Experimental and Clinical Immunogenetics

Exp Clin Immunogenet 2001;18:100–116 Received: August 31, 2000

# Nomenclature of the Human Immunoglobulin Heavy (IGH) Genes

Marie-Paule Lefranc

Chairperson, IMGT Nomenclature Committee, CNRS, Université Montpellier II, Montpellier, France

**Key Words**

Human · IMGT · Immunoglobulin · Heavy chain genes · Orphons

**Abstract**

'Nomenclature of the Human Immunoglobulin Heavy (IGH) Genes', the 16th report of the 'IMGT Locus in Focus' section, provides the first complete list of all the human IGH genes. The total number of human IGH genes per haploid genome is 170–176 (206–212 genes, if the orphons and the processed gene are included), of which 77–84 genes are functional. IMGT/Human Genome Organization (HUGO) gene names and definitions of the human IGH genes on chromosome 14q32.33, processed gene on chromosome 9 and IGH orphons on chromosomes 15 and 16 are provided with the gene functionality and the number of alleles, according to the rules of the IMGT Scientific chart, with the accession numbers of the IMGT reference sequences and with the accession ID of the Genome Database GDB and NCBI LocusLink databases, in which all the IMGT human IGH genes have been entered. The tables are available at the IMGT Marie-Paule page of **IMGT**, the international ImMunoGeneTics database **(http://imgt.cines.fr:8104)** created by Marie-Paule Lefranc, Université Montpellier II, CNRS, France.



## Introduction

'Nomenclature of the Human Immunoglobulin Heavy (IGH) Genes' is the 16th report of the 'IMGT Locus in Focus' section launched in the April 1998 issue of *Experimental and Clinical Immunogenetics* [1–16]. This report comprises two tables and three figures entitled, respectively: (1) 'Complete list of the human IGH genes on chromosome 14 at 14q32.33'; (2) 'Human IGH processed gene on chromosome 9 and IGH orphons on chromosomes 15 and 16'; (3) 'Chromosom-

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com


0254–9670/01/0182–0100$17.50/0

Accessible online at:
www.karger.com/journals/eci

Prof. Marie-Paule Lefranc, IMGT
Laboratoire d'ImmunoGénétique Moléculaire, LIGM, UPR CNRS 1142
IGH, 141 rue de la Cardonille, F–34396 Montpellier Cedex 5 (France)
Tel. +33 4 99 61 99 65, Fax +33 4 99 61 99 01
E-Mail lefranc@ligm.igh.cnrs.fr, IMGT http://imgt.cines.fr:8104

al localization of the human IGH locus at 14q32.33'; (4) 'Representation of the human IGH locus at 14q32.33', and (5) 'The "CLASSIFICATION" concept of the IMGT-ONTOLOGY exemplified for the IGHV genes'. The tables provide the first complete list of all the human IGH genes. The total number of human IGH genes per haploid genome is 170–176 (206–212 genes, if the orphons and the processed gene are included), of which 77–84 genes are functional. IMGT/Human Genome Organization (HUGO) gene names and definitions of the human IGH genes on chromosome 14q32.33, processed gene on chromosome 9 and IGH orphons on chromosomes 15 and 16 are provided with the gene functionality and the number of alleles, according to the rules of the IMGT Scientific chart, with the accession numbers of the IMGT reference sequences and with the accession ID of the Genome Database GDB and NCBI LocusLink databases, in which all the IMGT human IGH genes have been entered. Detailed references for individual IGHV, IGHD, IGHJ and IGHC genes are available in other reports [5–7, 17]. These tables and figures are available at the IMGT Marie-Paule page of **IMGT**, the international ImMunoGeneTics database **(http://imgt.cines.fr:8104)** created by Marie-Paule Lefranc, Université Montpellier II, CNRS, Montpellier, France [18–20].

## Human IGH Locus at 14q32.33

The human IGH locus is located on chromosome 14 [21] at band 14q32.33, at the telomeric extremity of the long arm [22, 23] (fig. 1). The orientation of the locus has been determined by analysis of translocations involving the IGH locus in leukemia and lymphoma.

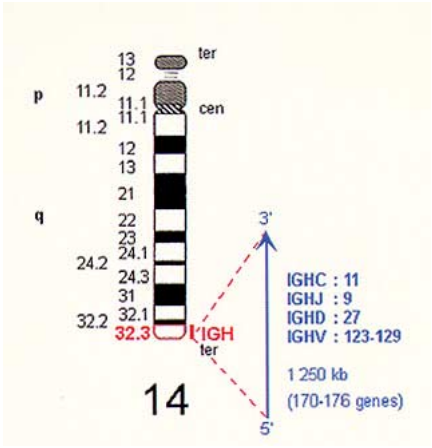


**Fig. 1.** Chromosomal localization of the human IGH locus at 14q32.33. The vertical line indicates the localization of the IGH locus at 14q32.33. The arrow indicates the orientation 5′ → 3′ of the locus, and the gene group order in the locus. The arrow is proportional to the size of the locus, indicated in kilobases (kb). The total number of genes in the locus is shown between parentheses. Seven unmapped IGHV genes which have a provisional designation are not included in this figure. The number of functional genes defines the potential IGH repertoire, which comprises 76–84 genes (38–46 IGHV, 23 IGHD, 6 IGHJ and, in the most frequent haplotype, 9 IGHC) per haploid genome.

**Fig. 2.** Representation of the human IGH locus at 14q32.33. The boxes representing the genes are not to scale. Exons are not shown. SWITCH sequences are represented by a shaded circle upstream of the IGHC genes. The 7 unmapped IGHV genes which correspond to insertion/deletion polymorphisms but which have not yet been precisely located are not shown. Pseudogenes which could not be assigned to subgroups with functional genes are designated by a Roman numeral between parentheses, corresponding to the clans, followed by a hyphen and a number for the localization from 3′ to 5′ in the locus. All these pseudogenes have truncations. Horizontal arrows indicate the extent of the allelic IGHC multigene deletions [36, 37, 50–55] or duplications [56–59] described in healthy individuals. I to VI refer to the type of deletion [for a review, see ref. 55], duplication or triplication, defined by the deleted, duplicated or triplicated genes. Example: deletion I (del G1-EP1-A1-GP-G2-G4).

