More specifically, IMGT® uses the same accession numbers for nucleotide sequences as ENA, GenBank and DDBJ. IMGT genes are officially approved by HUGO Nomenclature Committee (HGNC) since 1999 and used by Ensembl and Vega (Hinxton, UK), NCBI Gene (all have direct links to IMGT/GENE-DB which is the international reference database for IG and TR genes). IMGT® consists of **7 databases** (IMGT/LIGM-DB, IMGT/GENE-DB, etc) and **17 online interactive tools** for the analysis of sequences (IMGT/V-QUEST and its high throughput version IMGT/HighV-QUEST for Next Generation Sequencing (NGS), IMGT/DomainGapAlign, etc.) and the analysis of genes and three-dimensional (3D) structures. IMGT® also provides **one standalone** and more than 20,000 pages of **Web resources**. A schematic representation of the data exchanges is provided in the figure given below. IMGT® resource is unique: there is no equivalent in the USA or elsewhere in the world.




## 2. Principal investigators

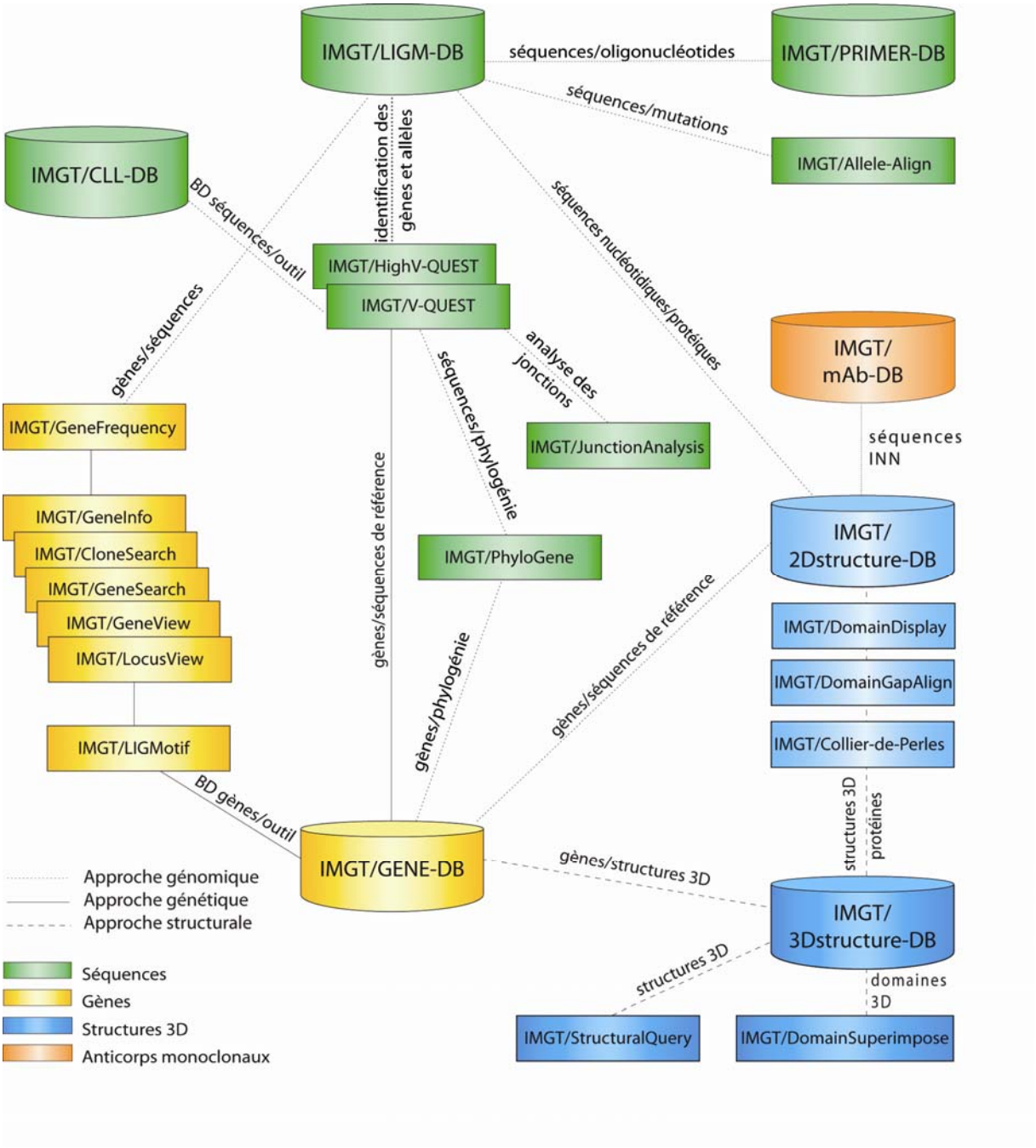
**Marie-Paule LEFRANC**, IMGT® Founder and Executive Director Emeritus, Professor Emeritus Université de Montpellier.
**Gérard LEFRANC**, IMGT® Scientist Emeritus, Professor Emeritus Université de Montpellier.
**Sofia KOSSIDA**, IMGT® Director, Professor Université de Montpellier.

