Immunoglobulin heavy variable (IGHV) genes and alleles: new entities, new names and implications for research and prognostication in chronic lymphocytic leukaemia

Aliki Xochelli · Andreas Agathangelidis · Ioannis Kavakiotis · Evangelia Minga ·
Lesley Ann Sutton · Panagiotis Baliakas · Ioanna Chouvarda · Véronique Giudicelli ·
Ioannis Vlahavas · Nikos Maglaveras · Lisa Bonello · Livio Trentin ·
Alessandra Tedeschi · Panagiotis Panagiotidis · Christian Geisler ·
Anton W. Langerak · Sarka Pospisilova · Diane F. Jelinek · David Oscier ·
Nicholas Chiorazzi · Nikos Darzentas · Fred Davi · Paolo Ghia ·
Richard Rosenquist · Anastasia Hadzidimitriou · Chrysoula Belessi ·
Marie-Paule Lefranc · Kostas Stamatopoulos

Abstract Next generation sequencing studies in *Homo sapiens* have identified novel immunoglobulin heavy variable (IGHV) genes and alleles necessitating changes in the international ImMunoGeneTics information system (IMGT) GENE-DB and reference directories of IMGT/V-QUEST. In chronic lymphocytic leukaemia (CLL), the somatic hypermutation (SHM) status of the clonotypic rearranged IGHV gene is strongly associated with patient outcome. Correct determination of this parameter strictly depends on the comparison of the nucleotide sequence of the clonotypic rearranged IGHV gene with that of the closest germline counterpart. Consequently, changes in the reference directories could, in principle, affect the correct interpretation of the IGHV mutational status in CLL. To this end, we analyzed 8066 productive IG heavy chain (IGH) rearrangement sequences from our consortium both before and after the latest update of the IMGT/V-QUEST reference directory. Differences were identified in 405 cases (5 % of the cohort).

Aliki Xochelli and Andreas Agathangelidis are co-first authors.

A. Xochelli · E. Minga · I. Chouvarda · N. Maglaveras ·
A. Hadzidimitriou · K. Stamatopoulos (✉)
Institute of Applied Biosciences, Center for Research and
Technology Hellas, Thessaloniki, Greece
e-mail: kostas.stamatopoulos@gmail.com

A. Agathangelidis · P. Ghia
Unit of B Cell Neoplasia, Division of Molecular Oncology, IRCCS
San Raffaele Scientific Institute, Milan, Italy

I. Kavakiotis · I. Vlahavas
Department of Informatics, Aristotle University of Thessaloniki,
Thessaloniki, Greece

L. A. Sutton · P. Baliakas · R. Rosenquist · A. Hadzidimitriou ·
K. Stamatopoulos
Department of Immunology, Genetics and Pathology, Science for
Life Laboratory, Uppsala University, Uppsala, Sweden

I. Chouvarda · N. Maglaveras
Lab of Medical Informatics, Aristotle University of Thessaloniki,
Thessaloniki, Greece

V. Giudicelli · M.-P. Lefranc
The International ImMunoGeneTics Information System (IMGT),
LIGM, Institut de Génétique Humaine IGH, Université Montpellier
2, Montpellier, France

L. Bonello
Center for Experimental Research and Medical Studies, S. Giovanni
Battista Hospital, Turin, Italy

L. Trentin
Department of Medicine, Hematology and Clinical Immunology
Branch, School of Medicine, Padua University, Padova, Italy

In 291/405 sequences (71.9 %), changes concerned only the IGHV gene or allele name, whereas a change in the percent germline identity (%GI) was noted in 114/405 (28.1 %) sequences; in 50/114 (43.8 %) sequences, changes in the %GI led to a change in the mutational set. In conclusion, recent changes in the IMGT reference directories affected the interpretation of SHM in a sizeable number of IGH rearrangement sequences from CLL patients. This indicates that both physicians and researchers should consider a re-evaluation of IG sequence data, especially for those IGH rearrangement sequences that, up to date, have a GI close to 98 %, where caution is warranted.