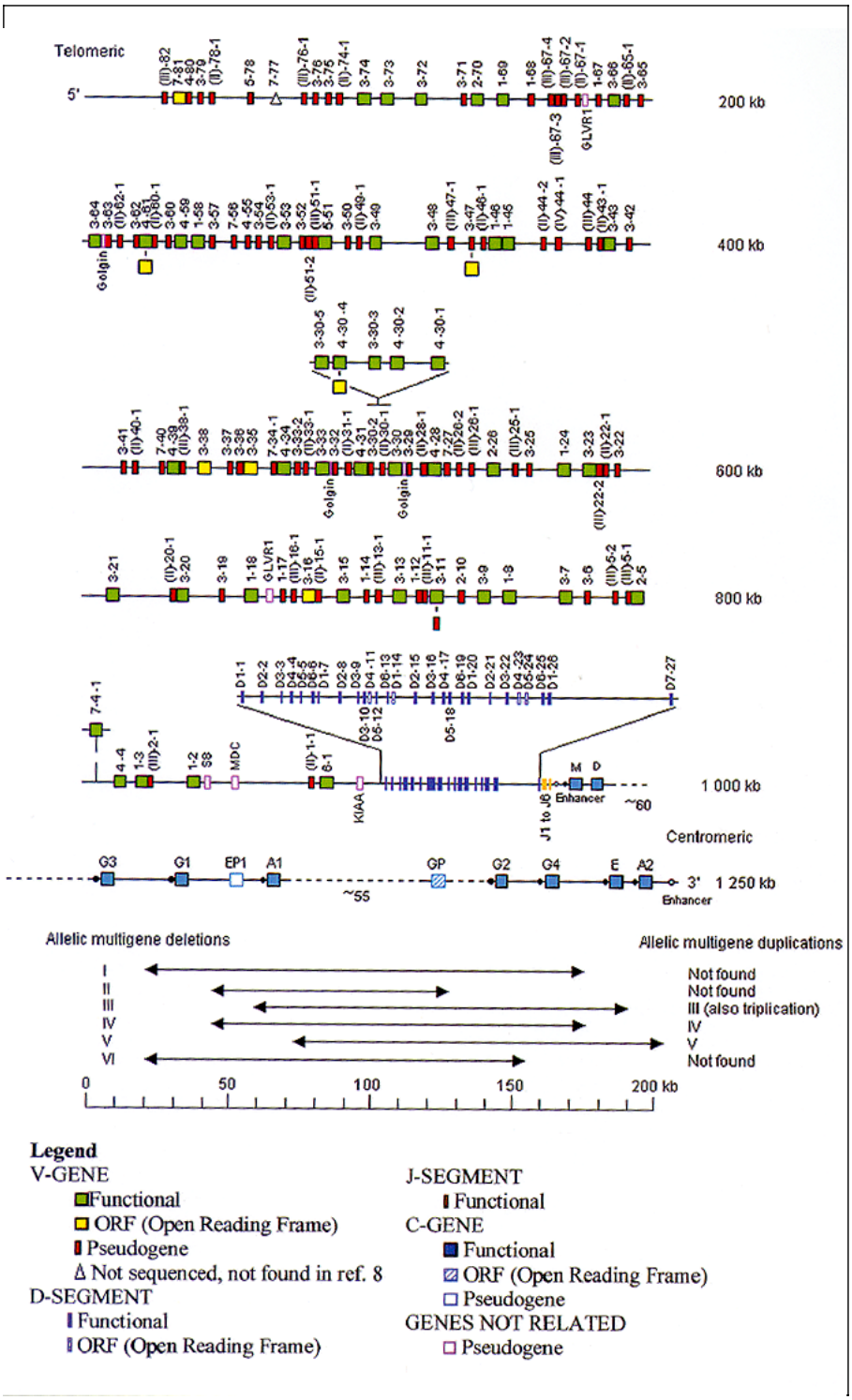


**Fig. 2.**
For legend see p. 101.

**Table 1.** Complete list of the human IGH genes on chromosome 14 at 14q32.33

IGHV gene nomenclature: IGHV genes are designated by a number for the subgroup, followed by a hyphen and a number for the localization from 3′ to 5′ in the locus. Seven genes which have been described as insertion/deletion polymorphisms but which have not been precisely located are designated by a number for the subgroup, followed by a hyphen and a small letter: IGHV1-c, IGHV1-f, IGHV3-d, IGHV3-g, IGHV3-h, IGHV4-b, IGHV5-a. A putative gene IGHV7-77 was not found by Matsuda et al., 1998. These genes are not counted in the potential repertoire [5, 17] and have a provisional designation.

Pseudogenes which could not be assigned to subgroups with functional genes are designated by a Roman numeral between parentheses, corresponding to the clans, followed by a hyphen and a number for the localization from 3′ to 5′ in the locus. All these pseudogenes have truncations.

Clans comprise, respectively:

– clan I: IGHV1, IGHV5 and IGHV7 subgroup genes
– clan II: IGHV2, IGHV4 and IGHV6 subgroup genes, and pseudogenes IGHV(II)
– clan III: IGHV3 subgroup genes, and pseudogenes IGHV(III)
– clan IV: one pseudogene IGHV(IV)-44

An asterisk (*) indicates allelic polymorphisms by insertion/deletion which concern:

– a 50 kb insertion of 5 genes (3–30–5, 4–30–4, 3–30–3, 4–30–2, 4–30–1) observed in 45% Caucasoids
– the IGHV7–4-1 gene

| IMGT gene groups | IMGT gene name (1) | IMGT func-tionality | IMGT reference sequence accession numbers | Number of alleles | IMGT gene definition (2) | GDB accession ID (3) | LocusLink Accession ID (3) |
|---|---|---|---|---|---|---|---|
| IGHC | IGHA1 | F | J00220, M60193 | 1 | Immunoglobulin heavy constant alpha 1 | GDB:119332 | 3493 |
| | IGHA2 | F | J00221 | 3 | Immunoglobulin heavy constant alpha 2 (A2m marker) | GDB:119333 | 3494 |
| | IGHD | F | K02875-K02882 | 2 | Immunoglobulin heavy constant delta | GDB:120084 | 3495 |
| | IGHE | F | J00222, X63693 | 3 | Immunoglobulin heavy constant epsilon | GDB:119335 | 3497 |
| | IGHEP1 | P | J00223 | 3 | Immunoglobulinheavy constant epsilon P1 | GDB:119336 | 3498 |
| | IGHG1 | F | J00228 | 2 | Immunoglobulin heavy constant gamma 1 (G1m marker) | GDB:120085 | 3500 |
| | IGHG2 | F | J00230, AB006775, AB006776 | 2 | Immunoglobulin heavy constant gamma 2 (G2m marker) | GDB:119338 | 3501 |
| | IGHG3 | F | M12958/X03604 | 4 | Immunoglobulin heavy constant gamma 3 (G3m marker) | GDB:119339 | 3502 |
| | IGHG4 | F | K01316 | 3 | Immunoglobulin heavy constant gamma 4 (G4m marker) | GDB:119340 | 3503 |
| | IGHGP | ORF | X06766 | 1 | Immunoglobulin heavy constant gamma P | GDB:120689 | 3505 |
| | IGHM | F | X14940, X14939 | 3 | Immunoglobulin heavy constant mu | GDB:120086 | 3507 |
| IGHD | IGHD1–1 | F | X97051 | 1 | Immunoglobulinheavy diversity 1–1 | GDB:9953175 | 28510 |
| | IGHD1–7 | F | X13972 | 1 | Immunoglobulin heavy diversity 1–7 | GDB:9953261 | 28509 |
| | IGHD1–14 | ORF | X13972 | 1 | Immunoglobulin heavy diversity 1–14 | GDB:9953263 | 28508 |
| | IGHD1–20 | F | X97051 | 1 | Immunoglobulin heavy diversity 1–20 | GDB:9953265 | 28507 |
| | IGHD1–26 | F | X97051 | 1 | Immunoglobulin heavy diversity 1–26 | GDB:9953266 | 28506 |
| | IGHD2–2 | F | J00232 | 3 | Immunoglobulinheavy diversity 2–2 | GDB:9953230 | 28505 |
| | IGHD2–8 | F | X13972 | 2 | Immunoglobulin heavy diversity 2–8 | GDB:9953278 | 28504 |

**Table 1** (continued)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | IGHD2–15 | F | J00234 | 1 | Immunoglobulinheavy diversity 2–15 | GDB:9953292 | 28503 |
| | IGHD2–21 | F | J00235 | 2 | Immunoglobulinheavy diversity 2–21 | GDB:9953294 | 28502 |
| | IGHD3–3 | F | X13972 | 2 | Immunoglobulinheavy diversity 3–3 | GDB:9953296 | 28501 |
| | IGHD3–9 | F | X13972 | 1 | Immunoglobulin heavy diversity 3–9 | GDB:9953298 | 28500 |
| | IGHD3–10 | F | X13972 | 2 | Immunoglobulinheavy diversity 3–10 | GDB:9953300 | 28499 |
| | IGHD3–16 | F | X93614 | 1 | Immunoglobulin heavy diversity 3–16 | GDB:9953302 | 28498 |
| | IGHD3–22 | F | X93616 | 1 | Immunoglobulin heavy diversity 3–22 | GDB:9953304 | 28497 |
| | IGHD4–4 | F | X13972 | 1 | Immunoglobulinheavy diversity 4–4 | GDB:9953306 | 28496 |
| | IGHD4–11 | ORF | X13972 | 1 | Immunoglobulin heavy diversity 4–11 | GDB:9953308 | 28495 |
| | IGHD4–17 | F | X97051 | 1 | Immunoglobulin heavy diversity 4–17 | GDB:9953310 | 28494 |
| | IGHD4–23 | ORF | X97051 | 1 | Immunoglobulin heavy diversity 4–23 | GDB:9953312 | 28493 |
| | IGHD5–5 | F | X13972 | 1 | Immunoglobulinheavy diversity 5–5 | GDB:9953314 | 28492 |
| | IGHD5–12 | F | X13972 | 1 | Immunoglobulinheavy diversity 5–12 | GDB:9953316 | 28491 |
| | IGHD5–18 | F | X97051 | 1 | Immunoglobulin heavy diversity 5–18 | GDB:9953318 | 28490 |
| | IGHD5–24 | ORF | X97051 | 1 | Immunoglobulin heavy diversity 5–24 | GDB:9953320 | 28489 |
| | IGHD6–6 | F | X13972 | 1 | Immunoglobulinheavy diversity 6–6 | GDB:9953322 | 28488 |
| | IGHD6–13 | F | X13972 | 1 | Immunoglobulinheavy diversity 6–13 | GDB:9953324 | 28487 |
| | IGHD6–19 | F | X97051 | 1 | Immunoglobulin heavy diversity 6–19 | GDB:9953326 | 28486 |
| | IGHD6–25 | F | X97051 | 1 | Immunoglobulin heavy diversity 6–25 | GDB:9953328 | 28485 |
| | IGHD7–27 | F | J00256 | 1 | Immunoglobulin heavy diversity 7–27 | GDB:9953330 | 28484 |
| IGHJ | IGHJ1 | F | J00256 | 1 | Immunoglobulin heavy joining 1 | GDB:9953332 | 28483 |
| | IGHJ1P | P | J00256 | – | Immunoglobulin heavy joining 1P | GDB:9953334 | 28482 |
| | IGHJ2 | F | J00256 | 1 | Immunoglobulin heavy joining 2 | GDB:9953336 | 28481 |
| | IGHJ2P | P | J00256 | – | Immunoglobulin heavy joining 2P | GDB:9953338 | 28480 |
| | IGHJ3 | F | J00256 | 2 | Immunoglobulin heavy joining 3 | GDB:9953340 | 28479 |
| | IGHJ3P | P | J00256 | – | Immunoglobulin heavy joining 3P | GDB:9953342 | 28478 |
| | IGHJ4 | F | J00256 | 3 | Immunoglobulin heavy joining 4 | GDB:9953344 | 28477 |
| | IGHJ5 | F | J00256 | 2 | Immunoglobulin heavy joining 5 | GDB:9953346 | 28476 |
| | IGHJ6 | F | J00256 | 3 | Immunoglobulin heavy joining 6 | GDB:9953348 | 28475 |
| IGHV | IGHV1–2 | F | X07448 | 4 | Immunoglobulin heavy variable 1–2 | GDB:9931660 | 28474 |
| | IGHV1–3 | F | X62109 | 2 | Immunoglobulin heavy variable 1–3 | GDB:9931661 | 28473 |
| | IGHV1–8 | F | M99637 | 1 | Immunoglobulin heavy variable 1–8 | GDB:9931662 | 28472 |
| | IGHV1–12 | P | X92210 | – | Immunoglobulinheavy variable 1–12 | GDB:9931663 | 28471 |
| | IGHV1–14 | P | M99639 | – | Immunoglobulin heavy variable 1–14 | GDB:9931664 | 28470 |
| | IGHV1–17 | P | M99640 | – | Immunoglobulin heavy variable 1–17 | GDB:9931665 | 28469 |
| | IGHV1–18 | F | M99641 | 2 | Immunoglobulin heavy variable 1–18 | GDB:9931666 | 28468 |
| | IGHV1–24 | F | M99642 | 1 | Immunoglobulin heavy variable 1–24 | GDB:9931667 | 28467 |
| | IGHV1–45 | F | X92209 | 3 | Immunoglobulin heavy variable 1–45 | GDB:9931668 | 28466 |
| | IGHV1–46 | F | X92343 | 3 | Immunoglobulin heavy variable 1–46 | GDB:9931669 | 28465 |
| | IGHV1–58 | F | M29809 | 1 | Immunoglobulin heavy variable 1–58 | GDB:9931670 | 28464 |
| | IGHV1–67 | P | X92212 | – | Immunoglobulin heavy variable 1–67 | GDB:9931671 | 28463 |

