IMGT/GENE-DB: genomic reference sequences for human and mouse IG and TR genes and alleles

Fatena Bellahcene, Géraldine Folch, Joumana Jabado-Michaloud, Chantal Ginestoux, Patrice Duroux, Véronique Giudicelli and Marie-Paule Lefranc

IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire LIGM, Université Montpellier 2, Institut de Génétique Humaine IGH, CNRS UPR 1142, 141 rue de la Cardonille, F-34396 Montpellier cedex 05, France

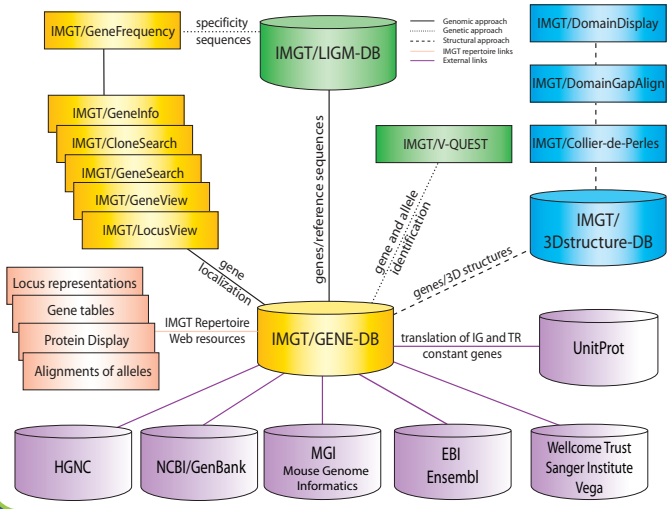
Im Muno Gene Tics Information system®

http://www.imgt.org

The immunoglobulin (IG) and T cell receptor (TR) major loci span about 6 Megabases (Mb) of the human genome on chromosomes 2, 7, 14 and 22, and 9 Mb in mouse on chromosomes 6, 12, 13, 14 and 16. There are seven major loci: three IG loci (IGH, IGK, IGL) and four TR loci (TRA, TRB, TRG, TRD), with a distinct repartition of the variable (V), diversity (D), joining (J) and constant (C) genes. The human genome comprises a total number of 608-665 IG and TR genes (371-422 IG and 237-243 TR), depending on the haplotypes, per haploid genome [1, 2] of which 531-588 genes are located in the major loci (distributed in 369-418 V, 32 D, 105-109 J and 25-29 C genes). There are also 77 orphons (68 IG and 9 TR) including two processed IG genes, outside the major loci. The number of functional IG and TR genes is 308-356 (136-171 IG and 172-185 TR) per haploid genome. The mouse genome comprises an approximate number of 876 IG and TR genes (624 IG and 252 TR). All these genomic data are available in the IMGT® gene database, IMGT/GENE-DB [3]. The major contribution of IMGT/GENE-DB has been to establish, for the first time, a standardized nomenclature of the IG and TR genes and alleles of humans and other vertebrates. In April 2009, IMGT/GENE-DB manages 1999 genes and 3026 alleles.

[1] Lefranc M.-P. and Lefranc G., The Immunoglobulin FactsBook, Academic Press, London, 458 pages (2001).
[2] Lefranc M.-P. and Lefranc G., The T cell receptor FactsBook, Academic Press, London, 398 pages (2001).
[3] Giudicelli V. et al. Nucleic Acids Res., 33, D256-261 (2005).

IMGT/GENE-DB DATABASE



IMGT/GENE-DB provides links to IMGT® sequence database (IMGT/LIGM-DB) and 3D structure database (IMGT/3Dstructure-DB), to IMGT® tools for sequences (IMGT/V-QUEST, IMGT/DomainGapAlign, etc.), genomic data (IMGT/GeneFrequency, etc.), structural data (IMGT/DomainDisplay, etc.), and to IMGT Repertoire Web resources (Chromosomal localizations, Locus representations, Gene tables, etc.).

Human and mouse IG and TR IMGT reference sequences have been provided to NCBI Entrez Gene, Mouse Genome Informatics (MGI), EBI Ensembl and Wellcome Trust Sanger Institute Vega, for the human and mouse genome IG and TR annotation. The translation of the IG and TR constant genes was provided to UniProt in 2008.

