NOMENCLATURE: AVOIDING BABYLONIAN SPEECH CONFUSION IN PRESENT DAY IMMUNOLOGY

EDITED BY: Menno C. van Zelm, Loems Ziegler-Heitbrock, Andrew M. Collins, Pablo Engel and Sanny K. Chan

PUBLISHED IN: Frontiers in Immunology

IUIS

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88966-445-0
DOI 10.3389/978-2-88966-445-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

NOMENCLATURE: AVOIDING BABYLONIAN SPEECH CONFUSION IN PRESENT DAY IMMUNOLOGY

Topic Editors:
Menno C. van Zelm, Monash University, Australia
Loems Ziegler-Heitbrock, Independent researcher, Germany
Andrew M. Collins, University of New South Wales, Australia
Pablo Engel, University of Barcelona, Spain
Sanny K. Chan, National Jewish Health (United States), United States

We acknowledge the initiation and support of this Research Topic by the International Union of Immunological Societies (IUIS). A/Prof. Menno van Zelm currently serves as the chairman for the IUIS Nomenclature Committee; Prof. Pablo Engel is the chair of the IUIS CD Nomenclature Sub-Committee; Prof. Loems Ziegler-Heitbrock is the chair of the IUIS Monocytes and Dendritic Cells in Blood Sub-Committee; Asst. Prof. Sanny Chan is a member of the WHO / IUIS Allergen Nomenclature Sub-Committee and A/Prof. Andrew Collins is co-chair of the Germline Gene Database (GLDB) Working Group of the Adaptive Immune Receptor Repertoire community (AIRR-C) and chair of the Inferred Allele Review Committee (IARC).

Citation: van Zelm, M. C., Ziegler-Heitbrock, L., Collins, A. M., Engel, P., Chan, S. K., eds. (2021). Nomenclature: Avoiding Babylonian Speech Confusion in Present Day Immunology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-445-0

Table of Contents

frontiers in Immunology

EDITORIAL
published: 14 December 2020
doi: 10.3389/fimmu.2020.621100

Editorial: Nomenclature - Avoiding Babylonian Speech Confusion in Present Day Immunology

Menno C. van Zelm[1,2,3*†], *Loems Ziegler-Heitbrock*[4*†], *Andrew M. Collins*[5*†], *Sanny K. Chan*[6,7*†] *and Pablo Engel*[8*†]

[1] Department of Immunology and Pathology, Central Clinical School, Monash University, Melbourne, VIC, Australia, [2] Department of Allergy, Immunology & Respiratory Medicine, Central Clinical School, Monash University, Melbourne, VIC, Australia, [3] Allergy, Asthma and Clinical Immunology, Alfred Health, Melbourne, IC, Australia, [4] Monocytomics Research, Munich, Germany, [5] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia, [6] Department of Pediatrics, Division of Allergy and Immunology, National Jewish Health, Denver, CO, United States, [7] Department of Biomedical Research, National Jewish Health, Denver, CO, United States, [8] Immunology Unit, Department of Biomedical Sciences, Medical School of the University of Barcelona, Barcelona, Spain

OPEN ACCESS

Edited by:
Harry W. Schroeder, University of Alabama at Birmingham, United States

Reviewed by:
Hans-Martin Jäck, University of Erlangen Nuremberg, Germany

***Correspondence:**
Menno C. van Zelm
menno.vanzelm@monash.edu
Loems Ziegler-Heitbrock
LZH@monocyte.eu
Andrew M. Collins
a.collins@unsw.edu.au
Sanny K. Chan
sannychan.ai@gmail.com
Pablo Engel
pengel@ub.edu

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *25 October 2020*
Accepted: *20 November 2020*
Published: *14 December 2020*

Citation:
van Zelm MC, Ziegler-Heitbrock L, Collins AM, Chan SK and Engel P (2020) Editorial: Nomenclature - Avoiding Babylonian Speech Confusion in Present Day Immunology. Front. Immunol. 11:621100. doi: 10.3389/fimmu.2020.621100

Keywords: B cell, clusters of differentiation marker, allergen, cytokine, complement, C-type lectin receptor, immunoglobulin gene, monocytes and dendritic cells

Editorial on the Research Topic

Nomenclature - Avoiding Babylonian Speech Confusion in Present Day Immunology

INTRODUCTION

The complexity of the immune system at the gene, protein, cell, and organism levels continues to provide a major challenge. Genomic landscaping, single-cell analysis and mass data acquisition including genome, transcriptome, metabolome, and proteome have now added new levels of complexity. With the rapid progress in these and other fields of immunology, it has become more important than ever to agree on uniform nomenclatures, i.e. to agree on how to name novel genes, proteins, cells, and biological reagents.

Names given initially might, in retrospect, not always be logical. For example, tumor necrosis factor (TNF) was named on the basis of the observation of central necrosis in an experimental subcutaneous mouse tumor model (1). It was only after many unsuccessful studies in cancer, that eventually the role of TNF as a master cytokine in inflammation emerged. By that time, it was too late to rename the molecule because that would cause renewed confusion.

Another cytokine has been successfully renamed. Interleukin-6 was initially known as B-cell Stimulatory Factor 2, Cytotoxic T lymphocyte Differentiation Factor, Hybridoma Growth Factor, Hepatocyte Stimulating Factor, and Interferon Beta-2. Obviously, such usage of different names for the same item can lead to confusion and may hinder progress in the field. These two examples demonstrate the need for a consensus nomenclature, which is timely applied.

Indeed, the impacts of early consensus nomenclature are enormous, and the immunological community has an excellent track record of conducting worldwide cooperative efforts on nomenclature issues (2). Remarkable examples of these include nomenclature for antigen receptor (IG and TR) genes (3), cytokines and chemokines and their receptors (4), as well as allergens (5), cell types (6, 7) and the CD nomenclature for monoclonal antibodies (8).

More recently, nomenclature initiatives agreed on an early consensus regarding classification of leukocytes. The monocyte nomenclature proposal in 2010 defined the population of "intermediate monocytes" (7), resulting in >2000 returns for this term under Google Scholar. Furthermore, consensus nomenclature has been defined for innate lymphoid cell (ILC) subsets (6), which will undoubtedly drive discoveries into their roles in tissue homeostasis, morphogenesis, metabolism, repair, and regeneration.

With the many achievements reached in the past 40 decades, there is a wealth of experience to draw upon, especially within the subcommittees of the Nomenclature Committee of the International Union of Immunological Societies (IUIS; https://iuis.org/). The IUIS Nomenclature Committee is fostering nomenclature efforts by providing a platform that currently includes the activities of altogether 11 nomenclature subcommittees (https://iuis.org/committees/nom/). Each subcommittee consists of a representative group of experts in the field, independently decides on nomenclature on the basis of consensus, and is typically endorsed by IUIS and one or more sister societies (e.g. Allergens by AAAI/EAACI/IUIS, Complement by ICS/IUIS).

The present Research Topic aims to highlight the need to address controversies and to stress the importance of nomenclature based on consensus within the immunology community. Twelve articles are included in this Research Topic, and are categorized into the following types: two Original Research (Kalina et al; Magadan et al.), four Reviews (Gunther et al.; Heger et al.; Ohlin et al.; Sanz et al.), two Mini Reviews (Chan et al.; Del Fresno et al.), two Opinions (Hsiao et al.; Zlotnik), and two Perspectives (Bohlson et al.; Busse et al.). The contributions are briefly covered below with the subtopics soluble mediators, cell surface receptors, immunoglobulin genes, cells of the immune system and allergens.

SOLUBLE MEDIATORS

The nomenclature of complement components dates back to the 1960s (9), followed by a formalized nomenclature of the later discovered alternative pathway in 1981 (10). In 2014, an updated nomenclature of these components with inclusion of newly-identified receptors was published (11). Despite these collaborative efforts, in this issue Bohlson et al. identify several unresolved naming issues. Most importantly, in their present proposal, the nomenclature for the cleavage fragments of C2 is brought into line with all other components (C3, C4, C5, Factor B), such that the smaller fragment is now designated "C2a"and the larger fragment "C2b". Additional updates to clusterin, C1 complex activation states recognition molecules (PRMs) and enzymes of the lectin pathway and regulatory proteins of the complement system are proposed as an update to the 2014 Nomenclature.

Nomenclature of soluble mediators of the immune system has been a major challenge, and this was recognized already in the 1970s, resulting in the consensus nomenclature of IL-1 and IL-2. Despite such activities and structured naming of cytokines, chemokines and their receptors, Zlotnik identifies several current issues and challenges. These include the 'neutral' nomenclature of interleukins, which does not relate to their (inflammatory or anti-inflammatory) biological activity, or their evolutionary relationships. Furthermore, the CXC and CC chemokines are well-defined, but challenges have arisen for more recently-identified chemokines that are located in other genomic regions. Thus, in this field there are ongoing nomenclature challenges, which require clear and perhaps updated definitions of what would qualify as a novel cytokine and/or chemokine.

CELL SURFACE RECEPTORS

Since the 1980s, Human Leukocyte Differentiation Antigen (HLDA) workshops have been organized to test and name clusters of antibodies that reacted with a specific antigen (8). These cluster of differentiation (CD) markers provided consistency and uniformity in manuscripts when referring to identical molecules. CD markers have proven critical for the identification and isolation of leukocytes and lymphocyte subsets, the diagnosis and follow-up of hematological malignancies, autoimmune diseases, and immunodeficiencies, and the monitoring of cancer immunotherapy. However, there are important gaps in our knowledge of CD molecule expression profiles, especially because of the major advances in multiparametric flow cytometry over the last 30 years. The paper by Kalina et al. presents a pilot study that shows the expression patterns of CD1 to CD100 on 47 leukocyte subsets from the blood, thymus and tonsil, using highly standardized eight-color flow cytometry. The resulting dataset includes median antibody binding capacities and percentage of positivity for all markers on all subsets, and can be explored online through an interactive CD Maps web resource (www.hcdm.org). The data presented in this paper will provide a better picture of the surfaceome of human leukocytes and increase our understanding of leukocyte biology.

The Ca++-dependent type lectin receptors (C-type lectin receptors; CLRs) offer an example for Babylonian speech confusion. Del Fresno et al. explain that there can be up to seven different names for a single CLR, and that the same name is used for different CLRs between man and mouse. Here they analyzed the frequency of use of the different names in the literature. They suggest the gene name be mentioned for a given CLR plus the most frequently used name in the abstract of every paper on the topic. This recommendation can help to overcome the nomenclature confusion in the field.

A nomenclature for adhesion G protein–coupled receptors (ADGR) was published several years ago (12). In this nomenclature the brain-specific angiogenesis inhibitor 1 (BAI1) has been renamed ADGRB1. For this receptor, expression in macrophages had been reported in 2007. Hsiao

et al. revisit this point and they have extracted data on ADGRs from proteome and transcriptome repositories. None of the available data sets contained a signal for ADGRB1 in monocytes/macrophages and this included RNAseq analyses, which can pick up low abundance transcripts. The study contributes to clarification of an important issue in the field of ADGRs.

IMMUNOGLOBULIN GENES

Previously-unreported *IGHV* alleles are often a conspicuous presence in human datasets of rearranged VDJ gene sequences (13, 14), but there has been no mechanism by which they can be named. With support from the ImMunoGeneTics (IMGT) Group, Ohlin et al. describe processes that can now lead to the official naming of such sequences.

Magadan et al. describe a new nomenclature to deal with the complexities of the *IGH* loci of salmonid species. This task is made particularly challenging by the fact that the *IGH* loci of salmonid species are duplicated on separate chromosomes, and both loci can rearrange to form functional VDJ genes. In this study, genomic assemblies of the *IGH* loci of the Atlantic salmon and Rainbow trout have been annotated, and IG genes have been named according to the IMGT positional-within-locus nomenclature rules.

Busse et al. address the nomenclature challenges that arise from structural variation in the *IGH* loci of laboratory mice. The IMGT positional-within-families mouse *IGH* nomenclature is based upon annotations of the *IGH* locus of the C57BL/6 mouse genome reference sequence, but this sequence is remarkably different to the *IGH* loci of other inbred strains (15, 16). Busse and colleagues outline the principles that should guide the development of a new nomenclature to deal with this challenge. They argue in favor of a non-positional nomenclature, for this would facilitate the naming of hundreds of mouse *IGHV* genes that are now known, but which remain unmapped and unnamed. Non-positional nomenclature would also avoid the need for the renaming of some IGHV sequences, when new genome assemblies identify errors in previous gene maps. Such changes have resulted in confusion within the research community in the past.

CELLS OF THE IMMUNE SYSTEM

With the availability of an increasing number of new markers, there is the temptation to define more and more cell subsets. While in the past a bimodal expression of a cell surface molecule on a given leukocyte was considered sufficient to define two subsets, we now require, in addition, a differential transcriptome, differences in function and, for consolidation, informative clinical associations.

The renewed attention to B cells during the last decade has resulted in the identification of many new subsets that are inconsistently defined and named. Thus, there is an urgent need for a consistent nomenclature of human B cells to allow for inter-laboratory interpretability. The very comprehensive review by Sanz et al. presents a unified approach of classification based on phenotypic standardization. The authors propose the use of seven surface markers, using multiparameter flow cytometry, to define a variety of functionally-distinct B-cell populations. They also discuss the need for awareness that not all current surface antigens being utilized for defining distinct B-cell subsets are sufficiently conclusive. This Perspective is meant to initiate a discussion in the B-cell community with the aim to reach an international consensus nomenclature for B cells.

The heterogeneity of monocytes and dendritic cells and the impact of extensive data sets is covered by Gunther et al.. The paper points out that many different myeloid cell types and subsets have been defined on the basis of morphology, cell surface marker expression and function. Much of this has been confirmed by mass cytometry, multicolor flow cytometry, and by single cell sequencing, but additional populations emerged. The pitfalls of these novel approaches, including misclassifications, are discussed and unbiased strategies for future research are presented.

The nomenclature of dendritic cells has been difficult because of a trend to name any cell "dendritic cell", as long as it could induce a T cell response. More recently, a more stringent definition has been used with pDCs, DC1s and DC2s being considered *bona fide* DCs. With respect to DC2s, a detailed analysis has demonstrated there is a subset, which lacks typical DC features but instead shows markers and functions that are characteristic of monocytes/macrophages. Heger et al. review the latest developments in this area and discuss whether or not these cells belong to the monocyte lineage.

ALLERGENS

Since at least the early 1980s leading allergists started to standardize the naming process for protein allergens that cause IgE-mediated reactions in humans. The use of the taxonomic name of the source organism now ensures a consistent nomenclature that enables communication within allergy research and clinical care and with external regulatory bodies (17). Today, applying for a unique name from the WHO/IUIS Allergen Nomenclature Sub-Committee is a critical step prior to the publication of data on a novel allergen (5).

The paper by Chan et al. reviews the current procedures and requirements for the submission of new proteins allergens and the reasons behind recent changes. These changes are related to advances regarding a) the amount and route of exposure that causes a protein to become an allergen; b) the structural biology of allergen subunits and their contribution to larger complex allergen structures; c) non-protein allergens such as complex carbohydrates. This paper will be helpful to colleagues, who plan on submitting new allergens to the Allergen Nomenclature Committee.

Together, the articles in this Research Topic illustrate the ongoing need for active governance of existing and assignment

of new nomenclatures. There is vast experience in the immunological research community to deal with such complex issues. The IUIS Nomenclature committee has a history of bringing leaders in the field together for timely and open discussion, and will remain committed to supporting current and new consensus nomenclature initiatives.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

MZ is supported by the Australian National Health and Medical Research Council (NHMRC; Senior Research Fellowship 1117687) and the Jeffrey Modell Foundation.

ACKNOWLEDGMENTS

The authors thank all authors, reviewers, and guest editors who contributed to the Research Topic Nomenclature: Avoiding Babylonian Speech Confusion in Present Day Immunology.

REFERENCES

1. Carswell EA, Old LJ, Kassel RL, Green S, Fiore N, Williamson B. An endotoxin-induced serum factor that causes necrosis of tumors. *Proc Natl Acad Sci USA* (1975) 72:3666–70. doi: 10.1073/pnas.72.9.3666
2. Mantovani A. Reflections on immunological nomenclature: in praise of imperfection. *Nat Immunol* (2016) 17:215–6. doi: 10.1038/ni.3354
3. Lefranc MP, Lefranc G. IMGT((R)) and 30 Years of Immunoinformatics Insight in Antibody V and C Domain Structure and Function. *Antibodies (Basel)* (2019) 8:1–21. doi: 10.3390/antib8020029
4. IUIS WHO Subcommittee on Chemokine Nomenclature. Chemokine/chemokine receptor nomenclature. *Cytokine* (2003) 21:48–9. doi: 10.1016/S1043-4666(02)00493-3
5. Goodman RE, Breiteneder H. The WHO/IUIS Allergen Nomenclature. *Allergy* (2019) 74:429–31. doi: 10.1111/all.13693
6. Vivier E, Artis D, Colonna M, Diefenbach A, Di Santo JP, Eberl G, et al. Innate Lymphoid Cells: 10 Years On. *Cell* (2018) 174:1054–66. doi: 10.1016/j.cell.2018.07.017
7. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood* (2010) 116:e74–80. doi: 10.1182/blood-2010-02-258558
8. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD Nomenclature 2015: Human Leukocyte Differentiation Antigen Workshops as a Driving Force in Immunology. *J Immunol* (2015) 195:4555–63. doi: 10.4049/jimmunol.1502033
9. Rapp HJ, Borsos T. Complement and Hemolysis. *Science* (1963) 141:738–40. doi: 10.1126/science.141.3582.738
10. Nomenclature of the alternative activating pathway of complement. *J Immunol* (1981) 127:1261–2.
11. Kemper C, Pangburn MK, Fishelson Z. Complement nomenclature 2014. *Mol Immunol* (2014) 61:56–8. doi: 10.1016/j.molimm.2014.07.004
12. Hamann J, Aust G, Arac D, Engel FB, Formstone C, Fredriksson R, et al. International Union of Basic andAdhesion G protein-coupled receptors. *Pharmacol Rev* (2015) 67:338–67. doi: 10.1124/pr.114.009647
13. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* (2010) 184:6986–92. doi: 10.4049/jimmunol.1000445
14. Gadala-Maria D, Gidoni M, Marquez S, Vander Heiden JA, Kos JT, Watson CT, et al. Identification of Subject-Specific Immunoglobulin Alleles From Expressed Repertoire Sequencing Data. *Front Immunol* (2019) 10:129. doi: 10.3389/fimmu.2019.00129
15. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos Trans R Soc London Ser B Biol Sci* (2015) 370:1–16. doi: 10.1098/rstb.2014.0236
16. Watson CT, Kos JT, Gibson WS, Newman L, Deikus G, Busse CE, et al. A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. *Immunol Cell Biol* (2019) 97:888–901. doi: 10.1111/imcb.12288
17. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TA. Allergen nomenclature. *Bull World Health Organ* (1986) 64:767–74.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 18 March 2019
doi: 10.3389/fimmu.2019.00435

Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming

Mats Ohlin[1], Cathrine Scheepers[2,3], Martin Corcoran[4], William D. Lees[5], Christian E. Busse[6], Davide Bagnara[7], Linnea Thörnqvist[1], Jean-Philippe Bürckert[8], Katherine J. L. Jackson[9], Duncan Ralph[10], Chaim A. Schramm[11], Nishanth Marthandan[12], Felix Breden[13], Jamie Scott[14], Frederick A. Matsen IV[10], Victor Greiff[15], Gur Yaari[16], Steven H. Kleinstein[17], Scott Christley[18], Jacob S. Sherkow[19], Sofia Kossida[20], Marie-Paule Lefranc[20], Menno C. van Zelm[21], Corey T. Watson[22] and Andrew M. Collins[23]**

[1] Department of Immunotechnology, Lund University, Lund, Sweden, [2] Center for HIV and STIs, National Institute for Communicable Diseases, Johannesburg, South Africa, [3] Faculty of Health Sciences, School of Pathology, University of the Witwatersrand, Johannesburg, South Africa, [4] Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden, [5] Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, United Kingdom, [6] Division of B Cell Immunology, German Cancer Research Center, Heidelberg, Germany, [7] Department of Experimental Medicine, University of Genoa, Genoa, Italy, [8] BISC Global Inc., Boston, MA, United States, [9] Immunology Division, The Garvan Institute of Medical Research, Darlinghurst, NSW, Australia, [10] Fred Hutchinson Cancer Research Center, Seattle, WA, United States, [11] Vaccine Research Center, National Institutes of Health, Washington, DC, United States, [12] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada, [13] Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada, [14] Department of Molecular Biology and Biochemistry, Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada, [15] Department of Immunology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [16] Faculty of Engineering, Bar Ilan University, Ramat Gan, Israel, [17] Department of Pathology, Yale University, New Haven, CT, United States, [18] Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States, [19] Innovation Center for Law and Technology, New York Law School, New York, NY, United States, [20] IMGT®, The International ImMunoGenetics information system® (IMGT), Laboratoire d'ImmunoGénétique Moléculaire (LIGM), CNRS, Institut de Génétique Humaine, Université de Montpellier, Montpellier, France, [21] Department of Immunology and Pathology, Central Clinical School, The Alfred Hospital, Monash University, Melbourne, VIC, Australia, [22] Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville, KY, United States, [23] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia*

OPEN ACCESS

Edited by:
Patrick C. Wilson, University of Chicago, United States

Reviewed by:
Jiwon Lee, University of Texas at Austin, United States
Masaki Hikida, Akita University, Japan

***Correspondence:**
Mats Ohlin
mats.ohlin@immun.lth.se
Andrew M. Collins
a.collins@unsw.edu.au

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *10 December 2018*
Accepted: *19 February 2019*
Published: *18 March 2019*

Citation:
Ohlin M, Scheepers C, Corcoran M, Lees WD, Busse CE, Bagnara D, Thörnqvist L, Bürckert J-P, Jackson KJL, Ralph D, Schramm CA, Marthandan N, Breden F, Scott J, Matsen FA IV , Greiff V, Yaari G, Kleinstein SH, Christley S, Sherkow JS, Kossida S, Lefranc M-P, van Zelm MC, Watson CT and Collins AM (2019) Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming. Front. Immunol. 10:435. doi: 10.3389/fimmu.2019.00435

Immunoglobulins or antibodies are the main effector molecules of the B-cell lineage and are encoded by hundreds of variable (V), diversity (D), and joining (J) germline genes, which recombine to generate enormous IG diversity. Recently, high-throughput adaptive immune receptor repertoire sequencing (AIRR-seq) of recombined V-(D)-J genes has offered unprecedented insights into the dynamics of IG repertoires in health and disease. Faithful biological interpretation of AIRR-seq studies depends upon the annotation of raw AIRR-seq data, using reference germline gene databases to identify the germline genes within each rearrangement. Existing reference databases are incomplete, as shown by recent AIRR-seq studies that have inferred the existence of many previously unreported polymorphisms. Completing the documentation of genetic variation in germline gene databases is therefore of crucial importance. Lymphocyte receptor genes and alleles are currently assigned by the Immunoglobulins, T cell Receptors and Major Histocompatibility Nomenclature Subcommittee of the International Union of Immunological Societies (IUIS)

and managed in IMGT®, the international ImMunoGeneTics information system® (IMGT). In 2017, the IMGT Group reached agreement with a group of AIRR-seq researchers on the principles of a streamlined process for identifying and naming inferred allelic sequences, for their incorporation into IMGT®. These researchers represented the AIRR Community, a network of over 300 researchers whose objective is to promote all aspects of immunoglobulin and T-cell receptor repertoire studies, including the standardization of experimental and computational aspects of AIRR-seq data generation and analysis. The Inferred Allele Review Committee (IARC) was established by the AIRR Community to devise policies, criteria, and procedures to perform this function. Formalized evaluations of novel inferred sequences have now begun and submissions are invited via a new dedicated portal (https://ogrdb.airr-community.org). Here, we summarize recommendations developed by the IARC—focusing, to begin with, on human IGHV genes—with the goal of facilitating the acceptance of inferred allelic variants of germline IGHV genes. We believe that this initiative will improve the quality of AIRR-seq studies by facilitating the description of human IG germline gene variation, and that in time, it will expand to the documentation of TR and IG genes in many vertebrate species.

Keywords: immunoglobulin, allelic variation, inference, AIRR-seq, IGHV, V(D)J rearrangement

INTRODUCTION

Immunoglobulins (IG) are the main antigen receptors and effector molecules of the B cell lineage, and are expressed either as a component of the membrane-bound B cell receptor (BCR) or as secreted antibodies. They are encoded by large numbers of variable (V), diversity (D), and joining (J) genes, which recombine in developing B cells to generate rearranged V-(D)-J genes. This process, referred to as V-(D)-J rearrangement, occurs at the DNA level and leads to an IG V domain repertoire of immense diversity. The study of such repertoires has recently been revolutionized by high-throughput sequencing (1–4), and this is termed Adaptive Immune Receptor Repertoire (AIRR) sequencing (AIRR-seq). The technical and biological interpretation of AIRR-seq data is facilitated by databases containing reference sequences of all known germline genes (**Figure 1**), but AIRR-seq studies have demonstrated that these databases are presently far from complete (5–8). This compromises analysis of AIRR-seq data in many ways. For example, it can lead to the inaccurate determination of gene utilization frequencies, and the extent to which sequences have been affected by the process of somatic point mutation.

The first complete nucleotide sequence of a human germline heavy chain variable gene was reported in 1980 (9). In 1989 at the Human Gene Mapping (HGM) (10) Workshop in New Haven, starting with the human T cell receptor gamma (TRG) locus genes as a paradigm, the variable, diversity and joining IG and TR genes were officially acknowledged as "genes" just like conventional genes, and under the HGM auspices, IMGT®, the international ImMunoGeneTics information system® (IMGT) was created by University of Montpellier and the Centre National de la Recherche Scientifique (CNRS) (10). Ten years of IMGT biocuration on sequences from human genomic cosmid and artificial chromosome libraries were key to the assembly of the IG loci and their annotation (11–13). The IG and TR gene names, available on the IMGT web site since 1995, were approved by the HUGO Nomenclature Committee (HGNC) in 1999 and are managed by the IMGT Nomenclature Committee (IMGT-NC), the IG, TR and MH nomenclature subcommittee of the International Union of Immunological Societies (IUIS). The functional and open reading frame (ORF) of approved human genes were published with their alleles (203 IG and 168 TR) in two FactsBooks in 2001 (14, 15), and the number of sequences now cataloged by IMGT is shown in **Table 1**.

With this description of the human IG germline genes, the gene identification and mutation description became an integral part of the study of V-(D)-J gene rearrangements. Over the next 20 years, hundreds of thousands of expressed V-(D)-J genes were reported, and dedicated tools and databases were established to facilitate research (10, 16, 17). It soon became possible to compile datasets of hundreds of rearranged human V-(D)-J gene sequences that could be used to analyse the process of V-(D)-J recombination (18, 19). These analyses also demonstrated that such datasets could be used to identify previously unreported allelic variants of known germline IG genes (20).

In 2009, AIRR-seq data were reported for the first time (21, 22). Even in the earliest AIRR-seq studies, thousands of independent V-(D)-J rearrangements could be identified from each subject investigated, and this facilitated the detection of previously unreported polymorphisms (5–8) (**Figure 2**). New allelic variants of IGHV genes were detectable in these AIRR-seq data because the crucial nucleotides that defined these alleles showed up as conspicuous patterns of shared mismatches within alignments to the known germline V gene sequences.

Utilities have now been developed to streamline the identification of allelic variants, and to assign measures of

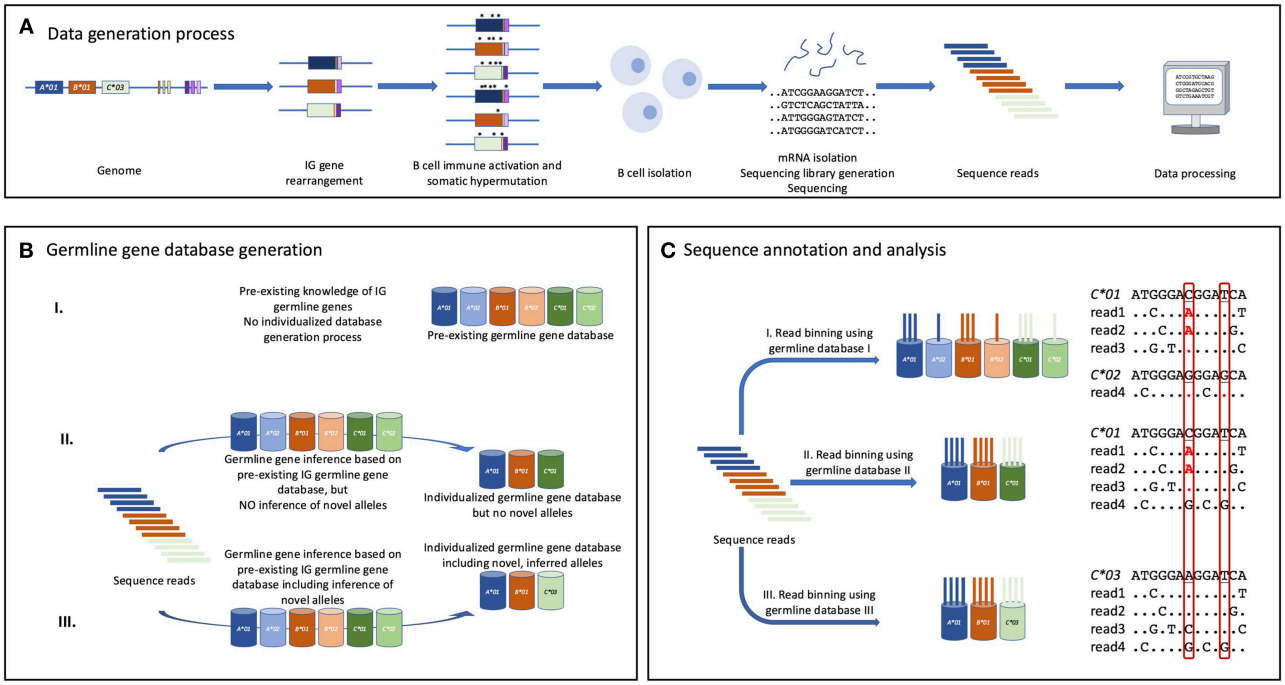

FIGURE 1 | The value of germline IGHV gene inference for detailed AIRR-seq annotation and analysis. **(A)** Germline genes of an individual [here represented by a very limited set of three IGHV genes (A, B, C), and a small number of IGHD (yellow/brown) and IGHJ (purple) genes] are rearranged in cells of the B cell lineage. Following stimulation with antigen many sequences undergo somatic hypermutation and acquire base substitutions (marked *) that may impact subsequent data analysis. An investigated subject's B and plasma cells are collected and typically the cells' transcriptomes are sequenced (e.g., using Illumina MiSeq technology) to generate reads that can be computationally processed. **(B)** A germline IGHV gene database [here represented only by three genes (A, B, C)] will facilitate data analysis, though it is possible to infer genes and alleles without reference to a starting database. The database could be a collection of all known germline IGHV gene alleles (I), or an individualized subset of these (II) that best fits the set of sequence reads that are to be analyzed. Finally, computationally inferred novel germline IGHV gene alleles can be introduced into the individualized germline gene database (III) to even better account for the diversity observed in the experimentally generated sequence dataset. **(C)** Each sequence read is binned to the most appropriate germline gene/allele available in the used germline gene database. If germline gene alleles are present in the database but not in the subject's genotype, some reads will be binned to them as a consequence of base changes introduced by somatic hypermutation (or sequencing errors), resulting in a partial incorrect assignment of germline gene allele origin and consequently of the associated analysis of the mutational pattern. Detailed annotations of part of the sequences are provided for reads binned to alleles of gene *C*. In this example, the investigated subject has an allele (*C*03*) of this gene that is not represented in the original pre-existing germline IGHV gene database. Two bases that differ between one or several alleles in the database and C*03, and thus may be misinterpreted in mutational analysis, are boxed. Unless valid inference of novel germline genes is also performed, the mutational analysis will substantially misinterpret the mutational pattern (highlighted in red letters/dots) targeting this gene. Dots indicate identity to the germline gene to which it is compared.

confidence to each inference (23–27). These utilities employ a variety of inference methodologies, as they have been designed for the analysis of different kinds of data. IgDiscover, for example, is best suited to the analysis of relatively unmutated sequences (23), whereas TIgGER (24) and partis (26) are specifically designed to analyse data that include both unmutated and mutated sequences. To date, 58 sequences have been inferred in this way (see **Table 1**), and can be found in the Immunoglobulin Polymorphism database (IgPdb) (http://cgi.cse.unsw.edu.au/~ihmmune/IgPdb/).

The identification of these previously unreported polymorphisms has remained unknown to many researchers because such variants lie outside the scope of the widely-used IMGT/V-QUEST reference directory of germline sequences (28). This emerged as an early concern of the AIRR Community (https://www.antibodysociety.org/the-airr-community/), a grassroots organization that was founded in 2015 to address the challenges surrounding the generation, analysis and use of AIRR-seq data (29). In 2018, this community formally joined The Antibody Society, a non-profit trade association dedicated to the field of antibody research and immunotherapeutics.

In 2017, the AIRR Community and IMGT agreed to an approach for evaluating the veracity of inferred germline-gene sequences, and for the incorporation of validated sequences into the IMGT Reference Directory. The Germline Database (GLDB) Working Group of the AIRR Community was formed to develop the necessary policies and procedures, and the Inferred Allele Review Committee (IARC) was formed to critically evaluate submitted inferences.

Here we present challenges faced in inferring novel IGHV sequences from AIRR-seq data, and outline strategies for their mitigation. The process for submitting inferred sequences to the IARC is also described. It is our aim that this initiative of the AIRR Community will contribute to a more complete description of human genetic variation, thereby improving the quality of AIRR-seq studies. Human IGHV genes are the focus of this

TABLE 1 | Numbers of human IGHV genes and alleles reported in the IMGT repertoire and in the IgPdb database of inferred alleles.

	IMGT[a]		IgPdb[b]	
Subgroup	**Genes**	**Alleles**	**Genes**	**Alleles**
IGHV1	12	45	8	21
IGHV2	4	29	2	4
IGHV3	30	110	11	18
IGHV4	11	79	8	13
IGHV5	2	9	1	2
IGHV6	1	2	0	0
IGHV7	2	6	0	0

[a] *IMGT genes and allele counts include sequences reported as Functional sequences and Open Reading Frames. The IMGT repertoire was accessed on 11/02/2019.*
[b] *Sequences in IgPdb that have only been identified by genomic sequencing, and sequences that extend previously reported but truncated sequences are not included. Eleven sequences (IGHV1-2*05, IGHV1-2*06, IGHV1-8*03, IGHV1-69*15, IGHV1-69*17, IGHV2-70*15, IGHV3-11*05, IGHV3-11*06, IGHV3-13*05, IGHV3-43D*04 and IGHV3-64D*06) that were first discovered by inference but are now present in the IMGT repertoire are also not included here.*

discussion, though the challenges surrounding the inference of other IG and TR germline genes in human and non-human species are likely to be similar. We anticipate that over time this initiative will expand to the documentation of IG and TR genes in all vertebrate species.

GERMLINE GENE INFERENCES: CHALLENGES, AND STRATEGIES FOR MINIMIZING ERRONEOUS INFERENCES

Reports of inferred antibody sequences have not been immediately and universally accepted, in part because alternative explanations can account for observed nucleotide differences in IG genes (see **Figure 2**). Uniquely, IG genes within activated B cells undergo secondary diversification by somatic hypermutation (SHM) (30). During an immune response, an IGHV gene with a 300 bp length will commonly accumulate 15–20 somatic point mutations (31, 32) and much higher levels of mutations can be observed (33).

The datasets of Sanger sequences that underpinned the first inferred IGHV sequences were very small—in some cases, just six or seven sequences (20). This raised the possibility that these sequences were mutated versions of known alleles. Importantly though, many of the early inferences have now been confirmed by genomic sequencing (20, 34, 35), lending support to the validity of the inference process. Today, the availability of large AIRR-seq datasets gives much greater confidence in the inference process, but challenges remain. These challenges have their origins in the biology of the B cell and of the antibody repertoire, as well as in technical issues affecting the preparation and sequencing of recombined V-(D)-J gene libraries.

The following strategies and tests will aid in the identification of real allelic variants while minimizing the reporting of erroneous inferences.

- Inferences must be made from AIRR-seq data of the highest quality. Experimental strategies to ensure such quality in library generation and sequencing of IG transcripts are now well-established (36, 37), and the assessment of the quality of library generation and of sequencing, using synthetic mRNA spike-ins, is a strategy that can build confidence in inferences made from a dataset (38–41). Proof-reading enzymes with minimal error rates should always be used (42), and putative polymorphisms should be assessed in light of the different types of sequencing errors (base insertions, deletions and substitutions) that are associated with the different sequencing technologies (43). Such errors can be specifically enriched at particular sequence motifs (44), and if these motifs are present in a germline gene, the errors may suggest the existence of a novel allele (7, 45).
- A vital step in the pre-processing of raw sequence data is the removal of reads with a low average quality, but Phred scores should also be assessed for critical nucleotides in individual reads that have contributed to a particular inference. Poor read quality of single nucleotides may result in erroneous inferences (7, 45).
- Correction of sequencing errors and PCR artifacts can be achieved by the use of unique molecular identifiers (UMI). UMIs are introduced during library preparation, labeling each individual transcript prior to amplification. Subsequent consensus building of reads employing identical UMIs can largely remove erroneous bases (46). Technical or biological replicates can also be used to validate sequences and increase confidence that artifacts have been properly discarded.
- Incomplete PCR amplifications create problems. An incompletely amplified product generated in one cycle may later anneal to a similar but distinct template, resulting in the amplification of a hybrid sequence (**Figure 3**) (47, 48). Such chimeric amplification products are often observed in datasets of IG transcripts (49), and unless appropriate filters are applied to AIRR-seq data, these chimeras can masquerade as novel alleles. Preparing libraries with minimally detectable PCR bands helps reduce the problem of chimerism (49, 50), but this strategy is incompatible with some research objectives.
- The detection and elimination of chimeric sequences can be a valuable step in the pre-processing of data. Manual identification of chimeric sequences involves assessment of the distribution of apparent mutations along the length of a sequence. Chimeric sequences often appear to have somatic point mutations clustered at one or the other end of the sequence, and utilities have been developed to automate the detection of sequences with such a non-random distribution of apparent mutations (51).
- Very large AIRR-seq datasets are required if variants of some IGHV genes are to be identified. Reports from analysis of peripheral blood B cells show that usage frequencies of particular IGHV genes in V-(D)-J rearrangements can be as high as 20% for IGHV3-23*01 (52), but as low as 0.01% for rearranged genes incorporating IGHV3-13, IGHV4-28, or IGHV7-81 (5). Rarely utilized IGHV genes will only be present in convincing numbers in the very largest V-(D)-J

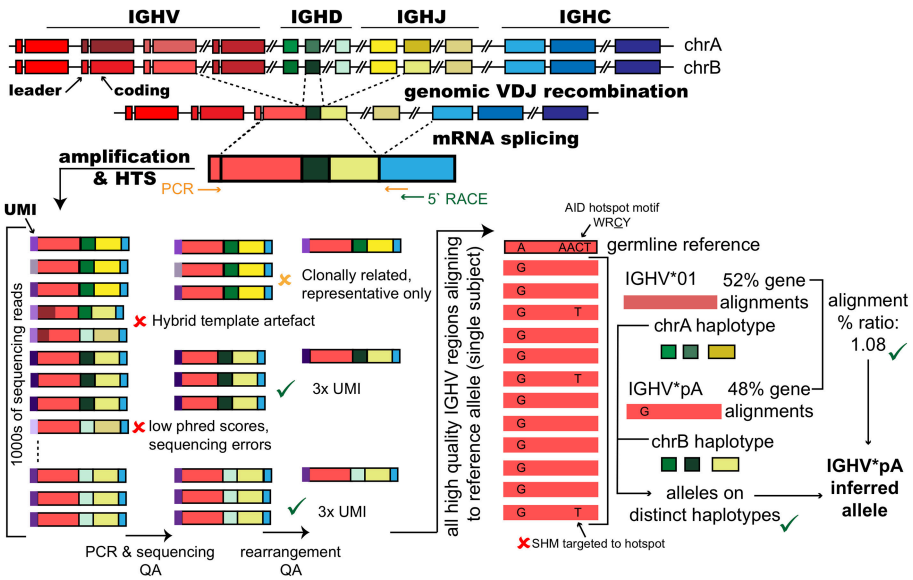

FIGURE 2 | Inference of polymorphic IGHV gene alleles from immunoglobulin repertoire sequencing datasets. The genes that encode immunoglobulin heavy chain variable regions are generated through genomic recombination of single genes of three different types; variable (IGHV), diversity (IGHD), and joining (IGHJ). Each of these gene types are present in the genome as a set of tandem genes that are both polymorphic and polygenic and include approximately 50 IGHV genes, 27 IGHD genes and 6 IGHJ genes. At the mRNA level, splicing joins the rearranged V-(D)-J gene with the IGH constant region genes that confer the isotype to the IG. Repertoire sequencing studies amplify mRNA transcripts, via cDNA, by methods such as targeted PCR using leader region forward primers coupled with reverse primers specific for the IGHC or by using 5′ RACE primed from the IGHC genes. To be suitable for inference, amplification protocols must capture the complete V-(D)-J rearranged coding region and cannot use primer that bind within the V region. Amplification strategies may add unique molecular identifiers (UMIs) as part of cDNA synthesis to tag individual RNA transcripts. High throughput sequencing of a V-(D)-J gene library can generate many thousands of reads for single subjects. Reads are quality assured (QA) to remove sequencing errors; for example, reads with low quality (phred) scores, and PCR artifacts such as hybrid amplicons that have resulted from single stranded DNA from previous PCR cycles acting as primers in future cycles creating hybrid or chimeric templates that are derived from two original amplicons. For inference, IGH from clonally expanded B cells, which have each originated from a single progenitor cell must be reduced to a single representative sequence (for example, the clone member with the fewest mutations) to prevent over-counting. IGH carrying the same UMI can greatly increase confidence that the V-(D)-J rearrangement was in the original pool. In the absence of UMIs, read counts for unique IGH can provide some confidence. Finally, V-(D)-Js are aligned to germline reference datasets that report alleles for the population. Shared "mismatches" relative to the closest germline gene among many sequences from the same subject at position(s) that are not motifs for somatic hypermutation (SHM) in sequences expected to be unmutated (naïve B cells) or have low mutation (IgM+) can be suggestive of the putative allelic variant in the subject's genotype. When a polymorphism is inferred at a heterozygous locus, confidence in the inference can be greatly increased if the putative allele is considered in the context of any other expressed alleles for the gene, with approximately equal expression of the two alleles, and the haplotype, which can phase the gene alleles to their respective chromosomes.

datasets. Large datasets are also needed if the final nucleotides of a germline IGHV sequence are to be determined. The uncertainties surrounding the nucleotides at the 3′ end of the sequence are a consequence of the variability of the gene ends, produced by the processes of exonuclease removal and N nucleotide addition. Biases in these processes can result in the generation of relatively common motifs that may be mistaken for germline-encoded nucleotides (53–56). Of special note, the last base of a germline sequence may not be the most common base in rearranged sequences.

- Somatic point mutations accumulate in IG-encoding genes at a rate of about one mutation per 1,000 bp per cell division within the germinal center reaction (57, 58). The existence of mutational hotspots (59–61) that can target specific germline IGHV genes (62, 63) means that it is inevitable that there will be some shared mutations in any dataset that includes mutated sequences. Some IGHV genes have positions that can be mutated in >30% of class-switched sequences (24). Very high levels of mutation can occur at positions far removed from the regions encoding complementarity determining regions (CDR) of an IGHV sequence, and even at positions outside conventional mutational hotspots (62, 64). For these reasons, inferences of new germline IGHV genes using datasets of mutated sequences are more likely to be erroneous.

Somatic point mutations may be mistaken for germline-encoded nucleotides, but this issue is substantially reduced if sequences are derived from less-mutated cell populations. This can be achieved by the amplification of IgM-encoding transcripts through the use of constant region-specific primers. The issue is partially addressed by the amplification of sequences from sorted B cells displaying a naïve phenotype. More highly mutated datasets can, however, still be the source of reliable inferences if appropriate analytical tools are used. Both the TIgGER and partis software suites, for instance, are designed to use patterns of apparent mutation to infer novel alleles (24, 26). While taking different overall approaches, they both use regression-based statistical tests to identify polymorphisms at positions that

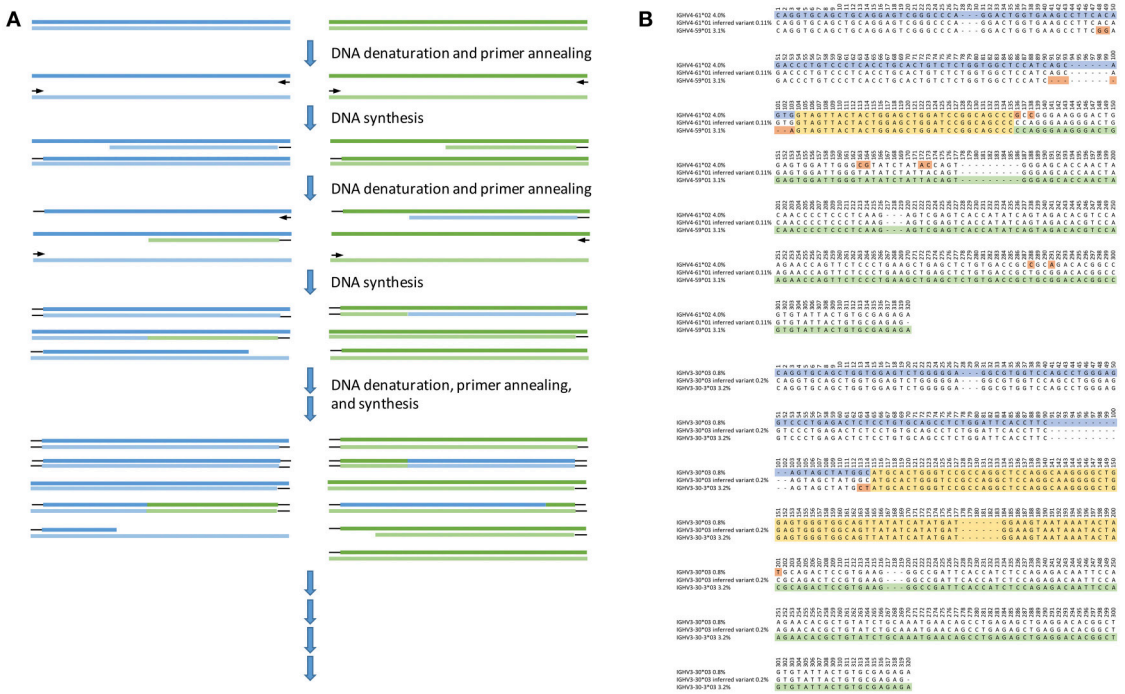

FIGURE 3 | Likely principle of generation of hybrid molecules through cross-over caused by incomplete products during previous PCR cycles **(A)**, and examples of inferred sequences **(B)** that may have an origin in cross-over events during the library generation process. Shown are the inferred sequences (middle) and the parts of other germline sequences that are inferred to be present in the same subject and that through an artifactual cross-over event may have created the inferred novel allele. The region in which cross-over events may have occurred is shown in yellow. The 5′- and 3′- sequences unique to one of the potential donor sequences are shown in blue and green, respectively. Mismatches between the two potential donor sequences are highlighted in orange. The frequency of reads associated with each inference is shown after the name, indicating that some but not all of these inferences may be removed by cut-offs defined in terms of allelic ratios. Base numbering is according to the IMGT numbering scheme.

appear to be recurrently mutated, in sequences that are otherwise relatively unmutated.

- A useful test of an inferred allelic variant is to consider the sequence in the context of other alleles of the gene that may be present in the dataset. For single-copy genes, only two alleles should be present in the genotype of an individual. If an inference suggests that three alleles of a particular gene are present in an individual genotype, the inference should be further investigated. More than two named variants of some genes can be present in an AIRR-seq dataset for gene names without genomic information on the haplotype copy number, as a result of copy number variation (CNV) (5, 6, 8, 35, 46, 65). It is highly likely, for example, that some named allelic variants of the IGHV1-69 gene are actually variants of the duplicate IGHV1-69D gene. Genotypes may also include three or more alleles of one or other of the highly similar IGHV4-4, IGHV4-59, and IGHV4-61 genes, for the genomic location of some sequences associated with these genes is uncertain (7, 45, 52, 66). It seems likely, for example, that the IGHV4-59*08 sequence in some subjects is actually a variant of the IGHV4-61 gene (66). In view of these complications, an evaluation of some inferences must be made with reference to alleles of several genes that may be present in the genotype of the individual. Genomic data of a single cell or individual will remain necessary to unambiguously assign expressed genes with CNV.
- Alleles at heterozygous loci are usually expressed at similar frequencies (52), while inferred sequences suggested by sequencing errors or somatic point mutations are usually present at relatively low frequencies. The calculation of the percentage of alignments to a gene that involve the inferred allele is therefore a simple test that can be used to identify false inferences. Although the IARC will affirm inferred alleles that are observed in just 10% of all alignments to a particular gene, additional analysis may be required to support inferences that imply expression at a low level. We recognize that this will make it more difficult to infer some alleles. CNV, mentioned above, will also complicate the interpretation of this measure of allele expression, whereas recombination signal (RS) sequence variation and other non-coding region variation could lead to abnormal allele expression levels (67). For all these reasons, the allele expression test has limitations.
- The validity of an inference can be demonstrated if all V-(D)-J sequences containing the inference in an AIRR-seq dataset are associated with just one of the two chromosomes of the individual. Such validation can be done using haplotype analysis as outlined in **Figure 4**. This is a method that was developed for human AIRR-seq studies (6, 53, 68), and it is

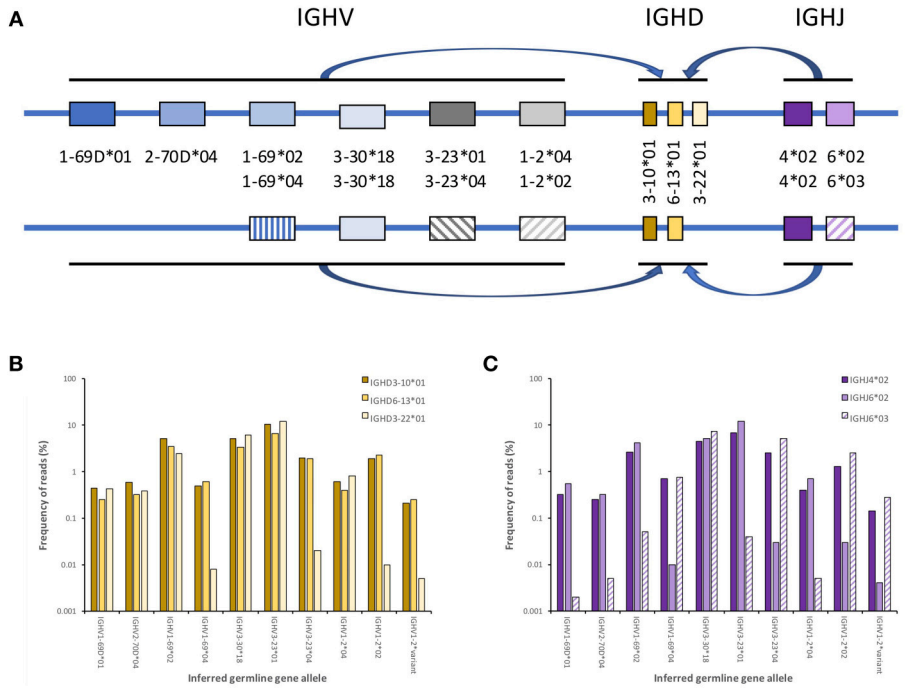

FIGURE 4 | Principle of inference haplotyping defining the AIRR-encoding genes associated with each of the chromosomes carrying the relevant locus, here demonstrated by genes encoding human IG heavy chains. Germline genes involved in V-(D)-J rearrangements are from genes harbored on each chromosome only (i.e., in *cis*), as illustrated with a small number of the genes that actually populate the IGH locus **(A)**. It is possible, using large sequence datasets, to computationally define the association of each IGHV allele to one of the two haplotypes using their association to e.g., different alleles of an IGHJ gene (commonly IGHJ6), if such different alleles are present in the genotype. These alleles serve as anchors in the haplotyping process. An IGHV allele that resides on both chromosomes will rearrange to both alleles of the heterozygous IGHJ gene, whereas IGHV alleles that reside on only one chromosome should primarily be found rearranged to one of the alleles of the heterozygous IGHJ gene **(C)**. In the case of haplotype differentiating expression of IGHD genes (or allelic differences in one or several IGHD genes) these differences can similarly be used as anchors to visualize IGHV allele distributions between haplotypes **(B)**. Such inference can be used to raise confidence in specific allele calls, as incorrectly inferred alleles are likely to associate with the same haplotype as another allele of the same (or a very similar allele) that also exist in the haplotype. This is exemplified here by the haplotyping of an artifactual inference of a novel allele (IGHV1-2*variant) that has a similar association to haplotypes as IGHV1-2*02. [For specific examples, see Kirik et al. (22, 53)].

increasingly being used to support reported inferences (7, 45, 52). Haplotyping is only possible for the validation of IGHV gene inferences in subjects who are heterozygous at IGH loci beyond the IGHV locus region. Anchors for the haplotype inference of IGHV genes are most commonly IGHJ6 alleles (IGHJ6*02 and IGHJ6*03), but heterozygosity at the IGHD2-8 and IGHD2-21 loci can also allow them to be used (7, 45, 52, 68). It is likely that novel long-read high-throughput sequencing platforms will soon make it possible to use IGH constant region genes as haplotype anchors as well.

SUBMISSION OF INFERENCES AND DATA DEPOSITION

IARC and the GLDB WG strive to provide the community with open, transparent and reusable information on inferred genes. To this end, a web-based service termed Open Germline Reference Database (OGRDB) has been set up to facilitate the submission and evaluation of inferences as well as the subsequent retrieval of inferred genes accepted by IARC. In addition, the inferred sequence and the NGS data supporting it have to be deposited in general purpose sequence repositories of the International Nucleotide Sequence Database Collaboration framework to allow re-analysis by third parties and ensure long-term availability of the data. The detailed workflow for data submission is available at OGRDB (https://ogrdb.airr-community.org). In brief, it covers the following steps:

- Verification that the complete raw data of the underlying experiment is available via the Sequence Read Archive (SRA). If possible, the SRA and associated metadata records should be compliant with the Minimal Information about Adaptive Immune Receptor Repertoire (MiAIRR) standard (69).
- Deposition of reads supporting the gene inference to SRA. Note that this submission is in addition to the publication of the complete read data of a given set of experiments.
- Submission of the inferred sequence to GenBank/TPA, depending on the origin of the data on which the inference is based:
  a. First-party data (the inference is performed on one's own datasets) is submitted to GenBank.
  b. Third-party data (inference performed on datasets produced by others) is submitted to GenBank's Third Party Annotation (TPA) section.

- Submission of the inferred sequence and the associated information about the inference procedure as well as the accession IDs of the INSDC submission to IARC via the OGRDB interface.

Each inference must be made from data that originates from a single individual. The standardized submission protocol incorporates metadata related to the individual, as well as to the generation, processing and analysis of the individual's sequences. It also provides data that gives the genotypic context in which an inference should be assessed, and helps identify confounding factors that should be considered.

Currently, data used for germline IGHV gene inference are often generated from PCR-amplified IG transcripts using Illumina's MiSeq technology, as it provides sufficient read length and depth. The IARC will, however, consider inferences and determinations made in other ways. The IMGT-NC requires genomic sequencing of IGHV genes, including the complete leader sequence and associated Recombination Signal sequence (V-RS). Genomic sequences that are not suitable for submission to IMGT-NC will be considered by the IARC if they include the complete IGHV coding region. Partial genomic sequences may also be considered by IARC as evidence in support of an inference from AIRR-seq data. Direct RNA sequencing (70) may also come to play an important role in defining germline IGHV genes in the future.

Inferences must be made from full-length sequencing reads. In contrast, many studies employ primers that anneal within the IGHV sequences themselves, such as the well-validated BIOMED-2 primer set (71). Although sequences generated in this way may be suitable for many research purposes, the partial sequences that can be inferred from such datasets are not suitable for submission to IARC. Submitted sequences must be full-length V-REGION sequences, from base 1 to at least base 318 of the IGHV sequence, according to the IMGT numbering system. Inferences generated using primers that anneal within the sequence should not be submitted to the IARC.

Inference may be carried out using a diversity of computational methodologies. The IARC is agnostic to the investigator's choice of inference methodology as long as it is validated, published, publicly available, and well-documented.

We believe that the identification of dependable, curated gene sets, to which this effort contributes, is a public good. To that end, affirmed sequences, and the submissions that support them are published by IARC under the Creative Commons CC0 license (https://creativecommons.org/publicdomain/zero/1.0/legalcode), allowing their use for any purpose without restriction under copyright or database law.

THE EVALUATION AND DECISION-MAKING PROCESS

The affirmation of submitted inferences requires the unanimous support of the IARC, and this may only be possible after the provision of additional information by the Submitter. The deliberations of this Committee may differ depending on the biological context in which particular sequences are observed and on the process of inference. Particular attention will be paid to:

- The frequency of V-(D)-J rearrangements that include the inferred sequence. Inferences that appear to be very rarely represented in the IG repertoire are at high risk of being incorrect inferences. To guard against this, inferences of sequences that are seen at a frequency of 0.05% or less will not generally be affirmed.
- The number and frequency of unmutated sequences representing the inferred sequence.
- The presence of the inferred IGHV sequence in a diversity of V-(D)-J rearrangements. The sequence needs to be seen in association with different IGHJ genes and in rearrangements with varying CDR3 lengths. This guards against the possibility that sequences that support the inference are clonally-related sequences.
- The number of alleles assigned to the relevant gene or to the set of highly similar genes.
- The distribution of reads between an inferred allele and other alleles of that particular gene, calculated using unmutated sequences. Inferences with low expression frequencies may require additional supporting evidence.
- The outcome of haplotype analysis, where such analysis is possible.
- Evidence that PCR artifacts, such as cross-over events involving other genes and alleles of the subject's genotype, do not explain the inference. Evidence could include a demonstration of the absence of cross-over effects in sequencing libraries of germline gene standards analyzed in parallel to the subject's expressed IG repertoire (38), or demonstration of the systematic identification and removal of sequences with evidence of cross-over effects prior to inference, or analysis of the extent of shared CDR3 sequences between different V-(D)-J gene rearrangements.
- Evidence supporting the reported 3′-end of an inferred germline IGHV gene. The final base of an IGHV gene sequence cannot be inferred with confidence (55, 56) unless additional investigations are undertaken. If a sequence is reported up to and including base 320, the final base will only be affirmed by IARC if supporting analysis is provided.
- Sequencing of part of an inferred allele, from non-B cell genomic DNA.

An assessment will result in one of three outcomes. If a sequence is affirmed as a valid inference, it will be assigned an IARC sequence name and a summary of evidence in support of the inference will be documented in an Inferred Sequence Documentation Sheet. This will be made publicly available at the AIRR community website. It will also be reported to IMGT-NC with an individual GenBank accession number for inclusion in the IMGT Reference Directory. When a sequence is affirmed for the first time, it will be reported as a Level 1 Sequence. If affirmed a second time, it will be reported as a Level 2 Sequence, and if affirmed a third time, it will be reported as a Level 3 Sequence. It is important that researchers continue to notify the IARC of later identification of Level 1 and Level 2 Sequences, so that

they can rise to higher tiers. This will promote acceptance of the inferences within the research community. The IARC will not consider additional inferences of a sequence following its elevation to Level 3.

If evidence in support of a sequence does not reach the level of certainty required for immediate affirmation, the sequence may remain "under review". An Inferred Sequence Documentation Sheet will be completed, and the sequence will be assigned an IARC name, but it will not be publicly reported. Such sequences will be re-assessed if additional supporting information becomes available, or if identical inferences are later submitted to IARC. If a later inference supports the elevation of the sequence to Level 1, the original inference will be credited in the documentation of the sequence.

If there is insufficient evidence to allow a sequence to remain "under review", details of the submission will be retained by IARC, but the submission will not be a part of any future re-assessments.

Inferred alleles will be named using a modification of the IMGT nomenclature (72), incorporating:

- the gene locus (e.g., IGHV, IGKV, IGLV for genes of heavy, kappa light, and lambda light chain loci, respectively);
- the most similar gene at the time of submission in the IMGT/V-QUEST reference directory (28), or in the case of multiple, most-similar genes, using the name with the lowest alphanumeric value;
- an allele number, preceded by an "i" to indicate its discovery by inference. Assigned allele numbers for any gene will be consecutive, and the first inferred allele will be designated the *i01 allele (e.g., IGHV1-2*i01).

A given allele number for a specific gene will be uniquely associated with a specific sequence. If the sequence is incorporated into the IMGT Reference Directory, it will be assigned a new name by IMGT-NC based on the chronological rule and reported to the IUIS/IMGT Nomenclature Committee. The inferred allele name will not be reused and records of the inference will be permanently maintained. Similarly, if evidence emerges suggesting that a particular inference was made in error, the sequence will be removed from any listing of affirmed sequences, but the name and documentation sheets will remain permanently associated with the sequence.

Germline gene databases currently include entries that are incomplete at the 5′ and/or the 3′ end. The inference process could allow the extension of incomplete sequences, as is the case with the sequence IGHV4-4*i01 that is reported here (see **Figure 5**). A sequence of this kind could be a longer representation of the previously reported allele, or it could be a very similar sequence that varies from the original sequence at its ends. The IARC will not attempt to resolve this ambiguity and will simply assign an inferred allele name to the new sequence.

AFFIRMED NOVEL ALLELES

Using the recommendations and policies outlined above, as of August 31, 2018, the IARC has approved five novel alleles at Level 1 (**Figure 5**) and nine inferred alleles remain "under review" (data not shown). Four of the inferred alleles were affirmed from data submitted by the data-generating author (73), of which three were from one donor and one was from a second donor.

IGHV1-2*i01 differs from IGHV1-2*02, its closest matching allele from IMGT, by a single substitution (t163c), resulting in an amino acid change (W55R). Exact matches to the inference were seen in 2.19% of those donor sequences that were determined to be unmutated rearrangements. A second allele for IGHV1-2 (IGHV1-2*04) was observed within the subject's genotype, however IGHV1-2*i01 was seen in 71% of alignments to IGHV1-2. This sequence has been previously described in multiple subjects from AIRR-seq (5, 7, 24), and from genomic DNA (8) and it is listed in the IgPdb database as IGHV1-2*p06. Since this inference was affirmed by IARC, it has been confirmed using full-length genomic DNA sequencing and was recently accepted (24 July 2018) by IMGT-NC as IGHV1-2*06 (Report 2018-1-0724) (http://www.imgt.org/IMGTindex/IMGT-NC.php).

IGHV1-3*i01 was present in 1.17% of the donor's sequences, and differs from IGHV1-3*01 by a single nucleotide (g172a), resulting in an amino acid change (A58T). This sequence has not been observed previously.

IGHV4-30-4*i01 was observed in 1.3% of the donor's sequences, and also has a single nucleotide difference (t120c) compared to its closest matching IMGT allele, IGHV4-30-4*01, however this did not result in an amino acid change. It has been observed in multiple individuals from genomic DNA sequencing (8) and in a single individual from AIRR-seq (63). It was previously listed as IGHV4-30-4*p08 in the IgPDb database.

IGHV4-4*i01 was observed in 0.6% of the donor's sequences. It may be an extension of the existing IGHV4-4*03 allele described in IMGT, involving bases 312-319.

The last of the five affirmed alleles, IGHV3-43D*i01, was submitted as a third party annotation dataset (74) and although it was observed at a low frequency (0.07%) in the subject's repertoire, it could be accepted as a Level 1 sequence. It has been observed previously in multiple individuals from AIRR-seq studies (7), and as genomic DNA (8), and is listed as IGHV3-43*p04 in IgPdb. It has also been observed in a fosmid clone (GenBank: AC242184) that was not annotated in detail. At the time of its acceptance by IARC, this sequence differed from its closest matching IMGT sequence IGHV3-43D*01 (now renamed as IGHV3-43D*03) by a single nucleotide (c195a), however this does not result in an amino acid change. Since the affirmation by IARC of this novel inferred allele, it has been accepted (October 4, 2018) by IMGT as IGHV3-43D*04, based on genomic evidence.

For all five affirmed alleles, the genotype and allele frequencies were within the IARC guidelines. Where possible, haplotype analysis confirmed the validity of the inferences, and cross-over artifacts were ruled out. The Inference Documentation Sheets for these inferred alleles can be found at the OGRDB website (https://ogrdb.airr-community.org).

CONCLUSION

Germline IGHV, IGHD, and IGHJ genes constitute the building blocks of IG V domain diversity, and so have a direct bearing on the functional B cell immune response. The formation of

A

Name of inferred allele	Closest allele in IMGT database at the time of inference	Modification relative to closest IMGT allele	Other names associated with the sequence	Frequency in donor's repertoire	Allele Percentage
*IGHV1-2*i01*	*IGHV1-2*02*	T163C	I*GHV1-2*p06; IGHV1-2*06*	2.2%	71%
*IGHV1-3*i01*	*IGHV1-3*01*	G172A	none	1.2%	57%
*IGHV4-30-4*i01*	*IGHV4-30-4*01*	T120C	*IGHV4-30-4*p08*	1.3%	100%
*IGHV4-4*i01*	*IGHV4-4*03*	Extension of sequence with bases 312-319	none	0.6%	25%
*IGHV3-43D*i01*	*IGHV3-43D*01*	C195A	*IGHV3-43*p04; IGHV3-43D*04*	0.07%	100%

B

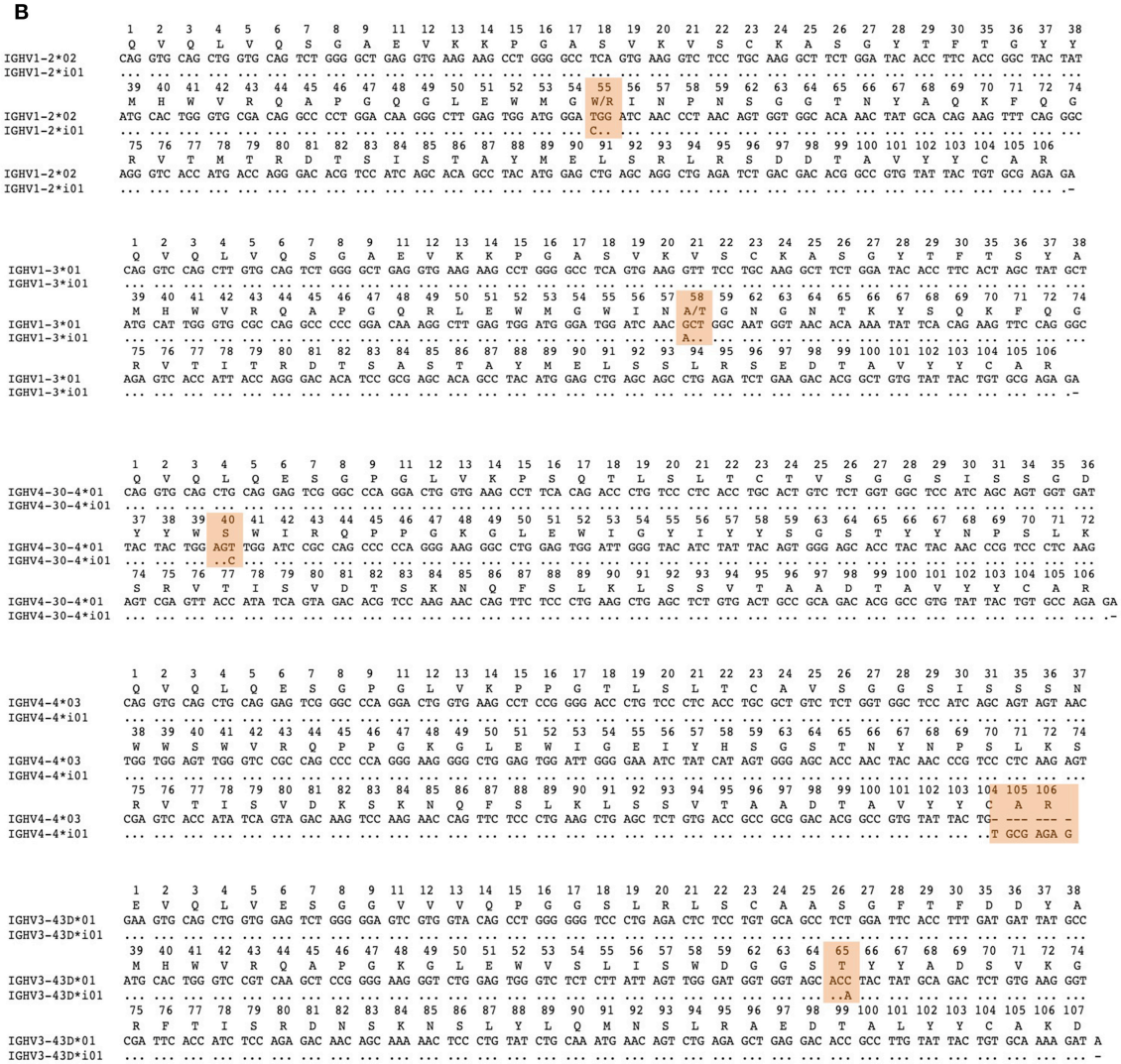

FIGURE 5 | Affirmed inferred alleles. **(A)** Table of inferred alleles. Shown are the names given to the inferred sequences by IARC; the closest matching IMGT alleles; the genetic differences observed in the inferred allele relative to the IMGT allele; any other name that has previously been associated with this sequence, if previously identified; the genotype frequency of the inferred sequences within the donor's genotype and the allele percentage of the inferred allele based on all of the alleles within the donors genotype for that particular gene. **(B)** Alignment of each inferred sequence relative to the closest matching IMGT allele with the differences between the sequences highlighted in orange. Numbering of the alignments are according to IMGT numbering.

IARC, and the establishment of processes for the evaluation of inferred sequences provides an important new avenue for cataloging germline gene variation at the population level. Ultimately, this should provide insights into how germline gene diversity influences functional immunity (75, 76). Here, we describe the prerequisites, procedures and potential outcome of the IARC-based review and evaluation process, and as proof of principle, we report five novel alleles.

The establishment of the IARC review process should help the research community to chart germline IGHV gene variation across human ethnicities and patient groups. This is an achievable goal if studies increasingly infer the germline gene repertoires of each of their study subjects. Such personalized references databases will also improve AIRR-seq studies, through the improved germline gene annotation and confidence in identification of SHMs that will result (**Figure 1**).

The AIRR Community and the IMGT group have attempted to provide a robust roadmap and conceptual framework for germline gene inference, but the challenge will now be to encourage the incorporation of germline gene inference software into preprocessing and data analytical workflows. This has not yet been widely adopted by the community of researchers who generate and analyze AIRR-seq data. To facilitate this, IARC aims to create detailed step-by-step experimental and bioinformatics tutorials, and will document case studies showing the manifold advantages that lie in this approach. To minimize human intervention and subjectivity, we will also work to further automate the evaluation process of putative germline gene alleles, and to improve the data submission toolchains to INSDC repositories and to IMGT. Finally, in the future, we intend to partner with other researchers, to extend this initiative to the validation of other adaptive immune receptor gene loci. Other IG and TR genes in humans and species of medical importance may be an early focus, but in time we anticipate that the process of inference can be used to extend our knowledge of antigen receptor genes in all vertebrate species.

Putative novel alleles may now be submitted to the IARC-managed web portal for evaluation.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: http://ogrdb.airr-community.org/.

AUTHOR CONTRIBUTIONS

The authors are all members of the Germline Database Working Group of the AIRR Community of the Antibody Society. All authors contributed to the development of the policies and procedures described. MO and AC drafted the manuscript, and all authors contributed to the editing of the manuscript.

FUNDING

DB was supported in part by the Fondazione Umberto Veronesi. SC was supported in part by an NIAID grant (R01-AI097403). SK was supported in part by an NIH grant (R01-AI104739). FM and DR were supported in part by NIH grants (R01-GM113246, R01-AI120961, R01-AI138709). DR was also supported by grants from the NIH and University of Washington/Fred Hutchinson CFAR (U19-AI117891, P30-AI027757). MO and LT were supported in part by a grant from the Swedish Research Council (2016-01720). CW was supported in part by NIAID grants (R24-AI138963, R21-AI142590). GY was supported in part by a grant from the United States-Israel Binational Science Foundation (2017253). MvZ was supported in part by a grant from the NHMRC (GNT1117687).

REFERENCES

1. Nielsen SCA, Boyd SD. Human adaptive immune receptor repertoire analysis-Past, present, and future. *Immunol Rev*. (2018) 284:9–23. doi: 10.1111/imr.12667
2. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. (2014) 32:158–68. doi: 10.1038/nbt.2782
3. Wardemann H, Busse CE. Novel approaches to analyze immunoglobulin repertoires. *Trends Immunol*. (2017) 38:471–82. doi: 10.1016/j.it.2017.05.003
4. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front Immunol*. (2018) 9:224. doi: 10.3389/fimmu.2018.00224
5. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol*. (2010) 184:6986–92. doi: 10.4049/jimmunol.1000445
6. Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, et al. The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J Immunol*. (2012) 188:1333–40. doi: 10.4049/jimmunol.1102097
7. Kirik U, Greiff L, Levander F, Ohlin M. Parallel antibody germline gene and haplotype analyses support the validity of immunoglobulin germline gene inference and discovery. *Mol Immunol*. (2017) 87:12–22. doi: 10.1016/j.molimm.2017.03.012
8. Scheepers C, Shrestha RK, Lambson BE, Jackson KJ, Wright IA, Naicker D, et al. Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. *J Immunol*. (2015) 194:4371–8. doi: 10.4049/jimmunol.1500118
9. Matthyssens G, Rabbitts TH. Structure and multiplicity of genes for the human immunoglobulin heavy chain variable region. *Proc Natl Acad Sci USA*. (1980) 77:6561-5.
10. Lefranc M-P. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. *Front Immunol*. (2014) 5:22. doi: 10.3389/fimmu.2014.00022
11. Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, et al. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med*. (1998) 188:2151-62.
12. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, et al. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res*. (1997) 7:250-61.
13. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Asakawa S, et al. Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. *Eur J Immunol*. (2001) 31:1017–28. doi: 10.1002/1521-4141(200104)31:43.3.CO;2-V
14. Lefranc M-P, Lefranc G. *The Immunoglobulin FactsBook*. London, UK: Academic Press (2001), pp 1–458.
15. Lefranc M-P, Lefranc G. *The T Cell Receptor FactsBook*. London, UK: Academic Press (2001), pp 1–398.
16. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res*. (2015) 43:D413–22. doi: 10.1093/nar/gku1056
17. Retter I, Althaus HH, Munch R, Muller W. VBASE2, an integrative V gene database. *Nucleic Acids Res*. (2005) 33:D671-4. doi: 10.1093/nar/gki088

18. Jackson KJ, Gaeta B, Sewell W, Collins AM. Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire. *BMC Immunol.* (2004) 5:19. doi: 10.1186/1471-2172-5-19
19. Ohm-Laursen L, Nielsen M, Larsen SR, Barington T. No evidence for the use of DIR, D-D fusions, chromosome 15 open reading frames or VH replacement in the peripheral repertoire was found on application of an improved algorithm, JointML, to 6329 human immunoglobulin H rearrangements. *Immunology*. (2006) 119:265–77. doi: 10.1111/j.1365-2567.2006.02431.x
20. Wang Y, Jackson KJ, Sewell WA, Collins AM. Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. *Immunol Cell Biol.* (2008) 86:111–5. doi: 10.1038/sj.icb.7100144
21. Weinstein JA, Jiang N, White RA, III, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. (2009) 324: 807–10. doi: 10.1126/science.1170020
22. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci USA*. (2009) 106:20216–21. doi: 10.1073/pnas.0909775106
23. Corcoran MM, Phad GE, Vazquez Bernat N, Stahl-Hennig C, Sumida N, Persson MA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nat Commun.* (2016) 7:13642. doi: 10.1038/ncomms13642
24. Gadala-Maria D, Yaari G, Uduman M, Kleinstein SH. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc Natl Acad Sci USA*. (2015) 112:E862–70. doi: 10.1073/pnas.1417683112
25. Zhang W, Wang IM, Wang C, Lin L, Chai X, Wu J, et al. IMPre: An accurate and efficient software for prediction of T- and B-cell receptor germline genes and alleles from rearranged repertoire data. *Front Immunol.* (2016) 7:457. doi: 10.3389/fimmu.2016.00457
26. Ralph DK, Matsen IV FA. *Per Sample Immunoglobulin Germline Inference From B cell Receptor Deep Sequencing Data*. arXiv. Available online at: https://arxiv.org/abs/1711.05843 (Accessed August 15, 2018).
27. Wendel BS, He C, Crompton PD, Pierce SK, Jiang N. A streamlined approach to antibody novel germline allele prediction and validation. *Front Immunol.* (2017) 8:1072. doi: 10.3389/fimmu.2017.01072
28. Brochet X, Lefranc M-P, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucl Acids Res.* (2008) 36:W503–508. doi: 10.1093/nar/gkn316
29. Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, et al. Reproducibility and reuse of adaptive immune receptor repertoire data. *Front Immunol.* (2017) 8:1418. doi: 10.3389/fimmu.2017.01418
30. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. *Ann Rev Biochem*. (2007) 76:1–22. doi: 10.1146/annurev.biochem.76.061705.090740
31. Jacob J, Kelsoe G, Rajewsky K, Weiss U. Intraclonal generation of antibody mutants in germinal centres. *Nature*. (1991) 354:389–92. doi: 10.1038/354389a0
32. Zheng NY, Wilson K, Jared M, Wilson PC. Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *J Exp Med*. (2005) 201:1467–78. doi: 10.1084/jem.20042483
33. Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. (2010) 329:856–61. doi: 10.1126/science.1187659
34. Wang Y, Jackson KJ, Chen Z, Gaeta BA, Siba PM, Pomat W, et al. IgE sequences in individuals living in an area of endemic parasitism show little mutational evidence of antigen selection. *Scand J Immunol.* (2011) 73:496–504. doi: 10.1111/j.1365-3083.2011.02525.x
35. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, et al. 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*. (2013) 92:530–46. doi: 10.1016/j.ajhg.2013.03.004
36. Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* (2015) 7:121. doi: 10.1186/s13073-015-0243-2
37. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trend Immunol.* (2015) 36:738–49. doi: 10.1016/j.it.2015.09.006
38. Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A, et al. Synthetic standards combined with error and bias correction improve the accuracy and quantitative resolution of antibody repertoire sequencing in human naive and memory B cells. *Front Immunol.* (2018) 9:1401. doi: 10.3389/fimmu.2018.01401
39. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nature Commun.* (2013) 4:2680. doi: 10.1038/ncomms3680
40. Khan TA, Friedensohn S, Gorter de Vries AR, Straszewski J, Ruscheweyh HJ, Reddy ST. Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv*. (2016) 2:e1501371. doi: 10.1126/sciadv.1501371
41. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, et al. Towards error-free profiling of immune repertoires. *Nat Methods*. (2014) 11:653–5. doi: 10.1038/nmeth.2960
42. McInerney P, Adams P, Hadi MZ. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol Biol Int*. (2014) 2014:287430. doi: 10.1155/2014/287430
43. Bolotin DA, Mamedov IZ, Britanova OV, Zvyagin IV, Shagin D, Ustyugova SV, et al. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol*. (2012) 42:3073–83. doi: 10.1002/eji.201242517
44. Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*. (2016) 17:125. doi: 10.1186/s12859-016-0976-y
45. Kirik U, Greiff L, Levander F, Ohlin M. Data on haplotype-supported immunoglobulin germline gene inference. *Data Brief*. (2017) 13: 620–40. doi: 10.1016/j.dib.2017.06.031
46. Vollmers C, Sit RV, Weinstein JA, Dekker CL, Quake SR. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc Natl Acad Sci USA*. (2013) 110:13463–8. doi: 10.1073/pnas.1312146110
47. Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. *Nucleic Acids Res*. (1990) 18:1687–91.
48. Judo MS, Wedel AB, Wilson C. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res*. (1998) 26:1819–25.
49. Zylstra P, Rothenfluh HS, Weiller GF, Blanden RV, Steele EJ. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunol Cell Biol*. (1998) 76:395–405. doi: 10.1046/j.1440-1711.1998.00772.x
50. Kalle E, Kubista M, Rensing C. Multi-template polymerase chain reaction. *Biomol Det Quant*. (2014) 2:11–29. doi: 10.1016/j.bdq.2014.11.002
51. Gupta NT, vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein, SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*. (2015) 31:3356–8. doi: 10.1093/bioinformatics/btv359
52. Gidoni M, Snir O, Peres A, Polak P, Lindeman I, Clouser C, et al. Mosaic deletion patterns of the human antibody heavy chain gene locus. *Nat Commun*. (2019) 10:628. doi: 10.1038/s41467-019-08489-3
53. Elhanati Y, Sethna Z, Marcou Q, Callan CG, Jr., Mora T, Walczak AM. Inferring processes underlying B-cell repertoire diversity. *Philos Trans R Soc Lond B Biol Sci*. (2015) 370:1676. doi: 10.1098/rstb.2014.0243
54. Ralph DK, Matsen IV FA. Consistency of VDJ rearrangement and substitution parameters enables accurate B cell receptor sequence annotation. *PLoS Comput Biol*. (2016) 12:e1004409. doi: 10.1371/journal.pcbi.1004409
55. Thornqvist L, Ohlin M. The functional 3′-end of immunoglobulin heavy chain variable (IGHV) genes. *Mol Immunol*. (2018) 96:61–8. doi: 10.1016/j.molimm.2018.02.013
56. Thornqvist L, Ohlin M. Data on the nucleotide composition of the first codons encoding the complementary determining region 3 (CDR3) in immunoglobulin heavy chains. *Data Brief*. (2018) 19:337–52. doi: 10.1016/j.dib.2018.04.125
57. Kleinstein SH, Louzoun Y, Shlomchik MJ. Estimating hypermutation rates from clonal tree data. *J Immunol*. (2003) 171:4639–49. doi: 10.4049/jimmunol.171.9.4639

58. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci USA*. (1984) 81:3180–4.
59. Chang B, Casali P. The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. *Immunol Today*. (1994) 15:367–73. doi: 10.1016/0167-5699(94)90175-9
60. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol*. (2004) 172:3382–4. doi: 10.4049/jimmunol.172.6.3382
61. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat Immunol*. (2001) 2:530–6. doi: 10.1038/88732
62. Kirik U, Persson H, Levander F, Greiff L, Ohlin M. Antibody heavy chain variable domains of different germline gene origins diversify through different paths. *Front Immunol*. (2017) 8:1433. doi: 10.3389/fimmu.2017.01433
63. Sheng Z, Schramm CA, Kong R, Program NCS, Mullikin JC, Mascola JR, et al. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front Immunol*. (2017) 8:537. doi: 10.3389/fimmu.2017.00537
64. Schramm CA, Douek DC. Beyond hot spots: Biases in antibody somatic hypermutation and implications for vaccine design. *Front Immunol*. (2018) 9:1876. doi: 10.3389/fimmu.2018.01876
65. Luo S, Yu JA, Song YS. Estimating copy number and allelic variation at the immunoglobulin heavy chain locus using short reads. *PLoS Comput Biol*. (2016) 12:e1005117. doi: 10.1371/journal.pcbi.1005117
66. Parks T, Mirabel MM, Kado J, Auckland K, Nowak J, Rautanen A, et al. Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. *Nat Commun*. (2017) 8:14946. doi: 10.1038/ncomms14946
67. Feeney AJ, Atkinson MJ, Cowan MJ, Escuro G, Lugo G. A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to haemophilus influenzae type b disease. *J Clin Invest*. (1996) 97:2277–82. doi: 10.1172/JCI118669
68. Kidd MJ, Jackson KJ, Boyd SD, Collins AM. DJ pairing during VDJ recombination shows positional biases that vary among individuals with differing IGHD locus immunogenotypes. *J Immunol*. (2016) 196:1158–64. doi: 10.4049/jimmunol.1501401
69. Rubelt F, Busse CE, Bukhari SAC, Burckert JP, Mariotti-Ferrandiz E, Cowell LG, et al. Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. *Nat Immunol*. (2017) 18:1274–8. doi: 10.1038/ni.3873
70. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*. (2018) 15:201–6. doi: 10.1038/nmeth.4577
71. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. (2003) 17:2257–317. doi: 10.1038/sj.leu.2403202
72. Lefranc MP. From IMGT-ontology classification axiom to IMGT standardized gene and allele nomenclature: for immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harbor Protocols*. (2011) 2011:627–32. doi: 10.1101/pdb.ip84
73. Vergani S, Korsunsky I, Mazzarello AN, Ferrer G, Chiorazzi N, Bagnara D. Novel method for high-throughput full-length IGHV-D-J sequencing of the immune repertoire from bulk B-cells with single-cell resolution. *Front Immunol*. (2017) 8:1157. doi: 10.3389/fimmu.2017.01157
74. Thornqvist L, Ohlin M. Critical steps for computational inference of the 3′-end of novel alleles of immunoglobulin heavy chain variable genes - illustrated by an allele of IGHV3-7. *Mol Immunol*. (2018) 103:1–6. doi: 10.1016/j.molimm.2018.08.018
75. Watson CT, Glanville J, Marasco WA. The Individual and Population Genetics of Antibody Immunity. *Trend Immunol*. (2017) 38:459–70. doi: 10.1016/j.it.2017.04.003
76. Avnir Y, Watson CT, Glanville J, Peterson EC, Tallarico AS, Bennett AS, et al. IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci Rep*. (2016) 6:20842. doi: 10.1038/srep20842

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

OPINION
published: 03 May 2019
doi: 10.3389/fimmu.2019.00962

Macrophages Do Not Express the Phagocytic Receptor BAI1/*ADGRB1*

Cheng-Chih Hsiao[1], *Marlijn van der Poel*[2], *Tjakko J. van Ham*[3] *and Jörg Hamann*[1,2]*

[1] *Department of Experimental Immunology, Amsterdam Infection and Immunity Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands,* [2] *Department of Neuroimmunology, Netherlands Institute for Neuroscience, Amsterdam, Netherlands,* [3] *Department of Clinical Genetics, Erasmus University Medical Center Rotterdam, Rotterdam, Netherlands*

Keywords: adhesion GPCRs, brain-specific angiogenesis inhibitors, macrophages, microglia, monocytes, phagocytic receptors

OPEN ACCESS

Edited by:
Loems Ziegler-Heitbrock, Independent Researcher, Munich, Germany

Reviewed by:
Jean-François Côté, Institute of Clinical Research De Montreal (IRCM), Canada

***Correspondence:**
Jörg Hamann
j.hamann@amc.uva.nl

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *05 February 2019*
Accepted: *15 April 2019*
Published: *03 May 2019*

Citation:
Hsiao C-C, van der Poel M, van Ham TJ and Hamann J (2019) Macrophages Do Not Express the Phagocytic Receptor BAI1/ADGRB1. Front. Immunol. 10:962. doi: 10.3389/fimmu.2019.00962

The highly organized life of metazoa requires the ability to remove cells that lose their function during embryonic and postnatal development or as part of routine tissue homeostasis (1, 2). Normally, these cells undergo programmed, apoptotic cell death, followed by their recognition, engulfment, and, finally, elimination through adjacent tissue cells and/or professional phagocytes. As preeminent phagocytic cells, resident macrophages and circulating monocytes are equipped with an arsenal of receptors that recognize the "eat-me" signals exposed by apoptotic corpses. These phagocytic receptors comprise scavenger receptors, immunoglobulin-containing proteins, and tyrosine kinases (1).