| | | | | | | |
|---|---|---|---|---|---|---|
| IGHV1–68 | P | AB019437 | – | Immunoglobulin heavy variable 1–68 | GDB:9931672 | 28462 |
| IGHV1–69 | F | L22582 | 7 | Immunoglobulin heavy variable 1–69 | GDB:9931673 | 28461 |
| IGHV1-c | ORF | Z18904 | 1 | Immunoglobulin heavy variable 1-c (provisional) | GDB:9931674 | 28460 |
| IGHV1-f | F | Z12305 | 2 | Immunoglobulin heavy variable 1-f (provisional) | GDB:9931676 | 28458 |
| IGHV2–5 | F | X62111 | 9 | Immunoglobulin heavy variable 2–5 | GDB:9931677 | 28457 |
| IGHV2–10 | P | M99647 | – | Immunoglobulinheavy variable 2–10 | GDB:9931678 | 28456 |
| IGHV2–26 | F | M99648 | 1 | Immunoglobulin heavy variable 2–26 | GDB:9931679 | 28455 |
| IGHV2–70 | F | L21969 | 12 | Immunoglobulin heavy variable 2–70 | GDB:9931680 | 28454 |
| IGHV3–6 | P | M99650 | – | Immunoglobulinheavy variable 3–6 | GDB:9931681 | 28453 |
| IGHV3–7 | F | M99649 | 2 | Immunoglobulin heavy variable 3–7 | GDB:9931682 | 28452 |
| IGHV3–9 | F | M99651 | 1 | Immunoglobulin heavy variable 3–9 | GDB:9931683 | 28451 |
| IGHV3–11 | F, P | M99652 | 3 | Immunoglobulin heavy variable 3–11 | GDB:9931684 | 28450 |
| IGHV3–13 | F | X92217 | 2 | Immunoglobulin heavy variable 3–13 | GDB:9931685 | 28449 |
| IGHV3–15 | F | X92216 | 8 | Immunoglobulin heavy variable 3–15 | GDB:9931686 | 28448 |
| IGHV3–16 | ORF | M99655 | 1 | Immunoglobulin heavy variable 3–16 | GDB:9931687 | 28447 |
| IGHV3–19 | P | M99656 | – | Immunoglobulin heavy variable 3–19 | GDB:9931688 | 28446 |
| IGHV3–20 | F | M99657 | 1 | Immunoglobulin heavy variable 3–20 | GDB:9931689 | 28445 |
| IGHV3–21 | F | Z14073 | 2 | Immunoglobulin heavy variable 3–21 | GDB:9931690 | 28444 |
| IGHV3–22 | P | M99659 | – | Immunoglobulin heavy variable 3–22 | GDB:9931691 | 28443 |
| IGHV3–23 | F | M99660 | 3 | Immunoglobulin heavy variable 3–23 | GDB:9931692 | 28442 |
| IGHV3–25 | P | M99661 | – | Immunoglobulin heavy variable 3–25 | GDB:9931693 | 28441 |
| IGHV3–29 | P | M99662 | – | Immunoglobulin heavy variable 3–29 | GDB:9931694 | 28440 |
| IGHV3–30 | F | M83134 | 19(a) | Immunoglobulin heavy variable 3–30 | GDB:9931735 | 28439 |
| IGHV3–30–2 | P | AB019439 | – | Immunoglobulin heavy variable 3–30–2 | GDB:9931695 | 28438 |
| (*)IGHV3–30–3 | F | X92283 | 2 | Immunoglobulin heavy variable 3–30–3 | GDB:9931696 | 28437 |
| (*)IGHV3–30–5 | F | (a) | (a) | Immunoglobulin heavy variable 3–30–5 | GDB:9931697 | 28436 |
| IGHV3–32 | P | M99664 | – | Immunoglobulin heavy variable 3–32 | GDB:9931698 | 28435 |
| IGHV3–33 | F | L06618 | 5 | Immunoglobulin heavy variable 3–33 | GDB:9931699 | 28434 |
| IGHV3–33–2 | P | AB019439 | – | Immunoglobulin heavy variable 3–33–2 | GDB:9931700 | 28433 |
| IGHV3–35 | ORF | M99666 | 1 | Immunoglobulin heavy variable 3–35 | GDB:9931701 | 28432 |
| IGHV3–36 | P | M99667 | – | Immunoglobulin heavy variable 3–36 | GDB:9931702 | 28431 |
| IGHV3–37 | P | M99668 | – | Immunoglobulin heavy variable 3–37 | GDB:9931703 | 28430 |
| IGHV3–38 | ORF | M99669 | 2 | Immunoglobulin heavy variable 3–38 | GDB:9931704 | 28429 |
| IGHV3–41 | P | M99670 | – | Immunoglobulin heavy variable 3–41 | GDB:9931705 | 28428 |
| IGHV3–42 | P | M99671 | – | Immunoglobulin heavy variable 3–42 | GDB:9931706 | 28427 |
| IGHV3–43 | F | M99672 | 2 | Immunoglobulin heavy variable 3–43 | GDB:9931707 | 28426 |
| IGHV3–47 | ORF, P | Z18900 | 3 | Immunoglobulin heavy variable 3–47 | GDB:9931708 | 28425 |
| IGHV3–48 | F | M99675 | 3 | Immunoglobulin heavy variable 3–48 | GDB:9931709 | 28424 |
| IGHV3–49 | F | M99676 | 3 | Immunoglobulin heavy variable 3–49 | GDB:9931710 | 28423 |
| IGHV3–50 | P | M99677 | – | Immunoglobulin heavy variable 3–50 | GDB:9931711 | 28422 |
| IGHV3–52 | P | M99678 | – | Immunoglobulin heavy variable 3–52 | GDB:9931712 | 28421 |
| IGHV3–53 | F | M99679 | 2 | Immunoglobulin heavy variable 3–53 | GDB:9931713 | 28420 |
| IGHV3–54 | P | M99680 | – | Immunoglobulin heavy variable 3–54 | GDB:9931714 | 28419 |
| IGHV3–57 | P | M29815 | – | Immunoglobulin heavy variable 3–57 | GDB:9931715 | 28418 |

