IMGT® databases and tools for biocuration of immunoglobulin and T cell receptor genes

Guilhem Zeitoun, Géraldine Folch, Joumana Jabado-Michaloud, Patrice Duroux, Véronique Giudicelli and Sofia Kossida

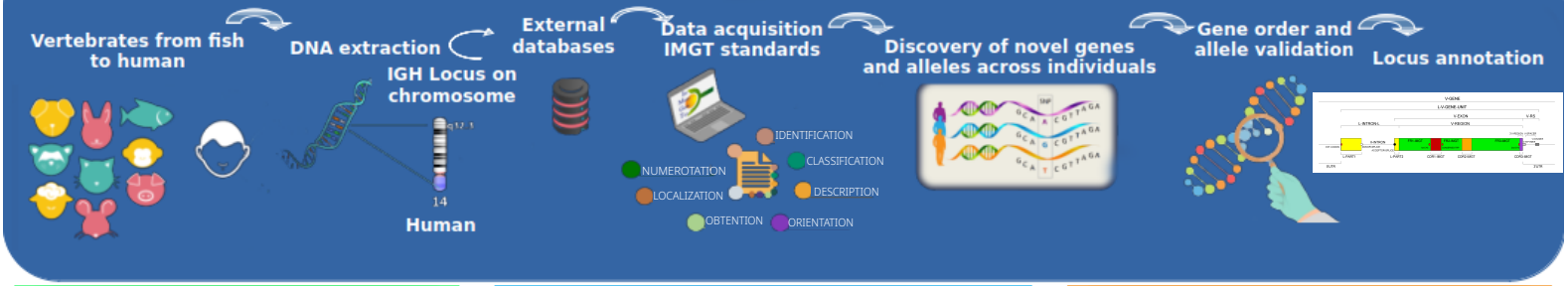
IMGT®, the International ImMunoGeneTics Information System,
University of Montpellier (UM) Institute of Human Genetics (IGH),
UMR 9002 CNRS-UM, Montpellier (France)
Contact: contact-imgt@igh.cnrs.fr

https://www.imgt.org

IMGT®, the International ImMunoGeneTics Information System https://www.imgt.org, is the global reference in immunogenetics and immunoinformatics 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) [1]. IMGT® is at the origin of immunoinformatics, a science at the interface between immunogenetics and bioinformatics. IMGT is based on the concepts of IMGT-ONTOLOGY[2]. 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[3] of IG and TR sequences includes IMGT/StatAssembly for quality evaluation of publicly available assemblies as well as IMGT/LIGMotif, for the analysis of large genomic DNA sequences. Analysis results are checked for consistency, both manually and by using IMGT internal tools. The annotated sequences are integrated in IMGT/LIGM-DB and include the seven ontology axioms: sequence IDENTIFICATION (IMGT keywords) and OBTENTION, the gene and allele CLASSIFICATION (IMGT nomenclature), the constitutive and specific motif DESCRIPTION (IMGT description), the localization and position of the locus (IMGT ORIENTATION and LOCALIZATION), the translation of the coding regions (IMGT NUMEROTATION). In parrallel, IMGT Repertoire is updated (Locus representations, Gene tables...), Protein displays (for new genes and alleles), Alignments of alleles (for new and/or confirmed alleles), Multiple Genome Viewer for assembly comparaison and the IMGT reference directory.

[1] Taciana Manso et al. Nucleic Acids Research, PUBMED: 34875068
[2] Giudicelli V, Lefranc, M-P. Front Genet 3:79 (2012) PMID: 22654892
[3] Pégorier Perrine et al. Front Immunol, PUBMED: 32431713
[4] Nurk, Sergey et al. Science (New York, N.Y.) PMID: 35357919

IMGT IG and TR biocuration pipeline

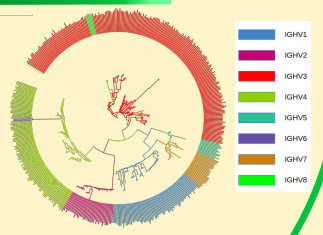


IMGT Nomenclature

1 IMGT genes and alleles

https://www.imgt.org/IMGTScientificChart/Nomenclature/IMGTnomenclature.php

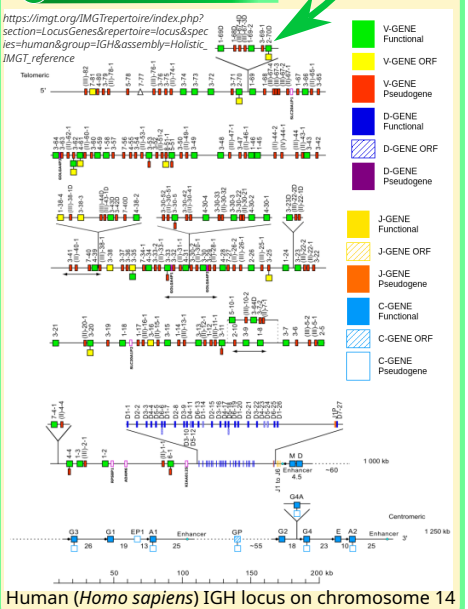
2 IMGT subgroups



V-REGION share at least 75% identity within the same subgroup

3 IMGT gene order

https://imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=locus&species=human&group=IGH&assembly=Holistic_IMGT_reference