IMGT/GENE-DB entry for Mus musculus IGHV1-58

An IMGT/GENE-DB entry includes IMGT gene name and definition, Chromosomal localization, Number of alleles, IMGT reference alleles, IMGT Repertoire links and table of annotated cDNA sequences in IMGT/LIGM-DB and, if available, protein structures in IMGT/3Dstructure-DB.

- IMGT gene name and definition
  - IMGT gene name : *Mus musculus* IGHV1-58
  - IMGT gene definition : Immunoglobulin heavy variable 1-58
- Chromosomal localization
  - Locus name : *Mus musculus* IGH locus
  - Chromosome : 12
  - Chromosomal localization : 12 (58.0 cM)
- Number of alleles: 2
- IMGT reference alleles :

IGHV1-58 allele names	Gene functionality	R	T	Pr	IMGT/LIGM-DB reference sequences			
					Strain	Clone names	Accession numbers	Molecule type
IGHV1-58*01	F		+		C57BL/6	-	AC087166	gDNA
IGHV1-58*02	P				A/J	IGHV1S48,VH1.3	M34978	gDNA

IGHV1-58 allele names	Gene functionality	IMGT/GENE-DB reference sequences (in FASTA format)				
		V-REGION			L-PART1+V-EXON	
		F+ORF+all P	F+ORF+in-frame P with IMGT gaps		F+ORF+all P	F+ORF+in-frame P
IGHV1-58*01	F	nucleotides	nucleotides	amino acids	nucleotides	amino acids
IGHV1-58*02	P					

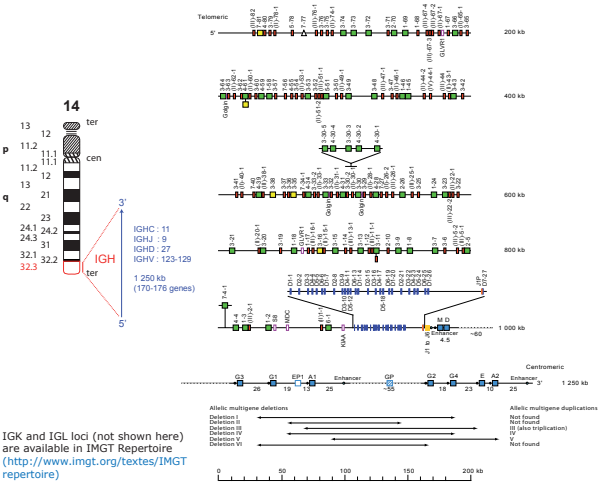
- IMGT Repertoire links :

Locus and genes	Proteins and alleles
Gene table Locus representation	Protein displays

- Annotated IMGT/LIGM-DB cDNA sequences for *Mus musculus* IGHV1-58: 2

HUMAN IMMUNOGLOBULIN GENES

IGH locus at 14q32.3



IGK and IGL loci (not shown here) are available in IMGT Repertoire (http://www.imgt.org/textes/IMGT repertoire)

Total number of IG genes

The human genome comprises 371-422 IG genes (303-354 genes located in the 3 major IG loci and 68 orphons), per haploid genome.

Locus	Major loci						Number of orphons	Total number of IG genes including orphons
	Chromosomal localization	V	D	J	C	Total number of genes		
IGH	14q32.33	123-129	27	9	11[a]	170-176b	36[c]	206-212[b,c]
IGK	2p11.2	(40[d] or) 76	0	5	1	(46d or) 82	25	(71[d] or) 107
IGL	22q11.2	73-74	0	7-11	7-11	87-96	7[e]	94-103
Total number of genes		236-279	27	21-25	19-23	303-354	68[c,e]	371-422[c,e]

Number of functional IG genes

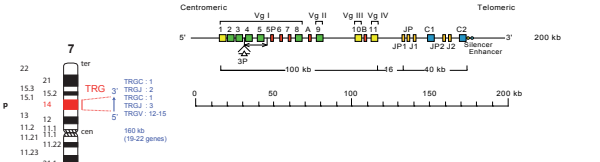
The functional IG genes (136-171 depending on the haplotypes) are located in the 3 major IG loci. Different molecular mechanisms (V-J and V-D-J rearrangements, N-diversity, and for IG, somatic hypermutations), unique to vertebrates, allow to create a huge repertoire of **2x10¹² IG** (or antibodies) per individual.