**Table 1** (continued)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | IGHV3–60 | P | M29813 | – | Immunoglobulin heavy variable 3–60 | GDB:9931716 | 28417 |
| | IGHV3–62 | P | M29814 | – | Immunoglobulin heavy variable 3–62 | GDB:9931717 | 28416 |
| | IGHV3–63 | P | M99681 | – | Immunoglobulin heavy variable 3–63 | GDB:9931718 | 28415 |
| | IGHV3–64 | F | M99682 | 5 | Immunoglobulin heavy variable 3–64 | GDB:9931719 | 28414 |
| | IGHV3–65 | P | Z27503 | – | Immunoglobulin heavy variable 3–65 | GDB:9931720 | 28413 |
| | IGHV3–66 | F | X92218 | 3 | Immunoglobulin heavy variable 3–66 | GDB:9931736 | 28412 |
| | IGHV3–71 | P | AB019437 | – | Immunoglobulin heavy variable 3–71 | GDB:9931721 | 28411 |
| | IGHV3–72 | F | X92206 | 2 | Immunoglobulin heavy variable 3–72 | GDB:9931722 | 28410 |
| | IGHV3–73 | F | Z27508 | 1 | Immunoglobulin heavy variable 3–73 | GDB:9931723 | 28409 |
| | IGHV3–74 | F | Z12353 | 3 | Immunoglobulin heavy variable 3–74 | GDB:9931724 | 28408 |
| | IGHV3–75 | P | Z27510 | – | Immunoglobulin heavy variable 3–75 | GDB:9931725 | 28407 |
| | IGHV3–76 | P | Z12341 | – | Immunoglobulin heavy variable 3–76 | GDB:9931726 | 28406 |
| | IGHV3–79 | P | AB019437 | – | Immunoglobulin heavy variable 3–79 | GDB:9931727 | 28405 |
| | IGHV3-d | F | Z18898 | 1 | Immunoglobulinheavy variable 3-d (provisional) | GDB:9931729 | 28404 |
| | IGHV3-g | P | Z12336 | – | Immunoglobulin heavy variable 3-g (provisional) | GDB:9931728 | 28403 |
| | IGHV3-h | n.d. | Z29981 | 1 | Immunoglobulin heavy variable 3-h (provisional) partial sequence | GDB:9931730 | 28402 |
| | IGHV4–4 | F | X05713 | 8 | Immunoglobulin heavy variable 4–4 | GDB:9931731 | 28401 |
| | IGHV4–28 | F | X05714 | 5 | Immunoglobulinheavy variable 4–28 | GDB:9931732 | 28400 |
| (*) | IGHV4–30–1 | F | (b) | (b) | Immunoglobulin heavy variable 4–30–1 | GDB:9931734 | 28399 |
| (*) | IGHV4–30–2 | F | L10089 | 4 | Immunoglobulin heavy variable 4–30–2 | GDB:9953352 | 28398 |
| (*) | IGHV4–30–4 | F | Z14238 | 6 | Immunoglobulin heavy variable 4–30–4 | GDB:9953354 | 28397 |
| | IGHV4–31 | F | L10098 | 10(b) | Immunoglobulin heavy variable 4–31 | GDB:9931737 | 28396 |
| | IGHV4–34 | F | X92278 | 13 | Immunoglobulin heavy variable 4–34 | GDB:9931738 | 28395 |
| | IGHV4–39 | F | L10094 | 6 | Immunoglobulin heavy variable 4–39 | GDB:9931739 | 28394 |
| | IGHV4–55 | P | M99685 | – | Immunoglobulin heavy variable 4–55 | GDB:9931740 | 28393 |
| | IGHV4–59 | F | L10088 | 9 | Immunoglobulin heavy variable 4–59 | GDB:9931741 | 28392 |
| | IGHV4–61 | F, ORF | M29811 | 7 | Immunoglobulin heavy variable 4–61 | GDB:9931742 | 28391 |
| | IGHV4–80 | P | AB019437 | – | Immunoglobulin heavy variable 4–80 | GDB:9931743 | 28390 |
| | IGHV4-b | F | Z12367 | 2 | Immunoglobulinheavy variable 4-b (provisional) | GDB:9931744 | 28389 |
| | IGHV5–51 | F | M99686 | 5 | Immunoglobulin heavy variable 5–51 | GDB:9931745 | 28388 |
| | IGHV5–78 | P | X92213 | – | Immunoglobulin heavy variable 5–78 | GDB:9931746 | 28387 |
| | IGHV5-a | F, P | X92227 | 4 | Immunoglobulin heavy variable 5-a (provisional) | GDB:9931747 | 28386 |
| | IGHV6–1 | F | X92224/J04097 | 2 | Immunoglobulin heavy variable 6–1 | GDB:9931748 | 28385 |
| (*) | IGHV7–4–1 | F | L10057 | 3 | Immunoglobulin heavy variable 7–4-1 | GDB:9931749 | 28384 |
| | IGHV7–27 | P | M99643 | – | Immunoglobulin heavy variable 7–27 | GDB:9931750 | 28383 |
| | IGHV7–34–1 | P | AB019739 | – | Immunoglobulin heavy variable 7–34–1 | GDB:9931751 | 28382 |
| | IGHV7–40 | P | M99644 | – | Immunoglobulin heavy variable 7–40 | GDB:9953358 | 28381 |
| | IGHV7–56 | P | M29810 | – | Immunoglobulin heavy variable 7–56 | GDB:9931752 | 28380 |
| | IGHV7–77 | | | | Immunoglobulin heavy variable 7–77 (provisional) | GDB:9931753 | 28379 |
| | IGHV7–81 | ORF | Z27509 | 1 | Immunoglobulin heavy variable 7–81 | GDB:9931754 | 28378 |

| | | | | | | |
|---|---|---|---|---|---|---|
| IGHV(II)-1–1 | P | AB019441 | – | Immunoglobulin heavy variable (II)-1–1 | GDB:9931755 | 28377 |
| IGHV(II)-15–1 | P | AB019440 | – | Immunoglobulin heavy variable (II)-15–1 | GDB:9931756 | 28376 |
| IGHV(II)-20–1 | P | AB019440 | – | Immunoglobulin heavy variable (II)-20–1 | GDB:9931757 | 28375 |
| IGHV(II)-22–1 | P | AB019439 | – | Immunoglobulin heavy variable (II)-22–1 | GDB:9931758 | 28374 |
| IGHV(II)-26–2 | P | AB019439 | – | Immunoglobulin heavy variable (II)-26–2 | GDB:9933358 | 28373 |
| IGHV(II)-28–1 | P | AB019439 | – | Immunoglobulin heavy variable (II)-28–1 | GDB:9931759 | 28372 |
| IGHV(II)-30–1 | P | AB019439 | – | Immunoglobulin heavy variable (II)-30–1 | GDB:9931760 | 28371 |
| IGHV(II)-31–1 | P | AB019439 | – | Immunoglobulin heavy variable (II)-31–1 | GDB:9931761 | 28370 |
| IGHV(II)-33–1 | P | AB019439 | – | Immunoglobulin heavy variable (II)-33–1 | GDB:9931762 | 28369 |
| IGHV(II)-40–1 | P | AB019438 | – | Immunoglobulin heavy variable (II)-40–1 | GDB:9953360 | 28368 |
| IGHV(II)-43–1 | P | AB019438 | – | Immunoglobulin heavy variable (II)-43–1 | GDB:9953362 | 28367 |
| IGHV(II)-44–2 | P | AB019438 | – | Immunoglobulin heavy variable (II)-44–2 | GDB:9931763 | 28366 |
| IGHV(II)-46–1 | P | AB019438 | – | Immunoglobulin heavy variable (II)-46–1 | GDB:9931764 | 28365 |
| IGHV(II)-49–1 | P | AB019438 | – | Immunoglobulin heavy variable (II)-49–1 | GDB:9931765 | 28364 |
| IGHV(II)-51–2 | P | AB019438 | – | Immunoglobulin heavy variable (II)-51–2 | GDB:9931766 | 28363 |
| IGHV(II)-53–1 | P | AB019438 | – | Immunoglobulin heavy variable (II)-53–1 | GDB:9931767 | 28362 |
| IGHV(II)-60–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-60–1 | GDB:9931768 | 28361 |
| IGHV(II)-62–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-62–1 | GDB:9931769 | 28360 |
| IGHV(II)-65–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-65–1 | GDB:9931770 | 28359 |
| IGHV(II)-67–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-67–1 | GDB:9953364 | 28358 |
| IGHV(II)-74–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-74–1 | GDB:9931771 | 28357 |
| IGHV(II)-78–1 | P | AB019437 | – | Immunoglobulin heavy variable (II)-78–1 | GDB:9931772 | 28356 |
| IGHV(III)-2–1 | P | AB019441 | – | Immunoglobulin heavy variable (III)-2–1 | GDB:9931773 | 28355 |
| IGHV(III)-5–1 | P | AB019440 | – | Immunoglobulin heavy variable (III)-5–1 | GDB:9931774 | 28354 |
| IGHV(III)-5–2 | P | AB019440 | – | Immunoglobulin heavy variable (III)-5–2 | GDB:9931775 | 28353 |
| IGHV(III)-11–1 | P | AB019440 | – | Immunoglobulin heavy variable (III)-11–1 | GDB:9931776 | 28352 |
| IGHV(III)-13–1 | P | AB019440 | – | Immunoglobulin heavy variable (III)-13–1 | GDB:9931777 | 28351 |
| IGHV(III)-16–1 | P | AB019440 | – | Immunoglobulin heavy variable (III)-16–1 | GDB:9931778 | 28350 |
| IGHV(III)-22–2 | P | AB019439 | – | Immunoglobulin heavy variable (III)-22–2 | GDB:9931779 | 28349 |
| IGHV(III)-25–1 | P | AB019439 | – | Immunoglobulin heavy variable (III)-25–1 | GDB:9931780 | 28348 |
| IGHV(III)-26–1 | P | AB019439 | – | Immunoglobulin heavy variable (III)-26–1 | GDB:9953366 | 28347 |
| IGHV(III)-38–1 | P | AB019439 | – | Immunoglobulin heavy variable (III)-38–1 | GDB:9953368 | 28346 |
| IGHV(III)-44 | P | M99673 | – | Immunoglobulin heavy variable (III)-44 | GDB:9953370 | 28345 |
| IGHV(III)-47–1 | P | AB019438 | – | Immunoglobulin heavy variable (III)-47–1 | GDB:9953372 | 28344 |
| IGHV(III)-51–1 | P | AB019438 | – | Immunoglobulin heavy variable (III)-51–1 | GDB:9953381 | 28343 |
| IGHV(III)-67–2 | P | AB019437 | – | Immunoglobulin heavy variable (III)-67–2 | GDB:9953382 | 28342 |
| IGHV(III)-67–3 | P | AB019437 | – | Immunoglobulin heavy variable (III)-67–3 | GDB:9953383 | 28341 |
| IGHV(III)-67–4 | P | AB019437 | – | Immunoglobulin heavy variable (III)-67–4 | GDB:9953385 | 28340 |
| IGHV(III)-76–1 | P | AB019437 | – | Immunoglobulin heavy variable (III)-76–1 | GDB:9953386 | 28339 |
| IGHV(III)-82 | P | AB019437 | – | Immunoglobulin heavy variable (III)-82 | GDB:9953388 | 28338 |
| IGHV(IV)-44–1 | P | AB019438 | – | Immunoglobulin heavy variable (IV)-44–1 | GDB:9953374 | 28337 |