Keywords CLL · SHM · New IGHV genes · New IGHV gene alleles · Prognostication

The development of novel high throughput sequencing (HTS) technologies has opened new ways for exploring and understanding the human genome. Recently, studies of genomic contigs obtained by next generation sequencing (NGS) in *Homo sapiens* have identified a series of novel duplicated immunoglobulin (IG) heavy variable (IGHV) genes and new alleles, refining our view of the immunoglobulin heavy chain (IGH) locus (Watson et al. 2013), particularly regarding copy number variation (CNV). IMGT gene and allele names were assigned to the novel IGHV genes and alleles by the IMGT Nomenclature Committee (IMGT-NC) (Lefranc 2014) and have been incorporated in IMGT/GENE-DB (Giudicelli et al. 2005) and in the reference directories of the international ImMunoGeneTics information system® (IMGT®) (http://www.imgt.org) the global reference in Immunogenetics and Immunoinformatics (Lefranc 2014; Lefranc et al. 2009).

Overall, the changes introduced within the last year to the IMGT/V-QUEST reference directory (Brochet et al. 2008) regarding the functional IGHV genes include the following: (i) the enrichment of the IGHV reference dataset through the addition of 5 novel genes and 17 new alleles (6 from the novel genes and 11 from pre-existing IGHV genes), (ii) the assignment of definitive IMGT gene names to 6 previously unmapped IGHV genes, (iii) the update of the reference sequence of 7 IGHV alleles and, (iv) the removal of two IGHV alleles which have become obsolete due to these changes (Table 1). These updates not only contribute to an improved characterization of the expressed repertoire but also raise the issue of whether re-evaluation of previous results is warranted.tgroup

Precise determination of somatic hypermutation (SHM) status strongly depends on alignment of the nucleotide sequence of the rearranged IGHV gene against that of its closest germline counterpart (Ghia et al. 2007; Langerak et al. 2011). Hence, changes in the relevant reference directories could, in principle, affect the IGHV gene and allele assignment and the interpretation of SHM status of IG rearrangement sequences in health and disease, including B cell malignancies, where immunogenetic analysis of the clonotypic IGHV genes has offered valuable insight into their ontogeny (Sutton et al. 2013).

Chronic lymphocytic leukaemia (CLL) amply exemplifies this concept. Indeed, in CLL, the SHM status of the clonotypic rearranged IGHV gene has been established as a strong and accurate prognosticator of patient outcome (Damle et al. 1999; Hamblin et al. 1999). Based on a cutoff value of 98 % germline identity (GI), clonotypic IGHV-IGHD-IGHJ gene rearrangement sequences are defined either as mutated (<98 % GI) or unmutated (≥98 % GI) and patients are, accordingly, assigned to a "mutated" (M-CLL) or "unmutated" (U-CLL) category, each associated with a distinct prognosis (Chiorazzi and

A. Tedeschi
Department of Oncology/Haematology, Division of Haematology, Niguarda Ca' Granda Hospital, Milan, Italy

P. Panagiotidis
First Department of Propaedeutic Medicine, University of Athens, Athens, Greece

C. Geisler
Department of Hematology and Pathology, Rigshospitalet, Copenhagen, Denmark

A. W. Langerak
Department of Immunology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands

S. Pospisilova · N. Darzentas
Central European Institute of Technology, Masaryk University, Brno, Czech Republic

D. F. Jelinek
Department of Immunology, Mayo Clinic, Rochester, MN, USA

D. Oscier
Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK

N. Chiorazzi
Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, NY, USA

F. Davi
Hematology Department and University Pierre et Marie Curie, Hopital Pitie-Salpetriere, Paris, France

P. Ghia
Università Vita-Salute San Raffaele, Milan, Italy

C. Belessi
Hematology Department, Nikea General Hospital, Piraeus, Greece

K. Stamatopoulos
Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece

Table 1 Synopsis of the changes (new genes and alleles and upgrade of previous data) recently introduced into the IMGT/V-QUEST reference directory for the IGHV functional genes based on comparison of the IMGT/V-QUEST reference directory release 201308-3 (20 February 2013) versus the IMGT/V-QUEST reference directory release 201408-4 (20 February 2014)