In a Nature paper in 2007, Park et al. described brain-specific angiogenesis inhibitor 1 (BAI1/*ADGRB1*) as a novel phagocytic receptor on macrophages (3). BAI1 is a member of the adhesion family of G protein-coupled receptors (GPCRs), which in humans comprises 33 non-canonical seven-span transmembrane receptors (4). Adhesion GPCRs possess large N-termini with various protein folds, equipped for (matri)cellular interactions, and a GPCR autoproteolysis-inducing (GAIN) domain that connects the extracellular part of the receptor to the seven-transmembrane region. A juxtamembranous GPCR-proteolysis site (GPS) within the GAIN domain facilitates autocatalytic cleavage of the majority of adhesion GPCRs into two fragments, which remain attached at the cell surface (5). Adhesion GPCRs are found in almost every cell type and adjust modalities in many organ systems. Based on their expression and function, adhesion GPCRs of subfamily E (EMR1/*ADGRE1*, EMR2/*ADGRE2*, EMR3/*ADGRE3*, EMR4/*ADGRE4*, and CD97/*ADGRE5*) and subfamily G (GPR56/*ADGRG1*, GPR97/*ADRGRG3*, and GPR114/*ADGRG5*) have been linked to the immune system (6, 7). BAI1 belongs to the subfamily B and is abundantly expressed in the brain, where it inhibits angiogenesis and, as recently reported, supports neurogenesis and synaptogenesis (8). The work by Park et al. and others established an additional function of BAI1 in apoptotic cell engulfment by macrophages and their brain equivalent, microglia (3, 9, 10). Through its N-terminal thrombospondin repeats, BAI1 binds phosphatidylserine, resulting in recruitment of ELMO1 and Dock180 to the C-terminus of the receptor, which function as guanine-exchange factors for Rac1 and thereby promote engulfment of apoptotic cells. Moreover, expression of BAI1 in primary human monocytes/macrophages and the mouse macrophage cell lines J774 and RAW264.7 was reported (3).

Ingestion of microbes, such as bacteria and fungi, is another phagocytic process executed by macrophages. A subsequent paper in 2011 described the ability of BAI1 to bind and engulf Gram-negative bacteria (11). Interaction of the thrombospondin repeats with bacterial membrane lipopolysaccharide triggered *Salmonella* engulfment via ELMO1/Dock180, similar to the uptake of

apoptotic cells. Subsequently, it has been reported that BAI1 mediates macrophage reactive oxygen species production and microbicidal activity through activation of the Rho family guanosine triphosphatase Rac1 (12). These observations further established BAI1 as a phagocytic receptor of macrophages.

Transcriptome (and proteome) analyses of purified cell populations and, more recently, even single cells is greatly deepening our knowledge about the spatial organization of gene expression. We noticed that omics studies directed at leukocytes consistently detect expression of subfamily E and G adhesion GPCRs, but fail to identify subfamily B receptors, including BAI1 (4, 6, 7). To clarify this discrepancy, we analyzed microarray, CAGE (cap analysis gene expression) and RNA sequencing, and protein mass spectrometry data of primary monocytes, monocytes maturated *in vitro* under stimulating conditions, macrophage cell lines, as well as bone marrow-derived and primary tissue-derived macrophages. We included all types of monocytes/macrophages, in which *Adgrb1*/*ADGRB1* expression has been reported, with the exception of gastric phagocytes (**Table 1**). Among other data sets, we evaluated adhesion GPCR transcriptomes (20) and proteomes (23) of classical, intermediate, and non-classical monocytes (**Figures 1A,B**). Moreover, we examined 299 transcriptomes of monocytes activated with 28 different stimuli, including pattern recognition receptor ligands, cytokines, and metabolic cues (19) (**Figure 1C**). In none of these and numerous other data sets (**Table 1**), we obtained evidence that monocytes or monocyte-derived macrophages express *Adgrb1*/*ADGRB1*, while known gene expression patterns of subfamily E adhesion GPCRs were fully confirmed (6, 7).

Knowledge of genome-wide gene expression in tissue-resident macrophages, so far, is mainly based on studies in mice. In transcriptomes of seven types of macrophages, *Adgrb1* was not detected (25) (**Figure 1D**). These transcriptomes also included microglia, for which a distinct role for BAI1 in the engulfment of neurons has been described in zebrafish (10). Zebrafish express homologs of most adhesion GPCRs, including BAI1 (34). Yet, by RNA sequencing highly pure microglia from zebrafish, we failed to detect significant levels of *Adgrb1* expression (27) (**Figure 1D**). Similarly, microglia from mouse and human express *Adgrg1*/*ADGRG1*, but not *Adgrb1*/*ADGRB1* (24, 28–32) (**Figures 1D,E**).

We also asked whether unusual mRNA properties, e.g., short poly(A) tails, could have hampered the detection of *Adgrb1*/*ADGRB1* transcripts. To exclude this possibility, we included in our comparison RNA sequencing data obtained by reduction of ubiquitously expressed ribosomal (r)RNAs in combination with random primer amplification (13, 14). Moreover, we were able to directly compare sequencing of human microglia RNAs obtained by poly(A) selection and rRNA depletion plus random primer amplification [(32) and Mizee et al., manuscript in preparation], but failed to detect *ADGRB1* transcripts with both methods (data not shown). Furthermore, *Adgrb1*/*ADGRB1* transcripts are found in mouse and human brain lysate (**Figure 1F**) as well as in mouse neurons, oligodendrocyte progenitors, and astrocytes (28), confirming their detectability.

Our data do not challenge the role of BAI1 as a phagocytic receptor. This biological activity is based on the binding capacity of the N-terminal thrombospondin repeats for "eat-me" signals on apoptotic cells and on the ability of the C-terminal tail to facilitate cytoskeletal rearrangements, and has been proven extensively (3, 11). We question, however, that BAI1 is part of the phagocytic machinery of macrophages. The link with macrophages has been established in primary cells and cell lines overexpressing BAI1 *in vitro*. More recently, Lee at al. investigated the role of BAI1 in the dextran sodium sulfate-induced model of colitis *in vivo*. *Adgrb1*-deficient mice had more pronounced colitis and lower survival, with many uncleared apoptotic cells and inflammatory cytokines within the colonic epithelium. Notably, transgenic overexpression of *Adgrb1* in epithelial, but not in myeloid cells, attenuated colitis severity (35), suggesting that BAI1 mediates clearance of apoptotic corpses within the colonic epithelium. Intestinal epithelial cells may not be the only non-professional phagocytes that engage BAI1. In astrocytes engulfing apoptotic targets, BAI1 showed accumulation within the phagocytic cup (26). Moreover, BAI1 and BAI3 have been described to promote myoblast fusion, a process possibly induced by dying myoblasts (36, 37).

In summary, monocytes and macrophages, including microglia, express the adhesion GPCRs EMR1, EMR2, EMR3, CD97, and GPR56 with different species and cell type specificity. BAI1, an adhesion GPCR with diverse and intriguing functions in angiogenesis, neural development, and apoptotic/microbial engulfment, is hardly expressed by

TABLE 1 | Studies reporting and studies failing to find expression of *Adgrb1/ADGRB1* (BAI1) in monocytes/macrophages.

Cell type	Reporting expression	Failing to find expression
Mouse monocyte/ macrophage cell lines J774A.1 and RAW264.7	RT-PCR, IB (3)	RNAseq (13–15)
Human monocyte/ macrophage cell line THP-1	RT-PCR, IB (9)	RNAseq (16, 17) (http://www.proteinatlas.org)
Monocytes and monocyte-derived macrophages	Microarray (18), IB (9)	Microarray (19), CAGEseq (20), RNAseq (17, 21, 22), MS (23)
Bone marrow-derived macrophages	RT-PCR (11)	RNAseq (14, 24)
Tissue-derived macrophages	RT-PCR, IB (9)	RNAseq (24, 25) (https://www.immgen.org/)
Microglia	IHC (26), ISH (10)	RNAseq (24, 25, 27–32)

CAGEseq, CAGE sequencing; IB, immunoblot; IHC, immunohistochemistry; ISH, in situ hybridization; MS, mass spectrometry; RNAseq, RNA sequencing; RT-PCR, reverse transcriptase-polymerase chain reaction.

Abbreviations: BAI, brain-specific angiogenesis inhibitor; GAIN, GPCR autoproteolysis-inducing; GPCR, G protein-coupled receptor; GPS, GPCR-proteolysis site.

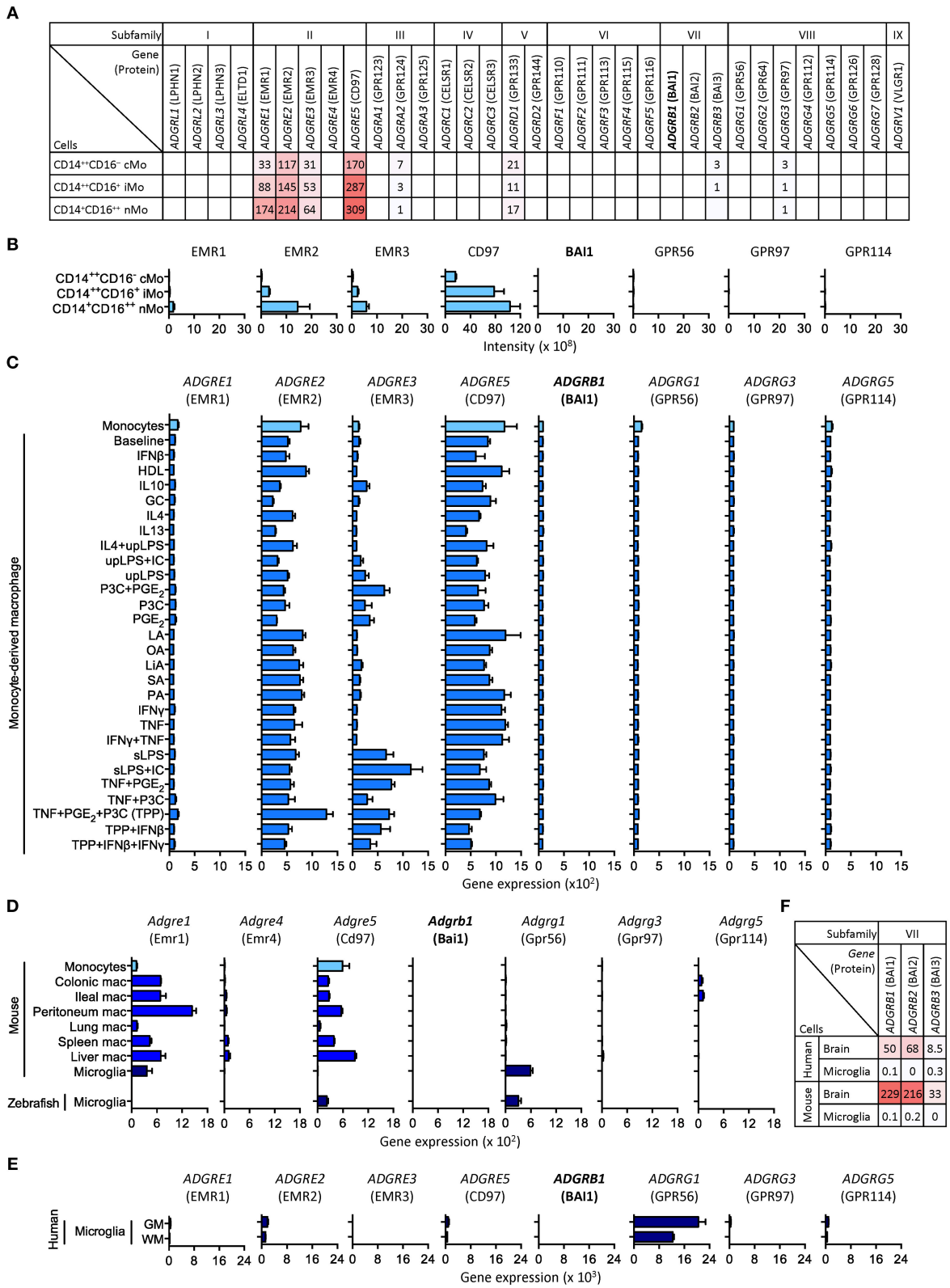

FIGURE 1 | Selected expression profiles of adhesion GPCRs in monocytes, monocyte-derived macrophages, and microglia. **(A)** CAGE sequencing of circulating human monocytes (20). **(B)** Protein mass spectrometry of circulating human monocytes (23). **(C)** Microarray of human monocytes activated with 28 different stimuli

(Continued)

FIGURE 1 | (19). **(D)** RNA sequencing of resident mouse macrophages as well as mouse and zebrafish microglia (25, 27). **(E)** RNA sequencing of resident human grey and white matter (GM and WM) microglia (32). **(F)** RNA sequencing of mouse and human brain lysates and microglia (30). Note the consistent lack of BAI1 (*Adgrb1/ADGRB1*) expression in all data sets. Expression of EMR1 to EMR4 (*Adgre1/ADGRE1* to *Adgre4/ADGRE4*) in human and mouse reflect their evolutionary diversification: (i) in contrast to its mouse homolog, F4/80, human EMR1 is weekly expressed by monocytes and macrophages; (ii) mice lack the genes encoding EMR2 and EMR3; (iii) the gene encoding EMR4 has become inactivated in human (33).

professional phagocytes, and we suggest to reassess the link between BAI1 and macrophage biology.

AUTHOR CONTRIBUTIONS

C-CH, MvdP, TvH, and JH generated and analyzed data. C-CH and JH wrote the paper.

ACKNOWLEDGMENTS

We thank Mark Mizee for sharing unpublished data and Tobias Langenhan for helpful comments. The study was supported by grants from the Thyssen Foundation (2015-00387), the MS Research Foundation (MS13-830), and the German Research Foundation (FOR 2149).

REFERENCES

1. Arandjelovic S, Ravichandran KS. Phagocytosis of apoptotic cells in homeostasis. *Nat Immunol*. (2015) 16:907–17. doi: 10.1038/ni.3253
2. Henson PM. Cell removal: efferocytosis. *Annu Rev Cell Dev Biol*. (2017) 33:111315–25315. doi: 10.1146/annurev-cellbio-111315-125315
3. Park D, Tosello-Trampont AC, Elliott MR, Lu M, Haney LB, Ma Z, et al. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature*. (2007) 450:430–4. doi: 10.1038/nature06329
4. Hamann J, Aust G, Araç D, Engel FB, Formstone C, Fredriksson R, et al. International union of basic and clinical pharmacology. XCIV. Adhesion G protein-coupled receptors. *Pharmacol Rev*. (2015) 67:338–67. doi: 10.1124/pr.114.009647
5. Prömel S, Langenhan T, Araç D. Matching structure with function: the GAIN domain of adhesion-GPCR and PKD1-like proteins. *Trends Pharmacol Sci*. (2013) 34:470–8. doi: 10.1016/j.tips.2013.06.002
6. Hamann J, Hsiao CC, Lee CS, Ravichandran KS, Lin HH. Adhesion GPCRs as modulators of immune cell function. *Handb Exp Pharmacol*. (2016) 234:329–50. doi: 10.1007/978-3-319-41523-9_15
7. Lin H-H, Hsiao C-C, Pabst C, Hébert J, Schöneberg T, Hamann J. Adhesion GPCRs in regulating immune responses and inflammation. *Adv Immunol*. (2017) 136:163–201. doi: 10.1016/bs.ai.2017.05.005
8. Stephenson JR, Purcell RH, Hall RA. The BAI subfamily of adhesion GPCRs: synaptic regulation and beyond. *Trends Pharmacol Sci*. (2014) 35:208–15. doi: 10.1016/j.tips.2014.02.002
9. Das S, Sarkar A, Ryan KA, Fox S, Berger AH, Juncadella IJ, et al. Brain angiogenesis inhibitor 1 is expressed by gastric phagocytes during infection with Helicobacter pylori and mediates the recognition and engulfment of human apoptotic gastric epithelial cells. *FASEB J*. (2014) 28:2214–24. doi: 10.1096/fj.13-243238
10. Mazaheri F, Breus O, Durdu S, Haas P, Wittbrodt J, Gilmour D, et al. Distinct roles for BAI1 and TIM-4 in the engulfment of dying neurons by microglia. *Nat Commun*. (2014) 5:4046. doi: 10.1038/ncomms5046
11. Das S, Owen KA, Ly KT, Park D, Black SG, Wilson JM, et al. Brain angiogenesis inhibitor 1 (BAI1) is a pattern recognition receptor that mediates macrophage binding and engulfment of Gram-negative bacteria. *Proc Natl Acad Sci USA*. (2011) 108:2136–41. doi: 10.1073/pnas.1014775108
12. Billings EA, Lee CS, Owen KA, D'Souza RS, Ravichandran KS, Casanova JE. The adhesion GPCR BAI1 mediates macrophage ROS production and microbicidal activity against Gram-negative bacteria. *Sci Signal*. (2016) 9:ra14. doi: 10.1126/scisignal.aac6250
13. Schott J, Reitter S, Philipp J, Haneke K, Schäfer H, Stoecklin G. Translational regulation of specific mRNAs controls feedback inhibition and survival during macrophage activation. *PLoS Genet*. (2014) 10:e1004368. doi: 10.1371/journal.pgen.1004368
14. Das A, Yang C-S, Arifuzzaman S, Kim S, Kim SY, Jung KH, et al. High-resolution mapping and dynamics of the transcriptome, transcription factors, and transcription co-factor networks in classically and alternatively activated macrophages. *Front Immunol*. (2018) 9:22. doi: 10.3389/fimmu.2018.00022
15. Andreu N, Phelan J, de Sessions PF, Cliff JM, Clark TG, Hibberd ML. Primary macrophages and J774 cells respond differently to infection with Mycobacterium tuberculosis. *Sci Rep*. (2017) 7:42225. doi: 10.1038/srep42225
16. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, et al. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Mol Cell*. (2017) 67:1037–48.e6. doi: 10.1016/j.molcel.2017.08.006
17. Liu H, Lorenzini PA, Zhang F, Xu S, Wong MSM, Zheng J, et al. Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator. *Nucleic Acids Res*. (2018) 46:6069–86. doi: 10.1093/nar/gky401
18. Cho HJ, Shashkin P, Gleissner CA, Dunson D, Jain N, Lee JK, et al. Induction of dendritic cell-like phenotype in macrophages during foam cell formation. *Physiol Genomics*. (2007) 29:149–60. doi: 10.1152/physiolgenomics.00051.2006
19. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. (2014) 40:274–88. doi: 10.1016/j.immuni.2014.01.006
20. Schmidl C, Renner K, Peter K, Eder R, Lassmann T, Balwierz PJ, et al. Transcription and enhancer profiling in human monocyte subsets. *Blood*. (2014) 123:e90–9. doi: 10.1182/blood-2013-02-484188
21. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell*. (2018) 175:1701–15.e16. doi: 10.1016/j.cell.2018.10.022
22. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carré C, et al. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell Rep*. (2019) 26:1627–40.e7. doi: 10.1016/j.celrep.2019.01.041
23. Rieckmann JC, Geiger R, Hornburg D, Wolf T, Kveler K, Jarrossay D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol*. (2017) 18:583–93. doi: 10.1038/ni.3693
24. Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*. (2014) 159:1327–40. doi: 10.1016/j.cell.2014.11.023
25. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell*. (2014) 159:1312–26. doi: 10.1016/j.cell.2014.11.018
26. Sokolowski JD, Nobles SL, Heffron DS, Park D, Ravichandran KS, Mandell JW. Brain-specific angiogenesis inhibitor-1 expression in astrocytes and neurons: implications for its dual function as an apoptotic engulfment receptor. *Brain Behav Immun*. (2011) 25:915–21. doi: 10.1016/j.bbi.2010.09.021

27. Oosterhof N, Holtman IR, Kuil LE, van der Linde HC, Boddeke EWGM, Eggen BJL, et al. Identification of a conserved and acute neurodegeneration-specific microglial transcriptome in the zebrafish. *Glia.* (2017) 65:138–49. doi: 10.1002/glia.23083
28. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci.* (2014) 34:11929–47. doi: 10.1523/JNEUROSCI.1860-14.2014
29. Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell.* (2017) 169:1276–90.e17. doi: 10.1016/j.cell.2017.05.018
30. Gosselin D, Skola D, Coufal NG, Holtman IR, Schlachetzki JCM, Sajti E, et al. An environment-dependent transcriptional network specifies human microglia identity. *Science.* (2017) 356:eaal3222. doi: 10.1126/science.aal3222
31. Galatro TF, Holtman IR, Lerario AM, Vainchtein ID, Brouwer N, Sola PR, et al. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat Neurosci.* (2017) 20:1162–71. doi: 10.1038/nn.4597
32. van der Poel M, Ulas T, Mizee MR, Hsiao C-C, Miedema SSM, Schuurman KG, et al. Transcriptional profiling of human microglia reveals grey-white matter heterogeneity and multiple sclerosis-associated changes. *Nat Commun.* (2019) 10:1139. doi: 10.1038/s41467-019-08976-7
33. Gordon S, Hamann J, Lin HH, Stacey M. F4/80 and the related adhesion-GPCRs. *Eur J Immunol.* (2011) 41:2472–6. doi: 10.1002/eji.201141715
34. Harty BL, Krishnan A, Sanchez NE, Schiöth HB, Monk KR. Defining the gene repertoire and spatiotemporal expression profiles of adhesion G protein-coupled receptors in zebrafish. *BMC Genomics.* (2015) 16:62. doi: 10.1186/s12864-015-1296-8
35. Lee CS, Penberthy KK, Wheeler KM, Juncadella IJ, Vandenabeele P, Lysiak JJ, et al. Boosting Apoptotic cell clearance by colonic epithelial cells attenuates inflammation *in vivo*. *Immunity.* (2016) 44:807–20. doi: 10.1016/j.immuni.2016.02.005
36. Hochreiter-Hufford AE, Lee CS, Kinchen JM, Sokolowski JD, Arandjelovic S, Call JA, et al. Phosphatidylserine receptor BAI1 and apoptotic cells as new promoters of myoblast fusion. *Nature.* (2013) 497:263–7. doi: 10.1038/nature12135
37. Hamoud N, Tran V, Croteau L-P, Kania A, Côté J-F. G-protein coupled receptor BAI3 promotes myoblast fusion in vertebrates. *Proc Natl Acad Sci USA.* (2014) 111:3745–50. doi: 10.1073/pnas.1313886111

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

PERSPECTIVE
published: 07 June 2019
doi: 10.3389/fimmu.2019.01308

Complement Nomenclature—Deconvoluted

Suzanne S. Bohlson[1], *Peter Garred*[2], *Claudia Kemper*[3] *and Andrea J. Tenner*[4]*

[1] Department of Microbiology and Immunology, Des Moines University, Des Moines, IA, United States, [2] Laboratory of Molecular Medicine, Department of Clinical Immunology, Rigshospitalet University Hospital of Copenhagen, Copenhagen, Denmark, [3] Laboratory of Molecular Immunology and the Immunology Center, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Bethesda, MD, United States, [4] Department of Molecular Biology and Biochemistry, Department of Neurobiology and Behavior, Department of Pathology and Laboratory Medicine, University of California, Irvine, Irvine, CA, United States

In 2014, specific recommendations for complement nomenclature were presented by the complement field. There remained some unresolved designations and new areas of ambiguity, and here we propose solutions to resolve these remaining issues. To enable rapid understanding of the intricate complement system and facilitate therapeutic development and application, a uniform nomenclature for cleavage fragments, pattern recognition molecules (PRMs) and enzymes of the lectin pathway and regulatory proteins of the complement system are proposed, and a standardization of language to designate different activation states of complement components is recommended.

Keywords: complement, nomenclature, C1, C1q, C2, lectin pathway, collectins, clusterin

OPEN ACCESS

Edited by:
Pablo Engel,
University of Barcelona, Spain

Reviewed by:
Anna Erdei,
Eötvös Loránd University, Hungary
Seppo Meri,
University of Helsinki, Finland

***Correspondence:**
Andrea J. Tenner
atenner@uci.edu

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *01 April 2019*
Accepted: *22 May 2019*
Published: *07 June 2019*

Citation:
Bohlson SS, Garred P, Kemper C and Tenner AJ (2019) Complement Nomenclature—Deconvoluted. Front. Immunol. 10:1308. doi: 10.3389/fimmu.2019.01308

INTRODUCTION

The complement system is composed of more than 50 different molecules and cleavage products including but not limited to pattern recognition molecules (PRMs), proenzymes, proteases, anaphylatoxins, opsonins, receptors, regulators, and multi-molecular complexes that are critical to host defense and maintenance of normal tissue homeostasis (1). While traditional functions of the complement system in host defense and clearance of cellular debris have long been appreciated, continued advancement in the field has revealed additional roles for complement from embryogenesis to aging, in both healthy and disease states (**Figure 1A**) (2–4). In addition, successful development of the anti-human C5 monoclonal antibody, eculizumab, for treatment in paroxysmal nocturnal hemoglobinuria, atypical hemolytic uremic syndrome and refractory myasthenia gravis has renewed interest in clinical applications for complement within the medical field for treatment as well as diagnosis. Consequently, researchers in multiple fields even beyond immunology are investigating various components and pathways of the complement system.

To facilitate advancement and communication in both basic research and clinical application in the field, it is important to standardize complement nomenclature. Following a joint effort of the International Complement Society (ICS) and the European Complement Network (ECN) to simplify and clarify complement nomenclature, a list of recommended names for complement pathways, proteins, protein complexes, and receptors was established in 2014 (5), and the recommendations from that effort are shown in **Table 1**. This update to complement nomenclature was the first since 1981. While comprehensive, consensus was not reached on several issues. Here we propose recommendations and updates for nomenclature regarding four of these unresolved issues: (1) the cleavage products of C2, (2) C1 complexes- activated molecules, native molecules, or proenzymes, (3) lectin pathway recognition proteins and enzymes and (4) Clusterin.

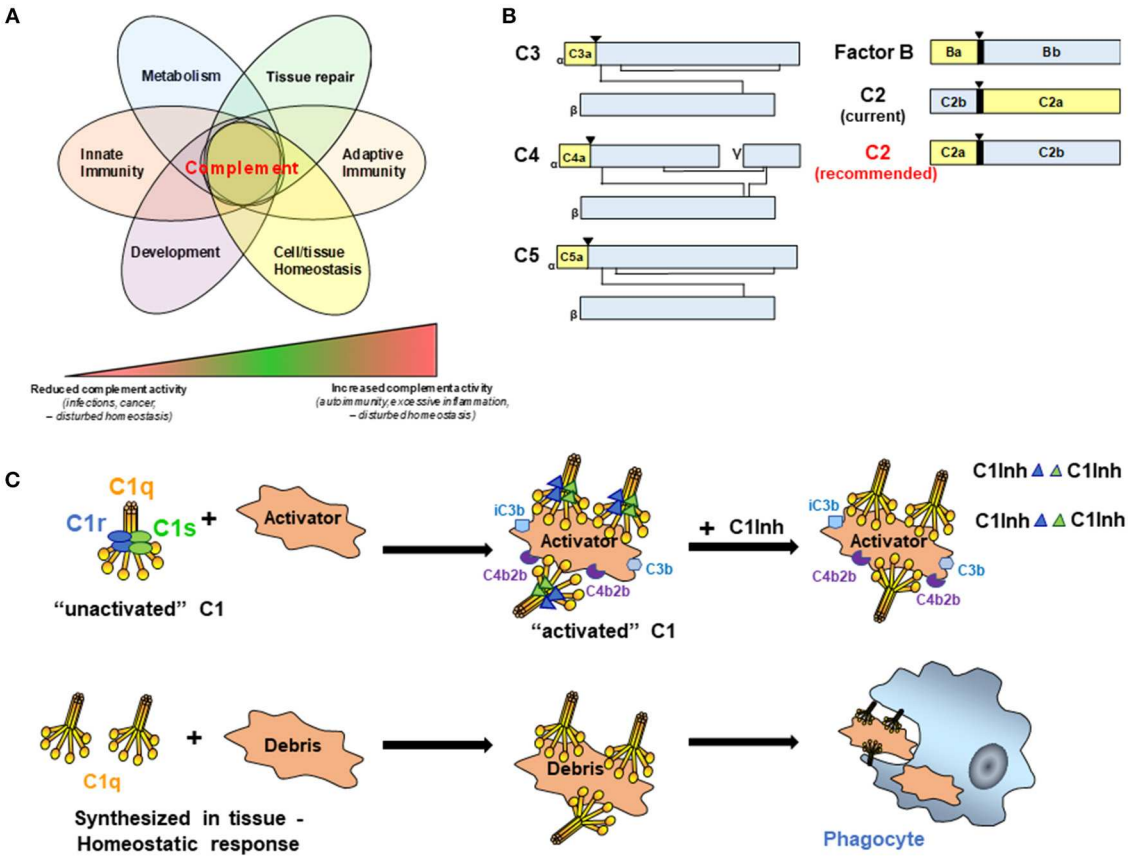

FIGURE 1 | (A) Complement activities control key cellular processes and contribute to a broad range of disease states. It is now broadly acknowledged that complement functions well beyond mere protection against invading pathogens but participates actively in the control of key physiological processes. Therefore, complement is key to normal cell and tissue homeostasis and aberrant complement activation (either too little or too much) can hence cause or contribute to a broad range of disease settings including recurrent infections and cancer (too little) or auto-immunity and fibrosis (too much). **(B)** Schematic representation of C3 family members, FB, and C2. While convention holds that smaller fragments (yellow for C3, C4, C5, and FB) retain earlier letters than larger fragments (Blue for C3, C4, C5, and FB), C2 breaks with convention (C2 current). To facilitate communication in the field and better standardize complement nomenclature, we recommend adopting standard convention for C2 (C2 recommended). (▼) Indicates cleavage site for liberation of smaller fragment. **(C)** C1 is a complex macromolecular structure consisting of C1q, C1r, and C1s. C1 circulates in blood as an "unactivated" complex of the recognition protein, C1q (yellow), and two molecules each of the proenzymes C1r and Cls (blue and green ellipses). Conformational changes induced by binding to an activator result in activated C1 due to the conversion of C1r and C1s to active serine proteases (blue and green triangles). C1s proceeds to cleave C4 and C2 which results ultimately in the formation of the classical pathway C3 convertase, C4b2b. Generation of opsonic C3b and iC3b, and subsequently C5a and C5b-9 (not shown) mediate the possible complement effector functions that follow. Four C1-INH molecules per C1 (only representative complexes are shown) are required to inactivate the serine proteases, C1r and C1s, and results in their dissociation from C1q, thereby regulating the amount of C4b2b generated. In the lower row, C1q can be synthesized in tissue to "silently" eliminate apoptotic cells and cellular debris.

THE CLEAVAGE PRODUCTS OF C2

In general, complement cleavage fragments are designated with letters according to their relative size with "a" fragments smaller than "b" fragments. **Figure 1B** is a schematic depicting members of the C3 family, as well as Factor B (FB) of the alternative pathway, for which the nomenclature follows this convention. In all of these cases (C3, C4, C5, FB), the larger fragment remains cell associated, and the smaller fragment diffuses from the original site of cleavage. Moreover, C3 family members (C3, C4, and C5) share similar structures, and the smaller cleavage products (C3a, C4a, and C5a) of these molecules all engage receptors on cells to trigger signaling pathways and activation processes. Convention is followed for FB, where the smaller, diffusible fragment, Ba, is liberated from the larger fragment, Bb, which remains associated with activator bound C3b. The serine protease domain of FB is within the Bb fragment.

Current nomenclature in popular use for C2 breaks convention in that the smaller fragment is often referred to as "C2b" and the larger fragment "C2a" (**Figure 1B**, current nomenclature). Originally, and prior to detailed knowledge of the activation mechanisms involved, the activated C2 molecule was designated as C'2a (6) and refers to the generated ability of C2 to enable/activate the cascade to continue through C3 and ultimately generate a hemolytic activity. However, a challenge to the fragment designation was debated as early as the late 70's, when the protein structure and function clearly showed the lack of conformity with the nomenclature of the

other complement activation fragments. Factor B and C2 are homologous proteins [39% sequence similarity (7)], and as such Ba is similar in structure and function to the smaller C2 fragment, and Bb is similar in structure and function to the larger C2 cleavage product. Previous attempts to amend this lack of consistency in the fragment designation [for example, as adopted in Fundamentals of Immunology (8) and editions 1–6 of Janeway's Immunobiology and other texts] were not sustained. We propose that it is time to align the C2 nomenclature with the other complement proteins (**Figure 1B**, recommended nomenclature). This becomes exceptionally apparent to the student (or instructor) of complement when given the challenge of understanding and communicating a robust system of pathways, receptors and regulators. C2 is clearly the outlier when working through the pathways of complement activation, and it adds ambiguity to a system of pathways that is already challenging to effectively communicate. The argument against adopting the conventional nomenclature for C2 is that it is established in the literature now in a non-conventional format. The counter argument is that it is estimated that our scientific output is doubling approximately every 9 years (9). With the recent resurgence of interest and therapeutic development in the complement field, it will benefit the next generation of complement biologists to learn and work within a system that is "as simple, as clear and as unambiguous as possible" (5).

C1 COMPLEXES (UNACTIVATED/ ACTIVATED/INACTIVATED)

There no longer is a specific designation for the classical complement pathway (CP) proenzymes C1r and C1s or the native zymogen C1 complex vs. activated C1r, C1s, or C1. However, to avoid confusion and therefore facilitate progress toward identification of effective therapeutic targets and therapeutic development, it is critical to accurately describe these various states of the complement components. C1 is a Ca^{++} dependent macromolecular complex comprised of C1q (itself a hexamer of trimers of 3 distinct protein chains C1qA, C1qB, and C1qC), and two molecules each of the proenzymes C1r and C1s (**Figure 1C**). In blood (or serum), most of the C1q (90%) is found in complex with the proenzymes $C1r_2C1s_2$ (10), and this is "unactivated C1" (or native zymogen C1). That is, under physiologic conditions (vs. a contrived *in vitro* situation), C1q is already complexed to $C1r_2C1s_2$ when it binds to an activator. C1q does not normally bind an activator and then "recruit C1r and C1s," as has been misstated in recent literature.

When C1q within this C1 binds to "activators," the C1q molecule is constrained in a conformation that enables C1r and C1s to be cleaved to active enzymes (C1r is autocatalytically cleaved and cleaves C1s). The activated C1s (which now converts the C1 complex to "activated C1") propagates CP activity by cleaving the next proteins in the cascade as illustrated (**Figure 1C**). C1 Inhibitor (C1-INH) is an important regulator of this enzymatic activity that rapidly binds covalently in the

TABLE 1 | (Top) Complement nomenclature as per International Complement Society (ICS) Complement Nomenclature Committee, and ICS and European Complement Network (ECN) boards recommendation from 2014 [reproduced with permission (5)].

Recommended 2014		Recommended 2014	
Name	**Comments**	**Name**	**Comments**
Pathways		**Proteins (cont.)**	
CP	Classical pathway	MBL	Mannose-binding lectin
AP	Alternative pathway	Ficolin-1	Ficolin M
LP	Lectin pathway	Ficolin-2	Ficolin L
TP	Terminal pathway (C5, C6, C7, C8, and C9)	Ficolin-3	Ficolin H
Proteins		MASP-1	MBL-associated serine protease 1
C1	Complex of C1q, 2C1r, 2C1s	MASP-2	MBL-associated serine protease 2
C1q		MASP-3	MBL-associated serine protease 3
C1r		FHL-1	Factor H-like protein 1
C1s		FHR-1	Factor H-related protein 1
C1-INH	C1 Esterase inhibitor	FHR-2	Factor H-related protein 2
C2		FHR-3	Factor H-related protein 3
C3		FHR-4	Factor H-related protein 4
C3(H2O)	Thioester-hydrolyzed form of C3	FHR-5	Factor H-related protein 5
C3a	Anaphylatoxin from C3	CD59	Protectin, Homologous restriction factor
C3b		**Protein complexes**	
iC3b	Inactivated C3b	C5b6	Terminal pathway complex of C5b + C6
C3dg		C5b-7	Terminal pathway complex of C5b6 + C7
C3d		C5b-8	Terminal pathway complex of C5b-7 + C8
C4		C5b-9	Terminal pathway complete complex
C4a		sC5b-9	Soluble C5b-9 with Vn bound
C4a-desArg	C4a without C-terminal arginine	C3bBb	AP C3 convertase
C4b		C3bBbP	AP C3 convertase with properdin
C4d		C3bBbC3b	AP C3/C5 convertase
C4BP	C4b binding protein	C4BP-Protein S	C4BP bound to protein S

(Continued)

TABLE 1 | Continued

Recommended 2014		Recommended 2014	
Name	**Comments**	**Name**	**Comments**
C5		**Receptors**	
C5a	Anaphylatoxin from C5	CR1	CD35, C3b/C4b receptor
C5a-desArg	C5a without C-terminal arginine	CR2	CD21, C3d receptor
C5b		CR3	CD11b/CD18 complex
C6		CR4	CD11c/CD18 complex
C7		C3aR	Requesting CD number
C8		C5aR1	C5aR, CD88
C9		C5aR2	C5L2, requesting CD number
Vn	Vitronectin, S protein, S40	CRIg	Complement receptor of the Ig family
FB	Factor B	C1qR	
FD	Factor D	gC1qR	Recognizes globular domains
FH	Factor H	cC1qR	Recognizes collagen domain, calreticulin
FI	Factor I	LHR	Long homologous repeat (CR1)
UPDATED COMPLEMENT NOMENCLATURE 2019			
Proteins		**Lectin pathway**	
C2a	Small C2 cleavage fragment	CL-10	Collectin-10
C2b	Large C2 cleavage fragment – enzyme	CL-11	Collectin-11
CLU	Clusterin (ApoJ, Sp40,40)	MAP-1	Previously Map44
Properdin		MAP-2	Previously Map19, sMAP
Protein complexes			
C1	$C1qr_2s_2$		
Activated C1	Activated complex (containing activated/cleaved C1s)		

One modification was made to remove 2014 proposed nomenclature for clusterin. (Bottom) (gray): updated complement nomenclature.

active catalytic site of each C1r and C1s in the activated C1 complex (i.e., four C1-INH molecules are needed to inhibit the activity of the two activated C1r and two activated C1s molecules per C1 complex). This interaction also mediates dissociation of C1r and C1s from the C1q molecule. However, there is no "inactive" form of C1q. Either after the dissociation of the activated enzymes C1r and C1s from C1q or if synthesized in tissues in the absence of C1r and C1s, C1q has many activities as described in a recent review (11) one of which, the silent clearance of apoptotic cells and cellular debris, is illustrated in **Figure 1C**.

LECTIN PATHWAY

The lectin pathway (LP) is activated by multiple PRMs and associated enzymes (12). The PRMs show specificity toward a variety of molecular patterns present on pathogens, but also on endogenous ligands. It is believed that it is the exposure, orientation and spatial distribution of the molecular structure that determines whether binding of the PRMs may lead to complement activation. The PRMs of the LP recognized so far comprise two protein families: the ficolins including ficolin-1, ficolin-2, and ficolin-3, formerly known as M-ficolin, L-ficolin, and H-ficolin, which are encoded by the *FCN1*, *FCN2,* and *FCN3* genes, respectively, and were previously assigned recommended names as shown in **Table 1**. The second PRM protein family of the LP is the so-called collectins comprising: mannose-binding lectin (also named MBL or mannan-binding lectin or protein), collectin-10 (also named CL-10, collectin liver-1, or CL-L1) and collectin-11 (also named CL-11 or collectin kidney-1, or CL-K1), which are encoded by the *MBL2*, *COLEC10,* and *COLEC11* genes, respectively. A large proportion of CL-10 and CL-11 are found as heteromeric complexes in the circulation (CL-10/CL-11 also named CL-LK). We propose that MBL remains as earlier designated and that collectin-10 and collectin-11 designate the latter two members of this family, using the abbreviations CL-10 and CL-11, respectively (**Table 1**). The above proposals relate to the PRMs of the LP in higher primates, while in lower primates and in other animal species the number and expression of LP PRMs might differ, and the nomenclature may not be directly comparable particularly for the ficolins and for MBL.

The LP PRMs circulate in the blood in complex with associated serine proteases abbreviated MASP-1,−2, and−3 after their original discovery of being associated with MBL. The *MASP1* gene encodes the serine proteases MASP-1 and MASP-3 as well as the non-enzymatically active MAP-1 (also named Map44), while the *MASP2* gene encodes MASP-2 and the non-enzymatically active MAP-2 (also named Map19 or sMAP). MASP-1,−2, and−3 are composed of an *N*-terminal heavy chain and a *C*-terminal light chain containing the serine protease domain, whereas the non-enzymatically active MAPs express unique exons, but only express part of the heavy chains and possess no serine protease domains. The different MASPs and MAPs arise from alternative splicing of the *MASP1* and *MASP2* genes. When the PRMs-MASPs complexes recognize ligands, LP complement activation is subsequently initiated upon MASP-2-mediated cleavage of C4 and C2. MASP-2 was thought mainly to be activated by autoactivation. However, recently it has been shown that MASP-1 may activate MASP-2 and cleave C2, but not C4 and is thus critical in the initiation of the LP. The function of MASP-3 has long been an enigma, but at least one of its functions appears to be cleaving pro-FD to mature active FD enabling activation of the alternative pathway. The MAPs are thought to be regulators of the activity of the LP, but this has so far only been convincingly demonstrated for MAP-1 (Map44). We propose to align the names of the proteins in the lectin pathway of complement with the common nomenclature in the gene databases. The 2014 proposed use of the terms MASP-1, MASP-2, and MASP-3 are reasonable, as are MAP-1 and MAP-2

used for the non-enzymatically active alternative splice variants of the *MASP1* and *MASP2* genes, respectively. However, we also suggest that the abbreviations MASP-1, MASP-2, MASP-3, MAP-1, and MAP-2 in the future will be the names of the proteins without the need for using the "MBL-associated" as that only explains a fraction of their associations.

CLUSTERIN

Clusterin is a multifunctional glycoprotein known in the complement field as binding to C5b-9 complexes sequestering soluble C5b-9 to prevent host membrane interaction and thus cell and tissue damage (13). However, it is also involved in clearance of misfolded proteins including amyloid ß (14), and clusterin has a risk variant associated with Alzheimer's disease in humans. There are now many publications that abbreviate clusterin as CLU (in line with its gene name, *CLU*), rather than the Cn that was recommended in the 2014 report (5), and CLU was used in the Complement Factsbook 2nd Edition (7) to designate clusterin. Thus, we recommend that the CLU, not Cn, be used as the abbreviation for clusterin.

PROPERDIN

Properdin, discovered 1954 by Pillemer et al. (15), is currently the only known positive regulator of complement activation. Properdin recognizes and binds to the C3 convertase leading to a 5–10 fold increase in the stability of this enzyme complex (16). Aside from increasing the half-life of the C3 convertase, properdin—after interacting with specific glycosaminoglycans—can also directly initiate complement activation on some altered self-surfaces, such as apoptotic cells, by directing C3b deposition (17). Abbreviating properdin as FP (Factor Properdin) to bring it in line with FB, FD, FH, and FI, was briefly considered among the complement community but was abandoned as the overwhelming majority of publications used, and uses, the term properdin without any arising issues. Thus, we recommend that properdin be used as the sole term.

To conclude, we hope that the updates proposed here for cleavage fragments, PRMs, activation states and regulatory proteins of the complement system will enhance communication and thus understanding of both the basic complement pathways and consequences of activation or lack thereof, as well as, the newly discovered nuances of the complement system in the classroom, in research, in pharma and in the boardroom. It is hoped that the simplified uniform nomenclature of this intricate system will facilitate therapeutic development and appropriate application to the clinic. We propose that the ICS consider and endorse these changes and submit them to the IUIS Nomenclature Committee.

AUTHOR CONTRIBUTIONS

AT initiated work. SB, PG, CK, and AT contributed to draft, revision, and approval of final version.

FUNDING

This contribution was supported by funds from NIH National Institute of Allergy and Infectious Diseases R15AI117474-01A1 (SB), the Division of Intramural Research, National Heart, Lung, and Blood Institute, NIH (CK) as well as Danish Research Foundation of Independent Research [DFF-6110-00489], the Novo Nordisk Research Foundation and the Sven Andersen Research Foundation (PG) and NIH National Institute on Aging R01 AG060148 (AT).

ACKNOWLEDGMENTS

The authors wish to acknowledge previous work from the International Complement Society (ICS) Complement Nomenclature Committee, and the ICS and European Complement Network (ECN) Boards that inspired continued improvement and refinement of complement nomenclature. The recommendations provided here are proposals from the authors to the complement community and were assembled independently of established committees and boards.

REFERENCES

1. Ricklin D, Lambris JD. Complement in immune and inflammatory disorders: pathophysiological mechanisms. *J Immunol.* (2013) 190:3831–8. doi: 10.4049/jimmunol.1203487
2. Harboe M, Mollnes TE. The alternative complement pathway revisited. *J Cell Mol Med.* (2008) 12:1074–84. doi: 10.1111/j.1582-4934.2008.00350.x
3. Hajishengallis G, Reis ES, Mastellos DC, Ricklin D, Lambris JD. Novel mechanisms and functions of complement. *Nat Immunol.* (2017) 18:1288–98. doi: 10.1038/ni.3858
4. Kemper C, Kohl J. Back to the future - non-canonical functions of complement. *Semin Immunol.* (2018) 37:1–3. doi: 10.1016/j.smim.2018.05.002
5. Kemper C, Pangburn MK, Fishelson Z. Complement nomenclature 2014. *Mol Immunol.* (2014) 61:56–8. doi: 10.1016/j.molimm.2014.07.004
6. Rapp HJ, Borsos T. Complement and Hemolysis. *Science.* (1963) 141:738–40. doi: 10.1126/science.141.3582.738
7. Barnum SR, Schein TN. *The Complement Factsbook.* London: Elsevier/Academic Press (2018).
8. Liszewski MK, Atkinson JP. The complement system. In: Paul WE, editor. *Fundamental Immunology.* New York, NY: Raven Press Ltd (1993). p. 917–39.
9. Bornmann L, Mutz R. Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. *J Assoc Inform Sci Technol.* (2015) 66:2215–22. doi: 10.1002/asi.23329
10. Ziccardi RJ, Tschopp J. The dissociation properties of native C1. *Biochem Biophys Res Commun.* (1982) 107:618–23. doi: 10.1016/0006-291X(82)91536-4
11. Thielens NM, Tedesco F, Bohlson SS, Gaboriaud C, Tenner AJ. C1q: a fresh look upon an old molecule. *Mol Immunol.* (2017) 89:73–83. doi: 10.1016/j.molimm.2017.05.025
12. Garred P, Genster N, Pilely K, Bayarri-Olmos R, Rosbjerg A, Ma YJ, et al. A journey through the lectin pathway of complement-MBL and beyond. *Immunol Rev.* (2016) 274:74–97. doi: 10.1111/imr.12468

13. Tschopp J, Chonn A, Hertig S, French LE. Clusterin, the human apolipoprotein and complement inhibitor, binds to complement C7, C8 beta, and the b domain of C9. *J Immunol.* (1993) 151:2 159–65.
14. Nelson AR, Sagare AP, Zlokovic BV. Role of clusterin in the brain vascular clearance of amyloid-beta. *Proc Natl Acad Sci USA.* (2017) 114:8681–2. doi: 10.1073/pnas.17113 57114
15. Pillemer L, Blum L, Lepow IH, Ross OA, Todd EW, Wardlaw AC. The properdin system and immunity. I Demonstration and isolation of a new serum protein, properdin, and its role in immune phenomena. *Science.* (1954) 120:279–85. doi: 10.1126/science.120.31 12.279
16. Fearon DT, Austen KF. Properdin: binding to C3b and stabilization of the C3b-dependent C3 convertase. *J Exp Med.* (1975) 142:856–63. doi: 10.1084/jem.142.4.856
17. Kemper C, Mitchell LM, Zhang L, Hourcade DE. The complement protein properdin binds apoptotic T cells and promotes complement activation and phagocytosis. *Proc Natl Acad Sci USA.* (2008) 105:9023–8. doi: 10.1073/pnas.0801015105

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

MINI REVIEW
published: 06 September 2019
doi: 10.3389/fimmu.2019.02098

A Proposal for Nomenclature in Myeloid C-Type Lectin Receptors

Carlos del Fresno, Francisco J. Cueto and David Sancho**

Immunobiology Lab, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain

Myeloid C-type lectin receptors (CLRs) comprise a family of receptors expressed by immune myeloid cells that share homologous C-type lectin domains. The implication of these CLRs in the regulation of homeostasis and activation of myeloid cells has generated a buoyant growth in the number of studies involving these receptors. Since their first description, diverse nomenclature has been used to refer to each of them, ranging from systematic classifications, such as gene name or cluster of differentiation, to non-systematic ones that include terminology based on gene expression patterns or function. In this review, we aim to summarize the different names used for the main myeloid CLRs and analyze which of them have been more frequently used in the literature. In addition, we have examined the evolution of the terminology applied to these myeloid CLRs over time. Based on this analysis, we propose a *consensus alias* for each of those myeloid CLRs. However, we acknowledge that systematicity is required beyond this terminology based on use frequency. Therefore, we have included gene names as the standardization tool to gather the maximum agreement. We suggest that a standard nomenclature consisting of both gene names and consensus alias should be included at least in scientific abstracts, which would help to identify relevant literature, saving time and effort and fostering the research in this field in a more systematic manner.

Keywords: lectin receptors, signaling, monocytes, macrophages, dendritic cells, innate immunity, nomenclature

OPEN ACCESS

Edited by:
Pablo Engel,
University of Barcelona, Spain

Reviewed by:
Georgina Clark,
University of Sydney, Australia
Cristina Lopez-Rodriguez,
Pompeu Fabra University, Spain

****Correspondence:***
Carlos del Fresno
carlos.delfresno@cnic.es
David Sancho
dsancho@cnic.es

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *21 June 2019*
Accepted: *20 August 2019*
Published: *06 September 2019*

Citation:
del Fresno C, Cueto FJ and Sancho D (2019) A Proposal for Nomenclature in Myeloid C-Type Lectin Receptors. Front. Immunol. 10:2098. doi: 10.3389/fimmu.2019.02098

BROAD DIVERSITY OF MYELOID C-TYPE LECTIN RECEPTORS AND BABYLONIAN CONFUSION IN THEIR NAMING

Innate immune cells surveil their nearby microenvironment, reacting against different challenges when they are identified. This reaction is based in a toolbox consisting of a plethora of pattern recognition receptors (PRRs) capable of sensing both pathogen-associated molecular patterns (PAMPs) present in invading microbes (1) and damage-associated molecular patterns (DAMPs), which are molecules released by stressed or necrotic cells (2). Ligation of PRRs by their ligands initiates intracellular signaling pathways that modulate innate and adaptive immune responses. Among the described PRRs, we will focus here in C-type lectin receptors (CLRs) expressed by myeloid cells such as monocytes, macrophages, or dendritic cells.

C-type lectin receptors (CLRs) comprise a large family of metazoan proteins (more than 1,000) characterized by containing at least one C-Type Lectin-like Domain (CTLD) (3). In order to organize the large number of receptors included in this superfamily, diverse classifications have been proposed. CLRs were early classified into 17 groups based on their structure (3).

As the functional relevance of this structural classification is limited, an alternative classification of myeloid transmembrane CLRs was proposed based on their intracellular signaling domains, namely Immunoreceptor Tyrosine-based Activation Motif (ITAM) domains, hemITAM domains, Immunoreceptor Tyrosine-based Inhibition Motif (ITIM) domains or CLRs without clear ITAM or ITIM domains (4, 5).

Moreover, the abundance and diversity in function and expression of myeloid CLRs has also contributed to a great dispersion in the way that researchers name these receptors. These myeloid CLRs were initially named based on their expression pattern or gene location (e.g., DEndritic cell-associated C-type lecTIN-1 for Dectin-1 (6), Dendritic cell NK lectin Group Receptor-1 for DNGR-1 (7), Macrophage-INducible C-type LEctin for MINCLE (8). Later on, a serial nomenclature was adopted for the naming of genes from the CLR family, based on their common domain, so that they are all cataloged as CLEC (C-type LECtin) followed by an alphanumerical identifier (9). In an attempt to avoid confusion when identifying surface markers, Human Leukocyte Differentiation Antigen Workshops were organized to standardize the naming of markers that are recognized by specific monoclonal antibodies. The result of this effort was the definition of the cluster of differentiation (CD) nomenclature (10), which also applies to CLRs, such as CD303 for *CLEC4C*/BDCA-2 (11). The last of these workshops hold in 2014 represented the tenth of these events and provided CD nomenclature for some more CLRs, reaching the CD371 for CLEC12A (12).

This diverse nomenclature used for myeloid CLRs may generate confusion when searching or disseminating information as illustrated for CLRs expressed on dendritic cells, where up to seven different names can be found for some of these receptors (13). In an attempt to systematically analyze potential solutions to this Babylonian confusion in the myeloid CLR field, we have herein listed the different names used to identify myeloid CLRs at the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov). We have studied the prevalence of each of these names and the evolution of their usage over time in title, abstract, and keywords of published works. We decided to delimit the survey to these sections of the manuscripts as they are the main sections where scientists search literature of their interest using the PubMed search tool from NCBI (14). Taking into account this information, we propose a standard nomenclature consisting of a *consensus alias* for each myeloid CLR based on the most frequently used in the current literature. In any case, our study illustrates the need for systematization in the naming of myeloid CLRs, and thus we propose that the gene name (based on the CLEC nomenclature) should always accompany the "consensus" identifier at least in the abstract. Importantly, the official nomenclature for naming of human genes is provided by the Human Genome Nomenclature Committee (HGNC) (www.genenames.org), while official mouse gene names come from the Mouse Genome Informatics (MGI) (www.informatics.jax.org).

ANALYSIS OF NAMES USED FOR MYELOID CLRS

In order to study the use of the different denominations, we selected a list of myeloid CLRs grounded on our former review on the flexibility of these sensors to trigger different signaling pathways (5). Based on the "Gene" resource of NCBI (https://www.ncbi.nlm.nih.gov/gene/) we listed all the potential names (aliases) used for the mouse and human version of each selected receptor, performing the search based on their gene names, and organizing the CLRs based on their functional intracellular domains as previously proposed (4, 5) (**Figure 1**).

Next, we completed a search based on the PubMed resource of NCBI (https://www.ncbi.nlm.nih.gov/pubmed/, performed along the second and third week of March 2019) for any of the provided aliases for each receptor. The scope of this survey was to obtain the total number of references where each specific name has been used either alone or combined with other aliases for the same CLR, generating usage frequencies for each of these terms (**Figure 1**). Colored bars represent names used in more than 5% of total articles referring to that CLR, except for gene names as a source of systematicity, which are always depicted in red independently of their frequency. An additional layer of confusion for naming myeloid CLRs relies on the use or not of hyphens in their names. For our study, those aliases found in both versions were clustered as a single search using the "OR" command (**Figure 1**).

A quick view clearly illustrates the variability of alias type across CLRs. It can range from receptors always appointed with the same alias, either their gene (*CLEC12B*) or alternative (DCAR1) name, to members identified with up to five different aliases in frequencies over 5% such as *CLEC10A*, CD301, MGL, CD301a, and MGL1 (and even found named as HML or CLECSF14 in minor proportions). This dispersion occurs regardless the classification based on intracellular domains.

This analysis highlights the need for a systematic nomenclature for myeloid CLRs. This is particularly important for those receptors expressed by diverse coding genes in different species, but commonly found with a shared alias. This is the case of ITIM-bearing DCIR, encoded by *CLEC4A* in humans and by *Clec4a2* in mice. In both cases, DCIR is one of the accepted aliases and is thus an unspecific name. Contrarily, and still using DCIR as an example, some names exclusively designate the human (LLIR) or mouse (DCIR1) receptors, but these aliases are not among the most frequent. Therefore, for our study, we have distinguished between "human DCIR" and "mouse DCIR," as they are encoded by genes with a different name. These same criteria were applied to an additional name for DCIR, CLECSF6, also shared for human and mouse receptors. In addition, mouse DCIR is linked to several isoforms, mainly those encoded by the genes *Clec4a2* and *Clec4a4* and literature on these isoforms may be confusing if the genes are not named.

The use of non-official aliases, even not included at the NCBI gene database, constitutes another anomalous situation. This is the case for the non-ITAM/ITIM coupled CD207, commonly known as LANGERIN, with CD207 used in 43,82% of references

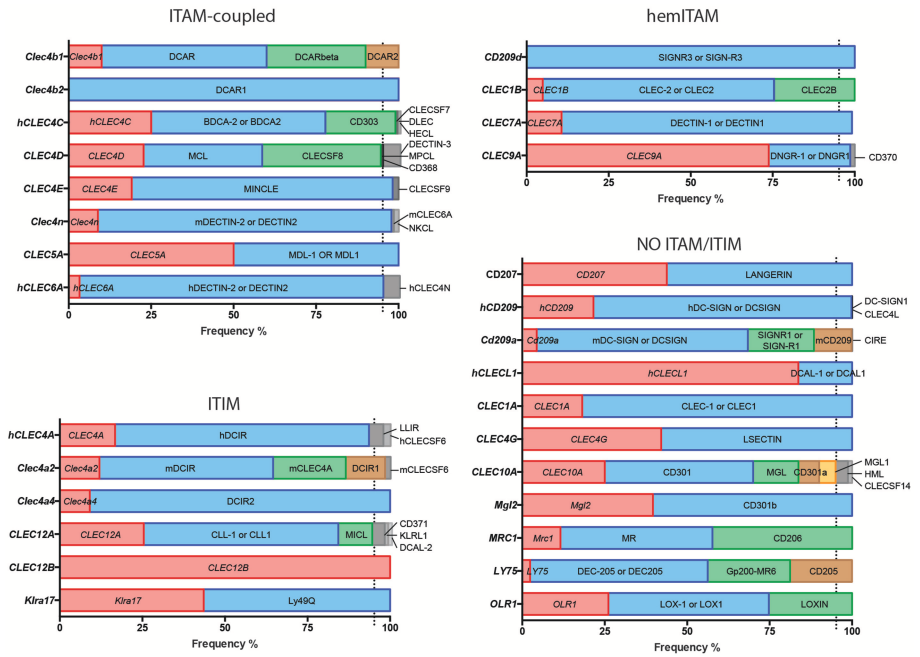

FIGURE 1 | Usage frequencies of the different aliases provided by NCBI for every myeloid C-type lectin receptor surveyed in this review. Receptors are grouped based on their intracellular domain and listed in alphabetical order of their gene name. Colored bars represent names with usage frequencies higher than 5%, except for gene names, which are always in red regardless their prevalence.

in PubMed and LANGERIN in the remaining 56,17%. Taking this into account, for this receptor in particular, we have performed our analysis including both names, even the one not recorded as a gene alias at the Gene tool (LANGERIN). The name LANGERIN comes from the specific expression of this CLR by Langerhans cells. These cells show unique intracellular structures called Birbeck granules whose presence is associated with LANGERIN/CD207 expression (15). Therefore, this CLR can be found both intracellularly and extracellularly (16). Taking this into account, we wondered whether the use of LANGERIN or CD207 correlates with the detection of the receptor by histologic methods often used for intracellular staining or flow cytometry, usually more associated to extracellular detection. To approach it, we crossed in Pubmed either LANGERIN or CD207 with "histology" or "flow cytometry." Interestingly, both aliases were more frequently found associated to histology (84,6% for LANGERIN and 82,86% for CD207). The same happened when the search was performed combining LANGERIN AND CD207 (84.7%). This would mean that histologic techniques are preferred for studies involving this CLR and that both names are used indistinctly.

A combination of these circumstances occurs for the ITAM-coupled DECTIN-2. According to the Gene resource, DECTIN-2 or DECTIN2 refers only to the product of the human *CLEC6A* gene but not to the mouse *Clec4n*. However, it is quite common to find references in the literature quoting DECTIN-2 for the mouse version, which can be specifically found as NKCL. Still, CLEC4N is also recorded as an alternative name for the human version. Therefore, we have differentiated between "human DECTIN-2 or DECTIN2" and "mouse DECTIN-2 or DECTIN2," applying the same criteria to CLEC4N and CLEC6A.

Another representative example of the divergence in mouse and human terms for a myeloid CLR is DC-SIGN, a member of these receptors not coupled to an identifiable ITAM/ITIM domain. The genes encoding for this receptor are named differently in human (*CD209*) and mouse (*Cd209a*). DC-SIGN/DCSIGN can be widely found referring to both forms, consequently, we have also differentiated between "human DC-SIGN or human DCSIGN" and "mouse DC-SIGN or mouse DCSIGN." In this case, the mouse receptor is specifically identified as SIGNR1 or SIGN-R1, but CD209 is also considered an alias for the mouse version. Therefore, we have differentiated between human and mouse CD209 for our analysis. In any case, to some extent, we assume some overlapping on the results generated by these searches based on "human" or "mouse." It is interesting to see how this CLR was initially described as a membrane-associated mannose-binding receptor for the HIV gp120 protein (17), with no specific name until it was first identified as DC-SIGN (18). Therefore, our analysis applies to the use of different nomenclatures for CLRs, although biological information about some of them could have been generated before their current names were coined.

It is also worthy to comment that the frequency of use of a name can be influenced by the contribution of a particular author to the study of a CLR, specially when the literature is not abundant. Therefore, the most

frequently used nomenclature may be biased by the publishing frequency of one author on a particular CLR. This notion emphasizes the need for an agreement in the way of naming myeloid CLRs.

Indeed, in light of the results shown in **Figure 1**, it is clear that defining a systematic manner to refer to myeloid CLRs is mandatory. In that sense, we encourage that, independently of how we prefer naming our favorite myeloid CLR, we should all

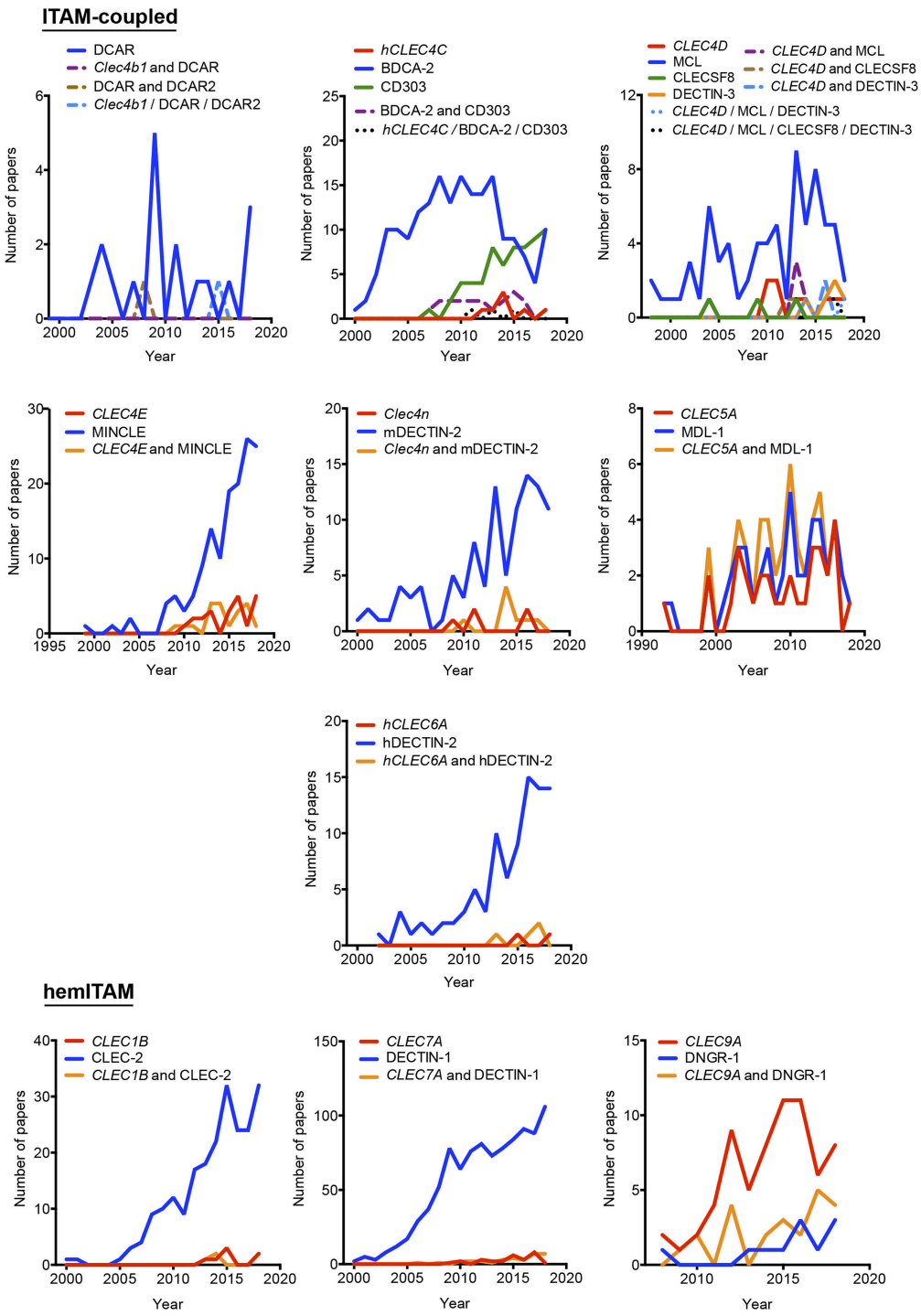

FIGURE 2 | Temporal analysis of the use of different aliases for ITAM- and hemITAM-coupled myeloid C-type lectin receptors. The frequency of each name was analyzed either alone or in combination with the alternative aliases.

include the gene name (as listed in **Figure 1**) to unequivocally and systematically identify the receptor, both in the abstract and the first time that they are mentioned in the text.

NOMENCLATURE EVOLUTION OF MYELOID C-TYPE LECTIN RECEPTORS

Our proposal assumes that there are some common and traditionally used nomenclatures that do not include the systematicity provided by the gene name. To define the temporal evolution of preferred names for myeloid CLRs, we analyzed the usage of these names (included in **Figure 1**) in the last years, which could suggest a consensus alias for each myeloid CLR. We analyzed the number of works naming a particular term either in the title, abstract, or keywords of published works. The search results were exported and a table including complete title, abstract, and keywords for each of the found items was generated using the reference manager software JabRef (www.jabref.org). Note that for each CLR, we incorporated the "NOT" command to search for literature where each specific alias had been used exclusively, but not the other aliases. We determined the use of each alias in any of these three sections for every paper. For the sake of clarity and unless it applies to gene names, we only analyzed aliases with frequencies over 5% for timeline analysis. Next, by using the "AND" command, we analyzed the number of simultaneous appearances among the selected aliases, in order to study their combined usage. Results are shown in **Figure 2** for ITAM- or hemITAM-coupled myeloid CLRs, **Figure 3** for ITIM-bearing and **Figure 4** for those receptors that do not bear any ITAM or ITIM domain. Those receptors with a single alias in **Figure 1** have not been included in this study.

Timeline analysis may not completely match the results included in **Figure 1**; Jabref exploits the Machine readable digital library (Mr. DLib) for their searches while PubMed search engine is based on the Medical subjet headings (MeSH) algorithm (19). This fact could introduce some variability. On the other hand, the analysis performed in **Figure 1** did not include the "NOT" command; therefore, those references where a myeloid CLR is identified by two (or more) aliases, were included in the total number accounting for both names. However, in the temporal analysis, those examples will be only included in the combinations, identified by the use of the "AND" command. The CLR devoid of ITAM/ITIM motif *MRC1* is a representative example, where all quotations in the temporal analysis using this identification are combined with other aliases. This would also suggest the inclusion of a consensus alias at least, in the abstract of manuscripts, because if we consider the growing amount of literature that we need to handle, representative abstracts are becoming fundamental for the selection of relevant papers in scientific research.

As indicated above, hyphens have been used almost randomly in some aliases to separate name and numbers. Both versions are documented in **Figure 1** as, for example, BDCA-2 or BDCA2. For this reason, we included both styles in the temporal analysis, and only the most prevalent version is depicted in **Figures 2–4**. This study illustrates that for all the instances except SIGNR3/SIGN-R3 and SIGNR1/SIGN-R1, the version including hyphen is the predominant one and consequently, we encourage its use. The limitation in the number of characters accepted in certain

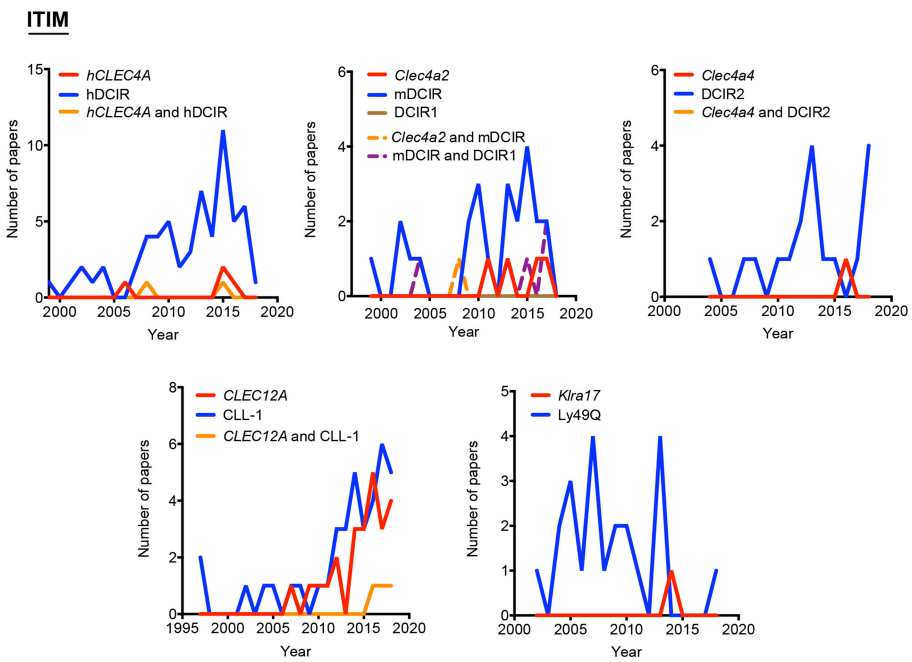

FIGURE 3 | Use of different names along time for myeloid C-type lectin receptors bearing an ITIM motif in their intracellular domain. The frequency of each alias was analyzed either alone or in combination with their alternatives.

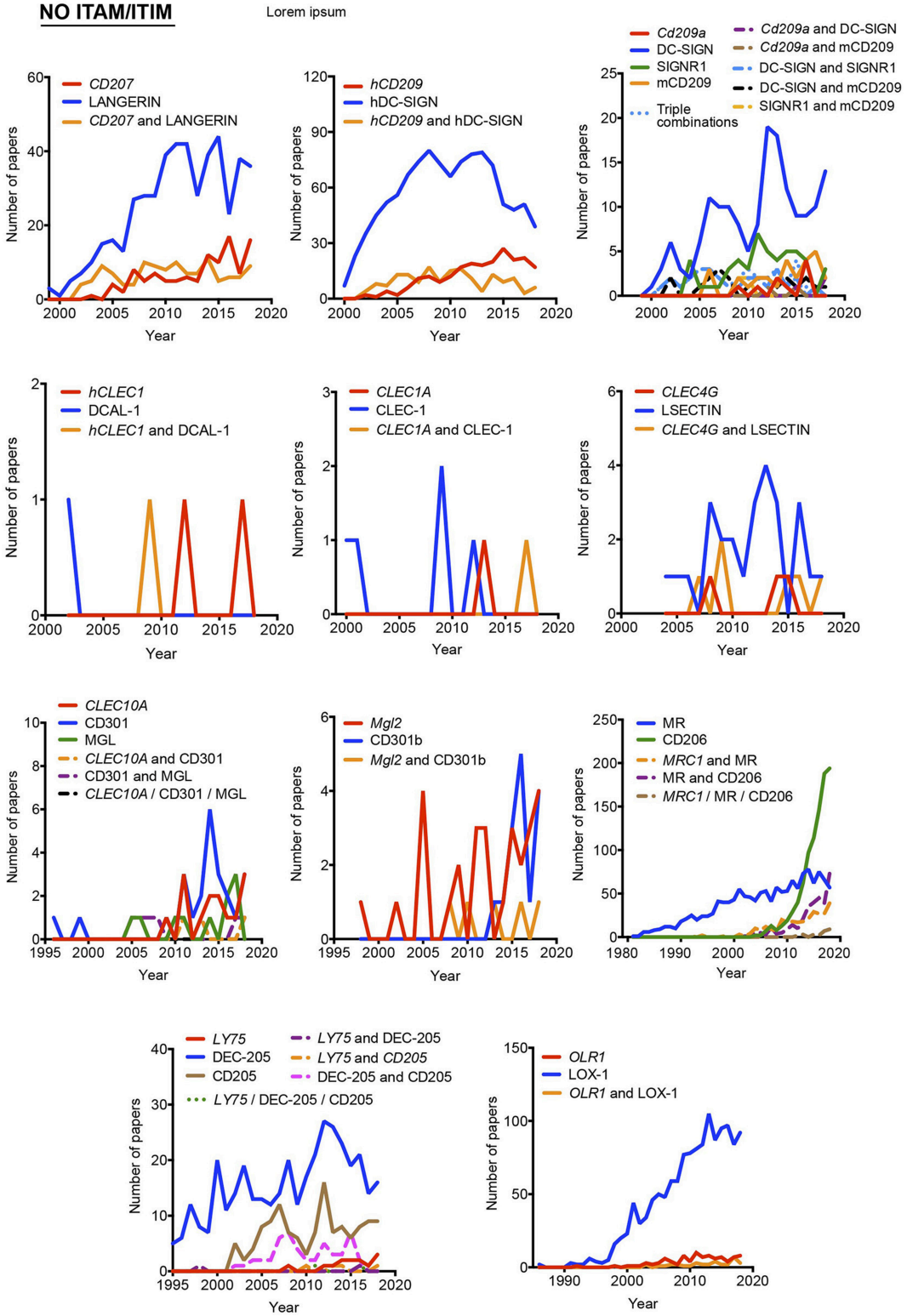

FIGURE 4 | Usage evolution of different aliases for C-type lectin receptors not coupled to an identifiable ITAM or ITIM motif in their intracellular domain. The frequency of each name was analyzed either alone or in combination with the alternative aliases.

TABLE 1 | Proposed consensus alias for myeloid C-type lectin receptors.

ITAM-COUPLED	
Clec4b1	DCAR
Clec4b2	DCAR1
CLEC4C	CD303
CLEC4D	MCL
CLEC4E	MINCLE
Clec4n	mDECTIN-2
CLEC5A	MDL-1
CLEC6A	hDECTIN-2
ITIM	
CLEC4A	hDCIR
Clec4a2	mDCIR
Clec4a4	DCIR2
CLEC12A	CLL-1
CLEC12B	CLEC12B
Klra17	Ly49Q
hemITAM	
CD209d	SIGNR3
CLEC1B	CLEC-2
CLEC7A	DECTIN-1
CLEC9A	DNGR-1
NO ITAM/ITIM	
CD207	LANGERIN
CD209	hDC-SIGN
CD209a	mDC-SIGN
CLECL1	DCAL-1
CLEC1A	CLEC-1
CLEC4G	LSECTIN
CLEC10A	CD301
Mgl2	CD301b
MCR1	CD206
LY75	DEC-205
OLR1	LOX-1

To facilitate identification, both gene name and alias should be named in the abstract of the manuscript.

journals might be behind the use of names lacking the hyphen. However, authors should adhere to consensus aliases to make their research more easily identifiable.

The temporal analysis illustrates that common names have been more frequently used than gene names as a systematic manner to identify myeloid CLRs. *CLEC9A* vs. DNGR-1 represent the only exception of a CLR more often identified by its gene name. The use of *CLEC9A* is extended as a gene marker for cDC1s (20, 21), while the use of DNGR-1 is linked to functional studies on this receptor (22, 23). *CLEC5A* shows a peculiar behavior, with virtually the same frequency for exclusive use of either the gene name or MDL-1 or the combination of both of them.

In general, gene names and main aliases are not combined together in the abstract, as we propose, which contributes to confusion. Thus, as indicated before, the same common name (DC-SIGN) has been applied to the mouse and human versions without distinction, although they are encoded by different genes and show different expression patterns (24).

Interestingly, the study of the temporal evolution also allows for the detection of preferred names for CLRs along the time, which could be an extra criterium to propose a consensus alias. In this line, it is remarkable the use of the cluster of differentiation nomenclature for some receptors such as *CLEC4C*/CD303, *CLEC10A*/CD301, *Mgl2*/CD301b, *MRC1*/CD206, and *LY75*/CD205. Once defined in a HLDA workshop, their frequencies overcame any of their other aliases. The best example is CD206 for the mannose receptor *MRC1*, used both alone and in combination with other aliases. This is because the first reference to a CD occurs after monoclonal antibodies are submitted to a HLDA workshop and, from there onwards, the use of the CD nomenclature for a specific receptor begins to be applied for multiple applications (25).

PROPOSAL OF CONSENSUS NOMENCLATURE FOR MYELOID CLRS

Considering all our analysis, **Table 1** compiles our proposal for the consensus alias that should be used when referring to any myeloid CLRs surveyed in this review. This proposal is based both in the total frequency (**Figure 1**) and temporal evolution (**Figures 2**–**4**) of their usage. In order to introduce systematicity, we propose that the current most frequent alias should be always accompanied by the specific gene name. Furthermore, we encourage the use of hyphens when required, as they are more frequently used.

Both the gene name and the proposed consensus alias (**Table 1**) should appear at least in the abstract of manuscripts or meetings. This practice would facilitate the identification of literature of interest, fostering the visibility of any work in their research field.

AUTHOR CONTRIBUTIONS

CF, FC, and DS: conceptualization. CF and FC: methodology, analysis, and investigation. CF: writing–original draft and preparation of figures. FC and DS: writing–review and editing. CF and DS: funding acquisition and supervision.

FUNDING

CF was supported by AECC Foundation (INVES192DELF). Work in the DS laboratory was funded by the CNIC and grant SAF2016-79040-R from Ministerio de Ciencia, Innovación e Universidades (MCIU), Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (FEDER); B2017/BMD-3733 Immunothercan-CM from Comunidad de Madrid; RD16/0015/0018-REEM from FIS-Instituto de Salud Carlos III, MICINN, and FEDER; Acteria Foundation; Constantes y Vitales prize (Atresmedia); La Marató de TV3 Foundation (201723); the European Research Council (ERC-2016-Consolidator Grant 725091); and the European Commission (635122-PROCROP H2020). The CNIC is supported by the MCIU and the Pro-CNIC Foundation, and is a Severo Ochoa Center of Excellence (SEV-2015-0505).