**Table 1** (continued)

(a) Sequences of the polymorphic IGHV3–30–5 gene cannot be differentiated from those of the IGHV3–30 gene. All sequences are described therefore as 'IGHV3–30 alleles' by comparison to the allele *01 of IGHV3–30 (M83134). However, it is not excluded that some of these 'alleles' belong exclusively to IGHV3–30–5.

(b) Sequences of the polymorphic IGHV4–30–1 gene cannot be differentiated from those of the IGHV4–31 gene. All sequences are described therefore as 'IGHV4–31' alleles by comparison to the allele *01 of IGHV4–31 (L10098). However, it is not excluded that some of these 'alleles' belong exclusively to IGHV4–30–1.

(1) IMGT gene names have been approved by the Human Genome Organization (HUGO) Nomenclature Committee in 1999. Note that, in the HUGO symbols, parentheses of the truncated pseudogene names are omitted, and capital letters replace the lowercase letters found in seven provisional IMGT gene names. Otherwise all the gene names (gene symbols) are identical in IMGT and HUGO nomenclatures.

(2) Gene definitions (full names) are identical (including parentheses) in IMGT and HUGO nomenclatures.

(3) Other entries concerning the IGH locus or groups, in the OMIM, GDB, and LocusLink genome databases, and in HUGO:

| IMGT designation | IMGT definition [a] | OMIM | GDB | LocusLink | HUGO |
|---|---|---|---|---|---|
| IGH locus | Immunoglobulin heavy locus | – | GDB:118731 | 3492 | IGH@ |
| IGHC group | Immunoglobulin heavy constant group | *b* | GDB:9992632 | *b* | *b* |
| IGHD group | Immunoglobulin heavy diversity group | 146910 | GDB:9992625 | 50648 | IGHD@ |
| IGHJ group | Immunoglobulin heavy joining group | 147010 | GDB:120741 | 3506 | IGHJ@ |
| IGHV group | Immunoglobulin heavy variable group | 147070 | GDB:128528 | 3509 | IGHV@ |

*a* Entry definitions are identical in IMGT, GDB, LocusLink and HUGO.

*b* There is no entry for the IGHC group in the OMIM and LocusLink databases and in HUGO. The individual OMIM entries for the IGHC genes are the following: IGHA1: 146900, IGHA2: 147000, IGHD: 147170, IGHE: 147180, IGHG1: 147100, IGHG2: 147110, IGHG3: 147120, IGHG4: 147130, IGHM: 147020. Individual GDB and LocusLink entries and HUGO symbols for the IGHC genes are reported in table 1.

The human IGH locus at 14q32.33 spans 1,250 kilobases (fig. 2). It consists of 123–129 IGHV genes [5, 7, 24–29], depending on the haplotypes, 27 IGHD genes [6, 30–33] belonging to 7 subgroups, 9 IGHJ genes [6, 34] and, in the most frequent haplotype, 11 IGHC genes [35–49] (table 1). Eighty-two to 88 IGHV genes belong to 7 subgroups, whereas 41 pseudogenes, which are too divergent to be assigned to subgroups, have been assigned to the 4 clans (see header of table 1). Seven unmapped IGHV genes have been described as insertion/deletion polymorphisms but have not yet been precisely located. Most 5′ IGHV genes occupy a position very close to the chromosome 14q telomere, whereas the IGHC genes are in a more centromeric position. The potential genomic IGH repertoire is more limited, since it comprises 38–46 functional IGHV genes belonging to 6 or 7 subgroups, depending on the haplotypes [7], 23 IGHD, 6 IGHJ and, in the most frequent haplotype, 9 IGHC genes. Allelic IGHC multigene deletions [36, 37, 50–55], duplications or triplications [56–59] have been described in healthy individuals. I to VI refer to the type of deletion [for a review, see ref. 55], duplication or triplication, defined by the deleted, duplicated or triplicated genes (fig. 2). For example, deletion I [36, 37, 50] is designated as del G1-EP1-A1-GP-G2-G4. The number of IGHC genes may vary from 5 in deletion I to probably up to 19 in triplication III. Duplicated and triplicated genes have, in all cases but one, not been sequenced.

## Orphons and Processed Gene

Thirty-five IGH genes have been found outside the main locus in other chromosomal localizations (table 2). These genes, designated as orphons, cannot contribute to the synthesis of the immunoglobulin chains, even if they have an open reading frame (ORF). Nine IGHV orphons and 10 IGHD orphons have been described on chromosome 15 (15q11.2), and 16 IGHV orphons on chromosome 16 (16p11.2) [5]. In addition, 1 IGHC processed gene, IGHEP2, is localized on chromosome 9 (9p24.2–p24.1) [60]. This is so far the only processed Ig gene which has been described.

## IGH Gene Nomenclature and IMGT Scientific Chart

*Gene Names*

Gene names (tables 1, 2) are designated according to the IMGT gene name nomenclature for Ig and T cell receptors (TcRs) of all vertebrates based on the 'CLASSIFICATION' concept of the IMGT-ONTOLOGY [61] (appendix 1) and according to the rules of the IMGT Scientific chart [18, 20], available at **http://imgt.cines.fr:8104**. IMGT gene names and IMGT gene definitions for the human Ig [62] and TcR genes [63] were approved by the HUGO Nomenclature Committee in 1999. Note that in the HUGO symbols **(http://www.gene.ucl.ac.uk/nomenclature)**, slashes and parentheses are omitted and capital letters replace the lowercase letters found in some provisional IMGT gene names. Otherwise, the gene symbols and all the full names (including slashes and parentheses) are identical in IMGT and HUGO nomenclatures.

*Functionality*

Criteria of functionality (F = functional; P = pseudogene; ORF = open reading frame) (tables 1 and 2) are described in the IMGT Scientific chart [1]. The definition of functionality is based on the sequence analysis. As examples, the instances functional (for germline V, D, J and for C sequences) mean that the coding regions have an ORF without a

stop codon, and that there is no described defect in the splicing sites and/or recombination signals and/or regulatory elements. According to the gravity of the identified defects, the functionality can be defined as ORF, pseudogene or vestigial (for germline V, D, J and for C genes) [1]. Complete definitions are available from the IMGT Scientific chart at the IMGT Marie-Paule page. Information on gene rearrangement, DNA transcription into mRNA and RNA translation into a polypeptide chain is provided in the IMGT 'Germline gene tables' in the IMGT Repertoire (columns designated as R, T and Pr, respectively), and has been published in a previous 'IMGT Locus in Focus' report [5]. This information is extracted from the literature and through an IMGT/LIGM-DB sequence database search [19, 20]. The IMGT/V-QUEST tool, available at the IMGT home page at **http://imgt.cines.fr:8104**, allows the identification of the germline IGHV, IGHD and IGHJ genes from IGHV-D-J genomic rearrangements and transcripts, and provides translation and two-dimensional representation (Collier de Perles) of the variable regions [19, 20, 64].

*Reference Sequences*

For each gene, an IMGT reference sequence accession number is given (tables 1 and 2). For the functional or ORF genes, the IMGT reference sequence accession number is that corresponding to the allele *01. Note that the number *01 does not necessarily mean that other alleles are already known, but it signifies that any new polymorphic sequence will be described by comparison to that allele *01. Although the IMGT accession numbers are the same as those from the EMBL/GenBank/DDBJ generalist databases, the content of the IMGT/LIGM-DB flat files differs in terms of the expert annotations added by IMGT.

*Alleles*

The number of alleles of the human IGHV, IGHD, IGHJ and IGHC genes (tables 1 and 2) is determined according to 'Tables of alleles' and 'Alignments of alleles', in the IMGT Repertoire, at **http://imgt.cines.fr:8104**. A dash (–) indicates that allele polymorphism of the pseudogenes has not been studied. Alignments of all known germline functional and ORF sequences assigned to the different alleles, by comparison to the allele *01, are displayed in another source [17]. Human IGH entries in this source include 99 genes and 285 alleles, with a total of 482 sequences [17].