Type of change	New functional genes and alleles		Upgrade of previous data			
	Enrichment of the IMGT/V-QUEST reference directory		Assignment of definitive IMGT nomenclature		Update of reference sequences	Removal of obsolete alleles
	New genes	New alleles	Temporary gene name	Definitive gene name		
Number of changes	5	17	6		7	2
IMGT genes and alleles	IGHV1-69D	IGHV1-18*04	IGHV1-c	IGHV1-38-4	IGHV1-2*04	IGHV2-5*07
	IGHV2-70D	IGHV1-69*14	IGHV1-f	IGHV1-69-2	IGHV2-5*01	IGHV2-5*10
	IGHV3-23D	IGHV1-69D*01	IGHV3-d	IGHV3-38-3	IGHV2-5*02	
	IGHV3-43D	IGHV2-70D*04	IGHV3-h	IGHV3-69-1	IGHV2-5*04	
	IGHV3-64D	IGHV2-70D*14	IGHV4-b	IGHV4-38-2	IGHV3-23*03	
		IGHV3-9*03	IGHV5-a	IGHV5-10-1	IGHV3-53*02	
		IGHV3-11*06			IGHV4-38-2*02	
		IGHV3-13*05				
		IGHV3-20*02				
		IGHV3-23D*01				
		IGHV3-30-3*03				
		IGHV3-43D*01				
		IGHV3-64D*06				
		IGHV4-4*08				
		IGHV4-28*07				
		IGHV4-30-2*06				
		IGHV4-30-4*07				

Table 2 Overall analysis involved releases 201308-3 (20 February 2013) and 201408-4 (20 February 2014), respectively (http://www.imgt.org/IMGT_vquest/share/textes/datareleases.html). For each analysis a different IMGT/HighV-QUEST version (http://www.imgt.org/HighV-QUEST/help.action?section=upgrades) was available and, thus, applied: namely versions 1.1.2 (01 October 2012) and 1.2.0 (14 November 2013) which in turn correspond to IMGT/V-QUEST program versions 3.2.30 (28 January 2013) and 3.3.0 (20 February 2014) (http://www.imgt.org/IMGT_vquest/share/textes/programversions.html). The different program versions correspond to improvements in the algorithms, the impact of which has not been analysed in this study

IMGT/V-QUEST reference directory release	IMGT/V-QUEST programme version	IMGT/HighV-QUEST programme version
201308-3 (20 February 2013)	3.2.30 (28 January 2013)	1.1.2 (01 October 2012)
201408-4 (20 February 2014)	3.3.0 (20 February 2014)	1.2.0 (14 November 2013)

Ferrarini 2011; Chiorazzi et al. 2005; Damle et al. 1999; Hamblin et al. 1999). For these reasons, CLL is an excellent candidate to explore the biological and clinical impact of recent developments regarding the IGH locus composition.

In order to address this issue, we took advantage of a large dataset of IGHV-IGHD-IGHJ gene rearrangement sequences from patients with CLL consolidated within the context of the multi-institutional IMGT/CLL-DB initiative (http://www.imgt.org/CLLDBInterface/Welcome.do). IGH rearrangement sequences were analyzed through the use of the IMGT/HighV-QUEST portal for NGS immunoglobulin (IG) and T cell receptor (TR) sequences (Alamyar et al. 2012a, b; Li et al. 2013). This tool is capable of analyzing up to 500,000 sequences per batch and provides the same high quality results
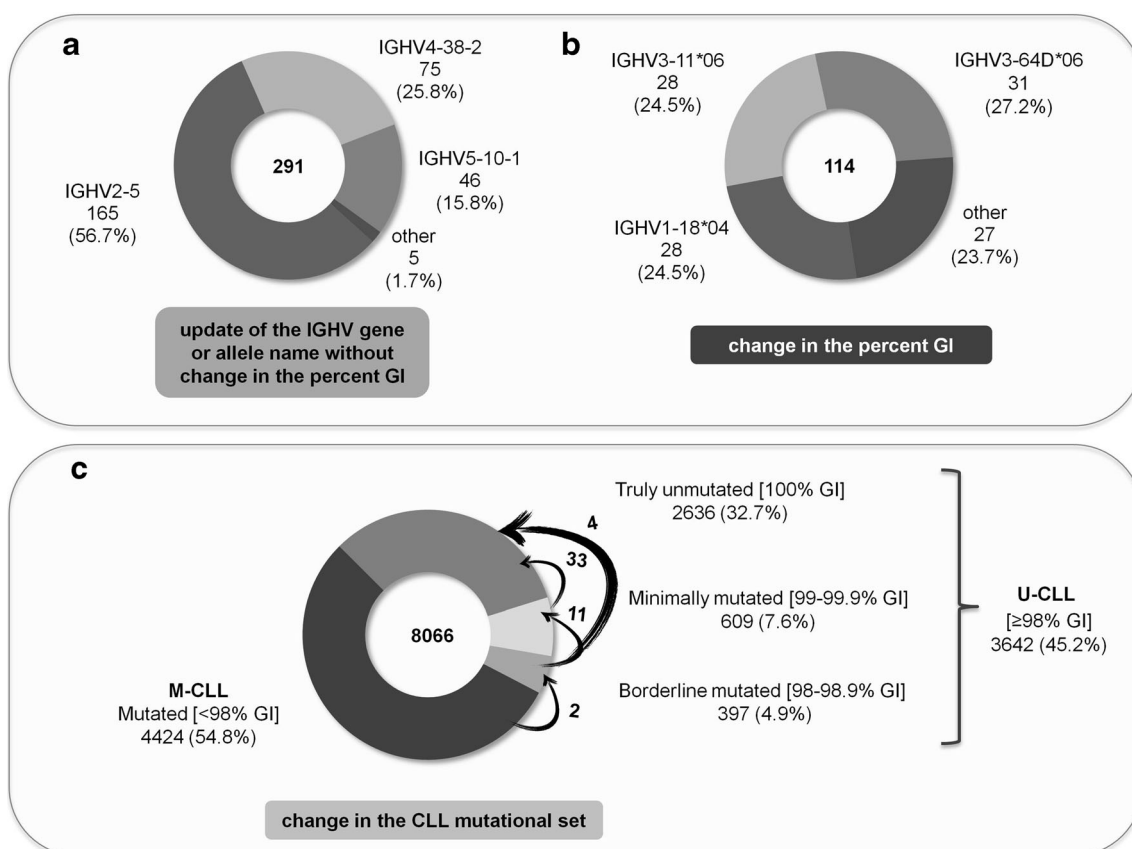