REFERENCES

1. Janeway CA. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol*. (1989) 54(Pt 1):1–13. doi: 10.1101/SQB.1989.054.01.003
2. Matzinger P. The danger model: a renewed sense of self. *Science*. (2002) 296:301–5. doi: 10.1126/science.1071059
3. Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. *FEBS J*. (2005) 272:6179–217. doi: 10.1111/j.1742-4658.2005.05031.x
4. Sancho D, Reis e Sousa C. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. *Annu Rev Immunol*. (2012) 30:491–529. doi: 10.1146/annurev-immunol-031210-101352
5. del Fresno C, Iborra S, Saz-Leal P, Martínez-López M, Sancho D. Flexible signaling of myeloid C-type lectin receptors in immunity and inflammation. *Front Immunol*. (2018) 9:804. doi: 10.3389/fimmu.2018.00804
6. Ariizumi K, Shen GL, Shikano S, Xu S, Ritter R, Kumamoto T, et al. Identification of a novel, dendritic cell-associated molecule, dectin-1, by subtractive cDNA cloning. *J Biol Chem*. (2000) 275:20157–67. doi: 10.1074/jbc.M909512199
7. Sancho D, Mourão-sá D, Joffre OP, Schulz O, Rogers NC, Pennington DJ, et al. Tumor therapy in mice via antigen targeting to a novel, DC-restricted C-type lectin. *J Clin Invest*. (2008) 118:2098–110. doi: 10.1172/JCI34584DS1
8. Matsumoto M, Tanaka T, Kaisho T, Sanjo H, Copeland NG, Gilbert DJ, et al. A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. *J Immunol*. (1999) 163:5039–48.
9. Colonna M, Samaridis J, Angman L. Molecular characterization of two novel C-type lectin-like receptors, one of which is selectively expressed in human dendritic cells. *Eur J Immunol*. (2000) 30:697–704. doi: 10.1002/1521-4141(200002)30:2<697::AID-IMMU697>3.0.CO;2-M
10. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD nomenclature 2015: human leukocyte differentiation antigen workshops as a driving force in immunology. *J Immunol*. (2015) 195:4555–63. doi: 10.4049/jimmunol.1502033
11. Röck J, Schneider E, Grün JR, Grützkau A, Küppers R, Schmitz J, et al. CD303 (BDCA-2) signals in plasmacytoid dendritic cells via a BCR-like signalosome involving Syk, Slp65 and PLCγ2. *Eur J Immunol*. (2007) 37:3564–75. doi: 10.1002/eji.200737711
12. Clark G, Stockinger H, Balderas R, van Zelm MC, Zola H, Hart D, et al. Nomenclature of CD molecules from the tenth human leucocyte differentiation antigen workshop. *Clin Transl Immunol*. (2016) 5:e57 doi: 10.1038/cti.2015.38
13. Kanazawa N. Dendritic cell immunoreceptors: C-type lectin receptors for pattern-recognition and signaling on antigen-presenting cells. *J Dermatol Sci*. (2007) 45:77–86. doi: 10.1016/j.jdermsci.2006.09.001
14. Delvin E, Pillay TS, Newman A. How to write a scientific paper: practical guidelines. *FEBS J*. (2016) 283:3882–5. doi: 10.1111/febs.13918
15. Valladeau J, Dezutter-Dambuyant C, Saeland S. Langerin/CD207 sheds light on formation of birbeck granules and their possible function in langerhans cells. *Immunol Res*. (2003) 28:93–107. doi: 10.1385/IR:28:2:93
16. Mc Dermott R. Birbeck granules are subdomains of endosomal recycling compartment in human epidermal langerhans cells, which form where langerin accumulates. *Mol Biol Cell*. (2002) 13:317–35. doi: 10.1091/mbc.01-06-0300
17. Curtis BM, Scharnowske S, Watson AJ. Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120. *Proc Natl Acad Sci USA*. (1992) 89:8356–60. doi: 10.1073/pnas.89.17.8356
18. Geijtenbeek TB, Kwon DS, Torensma R, van Vliet SJ, F van Duijnhoven GC, Middel J, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell*. (2000) 100:587–97. doi: 10.1016/s0092-8674(00)80694-7
19. Ecker ED, Skelly AC. Conducting a winning literature search. *Evid Based Spine Care J*. (2010) 1:9–14. doi: 10.1055/s-0028-1100887
20. Piva L, Tetlak P, Claser C, Karjalainen K, Renia L, Ruedl C. Cutting edge: Clec9A+ dendritic cells mediate the development of experimental cerebral malaria. *J Immunol*. (2012) 189:1128–32. doi: 10.4049/jimmunol.1201171
21. Gilfillan CB, Kuhn S, Baey C, Hyde EJ, Yang J, Ruedl C, et al. Clec9A + dendritic cells are not essential for antitumor CD8 + T cell responses induced by poly I:C immunotherapy. *J Immunol*. (2018) 200:2978–86. doi: 10.4049/jimmunol.1701593
22. del Fresno C, Martínez-cano S, Blanco-menéndez N, Schulz O, Gallizioli M. DNGR-1 in dendritic cells limits tissue damage by dampening neutrophil recruitment. *Science*. (2018) 356:351–6. doi: 10.1126/science.aan8423
23. Zelenay S, Keller AM, Whitney PG, Schraml BU, Deddouche S, Rogers NC, et al. The dendritic cell receptor DNGR-1 controls endocytic handling of necrotic cell antigens to favor cross-priming of CTLs in virus-infected mice. *J Clin Invest*. (2012) 122:1615–27. doi: 10.1172/JCI60644
24. Kalay H, Unger WWJ, García-Vallejo JJ, van Kooyk Y, Fehres CM. Glycan-based DC-SIGN targeting vaccines to enhance antigen cross-presentation. *Mol Immunol*. (2012) 55:143–5. doi: 10.1016/j.molimm.2012.10.031
25. Gez S, Barber N, Woolfson A, Belov L, Christopherson RI, Mulligan SP. Profiling CD antigens on leukaemias with an antibody microarray. *FEBS Lett*. (2009) 583:1785–91. doi: 10.1016/j.febslet.2009.03.018

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 04 October 2019
doi: 10.3389/fimmu.2019.02287

Mind the Map: Technology Shapes the Myeloid Cell Space

Patrick Günther[1,2] *and Joachim L. Schultze*[1,2]*

[1] Genomics and Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany, [2] Platform for Single Cell Genomics and Epigenomics, German Center for Neurodegenerative Diseases and University of Bonn, Bonn, Germany

The myeloid cell system shows very high plasticity, which is crucial to quickly adapt to changes during an immune response. From the beginning, this high plasticity has made cell type classification within the myeloid cell system difficult. Not surprising, naming schemes have been frequently changed. Recent advancements in multidimensional technologies, including mass cytometry and single-cell RNA sequencing, are challenging our current understanding of cell types, cell subsets, and functional states of cells. Despite the power of these technologies to create new reference maps for the myeloid cell system, it is essential to put these new results into context with previous knowledge that was established over decades. Here we report on earlier attempts of cell type classification in the myeloid cell system, discuss current approaches and their pros and cons, and propose future strategies for cell type classification within the myeloid cell system that can be easily extended to other cell types.

Keywords: monocytes, dendritic cells, human peripheral blood, multidimensional, single-cell RNA sequencing, mass cytometry

OPEN ACCESS

Edited by:
Loems Ziegler-Heitbrock, Independent Researcher, Munich, Germany

Reviewed by:
Marc Dalod, Centre National de la Recherche Scientifique (CNRS), France
Alexander Mildner, Max Delbrück Center for Molecular Medicine, Helmholtz Association of German Research Centers (HZ), Germany

***Correspondence:**
Joachim L. Schultze
j.schultze@uni-bonn.de

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *01 July 2019*
Accepted: *10 September 2019*
Published: *04 October 2019*

Citation:
Günther P and Schultze JL (2019) Mind the Map: Technology Shapes the Myeloid Cell Space. Front. Immunol. 10:2287. doi: 10.3389/fimmu.2019.02287

INTRODUCTION

Cell-type identification is an integral part of current immunology (1–5). The immune system as an organ is an assembly of an incredibly complex network of different types of immune cells including T and B lymphocytes, NK cells, innate lymphoid cells, monocytes, macrophages, and dendritic cells (DC), granulocytes including neutrophils, basophils and eosinophils, and mast cells (6). These cell types have specialized roles during homeostasis and infection. Moreover, it became clear that each of these significant immune cell types consists of cell type-specific cell subsets, for example, three monocyte subsets have been described in human peripheral blood, the so-called classical, intermediate, and non-classical monocyte (7). To understand the individual role of each of these subsets, it is crucial to understand the full heterogeneity of these cell types and their subsets to pinpoint the dedicated functions (8). This also needs to be considered in a spatiotemporal fashion, since immune cells are influenced in their function by their respective microenvironment as well as over time (9–11). For example, monocytes accumulate in peripheral reservoirs under homeostatic conditions, but during inflammation, they exert primarily pro-inflammatory effector functions (11–13). At a later time point during the repair phase of an inflammatory response, monocytes are characterized by regulatory properties necessary for tissue repair (14). During the last decade, technological advancements have been used to further refine our understanding of the diversity of cell types and subgroups within the immune system (15). These novel technologies must be put into context with the traditional way of defining cell types mostly relying on low-dimensional data

including microscopy, functional assays, and expression of single marker genes. In the first part of the review, we discuss the current principles and strategies of defining cell types and subsets, while highlighting the different aspects of resolving cellular heterogeneity. Here we want to outline how these principles have been applied to the DC/monocyte cell space. Moreover, we will provide a framework for the integration of these recent technological advances to define cell types, subsets, but also functional states of these subsets in an iterative process.

THE MONONUCLEAR MYELOID CELL SPACE AS AN EXAMPLE FOR CELL TYPE DEFINITION

Monocytes and DC arise from the myeloid lineage of the hematopoietic system and makeup about 11% of human blood leukocytes (monocytes ∼10%, DC ∼1%). In humans, monocytes and DC are defined as $MHCII^{+}CSF\text{-}1R^{+}$ cells, mostly generated through a cascade of continuously differentiating progenitors in the bone marrow. The last shared intermediate is the monocyte-DC progenitor, MDP, which is characterized as a $CD45RA^{+}CD123^{int}CD115^{+}$ fraction of a heterogeneous granulocyte-monocyte progenitor (GMP) population (16). Using CLEC12A and CD64 expression, a focused monocyte progenitor without DC potential, the common monocyte progenitor (cMoP), was described recently (17). This restricted precursor differentiates via pre-monocytes to monocytes, which in mice egress the bone marrow in a CCR2-dependent fashion (18).

Monopoiesis is highly dependent on the hematopoietic growth factor receptor CSF-1R and is enhanced, especially during infection or "sterile" inflammation (19–22). This phenomenon highlights the function of blood monocytes, which mainly serve as a reservoir for tissue-residing monocyte-derived macrophages and monocyte-derived DCs, especially during inflammation. Under homeostatic conditions, the majority of monocytes are weak phagocytic cells and are less efficient in antigen presentation when compared to DCs and macrophages (14, 23).

Initially described by Steinman and Cohn in the early 1970s DC have been extensively studied in recent decades (24, 25). Nevertheless, the high variability regarding ontogeny, phenotype, tissue localization, and function has hampered to find a comprehensive description of this cell type for a long time. On a functional level, DC are very efficient in phagocytosis and antigen presentation and are therefore crucial for the initiation of an adaptive immune response (23). DC are generated from MDPs giving rise to DC-committed precursor cells called common DC progenitors (CDP) which serve as precursor for plasmacytoid DCs and the two classical DC subtypes cDC1 and cDC2 (26, 27). Recently, a cDC-restricted progenitor cell, the pre-cDC, has been described in mouse and human (5, 28–30). Concerning pDCs, a new model has been recently suggested (1, 31). In fact, 70–90% of pDCs seem to be IRF8-dependent and derive from a different pre-pDC precursor. These cells actively produce type I interferons and do not present antigen very well. Further studies are required to corroborate these recent findings.

WHICH ASPECTS DEFINE CELLULAR IDENTITY?

The Traditional Approach: Morphology, Phenotype, and Function

Several characteristics have been used to describe and define cell types and subsets. Initially, morphological characterization by early microscopy and functional observations laid the ground for the idea of different categories of cells. Primarily, features like size, shape of the cell, and/or nucleus, density, and staining behavior for specific dyes were used to separate immune cells into several cell types and subsets (24, 32–37).

Collectively described as mononuclear phagocytic cells, macrophages and monocytes were defined by their unique morphology and ability to take up pathogens and debris (32, 33, 38, 39). Several experiments suggested that blood-derived monocytes will give rise to different types of tissue-resident macrophages, which was comprehended by van Furth and Cohn as the "mononuclear phagocyte system" (MPS) (40). Later, Ralph Steinmann described cells that display a characteristic morphology when cultured on glass surfaces (24). Due to their morphology, he termed them dendritic cells. These DCs were quickly found to be professional antigen presenting phagocytes and were incorporated into the definition of the MPS (25, 41, 42).

The MPS has been defined based on morphology and shared functionality of monocytes, DCs, and macrophages as a broader framework to describe the role of these cell types during homeostasis and immunity. However, the original definition of the MPS cannot adequately explain the heterogeneity of these cell types concerning their origin, tissue localization, disease association, regulation, and function. For example, contrary to the original ideas, blood monocytes are not the only reservoir for tissue-resident macrophages. An enormous body of research established that tissue-resident macrophages are mostly generated by early progenitors during embryogenesis and exhibit to a limited extend the partial ability for self-renewal (43–47). Only some tissues of barrier organs like the intestine rely on the replenishment of tissue-resident macrophages by differentiation of monocytes during adult life, especially during infection or inflammatory conditions (48). Nevertheless, when looking at monocyte-derived and tissue-resident macrophages, we must acknowledge that these cells have a high phenotypic and functional similarity. This redundancy is essential for the (functional) replacement of yolk-sac derived tissue-resident macrophages in some tissues but makes it difficult to find a unified classification.

The use of surface marker detection by monoclonal antibodies and flow cytometry has revolutionized the way of cell type definition throughout immunology. While a functional heterogeneity of monocytes was suggested by several earlier studies (34–37, 49), it was two-color flow cytometry that provided a tool to clearly define two major monocyte subsets by their expression of CD14 and CD16 (50, 51). About 80 to 90% percent of peripheral blood monocytes express CD14 but lack the expression of the Fcγ-receptor III (FcγRIII/CD16). This subset is characterized by a higher phagocytic activity compared to the minor subset expressing CD16 and intermediate levels

of CD14. Also, $CD16^+$ monocytes can be further separated based on their expression of CD14 into $CD14^{dim}$ $CD16^+$ population and a less frequent $CD14^+$ $CD16^+$ subset (52, 53). The $CD14^+CD16^-$ subset of monocytes is referred to as classical monocytes, monocytes expressing CD14 and CD16 as intermediate monocytes and non-classical monocytes are defined as the $CD14^{dim}CD16^+$ subset (7, 52, 53). Furthermore, during the last decade, several markers have been suggested for defining the monocyte cell heterogeneity, including Slan and CD2/FcεRI (54–56). However, these markers do not reach the specificity that would be required for an unambiguous definition of cell types or cell subsets (also see below and **Box 1**).

Like monocytes DC have been first described on the basis of their morphological and functional aspects. Here, pDCs are characterized as main type-I interferon (IFN-α/β) secreting cells with plasma cell-like morphology (57). Activation and secretion of type-I interferons are facilitated by recognition of virus-derived nucleic acids, especially by endosomal nucleic acid-sensing Toll-like receptors (TLRs) TLR7 and TLR9 (57). Initially, these cells were identified by several groups under different names, including natural interferon-producing cells, plasmacytoid monocytes, and plasmacytoid T-cells (58–61). Finally, a consensus name, the plasmacytoid DC was introduced and phenotypic markers were defined including human blood dendritic cell antigen (BDCA)-2, human IL-3Rα (CD123) and BDCA-4 (57, 62–65). However, as already mentioned before and described in more detail later, previously reported experiments suggest that this consensus is once again challenged (1, 31) strongly arguing for an iterative process of cell type definition continuously including new information.

Besides pDCs, there are two subsets of myeloid or classical DC (mDC/cDC) that can be distinguished in the Lin–MHC-II+CD11c+ fraction (66, 67) by using the non-overlapping markers CD1c (BDCA1) or CD141 (BDCA3) in flow cytometry (64, 65). These DC subsets have been termed cDC1 ($CD141^+$ DC) and cDC2 ($CD1c^+$ DCs), respectively, which have been reviewed extensively elsewhere (68–73). While these classical markers are widely used, further markers have also been suggested for subset classification of DCs (73, 74). For instance, $CD141^+$ cDC1 can be identified by using antibodies against XCR1 (75, 76), CLEC9A (77–79) and CADM1 (80). Interestingly, all DC populations vary regarding their expression of the pattern recognition receptor family toll-like receptors, which is highly correlated with the functional roles these cells play in T-cell

BOX 1 | Proposed framework for the definition of cell types, cell subsets, and functional states of cell types and subsets.

Cell type definition based on a single parameter space (e.g. only ontogeny) will be inferior to integrated approaches utilizing additional information (ontogeny, -omics data, phenotypic, and functional data). Nevertheless, even with such a large heterogeneous parameter space at hand, cell type definition is still not trivial. We propose a framework to define cell types and their subsets that is based on knowledge from decades of developmental and cell biology, further substantiated with recent developments and results in the field of single-cell omics (165–168). Certainly, such proposal will require larger community involvement and is mentioned as a starting point for discussion. This principle can be extended to define other cell types as well.

According to this framework, "**cell types**" would be defined as follows:

"Cell types" constitute the highest category. Cell types are defined by the lack of transdifferentiation capacity in more than 95% of all physiological and non-physiological conditions. Furthermore, cell types exhibit certain phenotypic, functional and genome-wide (transcriptome, epigenome, other) characteristics that are unique to all cells of a particular type. For immune cells that are terminally differentiated, cell types would include T and B lymphocytes, NK cells, monocytes, macrophages and DC, neutrophils, basophils and eosinophils, mast cells and innate lymphoid cells. For the stem cell and precursor compartment, the hematopoietic stem cell would be one cell type, while all precursors could be another cell type. Particularly in the precursor space, more research is required to define whether—based on this definition—further cell types or only cell subsets (see below) exist. This is similarly true for cell type development during embryogenesis. However, such a framework would certainly guide future research, specifically exploiting experimental systems that would allow answering the question, whether a cell is still capable of transdifferentiating toward another cell type.

"**Cell subsets**" would be defined as follows:

"Cell subsets" are a secondary category within any given cell type. Cell subsets share certain phenotypic, genome-wide (transcriptome, epigenome) and functional features within a given cell type, but are distinct in other phenotypic, functional, or genome-wide features that are unique to them within a cell type. In an ideal setting, these features should not overlap with those features that characterize the cell type. Furthermore, the feature set characterizing a cell subset should not change if cells are analyzed from different compartments (tissues, organs) and under differing conditions (homeostasis, acute inflammation, repair conditions, etc.). Cell subsets can be further distinguished from cell types in that cell subsets can change into another subset of the same cell type to the degree that is higher than 5%. For example, it is known that classical monocytes can further differentiate into non-classical monocytes via the intermediate monocyte subset.

"**Functional states**" are defined as follows:

"Functional states" are the overall current program of any given cell. Again, "functional states" would be defined by a specific pattern of phenotypic, functional and genome-wide characteristics, which ideally would exclude features characterizing cell types or subsets. "Functional states" rely on spatiotemporal information (e.g., location, the cell's individual age, the age of the organism), the activation state (homeostasis, acute, chronic inflammation, repair phase, etc.) and any combination thereof. Clearly, "functional states" can only be defined by integrated approaches and patterns or signatures of many parameters. Single parameter definitions for functional states are very unlikely. Any given cell can be described by combinations of "functional states." In other words, "functional states" can be linked to intracellular biological modules responsible for different cellular functions. A cell could express pro-inflammatory cytokines and have elevated migratory capacity. "Functional states" can even be shared among different cell types and cell subsets. However, together with the definition of the cell type and subset, a cell can be defined unambiguously according to the three levels of cell type classification.

"Cell types," "cell subsets," and "functional states" will be governed by transcriptional programs that are linked to defined and specific networks of transcription factors (TFs) not only single TFs. Therefore, the description of such networks might be another means of defining cells accordingly.

The introduction of functional states will reduce the excessive introduction of new cell types or subsets and—in our view—also represents the well-known plasticity of the myeloid cell space better.

activation. For example, human CD141$^+$cDC1 cells express high amounts of TLR3 (81), a pattern recognition receptor highly associated with cross-presentation (82) and thus cDC1s are specialized in presenting intracellular antigens to CD8$^+$ T-cells in human and mice (83).

The most abundant subset of blood DCs are CD1c$^+$ cDC2s, which can be defined analytically by expression of CD11c, CD1c (BDCA1), and FCεRIa (54, 64, 84). Furthermore, CD1c$^+$ cDC2 express high levels of class II MHC molecules like HLA-DR, HLA-DQ, and show a high endocytic capacity, which specializes this DC type for the presentation of exogenous antigen to CD4$^+$ T cells (64, 84). As we will outline below, future work will require community efforts to integrate the differential usage of cell subset classification markers to generate consensus nomenclatures.

Collectively, the definition of cell types of the MPS and their subsets was initially based on cellular morphology, further developed by introducing immunophenotyping using antibodies against the respective cell surface markers and complemented by a functional assessment of the cell subsets identified. We spare the many controversial findings throughout this period, which only reflects the limitations of these approaches to generate a widely accepted nomenclature of cell types and subsets.

Ontogeny as a Concept for Cell Type Definition

A group of leading experts in the field of monocyte, DC, and macrophage biology has recently proposed a nomenclature, which is based mainly on the ontogeny and tissue localization of cells (73). The proposed two-level model defines a cell type, first by its origin (level 1), which is further improved by adding a functional, phenotypic or location information (level 2) of the particular cell type. This aspect of cell type classification and the ontogeny of DCs and monocytes have been reviewed extensively (48, 72, 85).

The usage of cellular origin for cell type classification is beneficial since such approach already segregates distinct, functional units. For example, it was suggested that all phagocytes that are generated by yolk-sac derived progenitors should be referred to as macrophages and cells derived from the hematopoietic lineage as monocyte-derived cells (8, 68). A further advantage of using origin and development of immune cells as a guiding principle for cell type definition is the conservation of ontogeny across species. However, although there is a substantial overlap of ontogenies in human and murine macrophage, monocyte and DC development, there is also considerable disagreement (16, 83, 86–88). Additionally, the ontogeny of myeloid cells is difficult to study in humans, and most results are obtained by mice experiments and then projected to human myeloid cells. Clearly, the ontogeny approach is a very important aspect of cell type definition, but it needs to be combined with other characteristics of cells.

HIGH-DIMENSIONAL APPROACHES SHAPE THE MYELOID CELL SPACE

Here, we introduce the latest technological advancements that have made substantial contributions to clarify the monocyte/DC compartment. Furthermore, we want to discuss open questions and challenges associated with these new technologies. Multi-dimensional approaches have significantly improved our understanding of the myeloid cell space by providing more features resulting in higher resolution for cell typing. To contextualize this, we want to provide examples that outline how high-dimensional methods have shaped our understanding of heterogeneity in human blood-derived monocytes and DC.

Although conventional flow cytometry has revolutionized cell type classification, it is limited in the number of parameters (markers <20) being analyzed at the same time. In the early 2000s, there were a couple of technological advancements that paved the way to the development of mass cytometry enabling parallel analysis of up to 40 parameters (89–93). This higher depth of data simultaneously enabled a multitude of possibilities for immunological and biomedical sciences, including the high-dimensional assessment of cross-patient cell type dynamics during acute myeloid leukemia (94–97). More recently, multi-color flow cytometry (MCFC) has been introduced, increasing the parameter space to a similar range, as seen in mass cytometry. However, although mass cytometry and MCFC allow high-throughput protein profiling of thousands of cells, the restriction to <40 protein markers may be underrepresenting the true number of variables that are necessary to define the heterogeneity in highly complex biological samples. Besides, these markers have to be selected *a priori*, which may put a bias on the results obtained by mass spectrometry or MCFC. Another revolution was introduced by the development of high-throughput gene expression profiling methods like microarray-based technologies and RNA-sequencing enabling to profile thousands of genes in a single sample (98, 99). This second genomic revolution enables the genome-wide assessment of gene expression, which not only allows to characterize cellular subsets but also to investigate regulatory networks (20, 100–102).

One of the first studies that performed microarray analysis of human DCs compared the transcriptomes of sorted cDC1, cDC2, and pDCs populations from peripheral blood and tonsils to deeply characterize these subsets (103). Robbins et al. performed a comparative study to put the transcriptome data of DC subsets into context of other myeloid and lymphocyte populations in blood (104), which resulted in the identification of important conserved signature genes, thereby strengthening cDC1, cDC2, and pDC as distinct DC subsets. Moreover, assessing transcriptomic data of both murine and human immune cells allowed to align DC subsets across species (104, 105). Another important study performed transcriptome profiling of human blood CD14 and CD16 monocyte populations, three DC subsets pDC, cDC1, and cDC2 as well as their skin counterparts cDC1, cDC2, and skin derived CD14$^+$ cells (80).

Notably, cell types like skin cDC1 and cDC2 grouped together with their counterparts isolated from blood, suggesting a high similarity of DC subsets independent from the microenvironment. We extended these findings to compare different DC subsets in many individuals and different tissues [lymphohematopoietic (blood, thymus, spleen) and non-lymphohematopoietic (skin, lung)] allowing to characterize the impact of the microenvironment on the identity of a cell

type (74). Integration of immune phenotyping, gene expression profiling, and bioinformatic analysis revealed that DC subsets from blood, spleen, and thymus were transcriptionally conserved, with only minor transcriptomic differences between the same DC subsets across tissues. In contrast, the transcriptomic consequence of the respective microenvironment was stronger in lung and skin subsets. This suggests a higher tissue imprinting of non-lymphohematopoietic DC subsets in barrier organs like lung and skin, when comparing to the tissue imprinting that has been reported for tissue-resident macrophage subsets (47, 100, 106, 107). However, the difference between different DC subsets (cDC1 vs. cDC2) is still larger than the differences between the same DC subset among different tissues (e.g., skin cDC1 vs. blood cDC1).

Collectively, gene expression profiling and comparative bioinformatic analysis have substantially contributed to understand the complex DC networks across species further improving current descriptions of unified and more unbiased classifications (73, 105, 108).

Early transcriptomic approaches of human and mouse monocyte subsets not only helped to deeply characterize these cell types but also presented a framework to validate high conservation of gene expression profiles between mouse and humans (104, 109). For example, a combination of well-designed functional assays and gene expression profiling helped to refine the role of non-classical monocytes as the counterpart to murine "patrolling" Gr1$^-$ monocytes (110). Other studies sharpened the definition of the intermediate and non-classical monocytes as distinct cell subset (110–112). Interestingly, these studies revealed a high similarity of non-classical and intermediate monocytes, underlining the transitional nature of these cells, as they show intermediate expression for most of the marker genes differentially expressed between classical and non-classical monocytes. Interestingly, a unique module of class-II MHC genes was highest expressed in the intermediate monocyte population (111).

Measuring RNA rather than protein levels represents one of the major limitations of gene expression profiling methods. While the overall correlation of transcriptome and proteome is relatively high (113, 114), RNA-seq and microarrays do not allow to assess post-translational modifications, which represent a central part of cellular regulation (115, 116). To overcome this limitation, mass cytometry has been utilized to profile post-translational modifications like phosphorylation, methylation, and glycosylation (117, 118). A good example of the value of methods with larger feature size compared to single or few marker studies is the definition of cells expressing the carbohydrate modification 6-Sulfo LacNAc (*Slan*) on the PSGL1 protein. Indeed, myeloid cells presenting Slan initially were termed "SlanDCs" (119–121), while others described an overlap of Slan$^+$ cells with non-classical monocytes (122, 123). However, all these studies largely rely on low-dimensional marker assessment by flow cytometry and are not always directly comparable due to differences in their choice of markers or gating strategies. To investigate this in a more unbiased fashion, Roussel et al. defined a 38-marker panel to study human myeloid cells from peripheral blood by mass cytometry (124). A semi-supervised analysis of the data resulted in the identification of distinct monocyte populations, two subsets overlap with markers from classical and intermediate monocytes while there are two subsets of monocytes that are similar to non-classical monocytes. The multi-dimensional analysis maps Slan$^+$ cells to the non-classical monocytes and does not show alignment with any DC population. In this study, Slan separates the non-classical monocytes into a Slan^{high} and a Slan^{low} CD14^{dim}CD16$^+$ population. However, earlier genomic comparisons of sorted Slan^{high} vs. Slan^{low} subsets did not reveal a significant difference between those two populations (110). More recently, by combining index sorting and high-content single-cell RNA-sequencing, we show further evidence that Slan expression does not reflect different cell subsets as the underlying overall transcriptional program is not different between Slan^{high} and Slan^{low} cells. Moreover, we clearly show that Slan+ cells are all non-classical monocytes (125).

Manual gating of monocytes by CD14 and CD16 is biased by the investigator, which is a disadvantage for large multi-center clinical studies. Unsupervised and semi-supervised computational analyses improve the accuracy and reproducibility of subset definitions (95, 117, 124, 126–128). However, interpretation of these results must be performed with special care, since the primary analysis is still dependent on manual parameter settings by the investigator. For example, in contrast to an earlier study utilizing mass cytometry (124) similar profiling of human mononuclear myeloid cells revealed three subsets of human monocytes in two other studies, while others report significant heterogeneity including three non-classical, one intermediate and four classical subsets (22). Interestingly, Hamers et al. identified a non-classical population, which is quite different to other non-classical populations and expresses CD9$^+$ CD41$^+$ and CD61$^+$, which may represent an eosinophil/basophil contamination (129–131). Another interesting observation is the rather low inter-individual difference of human monocyte populations during homeostasis when assessed by mass cytometry (22, 132).

High-throughput gene expression profiling by microarray or RNA-seq has paved the way to understand the regulatory networks within human monocytes and DC. These technologies are indispensable for high-depth characterization of immune cell types. Nevertheless, these population-based methods are not designed to detect further cellular heterogeneity within a sample. The gene expression measurement in a population-based RNA-seq represents an average signal of typically more than 10,000 individual cells, resulting in leveling out any further heterogeneity. Frequently, samples are generated by flow cytometry assisted cell sorting, which relies on the information of a limited set of marker genes. However, if these markers are not sufficient for detecting the full heterogeneity of the tissue, the results may be underestimating the true heterogeneity.

Transcriptional profiling of individual cells by single-cell RNA-seq has been introduced in 2009 (133, 134) and has revolutionized cell type discovery in all fields of biology (135–142), therefore it may be claimed as "third genomic revolution." Single-cell RNA sequencing approaches allow transcriptional profiling of 10,000s of individual cells. In contrast

to population-based RNA-seq, the groups of cells are not defined *a priori*, rather the cell classification is based on the similarity of gene expression profiles.

A series of studies applied single-cell RNA-seq to understand the heterogeneity of human blood DCs and DC progenitors (5, 30, 143). See et al., as well as Villani et al., detected and characterized the conventional subsets, including cDC1, cDC2, and pDC. Surprisingly, beyond these similarities the results differed significantly, strongly arguing that such high-dimensional data require particular care when assigning cell types and cell subsets. We defined cell types and subsets by a combination of function, phenotype and transcriptional profile, which lead to the identification of precursors (pre-cDCs) for the cDC1 and cDC2 subsets in addition to the three main DC subsets (5). To reconcile these two major initial reports, we developed a strategy that allows developing cell type classification consensus based on phenotypic and transcriptional features also including prior knowledge (125). This approach revealed that (1) the $AXL^{+}Siglec6^{+}$ DCs (AS-DCs) described by Villani et al. are mainly pre-cDCs as described in (5), (2) Mono4 are contaminating $CD56^{dim}$ NK cells, and (3) cells introduced as $CD16^{+}$ $CSF1\text{-}R^{+}$ $CTSS^{+}$ DCs are not belonging to the DC lineage. This general strategy is not restricted to myeloid cells but can be applied to any cell type classification problem in any species (125).

Recently, single-cell RNA-seq has also been used for improving our knowledge about the generation of DCs from bone marrow-derived progenitors. There is evidence that there is much higher flexibility in the development of DC and monocytes than already appreciated. Hematopoietic models that are not based on repeating rounds of division and differentiation (72, 144, 145) allow for incorporation of recent findings that suggest that cDCs can be generated by lymphoid progenitors (146). Also, the latest reports show important evidence that the large majority of pDCs arise from lymphoid progenitors rather than CDPs (1, 31). Probably, a community effort to clarify future naming and nomenclature of these cells is now warranted. Importantly, the recent high-dimensional characterization of pDCs (5, 125, 132, 143) and new insights into their ontogeny in mice (1) could form the basis for such new discussions.

Clearly, this is only the beginning of applying these technologies to open questions concerning the plasticity of the myeloid cell compartment. We also recognize that single-cell RNA-seq data are currently challenging our view on cell type classification and function within the myeloid cell compartment. However, in the long run, we are convinced that the higher information content per cell will give us a much better understanding of individual cells within any given tissue, organ, or inflammatory response.

PROPOSAL OF GENERAL PRINCIPLES FOR CELL TYPE DEFINITIONS

Considering the apparent ease, with which different cell types were characterized based on morphological differences a century ago (39), our capabilities to simultaneously measure hundreds to thousands of parameters per single cell seem to decrease our ability to agree on defined cell types and cell subsets (1, 5, 31, 143). The ability to detect heterogeneity between individual cells has extended to biological differences that are not related to questions concerning cell type or cell subset. The best-characterized biological process in single cell –omics data being cell cycle in proliferating cells (147–149). Certainly, cell cycle differences should not classify two cells of the same type as different cell types or subsets. Stochastic behaviors of single cells, e.g., in transcription (150, 151) would be another biological phenomenon that should not impact on cell classification aspects. Furthermore, data sparsity, still very apparent in all sequencing-based single cell technologies, requires attention, when dealing with cell type definitions.

Similarly, important is the question, whether all biased approaches requiring feature selection (e.g., which markers to be analyzed) prior to analysis are good starting points for cell type definitions. These would include all multi-color flow cytometry and single-cell mass spectrometry approaches. Potentially a more appropriate approach would be the combination of markers (chosen by the investigator) with unbiased approaches provided by single cell sequencing-based technologies. This is crucial since it allows to link the enormous body of research that has been performed with flow cytometry-defined cell populations (e.g. ontogeny) with results obtained by analysis of high-dimensional data. For example, index sorting based on previously defined cell surface markers combined with scRNA-seq might be a better way of defining the cell population structure as well as the practicality of certain protein markers to capture the population structure (125, 152, 153). Alternative but significantly more expensive approaches are based on the combination of full transcriptome scRNA-seq and oligonucleotide-labeled antibodies (154, 155). It can be expected that these approaches require iterations of experiments until markers are identified that truly reflect the underlying population structure. In this context, it is important to note that even such large endeavors such as the Human Cell Atlas will require the integration of additional layers of information in addition to scRNA-seq data. Furthermore, we postulate that these iterations will lead to consensus maps as a basis for cell type definitions (125). Very much like the cluster of differentiation (CD) workshops for antibodies (156), a community effort will be necessary to agree on the different versions of such consensus maps of individual cell types.

However, even if the combination of truly unbiased single cell –omics approaches and antibody-based techniques leads to novel consensus maps of immune cells including the myeloid cell space, we propose that each cell type and more importantly each cell subset requires to be functionally characterized, as we have previously demonstrated for human DCs in blood (5). In other words, we strongly argue that a final definition of a cell subset should be validated on functional differences and not only on transcriptional and phenotypic differences.

Once cell types are defined under homeostatic conditions, which is a major goal of the Human Cell Atlas (157), an even more daunting task will be to define cell types and subsets under pathophysiological processes. While certain cell types will be under developmental trajectories (cell states) under physiological conditions, the space for different cell states in disease settings will further increase (158). More importantly, under these

conditions, there will be mainly changes in parameters related to biological function rather than features defining cell types or subset. A major goal for further cell type definitions will be to integrate these functional states and trajectories. In this context, we propose cell types as the highest level to distinguish cells. For example, DC, monocytes, and macrophages would qualify as individual cell types, while pDC, cDC1, and cDC2 would qualify as DC subsets (5, 125). Each of these subsets can exist in different functional states that depend on location, differentiation stage, acute or chronic activation signals, to name only a few (69, 74). Again, even for functional states, we would propose to define cells based on hundreds of parameters measured by single cell –omics technologies to be combined with classical marker strategies but finally also integrate functional readouts for these cellular states.

Even if we can agree on such an approach, the question remains, how this can be realized technically? In fact, this is not a mere technical question, as it requires to consider methods that are more independent of investigator bias. For example, we strongly suggest building approaches that will allow us to build cell type definitions based on machine learning rather than on investigator-driven and individualized analysis pipelines. Single-cell transcriptomics algorithms as they are implemented in singleR (159) or scMatch (160) are good starting points. Nevertheless, they still heavily rely on an investigator's interpretation of such high-feature data spaces. Cell type definition could be a classification problem requiring the respective machine learning as they are used for classifier generation in other areas (161, 162). We do not favor solely data-driven machine learning but would suggest the integration of prior knowledge. First attempts to develop such methods are currently underway, and we will soon know, whether the introduction of machine learning based cell class prediction will truly aid our attempts to make sense of the hundreds to thousands of parameters that we now can routinely measure from single cells.

SUMMARY AND OUTLOOK

Since the discovery of myeloid cells more than a century ago, we have learned a lot about these important immune cells. Their enormous plasticity is fascinating and challenging at the same time. Not surprisingly, cell type definitions and nomenclature—up to the day—have been changed or updated regularly (48, 68, 108, 163, 164). A unified nomenclature is the basis for an effective communication among scientists and will accelerate discovery of novel therapeutics. Moreover, high-dimensional profiling of samples will facilitate to compare results and cell types across experiments, tissues and species. Even with the highest number of parameters known per any given cell, we still differ in our interpretations of certain cell types within the myeloid compartment. While it will be rather critical to include prior knowledge when labeling cells based on high-dimensional single cell data, we need to develop better tools based on robust mathematical rules that help us to determine cellular phenotypes and functions less ambiguously. With the emergence of powerful machine learning and AI-based methodology, the time has probably come to utilize such approaches to our benefit when describing cell types, cell subsets, and their functional states. Irrespective of the power of such approaches, we also need to accept that we are far from a complete understanding of these cells. Additional layers of information, for example, epigenetic information, will have to be included in cell type definitions as they arise. Therefore, we foresee numerous iterations of defining cell types and their functions in the decades to come. In other words, consensus maps of cell types and subsets that we agree on today will form the basis for newer maps with updated information content in the future. A potential framework for such a community-based effort has been outlined here.

AUTHOR CONTRIBUTIONS

PG and JS conceived and wrote the manuscript.

FUNDING

This work was supported by the German Research Foundation to JS (GRK 2168, INST 217/577-1, EXC2151/1), by the HGF grant sparse2big to JS, the FASTGenomics grant of the German Federal Ministry for Economic Affairs and Energy to JS, and the EU project SYSCID under grant number 733100.

REFERENCES

1. Dress RJ, Dutertre C-A, Giladi A, Schlitzer A, Low I, Shadan NB, et al. Plasmacytoid dendritic cells develop from Ly6D+ lymphoid progenitors distinct from the myeloid lineage. *Nat Immunol*. (2019) 20:852–64. doi: 10.1038/s41590-019-0420-3
2. Van Galen P, Hovestadt V, Wadsworth MH, Aster JC, Lane AA, Bernstein Correspondence BE, et al. Single-Cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*. (2019) 176:1265–81.e24. doi: 10.1016/j.cell.2019.01.031
3. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The immune landscape of cancer. *Immunity*. (2018) 48:812–30.e14. doi: 10.1016/j.immuni.2018.03.023
4. Ibarra-Soria X, Jawaid W, Pijuan-Sala B, Ladopoulos V, Scialdone A, Jörg DJ, et al. Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol*. (2018) 20:127–34. doi: 10.1038/s41556-017-0013-z
5. See P, Dutertre CA, Chen J, Günther P, McGovern N, Irac SE, et al. Mapping the human DC lineage through the integration of high-dimensional techniques. *Science*. (2017) 356:eaag3009. doi: 10.1126/science.aag3009
6. Parkin J, Cohen B. An overview of the immune system. *Lancet*. (2001) 357:1777–89. doi: 10.1016/S0140-6736(00)04904-7
7. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DNDN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*. (2010) 116:e74–80. doi: 10.1182/blood-2010-02-258558
8. Guilliams M, Mildner A, Yona S. Developmental and functional heterogeneity of monocytes. *Immunity*. (2018) 49:595–613. doi: 10.1016/j.immuni.2018.10.005
9. Garcia-Bonilla L, Faraco G, Moore J, Murphy M, Racchumi G, Srinivasan J, et al. Spatio-temporal profile, phenotypic diversity, and fate of recruited monocytes into the post-ischemic brain. *J Neuroinflamm*. (2016) 13:285. doi: 10.1186/s12974-016-0750-0
10. Kepler TBB, Chan C. Spatiotemporal programming of a simple inflammatory process. *Immunol Rev*. (2007) 216:153–63. doi: 10.1111/j.1600-065X.2007.00500.x

11. Qi H, Kastenmüller W, Germain RNN. Spatiotemporal basis of innate and adaptive immunity in secondary lymphoid tissue. *Annu Rev Cell Dev Biol*. (2014) 30:141–67. doi: 10.1146/annurev-cellbio-100913-013254
12. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature*. (2018) 563:347–53. doi: 10.1038/s41586-018-0698-6
13. Schultze JL. Chromatin remodeling in monocyte and macrophage activation. *Adv Protein Chem Struct Biol*. (2017) 106:1–15. doi: 10.1016/bs.apcsb.2016.09.001
14. Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells. *Science*. (2010) 327:656–61. doi: 10.1126/science.1178331
15. Neu KEE, Tang Q, Wilson PCC, Khan AAA. Single-cell genomics: approaches and utility in immunology. *Trends Immunol*. (2017) 38:140–9. doi: 10.1016/j.it.2016.12.001
16. Lee J, Breton G, Oliveira TYK, Zhou YJ, Aljoufi A, Puhr S, et al. Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. *J Exp Med*. (2015) 212:385–99. doi: 10.1084/jem.20141442
17. Kawamura S, Onai N, Miya F, Sato T, Tsunoda T, Kurabayashi K, et al. Identification of a human clonogenic progenitor with strict monocyte differentiation potential: a counterpart of mouse cMoPs. *Immunity*. (2017) 46:835–48.e4. doi: 10.1016/j.immuni.2017.04.019
18. Jacquelin S, Licata F, Dorgham K, Hermand P, Poupel L, Guyon E, et al. CX3CR1 reduces Ly6Chigh-monocyte motility within and release from the bone marrow after chemotherapy in mice. *Blood*. (2013) 122:674–83. doi: 10.1182/blood-2013-01-480749
19. Patel AA, Zhang Y, Fullerton JN, Boelen L, Rongvaux A, Maini AA, et al. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J Exp Med*. (2017) 214:1913–23. doi: 10.1084/jem.20170355
20. Christ A, Günther P, Lauterbach MAR, Duewell P, Biswas D, Pelka K, et al. Western diet triggers NLRP3-dependent innate immune reprogramming. *Cell*. (2018) 172:162–75.e14. doi: 10.1016/j.cell.2017.12.013
21. Nagareddy PR, Kraakman M, Masters SL, Stirzaker RA, Gorman DJ, Grant RW, et al. Adipose tissue macrophages promote myelopoiesis and monocytosis in obesity. *Cell Metab*. (2014) 19:821–35. doi: 10.1016/j.cmet.2014.03.029
22. Hamers AAJ, Dinh HQ, Thomas GD, Marcovecchio P, Blatchley A, Nakao CS, et al. Human monocyte heterogeneity as revealed by high-dimensional mass cytometry. *Arterioscler Thromb Vasc Biol*. (2019) 39:25–36 doi: 10.1161/ATVBAHA.118.311022
23. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. *Nature*. (1998) 392:245–52. doi: 10.1038/32588
24. Steinman RM, Cohn ZA. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J Exp Med*. (1973) 137:1142–62. doi: 10.1084/jem.137.5.1142
25. Steinman RM, Witmer MD. Lymphoid dendritic cells are potent stimulators of the primary mixed leukocyte reaction in mice. *Proc Natl Acad Sci USA*. (1978) 75:5132–6. doi: 10.1073/pnas.75.10.5132
26. Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells. *Annu Rev Immunol*. (2009) 27:669–92. doi: 10.1146/annurev.immunol.021908.132557
27. Onai N, Kurabayashi K, Hosoi-Amaike M, Toyama-Sorimachi N, Matsushima K, Inaba K, et al. A clonogenic progenitor with prominent plasmacytoid dendritic cell developmental potential. *Immunity*. (2013) 38:943–57. doi: 10.1016/j.immuni.2013.04.006
28. Naik SH, Sathe P, Park H-Y, Metcalf D, Proietto AI, Dakic A, et al. Development of plasmacytoid and conventional dendritic cell subtypes from single precursor cells derived *in vitro* and *in vivo*. *Nat Immunol*. (2007) 8:1217–26. doi: 10.1038/ni1522
29. Schlitzer A, Sivakamasundari V, Chen J, Sumatoh HR Bin, Schreuder J, Lum J, et al. Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat Immunol*. (2015) 16:718–28. doi: 10.1038/ni.3200
30. Breton G, Zheng S, Valieris R, Tojal da Silva I, Satija R, Nussenzweig MC. Human dendritic cells (DCs) are derived from distinct circulating precursors that are precommitted to become CD1c+ or CD141+ DCs. *J Exp Med*. (2016) 213:2861–70. doi: 10.1084/jem.20161135
31. Rodrigues PFF, Alberti-Servera L, Eremin A, Grajales-Reyes GEE, Ivanek R, Tussiwand R. Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat Immunol*. (2018) 19:711–22. doi: 10.1038/s41590-018-0136-9
32. Murray EGD, Webb RA, Swann MBR. A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillusBacterium monocytogenes (n.sp.). *J Pathol Bacteriol*. (1926) 29:407–39. doi: 10.1002/path.1700290409
33. Metchnikoff E. Ueber den Kampf der Zellen gegen Erysipelkokken. Ein Beitrag zur Phagocytenlehre. *Arch für Pathol Anat und Physiol und für Klin Med*. (1887) 107:209–49. doi: 10.1007/BF01926053
34. Figdor C, Bont W, Touw I, de Roos J, Roosnek E, de Vries J. Isolation of functionally different human monocytes by counterflow centrifugation elutriation. *Blood*. (1982) 60:46–53.
35. Akiyama Y, Miller PJ, Thurman GB, Neubauer RH, Oliver C, Favilla T, et al. Characterization of a human blood monocyte subset with low peroxidase activity. *J Clin Invest*. (1983) 72:1093–105. doi: 10.1172/JCI111034
36. Akiyama Y, Stevenson GW, Schlick E, Matsushima K, Miller PJ, Stevenson HC. Differential ability of human blood monocyte subsets to release various cytokines. *J Leukoc Biol*. (1985) 37:519–30. doi: 10.1002/jlb.37.5.519
37. Elias J, Chien P, Gustilo K, Schreiber A. Differential interleukin-1 elaboration by density-defined human monocyte subpopulations. *Blood*. (1985) 66:298–301.
38. Ebert RH, Florey HW. THE extravascular development of the monocyte observed *in vivo*. *Br J Exp Pathol*. (1939) 20:342–56.
39. Metchnikoff E. *Leçons sur la Pathologie Comparée de L'inflammation : Faites à l'Institut Pasteur en Avril et mai 1891*. Paris: G Masson (1892).
40. van Furth R, Cohn ZA. The origin and kinetics of mononuclear phagocytes. *J Exp Med*. (1968) 128:415–35. doi: 10.1084/jem.128.3.415
41. Adams DO, Edelson PJ, Koren HS. *Methods for Studying Mononuclear Phagocytes*. New York, NY: Academic Press (1981). 1023 p.
42. Steinman RM. The Dendritic cell system and its role in immunogenicity. *Annu Rev Immunol*. (1991) 9:271–96. doi: 10.1146/annurev.iy.09.040191.001415
43. Takahashi K, Yamamura F, Naito M. Differentiation, maturation, and proliferation of macrophages in the mouse yolk sac: a light-microscopic, enzyme-cytochemical, immunohistochemical, and ultrastructural study. *J Leukoc Biol*. (1989) 45:87–96. doi: 10.1002/jlb.45.2.87
44. Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, Gokhan S, et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*. (2010) 330:841–5. doi: 10.1126/science.1194637
45. Schulz C, Perdiguero EG, Chorro L, Szabo-Rogers H, Cagnard N, Kierdorf K, et al. A lineage of myeloid cells independent of myb and hematopoietic stem cells. *Science*. (2012) 336:86–90. doi: 10.1126/science.1219179
46. Hashimoto D, Chow A, Noizat C, Teo P, Beasley MB, Leboeuf M, et al. Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity*. (2013) 38:792–804. doi: 10.1016/j.immuni.2013.04.004
47. Mass E, Ballesteros I, Farlik M, Halbritter F, Günther P, Crozet L, et al. Specification of tissue-resident macrophages during organogenesis. *Science*. (2016) 353:aaf4238. doi: 10.1126/science.aaf4238
48. Ginhoux F, Schultze JL, Murray PJ, Ochando J, Biswas SK. New insights into the multidimensional concept of macrophage ontogeny, activation and function. *Nat Immunol*. (2016) 17:34–40. doi: 10.1038/ni.3324
49. Yasaka T, Mantich NM, Boxer LA, Baehner RL. Functions of human monocyte and lymphocyte subsets obtained by countercurrent centrifugal elutriation: differing functional capacities of human monocyte subsets. *J Immunol*. (1981) 127:1515–8.
50. Passlick B, Flieger D, Ziegler-Heitbrock HW. Identification and characterization of a novel monocyte subpopulation in human peripheral blood. *Blood*. (1989) 74:2527–34.
51. Ziegler-Heitbrock HWL, Ulevitch RJ. CD14: cell surface receptor and differentiation marker. *Immunol Today*. (1993) 14:121–5. doi: 10.1016/0167-5699(93)90212-4
52. Grage-Griebenow E, Zawatzky R, Kahlert H, Brade L, Flad H-D, Ernst M. Identification of a novel dendritic cell-like subset of CD64+/CD16+ blood monocytes. *Eur J Immunol*. (2001) 31:48–56. doi: 10.1002/1521-4141(200101)31:1<48::AID-IMMU48>3.0.CO;2-5

53. Grage-Griebenow E, Flad H -D., Ernst M. Heterogeneity of human peripheral blood monocyte subsets. *J Leukoc Biol*. (2001) 69:11–20. doi: 10.1189/jlb.69.1.11
54. Maurer D, Fiebiger E, Reininger B, Wolff-Winiski B, Jouvin MH, Kilgus O, et al. Expression of functional high affinity immunoglobulin E receptors (Fc epsilon RI) on monocytes of atopic individuals. *J Exp Med*. (1994) 179:745–50. doi: 10.1084/jem.179.2.745
55. Crawford K, Gabuzda D, Pantazopoulos V, Xu J, Clement C, Reinherz E, et al. Cells monocytes are dendritic + circulating CD2. *J Immunol References*. (1999) 163:5920–8.
56. Cheng YX, Foster B, Holland SM, Klion AD, Nutman TB, Casale TB, et al. CD2 identifies a monocyte subpopulation with immunoglobulin E-dependent, high-level expression of Fc?RI. *Clin Exp Allergy*. (2006) 36:1436–45. doi: 10.1111/j.1365-2222.2006.02578.x
57. Reizis B, Bunin A, Ghosh HS, Lewis KL, Sisirak V. Plasmacytoid dendritic cells: recent progress and open questions. *Annu Rev Immunol*. (2011) 29:163–83. doi: 10.1146/annurev-immunol-031210-101345
58. Rönnblomo L, Ramstedt U, Alm G V. Properties of human natural interferon-producing cells stimulated by tumor cell lines. *Eur J Immunol*. (1983) 13:471–6. doi: 10.1002/eji.1830130608
59. Chehimi J, Starr SE, Kawashima H, Miller DS, Trinchieri G, Perussia B, et al. Dendritic cells and IFN-alpha-producing cells are two functionally distinct non-B, non-monocytic HLA-DR+ cell subsets in human peripheral blood. *Immunology*. (1989) 68:486–90.
60. Desmet; FFW-PKRV van den O. The American journal of surgical pathology. *Am J Surg Pathol*. (1990) 14:101–12. doi: 10.1097/00000478-199002000-00001
61. Grouard G, Rissoan M-C, Filgueira L, Durand I, Banchereau J, Liu Y-J. The enigmatic plasmacytoid T cells develop into dendritic cells with Interleukin (IL)-3 and CD40-ligand. *J Exp Med*. (1997) 185:1101. doi: 10.1084/jem.185.6.1101
62. Siegal FPP, Kadowaki N, Shodell M, Fitzgerald-Bocarsly PAA, Shah K, Ho S, et al. The nature of the principal type 1 interferon-producing cells in human blood. *Science*. (1999) 284:1835–7. doi: 10.1126/science.284.5421.1835
63. Cella M, Jarrossay D, Facchetti F, Alebardi O, Nakajima H, Lanzavecchia A, et al. Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. *Nat Med*. (1999) 5:919–23. doi: 10.1038/11360
64. Dzionek A, Fuchs A, Schmidt P, Cremer S, Zysk M, Miltenyi S, et al. BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood. *J Immunol*. (2000) 165:6037–46. doi: 10.4049/jimmunol.165.11.6037
65. MacDonald KPAPA, Munster DJJ, Clark GJJ, Dzionek A, Schmitz J, Hart DNJNJ. Characterization of human blood dendritic cell subsets. *Blood*. (2002) 100:4512–20. doi: 10.1182/blood-2001-11-0097
66. O'Doherty U, Peng M, Gezelter S, Swiggard WJ, Betjes M, Bhardwaj N, et al. Human blood contains two subsets of dendritic cells, one immunologically mature and the other immature. *Immunology*. (1994) 82:487–93.
67. Thomas R, Davis LSS, Lipsky PEE. Isolation and characterization of human peripheral blood dendritic cells. *J Immunol*. (1993) 150: 821–34.
68. Schlitzer A, McGovern N, Ginhoux F. Dendritic cells and monocyte-derived cells: Two complementary and integrated functional systems. *Semin Cell Dev Biol*. (2015) 41:9–22. doi: 10.1016/j.semcdb.2015.03.011
69. Schlitzer A, Zhang W, Song M, Ma X. Recent advances in understanding dendritic cell development, classification, and phenotype. *F1000Research*. (2018) 7:1558. doi: 10.12688/f1000research.14793.1
70. Vremec D, Shortman K. What's in a name? Some early and current issues in dendritic cell nomenclature. *Front Immunol*. (2015) 6:267. doi: 10.3389/fimmu.2015.00267
71. Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. *Annu Rev Immunol*. (2013) 31:563–604. doi: 10.1146/annurev-immunol-020711-074950
72. Bassler K, Schulte-Schrepping J, Warnat-Herresthal S, Aschenbrenner AC, Schultze JL. The myeloid cell compartment—cell by cell. *Annu Rev Immunol*. (2019) 37:269–93. doi: 10.1146/annurev-immunol-042718-041728
73. Guilliams M, Ginhoux F, Jakubzick C, Naik SH, Onai N, Schraml BU, et al. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat Rev Immunol*. (2014) 14:571–8. doi: 10.1038/nri3712
74. Heidkamp GF, Sander J, Lehmann CHK, Heger L, Eissing N, Baranska A, et al. Human lymphoid organ dendritic cell identity is predominantly dictated by ontogeny, not tissue microenvironment. *Sci Immunol*. (2016) 1:eaai7677. doi: 10.1126/sciimmunol.aai7677
75. Bachem A, Güttler S, Hartung E, Ebstein F, Schaefer M, Tannert A, et al. Superior antigen cross-presentation and XCR1 expression define human CD11c+CD141+ cells as homologues of mouse CD8+ dendritic cells. *J Exp Med*. (2010) 207:1273–81. doi: 10.1084/jem.20100348
76. Crozat K, Guiton R, Contreras V, Feuillet V, Dutertre C-A, Ventre E, et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8alpha+ dendritic cells. *J Exp Med*. (2010) 207:1283–92. doi: 10.1084/jem.20100223
77. Caminschi I, Proietto AII, Ahmet F, Kitsoulis S, Teh JSS, Lo JCYCY, et al. The dendritic cell subtype-restricted C-type lectin Clec9A is a target for vaccine enhancement. *Blood*. (2008) 112:3264–73. doi: 10.1182/blood-2008-05-155176
78. Huysamen C, Willment JA, Dennehy KM, Brown GD. CLEC9A Is a novel activation C-type lectin-like receptor expressed on BDCA3(+) dendritic cells and a subset of monocytes. *J Biol Chem*. (2008) 283:16693–701. doi: 10.1074/jbc.M709923200
79. Sancho D, Joffre OPP, Keller AMM, Rogers NCC, Martínez D, Hernanz-Falcón P, et al. Identification of a dendritic cell receptor that couples sensing of necrosis to immunity. *Nature*. (2009) 458:899–903. doi: 10.1038/nature07750
80. Haniffa M, Shin A, Bigley V, McGovern N, Teo P, See P, et al. Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. *Immunity*. (2012) 37:60–73. doi: 10.1016/j.immuni.2012.04.012
81. Sittig SPP, Bakdash G, Weiden J, Sköld AEE, Tel J, Figdor CGG, et al. A comparative study of the T cell stimulatory and polarizing capacity of human primary blood dendritic cell subsets. *Mediators Inflamm*. (2016) 2016:1–11. doi: 10.1155/2016/3605643
82. Schulz O, Diebold SSS, Chen M, Näslund TII, Nolte MAA, Alexopoulou L, et al. Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature*. (2005) 433:887–92. doi: 10.1038/nature03326
83. Schlitzer A, Ginhoux F. Organization of the mouse and human DC network. *Curr Opin Immunol*. (2014) 26:90–9. doi: 10.1016/j.coi.2013.11.002
84. Ito T, Inaba M, Inaba K, Toki J, Sogo S, Iguchi T, et al. A CD1a+/CD11c+ subset of human blood dendritic cells is a direct precursor of langerhans cells. *J Immunol*. (1999) 163:1409–19.
85. Ginhoux F, Guilliams M. Tissue-resident macrophage ontogeny and homeostasis. *Immunity*. (2016) 44:439–49. doi: 10.1016/j.immuni.2016.02.024
86. Kawamura S, Ohteki T. Monopoiesis in humans and mice. *Int Immunol*. (2018) 30:503–9. doi: 10.1093/intimm/dxy063
87. Ito T, Kanzler H, Duramad O, Cao W, Liu Y-J. Specialization, kinetics, and repertoire of type 1 interferon responses by human plasmacytoid predendritic cells. *Blood*. (2006) 107:2423–31. doi: 10.1182/blood-2005-07-2709
88. Dalod M, Salazar-Mather TP, Malmgaard L, Lewis C, Asselin-Paturel C, Brière F, et al. Interferon α/β and interleukin 12 responses to viral infections. *J Exp Med*. (2002) 195:517–28. doi: 10.1084/jem.20011672
89. Quinn ZA, Baranov VI, Tanner SD, Wrana JL. Simultaneous determination of proteins using an element-tagged immunoassay coupled with ICP-MS detection. *J Anal At Spectrom*. (2002) 17:892–6. doi: 10.1039/b202306g
90. Baranov VI, Quinn ZA, Bandura DR, Tanner SD. The potential for elemental analysis in biotechnology. *J Anal At Spectrom*. (2002) 17:1148–52. doi: 10.1039/B201494G
91. Ornatsky O, Baranov VI, Bandura DR, Tanner SD, Dick J. Multiple cellular antigen detection by ICP-MS. *J Immunol Methods*. (2006) 308:68–76. doi: 10.1016/j.jim.2005.09.020
92. Lou X, Zhang G, Herrera I, Kinach R, Ornatsky O, Baranov V, et al. Polymer-based elemental tags for sensitive bioassays. *Angew Chemie Int Ed*. (2007) 46:6111–4. doi: 10.1002/anie.200700796

93. Ornatsky O, Bandura D, Baranov V, Nitz M, Winnik MA, Tanner S. Highly multiparametric analysis by mass cytometry. *J Immunol Methods*. (2010) 361:1–20. doi: 10.1016/j.jim.2010.07.002
94. Bendall SC, Simonds EF, Qiu P, Amir E -a. D, Krutzik PO, Finck R, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*. (2011) 332:687–96. doi: 10.1126/science.1198704
95. Bendall SCC, Davis KLL, Amir EDD, Tadmor MDD, Simonds EFF, Chen TJJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*. (2014) 157:714–25. doi: 10.1016/j.cell.2014.04.005
96. Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity*. (2012) 36:142–52. doi: 10.1016/j.immuni.2012.01.002
97. Newell EW, Sigal N, Nair N, Kidd BA, Greenberg HB, Davis MM. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat Biotechnol*. (2013) 31:623–9. doi: 10.1038/nbt.2593
98. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. (2008) 320:1344–9. doi: 10.1126/science.1158441
99. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. (2009) 10:57–63. doi: 10.1038/nrg2484
100. Xue J, Schmidt S V, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. (2014) 40:274–88. doi: 10.1016/j.immuni.2014.01.006
101. Schmidt S V, Krebs W, Ulas T, Xue J, Baßler K, Günther P, et al. The transcriptional regulator network of human inflammatory macrophages is defined by open chromatin. *Cell Res*. (2016) 26:151–70. doi: 10.1038/cr.2016.1
102. Sander J, Schmidt S V, Cirovic B, McGovern N, Papantonopoulou O, Hardt A-L, et al. Cellular differentiation of human monocytes is regulated by time-dependent interleukin-4 signaling and the transcriptional regulator NCOR2. *Immunity*. (2017) 47:1051–66.e12. doi: 10.1016/j.immuni.2017.11.024
103. Lindstedt M, Lundberg K, Borrebaeck CAK. Gene family clustering identifies functionally associated subsets of human *in vivo* blood and tonsillar dendritic cells. *J Immunol*. (2005) 175:4839–46. doi: 10.4049/jimmunol.175.8.4839
104. Robbins SH, Walzer T, Dembélé D, Thibault C, Defays A, Bessou G, et al. Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol*. (2008) 9:R17. doi: 10.1186/gb-2008-9-1-r17
105. Crozat K, Guiton R, Guilliams M, Henri S, Baranek T, Schwartz-Cornil I, et al. Comparative genomics as a tool to reveal functional equivalences between human and mouse dendritic cell subsets. *Immunol Rev*. (2010) 234:177–98. doi: 10.1111/j.0105-2896.2009.00868.x
106. Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell*. (2014) 159:1312–26. doi: 10.1016/j.cell.2014.11.018
107. Gosselin D, Link VMM, Romanoski CEE, Fonseca GJJ, Eichenfield DZZ, Spann NJJ, et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell*. (2014) 159:1327–40. doi: 10.1016/j.cell.2014.11.023
108. Vu Manh T-P, Bertho N, Hosmalin A, Schwartz-Cornil I, Dalod M. Investigating evolutionary conservation of dendritic cell subset identity and functions. *Front Immunol*. (2015) 6:260. doi: 10.3389/fimmu.2015.00260
109. Ingersoll MAA, Spanbroek R, Lottaz C, Gautier ELL, Frankenberger M, Hoffmann R, et al. Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood*. (2010) 115:e10-9. doi: 10.1182/blood-2009-07-235028
110. Cros J, Cagnard N, Woollard K, Patey N, Zhang S-Y, Senechal B, et al. Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity*. (2010) 33:375–86. doi: 10.1016/j.immuni.2010.08.012
111. Wong KLL, Tai JJ-YJ-Y, Wong W-C, Han H, Sem X, Yeap W-H, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood*. (2011) 118:e16-31. doi: 10.1182/blood-2010-12-326355
112. Frankenberger M, Hofer TPJPJ, Marei A, Dayyani F, Schewe S, Strasser C, et al. Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. *Eur J Immunol*. (2012) 42:957–74. doi: 10.1002/eji.201141907
113. Vä L, Scheele C, Pedersen K, Nielsen J. Proteome-and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes accession numbers. *Cell Rep*. (2015) 11:921–33. doi: 10.1016/j.celrep.2015.04.010
114. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett*. (2009) 583:3966–73. doi: 10.1016/j.febslet.2009.10.036
115. Narita T, Weinert BTT, Choudhary C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol*. (2019) 20:156–74. doi: 10.1038/s41580-018-0081-3
116. Wang Y-C, Peterson SEE, Loring JFF. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res*. (2014) 24:143–60. doi: 10.1038/cr.2013.151
117. Spitzer MH, Nolan GP. Mass cytometry: single cells, many features. *Cell*. (2016) 165:780–91. doi: 10.1016/j.cell.2016.04.019
118. Bodenmiller B, Zunder ER, Finck R, Chen TJ, Savig ES, Bruggner R V, et al. Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators. *Nat Biotechnol*. (2012) 30:858–67. doi: 10.1038/nbt.2317
119. Schäkel K, Kannagi R, Kniep B, Goto Y, Mitsuoka C, Zwirner J, et al. 6-Sulfo LacNAc, a novel carbohydrate modification of PSGL-1, defines an inflammatory type of human dendritic cells. *Immunity*. (2002) 17:289–301. doi: 10.1016/S1074-7613(02)00393-X
120. Schäkel K, von Kietzell M, Hänsel A, Ebling A, Schulze L, Haase M, et al. Human 6-sulfo LacNAc-expressing dendritic cells are principal producers of early interleukin-12 and are controlled by erythrocytes. *Immunity*. (2006) 24:767–77. doi: 10.1016/j.immuni.2006.03.020
121. Micheletti A, Finotti G, Calzetti F, Lonardi S, Zoratti E, Bugatti M, et al. slan/M-DC8$_+$ cells constitute a distinct subset of dendritic cells in human tonsils. *Oncotarget*. (2016) 7:161–75. doi: 10.18632/oncotarget.12418
122. Hofer TPP, Zawada AMM, Frankenberger M, Skokann K, Satzl AAA, Gesierich W, et al. slan-defined subsets of CD16-positive monocytes: impact of granulomatous inflammation and M-CSF receptor mutation. *Blood*. (2015) 126:2601–10. doi: 10.1182/blood-2015-06-651331
123. Siedlarl M, Frankenberger M, Ziegler-Heitbrock LHWHW, Belge K-U. The M-DC8-positive leukocytes are a subpopulation of the CD14+CD16+monocytes. *Immunobiology*. (2000) 202:11–7. doi: 10.1016/S0171-2985(00)80047-9
124. Roussel M, Ferrell PB, Greenplate AR, Lhomme F, Le Gallou S, Diggins KE, et al. Mass cytometry deep phenotyping of human mononuclear phagocytes and myeloid-derived suppressor cells from human blood and bone marrow. *J Leukoc Biol*. (2017) 102:437–47. doi: 10.1189/jlb.5MA1116-457R
125. Günther P, Cirovic B, Baßler K, Händler K, Becker M, Dutertre CAA, et al. A rule-based data-informed cellular consensus map of the human mononuclear phagocyte cell space. *bioRxiv [Preprint]*. (2019) 658179. doi: 10.1101/658179
126. Thomas GD, Hamers AAJ, Nakao C, Marcovecchio P, Taylor AM, McSkimming C, et al. Human blood monocyte subsets. *Arterioscler Thromb Vasc Biol*. (2017) 37:1548–58. doi: 10.1161/ATVBAHA.117.309145
127. Amir ED, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol*. (2013) 31:545–52. doi: 10.1038/nbt.2594
128. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. *Nat Methods*. (2016) 13:493–6. doi: 10.1038/nmeth.3863
129. Johansson MWW. Eosinophil activation status in separate compartments and association with asthma. *Front Med*. (2017) 4:75. doi: 10.3389/fmed.2017.00075
130. Han X, Jorgensen JLL, Brahmandam A, Schlette E, Huh YOO, Shi Y, et al. Immunophenotypic study of basophils by multiparameter flow cytometry. *Arch Pathol Lab Med*. (2008) 132:813–9. doi: 10.1043/1543-2165(2008)132[813:ISOBBM]2.0.CO;2

131. Mukai K, Gaudenzio N, Gupta S, Vivanco N, Bendall SCC, Maecker HTT, et al. Assessing basophil activation by using flow cytometry and mass cytometry in blood stored 24 hours before analysis. *J Allergy Clin Immunol*. (2017) 139:889–99.e11. doi: 10.1016/j.jaci.2016.04.060
132. Alcántara-Hernández M, Leylek R, Wagar LE, Engleman EG, Keler T, Marinkovich MP, et al. High-dimensional phenotypic mapping of human dendritic cells reveals interindividual variation and tissue specialization. *Immunity*. (2017) 47:1037–50.e6. doi: 10.1016/j.immuni.2017.11.001
133. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*. (2009) 6:377–82. doi: 10.1038/nmeth.1315
134. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc*. (2010) 5:516–35. doi: 10.1038/nprot.2009.236
135. Aevermann BD, Novotny M, Bakken T, Miller JA, Diehl AD, Osumi-Sutherland D, et al. Cell type discovery using single-cell transcriptomics: implications for ontological representation. *Hum Mol Genet*. (2018) 27:R40–7. doi: 10.1093/hmg/ddy100
136. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*. (2017) 541:331–8. doi: 10.1038/nature21350
137. Junker JP, Van Oudenaarden A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell*. (2014) 157:8–11. doi: 10.1016/j.cell.2014.02.010
138. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet*. (2013) 14:618–30. doi: 10.1038/nrg3542
139. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. (2018) 562:367–72. doi: 10.1038/s41586-018-0590-4
140. Eberwine J, Sul J-Y, Bartfai T, Kim J. The promise of single-cell sequencing. *Nat Methods*. (2013) 11:25–7. doi: 10.1038/nmeth.2769
141. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*. (2017) 18:35–45. doi: 10.1038/nri.2017.76
142. Cao J, Packer JSS, Ramani V, Cusanovich DAA, Huynh C, Daza R, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. (2017) 357:661–7. doi: 10.1126/science.aam8940
143. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. (2017) 356:eaah4573. doi: 10.1126/science.aah4573
144. Collin M, Bigley V. Human dendritic cell subsets: an update. *Immunology*. (2018) 154:3–20. doi: 10.1111/imm.12888
145. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. (2018) 553:418–26. doi: 10.1038/nature25022
146. Helft J, Anjos-Afonso F, van der Veen AG, Chakravarty P, Bonnet D, Reis e Sousa C. Dendritic cell lineage potential in human early hematopoietic progenitors. *Cell Rep*. (2017) 20:529–37. doi: 10.1016/j.celrep.2017.06.075
147. Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, Natarajan KN, et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell*. (2015) 17:471–85. doi: 10.1016/j.stem.2015.09.011
148. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. (2015) 33:155–60. doi: 10.1038/nbt.3102
149. Wang J, Huang M, Torre E, Dueck H, Shaffer S, Murray J, et al. Gene expression distribution deconvolution in single-cell RNA sequencing. *Proc Natl Acad Sci USA*. (2018) 115:E6437–46. doi: 10.1073/pnas.1721085115
150. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. (2015) 16:133–45. doi: 10.1038/nrg3833
151. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, et al. Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*. (2016) 535:299–302. doi: 10.1038/nature18320
152. Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat Cell Biol*. (2018) 20:836–46. doi: 10.1038/s41556-018-0121-4
153. Franziska Paul A, Arkin ara, Giladi A, Torben Porse B, Tanay A, Amit I. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell*. (2015);163:1663–77. doi: 10.1016/j.cell.2015.11.013
154. Shahi P, Kim SCC, Haliburton JRR, Gartner ZJJ, Abate ARR. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep*. (2017) 7:44447. doi: 10.1038/srep44447
155. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PKK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*. (2017) Sep;14:865–8. doi: 10.1038/nmeth.4380
156. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD nomenclature 2015: human leukocyte differentiation antigen workshops as a driving force in immunology. *J Immunol*. (2015) 195:4555–63. doi: 10.4049/jimmunol.1502033
157. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas meeting participants. *bioRxiv [Preprint]*. (2017). doi: 10.1101/121202
158. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol*. (2016) 34:1145–60. doi: 10.1038/nbt.3711
159. Aran D, Looney APP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. (2019) 20:163–72. doi: 10.1038/s41590-018-0276-y
160. Hou R, Denisenko E, Forrest ARR. scMatch: a single-cell gene expression profile annotation tool using reference datasets. Kelso J, editor. *Bioinformatics*. (2019) btz292. doi: 10.1093/bioinformatics/btz292
161. Warnat-Herresthal S, Perrakis K, Taschler B, Becker M, Seep L, Baßler K, et al. Diagnostic value of blood gene expression-based classifiers as exemplified for acute myeloid leukemia. *bioRxiv [Preprint]*. (2018) 382143. doi: 10.1101/382143
162. Angra S, Ahuja S. Machine learning and its applications: a review. In: *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. Chirala: IEEE (2017). p. 57–60. doi: 10.1109/ICBDACI.2017.8070809
163. Guilliams M, Henri S, Tamoutounour S, Ardouin L, Schwartz-Cornil I, Dalod M, et al. From skin dendritic cells to a simplified classification of human and mouse dendritic cell subsets. *Eur J Immunol*. (2010) 40:2089–94. doi: 10.1002/eji.201040498
164. Ginhoux F, Jung S. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat Rev Immunol*. (2014) 14:392–404. doi: 10.1038/nri3671
165. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet*. (2008) 9:868–82. doi: 10.1038/nrg2416
166. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res*. (2015) 25:1491–8. doi: 10.1101/gr.190595.115
167. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. *Nat Rev Genet*. (2016) 17:744–57. doi: 10.1038/nrg.2016.127
168. Marioni JC, Arendt D. How single-cell genomics is changing evolutionary and developmental biology. *Annu Rev Cell Dev Biol*. (2017) 33:537–53. doi: 10.1146/annurev-cellbio-100616-060818

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 18 October 2019
doi: 10.3389/fimmu.2019.02458

Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations

Ignacio Sanz[1,2*], *Chungwen Wei*[1,2], *Scott A. Jenks*[1,2], *Kevin S. Cashman*[1,2], *Christopher Tipton*[1,2], *Matthew C. Woodruff*[1,2], *Jennifer Hom*[1,2] *and F. Eun-Hyung Lee*[1,3]

[1] Lowance Center for Human Immunology, Emory University, Atlanta, GA, United States, [2] Division of Rheumatology, Department of Medicine, Emory University, Atlanta, GA, United States, [3] Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, Emory University, Atlanta, GA, United States

The increasingly recognized role of different types of B cells and plasma cells in protective and pathogenic immune responses combined with technological advances have generated a plethora of information regarding the heterogeneity of this human immune compartment. Unfortunately, the lack of a consistent classification of human B cells also creates significant imprecision on the adjudication of different phenotypes to well-defined populations. Additional confusion in the field stems from: the use of non-discriminatory, overlapping markers to define some populations, the extrapolation of mouse concepts to humans, and the assignation of functional significance to populations often defined by insufficient surface markers. In this review, we shall discuss the current understanding of human B cell heterogeneity and define major parental populations and associated subsets while discussing their functional significance. We shall also identify current challenges and opportunities. It stands to reason that a unified approach will not only permit comparison of separate studies but also improve our ability to define deviations from normative values and to create a clean framework for the identification, functional significance, and disease association with new populations.

Keywords: B cells, naïve, memory, transitional, ABC, DN2, atypical B cells, Breg

OPEN ACCESS

Edited by:
Pablo Engel,
University of Barcelona, Spain

Reviewed by:
Shiv Pillai,
Harvard Medical School, United States
Martin Perez-Andres,
University of Salamanca, Spain

***Correspondence:**
Ignacio Sanz
ignacio.sanz@emory.edu

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *10 July 2019*
Accepted: *01 October 2019*
Published: *18 October 2019*

Citation:
Sanz I, Wei C, Jenks SA, Cashman KS, Tipton C, Woodruff MC, Hom J and Lee FE-H (2019) Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. Front. Immunol. 10:2458. doi: 10.3389/fimmu.2019.02458

INTRODUCTION

B cells constitute a critical arm of the immune system and are responsible for the short-term and long-term generation of humoral antibody responses. B cells also carry out antibody-independent functions including: antigen-presentation, modulation of T cell differentiation and survival, and production of both regulatory and pro-inflammatory cytokines (1–4). Finally, B cells play a critical role in the formation of secondary and tertiary lymphoid tissue. A growing number of publications reflect their indispensable role in the generation of protective and pathogenic antibodies and their functional versatility, as well as their potential utility as disease biomarkers and therapeutic targets. Their behavior and value as biomarkers during changes in disease activity, whether spontaneous or in response to treatment, has also been an area of strong exploration. Unfortunately, there is growing difficulty in reconciling different and often contradictory studies. At the core of this problem are: the use of limited and inconsistent phenotypic markers, the use of pauci-color flow

cytometry, inappropriate and imprecise extrapolation of the significance of individual markers, and the forced inclusion of multiple B cell subsets within larger and heterogeneous populations. These problems are compounded by the assignment of functional properties (regulation, activation, anergy) and developmental connotations on the basis of surface phenotype rather than precise functional characterization and/or molecular markers (transcriptional factors/networks).

PERIPHERAL B CELL DIVERSITY: EMPHASIS ON HUMAN PHENOTYPES

Widely accepted schemes of human B cell development are modeled on knowledge derived from mouse studies, and indeed, fractions equivalent to murine pro-B, pre-B, immature, transitional B, and naïve B cells have been identified in human bone marrow (5–7). Current knowledge of peripheral human B cells stems mostly from tonsil analyses and a plethora of peripheral blood studies complemented by more limited analyses of spleen cells and non-malignant lymph nodes (8–29). The combination of those studies has identified the human counterpart to transitional and mature B cell compartments with the latter containing naïve B cells; germinal center (GC) B cells; memory B cells; and antibody-secreting cells (ASC) (25). Of note, with the possible exception of GC cells, all human B cell populations found in lymphoid tissues can also be demonstrated in the peripheral blood lymphocytes (PBL) including the ambiguously coined tissue-based memory cells further discussed below. These populations can be considered as parental populations comprised of different B cell subsets as it will be discussed later in the review (**Figure 1**).

From a developmental standpoint, mouse B cells are generally classified into B1 and B2 (follicular) B cells and marginal zone (MZ) B cells, with B1 and MZ B cells sharing important functional properties and in particular, the ability to generate a fast and intense burst of ASC against particulate bacterial antigens (30). In addition, mouse studies have also defined two major pathways of B2 cell activation and differentiation into ASC: (1) the GC pathway leading to the generation of long-lived memory cells and plasma cells; and (2) the extra-follicular pathway responsible for the generation of B cell blasts and short-lived plasmablasts (PB) (31). Of note, the extrafollicular pathway can also generate long-term memory and long-lived PC in a T cell independent fashion (32–34). While MZ B cells have been well-described in the human spleen and a circulating MZ equivalent is also recognized in the peripheral blood, the existence, significance and phenotypic identifiers of human B1 cells remain in dispute. Indeed, while an original report identified a population of human B cells sharing phenotypic (CD20+, CD27+, CD43+, CD70–), B cell repertoire, and functional features of the murine B1 population, subsequent analysis by the same group and others shed some doubt on the quantitation of these cells (35–39).

PHENOTYPIC MARKERS OF HUMAN PERIPHERAL B CELLS. BASIC EXPERIMENTAL AND ANALYTICAL APPROACH

All major parental populations of human peripheral B cells can be identified using a relatively small number of surface phenotypic markers including CD19, CD20, IgD, CD27, CD38, and CD24 (25). With the exception of plasma cells, the expression of CD19 and CD20 follows a largely overlapping pattern and their concomitant measurement may be redundant. Generally, these markers are used in conjunction with non-B cell lineage markers (exclusion channel), for positive identification of members of the B cell lineage and exclusion of non-B cells from the analysis. When only a limited number of markers are available, we prefer to use CD19 since the intensity of expression of this marker may provide valuable information regarding B cell activation, and expression may help differentiate between short-lived and long-lived PC, as will be discussed in further detail below (40). In turn, while the absence of CD20 is a valuable indicator of ASC in peripheral B cell analysis, enumeration of all ASC can also be achieved through the combination of CD27 and CD38 with the CD27hiCD38hi fraction containing both CD20– ASC as well as a small fraction of CD20+ ASC (see below). In all, we believe that proper analysis of the major canonical human B cell subsets can be achieved through, and requires, the analysis of 7-markers combined with proper exclusion of dead cells and cellular doublets. The recommended markers include: (1) non-B cells exclusion markers such as CD3 and CD14; (2) CD19; (3) IgD; (4) CD27; (5) CD38, (6) CD24; and (7) CD21. This combination of markers enables the analysis of human B cells through a combination of the two more widely used classification schemes: IgD vs. CD27; and IgD vs. CD38 (**Figures 2A,B**). The latter approach classifies B cells using the Bm1–Bm5 nomenclature originally derived from the study of mature B cells (Bm) in the human tonsil (8). Of note, this nomenclature was designed for the classification of B cells in human lymphoid tissue with the help of additional markers including CD10, CD44, CD77, and CD23 (8). When applied to the peripheral blood on the basis of IgD and CD38 expression, this approach is less categorical than the IgD/CD27 approach: as it fails to separate transitional cells (IgD+ CD38hi) from pre-GC cells (Bm2'), coalesces different types of memory B cells (10), does not separate resting naïve cells (Bm1) from IgD+ unswitched memory cells (10), and does not distinguish between conventional CD27+ memory cells and IgD/CD27 double negative cells, which in turn contains a heterogeneous population of cells including atypical/tissue-based/exhausted memory cells and activated extrafollicular PB precursors (25, 41–46). Accordingly, we recommend to base the initial categorization of parental B cell populations on the combination of IgD, CD27, and CD38 together with CD24 (**Figures 2A,C**). In addition, we strongly recommend the inclusion of CD21 in the core marker set as its expression enables the recognition of activated cells within all parental B cell populations and possibly, as further discussed below, of developmentally distinct B cells (**Table 1**). While largely

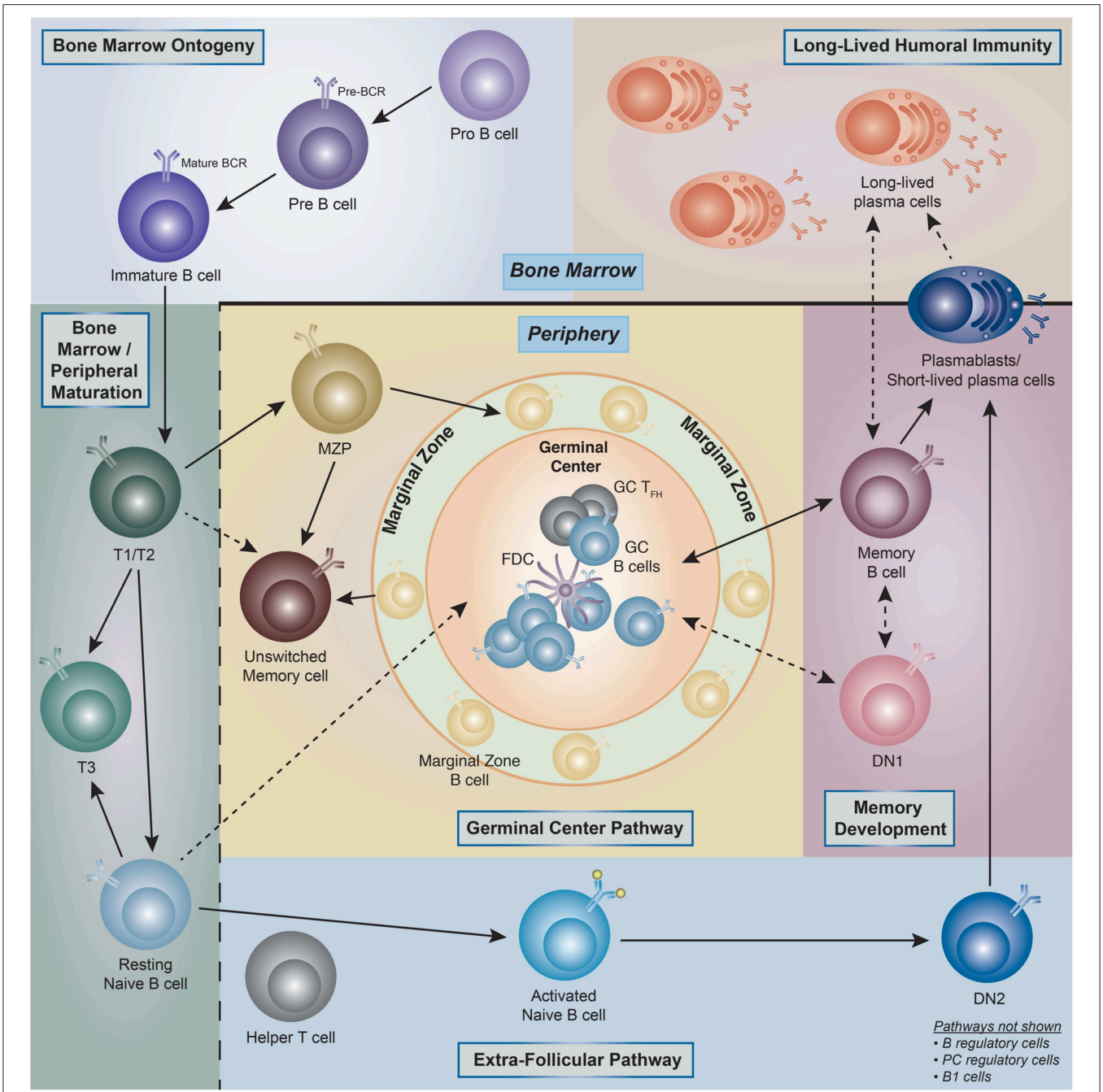

FIGURE 1 | Human B cell ontogeny. Illustration of the current understanding of B cell ontogeny in human B cells from late bone marrow (pro- and pre-B cells) through peripheral activation pathways, and into antibody secreting cells (ASC)/long-lived plasma cells compartments (PC). (→) Denotes clear literature supported associations, whereas (- - - >) represent associations which are theoretical but have literature supported plausibility. T1, transitional type 1 B cells; T2, transitional type 2 B cells; T3, transitional type 3 B cells; MZP, marginal zone precursor B cells; DN1, double-negative (IgD– CD27–)−1 B cells; DN2, double-negative (IgD– CD27–)−2 B cells; GC, germinal center; FDC, follicular dendritic cell.

overlapping with CD19, the inclusion of CD20 offers additional discriminating power to identify ASC.