*Genome Database Accession Numbers*

All IMGT/HUGO human IGH gene symbols, full names and reference sequence accession numbers have been entered into the Genome Database GDB, Toronto, Canada **(http://www.gdb.org)**, and into LocusLink at NCBI (National Center for Biotechnology Information), Bethesda, USA **(http://www.ncbi.nlm.nih.gov/LocusLink)**. Accession ID to these genome databases are provided in tables 1 and 2. Links to OMIM (Online Mendelian Inheritance in Man) **(http://www.ncbi.nlm.nih.gov/Omim)** are cited when there are existing entries in OMIM. Links to the individual IMGT, GDB and LocusLink gene entries are available from **http://imgt.cines.fr:8104** from IMGT Repertoire > List of human Ig and TcR genes > Immunoglobulins.

## Correspondences between Nomenclatures and Numberings

Correspondences between nomenclatures have previously been reported for the human IGHV [5] and IGHD genes [6].

In order to easily compare sequences of immunoglobulins and TcRs, a unique num-

**Table 2.** Human IGH processed gene on chromosome 9, and IGH orphons on chromosomes 15 and 16

**a** On chromosome 9 at 9p24.2–p24.1

| IMGT gene groups | IMGT gene name | IMGT func-tionality | IMGT reference sequence accession numbers | Number of alleles | IMGT gene definition | GDB accession ID | LocusLink accession ID |
|---|---|---|---|---|---|---|---|
| IGHC | IGHEP2 | P | K01241 | – | Immunoglobulin heavy constant epsilon P2 | GDB:119337 | 3499 |

**b** On chromosome 15 at 15q11.2

| IMGT gene groups | IMGT gene name (1) | IMGT func-tionality | IMGT reference sequence accession numbers | Number of alleles | IMGT gene definition (2) | GDB accession ID (3) | LocusLink accession ID (3) |
|---|---|---|---|---|---|---|---|
| IGHD | IGHD1/OR15–1a | ORF | X55575 | – | Immunoglobulin heavy diversity 1/OR15–1a | GDB:9953376 | 28335 |
| | IGHD1/OR15–1b | ORF | X55576 | – | Immunoglobulin heavy diversity 1/OR15–1b | GDB:9953378 | 28334 |
| | IGHD2/OR15–2a | ORF | X55577 | – | Immunoglobulin heavy diversity 2/OR15–2a | GDB:9953380 | 28333 |
| | IGHD2/OR15–2b | ORF | X55578 | – | Immunoglobulin heavy diversity 2/OR15–2b | GDB:9953382 | 28332 |
| | IGHD3/OR15–3a | ORF | X55579 | – | Immunoglobulin heavy diversity 3/OR15–3a | GDB:9953384 | 28331 |
| | IGHD3/OR15–3b | ORF | X55580 | – | Immunoglobulin heavy diversity 3/OR15–3b | GDB:9953386 | 28330 |
| | IGHD4/OR15–4a | ORF | X55581 | – | Immunoglobulin heavy diversity 4/OR15–4a | GDB:9953388 | 28329 |
| | IGHD4/OR15–4b | ORF | X55582 | – | Immunoglobulin heavy diversity 4/OR15–4b | GDB:9953390 | 28328 |
| | IGHD5/OR15–5a | ORF | X55583 | – | Immunoglobulin heavy diversity 5/OR15–5a | GDB:9953392 | 28327 |
| | IGHD5/OR15–5b | ORF | X55584 | – | Immunoglobulin heavy diversity 5/OR15–5b | GDB:9953394 | 28326 |
| IGHV | IGHV1/OR15–1 | ORF | Z29631 | – | Immunoglobulin heavy variable 1/OR15–1 | GDB:9931784 | 28325 |
| | IGHV1/OR15–2 | P | L25543 | – | Immunoglobulin heavy variable 1/OR15–2 | GDB:9931787 | 28324 |
| | IGHV1/OR15–3 | P, ORF | Z29595 | – | Immunoglobulin heavy variable 1/OR15–3 | GDB:9953396 | 28323 |
| | IGHV1/OR15–4 | P | Z29596 | – | Immunoglobulin heavy variable 1/OR15–4 | GDB:9931789 | 28322 |
| | IGHV1/OR15–5 | P, ORF | Z29633 | – | Immunoglobulin heavy variable 1/OR15–5 | GDB:9953398 | 28321 |
| | IGHV1/OR15–6 | P | Z29634 | – | Immunoglobulin heavy variable 1/OR15–6 | GDB:9931790 | 28320 |
| | IGHV1/OR15–9 | ORF | L25542 | – | Immunoglobulin heavy variable 1/OR15–9 | GDB:9931791 | 28319 |
| | IGHV3/OR15–7 | P, ORF | Z29597 | – | Immunoglobulin heavy variable 3/OR15–7 | GDB:9931792 | 28318 |
| | IGHV4/OR15–8 | ORF | Z29598 | – | Immunoglobulin heavy variable 4/OR15–8 | GDB:9953400 | 28317 |

**Table 2** (continued)

**c** On chromosome 16 at 16p11.2

| IMGT gene groups | IMGT gene name (1) | IMGT func-tionality | IMGT reference sequence accession numbers | Number of alleles | IMGT gene definition (2) | GDB accession ID (3) | LocusLink accession ID (3) |
|---|---|---|---|---|---|---|---|
| IGHV | IGHV1/OR16–1 | P | Z29599 | – | Immunoglobulin heavy variable 1/OR16–1 | GDB:9953402 | 28315 |
| | IGHV1/OR16–2 | P | Z29600 | – | Immunoglobulin heavy variable 1/OR16–2 | GDB:9953404 | 28314 |
| | IGHV1/OR16–3 | P | Z29639 | – | Immunoglobulin heavy variable 1/OR16–3 | GDB:9931793 | 28313 |
| | IGHV1/OR16–4 | P | Z17397 | – | Immunoglobulin heavy variable 1/OR16–4 | GDB:9931794 | 28312 |
| | IGHV2/OR16–5 | ORF | L25544 | – | Immunoglobulin heavy variable 2/OR16–5 | GDB:9931795 | 28311 |
| | IGHV3/OR16–6 | P | L25545 | – | Immunoglobulin heavy variable 3/OR16–6 | GDB:9953406 | 28310 |
| | IGHV3/OR16–7 | P | Z29604 | – | Immunoglobulin heavy variable 3/OR16–7 | GDB:9931796 | 28309 |
| | IGHV3/OR16–8 | ORF | Z29605 | – | Immunoglobulin heavy variable 3/OR16–8 | GDB:9931797 | 28308 |
| | IGHV3/OR16–9 | ORF | Z29606 | – | Immunoglobulin heavy variable 3/OR16–9 | GDB:9953408 | 28307 |
| | IGHV3/OR16–10 | ORF | Z29607 | – | Immunoglobulin heavy variable 3/OR16–10 | GDB:9953410 | 28306 |
| | IGHV3/OR16–11 | P | Z29608 | – | Immunoglobulin heavy variable 3/OR16–11 | GDB:9953412 | 28305 |
| | IGHV3/OR16–12 | ORF | Z29609 | – | Immunoglobulin heavy variable 3/OR16–12 | GDB:9953414 | 28304 |
| | IGHV3/OR16–13 | ORF | Z29610 | – | Immunoglobulin heavy variable 3/OR16–13 | GDB:9953416 | 28303 |
| | IGHV3/OR16–14 | P | Z29611 | – | Immunoglobulin heavy variable 3/OR16–14 | GDB:9953418 | 28302 |
| | IGHV3/OR16–15 | P | L25546 | – | Immunoglobulin heavy variable 3/OR16–15 | GDB:9953420 | 28301 |
| | IGHV3/OR16–16 | P | Z29613 | – | Immunoglobulin heavy variable 3/OR16–16 | GDB:9953422 | 28300 |

(1) Note that in the HUGO symbols, slashes of the orphon names are omitted. Otherwise the gene names (gene symbols) are identical in IMGT and HUGO nomenclatures.
(2) Gene definitions (full names) are identical (including slashes) in IMGT and HUGO nomenclatures.
(3) Other entries concerning the IGH orphons in the OMIM, GDB, and LocusLink genome databases and in HUGO:

| IMGT designation | IMGT definition | OMIM | GDB | LocusLink | HUGO |
|---|---|---|---|---|---|
| IGHD/OR15 | Immunoglobulin heavy diversity orphons on chromosome 15 | 146990 | GDB:119334 | 3496 | IGHDOR15@ |
| IGHV/OR15 | Immunoglobulin heavy variable orphons on chromosome 15 | 600949 | GDB:377735 | 3510 | IGHVOR15@ |
| IGHV/OR16 | Immunoglobulin heavy variable orphons on chromosome 16 | 600949 | GDB:377743 | 3511 | IGHVOR16@ |