## 3. Year of establishment

**1989**

## 4. Current overview

**October 2016**




**The IMGT®, the international ImMunoGeneTics information system®** (http://www.imgt.org)
IMGT® consists of 7 databases (cylinders) and 17 tools (rectangles). IMGT® also provides (not shown) IMGT/StatClonotype, a novel standalone tool which performs the statistical comparison of two sets of results from IMGT/HighV-QUEST for Next Generation Sequencing (NGS) (500,000 sequences per batch) and more 20,000 pages de Web resources.

## 5. Access

- *Databases*

Data provided by IMGT® come from academic sources and are publicly available. The seven IMGT databases are publicly available to academic users: IMGT/LIGM-DB (the IMGT® nucleotide database (178,296 sequences from 351 species in October 2016), IMGT/PRIMER-DB (the IMGT® primer database), IMGT/GENE-DB (the IMGT® gene database (3,926 genes and 5,627 alleles from 24 species, of which 712 genes and 1,474 alleles for *Homo sapiens* and 868 genes and 1,317 alleles for *Mus musculus* in October 2016), IMGT/2Dstructure-DB (for antibodies and other proteins for which the 3D structure is not available) and IMGT/3Dstructure-DB (for 3D structures, contact analysis and paratope/epitope interactions of IG/antigen and TR/peptide-MH complexes), IMGT/mAb-DB (interface for therapeutic antibodies and fusion proteins for immune applications (FPIA). IMGT/CLL-DB, requires a password as it is a working place for a consortium of clinicians and contains sequences from patients prior to submission to generalist databases (ENA, GenBank, DDBJ) and publications.

- *Standalone tools*

IMGT/StatClonotype, the IMGT® tool distributed as a standalone, is incorporated in the R package 'IMGTStatClonotype' under the LGPL licence. “The GNU Lesser General Public License (LGPL) is corporate-friendly and quite often used in R librairies. It allows for usage of certain library but modifications to it should be made public."

- *Online tools tools and web resources*

The seventeen IMGT® online tools and the web resources (>20,000 pages) are freely accessible to academic users. IMGT interactive tools accessible from the IMGT® Home page comprise: IMGT/V-QUEST (analysis of rearranged nucleotide sequence) and its high-throughput version IMGT/HighV-QUEST (500,000 sequences per batch), IMGT/JunctionAnalysis, IMGT/Allele-Align, IMGT/PhyloGene, IMGT/DomainDisplay (amino acid sequences), IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch, IMGT/CloneSearch, IMGT/GeneInfo, IMGT/GeneFrequency, IMGT/DomainGapAlign (for amino acid sequence analysis of IG and TR variable and constant domains and of MH groove domains), IMGT/Collier-de-Perles (for domain graphical 2D representation), IMGT/DomainSuperimpose, IMGT/StructuralQuery (for 3D structures). A password is required from the IMGT/HighV-QUEST users as this portal requires important computational resources for which IMGT® needs to apply twice a year (GENCI). All the IMGT® reference directories are publicly available online.