As further discussed elsewhere in this review, the use of additional markers may significantly increase the investigators' ability to discriminate finer subsets of different functional properties and developmental origin (**Table 1**). Valuable makers whose use and discriminatory power has been well-documented in multiple studies include: IgM, IgG, IgA, CD10, CD23, CD80, CD86, CD69, CD95, CD11c, CD22, CXCR5, FcRL4, and FcRL5. In addition, the use of CD138 identifies more mature PC. A

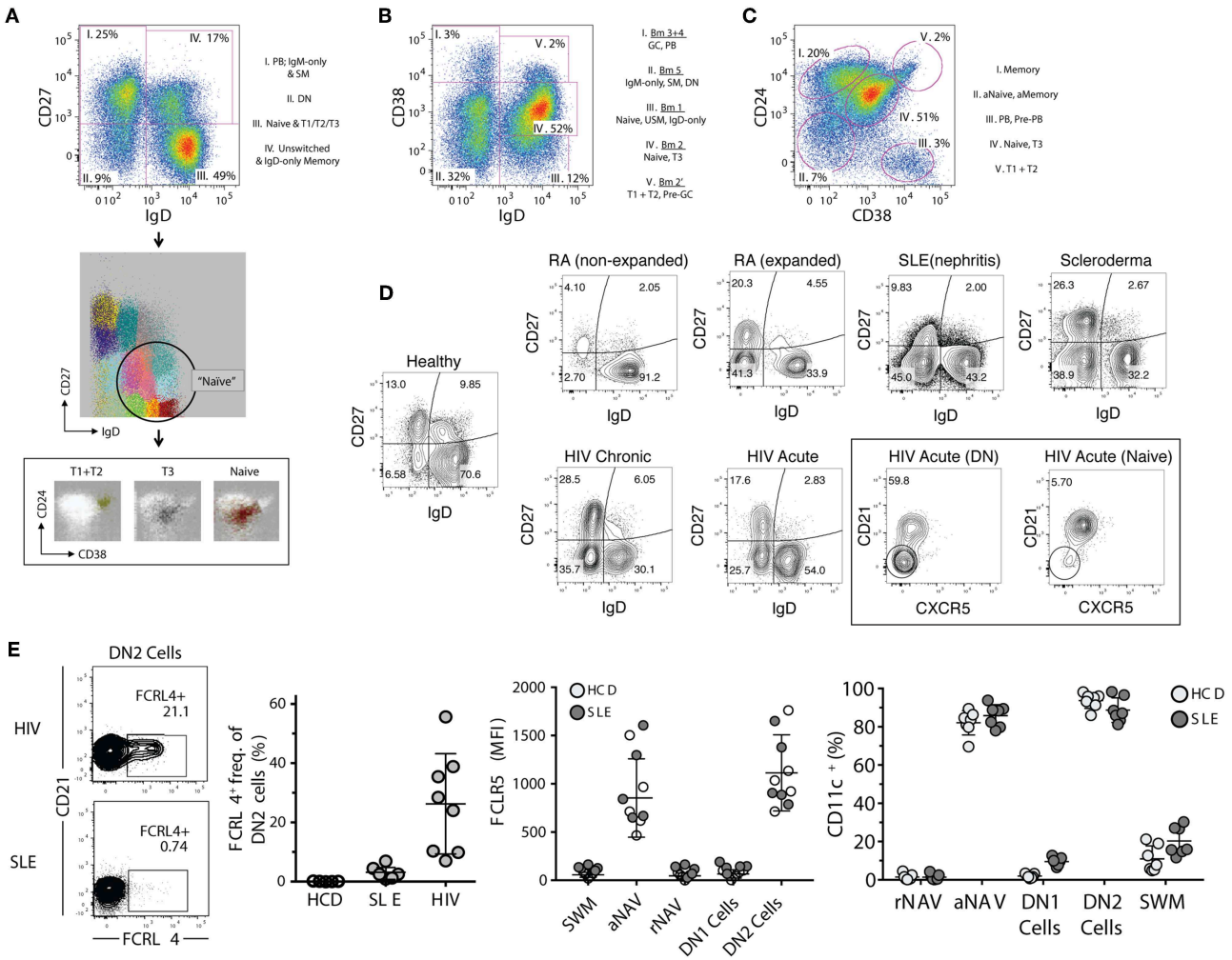

FIGURE 2 | Proposed B cell gating strategies. **(A)** CD27 vs. IgD gating strategy of CD19+ cells in a normal healthy donor representing four distinct B cell fractions: (I) switched memory (SM) B cells with associated plasmablasts (PB), (II) CD27– IgD– DN (double-negative) B cells, (III) global naïve B cell gate which includes transitional types 1–3 populations (associated FLOCK plots), (IV) Unswitched memory/IgD-only memory B cells. **(B)** Bm1–Bm5 gating strategy which identifies five B cell fractions which have more difficulty resolving memory from effector B cells, and naïve B cells from memory. **(C)** CD24 vs. CD38 gating strategy which defines five B cell fractions with good separation of memory from naïve (excluding the fraction II which represents combined activated memory and activate naïve) and naïve from transitional B cell populations. Better population resolution can be attained by subgating from the naïve population within the CD27 vs. IgD strategy followed by this CD24 vs. CD38 strategy. Of note is that transitional type-3 (T3) B cells cannot be resolved in any strategy without the use of MitoTracker® Green (MTG). **(D)** Representative FACS files showing (CD27– IgD–) DN B cells in a population of rheumatoid arthritis (RA), systemic lupus erythematous (SLE), scleroderma, and both acute and chronic human immunodeficiency virus (HIV) infected patients. Additionally, the DN population can be further defined by markers such as CXCR5. **(E)** Derived from Jenks et al. (41), Immunity, showing FcRL4 expression in atypical memory B cells from HIV patients but not SLE DN B cells, and the reciprocal expression pattern for FcRL5 (high in SLE but not HIV). This heterogeneity in expression patterns is indicative of multiple DN B cell populations, and can only be resolved by further marker subgating on DN B cells. Additionally, CD11c is highly expressed in SLE B cells and is indicative of DN2 B cells (activated effector cells) and are very different from the proposed atypical memory B cells of HIV which are thought to be anergic. PB, plasmablasts; DN, CD27– IgD– double negative B cells; T1, transitional type 1 B cells; T2, transitional type 2 B cells; T3, transitional type 3 B cells; GC, germinal center-like B cells; SM, switched memory B cells; USM, unswitched memory B cells; aNaive/aNAV, activated naïve B cells; Naïve/rNAV, resting naïve B cells; DN1, double negative 1 B cells; DN2, double negative 2 B cells, HIV, human immunodeficiency virus; SLE, systemic lupus erythematosus.

smaller number of studies have used CD25 as a marker of activation and in at least one study as a marker of B regulatory cells (47, 48). In addition, CD71 is a valuable marker of early memory B cell activation (49). Finally, CD5 and CD43 have been used to identify B1 cells. The value of CD5 for this purpose however, has been negated by multiple studies as this marker can be expressed by multiple B cell populations at least in part as a result of B cell activation (9, 50, 51).

In addition to surface markers, the classification of B cells can be powerfully aided by measuring the expression of proliferative markers (e.g., Ki-67); transcription factors (41, 52–60) and intracellular or secreted cytokines (1, 3).

TABLE 1 | Classification of blood human B cells using seven core markers.

B cell population (CD19+ unless otherwise noted)		Core markers	Additional markers	Function/properties
Transitional	T1/T2	IgD+ CD27– CD38++ CD24++	CD10++/+ IgM++ MTG+ CD10/IgM expression= T1 > T2 > T3	Developmental precursor
	T2-MZP	IgD+ CD27– CD38 ++ CD24++ CD21++	CD10+ IgM++	MZ precursor
	T3	IgD+ CD27– CD38+ CD24+ CD21+	CD10+/– IgM+ MTG+	Developmental precursor or activated naïve
Naïve	Resting	IgD+ CD27– CD38+ CD24+ CD21+	IgM+ MTG–	Antigen inexperienced mature cells
	Activated	IgD+ CD27– CD38– CD24– CD21–	IgM+ MTG+ CD95+ CD23– CD11c+ T-bet+ FcRL5+ SLAMF7+ CXCR5–	Precursor of short-lived PB and GC reactions
	Anergic	IgD+ CD27– CD38+/lo CD24+ CD21–	IgMlow/–	Hypo-responsive. Maintenance of tolerance
Memory	Unswitched	IgDlo CD27+ CD38+/lo CD24+ CD21+	CD1c+ IgM++	Natural memory. MZ equivalent
	Pre-switched	IgD– CD27+ CD38+/lo CD24+ CD21+	IgM+	Pre-switch memory; early IgM memory IgG memory precursor
	Switched resting	IgD– CD27+ CD38+/lo CD24+ CD21+	IgG/IgA+ CD95–	Pre-existing memory reservoir
	Switched activated	IgD– CD27+ CD38– CD24– CD21–	IgG/IgA+ CD95+ CD86+	Effector memory-PB/PC precursor
	Atypical tissue-based	IgD– CD27– CD38lo CD24lo CD21–	FcRL4+ IgM/IgG/IgA+ FcRL5+	Mucosal surveillance; exhausted memory; BCR hypo-responsive memory
Double negative (DN)	DN1	IgD– CD27– CD38+ CD24+ CD21+	FcRL4– IgM/IgG/IgA+ FcRL5– CXCR5+	Memory precursors
	DN2	IgD– CD27– CD38– CD24– CD21–	IgM/IgG/IgA+ T-bet+ CD11c+ FcRL5+ CXCR5– SLAMF7+	Extrafollicular ASC precursors
		IgD– CD27– CD38– CD24– CD21–	FcRL4+	Atypical/Tissue-based memory
Antibody secreting cells	Early PB	IgD– CD27lo/+ CD38++ CD24–	CD20+/– HLA-DR+ CD138–	Naïve and memory-derived PB precursors
	PB	IgD– CD27 ++ CD38+ + + CD24–	CD20– HLA–DR+ CD138– Ki67+	Antibody secretion
	PC	IgD– CD27 ++ CD38+ + + CD24–	CD19+/– CD20– CD138+	Antibody secretion
Regulatory B cells (Breg)		Multiple surface phenotypes corresponding to different core populations and subsets –Pro-B10 and B10 (no unique markers; defined by IL-10 production) –IL-35-producing Breg (defined by IL-35 production)		Down-regulation of T cell and monocyte inflammatory and autoimmune responses
Regulatory plasma cells (PCreg)		IgD– CD27++ CD38++	CD19+/– CD138+ IL−10+ or IL−35+ IgMhi or IgA+ (separate populations)	Suppress immune responses including anti-tumor responses
B1 cells		IgD++ CD27+	IgM+ CD43+ CD70– CD11c+ CD14+ CD5+/–	Production of natural autoantibodies

Core Markers*: CD19, IgM, IgD, CD27, CD38, CD24, CD21.* ***Additional Markers****: IgM, IgG, IgA, CD20, CD11c, FcRL5, FcRL4, CD138, CD95, CD80/86, CD23, T-bet, Ki-67, others.*

However, a comprehensive discussion of these valuable markers is outside the scope of this review. Accordingly, we will only address the significance of the expression of the transcription factor T-bet, given the prominence it has gained in the B cell literature over the last few years (61).

CLASSIFICATION OF CANONICAL B CELL POPULATIONS IN THE HUMAN PERIPHERAL BLOOD

As previously mentioned, a limited number of surface markers suffice for a consistent and quantitative measurement of all major B cell populations (**Table 1**). The canonical B cell populations thus recognized in the human peripheral blood are discussed in detail below and include: transitional cells, naïve B cells, memory B cells, circulating marginal zone (MZ) B cells, atypical memory B cells, and ASC. A more precise characterization of these canonical populations and of the distinct subsets they contain requires the use of additional markers (**Table 1**).

Transitional B Cells (T; Canonical Phenotype: CD19+, IgDlo/+, CD27–, CD24++, CD38++)

In the mouse, newly formed bone marrow immature B cells differentiate through sequential transitional stages of maturation into functionally competent follicular naïve B cells, a process originally thought to be limited to the spleen, but that it is now

known to also take place in the bone marrow (62, 63). The transitional B cell differentiation process is replicated in humans with T1, T2, and T3 circulating transitional fractions recognized in the peripheral blood (9, 22, 63). These consecutive fractions differ in their relative level of expression of CD24, CD38, CD10, and CD9 as well as IgM, all of which are gradually downregulated from T1 → T3 (**Figures 2A,C**) (9, 22). Transitional cells share with memory cells and activated naïve cells, the ability to retain mitochondrial dyes such as rhodamine or MitoTracker® Green owing to the absence of the ATP-binding cassette (ABC) B1 transporter (64). Similar to mice, the human spleen also contains a MZ precursor (MZP) representing a branching point of T2 cells characterized by high levels of CD21 (13, 22), a feature shared by a population of bone marrow transitional cells (65). Similar cells can be detected in the PBL on the basis of their expression of a CD45RB glycoform recognized by the MEM55 antibody (21, 66). In contrast to other transitional cells, a fraction of MZP express the (ABC)B1 transporter and accordingly, do not retain mitochondrial dyes.

The actual significance of T3 cells remains unknown as, in the mouse, these cells have been proposed to represent either anergic B cells (67), or the immediate precursor of mature naïve B cells (68); a characterization disputed by others (69). In turn, human cells with a T3 phenotype may represent either late transitional naïve precursors in the context of post-rituximab B cell repopulation (22) or early activated naïve B cells in active SLE (70).

Naïve B Cells (N: Canonical Phenotype: CD19+, IgD+, CD27–, CD38+/–, CD24+/–)

Human naïve B cells are typically defined by the expression of high levels of IgD and positive IgM staining, although at lower levels than in transitional cells. Using the core flow panel, N cells can be separated from transitional cells by their down-regulation of CD38 and CD24 which in N cells are both expressed as a continuum ranging from low-to-negative expression; and lack of CD27 expression (**Figure 2A**). Extended panels can further separate N cells based on their lack of expression of CD10 and lack of retention of mitochondrial dyes. Contrary to transitional cells, characterized by higher expression of IgM in a narrower MFI range, N cells express a wider continuum of surface IgM levels that include fractions with low-to-negative levels. The latter smaller fraction represents <2% of blood B cells and is characterized by the absence of IgM (IgD+, IgM– naïve, or BND), representing anergic autoreactive cells, a phenotype that has been extended to a larger fraction of IgMlow IgD+ cells encompassing 30% of all N cells (71). The larger low expression fraction is consistent with the abundance of autoreactive anergic B cells initially demonstrated in transgenic mice and then in wild-type mice (67, 72). It should be noted however, that IgM downregulation can also be induced by BCR activation and that expansions of IgMlow N cells with activated phenotype are observed in autoimmune diseases such as SLE (71). These observations should therefore, temper the use of IgM levels in isolation of other markers to define anergy or activation in the context of immune stimulation.

The use of additional markers recommended for the extended panels, such as CD21 and CD23 as well as traditional markers of activation such as CD80, CD86, CD95, and CD25, clearly discriminates distinct naïve B cell subsets. Thus, the characteristic phenotype of blood (resting) naïve B cells includes constitutive expression of CD21 and CD23 established during maturation and the absence of CD80, CD86, CD95, or CD25. The latter set of markers is only expressed by a minor fraction of N cells (<5%), in unperturbed healthy controls and gets differentially modulated upon B cell activation either through the BCR, CD40L, or TLR engagement in the context of IL-4 and other cytokines (73–79). It is important to recognize that significant differences exist in the intensity and kinetics of upregulation of these markers in response to different stimuli (80–83). Accordingly, it is recommended that the same marker and protocol be used when comparative analysis of naïve B cell activation is undertaken.

Valuable information can also be obtained through the expression of CD23 and CD21, two markers commonly used in the study of human B cells and whose expression is modulated during both B cell development and activation (65). CD21 and CD23, which in mouse and human spleen are fundamental for the adjudication of follicular and marginal zone phenotypes (13, 24, 84), are expressed universally by unperturbed human blood naïve B cells and their downregulation is observed in a major fraction of naïve B cells in several immunological conditions, including SLE (41, 85). Despite these observations, stemming from murine models and extrapolations made from the analysis of tonsil B cells and diverse *in vitro* stimulation conditions, the expression or lack thereof, of CD23 in human N cells has been interpreted in opposite ways. Thus, it was initially reported that tonsil naïve B cells upregulate CD23 during their differentiation (CD23– Bm1 to CD23+ Bm2) to GC centroblasts (Bm3). Yet, the same and subsequent studies also demonstrated the absence of CD23 in GC cells and in an activated (CD71+) intermediate population postulated to represent the early stages of naïve differentiation into GC cells or GC founders (86, 87). Consistent with an activated phenotype of CD23– N cells, multiple studies have identified expansions of CD23– B cell populations in SLE (40, 41, 71, 88, 89). These studies include recent detailed functional, transcriptional, and epigenetic characterization of activated naïve B cells marked by over-expression of T-bet, CD11c, SLAMF7, FcRL5 and other activation markers including CD80/CD86 and CD69, as well as downregulation of CD21 and CD23 (**Figures 2C**, **3A,B**) (40, 41, 90). However, expansion of activated naïve B cells in SLE has also been postulated on the basis of CD23 upregulation (74). This work however also described an expansion of CD23-negative naïve cells that were attributed to possible contamination with memory cells. Unfortunately, the absence of IgD, CD27, and CD23 co-staining precluded a conclusive identification of the relevant populations and even larger proportions of CD27– CD23– cells expressed CD80 and CD86 in SLE relative to CD23+ cells in healthy controls. Of interest, the recently described DN2 population (IgD– CD27– CD23– CD11c+ Tbet+), which is highly expanded in active

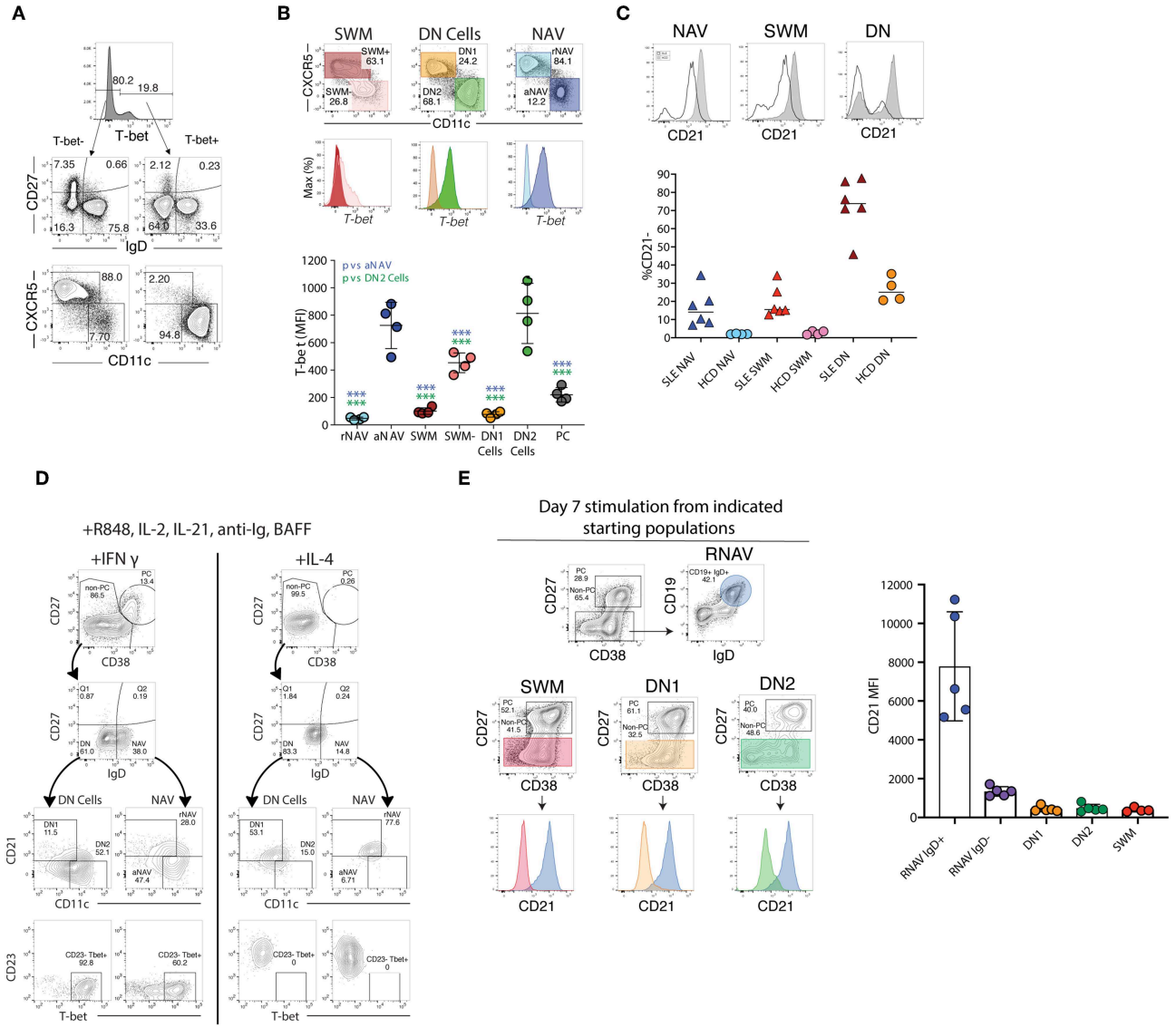

FIGURE 3 | T-bet, CD21, and CD11c expression *ex vivo* and after stimulation. **(A)** The majority of T-bet high B cells are IgD–CD27– DN or IgD+CD27– naïve B cells with a CD11c bright and CXCR5– phenotype characteristic of DN2 and activated naïve B cells, staining of CD19 B cells from a representative SLE patient. **(B)** Activated naïve and DN2 have the highest levels of intracellular T-bet staining. Gating and histograms are shown for a representative SLE patient is shown on top and quantification of T-bet mean fluorescence intensity for four SLE patients is shown below. Note, while CXCR5– SWM and CD27++ CD38++ PC express some T-bet their MFI is still significantly lower than that of DN2 and activated naïve. **(C)** Stimulation of HCD naïve B cells with TLR7 agonist R848, cytokines, and interferon gamma but not IL4 results in both plasma cell differentiation and increased T-bet and CD11c expression with concomitant loss of CD21 and CD23 expression. **(D)** Naive B cells from both HCD and SLE patients gain CD11c and lose CD21 in response to stimulation with interferon gamma, R848, and cytokines. **(E)** CD21 expression from flow sorted *in vitro* differentiated B cell populations (starting population indicated above center flow plots). There was a reduction of CD21 expression within all cultures (as compared to residual CD19+ IgD+ undifferentiated resting naïve B cells), independent of starting B cell population, suggesting that the loss of CD21 is indicative of a B cell activation state and recapitulates the *in vivo* phenotype of DN2 and activated naïve B cells. **(B,D)** were adapted from Jenks et al. (41).

SLE and which represents the progeny of activated extrafollicular naïve cells, could have accounted for the expansion of CD23– cells (42).

On the basis of what is known about the stimuli that upregulate CD23 expression on human B cells, it is certainly possible that activated naïve B cells could express different phenotypes in different conditions depending on a number of variables such as the duration of stimulation, type of T cell help, and cytokine milieu (75, 89, 91–93). Thus, IL-4 seems to be the main inducer of this marker after either BCR or CD40 stimulation and this induction is inhibited by IFNγ (77, 91, 92) (**Figure 3**). Notably, IL-4 and IFNγ exert a similarly reciprocal regulation on the differentiation of T-bet+ B cells induced in SLE (41, 94–103).

Similar to CD23, as further discussed below, CD21 downregulation identifies expanded fractions in several diseases

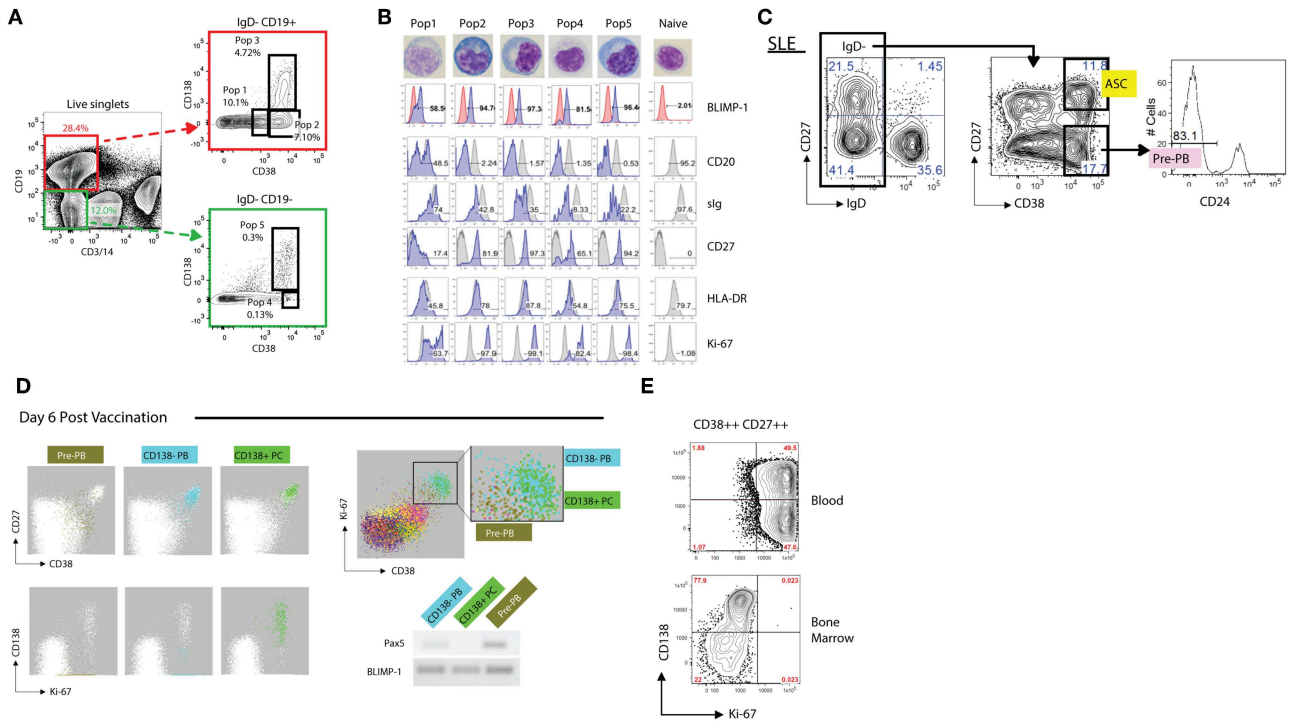

FIGURE 4 | Phenotypic analysis of antibody secreting cells (ASCs). **(A)** Representative FACS gating strategy (from an SLE patient) for discerning both CD19+ and CD19– ASCs populations from the blood. Utilizing CD3 and CD14 as "dump" channel markers CD19+ and CD19– (non-T cells, non-monocytes) B cell populations can be extracted from the data. Subgating on the IgD- fraction and utilizing CD138 vs. CD38 allows for the breakdown of ASC populations into five fractions: Pop1, CD19+ CD38mid CD138–; Pop2, CD19+ CD38hi CD138–; Pop3, CD19+ CD38hi CD138+; Pop4, CD19– CD38hi CD138–; Pop5, CD19– CD38hi CD138+. **(B)** Morphological and further phenotypic alterations in circulating ASC populations in **(A)** as compared to circulating naïve B cells, revealing plasmablast/plasma cell programming with high BLIMP-1, CD20lo/–, sIglo, and CD27++ expression, with the association of HLA-DR and Ki-67 positivity. **(C)** Gating strategy of an SLE patient revealing a sizeable presence of CD19+ IgD– CD27– CD38hi CD24– pre-PB B cells. **(D)** FLOCK analysis of a healthy donor 6 days post vaccination exhibiting CD38mid pre-PB, CD138– plasmablasts (PB), and CD138+ plasma cells (PC) on a CD27 vs. CD38 dot plot. In circulating ASCs Ki-67 was representative of both CD138– PB and CD138+ PC, whereas pre-PB are Ki-67+ without the CD138 expression. Pre-PB show an enrichment for BLIMP-1 expression while simultaneously expressing Pax5, signifying a transitional state toward an ASC fate. **(E)** CD138 vs. Ki-67 plot in subgated circulating and bone marrow ASCs representing the differences in proliferation status, with bone marrow ASCs exhibiting little Ki-67 expression. This dichotomy likely signifies newly generated vs. resident PC populations.

characterized by increased B cell autoreactivity including SLE, CVID, Sjogren's syndrome, and RA (12, 41, 85, 104–107). CD21 downregulation also marks activated memory cells whether in normal vaccination responses, HIV and malaria infection, or in memory cells expanded in response to checkpoint inhibitors. This feature is shared by other activated B cells including: CD11c+ activated naïve and DN2 cells, atypical and tissue-based memory cells, and T-bet+ B cells (15, 95, 108–111). Nevertheless, despite the preponderance of evidence for the activated phenotype of CD21low B cells within the naïve compartment and in particular in autoimmune diseases such as SLE, CD27– CD21low B cells have also been characterized as anergic B cells in a fraction of patients with RA and CVID (112). As previously discussed however for CD23, these studies focused on cells defined as CD27– CD21low without consideration of IgD expression and accordingly, they would have also included isotype-switched CD27– cells with hypo-responsiveness to anti-IgM stimulation (41, 93). One additional caveat that must be made in regards to CD21low B cells, is that although low in frequency, early transitional B cell populations (e.g., T1 B cells) may also be CD21low which can be isolated utilizing CD38, CD24, and/or CD10 (65).

Memory B Cells (M; Canonical Phenotype: CD19+, CD27+, CD38+/–, CD24+/–)

There is fairly universal consensus that human memory cells can be defined through the expression of CD19 and CD27 although CD19+ CD27+ blood cells also include antibody-secreting PB. In turn, M B cells and PB can be effectively differentiated with the core flow panel on the basis of CD38 expression. Thus, while M cells express intermediate to low levels of CD38 (Bm5 and Bm5' subsets, respectively, under the Bm classification), PB are CD38hi at levels that readily separate from memory cells. Also useful for this differentiation is the lack of expression of CD20 on most PB as further discussed below.

The assignation of a memory phenotype to circulating CD27+ B cells is based on the observation that, like germinal center cells which initiate CD27 upregulation (24, 113, 114), these cells accumulate significant rates of somatic hypermutation (14, 16,

17, 113, 115). Moreover, CD27+ B cells represent the main source of recall responses (49, 109, 116, 117). Finally, CD27+ B cells are endowed with enhanced differentiation into ASC and a distinct transcriptional program (118, 119).

Several memory subsets have been reported that can be identified with core markers (43, 120, 121). The first critical distinction separates IgD+ CD27+ and IgD– CD27+ cells. The IgD+ subset contains both IgM+ (CD27+ IgDlo IgM++; unswitched memory cells), and IgM-negative cells (IgD–only switched memory cells). Whereas, IgD– CD27+ cells comprise both IgM-only (pre-switch memory cells) and IgG, IgA, or IgE expressing cells (switched memory) (16, 17, 122–124). The unswitched nature of the IgD+ IgM+ populations and their separate origin relative to pre-switch memory, is consistent with the generation of the former subset in CD40L-deficient hyper-IgM patients unable to receive T cell help or form germinal centers and undergo class switch (although the potential for class-switching remains under normal conditions) (125). In turn, a pre-switched state capable of undergoing subsequent class switch has been suggested for IgM-only CD27+ memory cells (18, 126). It seems likely that IgM-only pre-switch cells represent the human counterpart of murine IgM memory cells generated in the early phases of the GC reaction that serve as a substrate for affinity matured isotype-switched cells during subsequent rounds of GC reactions (18, 117, 127). In contrast, despite lack of universal consensus (121), there is substantial evidence to indicate that IgD+ IgM+ unswitched memory cells may represent circulating marginal zone B cells (13, 84). Other studies however have proposed a GC origin for these cells on the basis of their mutational Bcl6 pattern (117). These observations taken together, lead us to propose that when IgM is available within an extended panel, memory B cell populations be further separated on the basis of both IgM and IgD expression as they represent distinct memory pools. Finally, IgD+ IgM–, CD27+ switched memory cells constitute a unique compartment representing a small but distinct fraction whose biology needs further clarification. Interestingly, IgD-only memory cells are characterized by an unusually large load of somatic hypermutation and may represent the origin of, heavily mutated IgD+ multiple myeloma (also uncommon) (128). While these cells appear to be enriched for autoreactivity (129, 130), their role in physiological and pathological conditions remains to be fully elucidated. Recent work however suggests that they may mediate important antimicrobial responses in the respiratory mucosa by arming basophils with IgD antibodies. In turn, faulty regulation of this mechanism could lead to the development of auto-inflammatory syndromes (131).

As for naïve B cells, resting and activated subpopulations can be identified through the use of markers included in the core and ancillary panels. Thus, constitutive expression of CD21 is maintained in resting memory cells and lost upon activation as illustrated by substantial increases of CD21lo CD27+ memory cells in HIV infection (108, 132, 133), influenza recall responses (106), and in patients with SLE and RA (110, 121, 134–136). Activated memory cells also upregulate CD95, CD80, and CD86 but as previously indicated, the correlation between these markers with CD21 downregulation is incomplete and accordingly, the same set of markers should be employed to establish changes in memory cell activation between samples (80, 110, 134, 137). Finally, upregulation of CD71 appears to be a helpful marker of early activation in proliferative antigen-specific memory cells and new germinal center products that differentiate into antibody-secreting PB (49, 106).

Atypical and Tissue-Based Memory B Cells (Canonical Phenotype: Double Negative, DN: CD19+, IgD–, CD27–, CD38+/–, CD24+/–)

SLE is characterized by large departures from normal immunological homeostasis that are commonly reflected in the peripheral blood. This behavior has facilitated the identification of distinct populations that under normal circumstances are under-represented in the blood, including a canonical subset characterized by the absence of both naïve (IgD) and conventional memory markers (CD27), termed double negative B cells (DN). Typically representing <10% of all PBL CD19+ cells, DN cells can contribute in excess of 40% of all B cells in active SLE and may become the largest circulating population of isotype switched IgD– cells (23, 41, 136) (**Figures 2A,D**). Distinguished by a CD19high activated phenotype, SLE DN cell expansion has been associated with: active disease, African-American descent, adverse clinical outcomes, and poor response to B cell depletion therapy (23, 41, 135). More recent work discussed below, has established their identify as activated effector ASC precursors enriched for SLE-associated autoreactivity (23, 41, 85).

While containing a relatively small fraction of IgM+ cells, DN cells are largely IgG+ or IgA+ and display a significant degree of somatic hypermutation albeit at a lower level than CD27+ memory cells. Although direct assignments are complicated by the omission of IgD staining in many studies, DN cells are also commonly expanded in patients with chronic HIV viremia and malaria infection (44, 108, 138, 139).

Overall, DN cells in different conditions are enriched in cells with an extended phenotype that includes: CD19hi; CD21lo; CD38+/–; CD24+/–; CD23–; FcRL5+; CD11c++; T-bet++. However, a notable difference between DN cells in SLE relative to HIV infection resides in the absence of FcRL4 in SLE (23, 41) (**Figure 2E**). The inhibitory FcRL4 receptor is a key phenotypic and functional feature of tissue-based memory cells, a population best defined in mucosal tissues that represented the first example of so-called atypical memory cells (26, 45). From a functional standpoint, the ability of FcRL4 to dampen B cell receptor signaling may account for the inhibited or exhausted function of DN cells reported in chronic HIV infection (93). Their ultimate functional properties however, are likely to be complex and context-dependent as both in these infections and in SLE, DN cells are characterized by high expression of multiple inhibitory receptors including CD32b, CD72, CD22, and PD-1 (138) and at least in HIV, of the inhibitory CD85j (132). Yet, despite the expression of FcRL4, CD85jhi DN cells appear to be activated and comprise the majority of the anti-gp140-specific responses in the early phases (<3 months) of acute HIV infection. However, their

relationship with the also CD21lo DN cells previously considered an unproductive repository in chronic HIV infection remains to be clarified (44, 108). Further suggesting that atypical memory cells can be functionally productive in certain settings, FcRL5+ cells with atypical memory markers can be generated through immunization and generate strong recall responses in malaria infection (140). Recent work indicates that the activation of DN cells in SLE may result not only from absence of FcRL4 but also from a more generalized defect in the function of inhibitory receptors and overall hyperresponsiveness to TLR stimulation (41).

Additional heterogeneity within the DN population has been recently established in SLE, where these cells comprise two major subsets on the basis of expression of CXCR5, CD21, and CD11c (41, 95) (**Figure 3A**). Thus, DN1 cells, representing the large majority of DN cells in healthy subjects and quiescent SLE, express a CXCR5+ CD21+ CD11c– phenotype. In contrast, DN2 cells representing the majority of expanded DN cells in active SLE, display a CXCR5– CD21– CD11c++ phenotype. Notably, DN2 cells also express FcRL5 and this marker can substitute for CD11c in their phenotypic characterization within the appropriate context of other markers. Immunologic, repertoire, and transcriptional characterization of DN1 and DN2 cells suggest that the former subset may represent early activated memory cells (41). Whereas, DN2 cells would represent a primed ASC precursor derived from newly activated naïve cells (which share essentially the same phenotype, with the exception of IgD expression), through an extra-follicular differentiation pathway; a fate consistent with that of CD11c+ extrafollicular plasmablasts in T-independent responses (41, 85, 141).

Currently, atypical memory and tissue-based memory nomenclature appears to be used in the literature interchangeably. Based on the evolving understanding of the different but overlapping populations comprised under these labels, we recommend the use of DN to denote a canonical population that is distinct from conventional naïve and memory cells on the basis of the defining core markers IgD and CD27. We would further advise the recognition of DN1 and DN2 cells based on the relative expression of CD21 or CXCR5 (which largely overlap) and CD11c or FcRL5. We also posit that the atypical memory definition may be unnecessary and possibly misleading, at least for patients with autoimmune diseases, as in such diseases these populations are largely comprised of non-memory cells but rather of activated DN2 effector cells generated through extrafollicular activation. Moreover, these populations are not atypical, but rather part of normal immune responses. We finally recommend that, if so desired, and as indicated by the specific condition under study, the terms atypical memory and/or tissue-based memory be reserved for DN cells with a FcRL4+ phenotype. Finally, we advise against classifying these populations on the basis of a CD27– CD21– phenotype in the absence of IgD staining as such cells would also include IgD+ activated naïve cells. Given the functional uncertainty and implications in different immunological conditions we would argue against attaching additional functional properties, including "exhaustion" or "inhibited," on the basis of these surface phenotypes alone.

The Conundrum of Human ABCs and Other Non-discreet Subsets Including T-bet+ and CD21low B Cells

Over the last few years, an interesting population of T-bet+ CD11c+ B cells has gained prominence in mice and by extension, in humans (61, 142–145). Initially described as Age-Associated B cells (ABC) on the basis of their prominence in aging mice, ABC expansions were then identified in younger animals in different autoimmune mouse models and demonstrated to be critical for viral clearance and autoimmune disease. Mouse ABC are TLR-7-driven and differentiate in response to IFNγ and IL-21 stimulation, a differentiation fate that is counteracted by IL-4 (140, 146, 147). More recently, similar populations (including the DN2 cells previously discussed), have been reported in humans and often lumped together as ABC despite growing evidence for the presence of a high degree of heterogeneity within these populations (95, 111, 132, 142). Thus, ABC and "ABC-like" phenotypes have been ascribed, often owing to the expression of one or more ABC markers (preferentially CD21, CD11c, or T-bet), to CD27+ memory cells with a number of studies proposing that human ABCs are predominantly of a CD27+ memory phenotype (108, 109, 111, 148). However, ABC-like populations have also been reported within atypical, tissue-based memory cells (132, 138). This confusion stems largely from the use of different markers to define ABC and in particular, from an over-emphasis on the significance of CD21 downregulation as an indicator of distinct B cell subsets beyond their activation status (**Table 1**). Indeed, the existing evidence supports the notion that both T-bet and CD11c upregulation as well as, and often times concurrently, CD21 downregulation, does occur in multiple B cell subsets under some activation conditions and does not by itself identify a specific B cell lineage or distinct population (**Figure 3**). In addition, the concentration of ABC within a given compartment may depend on the pre-determined study in some cases of only some B cell compartments and of the specific subjects and clinical situations studied. For example, it stands to reason that studies of memory recall responses would identify ABC-like cells predominantly within the responding memory cells that expand after immunization (109). Our own results provide some insight into the pattern of expression of T-bet and other associated ABC markers including CD11c, CD21, and FcRL5. In particular, studies in SLE, a condition that uniquely combines strong, simultaneous activation of both new naïve B cells and pre-existing memory cells, indicates that the majority of T-bet++, CD11c++, CXCR5–, CD21lo cells (all ABC-associated markers) reside within activated IgD+ naïve and IgD/CD27– DN2 cells with a smaller fraction of CD21lo cells with intermediate levels of expression of T-bet and CD11c, identified within the CD27+ memory population and in particular within the CD21lo activated memory subset (**Figure 3B**) (41, 42). Notably, *in vitro* stimulation of naïve, DN cells, and memory cells under conditions known to induce mouse ABC (TLR7, IFNγ, and IL-21), strongly induce a CD19++,

CD21lo, CD11c++, T-bet++, FcRL5+ phenotype that closely parallels the *in vivo* phenotype of activated naïve and DN2 cells (41, 95, 140) (**Figures 3C–E**).

In all, we postulate that the limited use of either CD21, T-bet or CD11c expression is inadequate to identify ABC or other distinct human B cell populations and that the present ABC assignment non-specifically integrates multiple B cell populations. This contention is further supported by recent evidence indicating that multilineage effector B cells can derive from T-bet+ memory precursors (149). A unifying view of the available evidence in humans supports the concept that ABC-like cells represent activated effector B cells induced by TLR7 and driven by IL-21 and IFNγ produced by T_{FH} cells in Th1-type responses within multiple and possibly, all B cell populations (41, 61, 95, 146, 147, 150). Whether the expression of the defining T-bet program or other transcriptional programs determines a distinct lineage or B cell differentiation fate beyond B cell activation and IgG1/IgG3 class switch remains to be determined (146, 151, 152). Given that the expansion of these cells appears to be age-independent in humans (41), we believe that there is no strong rationale nor biological basis for this designation in human B cells.

Antibody-Secreting Cells (ASC: Canonical Phenotype: CD19+/–, IgD–, CD27++, CD38++, CD24–)

Antibody-Secreting cells (ASC), encompass both proliferative cells (plasmablasts; PB), at different stages of differentiation from either naïve or memory cells as well as resting, mature plasma cells (PC). The ultimate classification of ASC as either PB or PC varies in the literature with some authors basing this adjudication on proliferative status (typically on the basis of Ki-67 expression), while others rely on the expression of CD138 to enumerate mature PC. BLIMP-1, IRF4, and XBP1 are the traditional transcription factors important for ASC differentiation (153–160). While it seems clear that CD138 expression identifies more mature PC and is globally expressed by bone marrow-long lived PC, it has also become evident that CD138 is expressed by a fraction of circulating, proliferative PC both in response to immunization as well as in active SLE (85, 161–164) (**Figures 4A–D**). Moreover, CD138 can be induced by *in vitro* stimulation of blood memory cells, thereby indicating that the expression of this marker is not restricted to fully differentiated, mature, resting PC (162, 164–166). As shown in **Figures 4B,E**, during active immune responses, including active SLE, proliferative CD138+ cells can account for a large fraction of all circulating ASC. By and large, the expression of Ki-67 overlaps with the expression of HLA-DR antigens and the latter therefore represents a useful surrogate for recently formed proliferative PB (161), obviating the need for intracellular staining. For the purpose of this review, we will focus on peripheral blood ASC subsets.

All peripheral human ASC (including those in lymphoid tissues and bone marrow), share high expression levels of CD38 and accordingly, this marker is critical for their identification (163, 167–170). Other human CD19+ B cells with high expression levels of CD38 include transitional cells which express IgD and CD24 and germinal center cells, which are usually absent in the blood. Although IgD+ ASC have been previously described, only a small fraction express BLIMP-1 and are rarely found in the periphery (131, 171, 172). Hence, all circulating ASC can be identified using the core markers under a CD19+, IgD–, CD38++ phenotype (**Figure 4A**). Human ASC also express high levels of CD27 and down-regulate CD20 expression. Of note, core markers also identify an additional population of proliferative IgD– CD38+/++ CD24– cells expressing low levels of CD27 (**Figure 4A**). Initially thought to be restricted to the human tonsil (169), circulating pre-PB have also been identified using automated multidimensional analysis of human vaccine responses where it behaves with a kinetics similar to conventional PB responses (163). These cells upregulate BLIMP-1 expression while maintaining expression of the B cell transcription factor Pax5 and are therefore, likely to represent PB precursors (pre-PB) (162). Of note, a significant fraction of pre-PB (up to 40%; **Figures 4C,D**), maintain expression of CD20 and accordingly, would be missed using a CD20– gate (162).

Finally, recent work has clearly identified populations of ASC lacking CD19 expression, a feature known to define terminally differentiated bone marrow plasma cells including multiple myeloma PC (173). However, in healthy subjects, CD19– PC are heterogenous and contain both CD138– and CD138+ cells with the latter fraction representing the source of human long-lived PC (40). CD19– PC can also be identified in the human blood in response to immunization and can be generated in culture (162, 164, 166).

B Regulatory Cells (Breg; No Specific Canonical Phenotype)

B regulatory cells (Breg), can be best defined as cells with the ability to inhibit pro-inflammatory monocytes and T cell responses although multiple other cellular targets have also been proposed, including plasmacytoid dendritic cells and anti-tumor cytotoxic T cells (48, 174). Following early descriptions for a central role of IL-10-producing Breg cells in the suppression of autoimmune diseases (175), multiple other cytokines including TGFβ and IL-35 as well as other mediators have been reported (48, 176). In addition Bregs can also suppress autoimmunity through the modulation of regulatory iNKT cells through CD1d-mediated lipid presentation (177). Intriguingly, CD1d Bregs were found in one study to be deficient in SLE and their expansion post-BCDT to correlate with favorable clinical response to rituximab (177). While CD1d+ Bregs concentrate within the CD24++ CD38++ transitional B cells proposed to represent a major IL-10-producing Breg human population (178), this study did not address the production of this cytokine.

From a phenotypic standpoint, Breg function has been identified within multiple human B cell populations including: CD38++C24++ transitional cells, naïve B cells, CD27+CD24high memory cells including B10 cells, CD27+ CD1d+ (marginal zone-like) memory cells, CD27+ CD5high PD-1high memory cells, and a TIM-1+ population with heterogeneous expression of CD27, CD24, CD38, CD1d and CD5 (48, 174, 178–186). Yet, even within those populations,

Bregs may account for <20% of B cells and thus, overall, they represent a small fraction of human B cells (48, 184). Currently, no set of surface phenotypic markers can identify Bregs and we agree therefore, with the expert recommendation that an accurate enumeration of Bregs, short of the ideal functional characterization, should rest on measurements of cytokine production at the single cell level by intracellular staining of regulatory cytokines such as IL-10 and IL-35 (48, 184). Substantial experimental evidence indicates that such measurements require *in vitro* short-term stimulation to reveal either cells containing pre-formed cytokines or longer stimulation to identify Breg precursors (B10 and pre-B10, respectively, for IL-10-producing Breg) (181, 184).

Finally, a number of studies have identified regulatory PB and PC whose function is mediated through either IL-10 or IL-35 and whose phenotype may include expression of PD-L1 and IgA (187–190). Regulatory PC appear to play an important role in autoimmune diseases and cancer immune responses where they may promote tumor progression (191).

CONCLUDING REMARKS

As our understanding of human B cell biology expands and new novel populations are discovered, it is critical to not lose sight of the need for phenotypic standardization. This all-important feature to experimental design will not only allow for the inter-laboratory interpretability but place the field of B cell immunology in a position to bound forward with limited hindrance on progress. Within this review we have offered a suggestion of standardizing B cell phenotyping with a core stain of seven surface markers, and also given strong support for the expandability of subgating beyond these core markers in a variety of B cell populations. Additionally, we raise the need for awareness that not all current surface antigens being utilized for B cell subgating strategies are likely sufficient to conclusively prove the existence of independent functional populations. Taken together we hope that this review may serve as a reference for future experimental designs and a springboard for phenotypic B cell normalization.

ETHICS STATEMENT

These studies were conducted from fully consented patients under ethical and safe protocols in accordance with the Declaration of Helsinki, and with the approval of the Institutional Review Boards of Emory University and the University of Rochester.

AUTHOR CONTRIBUTIONS

IS, CW, SJ, KC, and FL contributed to the literature review and manuscript preparation. IS, SJ, CW, CT, JH, KC, and MW were responsible for figure design and legend assembly.

FUNDING

This work was supported by National Institutes of Health: MW: T32-DK007656; IS: U19-AI110483 Autoimmunity Centers of Excellence, R37-AI049660; FL: P01-AI078907, R01-AI121252, U19-AI109962; IS and FL: P01-AI125180. United States-Israel Binational Science Foundation: IS: BSF-2013432.

ACKNOWLEDGMENTS

We would like to thank all members of the Sanz and Lee laboratories for their critical discussion and proofing of the manuscript.

REFERENCES

1. Lund FE, Garvy BA, Randall TD, Harris DP. Regulatory roles for cytokine-producing B cells in infection and autoimmune disease. *Curr Dir Autoimmun*. (2005) 8:25–54. doi: 10.1159/000082086
2. Lund FE. Cytokine-producing B lymphocytes – key regulators of immunity. *Curr Opin Immunol*. (2008) 20:332–8. doi: 10.1016/j.coi.2008.03.003
3. Shen P, Fillatreau S. Antibody-independent functions of B cells: a focus on cytokines. *Nat Rev Immunol*. (2015) 15:441–51. doi: 10.1038/nri3857
4. Lino AC, Dorner T, Bar-Or A, Fillatreau S. Cytokine-producing B cells: a translational view on their roles in human and mouse autoimmune diseases. *Immunol Rev*. (2016) 269:130–44. doi: 10.1111/imr.12374
5. LeBien TW. Fates of human B-cell precursors. *Blood*. (2000) 96:9–23.
6. Palanichamy A, Bauer JW, Yalavarthi S, Meednu N, Barnard J, Owen T, et al. Neutrophil-mediated IFN activation in the bone marrow alters B cell development in human and murine systemic lupus erythematosus. *J Immunol*. (2014) 192:906–18. doi: 10.4049/jimmunol.1302112
7. Duperray C, Boiron JM, Boucheix C, Cantaloube JF, Lavabre-Bertrand T, Attal M, et al. The CD24 antigen discriminates between pre-B and B cells in human bone marrow. *J Immunol*. (1990) 145:3678–83.
8. Pascual V, Liu YJ, Magalski A, de Bouteiller O, Banchereau J, Capra JD. Analysis of somatic mutation in five B cell subsets of human tonsil. *J Exp Med*. (1994) 180:329–39. doi: 10.1084/jem.180.1.329
9. Sims GP, Ettinger R, Shirota Y, Yarboro CH, Illei GG, Lipsky PE. Identification and characterization of circulating human transitional B cells. *Blood*. (2005) 105:4390–8. doi: 10.1182/blood-2004-11-4284
10. Bohnhorst JO, Bjorgan MB, Thoen JE, Natvig JB, Thompson KM. Bm1-Bm5 classification of peripheral blood B cells reveals circulating germinal center founder cells in healthy individuals and disturbance in the B cell subpopulations in patients with primary Sjogren's syndrome. *J Immunol*. (2001) 167:3610–8. doi: 10.4049/jimmunol.167.7.3610
11. Falini B, Tiacci E, Pucciarini A, Bigerna B, Kurth J, Hatzivassiliou G, et al. Expression of the IRTA1 receptor identifies intraepithelial and subepithelial marginal zone B cells of the mucosa-associated lymphoid tissue (MALT). *Blood*. (2003) 102:3684–92. doi: 10.1182/blood-2003-03-0750
12. Rakhmanov M, Keller B, Gutenberger S, Foerster C, Hoenig M, Driessen G, et al. Circulating CD21low B cells in common variable immunodeficiency resemble tissue homing, innate-like B cells. *Proc Natl Acad Sci USA*. (2009) 106:13451–6. doi: 10.1073/pnas.0901984106
13. Weill JC, Weller S, Reynaud CA. Human marginal zone B cells. *Annu Rev Immunol*. (2009) 27:267–85. doi: 10.1146/annurev.immunol.021908.132607

14. Seifert M, Kuppers R. Molecular footprints of a germinal center derivation of human IgM+(IgD+)CD27+ B cells and the dynamics of memory B cell generation. *J Exp Med*. (2009) 206:2659–69. doi: 10.1084/jem.20091087
15. Fecteau JF, Cote G, Neron S. A new memory CD27–IgG+ B cell population in peripheral blood expressing VH genes with low frequency of somatic mutation. *J Immunol*. (2006) 177:3728–36. doi: 10.4049/jimmunol.177.6.3728
16. Klein U, Rajewsky K, Kuppers R. Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med*. (1998) 188:1679–89. doi: 10.1084/jem.188.9.1679
17. Klein U, Kuppers R, Rajewsky K. Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood*. (1997) 89:1288–98.
18. Berkowska MA, Driessen GJA, Bikos V, Grosserichter-Wagener C, Stamatopoulos K, Cerutti A, et al. Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood*. (2011) 118:2150–8. doi: 10.1182/blood-2011-04-345579
19. Di Sabatino A, Rosado MM, Ciccocioppo R, Cazzola P, Morera R, Corazza GR, et al. Depletion of immunoglobulin M memory B cells is associated with splenic hypofunction in inflammatory bowel disease. *Am J Gastroenterol*. (2005) 100:1788–95. doi: 10.1111/j.1572-0241.2005.41939.x
20. Kuppers R, Klein U, Hansmann M-L, Rajewsky K. Cellular origin of human B-cell lymphomas. *N Engl J Med*. (1999) 341:1520–9. doi: 10.1056/NEJM199911113412007
21. Descatoire M, Weller S, Irtan S, Sarnacki S, Feuillard J, Storck S, et al. Identification of a human splenic marginal zone B cell precursor with NOTCH2-dependent differentiation properties. *J Exp Med*. (2014) 211:987–1000. doi: 10.1084/jem.20132203
22. Palanichamy A, Barnard J, Zheng B, Owen T, Quach T, Wei C, et al. Novel human transitional B cell populations revealed by B cell depletion therapy. *J Immunol*. (2009) 182:5982–93. doi: 10.4049/jimmunol.0801859
23. Wei C, Anolik J, Cappione A, Zheng B, Pugh-Bernard A, Brooks J, et al. A new population of cells lacking expression of CD27 represents a notable component of the B cell memory compartment in systemic lupus erythematosus. *J Immunol*. (2007) 178:6624–33. doi: 10.4049/jimmunol.178.10.6624
24. Cappione A III, Anolik JH, Pugh-Bernard A, Barnard J, Dutcher P, Silverman G, et al. Germinal center exclusion of autoreactive B cells is defective in human systemic lupus erythematosus. *J Clin Invest*. (2005) 115:3205–16. doi: 10.1172/JCI24179
25. Kaminski DA, Wei C, Qian Y, Rosenberg AF, Sanz I. Advances in human B cell phenotypic profiling. *Front Immunol*. (2012) 3:302. doi: 10.3389/fimmu.2012.00302
26. Ehrhardt GRA, Hsu JT, Gartland L, Leu C-M, Zhang S, Davis RS, et al. Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. *J Exp Med*. (2005) 202:783–91. doi: 10.1084/jem.20050879
27. Avery DT, Ellyard JI, Mackay F, Corcoran LM, Hodgkin PD, Tangye SG. Increased expression of CD27 on activated human memory B cells correlates with their commitment to the plasma cell lineage. *J Immunol*. (2005) 174:4034–42. doi: 10.4049/jimmunol.174.9.5885
28. Tangye SG, Liu Y-J, Aversa G, Phillips JH, de Vries JE. Identification of functional human splenic memory B cells by expression of CD148 and CD27. *J Exp Med*. (1998) 188:1691–703. doi: 10.1084/jem.188.9.1691
29. Jorge A, Wallace ZS, Zhang Y, Lu N, Costenbader KH, Choi HK. All-cause and cause-specific mortality trends of end-stage renal disease due to lupus nephritis from 1995 to 2014. *Arthritis Rheumatol*. (2019) 71:403–10. doi: 10.1002/art.40729
30. Martin F, Oliver AM, Kearney JF. Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity*. (2001) 14:617–29. doi: 10.1016/S1074-7613(01)00129-7
31. MacLennan IC. Extrafollicular antibody responses. *Immunol Rev*. (2003) 194:8–18. doi: 10.1034/j.1600-065X.2003.00058.x
32. Bortnick A, Allman D. What is and what should always have been: long-lived plasma cells induced by T cell-independent antigens. *J Immunol*. (2013) 190:5913–8. doi: 10.4049/jimmunol.1300161
33. Papillion AM, Kenderes KJ, Yates JL, Winslow GM. Early derivation of IgM memory cells and bone marrow plasmablasts. *PLoS ONE*. (2017) 12:e0178853. doi: 10.1371/journal.pone.0178853
34. Winslow GM, Papillion AM, Kenderes KJ, Levack RC. CD11c+ T-bet+ memory B cells: immune maintenance during chronic infection and inflammation? *Cell Immunol*. (2017) 321:8–17. doi: 10.1016/j.cellimm.2017.07.006
35. Griffin DO, Holodick NE, Rothstein TL. Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+CD27+CD43+CD70–. *J Exp Med*. (2011) 208:67–80. doi: 10.1084/jem.20101499
36. Rothstein TL, Quach TD. The human counterpart of mouse B-1 cells. *Ann N Y Acad Sci*. (2015) 1362:143–52. doi: 10.1111/nyas.12790
37. Quach TD, Rodriguez-Zhurbenko N, Hopkins TJ, Guo X, Hernandez AM, Li W, et al. Distinctions among circulating antibody-secreting cell populations, including B-1 cells, in human adult peripheral blood. *J Immunol*. (2016) 196:1060–9. doi: 10.4049/jimmunol.1501843
38. Perez-Andres M, Grosserichter-Wagener C, Teodosio C, van Dongen JJM, Orfao A, van Zelm MC. The nature of circulating CD27+CD43+ B cells. *J Exp Med*. (2011) 208:2565–6. doi: 10.1084/jem.20112203
39. Descatoire M, Weill J-C, Reynaud C-A, Weller S. A human equivalent of mouse B-1 cells? *J Exp Med*. (2011) 208:2563–4. doi: 10.1084/jem.20112232
40. Halliley JL, Tipton CM, Liesveld J, Rosenberg AF, Darce J, Gregoretti IV, et al. Long-lived plasma cells are contained within the CD19(–)CD38(hi)CD138(+) subset in human bone marrow. *Immunity*. (2015) 43:132–45. doi: 10.1016/j.immuni.2015.06.016
41. Jenks SA, Cashman KS, Zumaquero E, Marigorta UM, Patel AV, Wang X, et al. Distinct effector B cells induced by unregulated toll-like receptor 7 contribute to pathogenic responses in systemic lupus erythematosus. *Immunity*. (2018) 49:725–39.e6. doi: 10.1016/j.immuni.2018.08.015
42. Jenks SA, Cashman KS, Woodruff MC, Lee FE-H, Sanz I. Extrafollicular responses in humans and SLE. *Immunol Rev*. (2019) 288:136–48. doi: 10.1111/imr.12741
43. Wei C, Jung J, Sanz I. OMIP-003: phenotypic analysis of human memory B cells. *Cytometry A*. (2011) 79:894–6. doi: 10.1002/cyto.a.21112
44. Moir S, Ho J, Malaspina A, Wang W, DiPoto AC, O'Shea MA, et al. Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. *J Exp Med*. (2008) 205:1797–805. doi: 10.1084/jem.20072683
45. Davis RS, Wang Y-H, Kubagawa H, Cooper MD. Identification of a family of Fc receptor homologs with preferential B cell expression. *Proc Natl Acad Sci USA*. (2001) 98:9772–7. doi: 10.1073/pnas.171308498
46. Portugal S, Tipton CM, Sohn H, Kone Y, Wang J, Li S, et al. Malaria-associated atypical memory B cells exhibit markedly reduced B cell receptor signaling and effector function. *Elife*. (2015) 4:e07218. doi: 10.7554/eLife.07218
47. Brisslert M, Bokarewa M, Larsson P, Wing K, Collins LV, Tarkowski A. Phenotypic and functional characterization of human CD25+ B cells. *Immunology*. (2006) 117:548–57. doi: 10.1111/j.1365-2567.2006.02331.x
48. Mauri C, Menon M. Human regulatory B cells in health and disease: therapeutic potential. *J Clin Invest*. (2017) 127:772–9. doi: 10.1172/JCI85113
49. Ellebedy AH, Jackson KJ, Kissick HT, Nakaya HI, Davis CW, Roskin KM, et al. Defining antigen-specific plasmablast and memory B cell subsets in human blood after viral infection or vaccination. *Nat Immunol*. (2016) 17:1226–34. doi: 10.1038/ni.3533
50. Gagro A, McCloskey N, Challa A, Holder M, Grafton G, Pound JD, et al. CD5-positive and CD5-negative human B cells converge to an indistinguishable population on signalling through B-cell receptors and CD40. *Immunology*. (2000) 101:201–9. doi: 10.1046/j.1365-2567.2000.00098.x
51. Lee J, Kuchen S, Fischer R, Chang S, Lipsky PE. Identification and characterization of a human CD5+ pre-naive B cell population. *J Immunol*. (2009) 182:4116–26. doi: 10.4049/jimmunol.0803391
52. Suan D, Sundling C, Brink R. Plasma cell and memory B cell differentiation from the germinal center. *Curr Opin Immunol*. (2017) 45:97–102. doi: 10.1016/j.coi.2017.03.006

53. Basso K, Dalla-Favera R. Roles of BCL6 in normal and transformed germinal center B cells. *Immunol Rev*. (2012) 247:172–83. doi: 10.1111/j.1600-065X.2012.01112.x
54. Dengler HS, Baracho GV, Omori SA, Bruckner S, Arden KC, Castrillon DH, et al. Distinct functions for the transcription factor Foxo1 at various stages of B cell differentiation. *Nat Immunol*. (2008) 9:1388–98. doi: 10.1038/ni.1667
55. Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, et al. A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol*. (2010) 11:635–43. doi: 10.1038/ni.1891
56. Choukrallah MA, Matthias P. The interplay between chromatin and transcription factor networks during B cell development: who pulls the trigger first? *Front Immunol*. (2014) 5:156. doi: 10.3389/fimmu.2014.00156
57. Melchers F. Checkpoints that control B cell development. *J Clin Invest*. (2015) 125:2203–10. doi: 10.1172/JCI78083
58. Sciammas R, Li Y, Warmflash A, Song Y, Dinner AR, Singh H. An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. *Mol Syst Biol*. (2011) 7:495. doi: 10.1038/msb.2011.25
59. Ise W, Kohyama M, Schraml BU, Zhang T, Schwer B, Basu U, et al. The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat Immunol*. (2011) 12:536–43. doi: 10.1038/ni.2037
60. Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*. (1999) 401:556–62. doi: 10.1038/35005514
61. Myles A, Sanz I, Cancro MP. T-bet(+) B cells: a common denominator in protective and autoreactive antibody responses? *Curr Opin Immunol*. (2019) 57:40–5. doi: 10.1016/j.coi.2019.01.002
62. Allman D, Pillai S. Peripheral B cell subsets. *Curr Opin Immunol*. (2008) 20:149–57. doi: 10.1016/j.coi.2008.03.014
63. Srivastava B, Quinn WJ III, Hazard K, Erikson J, Allman D. Characterization of marginal zone B cell precursors. *J Exp Med*. (2005) 202:1225–34. doi: 10.1084/jem.20051038
64. Wirths S, Lanzavecchia A. ABCB1 transporter discriminates human resting naive B cells from cycling transitional and memory B cells. *Eur J Immunol*. (2005) 35:3433–41. doi: 10.1002/eji.200535364
65. Suryani S, Fulcher DA, Santner-Nanan B, Nanan R, Wong M, Shaw PJ, et al. Differential expression of CD21 identifies developmentally and functionally distinct subsets of human transitional B cells. *Blood*. (2010) 115:519–29. doi: 10.1182/blood-2009-07-234799
66. Koethe S, Zander L, Koster S, Annan A, Ebenfelt A, Spencer J, et al. Pivotal advance: CD45RB glycosylation is specifically regulated during human peripheral B cell differentiation. *J Leukoc Biol*. (2011) 90:5–19. doi: 10.1189/jlb.0710404
67. Merrell KT, Benschop RJ, Gauld SB, Aviszus K, Decote-Ricardo D, Wysocki LJ, et al. Identification of anergic B cells within a wild-type repertoire. *Immunity*. (2006) 25:953–62. doi: 10.1016/j.immuni.2006.10.017
68. Allman D, Lindsley RC, DeMuth W, Rudd K, Shinton SA, Hardy RR. Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J Immunol*. (2001) 167:6834–40. doi: 10.4049/jimmunol.167.12.6834
69. Teague BN, Pan Y, Mudd PA, Nakken B, Zhang Q, Szodoray P, et al. Cutting edge: transitional T3 B cells do not give rise to mature B cells, have undergone selection, and are reduced in murine lupus. *J Immunol*. (2007) 178:7511–5. doi: 10.4049/jimmunol.178.12.7511
70. Scharer CD, Blalock EL, Mi T, Barwick BG, Jenks SA, Deguchi T, et al. Epigenetic programming underpins B cell dysfunction in human SLE. *Nat Immunol*. (2019) 20:1071–82. doi: 10.1038/s41590-019-0419-9
71. Quach TD, Manjarrez-Orduno N, Adlowitz DG, Silver L, Yang H, Wei C, et al. Anergic responses characterize a large fraction of human autoreactive naive B cells expressing low levels of surface IgM. *J Immunol*. (2011) 186:4640–8. doi: 10.4049/jimmunol.1001946
72. Goodnow CC, Crosbie J, Jorgensen H, Brink RA, Basten A. Induction of self-tolerance in mature peripheral B lymphocytes. *Nature*. (1989) 342:385–91. doi: 10.1038/342385a0
73. Bijl M, Horst G, Limburg PC, Kallenberg CGM. Expression of costimulatory molecules on peripheral blood lymphocytes of patients with systemic lupus erythematosus. *Ann Rheum Dis*. (2001) 60:523–6. doi: 10.1136/ard.60.5.523
74. Chang N-H, McKenzie T, Bonventi G, Landolt-Marticorena C, Fortin PR, Gladman D, et al. Expanded population of activated antigen-engaged cells within the naive B cell compartment of patients with systemic lupus erythematosus. *J Immunol*. (2008) 180:1276–84. doi: 10.4049/jimmunol.180.2.1276
75. Hanten J, Vasilakos J, Riter C, Neys L, Lipson K, Alkan S, et al. Comparison of human B cell activation by TLR7 and TLR9 agonists. *BMC Immunol*. (2008) 9:39. doi: 10.1186/1471-2172-9-39
76. Evans DE, Munks MW, Purkerson JM, Parker DC. Resting B lymphocytes as APC for naive T lymphocytes: dependence on CD40 ligand/CD40. *J immunol*. (2000) 164:688–97. doi: 10.4049/jimmunol.164.2.688
77. Clark EA, Shu GL, Luscher B, Draves KE, Banchereau J, Ledbetter JA, et al. Activation of human B cells. Comparison of the signal transduced by IL-4 to four different competence signals. *J Immunol*. (1989) 143:3873–80.
78. Clark EA, Ledbetter JA. Activation of human B cells mediated through two distinct cell surface differentiation antigens, Bp35 and Bp50. *Proc Natl Acad Sci USA*. (1986). 83:4494–8. doi: 10.1073/pnas.83.12.4494
79. Ranheim EA, Kipps TJ. Activated T cells induce expression of B7/BB1 on normal or leukemic B cells through a CD40-dependent signal. *J Exp Med*. (1993) 177:925–35. doi: 10.1084/jem.177.4.925
80. Liu YJ, Barthelemy C, de Bouteiller O, Arpin C, Durand I, Banchereau J. Memory B cells from human tonsils colonize mucosal epithelium and directly present antigen to T cells by rapid up-regulation of B7-1 and B7-2. *Immunity*. (1995) 2:239–48. doi: 10.1016/1074-7613(95)90048-9
81. Hathcock KS, Laszlo G, Pucillo C, Linsley P, Hodes RJ. Comparative analysis of B7-1 and B7-2 costimulatory ligands: expression and function. *J Exp Med*. (1994) 180:631–40. doi: 10.1084/jem.180.2.631
82. Lenschow DJ, Sperling AI, Cooke MP, Freeman G, Rhee L, Decker DC, et al. Differential up-regulation of the B7-1 and B7-2 costimulatory molecules after Ig receptor engagement by antigen. *J Immunol*. (1994) 153:1990–7.
83. Azuma M, Ito D, Yagita H, Okumura K, Phillips JH, Lanier LL, et al. B70 antigen is a second ligand for CTLA-4 and CD28. *Nature*. (1993) 366:76–9. doi: 10.1038/366076a0
84. Weller S, Braun MC, Tan BK, Rosenwald A, Cordier C, Conley ME, et al. Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire. *Blood*. (2004) 104:3647–54. doi: 10.1182/blood-2004-01-0346
85. Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoretti I, et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat Immunol*. (2015) 16:755–65. doi: 10.1038/ni.3175
86. Kolar GR, Mehta D, Pelayo R, Capra JD. A novel human B cell subpopulation representing the initial germinal center population to express AID. *Blood*. (2007) 109:2545–52. doi: 10.1182/blood-2006-07-037150
87. Jackson SM, Harp N, Patel D, Wulf J, Spaeth ED, Dike UK, et al. Key developmental transitions in human germinal center B cells are revealed by differential CD45RB expression. *Blood*. (2009) 113:3999–4007. doi: 10.1182/blood-2008-03-145979
88. Fernandez-Gutierrez B, de Miguel S, Morado C, Hernandez-Garcia C, Banares A, Jover JA. Defective early T and T-dependent B cell activation in systemic lupus erythematosus. *Lupus*. (1998) 7:314–22. doi: 10.1191/096120398678920226
89. Kumagai S, Ishida H, Iwai K, Tsubata T, Umehara H, Ozaki S, et al. Possible different mechanisms of B cell activation in systemic lupus erythematosus and rheumatoid arthritis: opposite expression of low-affinity receptors for IgE (CD23) on their peripheral B cells. *Clin Exp Immunol*. (1989) 78:348–53.
90. Barwick BG, Scharer CD, Bally APR, Boss JM. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. *Nat Immunol*. (2016) 17:1216–25. doi: 10.1038/ni.3519
91. Defrance T, Aubry JP, Rousset F, Vanbervliet B, Bonnefoy JY, Arai N, et al. Human recombinant interleukin 4 induces Fc epsilon receptors (CD23) on normal human B lymphocytes. *J Exp Med*. (1987) 165:1459–67. doi: 10.1084/jem.165.6.1459
92. Fournier S, Delespesse G, Rubio M, Biron G, Sarfati M. CD23 antigen regulation and signaling in chronic lymphocytic leukemia. *J Clin Invest*. (1992) 89:1312–21. doi: 10.1172/JCI115717