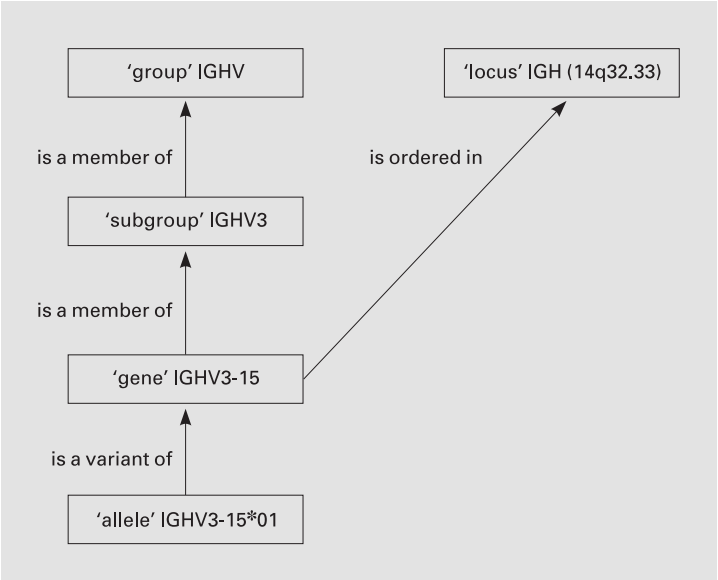


**Fig. 3.** The 'CLASSIFICATION' concept of the IMGT-ONTOLOGY, exemplified for the IGHV genes.

bering has been defined for the variable regions [64, 65]. Correspondence between the IMGT unique numbering and other numberings for the human IGHV genes is available from the IMGT Scientific chart and from a previous report [64]. The IMGT unique numbering relies on the high conservation of the structure of the variable region. This numbering takes into account and combines the definition of the framework (FR) and complementarity determining regions (CDR) [66], structural data from X-ray diffraction studies [67] and the characterization of the hypervariable loops [68]. The unique numbering has allowed the redefinition of the limits of the FR and CDR [64]. The FR-IMGT and CDR-IMGT lengths themselves become crucial information characterizing the variable regions belonging to a group, a subgroup and/or a gene. For example, for a germline gene of the human IGHV1 subgroup, the lengths of the 3 CDR-IMGT, expressed as the number of amino acids, are designated as [8.8.2] (IMGT Repertoire > 2D and 3D structures) [7, 64]. The unique numbering is used as the output of the IMGT/V-QUEST alignment tool, and in the 'Alignments of alleles' (IMGT Repertoire > Proteins and alleles) [17].

Correspondence between the IMGT CH (immunoglobulin heavy constant exon or domain) numberings and the protein EU heavy chain numbering [69] has been reported elsewhere [17].

## Acknowledgments

I thank Gérard Lefranc, Ruth Lovering, Donna Maglott, Chris Porter, Sue Povey, Conover Talbot and all my immunologist colleagues for their contribution to the IMGT Nomenclature Committee. I am grateful to Valérie Contet for the editorial work. IMGT is funded by the European Union's BIOTECH (BIO4CT96-0037) and 5th PCRDT (QLG2-2000-01287) programs, the CNRS (Centre National de la Recherche Scientifique), the Ministère de la Recherche and the Ministère de l'Education Nationale. Subventions have been received from ARC (Association pour la Recherche sur le Cancer) and the Région Languedoc-Roussillon.

## Appendix 1

*The 'CLASSIFICATION' Concept of the IMGT-ONTOLOGY*

The 'CLASSIFICATION' concept of the IMGT-ONTOLOGY (fig. 3) organizes the immunogenetics knowledge useful for the naming and classification of the immunoglobulin genes [63].

'**locus**': A locus is a group of immunoglobulin genes that are ordered and are localized in the same chromosomal location in a given species. The 'locus' IGH (14q32.33) is one of the three main immunoglobulin loci in the human genome. Immunoglobulin genes have also been identified in other chromosomal locations outside the main loci, which represent new instances of the concept locus. However, the genes they contain, designated as orphons, are not functional.

'**group**': A group is a set of genes which share the same 'gene type' (V, D, J or C) and potentially participate in the synthesis of a polypeptide of the same 'chain type'. By extension, a group includes the related pseudogenes and orphons. A 4-letter root designates the 'group', for example, IGHV, IGHD, IGHJ and IGHC for the immunoglobulin heavy genes.

'**subgroup**': A subgroup is a set of genes which belong to the same group, in a given species, and which share at least 75% identity at the nucleotide level (in the germline configuration for V, D and J).

'**gene**': A gene is defined as a DNA sequence that can be potentially transcribed and/or translated (this definition includes the regulatory elements in 5′ and 3′, and the introns, if present). Instances of the 'gene' concept are gene names. By extension, orphons and pseudogenes are also instances of the 'gene' concept. For each gene, IMGT has defined a reference sequence [18]. For the V, D and J genes, the reference sequence corresponds to a germline entity. The rules for the choice of the reference sequences are described at **http://imgt.cines.fr:8104** in the IMGT Scientific chart.

'**allele**': An allele is a polymorphic variant of a gene. Alleles are described, exhaustively and in a standardized way, for the four 'core' coding regions, that is, the germline V-REGIONs, D-REGIONs and J-REGIONs and for the C-REGIONs from immunoglobulin genes. These alleles refer to sequence polymorphisms, with mutations described at the sequence level [1]. Their sequences are compared to the reference sequence designated as *01 (see IMGT Scientific chart at **http://imgt.cines.fr:8104** for IMGT description of mutations and IMGT allele nomenclature for sequence polymorphisms).

## References


1 Lefranc MP: IMGT (ImMunoGeneTics) Locus on Focus. A new section of Experimental and Clinical Immunogenetics. Exp Clin Immunogenet 1998;15:1–7.
2 Pallarès N, Frippiat JP, Giudicelli V, Lefranc MP: The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp Clin Immunogenet 1998;15:8–18.
3 Barbié V, Lefranc MP: The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet 1998;15:171–183.
4 Martinez C, Lefranc MP: The mouse *(Mus musculus)* immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet 1998;15:184–193.
5 Pallarès N, Lefebvre S, Contet V, Matsuda F, Lefranc MP: The human immunoglobulin heavy variable genes. Exp Clin Immunogenet 1999;16:36–60.
6 Ruiz M, Pallarès N, Contet V, Barbié V, Lefranc MP: The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. Exp Clin Immunogenet 1999;16:173–184.
7 Scaviner D, Barbié V, Ruiz M, Lefranc MP: Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp Clin Immunogenet 1999;16:234–240.
8 Folch G, Lefranc MP: The human T cell receptor beta variable (TRBV) genes. Exp Clin Immunogenet 2000;17:42–54.
9 Scaviner D, Lefranc MP: The human T cell receptor alpha variable (TRAV) genes. Exp Clin Immunogenet 2000;17:83–96.
10 Scaviner D, Lefranc MP: The human T cell receptor alpha joining (TRAJ) genes. Exp Clin Immunogenet 2000;17:97–106.
11 Folch G, Lefranc MP: The human T cell receptor beta diversity (TRBD) and beta joining (TRBJ) genes. Exp Clin Immunogenet 2000;17:107–114.
12 Artero S, Lefranc MP: The Teleostei immunoglobulin heavy IGH genes. Exp Clin Immunogenet 2000;17:148–161.
13 Artero S, Lefranc MP: The Teleostei immunoglobulin light IGL1 and IGL2 V, J and C genes. Exp Clin Immunogenet 2000;17:162–172.
14 Folch G, Scaviner D, Contet V, Lefranc MP: Protein displays of the human T cell receptor alpha, beta, gamma and delta variable and joining regions. Exp Clin Immunogenet 2000;17:205–215.