## 6. Indicators used to measure the quality and impact

IMGT® uses the indicators of its Quality Management System. IMGT® has been approved by Lloyd's Register Quality Assurance France SAS to the Quality Management System Standards: ISO 9001:2008 since October, 18, 2010 and NFX 50-900 since October 13, 2014. “Research, development and provision of an integrated system (databases, tools and Web resources in immunogenetics and immunoinformatics”. These indicators are analyzed regularly, by the Pilots of the SMQ processes and every year in January at the Steering Committee.
Since the NFX certification, IMGT® has established Key Performance Indicators (KPI). These KPI will be reevaluated at the next Steering Committee in January 2017 to take into account the perspective of the NFX 50-900: 2016 and ISO 9001: 2015.

Scientific focus and quality of science are measured by the number of publications, participations to meetings, invitations, new external collaborations and new contracts.
Community served by the resource is measured by the annual survey (fundamental research, clinicians, biotechnology, big pharma, etc), e-mail exchanges, number of academic collaborations, educational workshops (IG-CLL), networks (EuroClonality), participation to education abroad on antibody informatics (USTH Hanoi, Vietnam), permanent education (Biocampus).
Quality of service is measured by the SMQ indicators (e.g., response time to user messages, survey).

## 7. Translational stories

*IMGT® is in charge, since its creation in 1989 (HGM10, New Haven), of providing the international immunoglobulin (IG) and T cell receptor (TR) gene nomenclature for vertebrate species from fishes to humans.*

By its creation in 1989, IMGT® marked the advent of immunoinformatics, which emerged at the interface between immunogenetics and bioinformatics. For the first time, immunoglobulin (IG) or antibody and T cell receptor (TR) variable (V), diversity (D), joining (J), and constant (C) genes were officially recognized as "genes" as well as the conventional genes. This major breakthrough allowed genes and data of the complex and highly diversified adaptive immune responses to be managed in genomic databases and tools.
The IMGT® *Homo sapiens* IG and TR gene names were approved by the Human Genome Organisation (HUGO) Nomenclature Committee (HGNC) in 1999 and were endorsed by the International Union of Immunological Societies (IUIS) Nomenclature Subcommittee for IG and TR. The IMGT® IG and TR gene names are the official international reference and, as such, have been entered in IMGT/GENE-DB, in the Genome Database (GDB), in LocusLink at the National Center for Biotechnology Information (NCBI) USA, in Entrez Gene (NCBI) when this database (now designated as "Gene") superseded LocusLink, in NCBI MapViewer, in Ensembl at the European Bioinformatics Institute (EBI), and in the Vertebrate Genome Annotation (Vega) Browser at the Wellcome Trust Sanger Institute (UK). HGNC, Gene NCBI, Ensembl, and Vega have direct links to IMGT/GENE-DB.
IMGT® human IG and TR genes were also integrated in IMGT-ONTOLOGY on the NCBO BioPortal and, on the same site, in the HUGO ontology and in the National Cancer Institute (NCI) Metathesaurus. Since 2007, IMGT® gene and allele names have been used for the description of the therapeutic mAb and FPIA of the WHO INN Programme. Amino acid sequences of human IG and TR constant genes (e.g., *Homo sapiens* IGHM, IGHG1, IGHG2) were provided to UniProt in 2008. The current collaboration has for aim the entry of the *Homo sapiens* germline IG variable genes in UniProt.
As an ELIXIR resource, IMGT® will actively pursue the current collaborations with the different genome databases of species newly sequenced and the laboratories engaged in the characterization of the IG and TR genes (UK, Italy, USA, etc.). Indeed the same IMGT concepts IDENTIFICATION (standardized keywords), DESCRIPTION (labels), CLASSIFICATION (gene and allele nomenclature), NUMEROTATION (IMGT unique numbering) and IMGT Scientific chart rules are used whatever the species from fishes to humans.