93. Sohn HW, Krueger PD, Davis RS, Pierce SK. FcRL4 acts as an adaptive to innate molecular switch dampening BCR signaling and enhancing TLR signaling. *Blood*. (2011) 118:6332–41. doi: 10.1182/blood-2011-05-353102
94. Myles A, Gearhart PJ, Cancro MP. Signals that drive T-bet expression in B cells. *Cell Immunol*. (2017) 321:3–7. doi: 10.1016/j.cellimm.2017.09.004
95. Wang S, Wang J, Kumar V, Karnell JL, Naiman B, Gross PS, et al. IL-21 drives expansion and plasma cell differentiation of autoreactive CD11c(hi)T-bet(+) B cells in SLE. *Nat Commun*. (2018) 9:1758. doi: 10.1038/s41467-018-03750-7
96. Ehlers M, Fukuyama H, McGaha TL, Aderem A, Ravetch JV. TLR9/MyD88 signaling is required for class switching to pathogenic IgG2a and 2b autoantibodies in SLE. *J Exp Med*. (2006) 203:553–61. doi: 10.1084/jem.20052438
97. Nakano S, Morimoto S, Suzuki J, Nozawa K, Amano H, Tokano Y, et al. Role of pathogenic auto-antibody production by Toll-like receptor 9 of B cells in active systemic lupus erythematosus. *Rheumatology*. (2007) 47:145–9. doi: 10.1093/rheumatology/kem327
98. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, et al. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. *Sci Transl Med*. (2011) 3:73ra20. doi: 10.1126/scitranslmed.3001201
99. Busconi L, Bauer JW, Tumang JR, Laws A, Perkins-Mesires K, Tabor AS, et al. Functional outcome of B cell activation by chromatin immune complex engagement of the B cell receptor and TLR9. *J Immunol*. (2007) 179:7397–405. doi: 10.4049/jimmunol.179.11.7397
100. Tian J, Avalos AM, Mao S-Y, Chen B, Senthil K, Wu H, et al. Toll-like receptor 9-dependent activation by DNA-containing immune complexes is mediated by HMGB1 and RAGE. *Nat Immunol*. (2007) 8:487–96. doi: 10.1038/ni0707-780b
101. Chiche L, Jourde-Chiche N, Whalen E, Presnell S, Gersuk V, Dang K, et al. Modular transcriptional repertoire analyses of adults with systemic lupus erythematosus reveal distinct type I and type II interferon signatures. *Arthritis Rheumatol*. (2014) 66:1583–95. doi: 10.1002/art.38628
102. Sau Lee K, Diego Silva G, Jaime Martin L, Pratama A, Hu X, Chang P-P, et al. Interferon-γ excess leads to pathogenic accumulation of follicular helper T cells and germinal centers. *Immunity*. (2012) 37:880–92. doi: 10.1016/j.immuni.2012.10.010
103. Pollard KM, Cauvi DM, Toomey CB, Morris KV, Kono DH. Interferon-gamma and systemic autoimmunity. *Discov Med*. (2013) 16:123–31.
104. Wehr C, Eibel H, Masilamani M, Illges H, Schlesier M, Peter HH, et al. A new CD21low B cell population in the peripheral blood of patients with SLE. *Clin Immunol*. (2004) 113:161–71. doi: 10.1016/j.clim.2004.05.010
105. Saadoun D, Terrier B, Bannock J, Vazquez T, Massad C, Kang I, et al. Expansion of autoreactive unresponsive CD21–/low B cells in Sjogren's syndrome-associated lymphoproliferation. *Arthritis Rheum*. (2013) 65:1085–96. doi: 10.1002/art.37828
106. Glauzy S, Boccitto M, Bannock JM, Delmotte FR, Saadoun D, Cacoub P, et al. Accumulation of antigen-driven lymphoproliferations in complement receptor 2/CD21(–/low) B cells from patients with Sjogren's syndrome. *Arthritis Rheumatol*. (2018) 70:298–307. doi: 10.1002/art.40352
107. Fleischer SJ, Giesecke C, Mei HE, Lipsky PE, Daridon C, Dorner T. Increased frequency of a unique spleen tyrosine kinase bright memory B cell population in systemic lupus erythematosus. *Arthritis Rheumatol*. (2014) 66:3424–35. doi: 10.1002/art.38854
108. Kardava L, Moir S, Shah N, Wang W, Wilson R, Buckner CM, et al. Abnormal B cell memory subsets dominate HIV-specific responses in infected individuals. *J Clin Invest*. (2014) 124:3252–62. doi: 10.1172/JCI74351
109. Lau D, Lan LY, Andrews SF, Henry C, Rojas KT, Neu KE, et al. Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell differentiation. *Sci Immunol*. (2017) 2:eaai8153. doi: 10.1126/sciimmunol.aai8153
110. Jacobi AM, Reiter K, Mackay M, Aranow C, Hiepe F, Radbruch A, et al. Activated memory B cell subsets correlate with disease activity in systemic lupus erythematosus: delineation by expression of CD27, IgD, and CD95. *Arthritis Rheum*. (2008) 58:1762–73. doi: 10.1002/art.23498
111. Das R, Bar N, Ferreira M, Newman AM, Zhang L, Bailur JK, et al. Early B cell changes predict autoimmunity following combination immune checkpoint blockade. *J Clin Invest*. (2018) 128:715–20. doi: 10.1172/JCI96798
112. Isnardi I, Ng Y-S, Menard L, Meyers G, Saadoun D, Srdanovic I, et al. Complement receptor 2/CD21– human naive B cells contain mostly autoreactive unresponsive clones. *Blood*. (2010) 115:5026–36. doi: 10.1182/blood-2009-09-243071
113. Przekopowitz M, Kuppers R, Weniger MA. A large fraction of human tonsillar B cells expressing CD27 are germinal center B cells. *Immunol Cell Biol*. (2015) 93:429–30. doi: 10.1038/icb.2015.6
114. Jung J, Choe J, Li L, Choi Y. Regulation of CD27 expression in the course of germinal center B cell differentiation: the pivotal role of IL-10. *Eur J Immunol*. (2000) 30:2437–43. doi: 10.1002/1521-4141(2000)30:8<2437::AID-IMMU2437>3.0.CO;2-M
115. Agematsu K, Hokibara S, Nagumo H, Komiyama A. CD27: a memory B-cell marker. *Immunol Today*. (2000) 21:204–6. doi: 10.1016/S0167-5699(00)01605-4
116. Crotty S, Aubert RD, Glidewell J, Ahmed R. Tracking human antigen-specific memory B cells: a sensitive and generalized ELISPOT system. *J Immunol Methods*. (2004) 286:111–22. doi: 10.1016/j.jim.2003.12.015
117. Seifert M, Przekopowitz M, Taudien S, Lollies A, Ronge V, Drees B, et al. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc Natl Acad Sci USA*. (2015) 112:E546–55. doi: 10.1073/pnas.1416276112
118. Good KL, Avery DT, Tangye SG. Resting human memory B cells are intrinsically programmed for enhanced survival and responsiveness to diverse stimuli compared to naive B cells. *J Immunol*. (2009) 182:890–901. doi: 10.4049/jimmunol.182.2.890
119. Tangye SG, Avery DT, Deenick EK, Hodgkin PD. Intrinsic differences in the proliferation of naive and memory human B cells as a mechanism for enhanced secondary immune responses. *J Immunol*. (2003) 170:686–94. doi: 10.4049/jimmunol.170.2.686
120. Seifert M, Kuppers R. Human memory B cells. *Leukemia*. (2016) 30:2283–92. doi: 10.1038/leu.2016.226
121. Sanz I, Wei C, Lee FE, Anolik J. Phenotypic and functional heterogeneity of human memory B cells. *Semin Immunol*. (2008) 20:67–82. doi: 10.1016/j.smim.2007.12.006
122. Bagnara D, Squillario M, Kipling D, Mora T, Walczak AM, Da Silva L, et al. A reassessment of IgM memory subsets in humans. *J Immunol*. (2015) 195:3716–24. doi: 10.4049/jimmunol.1500753
123. Agematsu K, Nagumo H, Yang FC, Nakazawa T, Fukushima K, Ito S, et al. B cell subpopulations separated by CD27 and crucial collaboration of CD27+ B cells and helper T cells in immunoglobulin production. *Eur J Immunol*. (1997) 27:2073–9. doi: 10.1002/eji.1830270835
124. Shi Y, Agematsu K, Ochs HD, Sugane K. Functional analysis of human memory B-cell subpopulations: IgD+CD27+ B cells are crucial in secondary immune response by producing high affinity IgM. *Clin Immunol*. (2003) 108:128–37. doi: 10.1016/S1521-6616(03)00092-5
125. Weller S, Faili A, Garcia C, Braun MC, Le Deist FF, de Saint Basile GG, et al. CD40-CD40L independent Ig gene hypermutation suggests a second B cell diversification pathway in humans. *Proc Natl Acad Sci USA*. (2001) 98:1166–70. doi: 10.1073/pnas.98.3.1166
126. Agematsu K, Nagumo H, Shinozaki K, Hokibara S, Yasui K, Terada K, et al. Absence of IgDCD27 memory B cell population in X-linked hyper-IgM syndrome. *J Clin Invest*. (1998) 102:853–60. doi: 10.1172/JCI3409
127. Dogan I, Bertocci B, Vilmont V, Delbos F, Megret J, Storck S, et al. Multiple layers of B cell memory with different effector functions. *Nat Immunol*. (2009) 10:1292–9. doi: 10.1038/ni.1814
128. Liu YJ, de Bouteiller O, Arpin C, Briere F, Galibert L, Ho S, et al. Normal human IgD+IgM– germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity*. (1996) 4:603–13. doi: 10.1016/S1074-7613(00)80486-0
129. Zheng NY, Wilson K, Wang X, Boston A, Kolar G, Jackson SM, et al. Human immunoglobulin selection associated with class switch and possible tolerogenic origins for C delta class-switched B cells. *J Clin Invest*. (2004) 113:1188–201. doi: 10.1172/JCI20255
130. Koelsch K, Zheng N-Y, Zhang Q, Duty A, Helms C, Mathias MD, et al. Mature B cells class switched to IgD are autoreactive in healthy individuals. *J Clin Invest*. (2007) 117:1558–65. doi: 10.1172/JCI27628
131. Chen K, Xu W, Wilson M, He B, Miller NW, Bengten E, et al. Immunoglobulin D enhances immune surveillance by activating

antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nat Immunol*. (2009) 10:889–98. doi: 10.1038/ni.1748
132. Knox JJ, Buggert M, Kardava L, Seaton KE, Eller MA, Canaday DH, et al. T-bet+ B cells are induced by human viral infections and dominate the HIV gp140 response. *JCI Insight*. (2017) 2:92943. doi: 10.1172/jci.insight.92943
133. Kardava L, Sohn H, Youn C, Austin JW, Wang W, Buckner CM, et al. IgG3 regulates tissue-like memory B cells in HIV-infected individuals. *Nat Immunol*. (2018) 19:1001–12. doi: 10.1038/s41590-018-0180-5
134. Adlowitz DG, Barnard J, Biear JN, Cistrone C, Owen T, Wang W, et al. Expansion of activated peripheral blood memory B cells in rheumatoid arthritis, impact of B cell depletion therapy, and biomarkers of response. *PLoS ONE*. (2015) 10:e0128269. doi: 10.1371/journal.pone.0128269
135. Nicholas MW, Dooley MA, Hogan SL, Anolik J, Looney J, Sanz I, et al. A novel subset of memory B cells is enriched in autoreactivity and correlates with adverse outcomes in SLE. *Clin Immunol*. (2008) 126:189–201. doi: 10.1016/j.clim.2007.10.004
136. Menard LC, Habte S, Gonsiorek W, Lee D, Banas D, Holloway DA, et al. B cells from African American lupus patients exhibit an activated phenotype. *JCI insight*. (2016) 1:e87310. doi: 10.1172/jci.insight.87310
137. Bar-Or A, Oliveira EML, Anderson DE, Krieger JI, Duddy M, O'Connor KC, et al. Immunological memory: contribution of memory B cells expressing costimulatory molecules in the resting state. *J Immunol*. (2001) 167:5669–77. doi: 10.4049/jimmunol.167.10.5669
138. Portugal S, Obeng-Adjei N, Moir S, Crompton PD, Pierce SK. Atypical memory B cells in human chronic infectious diseases: an interim report. *Cell Immunol*. (2017) 321:18–25. doi: 10.1016/j.cellimm.2017.07.003
139. Weiss GE, Crompton PD, Li S, Walsh LA, Moir S, Traore B, et al. Atypical memory B cells are greatly expanded in individuals living in a malaria-endemic area. *J Immunol*. (2009) 183:2176–82. doi: 10.4049/jimmunol.0901297
140. Zumaquero E, Stone SL, Scharer CD, Jenks SA, Nellore A, Mousseau B, et al. IFNgamma induces epigenetic programming of human T-bet(hi) B cells and promotesTLR7/8 and IL-21 induced differentiation. *Elife*. (2019) 8:e41641. doi: 10.7554/eLife.41641
141. Racine R, Chatterjee M, Winslow GM. CD11c Expression identifies a population of extrafollicular antigen-specific splenic plasmablasts responsible for CD4 T-independent antibody responses during intracellular bacterial infection. *J Immunol*. (2008) 181:1375–85. doi: 10.4049/jimmunol.181.2.1375
142. Rubtsov AV, Rubtsova K, Fischer A, Meehan RT, Gillis JZ, Kappler JW, et al. Toll-like receptor 7 (TLR7)–driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood*. (2011) 118:1305–15. doi: 10.1182/blood-2011-01-331462
143. Rubtsova K, Rubtsov AV, Thurman JM, Mennona JM, Kappler JW, Marrack P. B cells expressing the transcription factor T-bet drive lupus-like autoimmunity. *J Clin Invest*. (2017) 127:1392–404. doi: 10.1172/JCI91250
144. Rubtsova K, Rubtsov AV, Cancro MP, Marrack P. Age-associated B cells: a T-bet-dependent effector with roles in protective and pathogenic immunity. *J Immunol*. (2015) 195:1933–7. doi: 10.4049/jimmunol.1501209
145. Manni M, Gupta S, Ricker E, Chinenov Y, Park SH, Shi M, et al. Regulation of age-associated B cells by IRF5 in systemic autoimmunity. *Nat Immunol*. (2018) 19:407–19. doi: 10.1038/s41590-018-0056-8
146. Rubtsova K, Rubtsov AV, van Dyk LF, Kappler JW, Marrack P. T-box transcription factor T-bet, a key player in a unique type of B-cell activation essential for effective viral clearance. *Proc Natl Acad Sci USA*. (2013) 110:E3216–24. doi: 10.1073/pnas.1312348110
147. Naradikian MS, Myles A, Beiting DP, Roberts KJ, Dawson L, Herati RS, et al. Cutting edge: IL-4, IL-21, and IFN-gamma interact to govern T-bet and CD11c expression in TLR-activated B cells. *J Immunol*. (2016) 197:1023–8. doi: 10.4049/jimmunol.1600522
148. Thorarinsdottir K, Camponeschi A, Cavallini N, Grimsholm O, Jacobsson L, Gjertsson I, et al. CD21(–/low) B cells in human blood are memory cells. *Clin Exp Immunol*. (2016) 185:252–62. doi: 10.1111/cei.12795
149. Kenderes KJ, Levack RC, Papillion AM, Cabrera-Martinez B, Dishaw LM, Winslow GM. T-Bet(+) IgM memory cells generate multi-lineage effector B cells. *Cell Rep*. (2018) 24:824–37.e3. doi: 10.1016/j.celrep.2018.06.074
150. Song W, Craft J. T follicular helper cell heterogeneity: time, space, and function. *Immunol Rev*. (2019) 288:85–96. doi: 10.1111/imr.12740
151. Gerth AJ, Lin L, Peng SL. T-bet regulates T-independent IgG2a class switching. *Int Immunol*. (2003) 15:937–44. doi: 10.1093/intimm/dxg093
152. Du SW, Arkatkar T, Jacobs HM, Rawlings DJ, Jackson SW. Generation of functional murine CD11c(+) age-associated B cells in the absence of B cell T-bet expression. *Eur J Immunol*. (2019) 49:170–8. doi: 10.1002/eji.201847641
153. Shapiro-Shelef M, Lin KI, McHeyzer-Williams LJ, Liao J, McHeyzer-Williams MG, Calame K. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity*. (2003) 19:607–20. doi: 10.1016/S1074-7613(03)00267-X
154. Nutt SL, Taubenheim N, Hasbold J, Corcoran LM, Hodgkin PD. The genetic network controlling plasma cell differentiation. *Semin Immunol*. (2011) 23:341–9. doi: 10.1016/j.smim.2011.08.010
155. Nutt SL, Hodgkin PD, Tarlinton DM, Corcoran LM. The generation of antibody-secreting plasma cells. *Nat Rev Immunol*. (2015) 15:160–71. doi: 10.1038/nri3795
156. Tellier J, Shi W, Minnich M, Liao Y, Crawford S, Smyth GK, et al. Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response. *Nat Immunol*. (2016) 17:323–30. doi: 10.1038/ni.3348
157. Klein U, Casola S, Cattoretti G, Shen Q, Lia M, Mo T, et al. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nat Immunol*. (2006) 7:773–82. doi: 10.1038/ni1357
158. Minnich M, Tagoh H, Bonelt P, Axelsson E, Fischer M, Cebolla B, et al. Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat Immunol*. (2016) 17:331–43. doi: 10.1038/ni.3349
159. Sciammas R, Shaffer AL, Schatz JH, Zhao H, Staudt LM, Singh H. Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity*. (2006) 25:225–36. doi: 10.1016/j.immuni.2006.07.009
160. Reimold AM, Iwakoshi NN, Manis J, Vallabhajosyula P, Szomolanyi-Tsuda E, Gravallese EM, et al. Plasma cell differentiation requires the transcription factor XBP-1. *Nature*. (2001) 412:300–7. doi: 10.1038/35085509
161. Jacobi AM, Mei H, Hoyer BF, Mumtaz IM, Thiele K, Radbruch AH, et al. HLA-DRhigh/CD27high plasmablasts indicate active disease in patients with SLE. *Ann Rheum Dis*. (2009) 69:305–8. doi: 10.1136/ard.2008.096495
162. Garimilla S, Nguyen DC, Halliley JL, Tipton C, Rosenberg AF, Fucile CF, et al. Differential transcriptome and development of human peripheral plasma cell subsets. *JCI Insight*. (2019) 4:126732. doi: 10.1172/jci.insight.126732
163. Qian Y, Wei C, Eun-Hyung Lee F, Campbell J, Halliley J, Lee JA, et al. Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data. *Cytometry B Clin Cytom*. (2010) 78 (Suppl. 1):S69–82. doi: 10.1002/cyto.b.20554
164. Arumugakani G, Stephenson SJ, Newton DJ, Rawstron A, Emery P, Doody GM, et al. Early emergence of CD19-negative human antibody-secreting cells at the plasmablast to plasma cell transition. *J Immunol*. (2017) 198:4618–28. doi: 10.4049/jimmunol.1501761
165. Huggins J, Pellegrin T, Felgar RE, Wei C, Brown M, Zheng B, et al. CpG DNA activation and plasma-cell differentiation of CD27- naive human B cells. *Blood*. (2007) 109:1611–9. doi: 10.1182/blood-2006-03-008441
166. Cocco M, Stephenson S, Care MA, Newton D, Barnes NA, Davison A, et al. *In vitro* generation of long-lived human plasma cells. *J Immunol*. (2012) 189:5773–85. doi: 10.4049/jimmunol.1103720
167. Medina F, Segundo C, Campos-Caro A, Gonzalez-Garcia I, Brieva JA. The heterogeneity shown by human plasma cells from tonsil, blood, and bone marrow reveals graded stages of increasing maturity, but local profiles of adhesion molecule expression. *Blood*. (2002) 99:2154–61. doi: 10.1182/blood.V99.6.2154
168. Gonzalez-Garcia I, Ocana E, Jimenez-Gomez G, Campos-Caro A, Brieva JA. Immunization-induced perturbation of human blood plasma cell pool: progressive maturation, IL-6 responsiveness, and high

PRDI-BF1/BLIMP1 expression are critical distinctions between antigen-specific and nonspecific plasma cells. *J Immunol.* (2006) 176:4042–50. doi: 10.4049/jimmunol.176.7.4042

169. Arce S, Luger E, Muehlinghaus G, Cassese G, Hauser A, Horst A, et al. CD38 low IgG-secreting cells are precursors of various CD38 high-expressing plasma cell populations. *J Leuk Biol.* (2004) 75:1022–8. doi: 10.1189/jlb.0603279
170. Yu C, Liu Y, Chan JT, Tong J, Li Z, Shi M, et al. Identification of human plasma cells with a lamprey monoclonal antibody. *JCI Insight*. (2016) 1:84738. doi: 10.1172/jci.insight.84738
171. Shan M, Carrillo J, Yeste A, Gutzeit C, Segura-Garzon D, Walland AC, et al. Secreted IgD amplifies humoral T helper 2 cell responses by binding basophils via galectin-9 and CD44. *Immunity*. (2018) 49:709–24.e8. doi: 10.1016/j.immuni.2018.08.013
172. Choi JH, Wang KW, Zhang D, Zhan X, Wang T, Bu CH, et al. IgD class switching is initiated by microbiota and limited to mucosa-associated lymphoid tissue in mice. *Proc Natl Acad Sci USA*. (2017) 114:E1196–204. doi: 10.1073/pnas.1621258114
173. Mei HE, Wirries I, Frolich D, Brisslert M, Giesecke C, Grun JR, et al. A unique population of IgG-expressing plasma cells lacking CD19 is enriched in human bone marrow. *Blood*. (2015) 125:1739–48. doi: 10.1182/blood-2014-02-555169
174. Xiao X, Lao X-M, Chen M-M, Liu R-X, Wei Y, Ouyang F-Z, et al. PD-1hi identifies a novel regulatory B-cell population in human hepatoma that promotes disease progression. *Cancer Discov*. (2016) 6:546–59. doi: 10.1158/2159-8290.CD-15-1408
175. Fillatreau S, Sweenie CH, McGeachy MJ, Gray D, Anderton SM. B cells regulate autoimmunity by provision of IL-10. *Nat Immunol*. (2002) 3:944–50. doi: 10.1038/ni833
176. Shen P, Roch T, Lampropoulou V, O'Connor RA, Stervbo U, Hilgenberg E, et al. IL-35-producing B cells are critical regulators of immunity during autoimmune and infectious diseases. *Nature*. (2014) 507:366–70. doi: 10.1038/nature12979
177. Bosma A, Abdel-Gadir A, Isenberg DA, Jury EC, Mauri C. Lipid-antigen presentation by CD1d(+) B cells is essential for the maintenance of invariant natural killer T cells. *Immunity*. (2012) 36:477–90. doi: 10.1016/j.immuni.2012.02.008
178. Blair PA, Norena LY, Flores-Borja F, Rawlings DJ, Isenberg DA, Ehrenstein MR, et al. CD19(+)CD24(hi)CD38(hi) B cells exhibit regulatory capacity in healthy individuals but are functionally impaired in systemic Lupus Erythematosus patients. *Immunity*. (2010) 32:129–40. doi: 10.1016/j.immuni.2009.11.009
179. Flores-Borja F, Bosma A, Ng D, Reddy V, Ehrenstein MR, Isenberg DA, et al. CD19+CD24hiCD38hi B cells maintain regulatory T cells while limiting TH1 and TH17 differentiation. *Sci Transl Med*. (2013) 5:173ra23. doi: 10.1126/scitranslmed.3005407
180. Aravena O, Ferrier A, Menon M, Mauri C, Aguillon JC, Soto L, et al. TIM-1 defines a human regulatory B cell population that is altered in frequency and function in systemic sclerosis patients. *Arthritis Res Ther*. (2017) 19:8. doi: 10.1186/s13075-016-1213-9
181. Iwata Y, Matsushita T, Horikawa M, Dilillo DJ, Yanaba K, Venturi GM, et al. Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells. *Blood*. (2011) 117:530–41. doi: 10.1182/blood-2010-07-294249
182. Maseda D, Smith SH, DiLillo DJ, Bryant JM, Candando KM, Weaver CT, et al. Regulatory B10 cells differentiate into antibody-secreting cells after transient IL-10 production *in vivo*. *J Immunol*. (2012) 188:1036–48. doi: 10.4049/jimmunol.1102500
183. Candando KM, Lykken JM, Tedder TF. B10 cell regulation of health and disease. *Immunol Rev*. (2014) 259:259–72. doi: 10.1111/imr.12176
184. Tedder TF. B10 cells: a functionally defined regulatory B cell subset. *J Immunol*. (2015) 194:1395–401. doi: 10.4049/jimmunol.1401329
185. Liu J, Zhan W, Kim CJ, Clayton K, Zhao H, Lee E, et al. IL-10-producing B cells are induced early in HIV-1 infection and suppress HIV-1-specific T cell responses. *PLoS ONE*. (2014) 9:e89236. doi: 10.1371/journal.pone.0089236
186. Duddy M, Niino M, Adatia F, Hebert S, Freedman M, Atkins H, et al. Distinct effector cytokine profiles of memory and naive human B cell subsets and implication in multiple sclerosis. *J Immunol*. (2007) 178:6092–9. doi: 10.4049/jimmunol.178.10.6092
187. Fillatreau S. Regulatory plasma cells. *Curr Opin Pharmacol*. (2015) 23:1–5. doi: 10.1016/j.coph.2015.04.006
188. Fillatreau S. Natural regulatory plasma cells. *Curr Opin Immunol*. (2018) 55:62–6. doi: 10.1016/j.coi.2018.09.012
189. Lino AC, Dang VD, Lampropoulou V, Welle A, Joedicke J, Pohar J, et al. LAG-3 inhibitory receptor expression identifies immunosuppressive natural regulatory plasma cells. *Immunity*. (2018) 49:120–33.e9. doi: 10.1016/j.immuni.2018.06.007
190. Matsumoto M, Baba A, Yokota T, Nishikawa H, Ohkawa Y, Kayama H, et al. Interleukin-10-producing plasmablasts exert regulatory function in autoimmune inflammation. *Immunity*. (2014) 41:1040–51. doi: 10.1016/j.immuni.2014.10.016
191. Shalapour S, Font-Burgada J, Di Caro G, Zhong Z, Sanchez-Lopez E, Dhar D, et al. Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy. *Nature*. (2015) 521:94–8. doi: 10.1038/nature14395

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 23 October 2019
doi: 10.3389/fimmu.2019.02434

CD Maps—Dynamic Profiling of CD1–CD100 Surface Expression on Human Leukocyte and Lymphocyte Subsets

Tomas Kalina[1,2†], Karel Fišer[1†], Martin Pérez-Andrés[3,4], Daniela Kužílková[1], Marta Cuenca[5], Sophinus J. W. Bartol[6], Elena Blanco[3,4], Pablo Engel[5], and Menno C. van Zelm[6,7], on behalf of the Human Cell Differentiation Molecules (HCDM) organization*

[1] CLIP - Childhood Leukaemia Investigation Prague, Department of Paediatric Haematology and Oncology, Charles University, Prague, Czechia, [2] Department of Paediatric Haematology and Oncology, University Hospital Motol, Prague, Czechia, [3] Department of Medicine, Cancer Research Centre (IBMCC, USAL-CSIC), Cytometry Service (NUCLEUS), Institute of Biomedical Research of Salamanca, University of Salamanca, Salamanca, Spain, [4] Biomedical Research Networking Centre Consortium of Oncology (CIBERONC), Instituto de Salud Carlos III, Madrid, Spain, [5] Department of Biomedical Sciences, University of Barcelona, Barcelona, Spain, [6] Department of Immunology, Erasmus MC, University Medical Center, Rotterdam, Netherlands, [7] Department of Immunology and Pathology, Monash University and the Alfred Hospital, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Harry W. Schroeder, University of Alabama at Birmingham, United States

Reviewed by:
Kay L. Medina, Mayo Clinic, United States
Rong Jin, Peking University, China

***Correspondence:**
Menno C. van Zelm
menno.vanzelm@monash.edu

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *04 June 2019*
Accepted: *30 September 2019*
Published: *23 October 2019*

Citation:
Kalina T, Fišer K, Pérez-Andrés M, Kužílková D, Cuenca M, Bartol SJW, Blanco E, Engel P and van Zelm MC (2019) CD Maps—Dynamic Profiling of CD1–CD100 Surface Expression on Human Leukocyte and Lymphocyte Subsets. Front. Immunol. 10:2434. doi: 10.3389/fimmu.2019.02434

CD molecules are surface molecules expressed on cells of the immune system that play key roles in immune cell-cell communication and sensing the microenvironment. These molecules are essential markers for the identification and isolation of leukocytes and lymphocyte subsets. Here, we present the results of the first phase of the CD Maps study, mapping the expression of CD1–CD100 (n = 110) on 47 immune cell subsets from blood, thymus, and tonsil using an eight-color standardized EuroFlow approach and quantification of expression. The resulting dataset included median antibody binding capacities (ABCs) and percentage of positivity for all markers on all subsets and was developed into an interactive CD Maps web resource. Using the resource, we examined differentially expressed proteins between granulocyte, monocyte, and dendritic cell subsets, and profiled dynamic expression of markers during thymocyte differentiation, T-cell maturation, and between functionally distinct B-cell subset clusters. The CD Maps resource will serve as a benchmark of antibody reactivities ensuring improved reproducibility of flow cytometry-based research. Moreover, it will provide a full picture of the surfaceome of human immune cells and serves as a useful platform to increase our understanding of leukocyte biology, as well as to facilitate the identification of new biomarkers and therapeutic targets of immunological and hematological diseases.

Keywords: CD marker, surfaceome, lymphocyte, monocyte, flow cytometry, expression profiling, B-cell, T-cell

INTRODUCTION

Leukocytes display on their surface molecules that are crucial for sensing hazardous environmental changes and mediating cell adhesion and communication between cells both within the immune system and with stroma. These include receptors, transporters, channels, cell-adhesion proteins, and enzymes. The complexity of surface-expressed proteins, also called

the surfaceome, is emphasized by the fact that an estimated 26% of human genes encode transmembrane proteins (∼5,500) (1). However, recent *in silico* evaluations predict that 2,886 proteins are actually expressed at the outer cell membrane, i.e., the cell surface (2). Experimental evidence exists for ∼1,492 proteins across multiple tissues (3) and 1,015 proteins that are expressed in one or more immune cell type and lymphoid tissue (4).

Over the past four decades, a vast array of cell surface molecules has been discovered through the production of monoclonal antibodies (mAbs) (5). These mAbs, together with the development of multicolor flow cytometric analysis (6), have been instrumental to determine their expression and function. Human leukocyte differentiation antigen (HLDA) workshops have led to the characterization and formal designation of more than 400 surface molecules (7, 8), known as CD molecules (www.hcdm.org). CD nomenclature provides a unified designation system for mAbs, as well as for the cell surface molecules that they recognize. These molecules include receptors, adhesion molecules, membrane-bound enzymes, and glycans that play multiple roles in leukocyte development, activation, and differentiation. CD molecules are routinely used as cell markers, allowing the identification of the presence and proportions of specific leukocyte cell populations and lymphocyte subsets, and their isolation, using combinations of fluorochrome-labeled antibodies and flow cytometry. Importantly, analysis of CD molecules, known as immunophenotyping, is a fundamental component for the diagnosis, classification, and follow-up of hematological malignancies and immunodeficiencies, and the monitoring of immune system disorders such as autoimmune diseases. More recently, mAbs recognizing CD molecules have been established as invaluable tools for the treatment of cancer, such as checkpoint inhibitors (9), and autoimmune diseases (10). Development and testing of such therapeutics rely on accurate knowledge expression and function of the target molecule as has been negatively illustrated by the disaster in the Phase I TGN1412 study with an anti-CD28 superagonist (11).

Currently, there are extensive gaps in our knowledge of CD molecule expression patterns, mainly because of the discordancy in the setup of the expression studies and the major changes in flow cytometry technology over the last 30 years (12). As a result, there has been overinterpretation in summarizing tables, which can be misleading. Thus, there is an urgent need to construct a higher resolution and accurate map of the expression profiles of the CD molecules to visualize the surface of leukocyte landscape. Moreover, an important part of the bibliography is incorrect and often misleading.

To correct current misinterpretation and to overcome gaps in knowledge, the HCDM has initiated the CD Maps project, a multi-institute research program to generate a high-resolution map of the cell surface of human immune cells using standardized multicolor flow cytometry protocols. Here, we present the results of the first phase of the CD Maps study, which includes the expression signature of CD1–CD100 on 47 cell populations and subsets, 41 of which were non-overlapping. The data have been acquired across four expert flow cytometry laboratories to ensure reproducibility and have been built into an online web resource with free user access. Expression profiling of CD markers across immune cell subsets revealed dynamic changes in expression levels and hints at further immune cell diversity for markers that were expressed on a fraction of defined populations. These insights can prove critical for development of therapeutics targeting dysregulated immune responses or malignant cells.

MATERIALS AND METHODS

Human Tissue Samples

The use of human pediatric tissue and adult buffy coats was approved by the Human Ethics Committees of the Erasmus Medical Center, the University Hospital Motol, and the universities of Salamanca and Barcelona, and was contingent on informed consent in accordance with the Declaration of Helsinki. Thymus material was obtained from 12 children requiring surgery for congenital heart disease. These children did not have hematologic or immunologic diseases. Non-necrotizing tonsil tissue was obtained from seven donors, including two adults (32 and 34 years) and five children (4–8 years) who underwent scheduled tonsillectomy. Blood buffy coats of 12 healthy adult volunteer donors were obtained from the local blood banks.

Single Cell Isolation and Preparation

The blood leukocyte isolation protocol was optimized to minimize platelet adhesion (satellitism). Briefly, the buffy coat suspension was diluted 6× in PBS containing 2 mM EDTA, followed by adding an equal volume of a 4% dextran solution (Sigma-Aldrich, Saint Louis, MO, USA) in 0.9% NaCl. The mixture was left for 30 min for erythrocytes to sediment prior to collecting the supernatant containing the leukocytes. Following a spin (130 *g*, 15 min, RT) and removal of the supernatant, the residual erythrocytes in the pellet were lysed using hypotonic lysis with a 0.2% NaCl solution for 55 s, followed by supplementation of 1.2% NaCl to achieve an isotonic concentration of NaCl. Following addition of PBS and a spin (130 *g*, 15 min, RT), the lysis step was repeated. Finally, the suspension of leukocytes was washed and diluted with PBS/BSA (PBS with 0.5% BSA and 0.09% NaN_3) to a final concentration of 4×10^7/ml.

Thymocytes and tonsillar lymphocytes were isolated via gentle shaking from manually dissociated thymus and tonsil tissue, respectively, washed with RPMI 1640 with 25 mM HEPES, L-glutamine, 100 U/ml penicillin, and 100 mg/ml streptomycin (Lonza, Basel, Switzerland) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS, Thermo Fisher Scientific, Rockford, IL). Single-cell suspensions were either directly used for immunophenotyping, or stored in FBS with 10% DMSO in liquid nitrogen for analysis at a later stage. Live frozen cells were thawed by dropwise addition of 1 ml FBS, followed by addition of 8 ml of medium. Cells were washed twice, counted, washed, and diluted with PBS/BSA (PBS with 0.5% BSA and 0.09% NaN_3) to a final concentration of 1.25×10^7/ml. Whenever frozen and thawed thymocytes were used,

we observed a marked decrease of proportion of double-positive stage thymocytes, but their phenotype was similar to the fresh thymocytes.

Staining of Cells With Antibodies for Immunophenotyping

Cells were stained in V-bottom 96-well plates in a total suspension volume of 50 μl. First, one of each of the PE-labeled mAbs to CD1–CD100 were added to each well (details of each marker are provided in **Supplementary Table 2**). The amounts were according to the manufacturer's recommended titer and topped up to 10 μl with PBS/BSA. Subsequently, 40 μl of cell suspension (1.6×10^6 cells for buffy coats, 5×10^5 cells for thymus or tonsil) was added to each well. Following careful mixing, the suspensions were incubated for 30 min at room temperature in the dark. Next, 25 μl of backbone mAb reagent mix was added to each well, carefully mixed, and incubated again for 30 min (RT, in the dark). Four Ab backbone cocktails were prepared (two for blood, one for thymus, and one for tonsil), and the reagents were titrated beforehand (details provided in **Supplementary Table 1**). The cells were washed three times (8 min, 500 *g*, RT) in PBS/BSA and resuspended in 200 μl of PBS with 2 mM EDTA for acquisition. A detailed CD Maps standard operating protocol can be downloaded from www.hcdm.org. Although we aimed for the complete set of CD1–CD100 markers, we were limited to the 110 that were commercially available and that were not of the IgM isotype. The following CD markers were not included: (a) mAbs with IgM isotype against carbohydrate antigens that were not available as PE-conjugates: CDw12, CD15u, CD15s, CD15su, CD17, CD60a, CD60b, CD60c, CD65, CD65s, CD75, and CD75s; (b) mAbs that were validated by the HLDA workshops, but that were not commercially obtainable: CD1c, CD66a, CD66d, CD66e, CD66f, CD85a, CD92, and CD94. Furthermore, several CD markers were present as backbone markers in our panels potentially interfering with the PE staining. To mitigate the blocking effect on the PE-reagent, we (a) used a different clone known to bind a distinct epitope (e.g., CD16, CD45), and where no clone with a distinct epitope was available, we (b) incubated the cells first with the PE-conjugate for 15 min, prior to addition of the backbone cocktail. When the backbone marker was impacted, the gating strategy was manually adjusted using the PE-conjugated marker. The CD1–CD100 markers were assessed with commercially available reagents from three different vendors and used at vendor-recommended titers. Some reagents exhibited higher background staining than others, which is probably due to these having a lower antigen affinity and were therefore used at higher concentration. This could explain why the expression levels (MFI) for some CD markers were above that of the FMO in a subset that is known not to express it. Finally, some subsets (particularly myeloid cells and cells from tonsil) exhibited high background autofluorescence and some degree of non-specific binding (13).

Flow Cytometer Instrument Setup

Data acquisition was performed on four different sites on LSR II, LSR Fortessa, and FACS Canto instruments (BD Biosciences, San Jose, CA, USA) equipped with 405-nm, 488-nm, and 633/647-nm excitation lasers and an HTS loader. Cytometer Setup and Tracking (CS&T) beads (BD Biosciences) and 8-peak Rainbow bead calibration particles (Spherotech, Lake Forest, IL, USA) were used for PMT voltages and light scatter setup to achieve inter-laboratory standardization as developed by the EuroFlow consortium (14). Each panel was applied on a total of 12 donors, and 1 million events were acquired per staining (well). The EuroFlow Standard Operating Procedure (SOP) for Instrument Setup and Compensation can be downloaded from www.euroflow.org. Three out of four laboratories participate in the EuroFlow Quality Assessment scheme that investigates the MFI of selected cell subsets (15). The same concept was adopted to test the performance of the four laboratories on a testing cohort of three local donors using four reagents (CD8, CD21, CD25, and CD28) representing different staining intensities.

Conversion of PE Fluorescence Intensity to Antibody Binding Capacity (ABC)

To convert PE fluorescence to the amount of PE molecules bound to a target, we used the PE Fluorescence Quantitation Kit (BD Biosciences) with four known levels of PE. The pellet was resuspended in 500 μl of PBS/BSA and analyzed by flow cytometry in parallel with each experiment. The measured PE signals for all stainings on all cell subsets were fitted to the PE calibration curve to extract the number of PE molecules.

PE-conjugation of mAbs is quite consistent with a 1:1 ratio of fluorochrome:antibody. To test and correct for any deviations, we have measured and calculated a correction factor reflecting the amount of PE for each antibody (correction factors were in the range 0.73–1.32, mean – 1 SD to mean + 1 SD). A volume of 25 μl of UltraComp eBeads™ Compensation Beads (Thermo Fischer Scientific) was diluted with 15 μl of PBS/BSA, mixed with excess of tested PE-labeled antibody and incubated for 30 min, RT, in the dark. Compensation Beads were washed twice in PBS/BSA (8 min, 500 *g*, RT), resuspended in 70 μl of PBS with 2 mM EDTA, and analyzed by flow cytometry. All 116 mAbs were measured, and for each mAb, a ratio of individual median PE/(median of all medians) was calculated as a correction factor. A standard deviation of the correction factor was 0.3; a total of 26 mAbs (25%) of all mAbs yielded a correction factor above or below 1 standard deviation; thus, for mAbs with a correction factor <0.7 or above 1.3, the measurement was repeated to exclude any outliers. The average of all correction factor values (after exclusion of outliers) was used to recalculate the ABC for all 111 CD markers on all 47 defined subsets.

Analysis, Gating, and Export of Values

Leukocyte and lymphocyte subsets to be analyzed were pre-defined (**Supplementary Figures 1–4**), and all acquisitions for each of the four panels were gated by a single laboratory using FlowJo (version 9 or 10) or Infinicyt software. From each defined subset, the following set of statistics was extracted for the PE channel: median, mean, mode, CV, 10th, 25th,

50th (median), 75th, 90th percentile (**Supplementary Figure 5**). Furthermore, a gate was set to define the percentage of positive events, using the fluorescence minus one (FMO) staining as a negative control. The minimum cell count for statistical evaluation was set to 100, and subsets with lower cell counts were omitted from further analysis. Samples with <500,000 events in the leukocyte gate (CD45+) or samples with an apparent shift in CD45 expression with time during acquisition (indicative of clogging) were not used for analysis (manually curated).

The conversion from PE fluorescence to target molecule number (ABC unit) was performed as described above using the "define calibration" function in either the FlowJo or Infinicyt software packages. Descriptive statistics obtained from these software programs were exported for all defined subsets into one delimited flat table text file per tube. To these tables, additional information on material source, antibody characteristics, experiment details, etc. were added, as well as uniform cell subset identifiers: short machine friendly names, longer descriptive names.

Data Import and Pre-processing

All subsequent work was carried out in R Development Core Team (16). All used R packages are listed and references are provided in **Supplementary Table 4**. Data were imported into the R environment using standard import functions, converting data to R objects. Each of the four data flat tables from the four tubes was processed separately. After checks for duplicated data entries, these were converted into matrix-like formats and previously calculated median correction factors were applied. Sample wise centrality measures (means and medians) were calculated and data were converted from wide to long format for easier subsequent computation. Dictionaries of cell subset and statistics-related terms were built and combined from all sources. The processed and combined data were stored in binary format and were cleaned (all non-positive values were converted to the value one), a correction factor was applied, and group-based centrality statistics (mean and median) were calculated.

Distribution of Frequency of PE-Positive Cells

Sigmoidal fit and separation of markers into positive, intermediate, and negative groups on a per-cell subset basis was performed using R package sicegar (17). Simple sigmoidal fit was performed by logistic function

$$PE\left(cds\right) = f_{sig}\left(cds\right) = \frac{PE_{\max}}{1 + \exp\left(-a_1\left(cds - cds_{mid}\right)\right)}$$

where *PE(cds)* is the percentage of PE-positive cells, given as a function of sequence of CD markers *cds*. The CD markers are ordered based on rising median percentage of PE-positive cells. There are three parameters to be fitted: PE_{max}—maximum percentage of PE-positive cells, cds_{mid}—midpoint as half of maximum, and a_1. The a_1 parameter is related to the slope of *PE(cds)* at $cds = cds_{mid}$ via the formula.

$$\frac{d}{dcds}PE\left(cds\right)|_{cds=cds_{mid}} = \frac{a_1 PE_{\max}}{4}$$

Distribution of Median Fluorescence Intensity

Modeling of a turning point in a sequence of rising median fluorescence intensity per cell subset was done using Menger curvature adapted from Christopoulos (18).

The Menger curvature for $y = f(x)$ at (x_i, y_i) is:

$$DC\left(x_i\right) = \frac{\sqrt{A - B^2}}{\|pq\|\,\|qr\|\,\|rq\|}$$

where

$$\begin{aligned} A &= 4\|pq\|^2\|qr\|^2 \\ B &= \|pq\|^2 + \|qr\|^2 - \|rp\|^2 \\ \|pq\| &= \sqrt{(x_{i-1} - x_i)^2 + \left(y_{i-1} - y_i\right)^2} \\ \|qr\| &= \sqrt{(x_i - x_{i+1})^2 + \left(y_i - y_{i+1}\right)^2} \\ \|rp\| &= \sqrt{(x_{i+1} - x_{i-1})^2 + \left(y_{i+1} - y_{i-1}\right)^2} \end{aligned}$$

And the convex turning point at section of the curve is:

$$D = max\left\{DC\left(x_i\right), i = 2, \ldots, n-1\right\}$$

Hierarchical Clustering Analysis

For hierarchical clustering analysis (HCA), the pheatmap R package (https://github.com/raivokolde/pheatmap) was used. Per cell subset, median Qb values were $\log_{10}$ transformed after minimum median Qb values were raised above zero. Observations with missing values and FMO controls were removed and data were *z*-score scaled. For HCA, the Euclidean distance and Ward linkage (ward.D2) were used (19).

Generation and Utilities of a Dynamic Web Resource

To share CD Maps data as a resource with a user-friendly interface, an application with web page front-end was written in R using the R package Shiny. Shiny allows background computations in R serving results to a web-based front-end and uses a reactive programming paradigm. Reactive programming allows for dynamic user-directed content generation and therefore interactive data exploration and analysis. For enhanced user interactivity, several R packages were used that facilitate access to JavaScript libraries (e.g., d3heatmap, htmlwidgets). The resulting web page includes general CD Maps information, as well as several angles from which to interrogate CD Maps data (www.hcdm.org; **Figure 1**).

An example is the interrogation feature "What are protein levels of selected CD markers on selected sequence of cell subsets?" For this scenario, the user is able to select CD

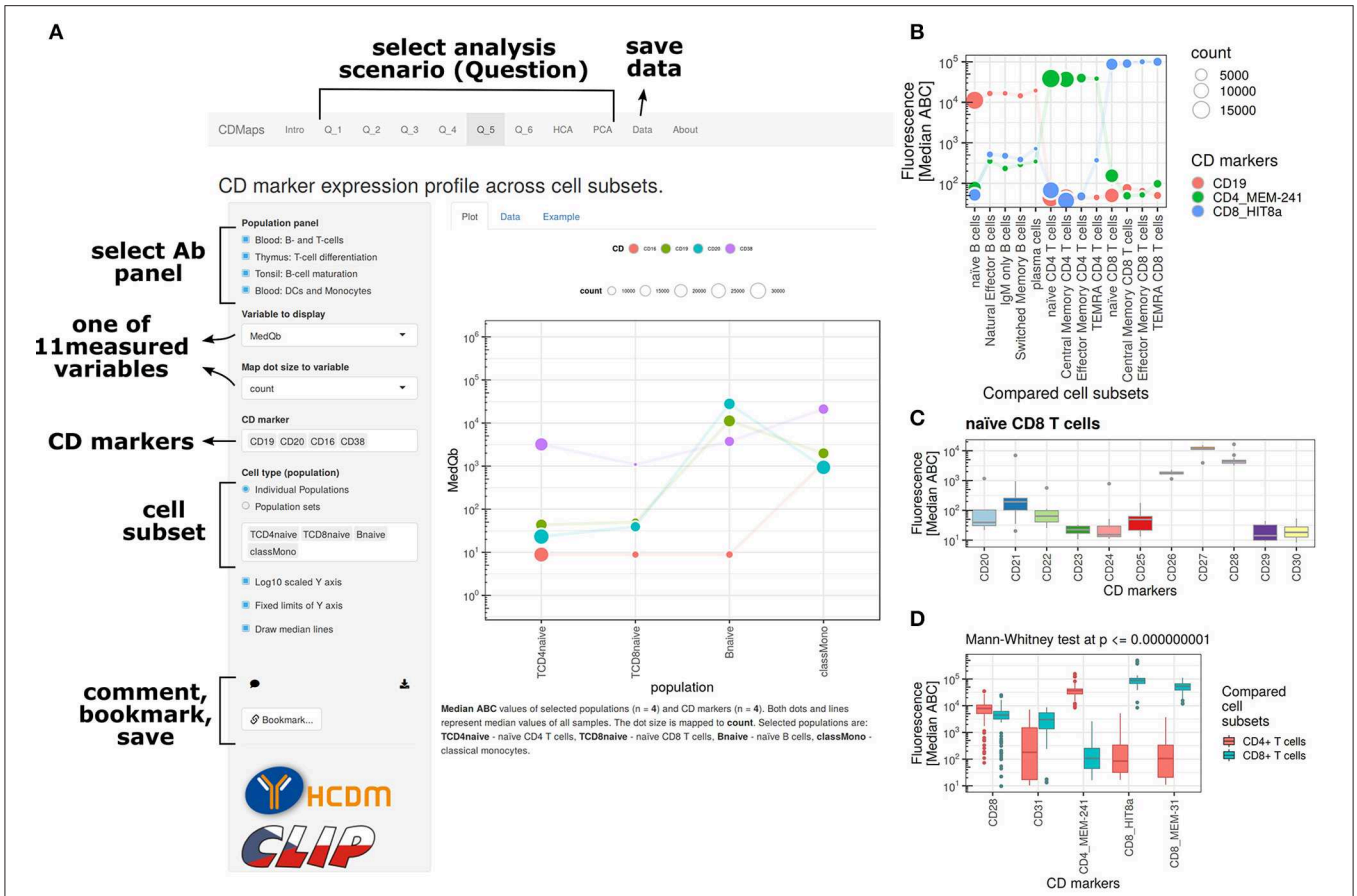

FIGURE 1 | Overview of the CD Maps web resource available at hcdm.org. **(A)** The user interface allows for interactive data interrogations. The resource uses several scenarios for data exploration and allows for data download, bookmarking of analysis state, and image export. Analysis examples: **(B)** paired analysis of CD marker expression on cell subsets, e.g., in a differentiation setting. **(C)** Visualization of expression of multiple CD markers including a measure of variation for a single subset. **(D)** Analysis and visualization of statistically differentially expressed CD markers between two subsets or two groups of cell subsets.

markers and a sequence of cell subsets to visualize expression in multiple subsets using a dot-line plot. The sequence of cell subsets is based on the order in which these have been selected, and the values on the *y* axis are by default the median ABC values from all biological repeats. The variable displayed on the *y* axis can be exchanged by the user for any of the available cell subset statistics. As the graph is also a dot plot, the size of the dots can be used to visualize an additional quantitative parameter per cell subset and can be selected by the user (e.g., percentage of PE-positive cells). The line plot uses unique colors for each selected CD marker. Besides the graph itself, the application also dynamically generates figure captions. Finally, the application also allows the user to "bookmark" the state selected settings in the application for later follow-up analysis. In conclusion, the web resource functions are based on the principle that the user specifies details for data interrogation within given scenario boundaries, and such details are sent to the web server, where R is used to compute and prepare outputs, and those outputs are sent back in real time to user, giving a smooth, dynamic, and interactive feeling to the user.

Reproducibility and Version Control

Reproducibility and version control of data processing and application development throughout the project were achieved using GIT versioning software (https://git-scm.com/) RStudio IDE (RStudio, Inc., Boston, MA, USA) and Bitbucket repository (Atlassian, Sydney, Australia). Deployment is facilitated via Docker virtualization (https://www.docker.com/, Docker, Inc., San Francisco, CA, USA).

RESULTS

Generation of a Web Resource for Expression Profiling of CD1–CD100 on Major Immune Cell Lineages and Their Subsets

To investigate the expression levels on major leukocytes, subsets of the first surface molecules that had been defined in the 1980s and early 1990s with CD markers 1–100 (20–24), we developed a multicolor immune phenotyping panel consisting of four tubes: (A) innate and (B) adaptive immune cells from blood (25, 26),

(C) B-cell subsets from tonsil (27, 28), and (D) T-cell progenitors in thymus (29, 30) (**Supplementary Table 1**). One channel was reserved for a PE-labeled drop-in mAb directed against one of the CD1–CD100 antigens (**Supplementary Table 2**). Twelve biological repeats were acquired, and after curation (detailed SOP in **Supplementary Data Sheet 1**), expression analysis was performed on nine biological repeats for tube A, 11 for B, 7 for C, and 5 for D. Multiple descriptors of CD marker expression were defined for each gated cell subsets and exported (**Supplementary Figure 5**), including the median fluorescence intensity, which was converted to ABC using the QuantiBRITE bead measurements, and the percentage of positive cells using the FMO control value as cutoff.

The resulting dataset consisted of over a million data points of derived statistics and annotation information that together form a quantitative insight into the cell surfaceome of the human immune system. To make the data accessible as a major resource for detailed studies by us and the scientific community, we constructed an interactive web-based application (**Figure 1**). The resource contains multiple features to visualize the complete dataset [e.g., principal component analysis (PCA)] and to examine specific cell lineages and/or subsets (e.g., pairwise comparisons and patterns of expression during cell maturation).

The combined information of CD marker expression levels and percentages of positive cells were depicted as a "drop plot" (**Figure 2**), in which colors represent the ABC and the dot sizes represent the percentage of positivity. The CD markers displayed a wide range of expression patterns. For example, CD44, CD45, CD46, and CD47 were highly expressed on nearly all cells within the majority of defined subsets, whereas CD49a, CD49b, and CD49c were typically expressed at low levels. Importantly, all markers showed positivity for at least one subset, and the expression patterns of molecules such as CD3, CD4, CD8, CD14, CD19, and CD20 agreed with their designation as well-defined lineage markers (**Figure 2**).

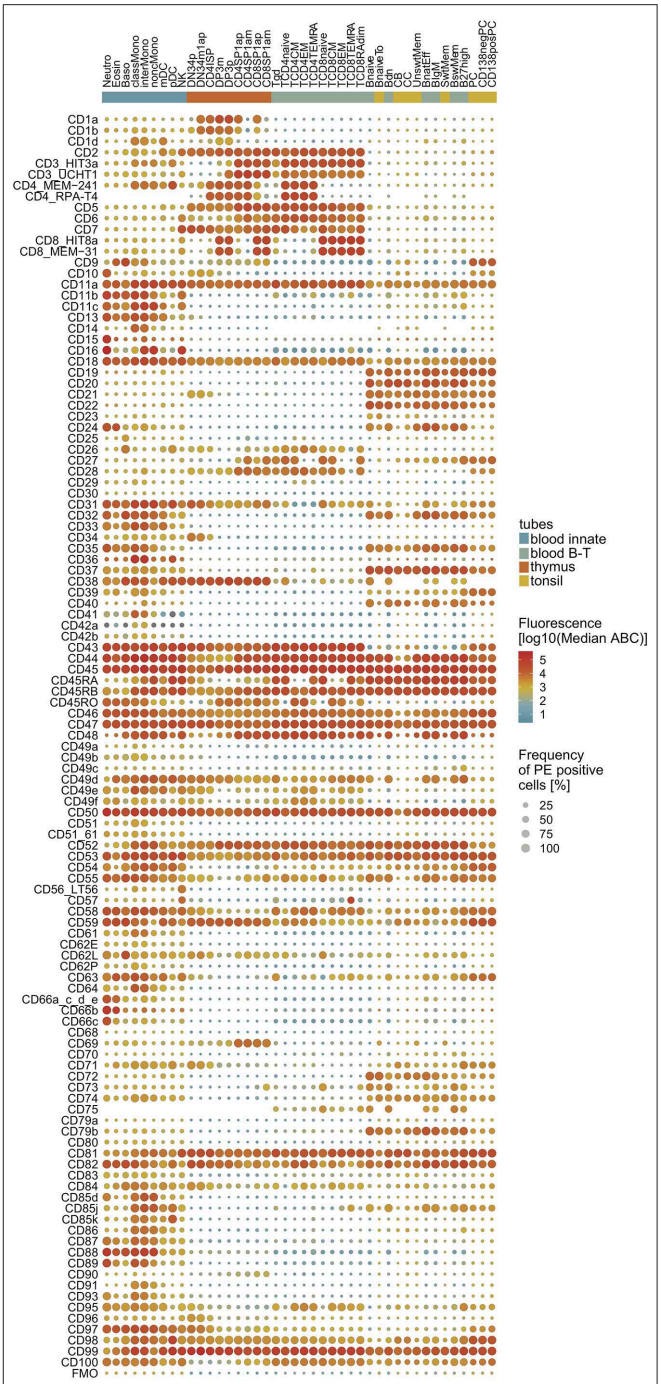

FIGURE 2 | Expression map of CD1–CD100 on all 42 non-overlapping cell subsets. CD markers are numerically ordered vertically with the FMO on the bottom row. The cell subsets are grouped (innate cells; thymocytes; T-cells; B cells) and sorted within lineage on their maturity. The median expression level is visualized by color, and the median percentage of positive cells is visualized by the size of the dot. For cell type abbreviations, see **Supplementary Table 3**.

Intra- and Inter-population Variation of CD Marker Expression

Further examination involved the relative intensity of expression of all CD markers in all defined cell subsets (**Supplementary Figure 6** and **Supplementary Table 4**). The most highly expressed markers (e.g., CD45 on naive CD4 T-cells; **Figure 3A**) reached 10^5 ABC units, with lower expression levels for, e.g., CD3 and CD27 at 10^4, and CD31 and CD49f at 10^3. Ubiquitously expressed molecules on immune cells such as CD44, CD45, CD46, CD47, CD50, CD98, and CD99 had a low coefficient of variation (CV) across the studied subsets (**Figure 3B**), as did some molecules with overall low expression levels (e.g., CD49c). In contrast, as expected, markers with lineage- and/or subset-specific expression patterns show a greater degree of heterogeneity in expression over the examined subsets (e.g., CD19, CD24, CD35).

To examine donor variation for expression all markers, CVs were calculated per cell subset for each marker and displayed as box whisker plots (**Figure 3C**). In general, the highly expressed markers were found to have relatively low inter-donor variability, whereas the CVs were higher for CD markers that were expressed at low levels (**Figure 3C**). Indeed, some of the markers with small boxes in **Figure 3A** (CD44, CD45, CD46, CD47, CD98, and

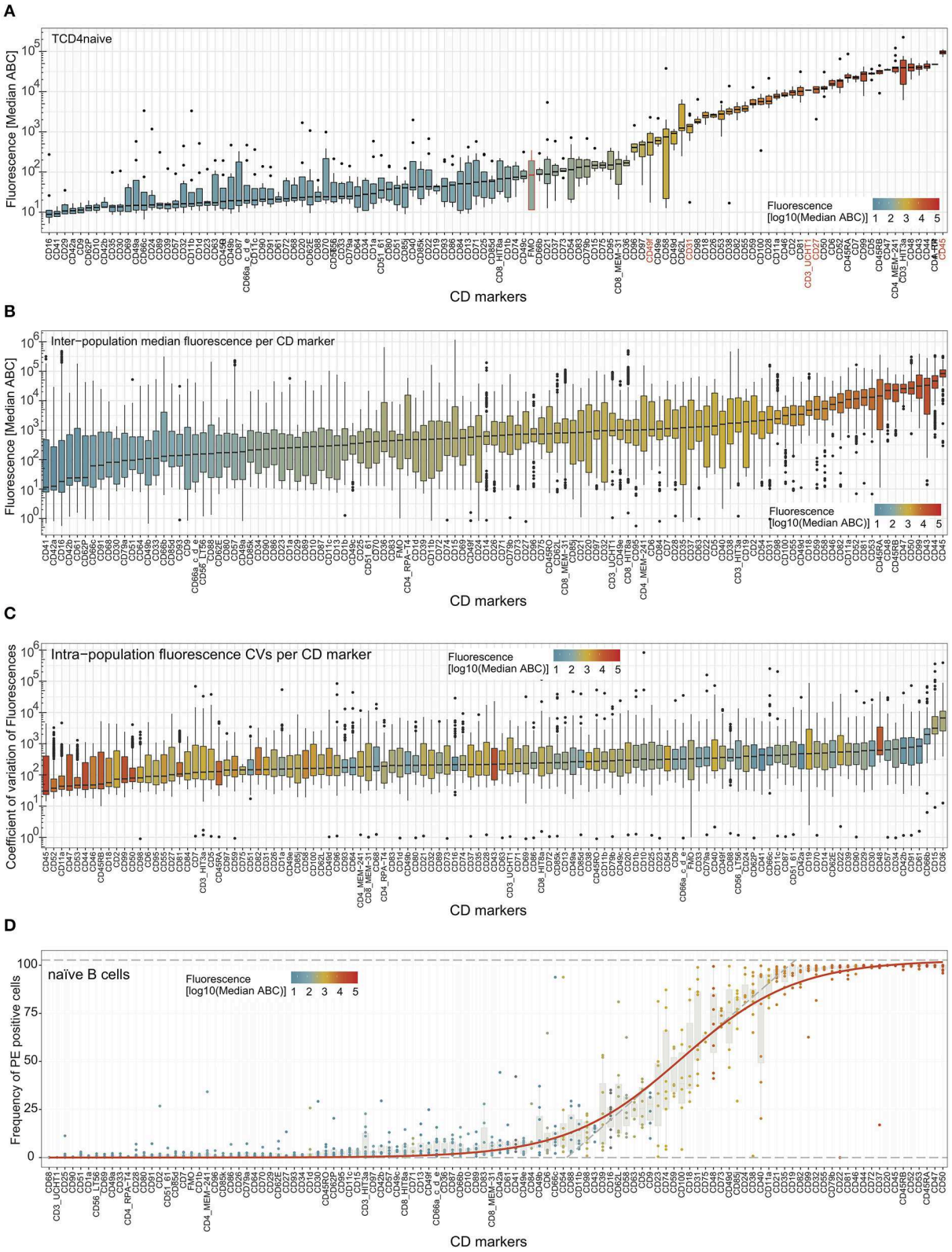

FIGURE 3 | Expression levels and heterogeneity of expression of cell surface markers across cell types. **(A)** Median fluorescence (in antibody binding capacity; ABC) for all markers on one cell subset (naïve CD4 T-cells) ordered from low to high median expression. CD markers in red font are discussed in the main text, FMO is highlighted in orange, and a horizontal orange line depicts the median FMO background. Similar plots for all cell subsets are provided in **Supplementary Figure 6**.

(Continued)

FIGURE 3 | (B) Fluorescence (in ABC) across all cell subsets per CD marker with box whisker plots (median, IQR, and range). The CD markers are ordered from low to high median expression (black horizontal lines). **(C)** Coefficients of variation (CV) of expression across all cell subsets per CD marker. The CD markers are ordered from low to high median CV (black horizontal lines) as box whisker plots with the color representing the median expression level ABC. **(D)** Frequency of positive cells for all markers on one cell subset (naive B-cells) ordered from low to high frequency. Similar plots for all cell subsets are provided in **Supplementary Figure 7**. In all plots, fluorescence intensity is also represented by the coloring of the boxes.

CD99) were highly expressed and showed a relatively low CV. Still, some CD markers had a higher variability of expression in all cell subsets (CD15, CD36, and CD66b), and some CD markers with higher ABC had also relatively high CVs (CD43 and CD48).

The amount of surface protein (here expressed as ABC) is perhaps the most used measure of protein expression in a cell subset and corresponds most closely to measures of expression in other forms of analysis with bulk cells. However, flow cytometry being a single-cell technique has the advantage of distinguishing individual cells that do or do not express a marker. This can be shown as percentage of positivity, and this has been defined relative to FMO for all measured CD markers in each cell subset (**Supplementary Figure 7** and **Supplementary Table 5**). Ordered visualization of markers with increased positivity revealed sigmoidal curves per cell subset (**Figure 3D** and **Supplementary Figure 7**), separating markers that were negative on all, positive on all, or positive on a fraction of the cells within the subset. The frequencies of positive cells were tightly associated with the fluorescence (shown by coloring), with some exceptions: e.g., low CD9 and high CD48 on naive B-cells (**Figure 3D**).

Clustering of Cell Subsets and CD Markers

To interrogate and visualize common expression patterns of markers and how these related on the defined cell subsets, we performed unsupervised HCA (**Figure 4**). The analysis revealed three main cell clusters: T-cells, B-cells, and myeloid cells. Within both B- and T-cells, the blood and tissue subsets were grouped into two separate subclusters.

Regarding CD marker patterns, CD19, CD20, CD21, CD22, CD72, and CD74 clustered together with predominant expression among B-cell subsets, whereas CD11b, CD11c, CD13, CD14, CD16, CD33, and CD88 were found to be expressed in the myeloid cell cluster (**Supplementary Figure 8**). The thymocyte cluster contained CD9, CD10, CD1a, CD1b, CD1d, CD71, CD69, CD90, and CD34, which are known markers for progenitor cells and for cell activation. A cluster of CD markers expressed on all subsets and at all stages included CD45, CD44, CD99, CD47, and CD50. Lastly, a T-cell cluster was apparent, containing CD2, CD3, CD4, CD5, CD6, CD7, CD8, CD26, CD28, CD49e, CD49f, CD62L, CD84, CD95, and CD96. In addition to these dominant clusters, the heatmap also clearly visualizes expression of CD markers outside of the dominant cluster, such as CD24 expression on neutrophils and eosinophils, and CD21 expression on immature thymocytes (**Figure 4** and **Supplementary Figure 8**).

Granulocyte, Monocyte, and Dendritic Cell Analysis

Three monocyte subsets can be typically defined based on differential expression of CD14 and CD16 (**Supplementary Figure 1**), and these subsets have been shown to be associated with distinct diseases (31, 32). Of the 111 CD markers tested, 31 were significantly different in ABC ($p < 0.01$) between any two of the three subsets (**Figure 5A**). Remarkably, multiple integrins (CD11b, CD49e) and other adhesion molecules (CD33, CD62P), as well as antigen-presentation molecule CD1d were specifically downregulated on non-classical monocytes as compared to the classical and intermediate subsets.

By definition, CD16 (FcγRIII) was upregulated on intermediate and non-classical monocytes. In contrast, CD64 (FcγRI) was specifically downregulated on non-classical monocytes, whereas all subsets expressed relatively similar levels of CD32 (FcγRIIa and FcγRIIb). The CD35 antigen (complement receptor 1) was specifically downregulated on non-classical monocytes. Within the family of tetraspanins, CD63 expression was specifically high on classical monocytes, and CD9 and CD82 expression levels were significantly reduced on non-classical monocytes, whereas no differences were seen for CD37, CD53, and CD81.

Similar to the monocyte subsets, we performed a detailed phenotypic comparison between the major two DC subsets in blood: myeloid (m)DC and plasmacytoid (p)DC. pDCs were defined on the basis of co-expression of HLA-DR and CD123 (**Supplementary Figure 1** and **Supplementary Table 3**). Due to the limitations in markers we could use in the backbone, we defined one mDC population on the basis of HLA-DR+CD11c+CD14–CD16–, which includes both the CD1c+ cDC1 and the CD141+ cDC2 subsets (33). Forty of the 111 CD molecules differed significantly in expression level between mDC and pDC ($p < 0.01$), and of these 19 with a $p < 0.001$ (**Figure 5B**). Most of the differences were the result of higher expression of markers on pDCs. Markers with low expression included molecules typically found on lymphocytes (CD3, CD10, and CD19), and this probably does not represent actual expression. In addition, pDC expressed higher levels of multiple integrins (CD29, CD49a, CD49c, CD49d) and adhesin molecule CD54 (ICAM-1), as well as the previously reported immunoregulatory receptor CD5 and tolerogenic receptors CD85d, CD85j, and CD85k (33), whereas the death receptor CD95 was significantly reduced on pDC (34). Expression levels of the previously reported CD11b, CD11c, and CD13 were reduced, but not with a significance of $p < 0.01$ (34).

Between neutrophils and eosinophils, 20 CD molecules were significantly different ($p < 0.01$) and all were lower on the latter subset (**Figure 5C**). These included the well-described CD10, CD15, and CD16, as well as integrins CD11b, CD11c, CD18; integrin ligand CD50; complement receptors CD35, CD88, and CD93; and the IgA receptor CD89. About half of the significantly different markers between basophils and eosinophils were around borderline expression

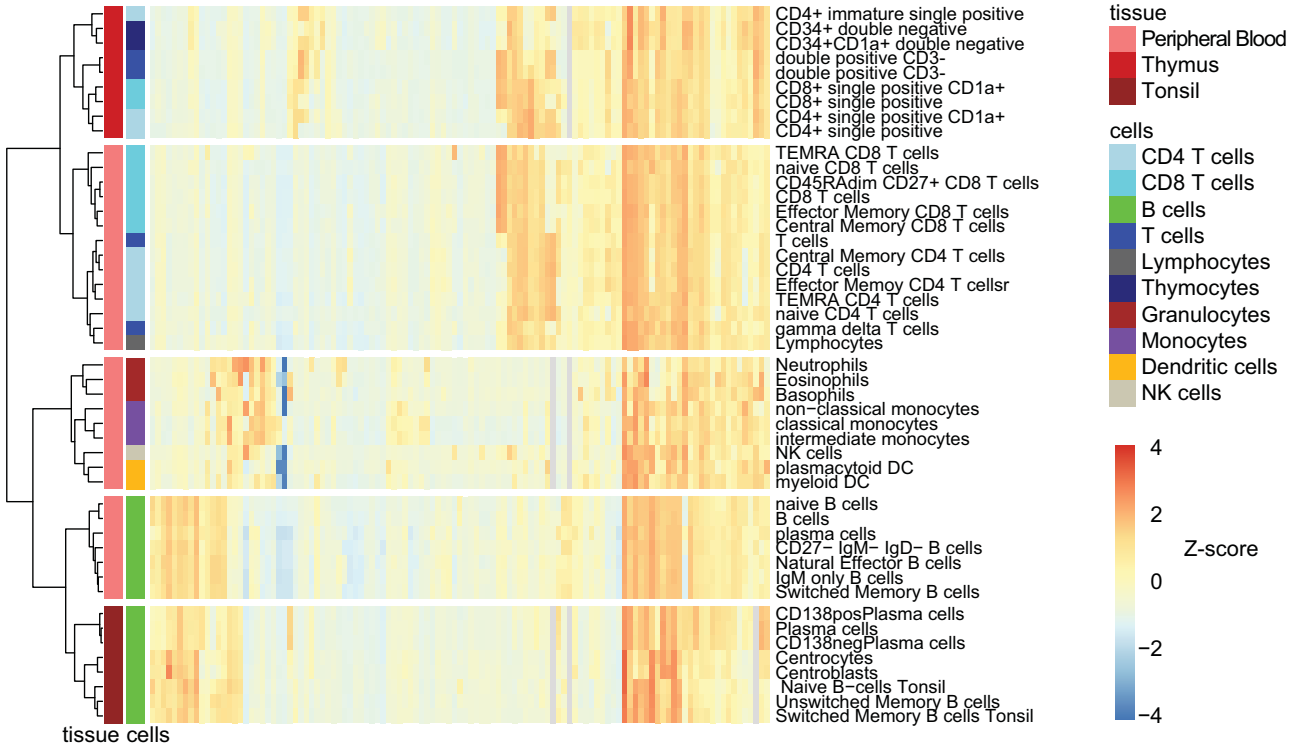

FIGURE 4 | Hierarchical clustering analysis of CD marker expression on leukocyte subsets. Unsupervised clustering was performed on all CD markers ($n = 117$) and all cell subsets ($n = 47$) based on $\log_{10}$ transformed Median ABC. Cell subsets are color coded based on their lineage and their tissue of origin. Hierarchical tree was algorithmically cut into clusters ($k = 5$). Each cell subset group could contain cell subset from one or more tissue type (peripheral blood, thymus, and tonsil). See **Supplementary Figure 8** for larger version of this figure, including CD marker labeling.

(10^3) (**Figure 5D**). Of the rest, 11 were significantly higher in basophils and included the tetraspanins CD9, CD53, and CD82; the FcγRII (CD32); multiple cell adhesion molecules (CD38, CD44, CD54, CD62L); complement decay factor CD55; and SLAM family member CD84. Conversely, eosinophils expressed significantly more CD15, glycoproteins CD22 and CD24, ectoenzyme CD39, TNF receptor CD40, and adhesion molecules CD49f and CD66c.

T-Cell Maturation

Within the CD3+ cells, the three main lineages (TCRγδ+, CD4+, and CD8+) were distinguished (**Supplementary Figure 2**). Pairwise analysis of parallel maturation stages between the CD4 and CD8 lineages for markers with significance of >0.01 and change of at least 10-fold (**Supplementary Figure 9A**) revealed consistently higher CD59 expression on CD4 T-cells (all stages, except for TemRA; CD45RA+CD27−) (35). Conversely, "senescence" marker CD57 and tetraspanin CD63 were both higher on CD8 T-cells in the central memory (Tcm) stage.

In addition, multiple CD markers were differentially expressed between stages of T-cell maturation. Naive CD8 T-cells (CD45RA+CD27+) were nearly all positive for the CD45RA isoform, CD31 (PECAM-1), and costimulatory molecules CD27 and CD28 (**Figure 6**) (36). While the integrins (CD18 and CD11c) were expressed on all T-cell subsets, their degree of expression increased with maturation (**Supplementary Figure 9B**). The relative amount of surface CD45RA was about twice as high as CD3, which in turn was nearly twice that of CD27 (**Supplementary Figure 9B**). The expression levels of regulators of activation were tightly controlled as evidenced by low CV within each subset (CD3, CD45RA, CD28, CD27, and CD31; **Supplementary Figure 9C**). By definition, CD8 Tcm and Tem cells lacked surface CD45RA, and all expressed the CD45RO isoform, generated by alternative splicing. CD95 was expressed on all memory subsets, whereas CD57 was gradually upregulated from Tcm to Tem subsets, which in turn gradually lost CD31. Furthermore, CD28 positivity decreased from Tcm to Tem. Finally, in TemRA, CD45RA was re-expressed with a concomitant loss of CD45RO, and a massive increase in CD57 positivity (37). In our gating strategy, a separate population ($CD45RA^{dim}CD27+$) was defined in-between CD8 Tnaive and Tcm. In contrast to Tnaive, $CD45RA^{dim}$ cells expressed CD95 and CD45RO and lower levels of CD27, and lacked CD38 expression. On the other hand, the $CD45RA^{dim}$ cells were distinct from TemRA, as they did express CD28, and not CD85j. The phenotype of $CD45RA^{dim}$ cells therefore seems to fit with that of antigen experienced T memory stem cell subset as has been suggested before (38, 39). Similar to CD8 T-cells, transition of naive CD4 T-cells to memory was accompanied by

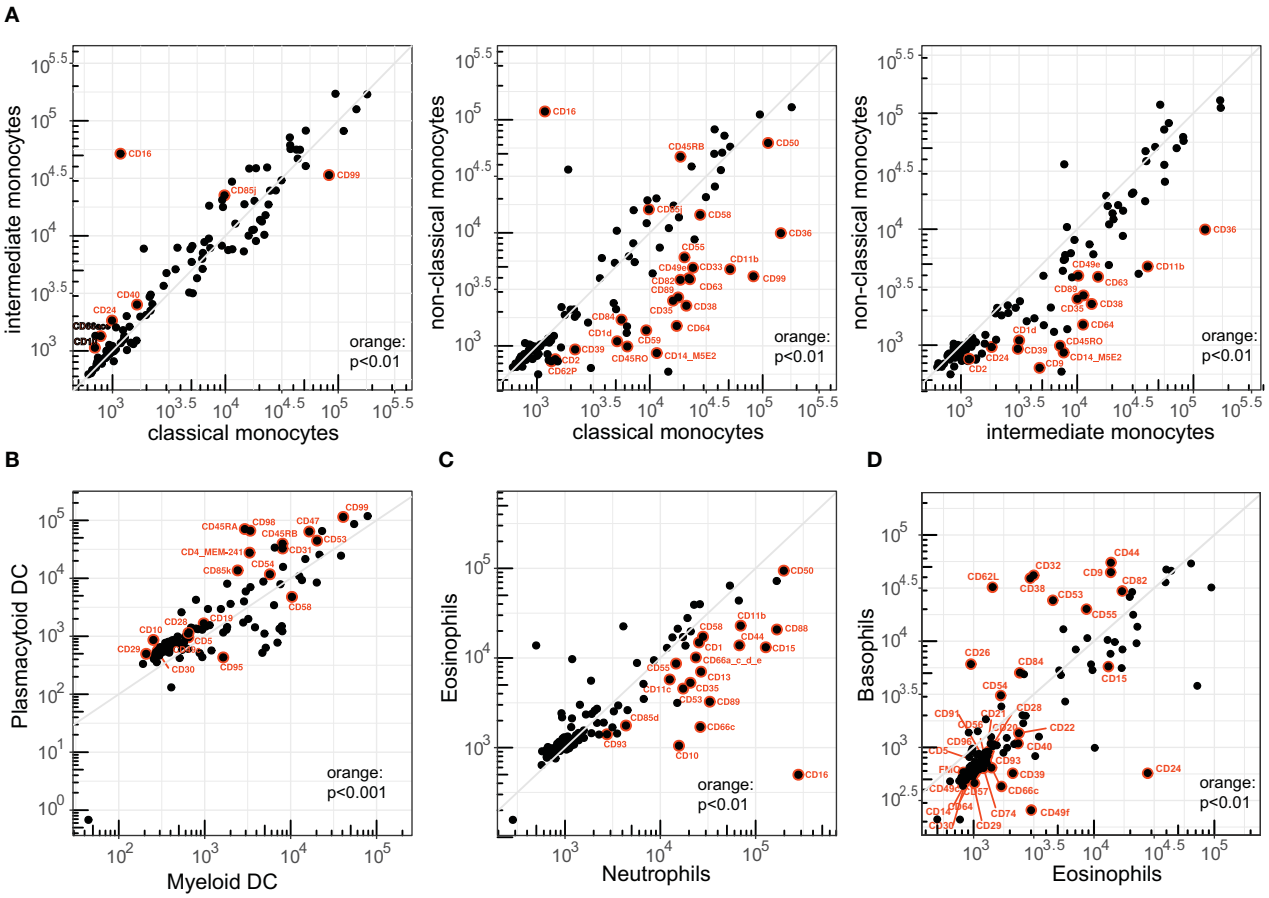

FIGURE 5 | Correlation of CD1–CD100 expression levels between related innate leukocytes. Pairwise analysis of expression levels (ABC) between **(A)** classical, intermediate, and non-classical monocytes; **(B)** myeloid DC and plasmacytoid DC; **(C)** neutrophils and eosinophils; and **(D)** eosinophils vs. basophils. CD41 and CD42b were excluded from the plots in **(B–D)**. Both markers were not expressed on any granulocyte or DC cell type (ABC $< 2 \times 10^2$).

a decrease in expression of CD31, CD38, and CD45RA, while CD45RO, CD95 (Fas-receptor), and CD84 (SLAMF5) were upregulated (data not shown) (40, 41).

Thymocyte Differentiation

In addition to mature T-cells in blood, T-cell progenitors in thymus were examined with a separate tube (**Supplementary Figure 3**) (29, 42), thereby enabling complete mapping of CD marker expression from early T-cell progenitors until effector memory cells (**Figure 7**) with the maturation tool in the web resource (**Figure 1**). This revealed that CD10 is gradually lost as cells differentiate from the double negative (DN) to the double positive (DP) stage, and is completely absent on single positive (SP) CD4+ T-cells. Distinct expression patterns were seen for costimulatory molecules CD27 and CD28. Early progenitors already expressed medium levels of CD28, which increased to a maximum after the DP stage, whereas CD27 was low or absent until the DP stage, reaching its maximum just before thymocytes exit to periphery at the CD1a-SP CD4 stage. All thymocytes expressed CD31, which was gradually lost on peripheral naive CD4 T-cells. CD11a was expressed on all stages of T-cell differentiation, with varying degrees of intensity, and a peak on effector memory T-cells.

Antigen-Dependent B-Cell Maturation in Tonsil

Within the total HCA of CD1–CD100 on all cell subsets (**Figure 4** and **Supplementary Figure 8**), the tonsil B-cell subsets were clustered together, and within this cluster, three subclusters were formed containing the three major functional compartments: (i) B-lymphocytes, including naive and unswitched and switched memory B-cells; (ii) germinal center (GC) cells, including centrocytes (CC) and centroblasts (CB); and (iii) plasma cells (PC), including CD138– and CD138+ PC. Over 30 CD markers showed statistically significant differences ($p < 0.01$) between any two of these three major subsets, and a $p < 0.001$ was observed for >20 CD markers. Populations within each of the three these subgroups were very homogeneous based with <5 CD markers significantly different ($p < 0.01$) between them.

PC and B-lymphocyte groups were most different with in CD marker expression ($p < 0.01$, 37 CD markers; $p < 0.001$, 27 CD markers). Those differences with a $p < 0.001$ included

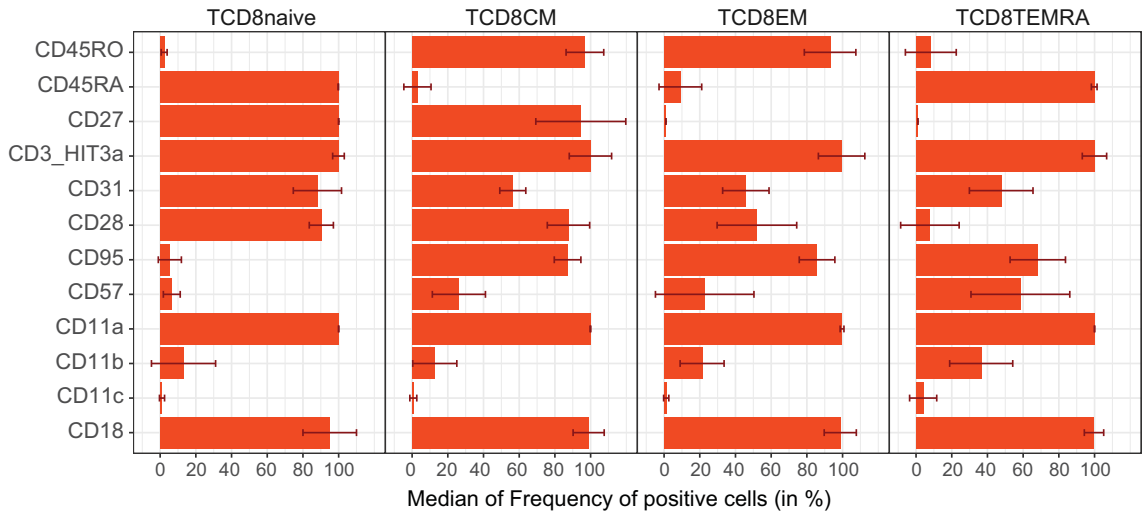

FIGURE 6 | Expression patterns of selected CD markers on naïve and memory CD8 T-cells. Selected CD molecules are shown for naive, central memory (CM), effector memory (EM), and CD45RA+ effector memory (TEMRA) CD8 T-cells. In contrast to co-stimulatory molecules (CD27, CD28) and FasR (CD95), the integrins (CD11a, CD11b, CD11c, CD18) are expressed on all stages of CD8 T-cells; however, their expression levels differ (see **Supplementary Figure 8**).

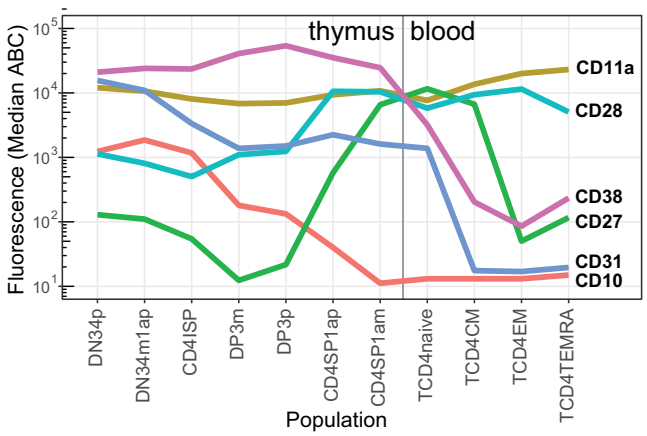

FIGURE 7 | Median expression levels of critical CD markers during thymocyte differentiation into peripheral CD4 T-cells. Expression levels (in ABC) of selected CD markers ($n = 6$) on thymocyte (red labels) and blood (pink labels) CD4 T-cell subsets ($n = 11$).

upregulation of a large set of adhesion and signaling molecules (CD18, CD31, CD54, CD97, CD98, and CD99) together with a different profile of expression of activation/signaling markers (CD9, CD24, CD27, CD28, CD37, CD39, CD43, CD44, CD45RA, CD52, CD53, CD63, CD79b, and CD81) and complement receptor proteins (CD35, CD46, CD55, and CD59) (28, 43). Visualization with the maturation tool from the CD Maps web resource (**Figure 1**) showed that some of these phenotypic features of an antibody-secreting cell signature were already acquired in the GC compartment (**Figure 8**). These phenotypic changes included upregulation of molecules involved in adhesion/migration (CD54, CD98) and enzymatic activity (CD10; pattern 3); changes in cell activation/signaling (CD24, CD44) and complement receptors (CD35, CD59; pattern 6), as compared to B-lymphocytes. PC and GC groups differed in 20 CD markers ($p < 0.001$), including those that were already upregulated during the GC phase (CD54, CD59, and CD98; pattern 3), reversion of phenotypic changes observed during GC reaction (CD20, CD31, CD32, CD40, CD47, and CD55; patterns 1, 2, and 4), and upregulation of markers that were absent on both B-lymphocytes and GC cells (CD9, CD28, CD43, CD63, and CD97). Finally, some markers were upregulated (CD46 and CD99; pattern 2) or decreased as compared to both B-lymphocytes and GC (CD37, CD45RA, and CD52; pattern 5).

DISCUSSION

We here examined 111 CD markers on 47 leukocyte subsets using multicolor flow cytometry with the marker of interest in the PE channel. The resulting expression profile is the largest quantitative dataset of surface protein expression levels on human immune cells.

The examined surface proteins represent those that were defined clustered mAbs in HLDA workshops I–V that were held in the 1980s and early 1990s (20–24). At that time, the protein expression patterns were defined in great detail. However, with advances in technologies and new insights into immune cell function and subsets, we deemed the expression data incomplete, not fully accurate, and lacking quantitative information. Indeed, when we compare our data with a CD chart of a major antibody vendor, we could find over 50 discrepancies and 25 missing values. In part, those discrepancies stem from a positivity and negativity definition on a broadly defined cell lineage: any positivity found at any stage and/or activation status is regarded as positivity on such chart. Our detailed analysis on well-defined subsets potentially clarifies this.

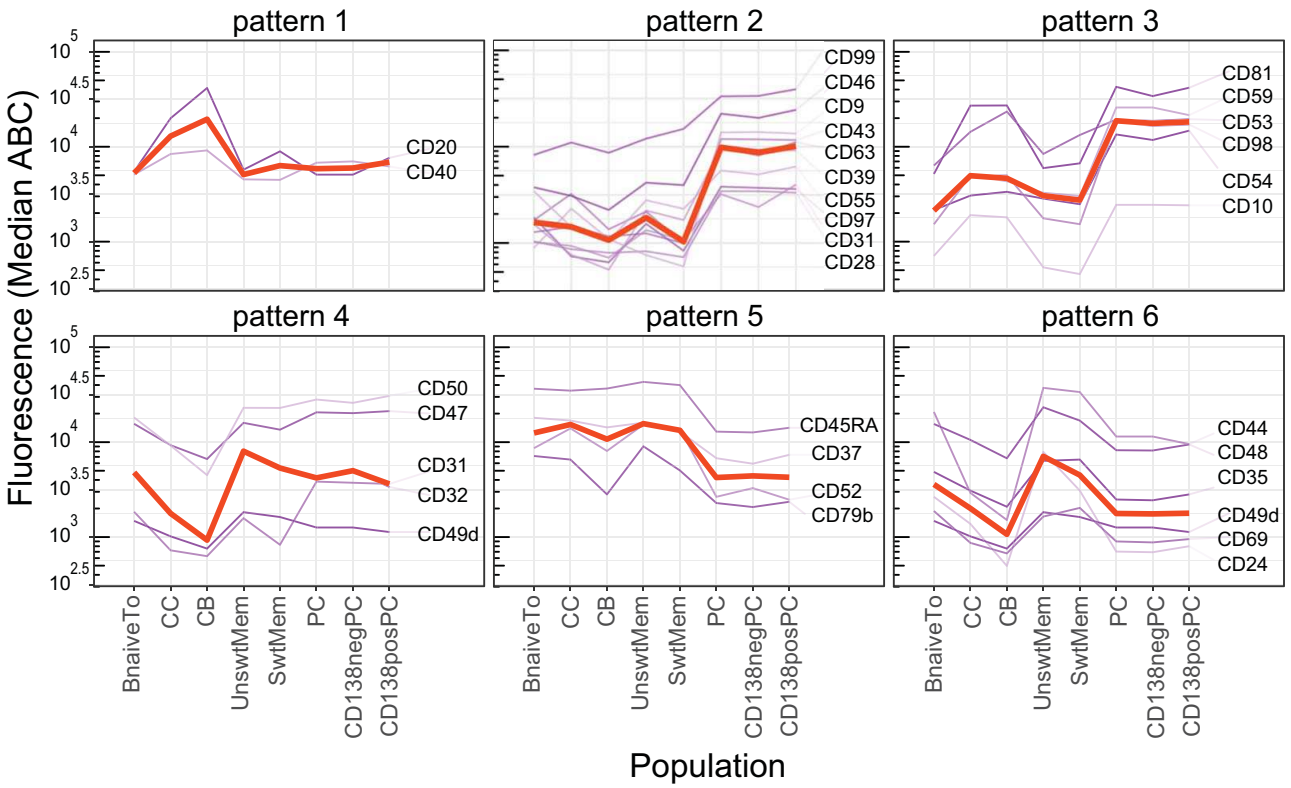

FIGURE 8 | Expression pattern of selected CD markers during B-cell maturation in tonsil. Expression patterns (median ABC) of CD markers showing statistically significant increased **(Upper)** or decreased **(Lower)** level of expression on germinal center (GC) cells (patterns 1 and 4), plasma cells (PC) (patterns 2 and 5), and both GC + PC simultaneously (patterns 3 and 6), as compared to B-lymphocytes (naive and memory B-cell subsets). Red lines indicate median expression levels of the depicted markers in that panel.

To ensure robustness and reproducibility of our data, we standardized our experimental procedures and flow cytometer setup according to the protocols that were established for clinical use by the EuroFlow consortium (www.EuroFlow.org) (14). Subsequently, the measurements were independently performed in three to four laboratories, each acquiring data from three to four donors with parallel acquisition of PE signal calibration particles. Indeed, gating of subsets using the backbone markers could be reliably performed on the data, irrespective of their origin. There are limitations in the interpretation of the signal near the background (a combination of autofluorescence, spillover spread, non-specific antibody binding, and antibody titer) that resulted in a "gray zone" at 200–700 ABC units in lymphocytes and 1,000–10,000 ABC units in myeloid cells that has to be evaluated by a more sensitive approach in future studies.

Thus, we have obtained a realistic dataset, which can be prepared reproducibly in any laboratory following the same operating procedure. Although we do not claim we have covered population variation with only 12 donors per CD marker, by displaying up to 12 donors using median values, outliers caused by, e.g., rare genetic polymorphisms (CD45 isoforms or CD39) or by accidental activation (CD69) would not overtly affect the results (44–46). Accurate quantification of CD marker expression levels is not only important for biological function, but can be utilized as well for a proper design of flow cytometry experiments, where also intensity of expression is essential information for a successful multicolor panel (47).

The unique feature of our data resource is the detailed information in expression levels and changes between diverse immune cell subsets, thus allowing interpretation of quantitative changes during thymocyte development, B-cell maturation in the tonsil, and between blood cell subsets that might share expression of the same marker but with different quantities.

In the present study, we quantitatively mapped the expression of 111 surface-expressed proteins on 41 non-overlapping leukocyte subsets from three human tissues. With this being a large-scale analysis and a systems approach, a few concessions had to be made in experimental design. Accuracy of exact quantification of CD marker expression is potentially skewed by the antibody binding occurring through either one or two Fab domains (48). Thus, the ABC unit that was used to quantitatively depict expression has an error margin of a factor 2 for the number of expressed molecules. Still, our measurements for CD4 yielded a median of 38,650 ABC (clone MEM-241) for naive CD4 T-cells, which was very similar to the previously published value of 42,000 ABC (clone SK3) (49). Finally, for this large-scale approach, we only could use one antibody reagent for each given CD marker. Selection criteria for these reagents included (1) being

a clone that was approved in the HLDA workshops and (2) good reactivity based on our in-house experience. Our pilot tests for two clones for CD4 (MEM-241 and RPA-T4) and CD8 (MEM-31 and HIT8a) showed differences of up to 20% in expression levels. As the clones we tested have been through the HLDA workshops, these will serve as a benchmark that can either be matched or can be surpassed by alternative reagents. The resource we have built will be appended in the future with new clones, new reagents, new CD markers, and new cell subsets. In the upcoming 11th HLDA workshop, this methodological framework will be used to measure and cluster antibody reactivities across subsets to help assign new CD nomenclature. This approach follows the strategy proposed by the International Working Group for Antibody Validation (IWGAV) that has documented expression patterns for 3,706 antibodies in immunoprecipitates (50, 51). Including future reactivity patterns of HLDA 11 in the CD Maps resource will enhance its role as a benchmark for the research community.