15 Bosc N, Lefranc MP: The mouse *(Mus musculus)* T cell receptor beta variable (TRBV), diversity (TRBD) and joining (TRBJ) genes. Exp Clin Immunogenet 2000;17:216–228.
16 Bosc N, Contet V, Lefranc MP: The mouse *(Mus musculus)* T cell receptor delta variable (TRDV), diversity (TRDD) and joining (TRDJ) genes. Exp Clin Immunogenet 2001;18: 51–58.
17 Lefranc MP, Lefranc G: The Immunoglobulin FactsBook. London, Academic Press, in press.
18 Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaître M, Malik A, Barbié V, Chaume D: IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 1999;27: 209–212.
19 Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bödmer J, Marsh S, Bontrop R, Lemaître M, Lefranc G, Chaume D, Lefranc MP: IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 2000;28:219–221.
20 Lefranc MP: IMGT ImMunoGeneTics database. Int BIOforum 2000;4:98–100.
21 Croce CM, Shander M, Martinis J, Cicurel L, D'Ancona GG, Dolby TW, Koprowski H: Chromosomal location of the genes for human immunoglobulin heavy chains. Proc Natl Acad Sci USA 1979;76:3416–3419.
22 Kirsch IR, Morton CC, Nakahara K, Leder P: Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. Science 1982;216: 301–303.
23 McBride OW, Battey J, Hollis GF, Swan DC, Siebenlist U, Leder P: Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res 1982;10:8155–8170.
24 Shin EK, Matsuda F, Nagaoka H, Fukita Y, Imai T, Yokoyama K, Soeda E, Honjo T: Physical map of the 3′ region of the human immunoglobulin heavy chain locus: Clustering of autoantibody-related variable segments in one haplotype. EMBO J 1991;10:3641–3645.
25 Matsuda F, Shin EK, Nagaoka H, Matsumura R, Haino M, Fukita Y, Taka-ishi S, Imai T, Riley JH, Anand R, Soeda E, Honjo T: Structure and physical map of 64 variable segments in the 3′ 0.8-megabase region of the human immunoglobulin heavy-chain locus. Nat Genet 1993; 3:88–94.
26 Cook GP, Tomlinson IM, Walter G, Riethman H, Carter NP, Buluwela L, Winter G, Rabbitts TH: A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat Genet 1994;7:162–168.
27 Matsuda F, Ishii K, Bouvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T: The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med 1998;188:2151–2162.
28 Matsuda F, Honjo T: The human immunoglobulin VH locus. Immunologist 1999;7:171–176.
29 Lefranc MP: Locus maps and genomic repertoire of the human Ig genes. Immunologist, in press.
30 Siebenlist U, Ravetch JV, Korsmeyer S, Waldmann T, Leder P: Human immunoglobulin D segments encoded in tandem multigenic families. Nature 1981;294:631–635.
31 Buluwela L, Albertson DG, Sherrington P, Rabbitts PH, Spurr N, Rabbitts TH: The use of chromosomal translocations to study human immunoglobulin gene organization: Mapping DH segments within 35 kb of the C mu gene and identification of a new DH locus. EMBO J 1988;7:2003–2010.
32 Ichihara Y, Matsuoka H, Kurosawa Y: Organization of human immunoglobulin heavy chain diversity gene loci. EMBO J 1988;7:4141–4150.
33 Corbett SJ, Tomlinson IM, Sonnhammer ELL, Buck D, Winter G: Sequence of the human immunoglobulin diversity (D) segment locus: A systematic analysis provides no evidence for the use of DIR segments, inverted D segments, 'minor' D segments or D-D recombination. J Mol Biol 1997;270:587–597.
34 Ravetch JV, Siebenlist U, Korsmeyer S, Waidmann T, Leder P: Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. Cell 1981;27:583–591.
35 Flanagan JG, Rabbitts TH: Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing γ, ε and α genes. Nature 1982;300:709–713.
36 Lefranc MP, Lefranc G, Rabbitts TH: Inherited deletion of immunoglobulin heavy chain constant region genes in normal human individuals. Nature 1982;300:760–762.
37 Lefranc MP, Lefranc G, de Lange G, Out TA, van den Broek PJ, van Nieuwkoop J, Radl J, Helal Chaabani H, van Loghem E, Rabbitts TH: Instability of the human immunoglobulin heavy chain constant region locus indicated by different inherited chromosomal deletions. Mol Biol Med 1983;1:207–217.
38 Rabbitts TH, Forster A, Milstein CP: Human immunoglobulin heavy chain genes: Evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences. Nucleic Acids Res 1981;9:4509–4524.
39 White MB, Shen AL, Word CJ, Tucker PW, Blattner FR: Human immunoglobulin D: Genomic sequence of the delta heavy chain. Science 1985;228:733–737.
40 Huck S, Fort P, Crawford DH, Lefranc MP, Lefranc G: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: Comparison with the other human C gamma genes. Nucleic Acids Res 1986;14:1779–1789.
41 Huck S, Lefranc G, Lefranc MP: A human immunoglobulin IGHG3 allele (Gmb0,b1,c3,c5,u) with an IGHG4 converted region and three hinge exons. Immunogenetics 1989; 30:250–257.
42 Ellison JW, Berson BJ, Hood LE: The nucleotide sequence of a human immunoglobulin Cγ1 gene. Nucleic Acids Res 1982;10:4071–4079.
43 Max EE, Battey J, Ney R, Kirsch IR, Leder P: Duplication and deletion in the human immunoglobulin ε genes. Cell 1982;29:691–699.

44 Bensmana M, Huck S, Lefranc G, Lefranc MP: The human immunoglobulin pseudo-gamma IGHGP gene shows no major structural defect. Nucleic Acids Res 1988;16: 3108.
45 Ellison JW, Hood LE: Linkage and sequence homology of two human immunoglobulin γ heavy chain constant region genes. Proc Natl Acad Sci USA 1982;79:1984–1988.
46 Ellison JW, Buxbaum J, Hood L: Nucleotide sequence of a human immunoglobulin Cγ4 gene. DNA 1981;1:11–18.
47 Flanagan JG, Rabbitts TH: The sequence of a human immunoglobulin epsilon heavy chain constant region gene, and evidence for three non-allelic genes. EMBO J 1982;1:655–660.
48 Flanagan JG, Lefranc MP, Rabbitts TH: Mechanisms of divergence and convergence of the human immunoglobulin alpha 1 and alpha 2 constant region gene sequences. Cell 1984;36:681–688.
49 Bensmana M, Chuchana P, Lefranc G, Lefranc MP: Sequence of the CH1 and hinge-CH2 exons of the human immunoglobulin IGHA2 A2m(2) allele: Comparison with the nonallelic and allelic IGHA genes. Cytogenet Cell Genet 1991;56:128.
50 Lefranc G, Chaabani H, Van Loghem E, Lefranc MP, De Lange G, Helal AN: Simultaneous absence of the human IgG1, IgG2, IgG4 and IgA1 subclasses: Immunological and immunogenetical considerations. Eur J Immunol 1983;13:240–244.
51 Migone N, Oliviero S, de Lange G, Delacroix DL, Boschis D, Altruda F, Silengo L, DeMarchi M, Carbonara AO: Multiple gene deletions within the human immunoglobulin heavy-chain cluster. Proc Natl Acad Sci USA 1984;81:5811–5815.
52 Lefranc MP, Lefranc G: Human immunoglobulin heavy-chain multigene deletions in healthy individuals. FEBS Lett 1987;213:231–237.
53 Lefranc MP, Hammarstrom L, Smith CI, Lefranc G: Gene deletions in the human immunoglobulin heavy chain constant region locus: Molecular and immunological analysis. Immunodefic Rev 1991;2:265–281.
54 Keyeux G, Lefranc G, Lefranc MP: A multigene deletion in the human IGH constant region locus involves highly homologous hot spots of recombination. Genomics 1989;5: 431–441.
55 Wiebe V, Helal A, Lefranc MP, Lefranc G: Molecular analysis of the T17 immunoglobulin CH multigene deletion (del A1-GP-G2-G4-E). Hum Genet 1994;93:520–528.
56 Bottaro A, Cariota U, DeMarchi M, Carbonara AO: Pulsed-field electrophoresis screening for immunoglobulin heavy-chain constant-region (IGHC) multigene deletions and duplications. Am J Hum Genet 1991; 48:745–756.
57 Bottaro A, Gallina R, Brusco A, Cariota U, Boccazzi C, Barilaro MR, Plebani A, Ugazio AG, van Leeuwen AM, DeLange GG, et al: Familial clustering of IGHC deletions and duplications: Functional and molecular analysis. Immunogenetics 1993;37:356–363.
58 Brusco A, Cariota U, Bottaro A, Boccazzi C, Delange G, Van Leeuwen AM, Galanello R, Plebani A, Ugazio AU, Guerra MG, et al: Deletions, duplications and triplications of the IGHC region in different Italian populations. Immunodeficiency 1993;4:243–244.
59 Brusco A, Cariota U, Bottaro A, Boccazzi C, Plebani A, Ugazio AG, Galanello R, van Leeuwen AM, DeLange GG, Depelchin S, et al: Structural and immunologic analysis of gene triplications in the Ig heavy chain constant region locus. J Immunol 1994;152:129–135.
60 Battey J, Max EE, McBride WO, Swan D, Leder P: A processed human immunoglobulin ε gene has moved to chromosome 9. Proc Natl Acad Sci USA 1982;79:5956–5960.
61 Giudicelli V, Lefranc MP: Ontology for immunogenetics: The IMGT-ONTOLOGY. Bioinformatics 1999;12:1047–1054.
62 Lefranc MP: Nomenclature of the human immunoglobulin genes. Curr Protoc Immunol, in press.
63 Lefranc MP: Nomenclature of the human T cell receptor genes. Curr Protoc Immunol, in press.
64 Lefranc MP: The IMGT unique numbering for immunoglobulins, T cell receptors and Ig-like domains. Immunologist 1999;7:132–136.
65 Lefranc MP: Unique database numbering system for immunogenetic analysis. Immunol Today 1997;8: 509.
66 Kabat EA, et al: Sequences of Proteins of Immunological Interest. Washington, Public Health Service, 1987.
67 Satow Y, Cohen GH, Padlan EA, Davies DR: Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 A. J Mol Biol 1986;190:593–604.
68 Chothia C, Lesk AM: Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987;196:901–917.
69 Edelman GM, Cunningham BA, Gall WE, Gottlieb PD, Rutishauser U, Waxdal MJ: The covalent structure of an entire gammaG immunoglobulin molecule. Proc Natl Acad Sci USA 1969;63:78–85.