## *IMGT® is the international reference resource for standardized analysis of the adaptive immune responses in humans and other vertebrate species*

IMGT® high-quality and integrated system provides the resource for the standardized analysis of the expressed IG and TR repertoire of the adaptive immune responses in humans and other vertebrate species (e.g., animal models). They are used in basic, veterinary, and medical research, and are at the forefront of major methodological advances and medical implications. Examples of clinical applications (mutation analysis in leukemia and lymphoma, NGS repertoire analysis of the adaptive immune responses) and of pharmaceutical research and biotechnology (therapeutic antibody engineering and humanization) are given below.

- Mutation analysis in leukemia and lymphoma

  IMGT/V-QUEST is frequently used by clinicians for the analysis of IG somatic hypermutations in leukemia, lymphoma, and myeloma, and more particularly in chronic lymphocytic leukemia (CLL) in which a low percentage of mutations of the rearranged IGHV gene in the VH of the leukemic clone has a poor prognostic value for the patients. For this evaluation, IMGT/V-QUEST is the standard recommended by the European Research Initiative on CLL (ERIC) for comparative analysis between laboratories. The sequences of the V-(D)-J junctions determined by IMGT/JunctionAnalysis are also used in the characterization of stereotypic patterns in CLL and for the synthesis of probes specific of the junction for the detection and follow-up of minimal residual diseases (MRD) in leukemias and lymphomas. A new era is opening up in hemato-oncology with the use of NGS for analysis of the clonality and MRD identification, making IMGT® standards usage more needed as ever.

- NGS repertoire analysis of the adaptive immune responses

  IMGT/HighV-QUEST is the first web portal for Next Generation Sequencing (NGS) immunoprofiling (online since October 2010). It is a paradigm for identification of IMGT clonotype diversity and standardized comparison of the NGS IG and TR repertoire analysis (diversity and expression) of the adaptive immune responses in normal (vaccination) and pathological situations (infectious diseases, cancers).
  - More than 10.7 billions of sequences analyzed to date (October 2016)
  - 1862 users from 46 countries (46% from USA, 34% from EU, 20% from the remaining world)
  - About 5.7 million hours of computational resources
  - 255 terabytes of results generated.

## *IMGT® is the international reference resource for standardized analysis of therapeutic antibody and T cell receptor engineering and antibody humanization*

The therapeutic monoclonal antibody engineering field represents the most promising potential in medicine. IMGT® standards usage is needed more than ever before in antibody humanization and engineering, IG and TR engineering for immunotherapy, search of new specificities and targets, paratope/epitope characterization as demonstrated by the IMGT/DomainGapAlign tool and the IMGT/2Dstructure-DB, IMGT/3Dstructure-DB and IMGT/mAb-DB databases. Indeed, a standardized analysis of IG genomic and expressed

sequences, structures, and interactions is crucial for a better molecular understanding and comparison of the mAb specificity, affinity, half-life, Fc effector properties, and potential immunogenicity. IMGT-ONTOLOGY concepts have become a necessity for IG loci description of newly sequenced genomes, antibody structure/function characterization, antibody engineering [single chain Fragment variable (scFv), phage displays, combinatorial libraries] and antibody humanization (chimeric, humanized, and human antibodies). IMGT® standardization allows repertoire analysis and antibody humanization studies to move to novel high-throughput methodologies with the same high-quality criteria. The CDR-IMGT lengths are now required for mAb INN applications and are included in the WHO INN definitions, bringing a new level of standardized information in the comparative analysis of therapeutic antibodies.

## 8. Funding

(1) IMGT® is a team of the Institut de Génétique Humaine (UPR CNRS 1142). The permanent positions depend on the French Ministère de la Recherche et de l'Enseignement Supérieur. The present and former IMGT® directors are Professors of the Université de Montpellier. There are 7 permanent positions: 4 from the University (1 PU, 2 Emeritus PU, 1 research engineer IR) and 3 from CNRS (1 research engineer IR, 2 engineers IE).

(2) The salaries of the 8-10 CDD IE CNRS (biocurators, bioinformaticians) (turn-over of 3 years) are financed on pharmaceutical industry agreements with CNRS.

(3) Biocampus contributes to the renewal of PC and/or servers.