Regarding the immunobiology, we did not exhaustively define all functionally defined immune cell subsets. With four tubes using seven channels for the backbone each, we were able to define 41 unique, non-overlapping subsets. Several cell types were not included, such as helper T-cell subsets, regulatory T-cells, NK T-cells, and mucosa-associated invariant T-cells (MAIT). With an extended panel using more fluorescent markers, such limitation can be overcome in future studies. However, rare cell populations such as innate lymphoid cells will remain a challenge as this would require the acquisition of more than a million events per staining.

In conclusion, we have demonstrated the possibility to systematically quantify the expression of surface-expressed proteins on the multitude of immune cells using standardized multicolor flow cytometry. There is a need for this standardized systems approach to avoid confusion from separate observations in individual laboratories, to correct potential mistakes in the literature, and to predict potential off-target effects of antibody-based therapies. The CD Maps web resource enables each user to explore the data and it has the capacity to function as a platform for surface molecule expression data that can be updated with newer CD markers and more leukocyte subsets. With the ongoing activities of the HLDA workshops, the CD Maps project can provide the means to get toward a full picture of the surfaceome of human immune cells.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the CD Maps web resource: http://www.hcdm.org.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the use of human pediatric tissue and adult buffy coats was approved by the Human Ethics Committees of the Erasmus Medical Center, the University Hospital Motol, and the universities of Salamanca and Barcelona. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

TK, MP-A, PE, and MZ conceptualized the study and designed the experiments. DK, MC, SB, and EB performed the experimental work. KF integrated all data, performed bioinformatics analysis, and built the online web resource. TK, KF, MP-A, PE, and MZ wrote the paper. All authors performed data analysis, commented on draft versions and approved the final manuscript.

FUNDING

The work was financially supported by the International Union of Immunological Societies (IUIS), projects 15-26588A (TK) and NV18-08-00385 (KF) of the Czech Republic Ministry of Health, Australian National Health and Medical Research Council (NHMRC) Fellowship GNT1117687 (MZ), and SAF2015-69829 from Ministerio Ciencia e Innovación (Spain to PE). Publishing costs for this article were partially covered by the International Union of Immunological Societies (IUIS).

ACKNOWLEDGMENTS

The authors are grateful to Robert Balderas (BD Biosciences), Kelly Lundsten (BioLegend), and Miloslav Suchanek (Exbio Praha, a.s.) for providing reagents, and to the Czech National Grid Infrastructure MetaCentrum under the program Projects of Large Research, Development, and Innovations Infrastructures (CESNET LM2015042) for access to computing and storage facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02434/full#supplementary-material

REFERENCES

1. Fagerberg L, Jonasson K, von Heijne G, Uhlen M, Berglund L. Prediction of the human membrane proteome. *Proteomics*. (2010) 10:1141–9. doi: 10.1002/pmic.200900258
2. Bausch-Fluck D, Goldmann U, Muller S, van Oostrum M, Muller M, Schubert, OT, et al. The *in silico* human surfaceome. *Proc Natl Acad Sci USA*. (2018) 115:E10988–97. doi: 10.1073/pnas.1808790115
3. Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, et al. A mass spectrometric-derived cell surface protein atlas. *PLoS ONE*. (2015) 10:e0121314. doi: 10.1371/journal.pone.0121314

4. Diaz-Ramos MC, Engel P, Bastos R. Towards a comprehensive human cell-surface immunome database. *Immunol Lett*. (2011) 134:183–7. doi: 10.1016/j.imlet.2010.09.016
5. Kohler, G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*. (1975) 256:495–7. doi: 10.1038/256495a0
6. Julius MH, Masuda T, Herzenberg LA. Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter. *Proc Natl Acad Sci USA*. (1972) 69:1934–8. doi: 10.1073/pnas.69.7.1934
7. Clark G, Stockinger H, Balderas R, van Zelm MC, Zola H, Hart D, et al. Nomenclature of CD molecules from the Tenth Human Leucocyte Differentiation Antigen Workshop. *Clin Transl Immunol*. (2016) 5:e57. doi: 10.1038/cti.2015.38
8. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD nomenclature 2015: human leukocyte differentiation antigen workshops as a driving force in immunology. *J Immunol*. (2015) 195:4555–63. doi: 10.4049/jimmunol.1502033
9. Iwai Y, Hamanishi J, Chamoto K, Honjo T. Cancer immunotherapies targeting the PD-1 signaling pathway. *J Biomed Sci*. (2017) 24:26. doi: 10.1186/s12929-017-0329-9
10. Rosman Z, Shoenfeld Y, Zandman-Goddard G. Biologic therapy for autoimmune diseases: an update. *BMC Med*. (2013) 11:88. doi: 10.1186/1741-7015-11-88
11. Horvath C, Andrews L, Baumann A, Black L, Blanset D, Cavagnaro J, et al. Storm forecasting: additional lessons from the CD28 superagonist TGN1412 trial. *Nat Rev Immunol*. (2012) 12:740; author reply 740. doi: 10.1038/nri3192-c1
12. Barnkob MS, Simon C, Olsen LR. Characterizing the human hematopoietic CDome. *Front Genet*. (2014) 5:331. doi: 10.3389/fgene.2014.00331
13. Andersen MN, Al-Karradi SN, Kragstrup TW, Hokland M. Elimination of erroneous results in flow cytometry caused by antibody binding to Fc receptors on human monocytes and macrophages. *Cytometry A*. (2016) 89:1001–9. doi: 10.1002/cyto.a.22995
14. Kalina T, Flores-Montero J, van der Velden VH, Martin-Ayuso M, Bottcher S, Ritgen M, et al. EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. *Leukemia*. (2012) 26:1986–2010. doi: 10.1038/leu.2012.122
15. Kalina T, Flores-Montero J, Lecrevisse Q, Pedreira CE, van der Velden VH, Novakova M, et al. Quality assessment program for EuroFlow protocols: summary results of four-year (2010-2013) quality assurance rounds. *Cytometry A*. (2015) 87:145–56. doi: 10.1002/cyto.a.22581
16. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014).
17. Caglar MU, Teufel AI, Wilke CO. Sicegar: R package for sigmoidal and double-sigmoidal curve fitting. *PeerJ*. (2018) 6:e4251. doi: 10.7717/peerj.4251
18. Christopoulos DT. A practical guide for using Unit Invariant Knee (UIK) as elbow point estimation in multivariate analysis. *ResearchGate*. (2014). doi: 10.13140/RG.2.2.10201.31846
19. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J Classif*. (2014) 31:274–95. doi: 10.1007/s00357-014-9161-z
20. Bernard A, Dausset J, Milstein C, Schlossmann SF, editors. *Leucocyte Typing*: *Human Leucocyte Differentiation Antigens Detected by Monoclonal Antibodies*. New York, NY: Springer-Verlag (1984).
21. Reinherz EL, Haynes BF, Nadler LM, Bernstein ID, editors. *Leukocyte Typing II*. New York, NY: Springer-Verlag (1986). doi: 10.1007/978-1-4612-4850-7
22. McMichael AJ, Beverley PCL, Cobbold S, Crumpton MJ, Gilks W, Gotch FM, et al. editors. *Leucocyte Typing III*: *White Cell Differentiation Antigens*. Oxford: Oxford University Press (1987).
23. Knapp W, Dörken B, Gilks WR, Rieber EP, Schimdt RE, Stein H, et al. editors. *Leucocyte Typing IV*: *White Cell Differentiation Antigens*. Oxford: Oxford University Press (1989).
24. Schlossman SF, Boumsell L, Gilks W, Harlan JM, Kishimoto T, Morimoto C, et al. editors. *Leucocyte Typing V*: *White Cell Differentiation Antigens*. Oxford: Oxford University Press (1995).
25. Clark GJ, Kupresanin F, Fromm PD, Ju X, Muusers L, Silveira PA, et al. New insights into the phenotype of human dendritic cell populations. *Clin Transl Immunol*. (2016) 5:e61. doi: 10.1038/cti.2015.40
26. van der Burg M, Kalina T, Perez-Andres M, Vlkova M, Lopez-Granados E, Blanco E, et al. The EuroFlow PID orientation tube for flow cytometric diagnostic screening of primary immunodeficiencies of the lymphoid system. *Front Immunol*. (2019) 10:246. doi: 10.3389/fimmu.2019.00246
27. van Zelm MC, Szczepanski T, van der Burg M, van Dongen JJ. Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion. *J Exp Med*. (2007) 204:645–55. doi: 10.1084/jem.20060964
28. Perez-Andres M, Paiva B, Nieto WG, Caraux A, Schmitz A, Almeida J, et al. Human peripheral blood B-cell compartments: a crossroad in B-cell traffic. *Cytometry B Clin Cytom*. (2010) 78 (Suppl 1):S47–60. doi: 10.1002/cyto.b.20547
29. Stepanek O, Kalina T, Draber P, Skopcova T, Svojgr K, Angelisova P, et al. Regulation of Src family kinases involved in T cell receptor signaling by protein-tyrosine phosphatase CD148. *J Biol Chem*. (2011) 286:22101–12. doi: 10.1074/jbc.M110.196733
30. Dik WA, Pike-Overzet K, Weerkamp F, de Ridder D, de Haas EF, Baert MR, et al. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J Exp Med*. (2005) 201:1715–23. doi: 10.1084/jem.20042524
31. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood*. (2010) 116:e74–80. doi: 10.1182/blood-2010-02-258558
32. Ziegler-Heitbrock L, Hofer TP. Toward a refined definition of monocyte subsets. *Front Immunol*. (2013) 4:23. doi: 10.3389/fimmu.2013.00023
33. Clark GJ, Silveira PA, Hogarth PM, Hart DNJ. The cell surface phenotype of human dendritic cells. *Semin Cell Dev Biol*. (2018) 86:3–14. doi: 10.1016/j.semcdb.2018.02.013
34. Piccioli D, Tavarini S, Borgogni E, Steri V, Nuti S, Sammicheli C, et al. Functional specialization of human circulating CD16 and CD1c myeloid dendritic-cell subsets. *Blood*. (2007) 109:5371–9. doi: 10.1182/blood-2006-08-038422
35. Hernandez-Campo PM, Almeida J, Sanchez ML, Malvezzi M, Orfao A. Normal patterns of expression of glycosylphosphatidylinositol-anchored proteins on different subsets of peripheral blood cells: a frame of reference for the diagnosis of paroxysmal nocturnal hemoglobinuria. *Cytometry B Clin Cytom*. (2006) 70:71–81. doi: 10.1002/cyto.b.20087
36. Larbi A, Fulop T. From "truly naive" to "exhausted senescent" T cells: when markers predict functionality. *Cytometry A*. (2014) 85:25–35. doi: 10.1002/cyto.a.22351
37. Kared H, Martelli S, Ng TP, Pender SL, Larbi A. CD57 in human natural killer cells and T-lymphocytes. *Cancer Immunol Immunother*. (2016) 65:441–52. doi: 10.1007/s00262-016-1803-z
38. Lugli E, Dominguez MH, Gattinoni L, Chattopadhyay PK, Bolton DL, Song K, et al. Superior T memory stem cell persistence supports long-lived T cell memory. *J Clin Invest*. (2013) 123:594–9. doi: 10.1172/JCI66327
39. Mahnke YD, Brodie TM, Sallusto F, Roederer M, Lugli E. The who's who of T-cell differentiation: human memory T-cell subsets. *Eur J Immunol*. (2013) 43:2797–809. doi: 10.1002/eji.201343751
40. Hamann D, Baars PA, Rep MH, Hooibrink B, Kerkhof-Garde SR, Klein MR, et al. Phenotypic and functional separation of memory and effector human CD8+ T cells. *J Exp Med*. (1997) 186:1407–18. doi: 10.1084/jem.186.9.1407
41. Romero X, Benitez D, March S, Vilella R, Miralpeix M, Engel P. Differential expression of SAP and EAT-2-binding leukocyte cell-surface molecules CD84, CD150 (SLAM), CD229 (Ly9) and CD244 (2B4). *Tissue Antigens*. (2004) 64:132–44. doi: 10.1111/j.1399-0039.2004.00247.x
42. Res P, Blom B, Hori T, Weijer, K, Spits H. Downregulation of CD1 marks acquisition of functional maturation of human thymocytes and defines a control point in late stages of human T cell development. *J Exp Med*. (1997) 185:141–51. doi: 10.1084/jem.185.1.141
43. Berkowska MA, Schickel JN, Grosserichter-Wagener C, de Ridder D, Ng YS, van Dongen JJ, et al. Circulating human CD27-IgA+ memory B cells recognize bacteria with polyreactive Igs. *J Immunol*. (2015) 195:1417–26. doi: 10.4049/jimmunol.1402708
44. Jacobsen M, Schweer D, Ziegler A, Gaber R, Schock S, Schwinzer R, et al. A point mutation in PTPRC is associated with the development of multiple sclerosis. *Nat Genet*. (2000) 26:495–9. doi: 10.1038/82659

45. Lopez-Cabrera M, Santis AG, Fernandez-Ruiz E, Blacher R, Esch F, Sanchez-Mateos P, et al. Molecular cloning, expression, and chromosomal localization of the human earliest lymphocyte activation antigen AIM/CD69, a new member of the C-type animal lectin superfamily of signal-transmitting receptors. *J Exp Med*.(1993) 178:537–47. doi: 10.1084/jem.178.2.537
46. Roederer M, Quaye L, Mangino M, Beddall MH, Mahnke Y, Chattopadhyay P, et al. The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell*. (2015) 161:387–403. doi: 10.1016/j.cell.2015.02.046
47. Cossarizza A, Chang HD, Radbruch A, Akdis M, Andra I, Annunziato F, et al. Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur J Immunol*. (2017) 47:1584–797. doi: 10.1002/eji.201646632
48. Davis KA, Abrams B, Iyer SB, Hoffman RA, Bishop JE. Determination of CD4 antigen density on cells: role of antibody valency, avidity, clones, and conjugation. *Cytometry*. (1998) 33:197–205. doi: 10.1002/(SICI)1097-0320(19981001)33:2<197::AID-CYTO14>3.0.CO;2-P
49. Wang L, Abbasi F, Gaigalas AK, Hoffman RA, Flagler, D, Marti GE. Discrepancy in measuring CD4 expression on T-lymphocytes using fluorescein conjugates in comparison with unimolar CD4-phycoerythrin conjugates. *Cytometry B Clin Cytom*. (2007) 72:442–9. doi: 10.1002/cyto.b.20354
50. Sikorski K, Mehta A, Inngjerdingen M, Thakor F, Kling S, Kalina T, et al. A high-throughput pipeline for validation of antibodies. *Nat Methods*. (2018) 15:909–12. doi: 10.1038/s41592-018-0179-8
51. Uhlen M, Bandrowski A, Carr S, Edwards A, Ellenberg J, Lundberg E, et al. A proposal for validation of antibodies. *Nat Methods*. (2016) 13:823–7. doi: 10.1038/nmeth.3995

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 12 November 2019
doi: 10.3389/fimmu.2019.02541

Standardized IMGT® Nomenclature of Salmonidae IGH Genes, the Paradigm of Atlantic Salmon and Rainbow Trout: From Genomics to Repertoires

Susana Magadan[1,2], Aleksei Krasnov[3], Saida Hadi-Saljoqi[4], Sergey Afanasyev[5], Stanislas Mondot[6], Delphine Lallias[7], Rosario Castro[8], Irene Salinas[2], Oriol Sunyer[9], John Hansen[10], Ben F. Koop[11], Marie-Paule Lefranc[4] and Pierre Boudinot[8]*

[1] Immunology Laboratory, Biomedical Research Center, University of Vigo, Vigo, Spain, [2] Department of Biology, Center of Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, NM, United States, [3] Nofima AS, Norwegian Institute of Food, Fisheries and Aquaculture Research, Tromsø, Norway, [4] IMGT®, The International ImMunoGeneTics Information System® (IMGT), Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), CNRS, University of Montpellier, Montpellier, France, [5] Sechenov Institute of Evolutionary Physiology and Biochemistry, Saint Petersburg, Russia, [6] MICALIS, Institut National de la Recherche Agronomique, Université Paris-Saclay, Jouy-en-Josas, France, [7] Génétique Animale et Biologie Intégrative (GABI), INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, [8] Virologie et Immunologie Moléculaires (VIM), Institut National de la Recherche Agronomique (INRA), Université Paris-Saclay, Jouy-en-Josas, France, [9] Pathobiology Department, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States, [10] Western Fisheries Research Center, U.S. Geological Survey, Seattle, WA, United States, [11] Department of Biology, University of Victoria, Victoria, BC, Canada

OPEN ACCESS

Edited by:
Andrew M. Collins, University of New South Wales, Australia

Reviewed by:
Peter Daniel Burrows, University of Alabama at Birmingham, United States
Michael Zemlin, Saarland University Hospital, Germany

***Correspondence:**
Susana Magadan
smaga@uvigo.es

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *30 August 2019*
Accepted: *14 October 2019*
Published: *12 November 2019*

Citation:
Magadan S, Krasnov A, Hadi-Saljoqi S, Afanasyev S, Mondot S, Lallias D, Castro R, Salinas I, Sunyer O, Hansen J, Koop BF, Lefranc M-P and Boudinot P (2019) Standardized IMGT® Nomenclature of Salmonidae IGH Genes, the Paradigm of Atlantic Salmon and Rainbow Trout: From Genomics to Repertoires. Front. Immunol. 10:2541. doi: 10.3389/fimmu.2019.02541

In teleost fish as in mammals, humoral adaptive immunity is based on B lymphocytes expressing highly diverse immunoglobulins (IG). During B cell differentiation, IG loci are subjected to genomic rearrangements of V, D, and J genes, producing a unique antigen receptor expressed on the surface of each lymphocyte. During the course of an immune response to infections or immunizations, B cell clones specific of epitopes from the immunogen are expanded and activated, leading to production of specific antibodies. Among teleost fish, salmonids comprise key species for aquaculture. Rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*) are especially important from a commercial point of view and have emerged as critical models for fish immunology. The growing interest to capture accurate and comprehensive antibody responses against common pathogens and vaccines has resulted in recent efforts to sequence the IG repertoire in these species. In this context, a unified and standardized nomenclature of salmonid IG heavy chain (IGH) genes is urgently required, to improve accuracy of annotation of adaptive immune receptor repertoire dataset generated by high-throughput sequencing (AIRRseq) and facilitate comparisons between studies and species. Interestingly, the assembly of salmonids IGH genomic sequences is challenging due to the presence of two large size duplicated IGH loci and high numbers of IG genes and pseudogenes. We used data available for Atlantic salmon to establish an IMGT standardized nomenclature of IGH genes in this species and then applied the IMGT rules to the rainbow trout IGH loci to set up a nomenclature, which takes into account

the specificities of Salmonid loci. This unique, consistent nomenclature for Salmonid IGH genes was then used to construct IMGT sequence reference directories allowing accurate annotation of AIRRseq data. The complex issues raised by the genetic diversity of salmon and trout strains are discussed in the context of IG repertoire annotation.

Keywords: immunoglobulin, antibody repertoire, salmonid fish, VDJ annotation, comparative immunology

INTRODUCTION

Vertebrate species with jaws (*Gnasthostomata*) that appeared more than 400 million years ago are all characterized by an adaptive immune system based on B and T cells along with the huge diversity and specificity of their antigen receptors, the immunoglobulins (IG) or antibodies and the T cell receptors (TR), respectively (1, 2). The analysis of the germline IGH locus defines the genomic repertoire with the identification of the functional variable (V), diversity (D), and joining (J) genes that participate in the synthesis of VH domains. It also allows the identification of the functional constant (C) genes that encode the constant regions of the heavy chains and define their isotypes (3–7).

In teleost fish, B cell clonal responses are induced by infection or immunization, as described in humans or mice. Antibodies constitute a key factor for fish specific immunity and for the protection afforded by vaccines. As key species in aquaculture, Salmonids (family Salmonidae) including rainbow trout (*Oncorhynchus mykiss; Oncmyk*) and Atlantic salmon (*Salmo salar; Salsal*) constitute important models for the study of antibodies and B cell responses in fish.

Several groups started to clone and sequence IGH cDNA from rainbow trout in the early 1990s (8–12). Comparison of VH domains (V-D-J-REGION) expressed in trout stocks from Sweden, France, and the US revealed differences in IGHV subgroup usage: subgroups named 8, 9, 10, and 11 were found only in Swedish stocks while subgroups 4 and 7 were only found in French stocks and subgroup 5 (now part of IGHV1) was found in Swedish, French, and US stocks. These observations suggested genetic differences between the IGHV gene germline repertoires of different populations, but this was not fully clear due to the very small numbers of sampled individuals. In 1996, expressed VH domain sequences were classified into a set of 11 IGHV subgroups, defining a first unified nomenclature for rainbow trout (13). A more extensive study performed in 2006 on American trout by the group of Steve Kaattari found all these subgroups expressed, indicating that IGHV subgroups may have a wider distribution than previously suggested. Two additional subgroups expressed at low frequency were also discovered in this survey (14), leading to a repertoire of 13 IGHV subgroups. These subgroups were used for an IMGT gene table created in 2009, with a provisional gene nomenclature (letter S) for rainbow trout IGHV [path to access: IMGT Repertoire (IG and TR) >1. Locus and genes > Gene tables > IGHV > Rainbow trout (*O. mykiss*)][1].

In Atlantic salmon, Solem et al. described in 2001 nine IGHV subgroups (15), seven of which corresponded to IGHV subgroups defined in rainbow trout (1, 2, 3, 6, 8, 9, and 11). Southern blot experiments suggested that the number of genes per subgroup could vary between 1 and 7 ± 10. This work also clearly established that Atlantic salmon IGHV genes were rearranged and transcribed from both of the two Atlantic salmon IGH loci (IGH locus A on chromosome 6 and IGH locus B on chromosome 3), which were most likely produced by the salmonid whole genome duplication. These data actually suggested that genes from some subgroups could be expressed only from a single locus, while genes from other subgroups were expressed from both A and B loci. This analysis was later extended and refined in 2010 by Yasuike et al. from a complete assembly of the Atlantic salmon IGH A and B loci based on sequences of 24 bacterial artificial chromosomes (BAC) (16). This study provided a first map of the organization of the duplicated IGH loci of a salmonid species. Ninety-nine IGHV genes were found in locus A, and 103 in locus B; 23 IGHV genes are functional in locus A, and 32 in locus B. Using the IMGT threshold of 75% identity for the V-REGION, 18 IGHV subgroups were defined in this work (16). Subgroups that did fit with the IGHV subgroups established in rainbow trout were given a subgroup number consistent with the online 2009 IMGT gene table [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHV > Atlantic salmon (*Salmo salar*)][1].

As new genome assemblies of Atlantic salmon and rainbow trout have been recently made available, we decided to annotate the IGH locus of these species and to establish a common nomenclature of IGH genes based on IMGT rules. We used data previously published for Atlantic salmon (16) to develop a prototype for the Salmonid IMGT standardized nomenclature. We also applied the IMGT rules to the rainbow trout IGH loci as a novel example of IMGT genomic annotation. The objective was to take into account the specificities of the Salmonid loci and to develop a unique, consistent nomenclature, while respecting the IMGT Scientific chart rules and standards. These standards are based on the concepts of identification (keywords), classification (gene and allele nomenclature), description (labels), and numbering (IMGT unique numbering and IMGT Collier de Perles) (3). It is important to note that a consistent nomenclature is crucial to build IMGT reference directory sets that are constituted by the V-REGION, D-REGION, and J-REGION of each IMGT reference allele from IMGT/LIGM-DB (same accession numbers as GenBank, ENA, and DDBJ) (17). These reference directory sets are the fundamental basis for annotation of repertoire datasets produced by high-throughput AIRRseq approaches for the analysis of expressed repertoires,

[1]http://www.imgt.org

in particular to define expressed clonotypes (18–20). The IMGT reference directories are built following the classification of the V, D, J, and C genes and alleles according to the IMGT rules and the assignment of the IMGT functionality: functional (F), open reading frame (ORF), or pseudogene (P) (IMGT Scientific chart > IMGT functionality)[1] (3). These rules ensure that the nomenclature is consistent within and between species, and can be updated when more sequence data become available. Reference directory sets are used by IMGT/V-QUEST and IMGT/JunctionAnalysis (21, 22) for detailed analysis of nucleotide (nt) sequences of V domains [V-(D)-J-REGION]; by IMGT/DomainGapAlign, which provides alignments of amino acid (AA) sequences with the closest V and J regions for V domains and the closest C exons for C domains (23); by IMGT Collier de Perles based on the IMGT unique numbering for V and C domains (24, 25); and by IMGT/HighV-QUEST (26, 27) for high-throughput sequence analysis of expressed IGH repertoires and clonotype definition (18–20). Importantly, IMGT reference directory sets are freely available for the academic community and can be used by other programs developed for repertoire analysis.

In this work, we produced reference directory sets for IGH loci of Atlantic salmon and rainbow trout, based on a unique nomenclature developed for salmonids and following IMGT rules. We show how the particularities of salmonid IGH loci (duplicated loci in each haplotype, large number of genes and pseudogenes) were taken into account and how reference directory sets can be used for annotation of IGH expression datasets. We also discuss how the nomenclature and reference directories can be updated with new data and extended to other salmonid species.

MATERIALS AND METHODS

GU129139 and GU129140 from GenBank, ENA, and DDBJ, entered in IMGT/LIGM-DB (Rel. 201839-1) and IMGT annotated (GU129139 in Rel. 201923-5, Last updated, Version 11 and GU129140 in Rel. 201930-1, Last updated, Version 10), were selected as *S. salar* (*Salsal*) IMGT IGH locus prototypes. Sequences from these entries are from Atlantic salmon BAC library (CHORI-214), constructed from a Norwegian aquaculture strain male, from BACPAC Resources, Children's Hospital Oakland Research Institute (CHORI) (16). GU129139 (931200 bp) (Salsal locus A, ssa06, IMGT locus ID: Salsal_IGH_1) is in reverse (REV) orientation on chromosome 6 whereas GU129140 (1063283 bp) (Salsal IGH locus B, ssa03, IMGT locus ID: Salsal_IGH_2) is in forward (FWD) orientation on chromosome 3.

For obtaining IMGT gene names, newly identified Atlantic salmon and rainbow trout IGH genes and alleles from genome assemblies were submitted to the IG, T cell receptors (TR), and major histocompatibility (MH) Nomenclature Sub-Committee (IMGT-NC) of the International Union of Immunological Societies (IUIS) Nomenclature Committee[2,3]. Two IMGT_NC reports #2019-5-0131 and #2019-7-0220[2] comprise the submission of 75 Atlantic salmon IGHV sequences from two accession numbers NC_027305.1 and NC_027302.1. These reports concern 75 different genes [35 Atlantic salmon IGHV on NC_027305.1 (Salsal locus A, ssa06) and 40 Atlantic salmon IGHV genes on NC_027302.1 (Salsal locus B, ssa03)] and correspond to 75 new alleles (61 of them are *01 and 14 are *02).

Two new entries were created in IMGT/LIGM-DB: IMGT000028 for Salsal locus A [*S. salar* (Atlantic salmon), taxon:8030, breed: double haploid, assembly GCF_000233375.1, GenBank assembly ID: GCA_000233375.4, chromosome 6, CM003284.1 (20520824–22238370, complement), IGH locus A] [this entry includes IMGT annotated genes from NC_027305.1 (Salsal ssa06)] and IMGT000029 for Salsal locus B [*S. salar* (Atlantic salmon), taxon:8030, breed: double haploid, assembly GCF_000233375.1, GenBank assembly ID: GCA_000233375.4, chromosome 3, CM003281.1 (77578187–79383607), IGH locus B] [this entry includes IMGT annotated genes from NC_027302.1 (Salsal ssa03)].

The rainbow trout genome (assembly: Omyk_1.0, June 2017; GenBank assembly accession GCA_002163495.1) obtained from the homozygous Swanson clonal line was examined to locate IGH locus. Two IGH loci were identified, locus A on chromosome 13 (Oncmyk chr13) and locus B on chromosome 12 (Oncmyk chr12), both of them are in forward (FWD) orientation. The IMGT-NC Report #2019-10-040[2] comprises the submission of 181 rainbow trout IGH gene sequences from NC_035089.1 (Oncmyk Omy13) and NC_035088.1 (Oncmyk Omy12). This IMGT-NC report concerns 181 different genes: 74 genes in locus A on Oncmyk chr 13 (49 IGHV, 11 IGHD, 10 IGHJ, and 4 IGHC on NC_035089.1) and 107 genes in locus B on Oncmyk chr 12 (80 IGHV, 13 IGHD, 9 IGHJ, and 5 IGHC on NC_035088.1) and corresponds to 181 new alleles *01. Two new entries were created in IMGT/LIGM-DB: IMGT000043 (IMGT/LIGM-DB) for Oncmyk locus A [*O. mykiss* (rainbow trout), taxon:8022, isolate: Swanson, assembly Omyk_1.0, GenBank assembly ID: GCF_002163495.1, chromosome 13: CM007947.1 (48012355–48422510), IGH locus A] [this entry includes IMGT annotated genes from NC_035089.1 (Oncmyk Omy13)] and IMGT000044 for Oncmyk locus B [*O. mykiss* (rainbow trout), taxon:8022, isolate: Swanson, assembly Omyk_1.0, GenBank assembly ID: GCF_002163495.1, chromosome 12: CM007946.1 (81302817–81805590), IGH locus B] [this entry includes IMGT annotated genes from NC_035088.1 (Oncmyk Omy12)].

RESULTS

The complete and correct assembly of the Salmonidae IGH loci is a significant challenge owing to (i) the existence of two duplicated loci due to the tetraploidization (named locus A and locus B), (ii) the large size of each locus, (iii) the high number of different IGHV subgroups compared to mammals, (iv) the internal amplification and potential gene conversion that occurred inside each locus during their evolution, and (v) the very high number of pseudogenes, many of them partial, relative to the functional genes.

We therefore explored how the standardized IMGT nomenclature could allow the identification and classification of genes and alleles in incomplete or not yet fully annotated genome

[2]http://www.imgt.org/IMGTindex/IMGT-NC.php
[3]http://www.imgt.org/IMGTindex/IUIS-NC.php

assemblies. The IGH data published for Atlantic salmon (16), largely based on BAC sequencing, were used as a prototype for establishing the standardized IMGT nomenclature for salmonids and for dealing, by comparison, with newly identified IGH genes from both Atlantic salmon and rainbow trout genome assemblies. The particularities of these IGH loci (in particular the tetraploidization) were taken into consideration for consistency between salmonid species.

From IG Classes to IMGT Constant (C) Gene Names

Three antibody classes have been identified in fish, namely, IgM, IgD, and IgT, while IgG, IgA, and IgE are absent (28). IgM and IgD are generally co-expressed at the cell surface of the same B cells through alternative splicing, as in mammals. Soluble IgM are tetrameric and constitute the main antibody class in serum. A third class, IgT, is expressed in most fish groups including salmonids. Interestingly, the IG-Heavy-Tau chains of IgT have a VH domain that results from independent V-D-J rearrangements, and is not obtained by a switch process (29). IgT has been found only in bony fish and is particularly involved in mucosal immunity and protection (30). IGHD was cloned and characterized in rainbow trout and Atlantic salmon, in parallel to the discovery of IGHT encoding the third fish IG-Heavy-Tau isotype (28, 29) and then in Atlantic salmon (31).

By convention, IMGT groups are designated by the locus and gene type. Based on the four gene types, V (variable), D (diversity), J (joining), and C (constant), the IGH genes belong to four groups: IGHV, IGHD, IGHJ, and IGHC. For the IGH locus, the constant genes are designated by the letter (and, if relevant, number) corresponding to the encoded isotype (IGHT, IGHM, and IGHD), instead of using the letter C.

The salmonid IGHC genes belong to three subgroups IGHM, IGHD, and IGHT and encode, when functional, the C-REGION of the heavy chain defining these three isotypes, IG-Heavy-Mu (heavy chain of the IgM class), IG-Heavy-Delta (heavy chain of the IgD class), and IG-Heavy-Tau (heavy chain of the IgT class) (**Table 1**). Salmonid locus A and locus B were assigned based on the literature, with the letter D (for "duplicated") added to the conventional gene names for locus B.

Atlantic Salmon IGH Constant Genes and Associated D and J Genes

The Atlantic salmon IGH locus A, which is in a reverse (REV) orientation on chromosome 6 and spans 660 kilobases (kb) (with the V genes encompassing 600 kb) (**Figure 1**) includes 7 IGHC genes with 17 associated IGHD genes and 13 IGHJ genes. The Atlantic salmon IGH locus B, which is in forward (FWD) orientation on chromosome 3 and spans 720 kb (with the V genes encompassing 670 kb) (**Figure 2**) includes 5 IGHC genes with 11 associated IGHD genes and 8 IGHJ genes. The constant region of the IG-Heavy-Mu chain and of the IG-Heavy-Delta are encoded by a unique gene per locus (IGHM and IGHD for locus A and IGHMD and IGHDD for locus B) preceded by a D-J cluster. There are several IG-Heavy-Tau genes (IGHT), but the associated D-J cluster may be incomplete (lacking D and/or J genes). In Atlantic salmon, there is only one IGHT functional (F) gene per locus, IGHT4 for locus A and IGHT2D for locus B, each one having a complete D-J cluster (**Table 2**).

In the Atlantic salmon locus A, the D and J genes associated to IGHT genes comprise two D (IGHD1T2 and IGHD2T2) and two J (IGHJ1T2 and IGHJ2T2) upstream of the pseudogene (P) IGHT2, two J (IGHJ1T3 and IGHJ2T3) upstream of IGHT3 (P), five D (IGHD1T4 to IGHD5T4) and two J (IGHJ1T4 and IGHJ2T4) genes, all of them functional, upstream of IGHT4 (F) and one D (IGHD1T5) and two J (IGHJ1T5 and IGHJ2T5) upstream of IGHT5 (P). There is no IGHD or IGHJ upstream of IGHT1 (P) (**Table 2**). The D and J associated to IGHM and IGHD comprise nine D (IGHD1 to IGHD9), all of them functional and five J genes, three of them functional (IGHJ1, IGHJ3, and IGHJ4), one with ORF, the IGHJ2, and one with alleles F or ORF (IGHJ5). They are located upstream of IGHM (F) and shared with the IGHD constant gene (F) (**Table 2** and **Figures S1**, **S2**). Eleven IGHD not directly associated to constant genes are dispersed in locus A (IGHD-1 to IGHD-11).

In the Atlantic salmon locus B, the D and J genes associated to IGHT genes comprise one J (IGHJ1T1D) upstream of IGHT1D (P), two D (IGHD1T2D and IGHD2T2D), and two J (IGHJ1T2D and IGHJ2T2D) all functional upstream of IGHT2D (F) and three D (IGHD1T3D, IGHD2T3D, and IGHD3T3D) downstream of IGHT3D (P) (**Table 2** and **Figures S1**, **S2**). The D

TABLE 1 | Salmonid IG receptor classes, heavy chain types, and IGHC gene names.

IG receptor class	IG heavy chain type	IG C-gene group	IG C-gene subgroup	IGHC gene names			
				Salmo salar		*Oncorhynchus mykiss*	
				Locus A	Locus B	Locus A	Locus B
IgM	IG-Heavy-Mu	IGHC	IGHM	IGHM	IGHMD	IGHM	IGHMD
IgD	IG-Heavy-Delta	IGHC	IGHD	IGHD	IGHDD	IGHD	IGHDD
IgT	IG-Heavy-Tau	IGHC	IGHT	IGHT1 IGHT2 IGHT3 IGHT4 IGHT5	IGHT1D IGHT2D IGHT3D	IGHT1 IGHT2	IGHT1D

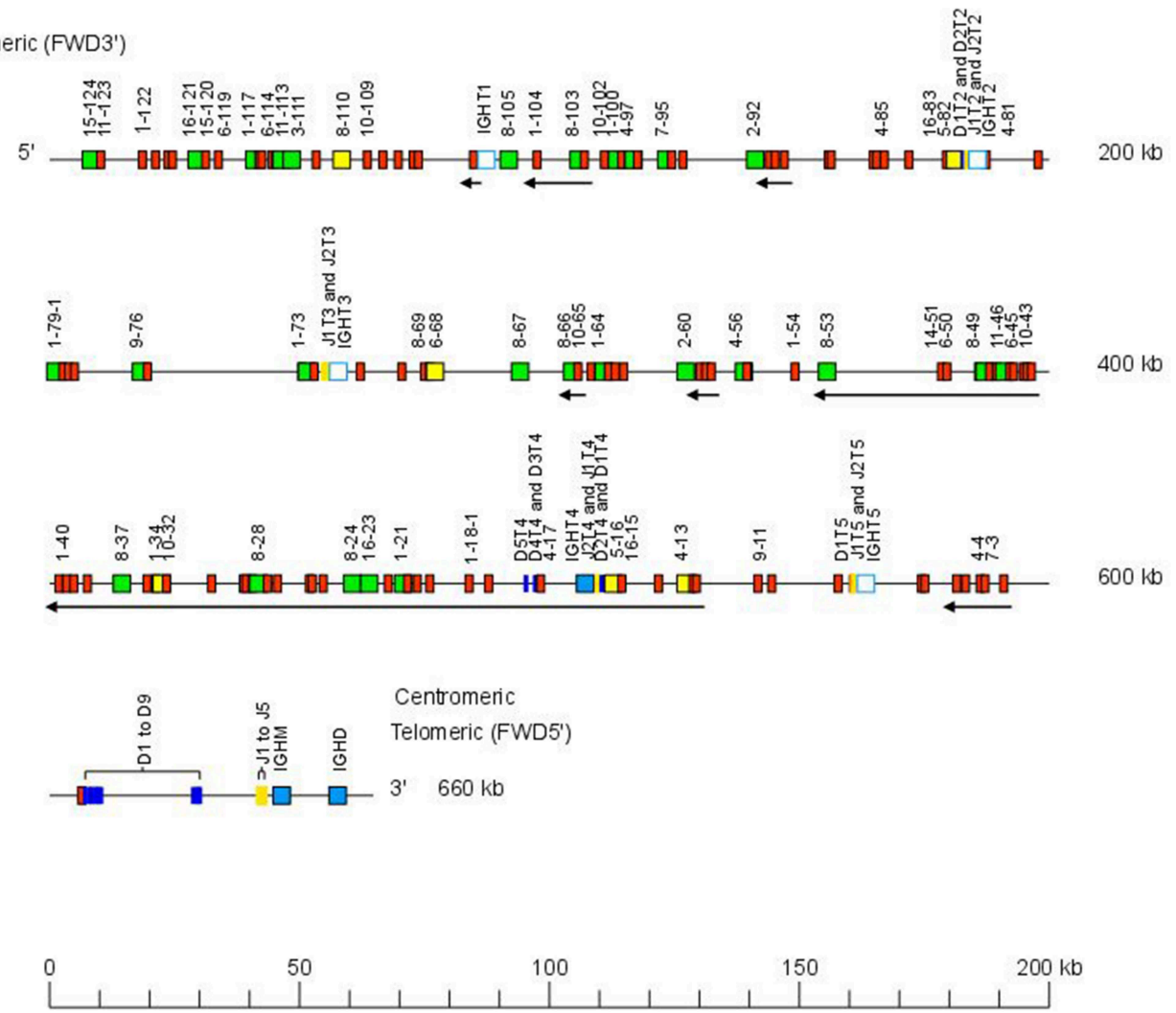

FIGURE 1 | Organization of the Atlantic salmon (*Salmo salar*) locus A. The orientation of the Atlantic salmon (*S. salar*) IGH locus on chromosome 6 (locus A—Salsal ssa06) is reverse (REV). IGH gene names are according to IMGT nomenclature (3). Single arrows show genes whose polarity is opposite to that of the D-J-C CLUSTER comprising IGHD (D1–D9)-IGHJ (J1–J5)-IGHM-IGHD. Gene names of IGHV pseudogenes with frameshift(s) in the V-REGION are not displayed due to lack of space in the locus map at the IMGT standardized scale (line = 200 kb). All gene names are displayed in a zoom (line = 100 kb) on the IMGT site [IMGT Repertoire (IG and TR) > Locus and genes > Locus representations][1]. Color coding of genes is according to the IMGT Color menu for genes (IMGT Scientific chart > 4. Representation Rules > IMGT color menu > 11. Color menu for genes)[1]: V-GENE (green: functional, yellow: ORF, red: pseudogene), D-GENE (blue: functional), J-GENE (yellow: functional), and C-GENE (blue: functional, white box: pseudogene). With permission of IMGT®, the international ImMunoGenetics information system®.

and J genes associated to IGHMD and IGHDD comprise six D (IGHD1D to IGHD6D, all functional) and five J genes (four functional, IGHJ1D to IGHJ4D) and one with alleles F or ORF (IGHJ5D). They are located upstream of IGHMD (F) and shared with the IGHDD constant gene (F) (**Table 2** and **Figures S1**, **S2**). Six IGHD not directly associated to constant genes are dispersed in locus B (IGHD-1D to IGHD-6D).

IGHD, IGHJ, and IGHC genes are reported in IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHD > Atlantic salmon (*S. salar*); ibid., IGHJ > Atlantic salmon (*S. salar*); ibid., IGHC > Atlantic salmon (*S. salar*)][1].

Rainbow Trout IGH Constant Genes and Associated D and J Genes

Similar to the Atlantic salmon, the rainbow trout has one functional gene per IGH locus encoding the constant region of the IG-Heavy-Mu (IGHM gene in locus A and IGHMD gene in locus B), the constant region of the IG-Heavy-Delta (IGHD gene in locus A and IGHDD gene in locus B), and the constant region of the IG-Heavy-Tau (IGHT2 gene in locus A and IGHT1D gene in locus B).

The rainbow trout IGH locus A, which spans 360 kb and is in a forward (FWD) orientation on chromosome 13, includes 11 IGHD genes, 10 IGHJ genes, and 4 IGHC genes (**Table 3**).

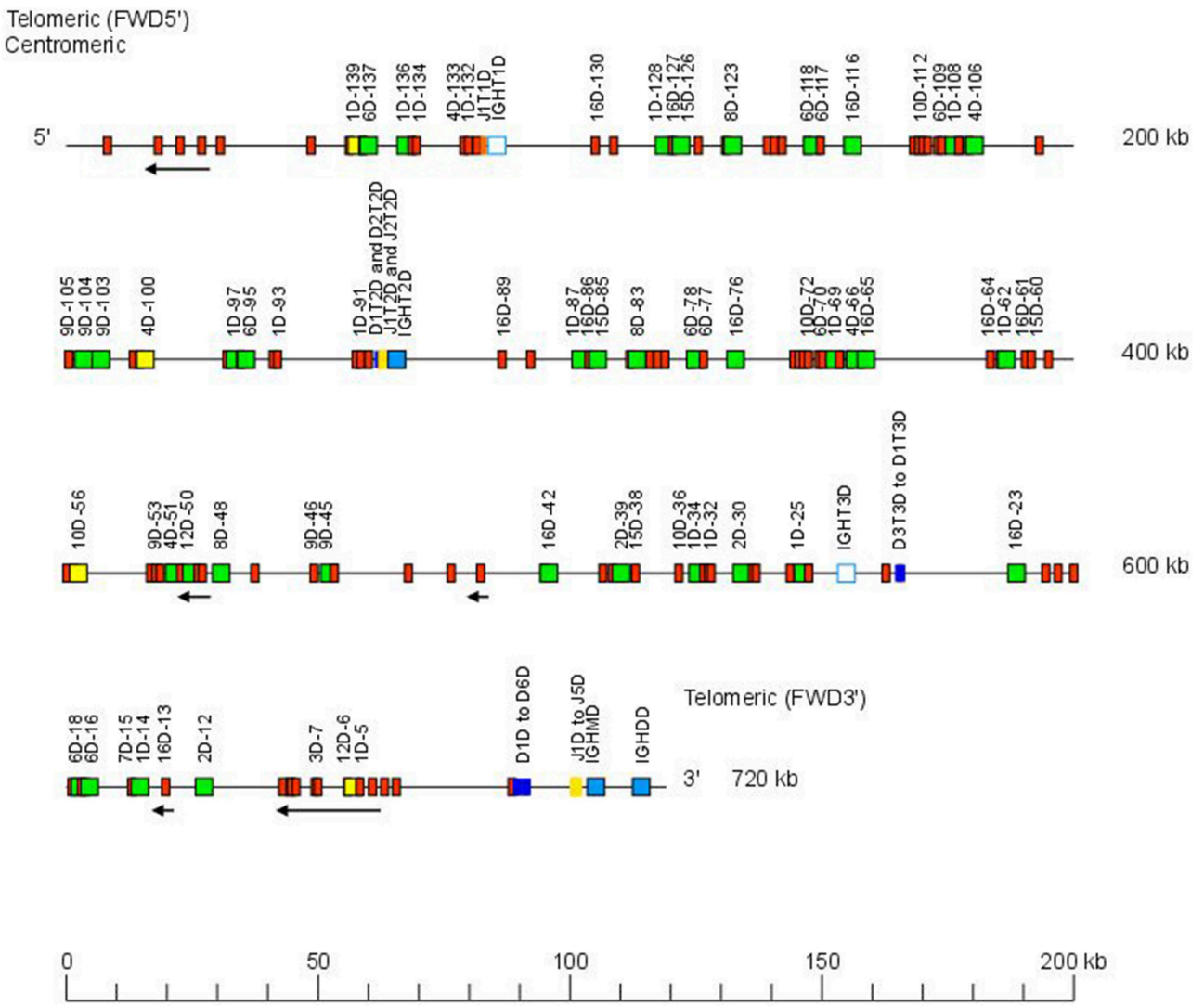

FIGURE 2 | Organization of the Atlantic salmon (*S. salar*) locus B. The orientation of the Atlantic salmon (*S. salar*) IGH locus on chromosome 3 (locus B—Salsal ssa03) is forward (FWD). IGH gene names are according to IMGT nomenclature (3). Single arrows show genes whose polarity is opposite to that of the D-J-C CLUSTER comprising IGHD (D1–D9)-IGHJ (J1–J5)-IGHM-IGHD. Gene names of IGHV pseudogenes with frameshift(s) in the V-REGION are not displayed due to lack of space in the locus map at the IMGT standardized scale (line = 200 kb). All gene names are displayed in a zoom (line = 100 kb) on the IMGT site [IMGT Repertoire (IG and TR) > Locus and genes > Locus representations][1]. Color coding of genes is according to the IMGT Color menu for genes (IMGT Scientific chart > 4. Representation Rules > IMGT color menu > 11. Color menu for genes)[1]: V-GENE (green: functional, yellow: ORF, red: pseudogene), D-GENE (blue: functional), J-GENE (yellow: functional), and C-GENE (blue: functional, white box: pseudogene). With permission of IMGT®, the international ImMunoGenetics information system®.

There are three D and two J genes upstream of IGHT1 (P), two D and two J genes upstream of IGHT2 (F), and six D and six J genes (all of them F) upstream of IGHM (F) and shared with the IGHD (F) constant gene (**Figures S1**, **S2**).

The rainbow trout IGH locus B, which spans 485 kb and is in a forward (FWD) orientation on chromosome 12, includes 13 IGHD genes, 9 IGHJ genes, and 3 IGHC genes (**Table 3**). There are four D genes (1 ORF and 3 F) and two J genes (both F) upstream of IGHT1D (F), and six D and seven J genes (all of them F) upstream of IGHMD (F) and shared with the IGHDD (F) constant gene (**Figures S1**, **S2**).

Sequences of rainbow trout IGHD and IGHJ genes and alleles are available in the downloadable IMGT reference directory sets from IMGT/GENE-DB (/download/GENE-DB)[1] and from IMGT/V-QUEST (/download/V-QUEST/IMGT_V-QUEST_reference_directory/Oncorhynchus_mykiss/IG/IGHD.fasta; ibid., /IGHJ.fasta)[1]. IGHD and IGHJ genes and alleles are reported in the IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHD > Rainbow trout (*O. mykiss*); ibid., IGHJ > Rainbow trout (*O. mykiss*)][1].

The demonstration that there is only one rainbow trout IG-Heavy-Delta complete gene per locus, IGHD in locus A and

TABLE 2 | Atlantic salmon (*Salmo salar*) IGH constant C genes and associated D and J genes.

***Salmo salar* locus A on chromosome 6 (Salsal ssa06)**						***Salmo salar* locus B on chromosome 3 (Salsal ssa03)**					
IGHD genes		**IGHJ genes**		**IGHC genes**		**IGHD genes**		**IGHJ genes**		**IGHC genes**	
–		–		IGHT1	P	–		IGHJ1T1D	P	IGHT1D	P
IGHD1T2	F	IGHJ1T2	F	IGHT2	P	IGHD1T2D	F	IGHJ1T2D	F	IGHT2D	F
IGHD2T2	F	IGHJ2T2	F			IGHD2T2D	F	IGHJ2T2D	F		
		IGHJ1T3	F	IGHT3	P	IGHD1T3D	F			IGHT3D	P
		IGHJ2T3	F			IGHD2T3D	F				
IGHD1T4	F	IGHJlT4	F	IGHT4	F	IGHD3T3D	F				
IGHD2T4	F	IGHJ2T4	F								
IGHD3T4	F										
IGHD4T4	F										
IGHD5T4	F										
IGHD1T5	F	IGHJlT5	P	IGHT5	P						
		IGHJ2T5	F								
IGHD1	F	IGHJ1	F	IGHM	F	IGHD1D	F	IGHJlD	F	IGHMD	F
IGHD2	F	IGHJ2	ORF			IGHD2D	F	IGHJ2D	F		
IGHD3	F	IGHJ3	F			IGHD3D	F	IGHJ3D	F		
IGHD4	F	IGHJ4	F			IGHD4D	F	IGHJ4D	F		
IGHD5	F	IGHJ5	F, ORF			IGHD5D	F	IGHJ5D	F, ORF		
IGHD6	F					IGHD6D	F				
IGHD7	F										
IGHD8	F										
IGHD9	F										
				IGHD	F					IGHDD	F

F, functional; ORF, open reading frame; P, pseudogene. The functionality is according to IMGT functionality (IMGT Scientific chart > IMGT functionality)[1] *(3).*

IGHDD in locus B, respectively, and that these two genes are functional, results from the analysis derived from applying the nomenclature of the Atlantic salmon IGH loci as well as the interpretation of expression data and published references (15, 16, 29, 31). The anomalies (partial IGHD and IGHDD genes with exons in aberrant localizations or in reverse-complementary orientation) are likely artifacts of the current genome assembly. For that reason, the functionality of the IGHD and IGHDD, deduced from literature data and supported by sequences external to the genome assembly, is shown in parentheses in **Table 3**.

Atlantic Salmon IGH Variable Genes

The Atlantic salmon IGH locus comprises a total of 303 IGH variable (IGHV) genes (145 IGHV in locus A on Salsal chromosome 6, spanning 600 kb, and 158 IGHV in locus B on Salsal chromosome 3, spanning 670 kb) (**Figures 1**, **2**). There are a total of 67–69 functional genes, 12 ORF, and 222–224 pseudogenes (**Table 4**).

Based on the percentage of identity between nucleotide sequences of the V-REGION (threshold 75%), the Atlantic salmon 303 IGHV genes can be classified into 16 IGHV subgroups. IGHV genes are reported in IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHV > Atlantic salmon (*S. salar*)][1]. Correspondence with previous gene names is indicated.

Translation of alleles *01 of F, ORF, and in-frame P are aligned according to the IMGT unique numbering in IMGT Protein display allowing the visualization of the FR-IMGT and CDR-IMGT [IMGT Repertoire (IG and TR) > 2. Proteins and alleles > Protein displays > IGHV > Atlantic salmon (*S. salar*)][1] and the comparison of the CDR-IMGT lengths per subgroup (3) {IMGT Repertoire (IG and TR) > 3. 2D and 3D structures > FR-IMGT and CDR-IMGT lengths (V-REGION and V-DOMAIN) > [CDR1-IMGT.CDR2-IMGT.] length per subgroup > IGHV > Atlantic salmon (*S. salar*)}[1] (3).

Rainbow Trout IGH Variable Genes

A total of 129 IGHV genes were identified in the rainbow trout genome, of which 57 can be considered fully functional or with an ORF without stop codon. A number of other sequences were identified as IGHV fragments in the assembly and were not included in the annotation. On chromosome 13 (locus A), 44 IGHV genes were found upstream of the functional IGHT2 gene, as well as 5 IGHV genes between the D-J-IGHT2 cluster and the D-J-IGHM-IGHD cluster. Eighty IGHV genes were found on chromosome 12 (locus B): 70 IGHV were located upstream of the functional IGHT1D gene and 10 IGHV were found between

TABLE 3 | Rainbow trout (*Oncorhynchus mykiss*) IGH constant C genes and associated D and J genes.

***Oncorhynchus mykiss* locus A on chromosome 13 (Oncmyk Omy13)**						***Oncorhynchus mykiss* locus B on chromosome 12 (Oncmyk Omy12)**					
IGHD genes		**IGHJ genes**		**IGHC genes**		**IGHD genes**		**IGHJ genes**		**IGHC genes**	
IGHD1T1	F	IGHJ1T1	F	IGHT1	P	IGHD1T1D	F	IGHJ1T1D	F	IGHT1D	F
IGHD2T1	F	IGHJ2T1	F			IGHD2T1D	F	IGHJ2T1D	F		
IGHD3T1	F					IGHD3T1D	ORF				
						IGHD4T1D	F				
IGHD1T2	F	IGHJ1T2	F	IGHT2	F						
IGHD2T2	F	IGHJ2T2	F								
IGHD1	F	IGHJ1	F	IGHM	F	IGHD1D	F	IGHJ1D	F	IGHMD	
IGHD2	F	IGHJ2	F			IGHD2D	F	IGHJ2D	F		
IGHD3	F	IGHJ3	F			IGHD3D	F	IGHJ3D	F		
IGHD4	F	IGHJ4	F			IGHD4D	F	IGHJ4D	F		
IGHD5	F	IGHJ5	F			IGHD5D	F	IGHJ5D	F		
IGHD6	F	IGHJ6	F			IGHD6D	F	IGHJ6D	F		
								IGHJ7D	F		
				IGHD	F					IGHDD	F

F, functional; ORF, open reading frame; P, pseudogene. The functionality is according to IMGT functionality (IMGT Scientific chart > IMGT functionality)[1] *(3).*

the D-J-IGHT1D cluster and the D-J-IGHMD-IGHDD cluster. The 129 rainbow trout IGHV genes could be classified into the same 16 subgroups defined for the Atlantic salmon IGHV genes, containing from only 1 pseudogene (i.e., IGHV5, IGHV13, and IGHV14 subgroups) to 35 genes, i.e., IGHV1 subgroup, which includes 12 F, 2 ORF, and 21 P IGHV genes. **Figure 3** shows a phylogenetic tree based on nucleotide sequences of IGHV genes (F and ORF) present in Atlantic salmon and rainbow trout IGH loci. While some IGHV subgroups are not represented in both species, as far as we know, this tree illustrates how rainbow trout IGHV genes nicely cluster with their Atlantic salmon counterparts.

Expressed Repertoire Analysis

IMGT/V-QUEST and its high-throughput version, IMGT/HighV-QUEST, can perform analysis of nucleotide sequences of the IG and TR variable domains (21, 22, 26, 27). These tools run against the IMGT/V-QUEST reference directory database that includes several sets (per group and per species) and are built based on the IMGT standards (3) (annotation in IMGT/LIGM-DB, Gene tables, Alignments of alleles, Protein display, entry in IMGT/GENE-DB). The IMGT/V-QUEST sets comprise IMGT reference sequences from all functional (F) and ORF genes and alleles (in Advanced parameters, Selection of IMGT reference directory set "F + ORF"). The sets also include IMGT reference sequences from pseudogenes (P) and alleles with an in-frame V-REGION for versatile genomic analysis (proposed by default, in Advanced parameters IMGT reference directory set "F + ORF + in-frame P").

Altogether, IMGT/V-QUEST reference directory for Atlantic salmon IGHV contains 150 alleles that include 76 F, 15 ORF, and 59 P in-frame (release: 201931-4, 1st August 2019) (**Table 5**). The 76 F comprise, in addition to the 67 F alleles *01 (28 from locus A and 39 from locus B), 8 alleles *02 and 1 allele *03. The 15 ORF comprise, in addition to the 12 ORF alleles *01 (7 from locus A and 5 from locus B), 3 alleles *02. The 59 in-frame P comprise, in addition to the 54 P alleles *01 (26 from locus A and 28 from locus B), 5 alleles *02. Alleles of closely related duplicated genes are managed in the same Alignments of alleles, as shown, for example, for IGHV1-64*01 F and IGHV1-100*01 F, which have identical V-REGION nucleotide sequences [IMGT Repertoire (IG and TR) 2. Proteins and alleles > Alignments of alleles > IGHV > Atlantic salmon (*S. salar*)][1].

IMGT/V-QUEST reference directory for rainbow trout IGHV contains sequences of 77 alleles that include 44 F, 11 ORF, and 22 P in-frame (/download/V-QUEST/IMGT_V-QUEST_reference_directory/Oncorhynchus_mykiss/IG/IGHV.fasta)[1]. All the alleles are *01 (release: 201931-4, 1st August 2019) (**Table 5**). All IGHV genes and alleles (including the P out-of-frame) are reported in IMGT Gene tables [IMGT Repertoire (IG and TR) > 1. Locus and genes > Gene tables > IGHV > Rainbow trout (*O. mykiss*)][1].

We then investigated the functionality and expression level of IGHV genes from the two species using the standardized nomenclature based on genomic annotation. To do so, adaptive immune receptor repertoire datasets generated by high-throughput sequencing (AIRRseq) were submitted to IMGT/HighV-QUEST analysis.

Atlantic Salmon

AIRRseq data from head kidney of Atlantic salmon were generated based on 5′RACE and specific primers for IGHM constant region [data from reference (32)]. Using the Atlantic salmon reference dataset updated in 2019, a total of 50 IGHV genes (42 functional "F," 4 "ORF" and 4 pseudogenes "P") were expressed in the dataset (**Figure 4A**). More than 80% of

TABLE 4 | Atlantic salmon IGH variable genes.

IMGT group	IMGT subgroup	Locus A on Salsal chromosome 6				Locus B on Salsal chromosome 3				Locus A + Locus B			
		Functional	ORF	Pseudogene	Total	Functional	ORF	Pseudogene	Total	Functional	ORF	Pseudogene	Total
IGHV	IGHV1	7	1	24	32	12	2	23	37	19	3	47	69
	IGHV2	2	0	0	2	2(+1)*	0	5(+1)*	8	4(+1)*	0	5(+1)*	10
	IGHV3	1	0	4	5	1	0	5	6	2	0	9	11
	IGHV4	2	2	12	16	3	1	14	18	5	3	26	34
	IGHV5	0	2	2	4	0	0	4	4	0	2	6	8
	IGHV6	1	1	16	18	6	0	20	26	7	1	36	44
	IGHV7	1	0	2	3	0	0	3	3	1	0	5	6
	IGHV8	10(+1)*	1	9(+1)*	21	3	0	5	8	13(+1)*	1	14(+1)*	29
	IGHV9	1	0	4	5	3	0	5	8	4	0	9	13
	IGHV10	0	0	14	14	0	1	8	9	0	1	22	23
	IGHV11	2	0	6	8	0	0	6	6	2	0	12	14
	IGHV12	0	0	1	1	1	1	0	2	1	1	1	3
	IGHV13	0	0	1	1	0	0	2	2	0	0	3	3
	IGHV14	0	0	1	1	0	0	1	1	0	0	2	2
	IGHV15	1	0	5	6	2	0	3	5	3	0	8	11
	IGHV16	1	0	7	8	5	0	10	15	6	0	17	23
	Total	29(+1)*	7	108(+1)*	145	38(+1)*	5	114(+1)*	158	67(+2)*	12	222(+2)*	303

Number of IGHV genes are given per subgroup and per locus A or B, and per IMGT functionality (functional, ORF, pseudogene) (3).
**An asterisk indicates that the following genes have alleles with different functionalities: Functional or Pseudogene (IGHV2D-12 and IGHV8-58).*

submitted sequences presented IGHV F genes. Interestingly, the majority of expressed V genes were from locus B (chromosome 3). This difference was reflected in the abundance of rearrangements (∼66% from locus B) and in the diversity of IGHV genes expressed: 25 IGHV from locus B vs. 17 IGHV from locus A (**Figure 4A**). On average, IGHV1D-25*01, IGHV6D-18*01, IGHV6D-16*01, and IGHV1-73*01 were the most abundant IGHV functional genes, accounting for 30% of the expressed repertoire.

Rainbow Trout

In this species, we analyzed AIRRseq datasets from fish intraperitoneally immunized with a killed bacterial pathogen, *Yersinia ruckeri* [data from reference (33)]. 5′RACE PCR products were produced from spleen of immunized fish, using specific primers for IGHM constant region and with unique molecular identifiers (UIDs) for better data normalization (33). Only in-frame productive rearrangements (CDR3-IMGT without stop codons) were analyzed. Trout used in this study belonged to the isogenic line derived from Swanson strain that was selected for the rainbow trout genome project. Hence, these AIRRseq data express IGH genes from the very same repertoire, which was annotated in the current IMGT reference directories. These data therefore provided a quantitative assessment of the expression of IGHV genes in the spleen of three genetically similar individuals responding to a pathogen.

In this dataset, IMGT/High V-QUEST unambiguously identified the IGHV gene in 94% of submitted sequences, 90% of them with at least 99% of sequence identity (52% with 100% of identity). A total of 55 IGHV genes (35 functional "F," 9 "ORF," and 7 pseudogenes "P") were expressed. Interestingly, these rearrangements are from both IGH loci (A and B) in relatively similar proportions.

In each trout sample, about 17% of sequences corresponded to IGHV ORF genes and 1.7–4.7% corresponded to IGHV pseudogenes (most of them correspond to IGHV1D-12*01 P or IGHV1-21*01 P) involved in-frame junction rearrangements. This feature could be detected because we selected the IMGT/HighV-QUEST directory sets "F + ORF + in-frame P," which also include pseudogenes with in-frame stop codon in V region or defect in the leader or recombination signal (RS) sequences (3). Although IGH transcripts with stop codon are generally rare in mammals, they are typically much more frequent in fish, perhaps because nonsense-mediated mRNA decay (NMD) may work differently (28, 32, 34, 35).

Hence, about 80% of submitted rainbow trout sequences presented functional IGHV genes (**Figure 4B**). IGHV4D-24*01 F, IGHV6D-40*01 F, IGHV1-18*01 F, and IGHV11-25*01 F were the most expressed on average, with a limited interindividual variation as expected from the genetic constitution of the fish analyzed. In this dataset, for about 6% of submitted sequences, IMGT/HighV-QUEST provided two results assigned to distinct duplicated germline IGHV with alleles having identical or close sequences (for example, IGHV12D56*01/IGHV12D57*01, or IGHV8-30*01/IGHV8-40*01) owing to the gene duplication in salmonids.

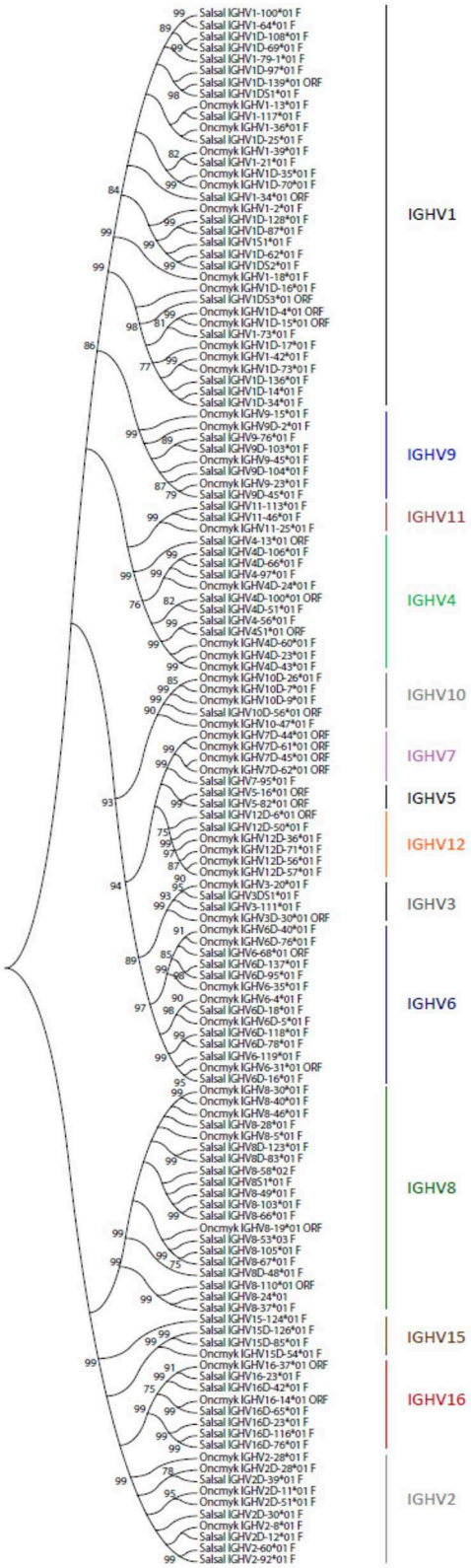

FIGURE 3 | Phylogenetic analysis of Atlantic salmon and rainbow trout IGHV gene sequences. A phylogenetic analysis of the V-REGION of Atlantic salmon (Salsal) and rainbow trout (Oncmyk) IGHV genes (F + ORF) was performed using the UPGMA method and a bootstrap based on 1,000 replicates. A total *(Continued)*

FIGURE 3 | of 136 genes, 81 from Salsal (69 F + 12 ORF) and 55 from Oncmyk (44 F + 11 ORF) from the IMGT/V-QUEST reference directory sets (release 201931-4, 1st August 2019) (**Table 5**), were compared. Only one allele per gene was included in the analysis (allele *01 for all but two IGHV8-58*02 F and IGHV8-53*03 F). Nodes with a bootstrap support higher than 75% are indicated.

Although the datasets analyzed here for salmon and trout were not selected for direct comparison, it suggests that these two species (at least, the fish strains analyzed here) do not use the two loci in the same way (see above). A rigorous and comprehensive comparison of expressed repertoires between rainbow trout and Atlantic salmon will require a systematic comparison of AIRRseq data from multiple strains.

Genetic Variability of IG Genes in Salmonids

Making available a full annotation and versatile nomenclature also offers the possibility to better integrate new data about variability of IG (or TR) genes. This issue is of particular interest in Salmonids for two main reasons: (1) variations of IG gene sequences may affect the repertoire of specificities targeted by Abs, in turn impacting the quality and efficiency of responses against pathogens, and (2) salmonid IG loci are particularly complex with high numbers of functional genes and pseudogenes located in two regions; therefore, they constitute interesting models to understand mechanisms of short-term evolution of such loci and the potential importance of homogenization vs. diversification of IG sequences.

To get preliminary data about IGHV variation in a salmonid species, we took advantage of the full genome sequencing of 19 isogenic lines of rainbow trout. These lines were produced using a mitogynogenesis-based strategy by Quillet et al. (36, 37). They represent 19 haplotypes randomly picked from the so-called INRA-SY "synthetic" population. This population was created about 35 years ago by a planned random mating (i.e., panmictic) mixture of French, Danish, and American domestic populations, and has been maintained since without any voluntary selection. The 19 isogenic lines analyzed here do not appear to be closely related to the Swanson trout generated at Washington State University using androgenesis, which has been sequenced and constitutes the reference genome (38, 39).

The numbers of indel and SNP detected within IGHV genes and pseudogenes are indicated in **Table 6**. Genetic variation between isogenic lines overall appears to be relatively modest at this level. It seems to be more frequent in the locus located in chromosome 13 (67 SNP and 1 indel for 29 functional genes, 41 SNP and 3 indel for 20 pseudogenes) compared to chromosome 12 (23 SNP and no indel for 29 functional genes, and 53 SNP and 10 indel for 51 pseudogenes). The proportion of silent vs. non-silent mutations was not significantly different between the two regions (40NS/67 SNP for chromosome 13 and 13NS/23 SNP for chromosome 12), suggesting that these genes did not evolve under strong positive selection. Indel

TABLE 5 | Atlantic salmon (*S. salar*) and rainbow trout (*O. mykiss*) IGHV alleles included in the IMGT/V-QUEST reference directory sets (release 201931-4, 1st August 2019).

IGHV subgroup	Atlantic salmon								Rainbow trout							
	Nb of genes	Nb of alleles	IGH locus A			IGH locus B			Nb of genes	Nb of alleles	IGH locus A			IGH locus B		
			F*	ORF+	P*	F*	ORF*	P*			F*	ORF*	P*	F*	ORF*	P*
IGHV1	33	38	7	1(2)	5(6)	12(14)	2	6(7)	17	17	6	0	2	5	2	2
IGHV2	5	5	2	0	0	3	0	0	8	8	2	0	1	3	0	2
IGHV3	3	4	1(2)	0	0	1	0	1	2	2	1	0	0	0	1	0
IGHV4	14	14	2	2	4	3	1	2	9	9	0	0	1	4	0	4
IGHV5	2	3	0	2(3)	0	0	0	0	0	0	0	0	0	0	0	0
IGHV6	14	16	1(2)	1	3	6	0	3(4)	8	8	2	1	0	3	0	2
IGHV7	3	3	1	0	1	0	0	1	4	4	0	0	0	0	4	0
IGHV8	16	20	11(14)	1(2)	1	3	0	0	6	6	4	1	1	0	0	0
IGHV9	8	8	1	0	1	3	0	3	5	5	3	0	1	1	0	0
IGHV10	9	9	0	0	5	0	1	3	4	4	1	0	0	3	0	0
IGHV11	3	4	2	0	1(2)	0	0	0	1	1	1	0	0	0	0	0
IGHV12	2	2	0	0	0	1	1	0	4	4	0	0	0	4	0	0
IGHV13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IGHV14	1	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0
IGHV15	6	7	1	0	1	2	0	2(3)	3	3	0	0	0	1	0	2
IGHV16	16	16	1	0	3	5	0	7	5	5	0	2	2	0	0	1
Total	135	150	30(35)	7(10)	26(28)	39(41)	5	28(31)	77	77	20	4	9	24	7	13
Locus A + Locus B	135	150	69(76)F + 12(15)ORF + 54(59)P						77	77	44 F + 11 ORF + 22 P					

**F, functional; ORF, open reading frame; P, pseudogene. Number of genes included in the IMGT reference directory, per subgroup and per functionality and total, are shown with, if relevant (more than one allele per gene), the corresponding number of alleles within brackets. The functionality is according to IMGT functionality (IMGT Scientific chart > IMGT functionality)*[1] *(3).*

and SNP were not significantly more frequent in pseudogenes. Variants were filtered to eliminate all assembly artifacts, but these data will have to be fully validated by resequencing, and the impact of variation on the gene status evaluated. We have indications that several new genes are present in productive and expressed rearrangements. This might be due to the absence of such genes in the genome of the Swanson strain or to gaps in the current reference genome assembly. In this context, it is of interest to evaluate the variability of IGHV gene numbers between the different haplotypes. Future assemblies will allow a more accurate description of the IGH diversity and variability. Incompleteness of the annotated repertoire may constitute a problem for repertoire analysis (for example, when a missing gene is used by a clonotype clonally selected in a response). Hence, sequences of genes that are not localized in the current assembly may be added to the IMGT Reference directories sets, providing that sufficient evidence is available to demonstrate their existence and expression. These sequences will be given a provisional name (with S) until their location and presence in the germline genomic sequence are validated. If new genes would appear, which do not belong to any of the IGHV subgroups identified and described in this work, a new subgroup may have to be defined. This is not impossible, but seems to be unlikely since we believe that the large set of IGHV sequences analyzed from Atlantic salmon and rainbow trout probably contains at least one representative of all subgroups. Such additions will be validated by the IG, TR, and MH Nomenclature Sub-Committee (IMGT-NC) (6, 7) of the IUIS Nomenclature Committee[2,3], following a procedure analogous to the one used for example for inferred alleles in human.

CONCLUSION

Genome assembly is available for both Atlantic salmon and rainbow trout, representing the two main genera of Salmonids (*Salmo* and *Oncorhynchus*). More genomic (and transcriptome) data are coming from a number of genomic backgrounds, which will provide a rich source of knowledge about variations of potential antibody repertoires in these species. We therefore revisited the description and annotation of the two IGH loci present in these two species, currently from cDNA and BAC clone sequences, based on the IMGT biocuration and nomenclature for Salmonid IGH genes that will facilitate the analysis of AIRRseq data.

The IG or antibody repertoire sequencing has started to develop both in rainbow trout and in Atlantic salmon, reflecting a growing interest for an accurate and comprehensive description of the response against common pathogens and vaccines. As full genome assemblies are now available for several salmonid species (Atlantic salmon, rainbow trout, coho salmon, and chinook salmon), comparative analysis of the IGH locus structure in

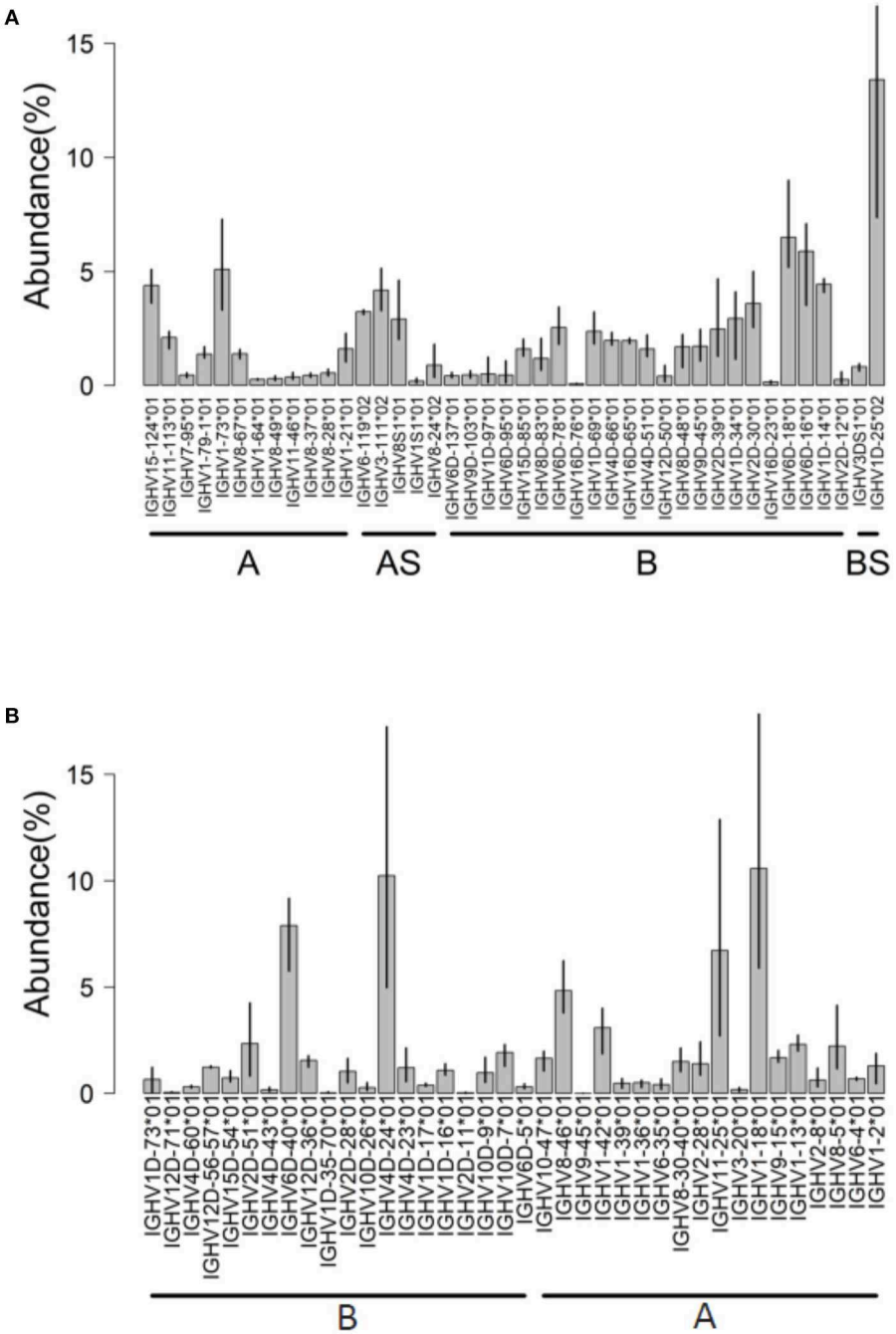

FIGURE 4 | IGHV usage determined by IMGT/HighV-QUEST tool. Analysis of AIRRseq datasets obtained previously from head kidney of three healthy Atlantic salmons **(A)** and from spleen of three rainbow trouts that were intraperitoneally immunized with killed *Yersinia ruckeri* **(B)**. Libraries were generated by 5′RACE using specific primers for IGHM constant region. IGHV usage is expressed as the percentage of total productive rearrangements.

these closely related tetraploidized species is of great interest. It also appears very important to investigate the level of variation between germline repertoires of IG genes across commercial and wild salmonid stocks. This variation may have significant implications for practical issues in aquaculture and conservation; it will also be of significant interest for the basic comparative immunology community, in particular to address accurately the mechanisms of gene conversion, somatic hypermutation, and memory in these species and during vertebrate evolution.

TABLE 6 | Number of SNP and variants in IGHV genes and pseudogenes across 19 isogenic rainbow trout lines.

	Functional genes					Pseudogenes				
Chrom	**Start**	**Stop**	**Name**	**SNPs number (NS)**	**Indel number**	**Start**	**Stop**	**Name**	**SNPs number**	**Indel number**
(A)										
Chr12	81 322 385	81 322 680	IGHV6D-76	2(1)	0	81 312 817	81 313 128	IGHV16D-79	3	0
Chr12	81 335 727	81 336 024	IGHV1D-73	2(2)	0	81 318 367	81 318 661	IGHV15D-78	5	1
Chr12	81 339 680	81 339 363	IGHV12D-71	1(0)	0	81 320 821	81 321 111	IGHV1D-77	1	0
Chr12	81 365 089	81 365 411	IGHV15D-69	0	0	81 332 711	81 333 058	IGHV3D-75	2	2
Chr12	81 365 950	81 366 255	IGHV2D-68	0	0	81 333 678	81 333 986	IGHV1D-74	1	0
Chr12	81 395 848	81 396 168	IGHV7D-62	0	0	81 336 977	81 337 276	IGHV4D-72	0	0
Chr12	81 397 765	81 398 073	IGHV4D-60	0	0	81 383 887	81 384 159	IGHV1D-65	0	0
Chr12	81 422 492	81 422 803	IGHV15D-54	0	0	81 384 684	81 384 968	IGHV6D-64	0	0
Chr12	81 436 861	81 437 166	IGHV2D-50	0	0	81 388 213	81 388 533	IGHV1D-63	1	1
Chr12	81 438 847	81 439 169	IGHV15D-49	0	0	81 396 766	81 397 086	IGHV7D-61	0	0
Chr12	81 464 500	81 464 820	IGHV7D-45	0	0	81 398 323	81 398 601	IGHV1D-59	0	0
Chr12	81 465 762	81 466 082	IGHV7D-44	0	0	81 399 513	81 399 812	IGHV4D-58	0	0
Chr12	81 466 761	81 467 069	IGHV4D-43	0	0	81 401 688	81 402 032	IGHV12D-57	0	0
Chr12	81 493 829	81 494 124	IGHV6D-40	0	0	81 402 044	81 401 727	IGHV12D-56	0	0
Chr12	81 529 173	81 529 478	IGHV1D-35	0	0	81 421 382	81 421 689	IGHV16D-55	11	0
Chr12	81 561 752	81 562 063	IGHV3D-30	0	0	81 434 530	81 434 810	IGHV2D-52	0	0
Chr12	81 568 866	81 569 171	IGHV2D-28	0	0	81 435 683	81 435 988	IGHV2D-51	0	0
Chr12	81 595 836	81 596 138	IGHV10D-26	2(0)	0	81 447 972	81 448 244	IGHV1D-48	0	0
Chr12	81 605 108	81 605 419	IGHV4D-24	1(0)	0	81 448 773	81 449 068	IGHV6D-47	0	0
Chr12	81 618 381	81 618 689	IGHV4D-23	2(2)	0	81 453 306	81 453 632	IGHV1D-46	0	0
Chr12	81 649 810	81 650 108	IGHV1D-17	5(3)	0	81 467 316	81 467 588	IGHV1D-42	0	0
Chr12	81 653 372	81 653 678	IGHV1D-16	0	0	81 481 949	81 482 241	IGHV16D-41	0	0
Chr12	81 661 001	81 661 301	IGHV1D-15	0	0	81 511 235	81 511 555	IGHV1D-39	2	0
Chr12	81 675 466	81 675 769	IGHV1D-12	2(2)	0	81 514 353	81 514 683	IGHV1D-38	0	0
Chr12	81 696 523	81 696 828	IGHV2D-11	1(0)	0	81 516 973	81 517 272	IGHV4D-37	0	0
Chr12	81 700 731	81 701 033	IGHV10D-9	0	0	81 519 248	81 519 565	IGHV12D-36	1	1
Chr12	81 705 764	81 706 066	IGHV10D-7	2(1)	0	81 548 116	81 548 425	IGHV16D-34	0	0
Chr12	81 717 198	81 717 494	IGHV6D-5	3(2)	0	81 550 863	81 551 152	IGHV16D-33	0	0
Chr12	81 737 673	81 737 973	IGHV1D-4	0	0	81 558 855	81 559 137	IGHV15D-32	0	0
						81 560 796	81 561 107	IGHV12D-31	0	0
						81 567 432	81 567 712	IGHV2D-29	0	0
						81 594 924	81 595 202	IGHV6D-27	1	0
						81 598 873	81 599 174	IGHV1D-25	0	0
						81 618 939	81 619 217	IGHV1D-22	11	2
						81 628 800	81 629 081	IGHV1D-21	2	0
						81 629 835	81 630 118	IGHV8D-20	0	0
						81 630 920	81 631 246	IGHV6D-19	0	0
						81 637 541	81 637 838	IGHV6D-18	0	0
						81 674 597	81 674 901	IGHV4D-13	0	0
						81 699 366	81 699 678	IGHV6D-10	0	0
						81 704 399	81 704 711	IGHV6D-8	0	0
						81 713 243	81 713 531	IGHV6D-6	3	1
						81 745 526	81 745 784	IGHV1D-3	2	0
						81 746 308	81 746 602	IGHV9D-2	1	0
						81 750 741	81 751 044	IGHV16D-1	6	2
						81 671 030	81 671 329	IGHV1D-14	0	0
						81 367 086	81 367 437	IGHV1D-67	0	0
						81 367 599	81 367 922	IGHV15D-66	0	0
						81 430 263	81 430 605	IGHV15D-53	0	0
						81 359 137	81 359 442	IGHV1D-70	0	0
						81 676 641	81 676 923	IGHV5D-11	0	0
Total				23(13)	0				53	10

(Continued)

TABLE 6 | Continued

	Functional genes					Pseudogenes				
Chrom	**Start**	**Stop**	**Name**	**SNPs number (NS)**	**Indel number**	**Start**	**Stop**	**Name**	**SNPs number**	**Indel number**
(B)										
Chr13	48 030 797	48 031 104	IGH IGHV10-47	2(1)	0	48 138 071	48 138 427	IGHV8-29	0	0
Chr13	48 034 515	48 034 814	IGHV8-46	1(0)	0	48 027 352	48 027 666	IGHV15-48	7	0
Chr13	48 054 181	48 054 484	IGHV1-42	0	0	48 046 874	48 047 207	IGHV9-45	1	0
Chr13	48 073 234	48 073 536	IGHV8-40	0	0	48 048 080	48 048 362	IGHV4-44	0	0
Chr13	48 077 115	48 077 414	IGHV1-39	1(1)	0	48 051 342	48 051 671	IGHV1-43	2	0
Chr13	48 082 080	48 082 391	IGHV16-37	3(3)	0	48 068 027	48 068 332	IGHV1-41	0	0
Chr13	48 093 298	48 093 597	IGHV1-36	0	0	48 079 966	48 080 277	IGHV16-38	1	0
Chr13	48 104 897	48 105 217	IGHV6-35	0	0	48 108 554	48 108 889	IGHV9-34	0	0
Chr13	48 109 683	48 109 994	IGHV14-33	0	0	48 146 928	48 147 231	IGHV2-27	3	0
Chr13	48 122 499	48 122 783	IGHV6-32	0	0	48 147 810	48 148 106	IGHV6-26	3	0
Chr13	48 127 168	48 127 466	IGHV6-31	0	0	48 157 816	48 158 168	IGHV9-24	5	2
Chr13	48 135 329	48 135 631	IGHV8-30	0	0	48 211 869	48 212 178	IGHV7-17	3	0
Chr13	48 145 742	48 146 047	IGHV2-28	7(3)	0	48 244 719	48 245 034	IGHV10-12	3	0
Chr13	48 148 981	48 149 284	IGHV11-25	0	0	48 245 867	48 246 172	IGHV8-11	2	0
Chr13	48 164 983	48 165 317	IGHV9-23	4(3)	0	48 254 879	48 254 570	IGHV16-9	0	0
Chr13	48 166 898	48 167 189	IGHV4-22	9(7)	0	48 279 548	48 279 841	IGHV1-7	0	0
Chr13	48 168 162	48 168 465	IGHV1-21	5(3)	0	48 280 387	48 280 681	IGHV4-6	4	1
Chr13	48 174 816	48 175 127	IGHV3-20	4(2)	0	48 316 059	48 315 761	IGHV6-3	3	0
Chr13	48 191 970	48 192 272	IGHV8-19	6(5)	0	48 339 577	48 339 885	IGHV4-1	0	0
Chr13	48 201 668	48 201 973	IGHV1-18	2(0)	0	48 076 216	48 076 475	IGHV13-39	4	0
Chr13	48 222 441	48 222 126	IGHV9-16	1(1)	1					
Chr13	48 223 844	48 223 544	IGHV9-15	4(2)	0					
Chr13	48 237 588	48 237 899	IGHV16-14	6(6)	0					
Chr13	48 243 688	48 243 987	IGHV1-13	3(0)	0					
Chr13	48 250 487	48 250 789	IGHV1-10	0	0					
Chr13	48 257 525	48 257 828	IGHV2-8	2(1)	0					
Chr13	48 307 972	48 307 670	IGHV8-5	7(2)	0					
Chr13	48 312 160	48 311 865	IGHV6-4	0	0					
Chr13	48 327 417	48 327 761	IGHV1-2	0	0					
Total				67(40)	1				41	3

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the www.imgt.org – accession numbers can be found within the manuscript. Any other data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SMa, AK, M-PL, and PB conceived the project and wrote the manuscript. SMa, AK, IS, and PB designed experiments. SMa, AK, SH-S, SA, SMo, DL, RC, IS, OS, JH, BK, M-PL, and PB performed data analysis. SMa, AK, DL, and IS provided resources. All authors contributed to manuscript revision, and read and approved the submitted version.