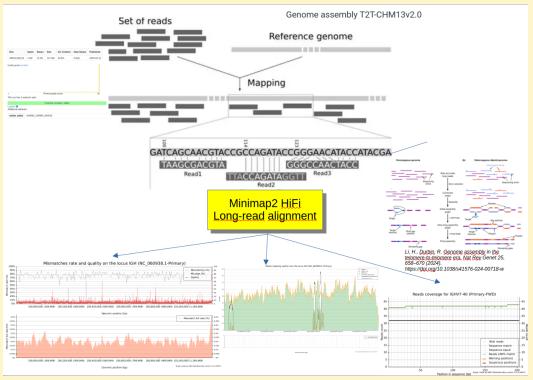


Human (*Homo sapiens*) IGH locus on chromosome 14 (14q32.33) Holistic IMGT reference map

IMGT/StatAssembly
Assessing assemblies quality of IG/TR loci

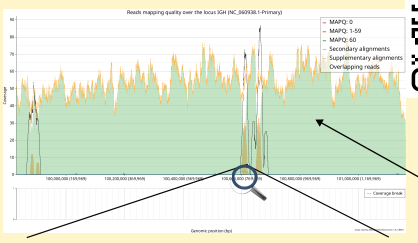
1 Pipeline of analysis

https://galaxyproject.github.io/training-material/topics/sequence-analysis/images/mapping/mapping.png



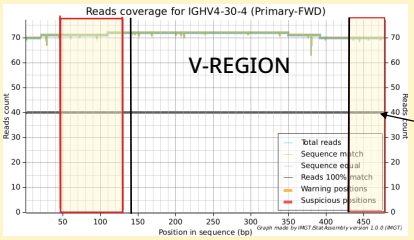
Realignment of HiFi reads into the genome

2 Results file



All figures comes from the software. v1.0.0 https://src.koda.cnrs.fr/imgt-igh/statassembly

Number of reads per position based on mapping quality



INTRONIC REGION

40 reads matching L-PART1 to V-EXON (L-V-GENE-UNIT)

3 Confirmation of locus and alleles

IMGT Rules for quality assessment

T2T-CHM13v2.0 IGH locus has no breaks and a good coverage. It is a full haplotype [4].

GOOD QUALITY LOCUS

Allele has a good coverage and 40 reads matching the full-length.

CONFIRMED ALLELE

https://www.imgt.org/IMGTScientificChart/Assemblies/IMGTassemblyquality.php

IMGT webpages and tools

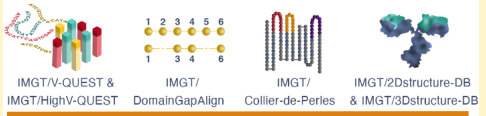
IMGT repertoire
- Locus representations
- Gene tables
- Protein display
- Alignments of alleles

IMGT reference directory
- IG & TR genomic reference sequences

IMGT/GENE-DB
- IG & TR genes and alleles database

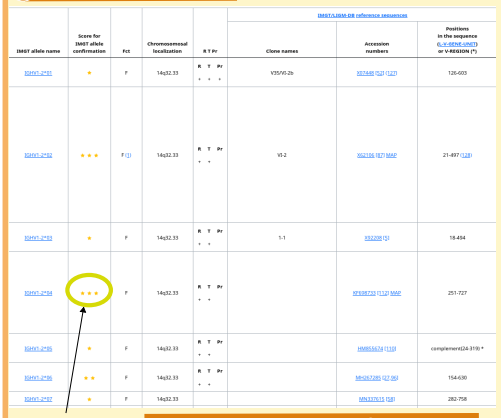
IMGT tools and databases using IMGT reference directory

From https://www.imgt.org/IMGTposters/IMGT_Axe1.pdf



IMGT webpages and tools from nucleotides to proteins

1 Dynamic gene table

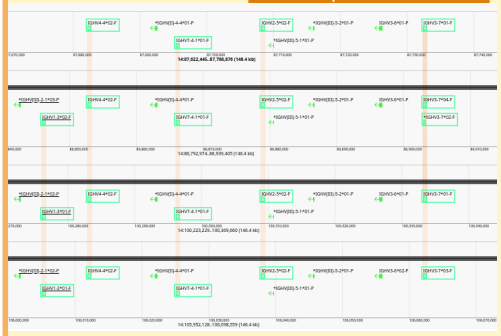
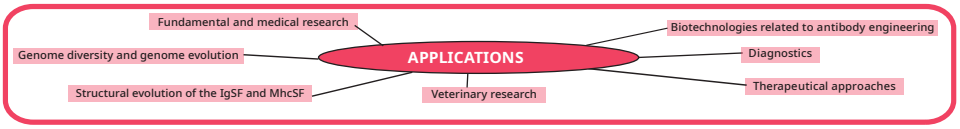


IMGT allele confirmation score (stars)

All IMGT genes and alleles, functionality, functionality notes, allele confidence score

2 Multi Genome Viewer

Visualization of genes among multiple assemblies



https://imgt.org/mgv