(4) IMGT/HighV-QUEST is granted access to the HPC@LR and to the High Performance Computing (HPC) resources of the Centre Informatique National de l'Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat l'Energie Atomique et aux Energies Alternatives (CEA) under the allocation [036029] (2010–2016) made by GENCI (Grand Equipement National de Calcul Intensif).

Previous funding sources include several EU grants (BIOMED, BIOTECH, 5th PCRDT, 7th PCRDT), IBiSA, ANR and Region Languedoc-Roussillon funding (listed on the web site).

## 9. Scientific evaluation

IMGT® has an international Scientific Committee since 1994. IMGT® is regularly evaluated scientifically by the SAB of the IGH.

## 10. Updates

IMGT® databases, tools and Web resources are constantly updated. New version numbers and dates are indicated on the web site and the users have the possibility to subscribe to the IMGT news RSS to be informed.

## 11. **Citations over the last three years (2014-2016)**


Aouinti S, Giudicelli V, Duroux P, Malouche D, Kossida S, Lefranc M-P. IMGT/StatClonotype for Pairwise Evaluation and Visualization of NGS IG and TR IMGT Clonotype (AA) Diversity or Expression from IMGT/HighV-QUEST. Front Immunol. 2016 Sep 9;7:339. doi: 10.3389/fimmu.2016.00339. eCollection 2016. Free PMC Article. PMID: 27667992

Aouinti S, Malouche D, Giudicelli V, Kossida S, Lefranc M-P. IMGT/HighV-QUEST statistical significance of IMGT clonotype (AA) diversity per gene for standardized comparisons of next generation sequencing immunoprofiles of immunoglobulins and T cell receptors. PLoS ONE. 2015 Nov 5;10(11):e0142353. doi: 10.1371/journal.pone.0142353. eCollection 2015. Free Article PMID: PMID: 26540440 Correction: PLoS ONE 2016 Jan 5;11(1): e0146702. doi: 10.1371/journal.pone.0146702 View correction. PMID: 26731095

Giudicelli V, Duroux P, Lavoie A, Aouinti S, Lefranc M-P, Kossida S. From IMGT-ONTOLOGY to IMGT/HighV-QUEST for NGS immunoglobulin (IG) and T cell receptor (TR) repertoires in autoimmune and infectious diseases. Autoimmun Infec Dis. 2015 Aug 10. 1(1): doi: 10.16966/aidoa.103. Free Article

Lefranc M-P. Antibody Informatics: IMGT, the International ImMunoGeneTics Information System, In: Crowe J, Boraschi D, Rappuoli R (Eds). Antibodies for Infectious Diseases. ASM Press, Washington, DC. doi: 10.1128/microbiolspec. AID-0001-2012, 2015, pp. 363-379.

Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G, Kossida S. IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. 2015 Jan;43(Database issue):D413-22. doi: 10.1093/nar/gku1056. Epub 2014 Nov 5 Free Article. PMID: 25378316

Lefranc M-P. IMGT® immunoglobulin repertoire analysis and antibody humanization. In: Alt, F.W, Honjo, T, Radbruch A. and Reth, M. (Eds.), Molecular Biology of B cells, Second edition, Academic Press, Elsevier Ltd, London, UK, Chapter 26, 2014, pp. 481-514. dx.doi.org, ISBN : 978-0-12-397933-9.

Lefranc M-P. How to use IMGT® for therapeutic antibody engineering. In: Dübel S. and Reichert J. (Eds), Handbook of Therapeutic Antibodies (4 vol.), Second edition, Wiley, Volume 1: Defining the right antibody composition, Chapter 10, 2014, pp. 229-264.

Lefranc M-P. Antibody informatics: IMGT®, the international ImMunoGeneTics information system®, the international ImMunoGeneTics information system®. Microbiol Spectrum 2014, 2(2):AID-0001-2012. doi:10.1128/microbiolspec.AID-0001-2012.

Lefranc M-P. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. Front Immunol. 2014 Feb 05;5:22. doi: 10.3389/fimmu.2014.00022. Open access. PMID: 24600447