Fig. 1 Impact of novel IGHV genes and alleles on the interpretation of immunogenetic results in CLL. 8066 productive IGHV-IGHD-IGHJ gene rearrangement sequences from the IMGT/CLL-DB initiative were analyzed before and after update of the IMGT/V-QUEST reference directory of IMGT/HighV-QUEST. Differences were identified in a total of 405 sequences (5 % of the cohort) falling into two categories: **a** Change in the IGHV gene or allele name (291 sequences) or **b** Change in the percent GI (114 sequences). **c** Repartition of the mutational sets in the cohort before the update. Changes in the percent GI led to changes in the mutational set (50 sequences) shown with *arrows* (additional information concerning cases with changes in the mutational set or cases that are referred to as "other" can be found in Supplemental file 1)

as IMGT/V-QUEST online. IMGT/HighV-QUEST uses, at a given time, the same tool version and the same reference directory release as IMGT/V-QUEST (Table 2).

Specifically, we analyzed 8066 productive IGHV-IGHD-IGHJ rearrangement sequences from our consortium both before and after the latest update to the IMGT/V-QUEST reference directory (Table 2).

Overall, differences due to the updates of the IMGT/V-QUEST reference directory were identified in a total of 405/8066 sequences (5 % of the cohort) and fell into two categories. The first category (291/405 sequences, 71.9 %) comprises rearranged sequences for which there were changes in the IGHV gene name or allele name without change in the percent GI (as expected for changes only due to the IMGT definitive nomenclature). The gene name changes were practically confined to two genes, IGHV4-38-2 (75/291 sequences, 25.8 %) and IGHV5-10-1 (46/291 sequences, 15.8 %). The allele changes mostly concern updates of the IGHV2-5 gene alleles (165/291 sequences, 56.7 %) (Fig. 1a). The second category (114/405 sequences, 28.1 %) comprises rearranged sequences for which there were changes in the percent GI. These are sequences for which a novel IGHV allele or an updated reference had a closest germline V-region identity; to a large extent, the changes were due to the addition of only three novel IGHV alleles, namely IGHV1-18*04 (28/114 sequences, 24.5 %), IGHV3-11*06 (28/114 sequences, 24.5 %) and IGHV3-64D*06 (31/114 sequences, 27.2 %) assigned to the novel IGHV3-64D gene (Fig. 1b) (Table 2).