Locus	Locus size (kb)	V	D	J	C	Number of functional IG genes
IGH	1250	38-46	23	6	9[a]	76-84
IGK	1820	34-38	0	5	1	40-44
	500	17-19[d]				23-25[d]
IGL	1050	29-33	0	4-5	4-5	37-43
Total of functional IG genes		101-117	23	15-16	14-15	136[d]-171
		84-98[d]				

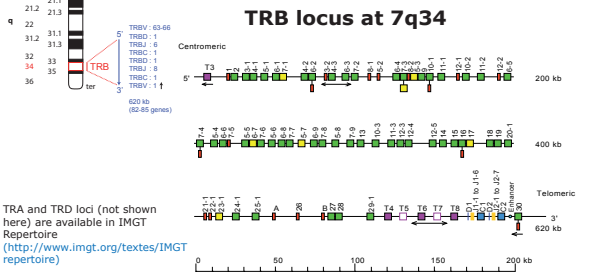
[a] Allelic IGHC multigene deletions, duplications and triplications have been described in healthy individuals. The number of IGHC genes is 5 (deletions I, III, and V), 6 (deletions IV and VI), or 8 (deletion II), per haploid genome.
In haplotypes with multigene duplication or triplication, the exact number of functional IGHC genes per haploid genome is not known.
[b] Not included the 7 non-mapped IGHV genes.
[c] Included the IGHC processed gene, IGHEP2, localized on chromosome 9 (9p24.2-24.1).
[d] Number of genes in the rare IGKV haplotype without the distal V-CLUSTER.
[e] Included the IGLJ-C/OR18 processed gene, localized on chromosome 18 (18p11.31).

HUMAN T CELL GENES

TRG locus at 7p14



TRB locus at 7q34



TRA and TRD loci (not shown here) are available in IMGT Repertoire (http://www.imgt.org/textes/IMGT repertoire)

Total number of TR genes

The human genome comprises 237-243 TR genes (228-234 genes located in the 4 major TR loci and 9 orphons), per haploid genome.

Locus	Major loci						Number of orphons	Total number of TR genes including orphons
	Chromosomal localization	V	D	J	C	Total number of genes		
TRA	14q11.2	54[a]	0	61	1	116a	0	116[a]
TRB	7q34	64-67	2	14	2	82-85	9	91-94
TRG	7p14	42339	0	5	2	19-22	0	19-22
TRD	14q11.2	3 (8[a])	3	4	1	11 (16[a])	0	11 (16[a])
Total number of genes		133-139	5	84	6	228-234	9	237-243

Number of functional TR genes

The functional TR genes (172-185 depending on the haplotypes) are located in the 4 major TR loci. Different molecular mechanisms (V-J and V-D-J rearrangements, N-diversity), unique to vertebrates, allow to create a huge repertoire of **2x10¹² TR** per individual.

Locus	Locus size (kb)	V	D	J	C	Number of functional TR genes
TRA	1000	43-45[a]	0	50	1	94-96[a]
TRB	620	40-48	2	41609	2	56-65
TRG	160	4-6	0	5	2	41579
TRD	60[b] (530[c])	3 (7-8[a])	3	4	1	11 (15-16[a])
Total of functional TR genes		90-102	5	71-72	6	172-185

[a] Including 5 TRAV/TRADV genes. These 5 genes are counted once. The number of functional TRAV and TRDV genes in the TRA/TRD locus is 46-48. The total number of the functional genes for the TRA/TRD locus is 105-107.
[b] Size of the cluster from TRDV2 to TRDV3.
[c] Distance between the most 5' TRAV/DV gene (TRAV14/DV4) and the most 3' gene of the TRD locus (TRDV3).
[d] Taking into account the rearrangement with 2 or 3 TRDD (4 possible combinations: D1,D2; D1,D3; D2,D3; D1,D2,D3).

IMGT founder and director: Marie-Paule Lefranc (Marie-Paule.Lefranc@igh.cnrs.fr)
Bioinformatics manager: Véronique Giudicelli (Veronique.Giudicelli@igh.cnrs.fr)
Computer manager: Patrice Duroux (Patrice.Duroux@igh.cnrs.fr)
Interface design: Chantal Ginestoux (Chantal.Ginestoux@igh.cnrs.fr)
© Copyright 1995-2009 IMGT®, the international ImMunoGeneTics information system®

IMGT® is funded by CNRS, MESR and EU ImmunoGrid IST-2004-028069

© 2009 F.Bellahcene and M-P.Lefranc