FUNDING

This work was supported by Xunta de Galicia (ED431C 2016/041), by the Norwegian Research Council (ImCom Grant 267644), and by USDA AFRI Grant 2DN70-2RDN7 (to IS). SMa also received funding from the People Programme Marie Curie Actions (FP7/2007-2013, REA Grant Agreement 600391). DL, RC, and PB were supported by institutional grants for Institut National de la Recherche Agronomique. RC and PB were supported by the ANR grant FishRNAvax.

ACKNOWLEDGMENTS

We are thankful to the entire IMGT® team in developing databases and tools. This work was granted access to the HPC@LR and to the High-Performance Computing (HPC) resources of the Centre Informatique National de l'Enseignement Supérieur (CINES) and to Très Grand Centre de Calcul (TGCC) of the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) under the allocation [036029] (2010–2019) made by GENCI (Grand Equipement National de Calcul Intensif). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02541/full#supplementary-material

Figure S1 | Alignment of the D-GENE-UNIT sequences of the IGHD (diversity) genes located upstream of the IGHM (locus A) and IGHMD (locus B) genes of *Salmo salar* (Salsal) and *Oncorhynchus mykiss* (Oncmyk) **(A)** and located upstream of IGHT genes **(B)**. Genes of the locus B genes are identified by the letter D which follows the gene number. Labels are according to the D-GENE prototype (IMGT Scientific chart > 1. Sequence and 3D structure identification and description > IMGT prototypes table > D-GENE)[1].

Figure S2 | Alignment of the J-REGION amino acid sequences of the IGHJ (joining) genes located upstream of the IGHM or IGHT (locus A) and IGHMD or IGHTD (locus B) genes of Salmo salar (Salsal) and *Oncorhynchus mykiss* (Oncmyk). Genes of the locus B are identified by the letter D which follows the gene number. Labels are according to the J-GENE prototype (IMGT Scientific chart > 1. Sequence and 3D structure identification and description > IMGT prototypes table > J-GENE)[1]. The highly conserved FDYWGKGTXVT motif is pink highlighted and those residues that deviated from it are in red.

REFERENCES

1. Lefranc M-P, Lefranc G. *The Immunoglobulin FactsBook*. London, UK: Academic Press (2001).
2. Lefranc M-P, Lefranc G. *The T Cell Receptor FactsBook*. London, UK: Academic Press (2001).
3. Lefranc M-P. Immunoglobulin and T cell receptor genes: IMGT® and the birth and rise of immunoinformatics. *Front Immunol.* (2014) 5:22. doi: 10.3389/fimmu.2014.00022
4. Lefranc M-P. Nomenclature of the human immunoglobulin genes. *Curr Protoc Immuno.* (2001) Appendix1: Appendix 1P.1-A1P37.
5. Lefranc M-P. Nomenclature of the human T cell receptor genes. *Curr Protoc Immunol.* (2001) Appendix 1:Appendix 1O. 1-A.1O.23.
6. Lefranc M-P. WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report. *Immunogenetics.* (2007) 59:899–902. doi: 10.1007/s00251-007-0260-4
7. Lefranc M-P. WHO-IUIS nomenclature subcommittee for immunoglobulins and T cell receptors report August 2007, 13th International Congress of Immunology, Rio de Janeiro, Brazil. *Dev Comp Immunol.* (2008) 32:461–3. doi: 10.1016/j.dci.2007.09.008
8. Matsunaga T, Chen T, Törmänen V. Characterization of a complete immunoglobulin heavy-chain variable region germ-line gene of rainbow trout. *Proc Natl Acad Sci USA.* (1990) 87:7767–71. doi: 10.1073/pnas.87.19.7767
9. Andersson E, Törmänen V, Matsunaga T. Evolution of a VH gene family in low vertebrates. *Int Immunol.* (1991) 3:527–33. doi: 10.1093/intimm/3.6.527
10. Lee MA, Bengtén E, Daggfeldt A, Rytting AS, Pilström L. Characterisation of rainbow trout cDNAs encoding a secreted and membrane-bound Ig heavy chain and the genomic intron upstream of the first constant exon. *Mol Immunol.* (1993) 30:641–8. doi: 10.1016/0161-5890(93)90075-M
11. Andersson E, Matsunaga T. Complete cDNA sequence of a rainbow trout IgM gene and evolution of vertebrate IgM constant domains. *Immunogenetics.* (1993) 38:243–50. doi: 10.1007/BF00188800
12. Roman T, Charlemagne J. The immunoglobulin repertoire of the rainbow trout (Oncorhynchus mykiss): definition of nine Igh-V families. *Immunogenetics.* (1994) 40:210–6. doi: 10.1007/BF00167081
13. Roman T, Andersson E, Bengtén E, Hansen J, Kaattari S, Pilström L, et al. Unified nomenclature of Ig VH genes in rainbow trout (*Oncorhynchus mykiss*): definition of eleven VH families. *Immunogenetics.* (1996) 43:325–6. doi: 10.1007/s002510050072
14. Brown GD, Kaattari IM, Kaattari SL. Two new Ig VH gene families in *Oncorhynchus mykiss*. *Immunogenetics.* (2006) 58:933–6. doi: 10.1007/s00251-006-0149-7
15. Solem ST, Hordvik I, Killie JA, Warr GW, Jørgensen TO. Diversity of the immunoglobulin heavy chain in the Atlantic salmon (Salmo salar L.) is contributed by genes from two parallel IgH isoloci. *Dev Comp Immunol.* (2001) 25:403–17. doi: 10.1016/S0145-305X(01)00008-8
16. Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Messmer A, et al. Evolution of duplicated IgH loci in Atlantic salmon, Salmo salar. *BMC Genom.* (2010) 11:486. doi: 10.1186/1471-2164-11-486
17. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* (2015) 43:D413–22. doi: 10.1093/nar/gku1056
18. Li S, Lefranc M-P, Miles JJ, Alamyar E, Giudicelli V, Duroux P, et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat Commun.* (2013) 4:2333. doi: 10.1038/ncomms3333
19. Aouinti S, Malouche D, Giudicelli V, Kossida S, Lefranc M-P. IMGT/HighV-QUEST statistical significance of IMGT clonotype (AA) diversity per gene for standardized comparisons of next generation sequencing immunoprofiles of immunoglobulins and T cell receptors. *PLoS ONE.* (2015) 10:e0142353. doi: 10.1371/journal.pone.0142353
20. Aouinti S, Giudicelli V, Duroux P, Malouche D, Kossida S, Lefranc M-P. IMGT/statclonotype for pairwise evaluation and visualization of NGS IG and TR IMGT clonotype (AA) diversity or expression from IMGT/HighV-QUEST. *Front Immunol.* (2016) 7:339. doi: 10.3389/fimmu.2016.00339
21. Brochet X, Lefranc M-P, Giudicelli V. 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.* (2008) 36:W503–8. doi: 10.1093/nar/gkn316
22. Monod MY, Giudicelli V, Chaume D, Lefranc M-P. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics.* (2004) 20 (Suppl_1):i379–85. doi: 10.1093/bioinformatics/bth945
23. Ehrenmann F, Lefranc M-P. IMGT/DomainGapAlign: the IMGT® tool for the analysis of IG, TR, MH, IgSF, and MhSF domain amino acid polymorphism. *Methods Mol Biol.* (2012) 882:605–33. doi: 10.1007/978-1-61779-842-9_33
24. Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol.* (2003) 27:55–77. doi: 10.1016/S0145-305X(02)00039-3
25. Lefranc M-P, Pommié C, Kaas Q, Duprat E, Bosc N, Guiraudou D, et al. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol.* (2005) 29:185–203. doi: 10.1016/j.dci.2004.07.003
26. Alamyar E, Duroux P, Lefranc M-P, Giudicelli V. IMGT® 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.* (2012) 882:569–604. doi: 10.1007/978-1-61779-842-9_32
27. Alamyar E, Giudicelli V, Shuo L, Duroux P, Lefranc M-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* (2012) 8:26.
28. Fillatreau S, Six A, Magadan S, Castro R, Sunyer JO, Boudinot P. The astonishing diversity of Ig classes and B cell repertoires in teleost fish. *Front Immunol.* (2013) 4:28. doi: 10.3389/fimmu.2013.00028
29. Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. *Proc Natl Acad Sci USA.* (2005) 102:6919–24. doi: 10.1073/pnas.0500027102
30. Zhang Y-A, Salinas I, Li J, Parra D, Bjork S, Xu Z, et al. IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat Immunol.* (2010) 11:827–35. doi: 10.1038/ni.1913
31. Hordvik I. Identification of a novel immunoglobulin delta transcript and comparative analysis of the genes encoding IgD in Atlantic

salmon and Atlantic halibut. *Mol Immunol.* (2002) 39:85–91. doi: 10.1016/S0161-5890(02)00043-3
32. Krasnov A, Jørgensen SM, Afanasyev S. Ig-seq: deep sequencing of the variable region of Atlantic salmon IgM heavy chain transcripts. *Mol Immunol.* (2017) 88:99–105. doi: 10.1016/j.molimm.2017.06.022
33. Magadan S, Jouneau L, Boudinot P, Salinas I. Nasal vaccination drives modifications of nasal and systemic antibody repertoires in rainbow trout. *J Immunol.* (2019) *J Immunol.* 203:1480–92. doi: 10.4049/jimmunol.1900157
34. Castro R, Jouneau L, Pham H-P, Bouchez O, Giudicelli V, Lefranc M-P, et al. Teleost fish mount complex clonal IgM *and* IgT responses in spleen upon systemic viral infection. *PLoS Pathog.* (2013) 9:e1003098. doi: 10.1371/journal.ppat.1003098
35. Magadán-Mompó S, Sánchez-Espinel C, Gambón-Deza F. Immunoglobulin heavy chains in medaka (*Oryzias latipes*). *BMC Evol Biol.* (2011) 11:165. doi: 10.1186/1471-2148-11-165
36. Quillet E, Dorson M, Leguillou S, Benmansour A, Boudinot P. Wide range of susceptibility to rhabdoviruses in homozygous clones of rainbow trout. *Fish Shellfish Immunol.* (2007) 22:510–9. doi: 10.1016/j.fsi.2006.07.002
37. Diter A, Quillet E, Chourrout D. Suppression of first egg mitosis induced by heat shocks in the rainbow trout. *J Fish Biol.* (1993) 42:777–86. doi: 10.1111/j.1095-8649.1993.tb00383.x
38. Palti Y, Genet C, Luo M-C, Charlet A, Gao G, Hu Y, et al. A first generation integrated map of the rainbow trout genome. *BMC Genom.* (2011) 12:180. doi: 10.1186/1471-2164-12-180
39. Palti Y, Gao G, Miller MR, Vallejo RL, Wheeler PA, Quillet E, et al. A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol Ecol Resour.* (2014) 14:588–96. doi: 10.1111/1755-0998.12204

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

MINI REVIEW
published: 19 November 2019
doi: 10.3389/fimmu.2019.02600

Keeping Allergen Names Clear and Defined

Sanny K. Chan[1], Anna Pomés[2], Christiane Hilger[3], Janet M. Davies[4,5], Geoffrey Mueller[6], Annette Kuehn[3], Andreas L. Lopata[7], Gabriele Gadermaier[8], Marianne van Hage[9], Monika Raulf[10] and Richard E. Goodman[11] on behalf of the WHO/IUIS Allergen Nomenclature Sub-Committee*

[1] Division of Allergy and Immunology, Department of Pediatrics, National Jewish Health, Denver, CO, United States, [2] INDOOR Biotechnologies, Inc. Charlottesville, VA, United States, [3] Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg, [4] Centre for Children's Health Research, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia, [5] Metro North Hospital and Health Service, Brisbane, QLD, Australia, [6] National Institute of Environmental Health Sciences, Durham, NC, United States, [7] Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia, [8] Department of Biosciences, University of Salzburg, Salzburg, Austria, [9] Division of Immunology and Allergy, Department of Medicine Solna, Karolinska Institutet and University Hospital, Solna, Sweden, [10] Institute of Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-Universitat Bochum, Bochum, Germany, [11] Food Allergy Research and Resource Program, Deptartment of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, OR, United States

OPEN ACCESS

Edited by:
Heiko Mühl,
Goethe University Frankfurt, Germany

Reviewed by:
Peisong Gao,
Johns Hopkins University, United States
Eva Untersmayr,
Medical University of Vienna, Austria

***Correspondence:**
Richard E. Goodman
Rgoodman2@unl.edu

Specialty section:
This article was submitted to Inflammation, a section of the journal Frontiers in Immunology

Received: *29 July 2019*
Accepted: *21 October 2019*
Published: *19 November 2019*

Citation:
Chan SK, Pomés A, Hilger C, Davies JM, Mueller G, Kuehn A, Lopata AL, Gadermaier G, van Hage M, Raulf M and Goodman RE (2019) Keeping Allergen Names Clear and Defined. Front. Immunol. 10:2600. doi: 10.3389/fimmu.2019.02600

The World Health Organization/International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee was established in 1986 by leading allergists to standardize names given to proteins that cause IgE-mediated reactions in humans. The Sub-Committee's objective is to assign unique names to allergens based on a critical analysis of confidentially submitted biochemical and clinical data from researchers, often prior to publication to preserve consistency. The Sub-Committee maintains and revises the database as the understanding of allergens evolves. This report summarizes recent developments that led to updates in classification of cockroach group 1 and 5 allergens to animal as well as environmental and occupational allergens. Interestingly, routes, doses, and frequency of exposure often affects allergenicity as does the biochemical properties of the proteins and similarity to self and other proteins. Information required by the Sub-Committee now is more extensive than previously as technology has improved. Identification of new allergens requires identification of the amino acid sequence and physical characteristics of the protein as well as demonstration of IgE binding from subjects verified by described clinical histories, proof of the presence of the protein in relevant exposure substances, and demonstration of biological activity (skin prick tests, activation of basophils, or mast cells). Names are assigned based on taxonomy with the abbreviation of genus and species and assignment of a number, which reflects the priority of discovery, but more often now, the relationships with homologous proteins in related species.

Keywords: allergen nomenclature, WHO/IUIS, taxonomy, diagnostic, airway, food, dermal

INTRODUCTION

Advances in molecular biology, recombinant protein technology, methods of genomic, and proteomic analysis, structural biology, and high throughput screening have led to a greater understanding of allergenic proteins over the past 30 years. Identification of new allergens and improved characterization dictate the need for revision of some allergens and updated requirements for allergen nomenclature.

In the 1980's key clinical and experimental allergists led by D Marsh at Johns Hopkins University, USA, L Goodfriend (McGill University, Canada), TP King (Rockefeller University, USA), H Lowenstein (University of Copenhagen (Denmark) and TAE Platts-Mills (University of Virginia, USA) framed basic rules for naming allergenic proteins in common allergenic sources (1). Over 2 years they agreed to a strategy of naming allergens based on taxonomy with identification using the first three letters of the genus, the first letter of the species and a number for order of identification. The first allergens were mostly inhalant proteins from pollens of weeds, trees and grasses and a few from animal dander, fungi, and venom proteins from stinging insects. The committee recognized that a systemic standardized nomenclature was needed for consistent identification in scientific publications. The standardized naming of allergens has become a formalized process where allergen names are assigned by the WHO/IUIS Allergen Nomenclature Sub-Committee after a defined submission process that includes data on IgE binding to identified novel target proteins or glycoproteins with the most recent major revision occurring in 2018 (2).

Allergen names are assigned with the first 3 or sometimes 4 letters of the genus, one or sometimes two letters for the species followed by an Arabic numeral, based on order of discovery (2). However, the same Arabic number is often used for conserved protein families in related taxa. Often species have evolved by duplication of genes with mutations including additions or deletion of nucleotides to produce alternative proteins of similar function, resulting in the occurrence of isoallergens as well as isoforms or variants within the individual organism or in related organisms of the same species. The isoallergens are designated by the addition of two digits after the decimal in the number and isoforms or variants by addition of two more digits (e.g., Amb a 1.0101). Since the evolutionary steps involved in generating each new isoform or variant is usually not known, the committee bases designations on the percent identity of amino acids compared to the first identified sequence. Sequences within ∼67% identity to the original allergen are designated as isoallergens and sequences differing by <90% identity are isoform or variants.

Expression of recombinant proteins for IgE tests and characterization has led to more accurate identification. Changes to the nomenclature have attempted to incorporate an increased understanding of allergens. Higher resolution studies and isolation techniques of previously defined allergens now shows some to be composed of multiple subunits from different genes; Fel d 1, or multimers; collagen, vicilins, and glycinins. This manuscript focuses on recent challenges and underlying reasons for name assignment by the WHO/IUIS Nomenclature Sub-Committee. We refer those working on allergen identification to **Figure 1**, from our 2018 publication with references to specific sections in this paper to explain the ideal process (2). The science of allergen determination has improved since 1990, for characterization of allergenic sources, protein sequences, post-translational modification determination, improved methods for measuring IgE binding, mediator release and measurements of specific bioactivity (skin tests and basophil assays). Scientists striving to characterize allergens should be aware of and using techniques appropriate to the types of allergens including those of foods, contact allergens, aeroallergens and venoms or salivary sources of allergy.

CHALLENGES FOR FUTURE ALLERGEN NAMES

Repeated Structure of Group 1 Cockroach Allergens Is Not Reflected in Current Allergen Nomenclature

Cockroach allergens from group 1 were difficult to name using the current WHO/IUIS paradigm. Cloning of the DNA encoding for Bla g 1 allergen revealed the existence of multiple DNA repeats encoding approximately 100 amino acids each (3). The molecular structure of Bla g 1 showed that two tandem amino acid repeats form a spherical fold (4). A recent analysis of the *Blatella germanica* genome (5) suggests there are 5 separate gene products, and each produce 1 to 5, partial or full repeats of the Bla g 1 structural motif. In total, there may be 9 other Bla g 1-like allergens, with 60–90% identity to the Bla g 1 structure. Prior to the genomic data, it was unclear whether the different Bla g 1 polymorphisms derived from a different gene. In addition, there was not a precedent for an allergen with this kind of structure and genomic origin.

Besides cockroaches, it is worthwhile to define new rules for assigning a nomenclature to this type of allergen, given the possibility that Bla g 1-like homologs exist in other sources, such as locusts and mealworms, recently included as food sources. Allergens belong to a selected group of protein families (6, 7), and it is likely that any Bla g 1-like homologs in these sources could also be an allergen. Indeed, a homolog of Bla g 1 was noted in the genome of *Locusta migratoria* (8). Further examination reveals that there are two genes with 2 and 3 copies of the Bla g 1-like structure, respectively. Other coleopteran species have been proposed as a substitute protein source in place of mammalian meat. A search of genomes of related species revealed that *Tribolium castaneum*, has one gene with 9 repeats similar to the Bla g 1- structure. This repeated sequence structure is likely to occur in other beetles and there are reported occupational and food allergies reported from a larval beetle (mealworm) of *Tenebrio molitor* (9). Homologs of Bla g 1 can be found in many other insect species as well (10). So far, the protein(s) responsible for reactions to mealworm are not yet identified except for tropomyosin with homology to shrimp proteins. The structural complexity of allergens like Bla g 1 are not reflected in the current allergen nomenclature. For similar cases, the

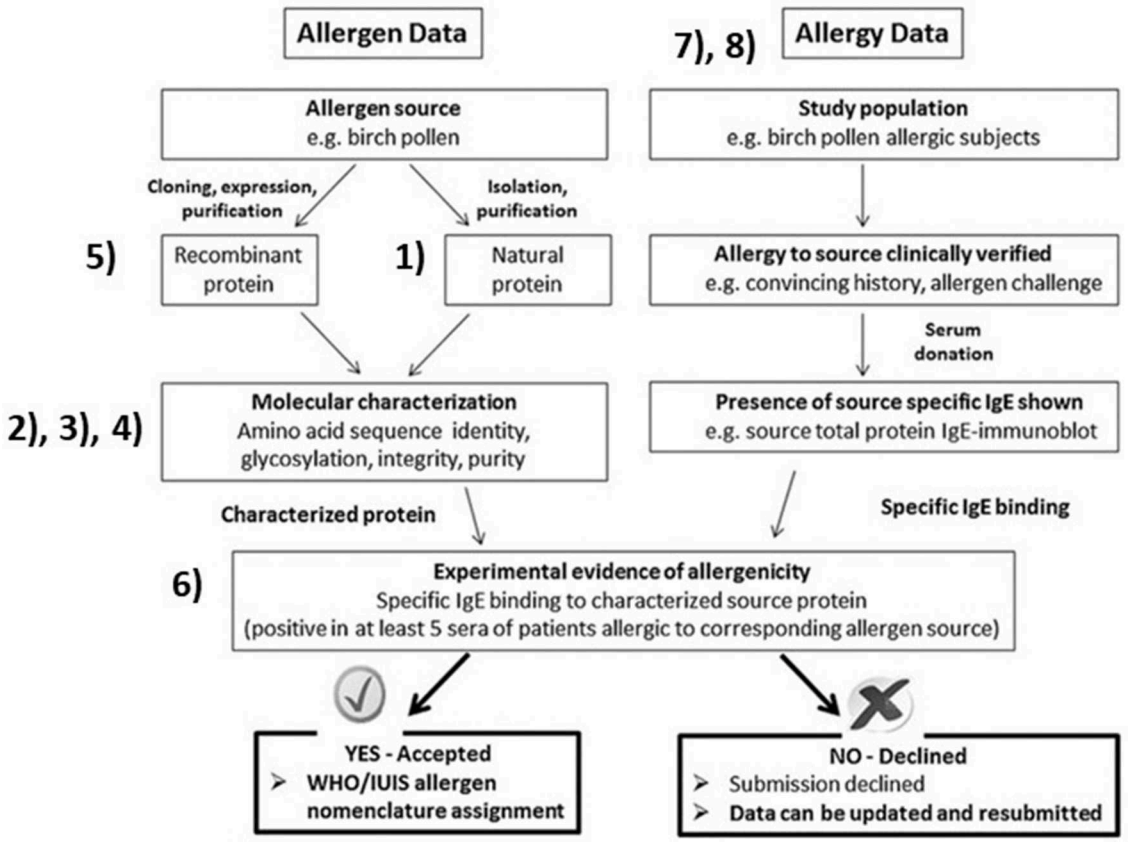

FIGURE 1 | Overall allergen characterization for submission to the WHO/IUIS Allergen Nomenclature Sub-Committee for possible name recognition. Figure with copyright allowance from Pomés et al. (2). Numbers are inserted beside key blocks, identifying topical areas and challenges related to our evaluations. (1) Nearly identical repeated sequences confuse characterization of the natural protein. (2) Amino acid identity matches across distant taxonomic relations are unlikely to represent cross-reactive proteins. (3) Native allergens of co-valently linked independent proteins such as Fel d 1 and Ory 3. (4) Carbohydrate IgE binding epitopes (CCD, alpha-gal) are not (yet) named in this system. (5) Studies testing cloned, recombinant proteins without native protein bridging can leave doubts. (6) Biological evidence, either basophil histamine release or activation, or skin prick test with pure protein. (7) Aeroallergen identities in clinical tests are typically not verified in allergen diagnostic testing due to lack of available components. (8) Occupational and food allergen testing typically involve few study subjects.

Allergen Nomenclature Sub-Committee discussed the possibility (not yet approved) of using capital letters behind the allergen name to signify either the different genes from which the allergen originated, or different duplexes derived from a gene (i.e., Bla g 1.0101A). Knowledge of the origin of unique allergen structures may help to define a more informative allergen nomenclature for these unusual allergens.

Group 5 Cockroach Allergens Are Glutathione S-Transferases (GSTs), Consideration of the Degree of Amino Acid Identity

In recent years the WHO/IUIS Sub-Committee has used similar allergen numbers for homologous proteins across related taxonomic groups when possible. Glutathione *S*-transferases (GST) are common enzymes present in multiple organisms from different taxonomic groups. One of the main functions of GSTs is detoxification by catalyzing the addition of a glutathione (GSH) molecules to other commonly toxic compounds that are ultimately removed from the cell (11, 12). Nine groups of allergens are recognized as GSTs and are listed in the WHO/IUIS Allergen Nomenclature database (www.allergen.org), including those from cockroaches (Bla g 5 and Per a 5), mites (Der p 8, Der f 8 and Blo t 8), helminths (Asc s 13, Asc l 13), fungi (Alt a 13), and plants (Bet v 8) (**Figure 2A**).

In general, homologous allergens from the same species are included in the same allergen group as isoallergens (e.g., Cyn d 1.0101 and Cyn d 1.0201 in **Table S1**) if they share a suggested threshold of 67% amino acid identity (2). At 67% identity, IgE cross-reactivity is possible. However, naming additional GSTs from cockroach in group 5, has been challenging due to their low sequence identities (down to 18% identity). For those identified as binding IgE, they could be given a different name and number (**Figure 2B**). We considered these enzymes belong to different protein classes (sigma, delta, theta, mu) as functionally related. The Bla g 5 from German cockroach (*Blattella germanica*) was the first cockroach GST to be identified (O18598), and it is a

A GST allergens:

	1	2	3	4	5	6	7	8	9
1: Alt a 13.0101-AY514673-Arora	100.00	10.93	16.40	16.40	17.71	16.76	16.24	16.13	16.27
2: Bet v 8.0101-KF246508-Bohle	10.93	100.00	20.21	20.21	18.64	17.58	14.00	13.71	12.85
3: Per a 5.0101-Delta-MG255130- Tungtrongchitr	16.40	20.21	100.00	99.54	20.99	18.78	21.74	16.67	20.44
4: Per a 5.0102-Delta-AEV23867-Chew	16.40	20.21	99.54	100.00	20.99	18.78	21.74	16.67	20.44
5: Asc s 13.0101-Sigma-X75502-Leibau	17.71	18.64	20.99	20.99	100.00	37.19	23.15	26.15	25.14
6: Bla g 5.0101-Sigma-O18598-Chapman	16.76	17.58	18.78	18.78	37.19	100.00	24.24	29.17	28.49
7: Blo t 8.0101-Mu-GQ398117-Caraballo	16.24	14.00	21.74	21.74	23.15	24.24	100.00	35.48	34.36
8: Der p 8.0101-Mu-S75286-Baldo	16.13	13.71	16.67	16.67	26.15	29.17	35.48	100.00	70.56
9: Der f 8.0101-Mu-KC305499-An	16.27	12.85	20.44	20.44	25.14	28.49	34.36	70.56	100.00

B Cockroach GSTs:

	1	2	3	4	5	6	7
1: Sigma-German CR-Bla g 5.0101-O18598-Chapman	100.00	99.50	18.89	20.11	20.11	18.99	18.99
2: Sigma-German CR-EF202178-Yong	99.50	100.00	18.75	20.57	20.57	19.43	19.43
3: Theta-German CR-AEV23882-Chew	18.89	18.75	100.00	57.21	57.67	58.14	58.60
4: Delta-American CR-Per a 5.0101-MG255130- Tungtrongchitr	20.11	20.57	57.21	100.00	99.54	81.48	81.94
5: Delta-American CR-Per a 5.0102-Delta-AEV23867-Chew	20.11	20.57	57.67	99.54	100.00	81.94	82.41
6: Delta-German CR-ABX57814-Yong	18.99	19.43	58.14	81.48	81.94	100.00	98.61
7: Delta-German CR-AEV23880-Chew	18.99	19.43	58.60	81.94	82.41	98.61	100.00

Background color code according to percentage of amino acid identities:

Black: 100%, or Dark grey: Medium grey >70%: Light grey > 30% Very light grey: > 20 %: White: < 20 %.

FIGURE 2 | Homologous protein GST allergens are currently categorized as isoallergens with a threshold of 67% amino acid identity although they may belong to different protein classes (sigma, delta, theta, mu). Percent identities are shown. **(A)** Nine groups of GST allergens from different taxa to understand sequence identities across sources. **(B)** Cockroach GSTs within German and American cockroaches with known allergens and genomic sequences with comparison of amino acid sequence identities.

sigma class GSTs (11). Additional GSTs from German cockroach have been identified, though these have not been submitted to the WHO/IUIS committee with proof of IgE binding: another sigma GST (EF202178), a theta GST (AEV23882) and two delta GSTs (ABX57814 and AEV23880). Two GSTs from the American cockroach (*Periplaneta americana*) were identified as binding IgE, and they belong to the delta class of GSTs [MG255130 (13) and AEV23867] and share 99.54% amino acid identity. When these GSTs were submitted, we decided to include them as Per a group 5 proteins even though their identity to Bla g 5 was only 19% identity with Bla g 5. They were named Per a 5.0101 and Per a 5.0102. If other cockroach GSTs are submitted to the database in the future, they will be assigned as isoallergens if fitting within the 67% identity range to the respective group 5 allergens.

GSTs from other species are and have been assigned to other numbers as seen in **Figure 2A** (Alt a 13, Bet v 8, Asc a 13, Blo t 8, Der p 8 amd Der f 8. Isoallergens of each of these will be assigned, based on sequence comparison, to currently named members of the respective source species. The committee now prefers to retain primary allergen numbers the same across related taxonomic groups, but that is not always possible. It is important to note that while enzyme functions may be retained with great diversity in amino acid sequences, the ability for IgE to bind across those species and meet the criteria of allergenicity is not retained.

Covalently Linked Dimeric Allergens and Allergen Nomenclature

Mammalian allergens of the secretoglobin family are heterodimers, encoded by separate genes. Secretoglobins (SCGB) represent a large protein family found in mammals and marsupials. Common characteristics are secretion in many body fluids, small size, alpha-helical, and dimeric structure creating a hydrophobic binding pocket (14, 15). Major allergens from cat and rabbit are Fel d 1 and Ory c 3. Both are glycosylated heterodimers linked by three disulfid bridges (16, 17). Fel d 1 also forms tetramers composed of two heterodimers. Each protein chain of the heterodimer is encoded by a single gene and both are needed to compose the natural molecule and may be needed for achieving full allergenic activity. The current allergen nomenclature system did not foresee linked heteromeric allergens and the names allocation for Fel d 1 and Ory c 3 were not consistent. Whereas, Fel d 1 is listed as Fel d 1.0101 for the dimer, composed of a chain 1 and a chain 2, products of two separate genes (17), Ory c 3 is listed as 2 isoallergens with Ory c 3.A.0101, chain A (lypophillin CL2) and Ory c 3.B.0101, chain B (lypophillin AL). The nomenclature for Ory c 3 suggests that both isoallergens have to be considered together as a full allergen (17). The two proteins do not exist naturally as monomers and allergenicity can not be assessed individually. The Sub-Committee recently renamed the Fel d 1 proteins as Fel d 1.A.0101 for chain 1 and Fel d 1.B.0101 for chain 2, consistent with the rabbit allergen.

Other examples are recently added allergens, barramundi collagen Lat c 6 and salmon allergen Sal s 6 proteins. Collagen is a triple-stranded rope-like coiled structure, composed of 3 protein chains. Each chain has been given an isoallergen name, despite the fact that they do not exist as monomers. Currently it is only possible to evaluate separate IgE binding if using recombinant monomeric proteins and thus it is not possible to assess the allergenicity of each of the entire native proteins. The committee may reevaluate this policy in the future.

Carbohydrates, New Allergen Epitopes Not Represented in the Allergen Nomenclature Database

Cross-reactive carbohydrate determinants (CCDs) are carbohydrate epitopes carrying an α-1,3-fucose and, or β 1,2 xylose on complex carbohydrates of some glycoproteins are the main targets of IgE recognition. They are present on many plant and insect proteins and responsible for extensive IgE cross-reactivity between different allergen sources (18). However, there is a general consensus that those glycans do not trigger noticeable clinical symptoms. Two other carbohydrate groups have been associated with food allergy and anaphylaxis (19). Galactose-α-1,3-galactose (α-gal), a carbohydrate present on mammalian proteins, has been reported to trigger delayed severe allergic reactions to red meat and to induce acute allergic reactions upon injection of therapeutic antibodies carrying α-gal (20, 21). Galacto-oligosaccharides (GOS) are heat-stable, non-digestable carbohydrates present in milk formula and as supplement in different beverages. Food allergy to GOS are reported in Singapore, Vietnam and Japan (19). The structure and characterization of carbohydrates that are the reported targets of IgE binding have not been demonstrated, although they have been assumed to be related to the synthesis of GOS complexes. Recent work failed to identify the structures or link them to activation of basophils (22). In contrast to α-gal, the GOS carbohydrates have not been linked to a protein backbone (19). So far, only allergens representing protein epitopes have been named by the WHO/IUIS Nomenclature Sub-Committee. The committee is currently considering whether there is enough definition of carbohydrate epitopes to provide unique names and whether names would be scientifically useful.

Updating Existing Allergen Entries Named From Partial Allergen Sequences

Some allergens are characterized based on peptide sequences from native molecules. In some cases, full-lengh sequences were not available, but are based on peptide matches to genomic sequences. That was the case for fish allergens, enolase Sal s 2 and aldolase Sal s 3 from salmon (*Salmon salar*) based on public genomic sequences (23). For the salmon allergens Sal s 2 and Sal s 3, complete sequences were attributed without testing of the allergenic properties of the corresponding proteins. In the future, we recommend that an alignment of the complete sequences of natural or recombinant protein be used to determine IgE binding be compared to the genome sequence, or at the least, a test of the recombinant protein be comparison to the native counterpart. Importantly, different isoforms and variants can exist in the same source and although highly similar, those proteins can differ in their allergenic potency (24). Verified information will be added to the existing database entry in order to further improve its quality.

Numbering of Allergens

The allergen nomenclature system was originally based on the order of allergen discovery to provide the allergen number, but the commitee recognized the conservation of sequences and structures are important. But sometimes, very similar allergens from related species receive different numbers (e.g., Fel d 2 and Can f 3) (25). It would be disruptive to change allergen names after a number of publications have used established allergen names and only allergen numbers have been revised and changed in a few cases (26).

OVERVIEW OF OUTDOOR ENVIRONMENTAL AEROALLERGENS

Most outdoor aeroallergens are derived from grass, weed and tree pollen or fungal spore sources. Although Blackley discovered that pollen caused hay fever in the nineteenth century (27), it was not until allergenic fractions were purified by chromatography from ryegrass pollen in the 1960s (28) and cloned in 1990s (29) that the first pollen allergens were characterized. Because the sources are typically not recognized due to small sizes of pollen (10–50 microns) and mold spores (2–50 microns), there is often uncertainty regarding the causal link between exposure to a given species and allergic reactions. This is specifally difficult given the cross-reactivities that exists among allergens from different species.

The IUIS Allergen Nomenclature Subcommittee recognize 43, 55, and 45 allergens for pollen of grasses, trees and weeds, respectively. Another four pollen allergens of herbaceous plants included; Lig v 1 of *Ligustrum vulgare* (Common privet), Syr v 1 and Syr v 3 of *Syringa vulgaris* (Lilac) and Hum j 1 of *Humulus japonicus* (Japanese hop).

Grass pollen allergens belong to 12 different protein families. The most abundant proteins of mature pollen that elicits a high frequency of IgE reactivity are the major group 1 beta-expansin allergens. The purpose of beta-expansins is to degrade cell walls and allow pollen tube extension which is essential for fertilization. A number of IgE and T cell epitope regions are found in close proximity with the enzymatic domain and specifically the HFD motif of the N terminal domain (30–32). While allergens of temperate grass pollens were characterized in 1990s, allergens of subtropical grass pollens continue to emerge. Grass pollen group 1 allergens among the Pooideae temperate grasses share between 84 and 91% amino acid identity, while subtropical grasses of Panicoideae, Chloridoideae, and Oryzoideae share between 49 and 86% identity (**Table S1**). Separate gene loci encode multiple Poaceae group 1 isoforms. The pollen expansin proteins show significant sequence diversity, but they apparently share similar functions exemplified by Cyn d 1 and Sor h 1 isoform diversity (**Table S1**). While research shows that there is IgE and T cell cross-reactivity between group 1 and group 5 grass pollen allergens across broad taxa, their heterogeneity contributes to the diversity of immune recognition (33, 34). The most recently listed (grass pollen allergen is Uro m 1 of the subtropical Panicoideae grass *Urochloa mutica* (Para grass). Group 2 and 3 grass pollen allergens share approximately 30% identity with the carboxyl domain of the beta-expansin protein family. Consequently, grass pollen group 2/3 pollen allergens are smaller in size (10–12 kDa), but their function is unknown. Sor h 2 and Ory s 2 are described

in subtropical species, but their clinical importance is less well-demonstrated. Several isoforms and multiple variants exist for Group 5 allergens (e.g., for Phl p 5) and these share between 53 and 78% identity between six species for which these allergens have been characterized (**Table S2**). Group 5 allergens are absent from subtropical grass pollen.

Weed pollen allergens arise from 16 species including ragweed (*Ambrosia artemisiifolia*), mugwort (*Artemisia vulgaris*), pellitory (*Parietaria judaica*), amaranth, thistle (*Salsola kali*), and sunflower (*Helianthus annuus*) and belong to 10 allergen families. Clinically important allergens of the Asteraceae family include pectate lyases (Amb a 1 and Art v 6), defensin-like proteins (Amb a 4 and Art v 1), non-specific lipid transfer proteins (Art v 3 and Par j 2), whereas allergens of the Amaranthaceae family include Ole e 1-like proteins and Che a 1, and pectin methylesterase (Sal k 1).

Trees from multiple orders; Fagales; for example, Birch (*Betula verrucosa*), Lamiales; Ash (*Fraxinus excelsior*) and Olive (*Olea europaea*), Proteales; Plane (*Plantanus acerifolia*) and Cupressales; Japanese Cedar (*Cryptomeria japonica*), encompassing 21 species contribute allergens belonging to 18 protein families. Clinically important allergens include Pathogenesis Related protein-10 (Bet v 1, Ole e 1-like protein); Fra e 1, polygalacturonase; Cry j 2 and Pla a 2, and pectate lyase; Cry j 1. Profilins are ubiquitous and highly conserved in sequence and structure (e.g., Phl p 12, Amb a 8 and Bet v 2) and polcalcins (Phl p 7, Amb a 9, and Ole e 8) are pan-allergens common to grass, weed and tree pollen, though these allergens are generally less potent than PR-10 proteins.

A variety of fungi produce the 112 listed WHO/IUIS allergens from as many as 36 protein families. While fungal hyphae or spores may be used as food or drug sources, many are common sources of indoor and outdoor allergy and a small number colonize human tissues including skin and lungs. Spores of 15 Ascomycota species belonging to diverse fungal families; Aspergillaceae, Cladosporiaceae, Didymellaceae, Nectriaceae, and Pleosporaceae produce 72 airborne outdoor fungal spore allergens. There are 23 allergens listed for *Aspergillus fumigatus* including the ribotoxin (Asp f 1), peroxisomal protein (Asp f 3), metalloprotease (Asp f 5), cyclophilin (Asp f 11), vacuolar serine protease (Asp f 18), and enolase (Asp f 22), which are all major allergens. *Alternaria alternata* spores, one of commonly recognized fungal spores contains 12 listed allergens including its major allergens (Alt a 1), and, a glutathione S transferase (Alt a 13).

OVERVIEW OF OCCUPATIONAL ALLERGENS FROM FOOD SOURCES

Occupational sensitization has been observed for centuries in historical and medical writings. The first description of baker's asthma was probably by Ramazini around 1700; the first report of allergic responses to fish was in 1937 by De Besche relating to a fisherman who developed allergic symptoms when handling codfish (35). Sensitization to allergens at the workplace occurs usually via the skin or the respiratory tract. Resulting cutaneous and respiratory diseases are a global health problem, with an estimate that 25% of all asthma cases related to work (36).

The most common workplace allergens are relatively high-molecular-weight proteins derived from cereal flour, natural rubber latex, wood dust, livestock and laboratory animals, seafood, industrial enzymes and mold (37). Animal and vegetable high-molecular-weight proteins present in the aerosol and often in aerosolized foods during food processing are the main inhalant allergen sources. This type of IgE-mediated respiratory food allergy usually does not generate any symptoms upon ingestion of the source material and is suggested to reflect a new type of allergy, termed Class 3 food allergy (37).

Over 400 occupational allergenic sources are identified and documented as sensitizers, derived from plants, animals and microbes. However, few have been characterized on the molecular level and only a limited number of purified native or recombinant allergens are available for testing (37). The current WHO/IUIS database categorizes allergens by typical route of exposure. However, in the case of occupational allergens the specific route is often not clear, and the route of sensitization may differ from the common route of elicitation and may be listed as "unknown."

A good example of the complex nature of identifying "occupational allergens" among plants is wheat (*Triticum aestivum*), where 28 allergens are registered with the WHO/IUIS, however only 13 are confirmed food allergens. The route of sensitization is categorized as "unknown" for 15, with 14 being relevant in baker's asthma and several allergens are relevant for grass-pollen cross-reactivity, including profilin (Tri a 12) and a subunit of the tetrameric heterologous a-amylase inhibitor chloroform/methanol-soluble CM17 protein (Tri a 40).

Among animal derived allergens, tropomyosin is an abundant and clear food allergen from crustacean shellfish and molluscs that is also implicated in IgE binding and possibly in triggering airway allergy from mites and some insects. *In vitro* IgE binding studies and skin tests without clear clinical histories of the patient can lead to uncertain diagnosis due to high amino acid homology across diverse taxonomic groups. Clear differentiation is often only possible if IgE inhibition studies are performed and verification of biological activity. Dose and exposure are important. The relevance of IgE binding to environmental sources (house dust mite, cockroach and other organisms) often leads to speculation about route of exposure in sensitization. Of the 35 registered tropomyosins, 22 are known food allergens, while 13 are not known to cause food allergies. Of these 10 are derived from insects or mites and the route of sensitization is most likely via inhalation. Among the food allergens 19 tropomyosin are derived from different seafood species and are relevant occupational allergens, however not identified as such (38). Some animal allergens are not commonly recognized by patients or clinicians due to limited exposure for most people to the source. Hemoglobins from the insect "harlequin fly larvae," *Chironomus thummii* bind IgE from sera of those who raise the insects for fish food (39). A few recent cases of occupational allergy have been noted for workers who raise mealworms (*Tenebrio molitor*) as a source of protein for processed human food (9, 40). Some individuals have been newly sensitized at work,

while for others it appears, they were sensitized by consumption of shrimp or other crustaceans. Highly conserved tropomyosin and a few enzymes such as arginine kinase may be responsible for cross reactivity.

Many other foods have been reported in single case reports as the cause of occupational allergies and asthma, and sensitization by inhaled allergen exposure is therefore quite likely (41). Importantly, though these proteins may be tolerated during exposure by one route or by most people, they may not be tolerated if exposure occurs differently and as such, the WHO/IUIS nomenclature committee has not asked for *in vivo* proof that new allergens illicit allergic responses to date, only *in vitro* evidence and a good clinical history of symptoms.

SUMMARY AND CONCLUSIONS

Though molecular and scientific advances have led to improved characterization of allergens and with-it increased clarity in naming and categorizing them, the heterogeneous nature of human data remains a challenge. This Sub-Committee has in the past asked for minimal data characterizing protein sequences, clinical symptoms of serum donors or test subjects, IgE binding and allergic functionality. We see great diversity in data submitted with candidate proteins. Our recent publications are intended to help researchers focus on relevant questions to improve the evidence that individual proteins are allergens, and not simply cross-reactive in IgE binding with only mild clinical consequences. Sequences of the allergens and demonstration of IgE binding from relevant symptomatic individuals are essential for defining allergens. Some researchers would like to simply submit to the Sub-Committee information about recombinant proteins based only on genomic sequences of potential allergen sources. While that activity may help predict possible allergens, our goal is to call for direct proof of allergenicity using subjects with clear clinical reactivity to source materials, and with clear IgE binding to proteins presented at representative concentrations.

The purpose of the allergen nomenclature system is to help researchers, clinicians, pharmaceutical companies, regulators, and the public clearly understand the identity of clinically important allergens for diagnosis and to help ensure compliance for improved safety.

However, our system is not perfect. Recent research shows that some specific glycans are important allergens including α-gal that is bound to tick and mammalian proteins and to glycol-lipids. We do not define these structures as allergens but hope to help publicize risks and educate consumers and allergists about risks.

AUTHOR CONTRIBUTIONS

Each author contributed a significant portion of writing for the manuscript and reviewed the final document. This collection of authors is a representative segment of the WHO/IUIS Allergen Nomenclature Sub-Committee. AK is not a Sub-Committee member, but has contributed to identification and characterization of allergens in this manuscript.

FUNDING

This work was supported by American Academy of Allergy Asthma and Immunology (AAAAI), European Academy of Allergy and Clinical Immunology (EAACI), and International Union of Immunological Societies (IUIS).

ACKNOWLEDGMENTS

This article was prepared on behalf of the WHO/IUIS Allergen Nomenclature Sub-Committee shown at www.allergen.org.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.02600/full#supplementary-material

REFERENCES

1. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TA. Allergen nomenclature. *Bull World Health Organ.* (1986) 64:767–74.
2. Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS allergen nomenclature: providing a common language. *Mol Immunol.* (2018) 100:3–13. doi: 10.1016/j.molimm.2018.03.003
3. Pomés A, Melén E, Vailes LD, Retief JD, Arruda LK, Chapman MD. Novel allergen structures with tandem amino acid repeats derived from German and American cockroach. *J Biol Chem.* (1998) 273:30801–7. doi: 10.1074/jbc.273.46.30801
4. Mueller GA, Pedersen LC, Lih FB, Glesner J, Moon AF, Chapman MD, et al. The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. *J Allergy Clin Immunol.* (2013) 132:1420–6. doi: 10.1016/j.jaci.2013.06.014
5. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol.* (2018) 2:557–66. doi: 10.1038/s41559-017-0459-1
6. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol.* (2008) 121:847–52. doi: 10.1016/j.jaci.2008.01.025
7. Radauer C, Breiteneder H. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J Allergy Clin Immunol.* (2006) 117:141–7. doi: 10.1016/j.jaci.2005.09.010
8. Spit J, Badisco L, Vergauwen L, Knapen D, Vanden Broeck J. Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. *Insect Mol Biol.* (2016) 25:745–56. doi: 10.1111/imb.12258
9. Broekman HCHP, Knulst AC, den Hartog Jager CF, van Bilsen JHM, Raymakers FML, Kruizinga AG, et al. Primary respiratory and food allergy to mealworm. *J Allergy Clin Immunol.* (2017) 140:600–3.e7. doi: 10.1016/j.jaci.2017.01.035
10. Randall TA, Perera L, London RE, Mueller GA. Genomic, RNAseq, and molecular modeling evidence suggests that the major allergen domain in insects evolved from a homodimeric origin. *Genome Biol Evol.* (2013) 5:2344–58. doi: 10.1093/gbe/evt182

11. Arruda LK, Vailes LD, Platts-Mills TA, Hayden ML, Chapman MD. Induction of IgE antibody responses by glutathione S-transferase from the German cockroach (*Blattella germanica*). *J Biol Chem.* (1997) 272:20907–12. doi: 10.1074/jbc.272.33.20907
12. Mueller GA, Pedersen LC, Glesner J, Edwards LL, Zakzuk J, London RE, et al. Analysis of glutathione S-transferase allergen cross-reactivity in a North American population: relevance for molecular diagnosis. *J Allergy Clin Immunol.* (2015) 136:1369–77. doi: 10.1016/j.jaci.2015.03.015
13. Sookrung N, Reamtong O, Poolphol R, Indrawattana N, Seesuay W, Saelim N, et al. Glutathione S-transferase (GST) of American cockroach, *Periplaneta americana*: classes, isoforms, and allergenicity. *Sci Rep.* (2018) 8:484. doi: 10.1038/s41598-017-18759-z
14. Klug J, Beier HM, Bernard A, Chilton BS, Fleming TP, Lehrer RI, et al. Uteroglobin/Clara cell 10-kDa family of proteins: nomenclature committee report. *Ann N Y Acad Sci.* (2000) 923:348–54. doi: 10.1111/j.1749-6632.2000.tb05549.x
15. Jackson BC, Thompson DC, Wright MW, McAndrews M, Bernard A, Nebert DW, et al. Update of the human secretoglobin (SCGB) gene superfamily and an example of "evolutionary bloom" of androgen-binding protein genes within the mouse Scgb gene superfamily. *Hum Genomics.* (2011) 5:691–702. doi: 10.1186/1479-7364-5-6-691
16. Kaiser L, Velickovic TC, Badia-Martinez D, Adedoyin J, Thunberg S, Hallén D, et al. Structural characterization of the tetrameric form of the major cat allergen Fel d 1. *J Mol Biol.* (2007) 370:714–27. doi: 10.1016/j.jmb.2007.04.074
17. Hilger C, Kler S, Arumugam K, Revets D, Muller CP, Charpentier C, et al. Identification and isolation of a Fel d 1-like molecule as a major rabbit allergen. *J Allergy Clin Immunol.* (2014) 133:759–66. doi: 10.1016/j.jaci.2013.04.034
18. Altmann F. Coping with cross-reactive carbohydrate determinants in allergy diagnosis. *Allergo J Int.* (2016) 25:98–105. doi: 10.1007/s40629-016-0115-3
19. Soh JY, Huang CH, Lee BW. Carbohydrates as food allergens. *Asia Pac Allergy.* (2015) 5:17–24. doi: 10.5415/apallergy.2015.5.1.17
20. Hilger C, Fischer J, Wölbing F, Biedermann T. Role and mechanism of galactose-alpha-1,3-galactose in the elicitation of delayed anaphylactic reactions to red meat. *Curr Allergy Asthma Rep.* (2019) 19:3. doi: 10.1007/s11882-019-0835-9
21. Levin M, Apostolovic D, Biedermann T, Commins SP, Iweala OI, Platts-Mills TAE, et al. Galactose α-1,3-galactose phenotypes. *Ann Allergy Asthma Immunol.* (2019) 122:598–602. doi: 10.1016/j.anai.2019.03.021
22. Elferink H, Rossing E, Huang CH, Lee BW, Cao L, Delsing DJ, et al. Modular synthesis and immunological evaluation of suspected allergenic galactooligosaccharides. *Org Biomol Chem.* (2019) 17:3108–12. doi: 10.1039/C9OB00108E
23. Kuehn A, Hilger C, Lehners-Weber C, Codreanu-Morel F, Morisset M, Metz-Favre C, et al. Identification of enolases and aldolases as important fish allergens in cod, salmon and tuna: component resolved diagnosis using parvalbumin and the new allergens. *Clin Exp Allergy.* (2013) 43:811–22. doi: 10.1111/cea.12117
24. Ferreira F, Hirtenlehner K, Jilek A, Godnik-Cvar J, Breiteneder H, Grimm R, et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med.* (1996) 183:599–609. doi: 10.1084/jem.183.2.599
25. Hilger C, van Hage M, Kuehn A. Diagnosis of allergy to mammals and fish: cross-reactive vs. specific markers. *Curr Allergy Asthma Rep.* (2017) 17:64. doi: 10.1007/s11882-017-0732-z
26. Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy.* (2014) 69:413–9. doi: 10.1111/all.12348
27. Waite KJ. Blackley and the development of hay fever as a disease of civilization in the nineteenth century. *Med Hist.* (1995) 39:186–96. doi: 10.1017/S0025727300059834
28. Johnson P, Marsh DG. Allergens from common rye grass pollen (*Lolium perenne*). I. Chemical composition and structure. *Immunochemistry.* (1966) 3:91–100. doi: 10.1016/0019-2791(66)90290-4
29. Waldrop SL, Pitcher CJ, Peterson DM, Maino VC, Picker LJ. Determination of antigen-specific memory/effector CD4+ T cell frequencies by flow cytometry: evidence for a novel, antigen-specific homeostatic mechanism in HIV-associated immunodeficiency. *J Clin Invest.* (1997) 99:1739–50. doi: 10.1172/JCI119338
30. Ball T, Fuchs T, Sperr WR, Valent P, Vangelista L, Kraft D, et al. B cell epitopes of the major timothy grass pollen allergen, phl p 1, revealed by gene fragmentation as candidates for immunotherapy. *FASEB J.* (1999) 13:1277–90. doi: 10.1096/fasebj.13.11.1277
31. Etto T, de Boer C, Prickett S, Gardner LM, Voskamp A, Davies JM, et al. Unique and cross-reactive T cell epitope peptides of the major Bahia grass pollen allergen, Pas n 1. *Int Arch Allergy Immunol.* (2012) 159:355–66. doi: 10.1159/000338290
32. Schenk S, Breiteneder H, Susani M, Najafian N, Laffer S, Duchêne M, et al. T-cell epitopes of Phl p 1, major pollen allergen of timothy grass (*Phleum pratense*): evidence for crossreacting and non-crossreacting T-cell epitopes within grass group I allergens. *J Allergy Clin Immunol.* (1995) 96:986–96. doi: 10.1016/S0091-6749(95)70237-7
33. Chabre H, Gouyon B, Huet A, Baron-Bodo V, Boran-Bodo V, Nony E, et al. Molecular variability of group 1 and 5 grass pollen allergens between Pooideae species: implications for immunotherapy. *Clin Exp Allergy.* (2010) 40:505–19. doi: 10.1111/j.1365-2222.2009.03380.x
34. Nony E, Timbrell V, Hrabina M, Boutron M, Solley G, Moingeon P, et al. Specific IgE recognition of pollen allergens from subtropic grasses in patients from the subtropics. *Ann Allergy Asthma Immunol.* (2015) 114:214–20.e2. doi: 10.1016/j.anai.2014.12.005
35. Besche A. On Asthma bronchiale in Man provoked by cat, dog and different other animals. *Acta Med Scand.* (2009) 92:237–55. doi: 10.1111/j.0954-6820.1937.tb13652.x
36. Kogevinas M, Zock JP, Jarvis D, Kromhout H, Lillienberg L, Plana E, et al. Exposure to substances in the workplace and new-onset asthma: an international prospective population-based study (ECRHS-II). *Lancet.* (2007) 370:336–41. doi: 10.1016/S0140-6736(07)61164-7
37. Raulf M. Allergen component analysis as a tool in the diagnosis and management of occupational allergy. *Mol Immunol.* (2018) 100:21–7. doi: 10.1016/j.molimm.2018.03.013
38. Lopata AL, Jeebhay MF. Airborne seafood allergens as a cause of occupational allergy and asthma. *Curr Allergy Asthma Rep.* (2013) 13:288–97. doi: 10.1007/s11882-013-0347-y
39. Meseguer Arce J, Villajos IM, Iraola V, Carnés J, Fernández Caldas E. Occupational allergy to aquarium fish food: red midge larva, freshwater shrimp, and earthworm. A clinical and immunological study. *J Investig Allergol Clin Immunol.* (2013) 23:462–70.
40. Lopata AL, Fenemore B, Jeebhay MF, Gäde G, Potter PC. Occupational allergy in laboratory workers caused by the African migratory grasshopper *Locusta migratoria*. *Allergy.* (2005) 60:200–5. doi: 10.1111/j.1398-9995.2005.00661.x
41. Jeebhay MF, Moscato G, Bang BE, Folletti I, Lipinska-Ojrzanowska A, Lopata AL, et al. Food processing and occupational respiratory allergy-A EAACI position paper. *Allergy.* (2019). doi: 10.1111/all.13807. [Epub ahead of print].

Conflict of Interest: AP is employed by the company Indoor Biotechnologies, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

PERSPECTIVE
published: 18 December 2019
doi: 10.3389/fimmu.2019.02961

A Proposed New Nomenclature for the Immunoglobulin Genes of *Mus musculus*

Christian E. Busse[1]*, *Katherine J. L. Jackson*[2], *Corey T. Watson*[3] and *Andrew M. Collins*[4]*

[1] Division of B Cell Immunology, German Cancer Research Center, Heidelberg, Germany, [2] Immunology Division, The Garvan Institute of Medical Research, Darlinghurst, NSW, Australia, [3] Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville, KY, United States, [4] School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

Mammalian immunoglobulin (IG) genes are found in complex loci that contain hundreds of highly similar pseudogenes, functional genes and repetitive elements, which has made their investigation particularly challenging. High-throughput sequencing has provided new avenues for the investigation of these loci, and has recently been applied to study the IG genes of important inbred mouse strains, revealing unexpected differences between their IG loci. This demonstrated that the structural differences are of such magnitude that they call into question the merits of the current mouse IG gene nomenclatures. Three nomenclatures for the mouse IG heavy chain locus (*Igh*) are presently in use, and they are all positional nomenclatures using the C57BL/6 genome reference sequence as their template. The continued use of these nomenclatures requires that genes of other inbred strains be confidently identified as allelic variants of C57BL/6 genes, but this is clearly impossible. The unusual breeding histories of inbred mouse strains mean that, regardless of the genetics of wild mice, no single ancestral origin for the IG loci exists for laboratory mice. Here we present a general discussion of the challenges this presents for any IG nomenclature. Furthermore, we describe principles that could be followed in the formulation of a solution to these challenges. Finally, we propose a non-positional nomenclature that accords with the guidelines of the International Mouse Nomenclature Committee, and outline strategies that can be adopted to meet the nomenclature challenges if three systems are to give way to a new one.

Keywords: immunoglobulin, nomenclature, V genes, B cell, IGH, IGK, IGL

OPEN ACCESS

Edited by:
Amy L. Kenter, University of Illinois at Chicago, United States

Reviewed by:
Claudia Berek, Charité Medical University of Berlin, Germany
Ann Feeney, The Scripps Research Institute, United States

***Correspondence:**
Christian E. Busse
christian.busse@dkfz-heidelberg.de
Andrew M. Collins
a.collins@unsw.edu.au

Specialty section:
This article was submitted to B Cell Biology, a section of the journal Frontiers in Immunology

Received: *11 September 2019*
Accepted: *03 December 2019*
Published: *18 December 2019*

Citation:
Busse CE, Jackson KJL, Watson CT and Collins AM (2019) A Proposed New Nomenclature for the Immunoglobulin Genes of Mus musculus. Front. Immunol. 10:2961. doi: 10.3389/fimmu.2019.02961

INTRODUCTION

The generation of antibody diversity relies in part on the use of genes from extensive gene families residing in the immunoglobulin (IG) loci of the mammalian genome. Remarkably, a comprehensive understanding of the organization of these gene sets emerged long before a detailed knowledge of antibody gene sequences was available (1, 2), and much of the research that led to this understanding was performed in mice. The IG gene loci – being polymorphic and polygenic—are especially complex, which has created challenges for the development of a gene nomenclature that is both logical and sustainable. This manuscript presents a new proposal to meet this challenge.

Mouse antibody polypeptide chains, and the genes that encode them, were first identified using mineral-oil induced plasmacytoma cell lines derived from BALB/c mice (3). It was soon realized that functional mouse IG heavy and kappa light chain variable (*Ighv*, *Igkv*) genes exist as multigene families (4, 5). *Ighv* genes, e.g., belong to 15 gene families (5, 6), and the first names given to the *Igkv* and *Ighv* gene families came from the names of the cell lines that were used in their identification. For example, the anti-dextran antibody-producing cell line J558 was used to generate a DNA probe by which a family of *Ighv* genes was identified by Southern blot analysis (7). This family was subsequently termed "J558," and today is generally known as the *Ighv1* family.

Ighv gene probes were also used in Southern blot analysis to explore the loci of different inbred mouse strains, with eight different haplotypes being identified (7). This expanded a system of classification that began with serologically-defined allotypic variation in the immunoglobulin constant regions (8). In this system, the BALB/c and C57BL/6 *Igh* haplotypes were designated Igh^a and Igh^b, respectively.

Although a later investigation of 72 inbred mouse strains found that 13 strains carried variant haplotypes involving substantial structural variation (9), broad similarities were believed to exist between the *Igh* loci of strains carrying shared haplotypes (6, 10). After the sequencing and annotation of the *Igk* locus (11, 12) and the *Igh* locus (13, 14) of the C57BL/6 strain, the earlier Southern blot studies provided justification for comparisons of sequences from other strains with those of the C57BL/6 reference genome, and for the identification of sequences as allelic variants of their most similar sequences in the reference genome. The sequencing of the loci also led to the development of new nomenclatures for both the heavy (14–16) and the light chain (17). Discussion here will focus on the nomenclature of the variable genes of the heavy chain, *Ighv*.

A positional nomenclature was developed by the International ImMunoGeneTics Information System (IMGT) group (18), based upon the mouse genome reference sequence (C57BL/6), in which each gene was assigned a name reflecting its gene family, and the sequential position of the gene within the locus from proximal to distal positions, relative to other genes of that family. The most proximal gene of a family was given the position number "1," counting up to the most distally located gene in that family, which was given the number equalling the total number of genes in that family (a scheme referred to as *family-centric* below, also see **Figure 1**). This is different to the IMGT nomenclature for human *IGHV* genes, in which the position number refers to the position of the gene within the entire set of *IGHV* genes, with the most proximal gene being numbered "1," and the most distal gene being numbered "81" (referred to as *locus-centric* scheme below). In the IMGT nomenclature, the locus name is included in the gene name (e.g., IGHV1-18), and the old *Ighv* family names are replaced with a numbering system proposed by Honjo and Matsuda (19).

A positional nomenclature was also developed by Johnston and colleagues, based upon their alternative genome assembly of the C57BL/6 *Igh* locus (14). The Johnston nomenclature utilizes the earlier gene family names (7183, J558, 36-60, etc.), a number representing the position of the gene within the gene family, and a second number representing the position of the gene amongst all genes of the locus (e.g., J558.31.121, 7183.7.10). In this nomenclature, pseudogenes are indicated by an additional "pg" tag (e.g., 36-60.7pg.72). A study of the *Ighv* locus of the 129S1 strain led to the development of a variant of the Johnston nomenclature by Retter et al. (16). While still following the basic rules set by Johnston et al., Retter et al. constructed the names using a locus descriptor ("VH"), the earlier *Ighv* gene family name, a letter referring to the *Igh* haplotype of the inbred strain, a number representing the position of the gene within the gene family, followed by the "psi" tag for pseudogenes, and a second

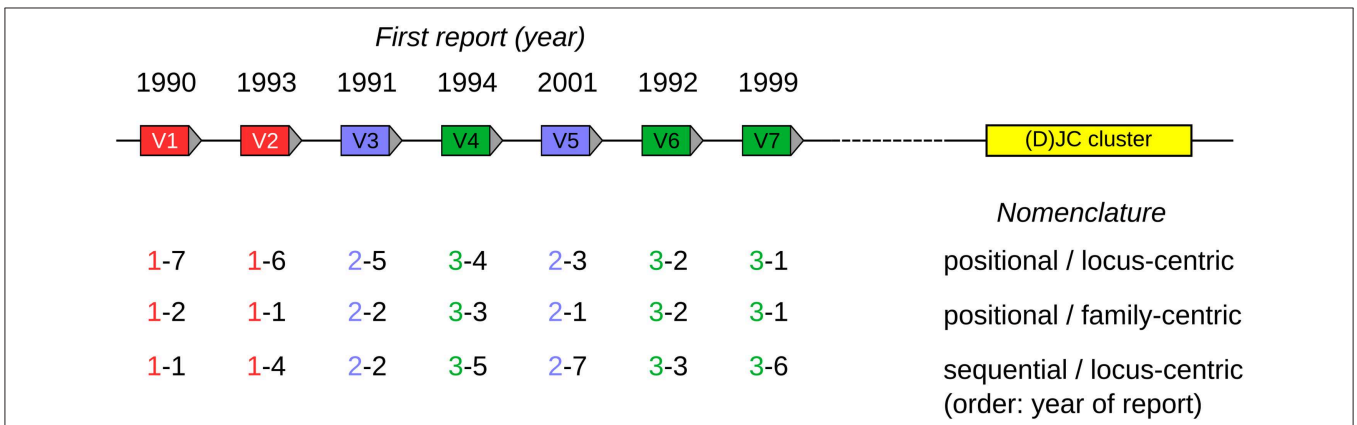

FIGURE 1 | Visualized scheme of three nomenclature strategies, using a hypothetical locus encompassing seven V genes (labeled V1–V7) belonging to three V gene families (indicated as red, blue, green). The year of the first report is indicated above the genes. The (D)JC region is shown as a yellow box and provides orientation for the positional strategies. The designations beneath the individual V genes follow the `<family>-<member>` format discussed in the text. To increase the readability, the `<stem>` component has been omitted from the designations, as it would be identical for all designations, since only a single locus is shown here. For better clarity, gene family designations are also indicated by text color.

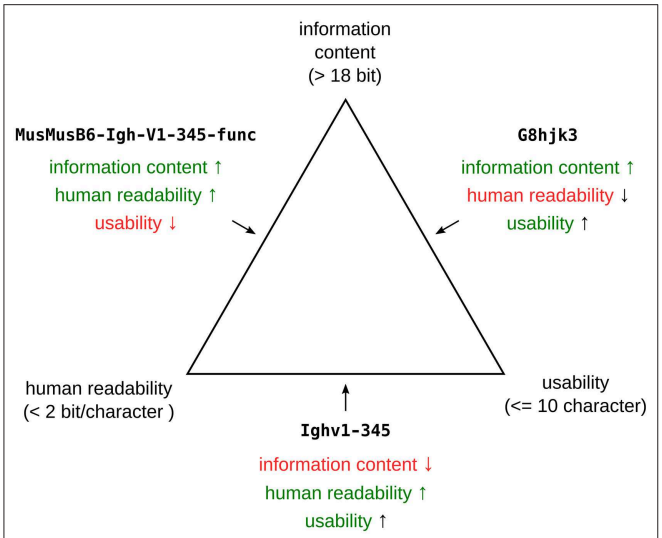

FIGURE 2 | Ternary plot depicting the constraints for gene symbols. The individual properties and their boundaries are located on the three corners. The minimal information content is based on the requirement to be able to encode at least 4 loci, 3 different types of gene segments (V, D, J), 32 gene families and 1024 members. The 10 character limit (usability) is based on current IMNC guidelines. The limits for human readability are a compromise between standard English language entropy (≈ 1 bit per character) and pure numerical representation (3.3 bit per character). Examples that optimize two properties (shown in green) at the expense of the third one (red) are shown on the median of the respective edges.

number denoting the position of the gene within the locus (e.g., VH7183.a3psi.5). Both the Johnston and the Retter reference data sets can be readily accessed for analysis e.g., via IgBLAST (20). Finally, while Retter and colleagues also developed a further designation system for their VBASE2 sequence repository (21), it should be noted that we consider these to be primarily sequence identifiers, rather than a genetic nomenclature in the strict sense.

All three mouse nomenclatures are currently in use, and all are challenged by recent findings that show that there are substantial differences, including structural differences, between the *Igh* loci of different classical inbred mouse strains (22). It has been proposed that the differences between the genes of the BALB/c and C57BL/6 strains could have resulted from their loci having originated in different subspecies of the house mouse (22). However, investigations of *Ighv* genes in wild-derived strains now suggest this is not the case (23). Instead, the *Ighv* loci of classical inbred strains appear to be mosaics, made up of many relatively short haplotype blocks that may have their origins in disparate subspecies of the house mouse, and even in other *Mus* species.

The discovery of variation in BALB/c and wild-derived strains was made using inferential techniques. These techniques are widely used in human and other species to identify allelic variants (24–26), but it is already clear that it will not be possible to unequivocally associate most inferred variants in the mouse to any particular gene in the C57BL/6 genome reference sequence (22, 23). For example, amongst the set of inferred BALB/c *Ighv* sequences, there are instances in which three or more sequences are most closely aligned with a single C57BL/6 *Ighv* gene. In such circumstances, it is impossible to discern whether these BALB/c sequences represent allelic variants of the C57BL/6 genes or distinct gene loci.

It therefore has become increasingly clear that, given the extent of IG diversity likely to be encountered among commonly used inbred mouse strains, our ability to effectively characterize and catalog mouse IG genes and alleles will be constrained by the current nomenclature systems. There is therefore a clear requirement for a new, more flexible nomenclature that will better meet the needs of the community. Here, in light of the challenges we face with the curation of mouse IG sequencing data, we discuss the key aspects that should be considered in the establishment of any nomenclature system. We use this discussion to motivate the proposal of a new non-positional mouse IG nomenclature.

CONSIDERATION FOR A NEW MOUSE IG NOMENCLATURE

Challenges and Constraints

Before considering the specifics of *Igh* nomenclature, it is worthwhile to consider what gene symbols (27) and their use must, should and should not try to achieve. Note that gene names (27) (e.g., Immunoglobulin Heavy Variable 1–2) will not be discussed separately here, as they are rarely used in scientific communication, and gene symbols (e.g., IGHV1-2) are synonymous with them. Gene symbols in general aim to provide designations to hereditary units, which in virtually all cases refer to specific physical regions in the genome. This nowadays often translates into linear base-pair sequences. Gene symbols serve as handles for this information, and have to balance three interdependent properties (**Figure 2**): a symbol should be unique (i.e., refer to a single specific gene), human-decodable and short enough for everyday use. The criterion of human-decodability requires a formalized system, e.g., that all IG symbols start with "Ig." This creates redundancy and thus reduces the potential information content. In combination with the limited length for a symbol—the International Committee on Standardized Genetic Nomenclature for Mice (IMNC) [https://perma.cc/6F9S-6H4U] recommends a maximum of ten characters (27, 28)—this means that the overall information content of a symbol is limited. From this it follows that a gene symbol should encode only the minimal information required for the unambiguous identification of each particular gene. We will refer to this conclusion as the *lean designator principle* below.

Based on these theoretical limitations, we now need to consider what information a gene symbol should not attempt to encode. Firstly, a gene symbol is not required to be a synonym for a specific physical location on the genome (e.g., "Chromosome 12; BPs 114,048,536-114,048,547"). Indeed, before the current era, in which the complete sequencing of genomes is now commonplace, having a fine-grained physical mapping for a gene in an organism was the exception rather than the rule. Secondly, a gene symbol is not expected to be used without some biological context. This implies, on the one hand,

that it should not encode information that can be stored and accessed elsewhere using the gene symbol as identifier (e.g., from a reference database). On the other hand, it means that a gene symbol is not a globally unique identifier, but only a unique identifier within a single organism, as it is safe to assume that the biological context provides knowledge of the species. Thirdly, harmonizing gene symbols between organisms has been a long-standing but ultimately futile endeavor. It is critical to recognize that any attempt at harmonization is at variance with the *lean designator principle*, as it tries to encode non-essential information about communality within a gene symbol.

With this basic theoretical understanding regarding the general design of gene symbols, we now need to understand how these symbols are best assigned to real-life data. While the following considerations can—in theory—be applied to all genetic loci, they are clearly most relevant to loci harboring large ensembles of genes from one or multiple related gene families (e.g., immunoglobulins). We will refer hereafter to these loci as "polymorphic, polygenic and repetitive loci" (PPRL). As discussed in the Introduction, various individuals of a species can exhibit substantial diversity in the form of single nucleotide polymorphisms (SNPs) as well as larger structural variants (SVs; e.g., deletions and duplications). Therefore, the sequencing of a PPRL in a new individual will often result in the observation of novel sequence variants for a number of genes. The frequently employed naming paradigm for such new genes is to *lump*, i.e., to assign the gene symbol of the closest related known gene to the novel sequence. However, this operation implies that it is most likely that the two sequences are allelic variants of the same gene, which is a claim that should require evidence in its support or otherwise be rejected. Importantly, with the increased structural variation observed in PPRL, the alternative scenario of a paralogous relationship existing between two observed sequences not only becomes more likely, it also becomes more challenging to detect due to the high redundancy of the surrounding sequences. In these cases, the opposite naming paradigm of *splitting*, i.e., assigning new gene symbols to a novel sequence, unless the allelic relationship of two sequences is proven beyond doubt, should be the preferred mode of action. While both naming paradigms can exhibit false-positive and false-negative errors, the key consideration should always be to protect the integrity of the scientific record in the most reliable manner. The consequence of these different errors is illustrated by the following complementary scenarios:

- False-negative gene assignment: a novel sequence is observed, which has a high but not perfect homology to a known gene. Under the *lump* paradigm, the sequence is assigned the gene symbol of the closest related known gene and considered to be an allelic variant of it. Five years later, it is recognized that the two "alleles" instead represent distinct genes. Therefore the more recently characterized sequence is given a new gene designation and the initially assigned allele symbol is retired. All past scholarly communication that does not clearly provide an allele designation will require reconsideration as the report could be referring to either gene. Continued use of the initial faulty nomenclature will continue to create confusion.
- False-positive gene assignment: As above, a novel sequence is observed, which has a high but not perfect homology to a known gene. Under the *split* paradigm, the sequence is assigned a new gene name. Five years later, it is recognized that the two "genes" actually represent alleles of the same gene. Therefore the initial designation is now retired and the newly assigned allele symbol is linked to the older designation. All scholarly communication using the erroneously assigned gene symbol can be easily understood. Continued use of the erroneously assigned designation is bad practice but not harmful.

These examples show that the potential loss of information regarding "inheritance by descent" in the *split* paradigm can be dealt with more easily than the lack of accuracy imposed by the *lump* paradigm.

The complexity of PPRLs might also require a revision of the current allele designation strategy: The IMNC currently assigns lower-case letters to mouse alleles of all loci, which are based on the reference strain in which a given sequence is observed. Importantly, this system mixes allele information with haplotype information. While this might be appropriate for stable parts of the genome, recent studies (23) suggest that novel haplotypes of the *Igh* locus will likely be identified with the analysis of each new inbred strain. This creates a situation in which multiple distinct haplotypes can share the same sequence. This is an example of the inappropriate use of gene symbols whereby too much information is being encoded in too little space, based on the assumption of relatively high stability and homogeneity between strains. Haplotypes are better stored in reference databases and/or the metadata for an allele, rather than in the allele symbol. We believe that this should be implemented in any new IG allele nomenclature.

Finally, it should be noted that gene symbols should follow the general nomenclature guidelines for a given species. This not only reduces potential ambiguity in scholarly communication and facilitates simplified distinctions between species (e.g., human and mouse), but it also allows for automated formatting. Of note, the nomenclatures of Johnston et al., Retter et al. and IMGT all fail to comply with IMNC guidelines. The nomenclatures of Johnston et al. and Retter et al. use punctuation, while the IMGT nomenclature uses gene symbols in all-caps with a numeric representation of alleles. All these features are at variance with IMNC guidelines.

In summary, we hope that we have established five central aspects for the curation of sequences in PPRL. Firstly, gene symbols need to be human-decodable, hence overall information needs to be minimized. Secondly, the best way for gene symbols to be human-decodable is to consider them to be designations for sequences, and nothing else. Thirdly, gene assignments should in general follow the *split* paradigm, as it is more robust to changes over time. Fourthly, the IMNC-recommended mouse allele nomenclature needs revision. Finally, gene symbols need to follow the established nomenclature rules for a given species.

Potential Numbering Strategies

The IMNC-recommended way to assign symbols to individual genes of a gene family follows a <stem><family><member> format (27), in which the <stem> field denotes the super-group of genes, the <family> field indicates the gene family and <member> the individual gene. The <stem> gene super-group usually closely follows the common concept of a "gene locus," e.g., *Ighv*, the special case of off-loci genes will be discussed separately below. However, there is no common standard stating whether <family> and <member> should be represented by letters or Arabic numerals: Protocadherins (*Pcdh*) use a "Letter Number" format, olfactory receptors (*Olfr*) use "Number Letter Number" in humans, but only numbers in mice. The Human Genome Nomenclature Committee (HGNC)-approved IMGT naming scheme for human IG and TR uses a "Number – Number" format, in which the fields are separated by a hyphen (the usage of which is explicitly allowed by HGNC for these loci). As there is no general problem with this format, we believe this component of the nomenclature should be retained.

The assignment of individual genes to families is usually based on arbitrary thresholds of sequence homology. However, it should be noted that this assignment procedure constitutes a non-trivial partitioning problem, especially when facing an increasing number of elements. As a detailed discussion of this problem is beyond the scope of this manuscript, we will assume that these assignments can be performed in a deterministic and stable fashion. An assignment is considered deterministic if each element will always be assigned to the same family, and stable if the addition of an $n + 1^{th}$ element does not alter the assignments of any of the other n elements.

There are multiple ways in which the number in the <member> field can be assigned and most of the existing strategies reflect differing responses to two questions:

1. Should the <member> field indicate the position of the gene in its locus, based on a reference genome assembly ("positional"), or just be assigned in an incremental way ("sequential"), e.g., according to the order of discovery?
2. Does the <member> field partition the namespace of all genes of the locus (*locus-centric*) or just the namespace of the members in a given family (*family-centric*)?

In the past—as discussed in the introduction—a multitude of approaches have been used for PPRL in general and for the IG loci in particular. These are depicted in a schematic way in **Figure 1**.

It is critical to recognize that positional schemes are problematic in general and especially when used for nomenclature of PPRL. Firstly, they violate the *lean designator principle* as by definition they encode positional information that is not strictly necessary for the gene symbol. Secondly, as they struggle to deal with duplications and other additive SVs, they are not well suited for application of the *split* paradigm, which we have established is appropriate for any PPRL nomenclature. Thirdly, not only do they encourage the use of the *lump* paradigm, but they also fail catastrophically once a downstream *split* is required. This is due to the fact that a *split* within an—initially—positional scheme requires an extension of the format to perform the required subpartitioning, which then by itself can lead the positional numbering *ad absurdum* (e.g., if *Ighv1-23a* and *Ighv1-23b* are not located next to each other). Finally, positional schemes usually assume that all genes are located in a single continuous locus, which—as discussed below—might not be the case. Because of these problems, we believe that only a sequential scheme can provide the flexibility required by PPRLs.

In regard to the question of which space the <member> field should actually partition, we favor a *locus-centric* scheme as this means that the <member> field becomes a unique identifier for a given gene, independent of the gene family assignment. This not only provides for more error-tolerant designations, as no two families share a gene with the same <member> field, but it also allows for more flexibility should a reassignment of families become necessary.

Having decided on a sequential and *locus-centric* scheme, we must now consider the actual assignment procedure that would be performed once a novel sequence is observed. As we have argued before, we should assume incomplete knowledge of the locus structure and therefore in general follow a *split* rather than a *lump* paradigm. Assuming that a novel sequence can always be grouped into a family, a new member number should therefore be assigned by default. However, this does not mean that all pre-existing information needs to be rejected. Taken to the extreme, such a rejection would mean that sequences from a well-known line of an inbred mouse strain that had been kept for numerous generations at a particular facility, would all need to be assigned new gene designations, as there may have been genetic divergence of the colony since its founding. We therefore consider it prudent to introduce a principle of parsimony, which implies that above a certain threshold, "Identity by descent" of two sequences will be considered to be likely. To maintain the stringency of this approach, we propose setting the threshold at 100% identity of the coding sequence. This allows on the one hand to collapse the majority of sequences observed when re-sequencing lines of existing strains but on the other hand follows the *split* paradigm as closely as possible. While thresholds slightly below 100% might seem attractive, as they could accommodate potential sequencing errors, we reject such thresholds as being arbitrary and situations are known to exist where two genes reside at distinct genomic locations but differ by just a single nucleotide. Furthermore, we consider both sequencing and inference technologies to be advanced enough by now that appropriate error correction should be in place.

In summary, we here propose a sequential and *locus-centric* nomenclature based on a parsimonious *split* paradigm.