In order to analyze if the changes in the percent had impact on the CLL prognosis and following our previously published definitions (Murray et al. 2008), we subdivided the 8066 sequences of the cohort into four different sets based on the level of SHM: truly unmutated (100 % GI) (n=2636, 32.7 %), minimally mutated (99–99.9 % GI) (n=609, 7.6 %), borderline mutated (98–98.9 % GI) (n=397, 4.9 %) and mutated (<98 % GI) (n=4424, 54.8 %). As expected, the recent updates in the IMGT reference directories did not affect the percent GI of those IGHV-IGHD-IGHJ rearranged sequences that were reported as truly unmutated from the previous IMGT/HighV-QUEST analysis. Changes in the percent GI led to a different mutational set assignment for 50/114 sequences (43.8 %) and 50/8066 sequences when considering the cohort (0.61 %). In detail: (i) 2 sequences moved from the mutated to the borderline mutated set, thus shifting from M-CLL to U-CLL following the 2 % GI cutoff value; (ii) 15 sequences moved from the borderline mutated set to either the minimally mutated (11/15) or the truly unmutated (4/15) set; and (iii) finally, 33 sequences moved from the minimally mutated to the truly unmutated set (Fig. 1c) (Supplemental file 1).

The advent of HTS technologies has led to an explosion of information at a scale unprecedented in cancer research. Within the field of immunogenetics, recent updates of the IMGT/V-QUEST reference directories, brought about by haplotype HTS studies, identified novel duplicated IGHV genes (CNV) and new alleles. These recent developments have implications for immunology at large since the reliable identification of SHM is fundamental for the understanding of immune processes occurring in both healthy and pathological states, where even a single change can make a big difference (Barbas et al. 1995; Murray et al. 2008). Therefore, discriminating true SHM from unrecognized polymorphisms is essential for biological inference based on immunogenetics, and as shown here, this requires up-to-date approaches and reference directories (Lefranc 2014). Given the biological and prognostic significance of SHM, our findings indicate that, mainly, researchers and physicians need to be alerted to these developments and consider re-evaluating sequence data, especially for IGHV-IGHD-IGHJ gene rearrangements that up until now were viewed as borderline mutated (GI close to 98 %), where caution is warranted.

References

Alamyar E, Duroux P, Lefranc MP, Giudicelli V (2012a) IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. Methods Mol Biol 882:569–604

Alamyar E, Giudicelli V, Shuo L, Duroux P, Lefranc M (2012b) IMGT(R) web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Res 8:26

Barbas SM, Ditzel HJ, Salonen EM, Yang WP, Silverman GJ, Burton DR (1995) Human autoantibody recognition of DNA. Proc Natl Acad Sci U S A 92:2529–2533

Brochet X, Lefranc MP, Giudicelli V (2008) IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 36:W503–W508

Chiorazzi N, Ferrarini M (2011) Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. Blood 117:1781–1791

Chiorazzi N, Rai KR, Ferrarini M (2005) Chronic lymphocytic leukemia. N Engl J Med 352:804–815

Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M, Chiorazzi N (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94:1840–1847

Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson F, Davi F, Rosenquist R (2007) ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. Leukemia 21:1–3

Giudicelli V, Chaume D, Lefranc MP (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res 33:D256–D261

Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94:1848–1854

Langerak AW, Davi F, Ghia P, Hadzidimitriou A, Murray F, Potter KN, Rosenquist R, Stamatopoulos K, Belessi C (2011) Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. Leukemia 25:979–984

Lefranc MP (2014) Immunoglobulin and T cell receptor genes: IMGT((R)) and the birth and rise of immunoinformatics. Front Immunol 5:22

Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, Regnier L, Ehrenmann F, Lefranc G, Duroux P (2009) IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res 37: D1006–D1012

Li S, Lefranc MP, Miles JJ, Alamyar E, Giudicelli V, Duroux P, Freeman JD, Corbin VD, Scheerlinck JP, Frohman MA, Cameron PU, Plebanski M, Loveland B, Burrows SR, Papenfuss AT, Gowans EJ (2013) IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. Nat Commun 4:2333

Murray F, Darzentas N, Hadzidimitriou A, Tobin G, Boudjogra M, Scielzo C, Laoutaris N, Karlsson K, Baran-Marzsak F, Tsaftaris A, Moreno C, Anagnostopoulos A, Caligaris-Cappio F, Vaur D, Ouzounis C, Belessi C, Ghia P, Davi F, Rosenquist R, Stamatopoulos K (2008) Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. Blood 111:1524–1533

Sutton LA, Agathangelidis A, Belessi C, Darzentas N, Davi F, Ghia P, Rosenquist R, Stamatopoulos K (2013) Antigen selection in B-cell lymphomas—tracing the evidence. Semin Cancer Biol 23:399–409

Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, Willsey AJ, Joy JB, Scott JK, Graves TA, Wilson RK, Holt RA, Eichler EE, Breden F (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. Am J Hum Genet 92:530–546