Handling of Existing Designations

The introduction of a revised nomenclature naturally raises the question of how to handle legacy designations. In general, as the current IMGT designations of C57BL/6 genes do not contain any obvious errors, these names should remain in place. However, all other alleles that are not present in GRCm38 should be subject to renaming, based on the scheme described here. The strict use of IMNC formatting will avoid potential confusion by clearly distinguishing legacy IMGT names from revised names. In addition, use of an initial value for the <member> field of

200 or above would avoid collisions in cases where the IMNC formatting is not used.

The proposed consecutive assignment of gene numbers might for some readers be reminiscent of the "S"-nomenclature used by IMGT for many unmapped sequences. This could lead to the perception that the "S"-nomenclature might present a more conservative way of addressing the existing nomenclature problems than the much more drastic revised nomenclature proposed here. However, it should be pointed out that IMGT considers the "S"-nomenclature to be a temporary designation that precedes the assignment of a positional-based gene symbol. In contrast to this, the nomenclature proposed here rejects the presumption that we will be able to map relevant genes with sufficient certainty.

The nomenclature for so-called "orphaned" V genes also needs to be addressed. These are genes residing at substantial distance from the main gene loci, often on other chromosomes. There are two general strategies that might be used to handle these genes, based on the concept of what the <stem> field refers to. On the one hand, the <stem> field could designate a single and continuous physical location in the genome. In this case, e.g., the "Igh" prefix would be considered a shorthand for "Chromosome 12; BPs 113,225,000-116,024,999," and all genes outside of this region would bear another designation. Based on strategies used for other gene families, these genes could for example be prefixed as "Ighvl" ("Immunoglobulin heavy variable-like"). On the other hand, <stem> could be considered a designation that a gene is part of a super-group of gene families (based on homology), with information about the physical location being stored elsewhere. In this case, "orphaned" genes would use the usual <stem> (e.g., "Ighv"), a <family> number based on the general homology thresholds as discussed above and unique <member> number. It should be noted that IMNC defines a "locus" as a mappable "point in the genome" (27) and both concepts of the <stem> field are compatible with this definition. The main differences between the two concepts are based on the interpretation of a locus as a continuous region vs. the grouping of genes based on homology independent of their location in the genome. As the currently available data does not show support for any claimed utilization of "orphaned" genes in V(D)J rearrangements, we think that *lean designator principle* is eclipsed by the aim of a stricter definition of the locus. Therefore we would argue for the reassignment of the off-loci genes as "-likes."

THE WAY FORWARD

The challenges to existing nomenclatures that stimulated this manuscript were studies that identified new mouse IG genes by inference from rearranged V(D)J sequences, rather than new genes that were identified by genomic sequencing. As we expect that the inference process will likely dominate mouse IG gene studies for some time, we believe that the development of a new nomenclature should go hand in hand with the development of a system for the validation of inferences by the research community. Procedures have recently been established for the validation of genes of the human *IGH* locus, through the establishment of the Inferred Allele Review Committee [IARC; (29)]. We would like to propose the creation of a Mouse Immunoglobulin Gene and Allele Review Committee, in cooperation with IMGT, IMNC and the AIRR Community, and under the auspices of the International Union of Immunological Societies (IUIS) Nomenclature Committee.

In conclusion, we hope that this proposal will stimulate discussion among and action by the stakeholders involved in the mouse IG nomenclature, to resolve these critical issues. Long live *Igh*!

AUTHOR CONTRIBUTIONS

CB and AC drafted the manuscript. All authors contributed to the editing of the manuscript and contributed to the development of the procedures described.

ACKNOWLEDGMENTS

The authors would like to thank the members of AIRR Community Germline Database Working Group (GLDB WG) for discussions and feedback.

REFERENCES

1. Adams JM, Kemp DJ, Bernard O, Gough N, Webb E, Tyler B, et al. Organization and expression of murine immunoglobulin genes. *Immunol Rev*. (1981) 59:5–32.
2. Tonegawa S. Somatic generation of antibody diversity. *Nature*. (1983) 302:575–81.
3. Potter M. Antigen-binding myeloma proteins of mice. *Adv Immunol*. (1977) 25:141–211.
4. Joho R, Weissman IL, Early P, Cole J, Hood L. Organization of kappa light chain genes in germ-line and somatic tissue. *Proc Natl Acad Sci USA*. (1980) 77:1106–10.
5. Kabat EA. *Sequences of Proteins of Immunological Interest, 4th ed*. Washington, DC: NIH (1987).
6. Mainville CA, Sheehan KM, Klaman LD, Giorgetti CA, Press JL, Brodeur PH. Deletional mapping of fifteen mouse VH gene families reveals a common organization for three Igh haplotypes. *J Immunol*. (1996) 156:1038–46.
7. Brodeur PH, Riblet R. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur J Immunol*. (1984) 14:922–30.
8. Blomberg B, Geckeler WR, Weigert M. Genetics of the antibody response to dextran in mice. *Science*. (1972) 177:178–80.
9. Tutter A, Riblet R. Duplications and deletions of Vh genes in inbred strains of mice. *Immunogenetics*. (1988) 28:125–35.
10. Brodeur PH, Osman GE, Mackle JJ, Lalor TM. The organization of the mouse Igh-V locus. Dispersion, interspersion, and the evolution of VH gene family clusters. *J Exp Med*. (1988) 168:2261–78.
11. Kirschbaum T, Pourrajabi S, Zocher I, Schwendinger J, Heim V, Röschenthaler F, et al. The 3′ part of the immunoglobulin kappa locus of the mouse. *Eur J Immunol*. (1998) 28:1458–66.

12. Röschenthaler F, Kirschbaum T, Heim V, Kirschbaum V, Schäble KF, Schwendinger J, et al. The 5' part of the mouse immunoglobulin kappa locus. *Eur J Immunol*. (1999) 29:2065–71.
13. Riblet R. Immunoglobulin heavy chain genes of mouse. In: Honjo T, Alt FW, Neuberger MS, editors. *Molecular Biology of B Cells*. Burlington: Elsevier (2004). p. 19–26. Available online at: https://linkinghub.elsevier.com/retrieve/pii/B9780120536412500034
14. Johnston CM, Wood AL, Bolland DJ, Corcoran AE. Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J Immunol*. (2006) 176:4221–34. doi: 10.4049/jimmunol.176.7.4221
15. Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, Chaume D, et al. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res*. (2006) 34:D781–4. doi: 10.1093/nar/gkj088
16. Retter I, Chevillard C, Scharfe M, Conrad A, Hafner M, Im TH, et al. Sequence and characterization of the Ig heavy chain constant and partial variable region of the mouse strain 129S1. *J Immunol*. (2007) 179:2419–27. doi: 10.4049/jimmunol.179.4.2419
17. Martinez-Jean C, Folch G, Lefranc MP. Nomenclature and overview of the mouse (Mus musculus and Mus sp.) immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet*. (2001) 18:255–79. doi: 10.1159/000049204
18. Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, et al. IMGT, the international ImMunoGeneTics information system 25 years on. *Nucleic Acids Res*. (2015) 43:D413–22. doi: 10.1093/nar/gku1056
19. Honjo T, Matsuda F. Immunoglobulin heavy chain loci of mouse and human. In: Honjo T, Alt FW, editors. *Immunoglobulin Genes*. London: Elsevier (1995). p. 145–71. Available online at: https://linkinghub.elsevier.com/retrieve/pii/B9780120536405500100
20. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*. (2013) 41:W34–40. doi: 10.1093/nar/gkt382
21. Retter I, Althaus HH, Münch R, Müller W. VBASE2, an integrative V gene database. *Nucleic Acids Res*. (2005) 33:D671–4. doi: 10.1093/nar/gki088
22. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJL. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos Trans R Soc Lond B Biol Sci*. (2015) 370:20140236. doi: 10.1098/rstb.2014.0236
23. Watson CT, Kos JT, Gibson WS, Busse CE, Newman L, Deikus G, et al. A comparison of immunoglobulin IGHV, IGHD and IGHJ genes in wild-derived and classical inbred mouse strains. *Immunol Cell Biol*. (2019) 97:888–901. doi: 10.1111/imcb.12288
24. Gadala-Maria D, Yaari G, Uduman M, Kleinstein SH. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc Natl Acad Sci USA*. (2015) 112:E862–70. doi: 10.1073/pnas.1417683112
25. Corcoran MM, Phad GE, Bernat NV, Stahl-Hennig C, Sumida N, Persson MAA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nature Commun*. (2016) 7:13642. doi: 10.1038/ncomms13642
26. Ralph DK, Matsen FAt. Per-sample immunoglobulin germline inference from B cell receptor deep sequencing data. *PLoS Comp Biol*. (2019) 15:e1007133. doi: 10.1371/journal.pcbi.1007133
27. International Committee on Standardized Genetic Nomenclature for Mice. *Guidelines for Nomenclature of Genes, Genetic Markers, Alleles, & Mutations in Mouse & Rat*. (2018). Available online at: https://perma.cc/Y7JQ-6TV7
28. Maltais LJ, Blake JA, Eppig JT, Davisson MT. Rules and guidelines for mouse gene nomenclature: a condensed version. International Committee on Standardized Genetic Nomenclature for Mice. *Genomics*. (1997) 45:471–6.
29. Ohlin M, Scheepers C, Corcoran M, Lees WD, Busse CE, Bagnara D, et al. Inferred allelic variants of immunoglobulin receptor genes: a system for their evaluation, documentation, and naming. *Front Immunol*. (2019) 10:435. doi: 10.3389/fimmu.2019.00435

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

OPINION
published: 15 May 2020
doi: 10.3389/fimmu.2020.00908

Perspective: Insights on the Nomenclature of Cytokines and Chemokines

*Albert Zlotnik**

Department of Physiology and Biophysics, University of California, Irvine, Irvine, CA, United States

Keywords: cytokines, chemokines, nomenclature, evolution, function

SUMMARY

The human genome contains some 23,000 genes. Many of these are important in immunology and we have witnessed a very large increase in the characterization of novel genes important in the function of the immune system. Along with these discoveries, issues related to nomenclature have arisen. Often the names proposed for these novel genes and the proteins they encode result in confusion for a new field of research. Here I explain how nomenclature can also help bring important biological insights into the functions of cytokines and chemokines.

INTRODUCTION

Immunology has advanced dramatically in the last 30 years and along with this progress we have witnessed the identification of many novel genes encoding proteins that have important functions in the immune system. Among these are the cytokines, which represent small secreted proteins (10–30 KDa) that are typically produced by cells of the immune system upon activation, and which play pivotal roles in the development and control of immune responses. The history of the cytokines starts in the second half of the 1970's when many groups realized that activated lymphocytes produced secreted proteins that had dramatic effects on other leukocytes. The typical experiment involved the activation of spleen cells with mitogens and the characterization of the biological activities of the supernatants derived thereof. The soluble mediators were given names of the assays that detected their activities like "macrophage activation factor" or "macrophage inhibitory factor." Several teams started to apply biochemical efforts to distinguish or molecularly characterize the mediators of these activities and this led to the realization that two of the earliest cytokines exhibited specific biochemical characteristics. This led to the identification of the first two interleukins, interleukin 1 and interleukin 2, which were named at the Second International Lymphokine Conference (which was held in Interlaken, Switzerland). Doubtless the venue site inspired the participants to come up with the term "interleukin" which suggests interactions between leukocytes. This example highlights that the issues of nomenclature have been relevant in immunology from the very start of the cytokine field.

Another dramatic step forward was the development of molecular biology tools which led to the initial efforts to "clone" the genes encoding important cytokines. One of the first to be cloned was interferon gamma (by Genentech). At that time biotechnology companies became players in the field and companies like DNAX and Immunex joined the efforts to clone genes of new cytokines. The roster of chemokines by the late 1980s had expanded significantly, up to Interleukin 10 (1). The molecular characterization of these cytokines led in turn to the availability of more molecular tools (Recombinant cytokines, monoclonal antibodies against them), that led to milestone discoveries in immunology like the definition of Th1 and Th2 immune responses (2).

OPEN ACCESS

Edited by:
Menno C. van Zelm,
Monash University, Australia

Reviewed by:
Brian F. Volkman,
Medical College of Wisconsin,
United States

***Correspondence:**
Albert Zlotnik
azlotnik@uci.edu

Specialty section:
This article was submitted to Cytokines and Soluble Mediators in Immunity, a section of the journal Frontiers in Immunology

Received: *26 December 2019*
Accepted: *20 April 2020*
Published: *15 May 2020*

Citation:
Zlotnik A (2020) Perspective: Insights on the Nomenclature of Cytokines and Chemokines. Front. Immunol. 11:908. doi: 10.3389/fimmu.2020.00908

A common belief was that cytokine biology held the key to novel therapeutics. This turned out to be correct, but not as originally conceived. Initial excitement about IL-1 and IL-2 as therapeutics did not yield hoped for breakthroughs. On the other hand, the cytokine field has yielded several very important therapeutics including anti-TNFα antibodies (3), RANKL (4), Erythropoietin or G-CSF (5). The development of these therapeutics has validated the original hopes in the field.

Unfortunately, the cytokine field remains a nomenclature minefield. The interleukins ended up being very difficult to organize. It is still unclear what qualifies a novel cytokine to receive the "interleukin" designation. For example, among the >40 human chemokines (chemotactic cytokines) only one received interleukin designation (interleukin 8). Conversely, there are many interleukins that are related evolutionarily to each other but this is not apparent from their names (IL-4 and IL-13, IL-2, IL-15 and IL-21, IL-10, and IL-22, etc.). In retrospect, the term "interleukin" had a significant advantage: it is a "neutral" designation, one that does not describe a specific characteristic or biological activity. In contrast, consider cytokines like interferon gamma (IFNγ); which is a major immunoregulatory cytokine, and this is what it is known for (not its "interferon" bioactivity). It is a major macrophage activator [including induction of antigen presenting activity (6)]. Thus, this is an example of a cytokine that received a name based on one of the first biological activities detected, even though it is not one of the most relevant (that it would eventually be shown to have).

CYTOKINES, EVOLUTION, AND NOMENCLATURE

The importance of good nomenclature can be explained by reviewing the experience of naming an important subfamily of cytokines, namely, the chemotactic cytokines or chemokines. As we shall see, the development of a systematic nomenclature for this subfamily lead to important insights into its evolution.

The chemokines represent one of the largest subfamilies of cytokines. There are more than 48 human chemokines described. This family is an excellent example of both nomenclature pitfalls as well as the power of studying a family in the context of its molecular evolution. When the first chemokines were identified, all of them were found to belong to two subclasses: the CXC family (where the first two cysteines were separated by another aminoacid) that tended to attract neutrophils, and the CC family that attracted monocytes and selected T cell subpopulations. Importantly, all the genes encoding CXC chemokines were located in a cluster in human chromosome 4 while the CC chemokines were located in a cluster in human chromosome 17 (7). However, later on other chemokines were identified, and a highly significant one was lymphotactin (now called XCL1) whose encoding gene was located not in any of those clusters but instead in chromosome 1 (8). Subsequently many other chemokines were identified and their genes, like lymphotactin, were located all over the genome (not in the original CXC or CC chemokine clusters).

Now that we know most (if not all) the members of the chemokine superfamily, an interesting evolutionary story has emerged. The chemokines can be subdivided into inflammatory and homeostatic, depending on their expression patterns (homeostatic are expressed without apparent stimuli in selected tissues or organs while the inflammatory typically are expressed during inflammatory conditions). Interestingly, the chemokines whose genes were located in clusters were the inflammatory chemokines, while the genes encoding the homeostatic chemokines were instead located in isolated chromosomal locations away from the clusters. This genomic arrangement can be explained evolutionarily as follows: The oldest and most conserved chemokines are the homeostatic subfamily, and their genes are located in isolated chromosomal locations because of the process through which the chemokine superfamily arose (gene duplication). In this process, a given chemokine gene would undergo duplication, and the resulting offspring genes would be located in the same chromosomal location and their encoded chemokines would bind the same receptor. These "offspring" chemokines would be free to undergo their own individual evolution (as a result of mutations) that would make them valuable to the host and favor its survival. However, if such a process occurred in a chemokine gene with an important function in either homeostasis or development, the chances that the affected organism would survive and pass on this trait to its offspring were not very good. This explains why chemokines with important developmental functions are very well conserved. An excellent example is CXCL12, which is very important during fetal development of various organs (7). In contrast, chemokines of the inflammatory class regularly underwent gene duplication (probably in recent evolutionary times) and therefore their genes are still located in the same location (clusters). Furthermore, the "offspring" genes of these events still bind the same receptors as the original unduplicated precursor. Thus, the evolution of homeostatic chemokines was likely conservative or static while the evolution of inflammatory chemokines was very dynamic. This explains why the inflammatory chemokines tend to share receptors, while the homeostatic chemokines mostly exhibit a single chemokine-receptor relationship (7). The reason most inflammatory chemokines arose was most likely to confer protection from a particular pathogen that a given species may have encountered. For the latter reason, deletion of a particular inflammatory chemokine is unlikely to result in heavily compromised survival of the mutated organism. In humans, this effect is evident in the delta-32 mutation of the CCR5 receptor. Humans affected with this mutation (which results in lack of expression of CCR5) cannot be infected with the AIDS virus (HIV) (9). Conversely, however, individuals carrying the delta 32 CCR5 mutation can be very susceptible to West Nile virus (10).

The conclusion that inflammatory chemokines likely arose recently in evolution is also supported by the observation that they often do not correspond well between species. For example, CXCL8 (Interleukin 8) exists in humans but not in mice (11). This observation can be explained by postulating that CXCL8 arose after the evolutionary separation of the ancestors that gave rise to humans and mice. Following this event, human ancestors

have had different ":infectious experiences" than the ancestors of mice. Hence, the inflammatory chemokines present in each species today reflect the "infectious experience" of the ancestors of each species.

This evolutionary model has important implications. For example, the chromosomal location of a particular chemokine can allow us to make predictions about the phenotype of, for example, knockout mice for each chemokine. Knockouts of homeostatic chemokines will likely show a more dramatic phenotype than inflammatory chemokines. Furthermore, if two chemokines share the same chromosomal location, they are likely to share the same receptor (for example, both CCL19 and CCL21 bind CCR7) (11).

Importantly, this evolutionary model is applicable to many superfamilies in the genome and particularly to other cytokines. For example, the genes for IL-4 and IL-13 are located close to each other in human chromosome 5 and their receptors share several features (12).

I can now explain why it was important to talk about gene evolution in immunology in an article focused on nomenclature. The reason is that it was precisely because of nomenclature issues in the chemokine superfamily that we came to understand the evolution of this superfamily. By the year 2000, the nomenclature of the chemokines had become so complicated and confusing that even among experts, the only way to figure out which chemokine we were talking about was to refer back to its sequence. At this point it became obvious that we needed a new standardized nomenclature. The new proposed nomenclature built on the chemokine receptor nomenclature which already existed, but replaced "receptor" for "ligand" (i.e., R for L). Thus, the ligands became CXCL (+ a number) or CCL (+ a number). Luckily, the groups annotating the genome had already allocated numbers to the chemokines but had used a different abbreviation (Small Cytokine subfamily A: SCYA for CC chemokines chemokines or SCYB for CXC chemokines chemokines). Thus, we ended up with CCL21, for example, for a CC chemokine ligand whose gene was originally designated *SCYA21*.

The availability of this new nomenclature allowed experts in the field to produce new figures depicting all the superfamily. Some of these showed the correspondence between receptors and ligands, and the chromosomal locations of the latter. What became immediately apparent was that chemokines whose genes were in the same chromosomal location tended to have the same chemokine receptors; it also became obvious that the genes of the homeostatic chemokines were located throughout the genome while the inflammatory chemokines were in clusters and the latter did not correspond well between species (11). In other words, the new nomenclature allowed us to take a "global view" of this superfamily that fit a compelling evolutionary model for this subfamily of cytokines.

This is therefore a nomenclature story that led to a significant scientific advance. It also underscores the importance of developing a logical nomenclature that has the strong potential to facilitate the study of a particular field.

In the case of the cytokines, there are several superfamilies whose evolution parallel the chemokines. These include the Tumor necrosis factors, the transforming growth factors, and the interferons, among others. I should point out that there are still new cytokines to be identified, if not specifically of importance in immunology, certainly produced in other organs where they likely play an important function. Recently, a study highlighted the fact that most researchers work only on a minority of human genes (13). This situation suggests that there are still many novel genes to be identified and they will need names. We recently identified one of these novel genes (*C17ORF99*) which encodes a novel cytokine we called Interleukin 40 (14).

I think that it is important, when describing a novel gene/molecule, to carefully think about the implications of the name proposed for such a molecule, because it will likely affect the field of research that its discovery will generate. It may be especially important to avoid cheeky or philosophical names. Perhaps a systematic nomenclature that takes structural features or relation to a particular protein family (derived from analyses of characteristics of the encoded protein), where the gene is expressed, rather than its nascent function may be the most likely to prevent a future confusing situation. An estimated 10% of the human genome encodes secreted proteins, and therefore there likely remain many cytokine-like proteins to be described. I hope that these insights may help choose better nomenclature for these proteins.

AUTHOR CONTRIBUTIONS

AZ conceived and wrote this opinion on cytokine and chemokine nomenclature based on his experience in the field.

FUNDING

Publishing costs for this article were partially covered by the International Union of Immunological Societies (IUIS).

ACKNOWLEDGMENTS

I would like to thank Dr. Menno Van Zelm for interesting conversations on this topic and Dr. Osamu Yoshie for his insights and contributions to the nomenclature of chemokine ligands.

REFERENCES

1. Fiorentino, D. F., Zlotnik, A., Vieira, P., Mosmann, T. R., Howard, M., Moore, K. W., et al. (2016). Pillars article: IL-10 acts on the antigen-presenting cell to inhibit cytokine production by Thl cells. J. Immunol. 1991, 146, 3444–3451. *J Immunol.* 197, 1531–1538.
2. Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A., and Coffman, R. L. (1986). Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* 136, 2348–2357.
3. Burmester, G. R., Lin, Y., Patel, R., Van Adelsberg, J., Mangan, E. K., Graham, N. M., et al. (2017). Efficacy and safety of sarilumab monotherapy versus adalimumab monotherapy for the treatment of patients

with active rheumatoid arthritis (MONARCH): a randomised, double-blind, parallel-group phase III trial. *Ann. Rheum. Dis.* 76, 840–847. doi: 10.1136/annrheumdis-2016-210310
4. Tella, S. H., and Gallagher, J. C. (2014). Prevention and treatment of postmenopausal osteoporosis. *J. Steroid Biochem. Mol. Biol.* 142, 155–170. doi: 10.1016/j.jsbmb.2013.09.008
5. Rupolo, M., Lleshi, A., Cacopardo, B., Michieli, M., and Berretta, M. (2013). Hematopoietic growth factors support in the elderly cancer patients treated with antiblastic chemotherapy. *Anticancer Agents Med. Chem.* 13, 1438–1443. doi: 10.2174/18715206113136660352
6. Zlotnik, A., Shimonkevitz, R. P., Gefter, M. L., Kappler, J., and Marrack, P. (1983). Characterization of the gamma-interferon-mediated induction of antigen-presenting ability in P388D1 cells. *J. Immunol.* 131, 2814–2820.
7. Zlotnik, A., and Yoshie, O. (2000). Chemokines: a new classification system and their role in immunity. *Immunity* 12, 121–127. doi: 10.1016/S1074-7613(00)80165-X
8. Kelner, G. S., Kennedy, J., Bacon, K. B., Kleyensteuber, S., Largaespada, D. A., Jenkins, N. A., et al. (1994). Lymphotactin: a cytokine that represents a new class of chemokine. *Science* 266, 1395–1399. doi: 10.1126/science.7973732
9. Hutter, G., Bodor, J., Ledger, S., Boyd, M., Millington, M., Tsie, M., et al. (2015). CCR5 targeted cell therapy for HIV and prevention of viral escape. *Viruses* 7, 4186–4203. doi: 10.3390/v7082816
10. Lim, J. K., and Murphy, P. M. (2011). Chemokine control of West Nile virus infection. *Exp. Cell Res.* 317, 569–574. doi: 10.1016/j.yexcr.2011.01.009
11. Zlotnik, A., and Yoshie, O. (2012). The chemokine superfamily revisited. *Immunity* 36, 705–716. doi: 10.1016/j.immuni.2012.05.008
12. Wang, T., and Secombes, C. J. (2015). The evolution of IL-4 and IL-13 and their receptor subunits. *Cytokine* 75, 8–13. doi: 10.1016/j.cyto.2015.04.012
13. Stoeger, T., Gerlach, M., Morimoto, R. I., and Nunes Amaral, L. A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biol.* 16:e2006643. doi: 10.1371/journal.pbio.2006643
14. Catalan-Dibene, J., Vazquez, M. I., Luu, V. P., Nuccio, S. P., Karimzadeh, A., Kastenschmidt, J. M., et al. (2017). Identification of IL-40, a novel B cell-associated cytokine. *J. Immunol.* 199, 3326–3335. doi: 10.4049/jimmunol.1700534

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

REVIEW
published: 30 September 2020
doi: 10.3389/fimmu.2020.559166

Subsets of CD1c$^+$ DCs: Dendritic Cell Versus Monocyte Lineage

Lukas Heger[1†], Thomas P. Hofer[2†], Venetia Bigley[3], I. Jolanda M. de Vries[4,5], Marc Dalod[6], Diana Dudziak[1,7,8,9†] and Loems Ziegler-Heitbrock[10*†]*

[1] Laboratory of Dendritic Cell Biology, Department of Dermatology, Friedrich-Alexander University of Erlangen-Nürnberg (FAU), University Hospital Erlangen, Erlangen, Germany, [2] Immunoanalytics-Tissue Control of Immunocytes and Core Facility, Helmholtz Centre Munich, Munich, Germany, [3] Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom, [4] Department of Tumor Immunology, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, Netherlands, [5] Department of Medical Oncology, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, Netherlands, [6] Aix Marseille Univ, CNRS, INSERM, CIML, Centre d'Immunologie de Marseille-Luminy, Marseille, France, [7] Deutsches Zentrum Immuntherapie (DZI), Erlangen, Germany, [8] Comprehensive Cancer Center Erlangen-European Metropolitan Area of Nuremberg (CCC ER-EMN), Erlangen, Germany, [9] Medical Immunology Campus Erlangen, Erlangen, Germany, [10] Monocytomics Research, Munich, Germany

Currently three bona fide dendritic cell (DC) types are distinguished in human blood. Herein we focus on type 2 DCs (DC2s) and compare the three defining markers CD1c, CD172, and CD301. When using CD1c to define DC2s, a CD14$^+$ and a CD14$^-$ subset can be detected. The CD14$^+$ subset shares features with monocytes, and this includes substantially higher expression levels for CD64, CD115, CD163, and S100A8/9. We review the current knowledge of these CD1c$^+$CD14$^+$ cells as compared to the CD1c$^+$CD14$^-$ cells with respect to phenotype, function, transcriptomics, and ontogeny. Here, we discuss informative mutations, which suggest that two populations have different developmental requirements. In addition, we cover subsets of CD11c$^+$CD8$^-$ DC2s in the mouse, where CLEC12A$^+$ESAM^{low} cells, as compared to the CLEC12A$^-$ESAM^{high} subset, also express higher levels of monocyte-associated markers CD14, CD3, and CD115. Finally, we summarize, for both man and mouse, the data on lower antigen presentation and higher cytokine production in the monocyte-marker expressing DC2 subset, which demonstrate that the DC2 subsets are also functionally distinct.

Keywords: DC2, CD1c, CD172, CD301, CD14, dendritic cells, DC subsets

OPEN ACCESS

Edited by:
Francesca Granucci, University of Milano-Bicocca, Italy

Reviewed by:
Francesco Borriello, Boston Children's Hospital and Harvard Medical School, United States
Muriel Moser, Université libre de Bruxelles, Belgium

***Correspondence:**
Diana Dudziak
diana.dudziak@uk-erlangen.de
Loems Ziegler-Heitbrock
LZH@monocyte.eu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Molecular Innate Immunity, a section of the journal Frontiers in Immunology

Received: *05 May 2020*
Accepted: *14 September 2020*
Published: *30 September 2020*

Citation:
Heger L, Hofer TP, Bigley V, de Vries IJM, Dalod M, Dudziak D and Ziegler-Heitbrock L (2020) Subsets of CD1c$^+$ DCs: Dendritic Cell Versus Monocyte Lineage. Front. Immunol. 11:559166. doi: 10.3389/fimmu.2020.559166

INTRODUCTION

In human blood, cells with dendritic cell (DC) properties have been classified as plasmacytoid DCs (pDCs), as CD141$^+$ DCs and as CD1c$^+$ DCs (1–3). CD141$^+$ DCs are also termed DC1s or cDC1s, while CD1c$^+$ DCs are defined as DC2s or cDC2s, with "c" standing variously for conventional or classical (4).

The pDCs express CD123 and CD303 and are characterized by their unique ability to produce high amounts of type I Interferon (5).

The $CD141^+$ DCs co-express CD370 (CLEC9A) (6–8) and they are described to activate $CD8^+$ T cells *via* the MHC class I pathway including cross-presentation of exogenous antigen to $CD8^+$ T cells (9–11). Their high ability to cross-present antigen from necrotic cells may be due to the expression of CLEC9A, since this receptor was shown to efficiently bind necrotic cells (12) *via* binding to actin filaments (13).

$CD1c^+$ DCs can present antigen to both $CD4^+$ and to $CD8^+$ T cells (9, 14), however, when cultured with necrotic cells then they are inferior to $CD141^+$ DCs in cross-presentation of necrotic cell derived antigen (9). The $CD1c^+$ DCs form the largest DC subset in human lympho-hematopoietic tissues (8). Due to their efficacy in antigen presentation and T cell activation, $CD1c^+$ as well as $CD141^+$ DCs are attractive cell populations for vaccination studies with primary blood DCs (15, 16).

For all of these three DC types, at least two subsets have been described: for the pDCs a $CD2^-$ and a $CD2^+$ subset has been reported (17), for $CD141^+$ DC there is a $XCR1^-$ and a $XCR1^+$ subset with the $XCR1^-$ cells being the putative precursors of the $XCR1^+$ DCs (18). Finally, within the $CD1c^+$ DC population a differential expression of CD5 and of the monocyte-associated CD14 molecule has been reported. The $CD14^+$ subset shows higher expression levels for several additional monocyte associated markers. This prompts the question whether the $CD14^+$ and $CD14^-$ subsets have a different ontogeny and specifically whether the $CD1c^+$ $CD14^+$ cells are linked to the monocyte lineage. With a focus on man and mouse, these questions will be addressed herein.

Markers to Define DC2 Cells

The initial question is, whether there are reliable markers in man and mouse to define DC2s as compared to $CD141^+$ DCs and to monocytes/macrophages. There are three markers used for DC2s and these are i) CD1c, ii) SIRPα (CD172a) and iii) CLEC10A (MGL or CD301). For the purpose of this review, we will preferentially use the CD nomenclature.

CD1c is a frequently employed marker for DCs in man (1). CD1c is part of the MHC-like CD1 family of genes and it is involved in the presentation of lipid-based antigens to T cells (19). Importantly, while CD1c is found in many species including horses and panda bears, no murine homologue could be identified.

In human blood, CD1c was consistently found to label a population distinct from $CD141^+$ DCs and from classical monocytes (20). In addition, CD1c expression is strongly expressed on almost all B cells (21), making it important to exclude $CD19^+CD20^+$ B cells when defining $CD1c^+$ DCs. Moreover, it had been noted early on that CD1c, even after exclusion of B cells, is not restricted to DCs since it can be induced readily on monocytes by culture with GM-CSF within one day (22). Also, CD1c can be found on $CD141^+$ DCs after FLT3L injection into apparently healthy volunteers (23). Of note, even $CD141^+$ cells isolated from human skin appeared to co-express CD1c (24). Taken together, although the marker CD1c is widely used for the description of the DC2 subset, one should be aware of the fact that the molecule is not uniquely expressed on the DC2s, when performing flow cytometry or immunohistological analyses.

CD172a (SIRP-α) is another marker frequently used to define DC2s. CD172a is a transmembrane glycoprotein, consisting of three extracellular Ig-domains and two intracellular ITIM motifs that mediate negative signals after binding of CD47 to the N-terminal Ig-domain (25).

In man, CD172a is expressed by blood and tissue $CD1c^+$ cells but it is low on $CD141^+$ DC1s in various tissues (26). However, CD172a is also expressed by granulocytes and by monocytes (27) and this is also the case in pigs (28). Therefore, several additional markers are needed for unequivocal identification of DC2 cells in blood and tissue when using CD172a.

Similar to man, CD172a also selectively stains mouse DC2s but not DC1s. In the mouse spleen, CD172a is strongly expressed by lineage-negative $CD11c^+CD4^+$ but not by $CD11c^+CD8^+$ DCs (29), which is consistent with a selective staining of DC2s. Also, in mouse thymus a $CD11c^+CD8^+SIRP\alpha^-$ and a $CD11c^+$ $CD11b^+CD8^-SIRP\alpha^+$ (=$CD172a^+$) cell population was described, and these represent the DC1 and DC2 subsets, respectively (30, 31). Others found, however, that CD172a is not completely absent from $CD103^+$ DC1 cells, since in ocular mucosa it is expressed at a low level by these cells (32). Still, it was suggested that mouse $CD172a^+$ DC2s can be clearly separated from $CD172a^{-/low}$ DC1s when the latter are defined *via* XCR1 (33). In conclusion, CD172a is a suitable marker for the definition of DC2s, but as several other cell types express this marker, these cells need to be carefully excluded in flow cytometry analyses.

CD301 (CLEC10A, MGL) has been suggested more recently as a defining marker for human $CD1c^+$ cells (8, 34). CD301 is identical to MGL (macrophage C-type galactose/*N*-acetylgalactosamine-specific lectin) and it acts as an endocytic receptor. In the mouse, it was cloned from elicited peritoneal cells (35) and the human gene was cloned from monocytes after 7-day culture with IL-2 (36). CD301 is expressed by monocytes cultured with GM-CSF and IL-4 for 7d (37). Also, very low levels of CD301 mRNA and protein were reported for intermediate monocytes (38). Hence, there is an apparent association of CD301 with monocytes/macrophages.

In this context, Heger et al. (34) have assessed CD301/CLEC10A for its suitability as a marker for DC2s. In these studies, CD301 was highly selective for $CD1c^+$ DCs. Only a small fraction of thymic B cells and a subset of monocytes/macrophages in the spleen was found positive for CD301 under steady-state conditions (34). Therefore, CD301 appears to have great potential as a unique marker for DC2s in man, with only some expression on monocytes/macrophages to be considered.

In the mouse, CD301 exists in two forms with different carbohydrate specificities, namely MGL1 and MGL2 (39). Based on structure, expression pattern, and carbohydrate specificity, mouse MGL2 (also termed CD301b) appears to be the homolog of human MGL (CLEC10A, CD301) (40). Staining of bone marrow cells with anti-MGL antibodies identifies a population of cells that is solely positive for MGL1 and another population positive for both MGL1 and MGL2. Hence, antibodies against

MGL1 stain more cells, and this includes pDCs and macrophages. Also, anti-MGL2 stains peritoneal macrophages (40). Additional studies using a Mgl2-DTR/GFP DTR-cell-depleting mouse model suggest a role of MHC class II$^+$ CD11c$^+$CD301b$^+$ cells in resistance against HSV2, and these cells were suggested to be DCs (41). Further studies reported on CD11c$^+$CD301b$^+$ cells, which were addressed as DCs (42), while F4/80$^+$CD11b$^+$CD206$^+$CD301b$^+$ were defined as macrophages (43). Also, CD301b$^+$CD11c$^+$CD11b$^+$ MHCclassII^{high}F4/80^{int}CD206$^+$ mononuclear phagocytes were described in various tissues including fat, liver, and muscle with very few cells seen in blood (44). In addition, Langerhans cells in the mouse skin show a strong signal for MGL (45) but it still needs to be determined whether this is CD301a or CD301b. Taken together, CD301 in the human system and CD301b in the mouse system are promising identifiers of DC2s, but additional markers and a careful approach are required for correct identification of these cells.

Summary Statement on DC2 Markers

While in the past, CD1c has been the main marker to define human DC2s, it may well be that CD172a and CD301 might serve a similar function. Comparative analysis may be helpful to define, which of these markers or which combination thereof is most appropriate to define DC2 cells.

In summary, the three markers that can be used to define DC2 cells (CD1c, CD172a/SIRPα, and CD301/CLEC10A/MGL) are not exclusively expressed by these cells. Therefore, they need to be combined with additional markers to exclude B cells, pDCs, and monocytes/macrophages, as appropriate.

DC2 Markers in Inflammation

While the expression profiles of CD1c, CD172a, and CD301 apply to homeostasis, additional markers may have to be added in inflammatory disease where cytokines can induce DC2-associated markers on other cell types. For example, as mentioned earlier CD1c can be found on CD141$^+$ DCs after FLT3L injection into apparently healthy volunteers (23). With the singular use of CD1c as a DC2 marker, such FLT3L-induced cells would be wrongly assigned to the DC2 lineage.

Since monocytes are CD172a-positive in the steady state and since *in vitro* culture of monocytes with GM-CSF can induce expression of CD1c (22) and of CD301 (37) it is important to exclude monocytes/macrophages, when defining DC2 cells in blood and more importantly in tissue. This is particularly relevant in the context of inflammatory diseases when cytokines like GM-CSF are increased (46). An informative example is sickle cell disease, which goes along with increased blood GM-CSF levels (47), with increased numbers of CD16$^+$ monocytes (48) and with expression of CD1c on monocytes (49). It remains to be determined whether these monocytes in the blood of sickle cell disease patients are akin to the CD14$^+$ subset of CD1c$^+$ DCs. In any event, these cells may contribute to the pathophysiology of the disease *via* production of inflammatory cytokines.

Overall, these deliberations show that determination of DC2s in inflammatory conditions requires additional steps in order to unequivocally define these cells.

CHARACTERIZATION OF SUBSETS OF DC2 IN MAN

DC2 Subsets in Human Blood

In the 2010 nomenclature proposal it had been noted for the CD1c$^+$ myeloid DCs in human blood that these DCs can be separated into CD14$^-$ and CD14^{low} cells (1). This was based on the original studies by Thomas and Lipsky (50) demonstrating antigen presenting activity in a population of CD33^{++}CD16$^-$CD14$^+$ cells, cells that later were shown to be CD1c$^+$ (51). Of note, the typical approach to DC analysis in human blood starts with the exclusion of lymphocytes and of monocytes. Monocytes are excluded using a CD14 antibody, but depending on the reagent and the conditions used, this may or may not remove CD14^{low} cells.

An example of the CD14 expression pattern of CD1c$^+$ DCs is given in **Figure 1**. Compared to the isotype control, there is strong CD1c-staining among the CD14$^-$ cells and a gradually decreasing expression of CD1c among the CD14^{low} cells (**Figure 1A**). The isotype control shows a few events within the CD1c$^+$ DC2 gate (**Figure 1B**).

The cells within the CD1c gate can then be separated into CD1c$^+$CD14$^+$ and CD1c$^+$CD14$^-$ cells (**Figure 1C**) with the CD14$^+$ cells (green in **Figure 1A**) accounting for about 40% of all CD1c$^+$ DCs. The CD5$^+$ cells are distinct from the CD14$^+$ cells.

Recently, in a study not excluding CD14$^+$ cells, it was reported that in apparently healthy donors about one third of the CD19$^-$CD1c$^+$ cells are CD14$^+$ (15). These CD1c$^+$CD14$^+$ cells, compared to the CD1c$^+$CD14$^-$ subset, were shown to express similar levels of HLA-DR and CD33 but higher levels of CD11b and clearly higher levels of PD-L1 (CD274) (15). Upon LPS stimulation, these cells showed a trend to produce higher amounts of TNF and IL-10, but they were less efficient in inducing T cell proliferation induced by allogeneic leukocytes. The T cell proliferation could be improved by addition of an anti-PD-L1 antibody (15). Furthermore, while the CD1c$^+$CD14$^-$ subset readily induced IFNγ production in CD4$^+$ T cells, the CD1c$^+$CD14$^+$ subset completely failed to do so. Only when CD1c$^+$CD14$^+$ DCs were stimulated with GM-CSF or LPS then a low level of IFNγ production could be induced. This suggests that the CD1c$^+$CD14$^+$ do not induce but rather impede T cell proliferation and differentiation toward the TH1 lineage.

In blood of melanoma patients with metastatic disease, the frequency of the CD14$^+$ subset of CD1c$^+$ cells in blood was found to be increased more than three-fold. Upon vaccination with antigen-loaded CD1c$^+$ DCs, patients with a high proportion of the CD1c$^+$CD14$^+$ subset showed lower T cell proliferation to control antigen (15). This underscores the notion that CD1c$^+$CD14$^+$ cells in cancer patients are not potent T cell stimulators but rather show suppressive activity. Together, these findings were taken to design an optimized cellular vaccine, in which the CD14$^+$ subset is removed from the CD1c$^+$ DC product for vaccination of patients with melanoma and other malignancies (15, 52).

Transcriptome analyses comparing the CD1c$^+$CD14$^+$ cells to CD1c$^+$CD14$^-$ cells showed the CD1c$^+$CD14$^+$ cells to express

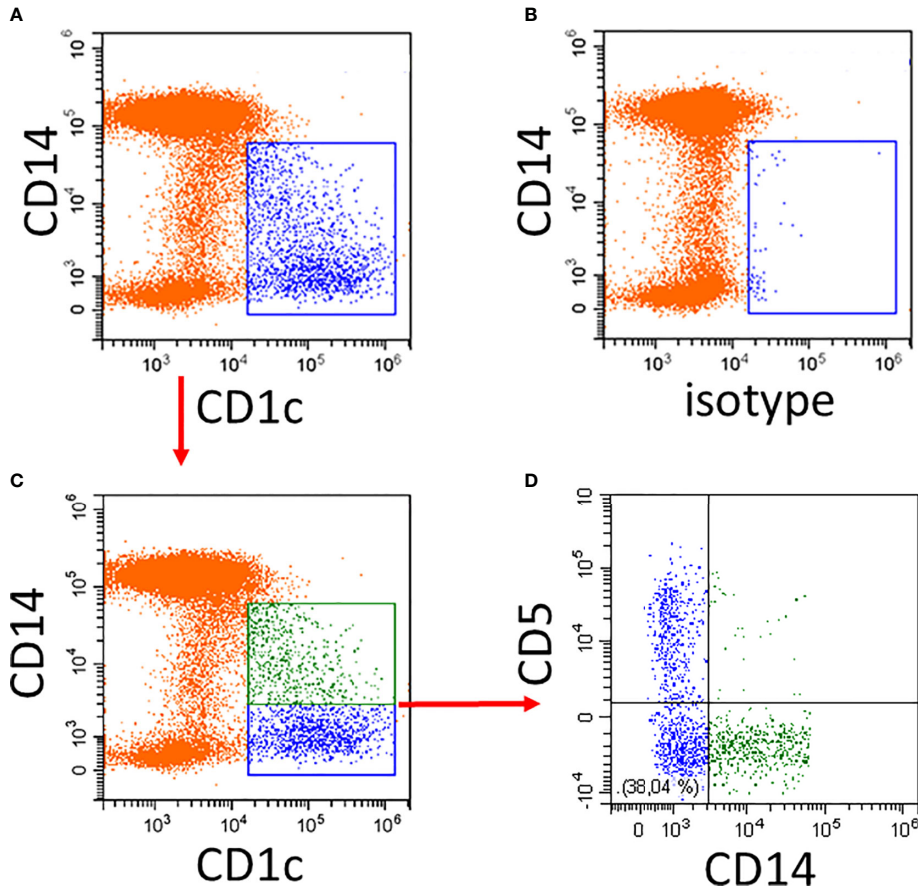

FIGURE 1 | Illustration of CD1c$^+$ DCs and its subsets in human blood. Whole blood was stained with CD14, CD16, CD19, HLA-DR, and CD1c antibodies and the expression of CD1c **(A)** compared to isotype control **(B)** was analyzed on HLA-DR$^+$ non-B cells. Of note, the CD14^{low} CD1c$^-$ cells in **(A)** represent the CD16$^+$ monocytes. **(C, D)** show additional staining for CD5. In the example in **(C)**, the CD1c$^+$ cells are divided into a CD14$^+$ subset (green) and a CD14$^-$ subset (blue). As shown in **(D)**, the CD14$^-$ subset in blue can be further subdivided into CD5$^+$ and CD5$^-$ cells. In the lower left there is a population of CD14$^-$ CD5$^-$ cells. Red arrows indicate the gating sequence.

higher mRNA levels for MafB and the CSF1-receptor (CD115) and lower levels for FLT3 and IRF4. In addition, higher levels for TLR7, TLR8, CLEC7A (CD369), CLEC12A (CD371), and CLEC12B were found for the CD1c$^+$CD14$^+$ DCs (15).

Hierarchical clustering using these transcriptome data suggested that the CD1c$^+$CD14$^+$ DC2 subset to be in between classical monocytes and the CD1c$^+$CD14$^-$ cells but closer to the classical monocytes (15). However, a comparison to a comprehensive set of blood DCs and monocytes is still required in order to assign them to either monocytes or DCs, when using transcriptomics as a tool. The central features of DC2 subsets in this study by Bakdash et al. are summarized in **Figure 2**.

The existence of two subsets of DC2 in man was confirmed recently in single cell sequencing studies on peripheral blood mononuclear cells (53). Here, both subsets were positive for CD1c, CLEC10A and FcεR1A. One subset expressed higher transcript levels for MHC class II molecules and CD1c, while the other was higher for S100A8/9, ANXA1, F13A, VCAN (versican), FCN1 (ficolin 1), RNase2, CD163, and CD14. Many of the latter molecules are associated with monocytes but both CD1c$^+$ subsets clustered separately from monocytes in this study. This work by Villani et al. is also listed in **Figure 2**.

Furthermore, Schroder et al. have studied the properties of cells isolated with CD1c-magnetic beads from human blood mononuclear cells, and they noted a CD1c$^+$CD14$^-$ population and a CD1c$^+$CD14$^+$ population, with the latter showing low level CD1c (54). Here, a higher expression of CD135 (FLT3) on the CD1c$^+$CD14$^-$ cells and a higher expression of CD115 (M-CSFR) on the CD1c$^+$CD14$^+$ cells was observed, and the CD1c$^+$CD14$^+$ subset was interpreted to represent monocytes.

In early studies, a differential expression of CD5 had been reported on human blood DCs (55, 56). More recently in a 2017 study, the lineage-negative HLA-DR$^+$CD123$^-$CD11c$^+$CD1c$^+$ cells have been subdivided into CD5^{low} and CD5^{high} cells (57). Gene expression analysis showed higher SIGLEC6 and IRF4 transcripts in the CD1c$^+$CD5^{high} cells, while the CD1c$^+$CD5^{low}

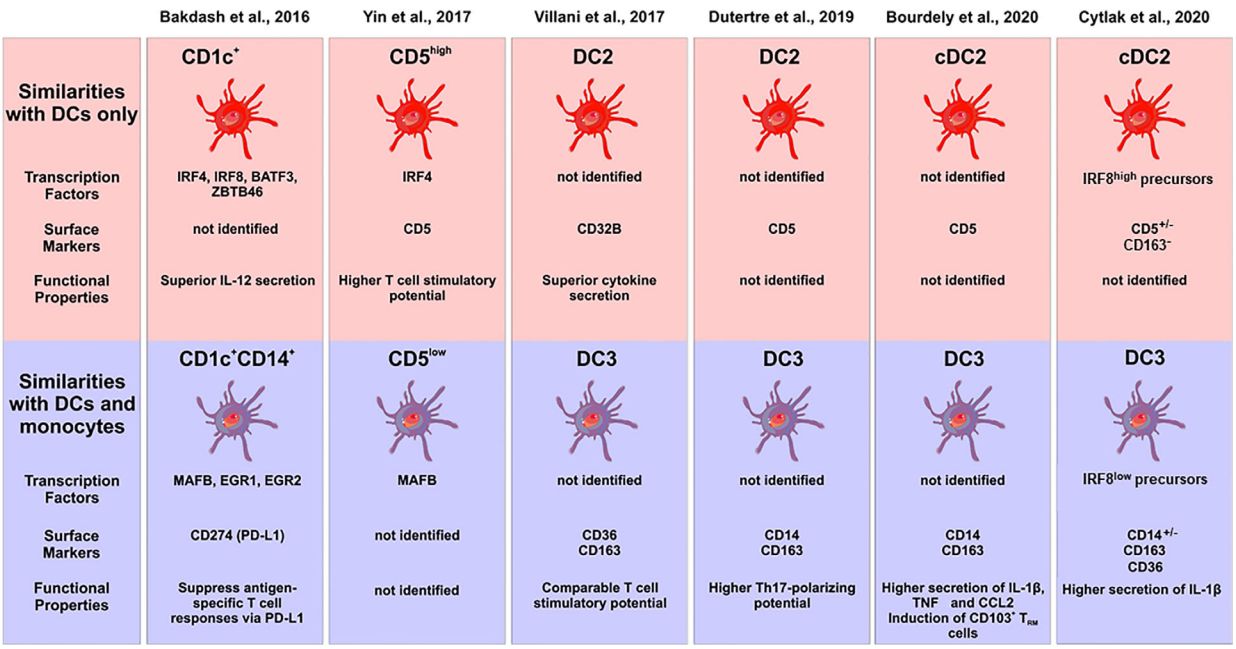

FIGURE 2 | Characterization of subsets of human DC2s in the recent literature. The different studies are listed at the top, the upper panel gives the subsets with pure DC features, the lower panel shows the subsets with monocyte features. Characteristic transcription factors, cell surface markers and functional properties are given when available. The cellular images are provided and adapted from Servier Medical Art (smart.servier.com).

cells expressed higher levels of CD14, MAFB, S100A8/9, RNAse2, CD163, and Ficolin1. A few of these transcripts were tested at the protein level and here the higher expression of CD14 and S100A9 was confirmed for the $CD1c^+CD5^{low}$ cells (57). This work by Yin et al. is listed in **Figure 2**.

The reciprocal gene expression pattern for $CD1c^+CD5^{high}$ cells (57) and the $CD1c^+CD14^{low}$ cells (15) suggests that these two subsets might be mutually exclusive and that $CD1c^+$ cells might consist of $CD14^-CD5^{high}$ and $CD14^+CD5^{low}$ cells. As illustrated in **Figure 1**, CD5 and CD14 are expressed on distinct cell subsets (see **Figure 1D**).

This pattern is consistent with what has been described by Meyerson et al. (58). The latter study and our illustrative figure demonstrate a population of $CD1c^+CD5^-CD14^-$ cells and the question is, whether this subset represents a distinct population. In this context, Dutertre et al. (59) analyzed DCs and their subsets with an extensive panel of cell surface markers. In this study, more than 300 protein markers were employed and markers HLA-DQ and FcεRIA on DCs and CD88 and CD89 on monocytes were identified as best discriminating markers. On this basis, the DC2s including the $CD14^+$ subset could be phenotypically separated from classical monocytes, albeit there is low level expression of both CD88 and CD89 on DC2s.

DCs are thought to be specifically governed by FLT3 (Fms-Like Tyrosine Kinase 3) and injection of this growth factor into patients was shown to result in a shift of the proportions of DC2 versus classical monocytes. Here, both $CD1c^+CD14^+$ and $CD1c^+CD14^-$ DC2s increased relative to the classical monocytes arguing for a DC nature of the $CD1c^+CD14^+$ cells (59). A more detailed analysis in the same report then revealed four distinct subsets of $CD1c^+$ cells, which are one $CD5^+$ subset and three $CD5^-$ subsets, the latter consisting of $CD14^-CD163^-$, $CD14^-CD163^+$ and $CD14^+CD163^+$ cells. Some salient features of the typical DC2 and the subset with monocyte features is given in **Figure 2** (see Dutertre et al).

Three different phenotypes of DC2 cells in human blood were also described in a 38-marker CYTOF analysis (60). The DC2s showed differential expression levels for CD172a and CD163 and the authors concluded that there were $CD172^{high}CD163^{low}$ and $CD172^{low}CD163^{med}$ and $CD172^{high}CD163^{high}$ DCs. The relationship of these three phenotypes to the $CD14^+$ und $CD5^+$ subsets remains to be determined.

Taken together, among the $CD1c^+$ DCs there is higher expression for several monocyte-associated genes (CD14, CD115, MAFB, S100A8/9, CD163, and Ficolin1) in cells defined either as $CD14^+$ cells or as $CD5^-$.

Lineage Assignment of DC2 Subsets

In order to appropriately address the question of lineage assignment of the subsets of $CD1c^+$ DCs, approaches using a broad panel of different monocytes, macrophages, and dendritic cells are required.

In a recent report on human $CD1c^+$ DC2 subsets, the $CD5^+$ cells were defined as cDC2, while cells positive for CD14 and CD163 were termed DC3 (61). The $CD14^+$ subset of $CD1c^+$ cells, when compared to the $CD1c^+CD5^+$ DCs, was shown to express higher levels of TNF and CCL2 and to induce features of tissue-

resident memory cells in $CD8^+$ T cells (see summary in **Figure 2**). Furthermore, it was demonstrated that GM-CSF but not FLT3L is able to support development of the $CD1c^+CD14^+$ DCs in a humanized mouse model. In *in vitro* studies, GM-CSF was able to induce these cells from a granulocyte-monocyte-dendritic cells precursor (GMDP) but not from common dendritic cell precursor or cMoP, indicating that this subset may have a distinct developmental pathway (61).

In addition, functional studies might be able to address the question of lineage assignment. To this end, recently the ability of DCs to activate the inflammasome and induce the release of IL-1β has been revisited (62) and it was shown that among the GM-CSF induced mouse bone marrow-derived cells only the macrophages but not the DCs were efficient producers of IL-1β. If a relevant IL-1β production by the $CD14^+$ subset of DC2 can be demonstrated then this would add another monocyte characteristic to these cells. Cytlak et al. addressed this question by comparing human $CD1c^+CD5^{+/-}CD163^-$ DC2, and $CD1c^+CD163^+$ DCs (termed DC3) (63). Here, it was noted that the $CD1c^+CD163^+$ cells, when stimulated with a mixture of TLR ligands followed by intracellular staining and flow cytometry, showed IL-1β production as high as monocytes, while the $CD1c^+CD163^-$ cells showed a low level production of this cytokine (see summary in **Figure 2**). Similar results were obtained for IL-10, while the two DC subsets produced comparable amounts of IL-12 (63).

Moreover, DC2s can be generated *in vitro* from $CD34^+$ hematopoietic stem cells (64) and more specifically from cells with the phenotype of MLPs, CMPs, and GMPs (65). The generation of subsets of DC2 was only studied recently (63). Here, CMPs and $CD33^+$ GMPs were found to have $CD1c^+CD14^+$ DC2 potential and the $CD14^+$ subset was shown to segregate with monocytes. In contrast, $CD1c^+CD14^-$ DC2s could be generated from $CD123^+$ GMPs and segregated with pDC and cDC1 potential.

Mutations of genes involved in development of DCs can be informative as to lineage assignment. Homozygous and heterozygous loss of function and dominant negative mutations of the *IRF8* gene have been described. It was shown that bi-allelic loss of function mutations of the *IRF8* gene led to a complete absence of DCs and monocytes (66, 67).

For two independent cases with recurrent disseminated BCG infection and heterozygous dominant negative mutation in the *IRF8* gene, normal numbers of monocyte subsets and no decrease of pDCs and DC1s were initially reported, while there was an apparent reduction in CD1c expression on $CD11c^+$ cells (66). In a later study by the same researchers, in a more detailed analysis using CLEC9A in addition to CD141 for DC1 definition and CD1c for DC2 identification, it was noted that DC1s are, in fact, decreased and the decrease of CD1c expression was confirmed (68).

Cytlak et al. looked at $CD1c^+$ DCs in a kindred with a heterozygous dominant-negative mutation of *IRF8*, which led to a moderate deficiency of this transcription factor (63). This went along with depletion of cDC1 and pDC but increased blood monocytes and normal numbers of $CD1c^+$ DC2s. Analysis of subsets, however, revealed an almost complete absence of the $CD1c^+CD5^+$ cells, while the $CD1c^+CD5^-$ subset was expanded (63). Whereas the moderate level activity of IRF8 is apparently sufficient to allow for development of the $CD14^+$ monocyte subsets and of the $CD1c^+CD14^+$ DC subset, it is not sufficient to allow for generation of $CD1c^+CD14^-CD5^+$ DCs. This indicates that the $CD14^+$ and $CD5^+$ subsets of $CD1c^+$DCs have distinct developmental requirements and that the $CD1c^+CD14^+$ DC2 subset is associated with monocytes. Analysis of additional mutations of genes involved in DC development in man may help to further support this point.

When looking at IRF8 CRISPR/Cas9 knock-out mutation in human *in vitro* induced pluripotent stem cells (iPS), it was noted that the generation of pDC and DC1 cells driven by FLT3L, SCF, GM-CSF, and IL-4 (FSG4) was ablated but the generation of $CD1c^+$ DC2s was unaffected (69). These data indicate that this intriguing *in vitro* system more closely mimics partial IRF8 deficiency *in vivo*. Still, manipulation of the IRF8 gene penetrance may help in unravelling the *in vitro* development of DC2s and their subsets. On the other hand, the generation of DC2s in the absence of IRF8 in this *in vitro* system may be due to the strong effects of the exogenous cytokines, since without lineage specifying cytokines, DC2 cells are absent in these $IRF8^{-/-}$ cultures similar to what is seen in immunodeficient patients (69). Again, it will be important to analyze DC2 subsets in this system.

Moreover, Borriello et al. reported on the expression of the receptor for thymic stromal lymphopoietin (TSLPR), which is induced by stimulation *via* TLR4 in $CD1c^+CD14^{low}$ cells among human blood mononuclear cells (70). TSLP directs type-2 immune reactions and it acts on a broad range of leukocytes including T cells, macrophages, DCs, mast cells, and ILCs indicating that its receptor is widely expressed (71). Borriello et al. noted that induction is specific to the $CD1c^+CD14^{low}$ cells and that there is little induction in $CD16^+$ monocytes. It will be important to assess the induction of the TSLPR in the $CD14^-$ subsets of $CD1c^+$ DC2s, in order to determine whether expression might be specific to the monocyte-related subset.

Taken together, while in the studies by Borriello et al. and Sontag et al. (69, 70) the analysis of DC2 subsets is still outstanding, the data on development of DC2 subsets show evidence of co-segregation of the monocyte-marker expressing DC2 subset with monocytes.

DC2 Subsets in Human Tissue

Because of the ease of accessibility, many human studies are performed on blood samples, but there are also a number of studies that look at DCs in human lymphoid and non-lymphoid tissues (8, 24, 60, 72–75) but only a few studies address subsets of DC2s in tissue.

Of note, interpretation of data in human tissue has to be done with caution: While in blood the monocyte subsets can be clearly defined and dissected from DCs, it is more difficult in tissues to exclude macrophages that may have up-regulated DC2-associated markers. Furthermore, such DC2-marker-positive macrophages may or may not co-exist with bona fide monocyte-lineage derived cells, and this can make lineage-assignment very demanding.

With respect to subsets of DC2s Yin et al. noted $CD5^+$ and $CD5^-$ subsets of $CD1c^+$ DCs in human tonsils and similar to blood the $CD5^+$ cells formed the minor subset (57).

Looking at single cell suspensions isolated from the nasal mucosa, all of the CD1c$^+$ cells expressed low levels of CD14 (76). Another study on nasal, bronchial, and intestinal mucosa showed low-level CD1c expression on the CD11c^{high} subset of CD14$^+$ cells (54).

In the human lung, CD14$^+$ and CD14$^-$ subsets of CD1c were observed in lavage samples and transcriptome analysis demonstrated higher ZBTB46, FLT3, CD83, and CCR7 mRNA levels in the CD14$^-$ subset, while the CD1c$^+$CD14$^+$ subset showed higher CD36, CD163, CD369 (=CLEC7A=Dectin-1), and S100A8/9 (77). Hence, the CD14$^+$ subset of CD1c$^+$ cells in the lung alveoli enriched for monocyte-associated genes. Also, in an analysis of bronchoalveolar lavage samples from healthy volunteers, a BTLA$^+$ and a BTLA$^-$ subset of CD1c$^+$CD11c$^+$ cells were noted and here a higher expression of monocyte/macrophage-associated genes CD14, CD163, S100A8, and CD115 was noted at the transcript level for the BTLA$^-$ cells (78). For the human intestine, Watchmaker et al. identified among the lineage-negative, HLA-DR$^+$CD11c$^+$ cells two subsets of CD172a$^+$ cells, one CD103$^+$ and one CD103$^-$. Hierarchical clustering using transcriptome data demonstrated that the CD103$^-$CD172$^+$ cells clustered with blood monocytes and the authors suggested that they might be monocyte-derived (72).

In summary, there is evidence for a monocyte-marker positive DC2 subset in various human tissues like tonsil, lung, and intestine.

CHARACTERIZATION OF SUBSETS OF DC2 IN THE MOUSE

Differential ESAM Expression for Definition of Mouse DC2 Subsets

For the mouse similar to man, pDCs, DC1s, and DC2s have been described (79–82). However, data on DCs in mouse blood are scarce, and most studies are done on tissue samples. Regarding subsets of DC2, CD11c$^+$CD8$^-$ cells in mouse spleen and mesenteric lymphnodes Kasahara and Clark described subsets of DCIR2$^+$DCAL2$^-$ and DCIR2$^-$DCAL2$^+$, i.e. CLEC4A4$^+$CLEC12A$^-$ and CLEC4A4$^-$CLEC12A$^+$, respectively (83). Here, the CLEC4A4$^-$CLEC12A$^+$ cells exclusively produced TNF and IL12 in response to the TLR9 ligand CpG. Separate studies by Lewis et al. on cells from mouse spleen and intestine revealed that DC2-type cells with the phenotype CD8$^-$CD11c^{++}CD11b$^+$ cells can be subdivided into an ESAM^{high} and an ESAM^{low} subset (ESAM = Endothelial cell–selective adhesion molecule) (84). Here, the ESAM^{high} cells expressed higher levels of MHC class II molecules, while the ESAM^{low} cells showed higher transcript levels for M-CSFR, CCR2, Lysozyme, CD14 and CD36, markers which are typical of the monocyte lineage. Furthermore, the ESAM^{low} subset was able to produce higher levels of TNF and IL-12, when stimulated with the TLR-9 ligand CpG DNA. Also, activation *via* TLR-2, using heat-killed Listeria monocytogenes, led to superior TNF production by ESAM^{low} cells. Phagocytosis of latex beads was similar for both subsets as was the capacity to present *in vitro* OVA peptide using transgenic responder T cells expressing OVA-specific TCR (OT-II cells) (84). However, when testing antigen presentation *in vivo* a robust response of OT-II cells was noted in wild type animals, while in knock-out mice, lacking ESAM^{high} cells, only a minimal response of OT-II cells was observed. This suggests that the ESAM^{high} cells are required for an efficient antigen presentation *in vivo* (84).

Importantly, the ESAM^{low} cells were noted to be CLEC12A^{high} (84) thereby linking the studies by Lewis et al. (84) and Kasahara et al. (83). The conclusion from both studies is that the ESAM^{low}CLEC12A^{high} subset of CD11c$^+$CD8$^-$ mouse DC2s expresses monocyte-associated markers and is more potent in producing cytokines TNF and IL-12.

Also, Lewis et al. (84) demonstrated that blockade of Notch2 signalling led to selective ablation of the ESAM^{high} subset. This ESAM^{high} subset was shown to derive from DC precursors, while the ESAM^{low} cells were suggested to be myeloid progenitor-derived. In other words, Notch2 signalling is required for the development of ESAM^{high} DC2s but not for ESAM^{low} DC2s (84). In an *in vitro* study, DCs were generated from progenitor cell lines by culture with a FLT3L and Notch ligand expressing cell line. Here, transcriptome analysis revealed that with Notch2 activation the CD11b$^+$CD24$^-$CD8$^-$ DC2s were similar to the splenic ESAM^{high} DC2 subset (85).

Mutations of human Notch2 have been described in two families with the patients showing multiple abnormalities of liver, lung, heart, and kidneys and typical facial features (86). At this point, no immunological workup has been published on such patients, and it is therefore not known whether these patients have abnormalities of the immune system.

As mentioned above, a bi-allelic mutation of the IRF8 gene leads to a depletion of DC2 cells in man (66). By contrast, in the mouse it was shown that *Icsbp* (=*Irf8*)$^{-/-}$ animals lack pDC and CD8$^+$ DCs (DC1), while CD4$^+$ DCs (DC2) were readily detected (87, 88). No information on subsets of DC2s is available, and it remains to be determined whether there is a differential effect on ESAM^{high} versus ESAM^{low} subsets in in *Irf8*$^{-/-}$ animals.

Taken together, in the mouse the ESAM expression level can be used to define subsets of DC2 with ESAM^{low} subsets showing features of monocytes.

T-Bet Expressing DC2s

Another recent study by Brown et al. defined in mouse lymphoid organs two subsets of lin$^-$CD90$^-$CD64$^-$Ly6C$^-$CD11c$^+$MHCII$^+$XCR1$^-$CD11b$^+$cDC2s based on their differential expression of the transcription factor T-bet (89). The T-bet$^+$ DC2s, dubbed DC2A, were found to overlap with the ESAM$^+$ DC2s previously described by Lewis et al. (84). The T-bet$^-$ cDC2s, called DC2B, showed higher expression of monocyte-associated genes. However, only a subset of these cells expressed M-CSF-R (CD115) and lysozyme mRNA, reminiscent of monocyte-related DC2 subset, which we discussed earlier in man and mouse. The authors also noted that the T-bet$^-$ population was heterogeneous with respect to expression of C-type lectin receptors and consisted of a CLEC10A$^-$CLEC12A$^-$, of a CLEC10A$^-$ CLEC12A$^+$ and of a CLEC10A$^+$CLEC12A$^+$ population.

Interestingly the T-bet$^+$ cDC2 gene expression signature subset was neither detected in mouse nor in human blood (89). However, gene expression analysis of two samples from melanoma patients identified a cluster of myeloid cells that expressed T-bet target genes

and additional signature genes like AREG and NR4A3. Also, in human spleen CD301$^-$ DCs with transcriptomic similarities to murine T-bet$^+$ DC2A cells were described.

Brown et al. (89) suggest that the human blood CD1c$^+$ cDC2s might be analogous to the mouse T-bet$^-$ CLEC10A$^+$ DC2B, while in human blood there is no homologue of the mouse T-bet$^+$ cDC2s. Clearly more studies, including functional tests, are required in order to collate these recent findings into a unified scheme and to align subsets across species.

DC2 Subsets in the Mouse in Lung

For the lung of wild type mice, it was shown that there are CD14$^+$ cells among a population of CD11b$^+$CD24$^+$CD64$^-$ DC2s (90). Also, the DC2 cells were either Irf4$^+$ or Irf4$^-$ and in *Irf4*$^{-/-}$animals the CD24$^+$ cells were decreased with a residual population remaining. Hence, one might speculate that the CD14$^+$ cells represent an IRF4-independent population of mouse DC2 cells.

Recently, Bosteels et al. have shown that in the mouse lung CD26$^+$CD172$^+$ DC2s in response to type I interferon can up-regulate IRF8 and CD64 and acquire features of both DC1s and monocytes (91). This suggests that there is further complexity of dendritic cells in inflammation, and it stresses the necessity to carefully delineate DC subsets in disease settings, where such inflammatory DC2s increase. Analysis of ESAM and CLEC12A (CD371) in these cells, and their progenitors may help to elucidate the relationship to the two DC2 subsets described for the mouse (83, 84).

Taken together, also in the mouse lung, DC2s and their subsets can be readily detected.

CONCLUDING REMARKS

There is a long-standing effort to dissect monocytes/macrophages and dendritic cells (3). This task becomes easier when the focus is set on bona fide DCs, i.e. the pDCs, CD1c$^+$ DCs, and the CD141$^+$ DCs in human blood. Still, there is concern that there might be overlap between the monocyte and DC lineage in some areas.

Regarding the CD1c$^+$ DCs in blood, the open question then remains whether the CD14$^+$ subset represents bona fide DCs or whether they belong to the monocyte lineage. The expression of markers like CD115 and S100A by these cells as well as their low antigen presenting capacity would support a monocyte nature.

A comparison between man and mouse of the DC2 subsets without and with monocyte features is given in **Figure 3**. In both species the monocyte-related subset (green in **Figure 3**) is characterized by higher expression of cell surface molecules like CD115 and cytokines like TNF.

It remains to be resolved, whether these monocyte features may be explained by i) DC-lineage cells developing monocyte features, by ii) the cells being derived from mature classical monocytes similar to monocyte-derived DCs generated in GM-CSF cultures *in vitro* or iii) whether these cells are derived from a bone marrow precursor in common with monocytes.

Steps to be taken to resolve this issue are

a. comparative single cell transcriptomics including CITE-seq approaches with a large set of prototypic monocytes/macrophages and dendritic cells from the same donor across different tissues,
b. analysis of informative immune-deficiencies including knock-outs to study the mouse CD8$^-$CD11c$^+$CD11b$^+$ ESAM^{low} DCs,
c. development from hematopoietic progenitor cells with or without bar coding, and
d. analysis of informative mutations in clonal hematopoiesis.

As mentioned earlier, a recent study has addressed some of these points and has looked into IRF8 immunodeficiencies and into *in vitro* development of DCs from progenitors. Here,

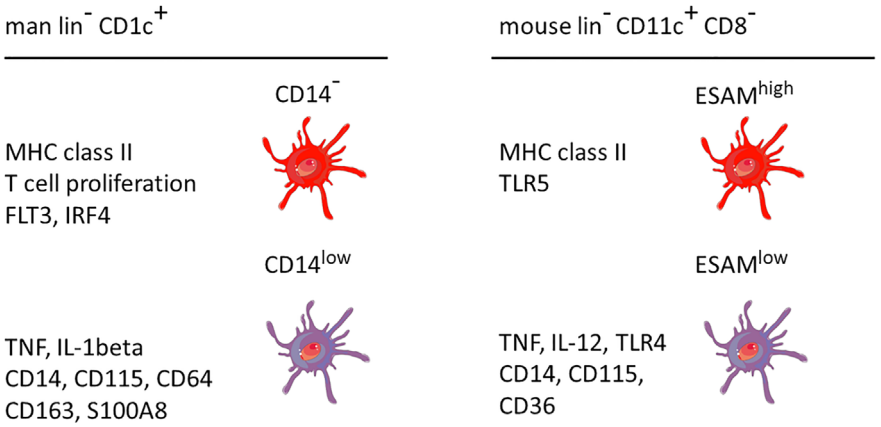

FIGURE 3 | Properties of two main DC2 subsets in man and mouse. The markers are listed based on a higher expression in the respective subset compared to the other subset, i.e. the other subset can also be positive but at a lower level. This cartoon is restricted to the subset with monocyte features as compared to a subset covering the remaining DC2 cells. The latter has been reported to include up to three distinct populations as detailed in the text. lin$^-$ = lineage negative. The human data are a summary of a series of studies compiled in **Figure 2**. The mouse data refer to Lewis et al., 2011, and Kasahara et al., 2012. The cellular images are provided and adapted from Servier Medical Art (smart.servier.com).

differential transcription factor requirements for the $CD14^+$ and the $CD5^+$ subset of DC2s were apparent, and the $CD1c^+CD14^+$ DC2s showed co-segregation of with monocytes (63). While the issue of lineage assignment of the $CD1c^+CD14^+$ cells in human blood still needs to be resolved, it has become clear that there may be more than two subsets of DC2. Therefore, future work will have to address, in man and in the mouse, all subsets of DC2s and their role in homeostasis and in disease in order to arrive at an adequate nomenclature.

AUTHOR CONTRIBUTIONS

LZ-H conceived the project. LZ-H, MD, and DD drafted the paper. All authors contributed to the article and approved the submitted version.

FUNDING

DD was partly supported by grants from the German Research Foundation [Deutsche Forschungsgemeinschaft (DFG)] (CRC1181-TPA7 and DU548/5-1), Agency national research (ANR)/German Research Foundation program (DU548/6-1) Bayresq.Net (Bavarian Ministry of Sciences), and Interdisziplinäres Zentrum für Klinische Forschung (IZKF-A80). LH was supported by Erlanger Leistungsbezogene Anschubfinanzierung und Nachwuchsförderung (ELAN) (DE-17-09-15-1-Heger). VB was supported by the Wellcome Trust grant 101155/Z13/Z.

ACKNOWLEDGMENTS

We thank the members of Dudziak team for their critical comments.

REFERENCES

1. Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of monocytes and dendritic cells in blood. *Blood* (2010) 116 (16):e74–80. doi: 10.1182/blood-2010-02-258558
2. Heidkamp GF, Lehmann CHK, Heger L, Baranska A, Hoffmann A, Lühr J, et al. Functional Specialization of Dendritic Cell Subsets. *Encyclopedia Cell Biol* (2016) 3:588–604. doi: 10.1016/B978-0-12-394447-4.30076-1
3. Amon L, Lehmann CHK, Baranska A, Schoen J, Heger L, Dudziak D. Transcriptional control of dendritic cell development and functions. *Int Rev Cell Mol Biol* (2019) 349:55–151. doi: 10.1016/bs.ircmb.2019.10.001
4. Guilliams M, Ginhoux F, Jakubzick C, Naik SH, Onai N, Schraml BU, et al. Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat Rev Immunol* (2014) 14(8):571–8. doi: 10.1038/nri3712
5. Swiecki M, Colonna M. The multifaceted biology of plasmacytoid dendritic cells. *Nat Rev Immunol* (2015) 15(8):471–85. doi: 10.1038/nri3865
6. Caminschi I, Proietto AII, Ahmet F, Kitsoulis S, Shin Teh J, Lo JC, et al. The dendritic cell subtype-restricted C-type lectin Clec9A is a target for vaccine enhancement. *Blood* (2008) 112(8):3264–73. doi: 10.1182/blood-2008-05-155176
7. Poulin LF, Salio M, Griessinger E, Anjos-Afonso F, Craciun L, Chen JL, et al. Characterization of human DNGR-1+ BDCA3+ leukocytes as putative equivalents of mouse CD8alpha+ dendritic cells. *J Exp Med* (2010) 207 (6):1261–71. doi: 10.1084/jem.20092618
8. Heidkamp GF, Sander J, Lehmann CHK, Heger L, Eissing N, Baranska A, et al. Human lymphoid organ dendritic cell identity is predominantly dictated by ontogeny, not tissue microenvironment. *Sci Immunol* (2016) 1(6). doi: 10.1126/sciimmunol.aai7677
9. Jongbloed SL, Kassianos AJ, McDonald KJ, Clark GJ, Ju X, Angel CE, et al. Human CD141+ (BDCA-3)+ dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. *J Exp Med* (2010) 207(6):1247–60. doi: 10.1084/jem.20092140
10. Bachem A, Guttler S, Hartung E, Ebstein F, Schaefer M, Tannert A, et al. Superior antigen cross-presentation and XCR1 expression define human CD11c+CD141+ cells as homologues of mouse CD8+ dendritic cells. *J Exp Med* (2010) 207(6):1273–81. doi: 10.1084/jem.20100348
11. Crozat K, Guiton R, Contreras V, Feuillet V, Dutertre CA, Ventre E, et al. The XC chemokine receptor 1 is a conserved selective marker of mammalian cells homologous to mouse CD8alpha+ dendritic cells. *J Exp Med* (2010) 207 (6):1283–92. doi: 10.1084/jem.20100223
12. Sancho D, Joffre OP, Keller AM, Rogers NC, Martinez D, Hernanz-Falcon P, et al. Identification of a dendritic cell receptor that couples sensing of necrosis to immunity. *Nature* (2009) 458(7240):899–903. doi: 10.1038/nature07750
13. Zhang JG, Czabotar PE, Policheni AN, Caminschi I, Wan SS, Kitsoulis S, et al. The dendritic cell receptor Clec9A binds damaged cells via exposed actin filaments. *Immunity* (2012) 36(4):646–57. doi: 10.1016/j.immuni.2012.03.009
14. Nizzoli G, Krietsch J, Weick A, Steinfelder S, Facciotti F, Gruarin P, et al. Human CD1c+ dendritic cells secrete high levels of IL-12 and potently prime cytotoxic T-cell responses. *Blood* (2013) 122(6):932–42. doi: 10.1182/blood-2013-04-495424
15. Bakdash G, Buschow SII, Gorris MA, Halilovic A, Hato SV, Skold AE, et al. Expansion of a BDCA1+CD14+ Myeloid Cell Population in Melanoma Patients May Attenuate the Efficacy of Dendritic Cell Vaccines. *Cancer Res* (2016) 76(15):4332–46. doi: 10.1158/0008-5472.CAN-15-1695
16. Bol KF, Schreibelt G, Rabold K, Wculek SK, Schwarze JK, Dzionek A, et al. The clinical application of cancer immunotherapy based on naturally circulating dendritic cells. *J Immunother Cancer* (2019) 7(1):109. doi: 10.1186/s40425-019-0580-6
17. Matsui T, Connolly JE, Michnevitz M, Chaussabel D, Yu CII, Glaser C, et al. CD2 distinguishes two subsets of human plasmacytoid dendritic cells with distinct phenotype and functions. *J Immunol* (2009) 182(11):6815–23. doi: 10.4049/jimmunol.0802008
18. Balan S, Arnold-Schrauf C, Abbas A, Couespel N, Savoret J, Imperatore F, et al. Large-Scale Human Dendritic Cell Differentiation Revealing Notch-Dependent Lineage Bifurcation and Heterogeneity. *Cell Rep* (2018) 24 (7):1902–15 e6. doi: 10.1016/j.celrep.2018.07.033
19. Moody DB, Ulrichs T, Muhlecker W, Young DC, Gurcha SS, Grant E, et al. CD1c-mediated T-cell recognition of isoprenoid glycolipids in Mycobacterium tuberculosis infection. *Nature* (2000) 404(6780):884–8. doi: 10.1038/35009119
20. Collin M, Bigley V. Monocyte, Macrophage, and Dendritic Cell Development: the Human Perspective. *Microbiol Spectr* (2016) 4(5):79–97. doi: 10.1128/microbiolspec.MCHD-0015-2015
21. Small TN, Knowles RW, Keever C, Kernan NA, Collins N, O'Reilly RJ, et al. M241 (CD1) expression on B lymphocytes. *J Immunol* (1987) 138 (9):2864–8.
22. Kasinrerk W, Baumruker T, Majdic O, Knapp W, Stockinger H. CD1 molecule expression on human monocytes induced by granulocyte-macrophage colony-stimulating factor. *J Immunol* (1993) 150(2):579–84.
23. Breton G, Lee J, Zhou YJ, Schreiber JJ, Keler T, Puhr S, et al. Circulating precursors of human CD1c+ and CD141+ dendritic cells. *J Exp Med* (2015) 212(3):401–13. doi: 10.1084/jem.20141441
24. Haniffa M, Shin A, Bigley V, McGovern N, Teo P, See P, et al. Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. *Immunity* (2012) 37(1):60–73. doi: 10.1016/j.immuni.2012.04.012
25. Cant CA, Ullrich A. Signal regulation by family conspiracy. *Cell Mol Life Sci* (2001) 58(1):117–24. doi: 10.1007/PL00000771
26. Granot T, Senda T, Carpenter DJ, Matsuoka N, Weiner J, Gordon CL, et al. Dendritic Cells Display Subset and Tissue-Specific Maturation Dynamics over Human Life. *Immunity* (2017) 46(3):504–15. doi: 10.1016/j.immuni.2017.02.019

27. Seiffert M, Cant C, Chen Z, Rappold I, Brugger W, Kanz L, et al. Human signal-regulatory protein is expressed on normal, but not on subsets of leukemic myeloid cells and mediates cellular adhesion involving its counterreceptor CD47. *Blood* (1999) 94(11):3633–43. doi: 10.1182/blood.V94.11.3633.423k01_3633_3643
28. Ezquerra A, Revilla C, Alvarez B, Perez C, Alonso F, Dominguez J. Porcine myelomonocytic markers and cell populations. *Dev Comp Immunol* (2009) 33 (3):284–98. doi: 10.1016/j.dci.2008.06.002
29. Lahoud MH, Proietto AII, Gartlan KH, Kitsoulis S, Curtis J, Wettenhall J, et al. Signal regulatory protein molecules are differentially expressed by CD8-dendritic cells. *J Immunol* (2006) 177(1):372–82. doi: 10.4049/jimmunol.177.1.372
30. Baba T, Nakamoto Y, Mukaida N. Crucial contribution of thymic Sirp alpha+ conventional dendritic cells to central tolerance against blood-borne antigens in a CCR2-dependent manner. *J Immunol* (2009) 183(5):3053–63. doi: 10.4049/jimmunol.0900438
31. Ardouin L, Luche H, Chelbi R, Carpentier S, Shawket A, Montanana Sanchis F, et al. Broad and Largely Concordant Molecular Changes Characterize Tolerogenic and Immunogenic Dendritic Cell Maturation in Thymus and Periphery. *Immunity* (2016) 45(2):305–18. doi: 10.1016/j.immuni.2016.07.019
32. Khandelwal P, Blanco-Mezquita T, Emami P, Lee HS, Reyes NJ, Mathew R, et al. Ocular mucosal CD11b+ and CD103+ mouse dendritic cells under normal conditions and in allergic immune responses. *PloS One* (2013) 8(5): e64193. doi: 10.1371/journal.pone.0064193
33. Gurka S, Hartung E, Becker M, Kroczek RA. Mouse Conventional Dendritic Cells Can be Universally Classified Based on the Mutually Exclusive Expression of XCR1 and SIRPalpha. *Front Immunol* (2015) 6:35. doi: 10.3389/fimmu.2015.00035
34. Heger L, Balk S, Luhr JJ, Heidkamp GF, Lehmann CHK, Hatscher L, et al. CLEC10A Is a Specific Marker for Human CD1c(+) Dendritic Cells and Enhances Their Toll-Like Receptor 7/8-Induced Cytokine Secretion. *Front Immunol* (2018) 9:744. doi: 10.3389/fimmu.2018.00744
35. Sato M, Kawakami K, Osawa T, Toyoshima S. Molecular cloning and expression of cDNA encoding a galactose/N-acetylgalactosamine-specific lectin on mouse tumoricidal macrophages. *J Biochem* (1992) 111(3):331–6. doi: 10.1093/oxfordjournals.jbchem.a123758
36. Suzuki N, Yamamoto K, Toyoshima S, Osawa T, Irimura T. Molecular cloning and expression of cDNA encoding human macrophage C-type lectin. Its unique carbohydrate binding specificity for Tn antigen. *J Immunol* (1996) 156 (1):128–35.
37. Higashi N, Fujioka K, Denda-Nagai K, Hashimoto S, Nagai S, Sato T, et al. The macrophage C-type lectin specific for galactose/N-acetylgalactosamine is an endocytic receptor expressed on monocyte-derived immature dendritic cells. *J Biol Chem* (2002) 277(23):20686–93. doi: 10.1074/jbc.M202104200
38. Wong KL, Tai JJ, Wong WC, Han H, Sem X, Yeap WH, et al. Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets. *Blood* (2011) 118(5):e16–31. doi: 10.1182/blood-2010-12-326355
39. van Vliet SJ, Saeland E, van Kooyk Y. Sweet preferences of MGL: carbohydrate specificity and function. *Trends Immunol* (2008) 29(2):83–90. doi: 10.1016/j.it.2007.10.010
40. Denda-Nagai K, Aida S, Saba K, Suzuki K, Moriyama S, Oo-Puthinan S, et al. Distribution and function of macrophage galactose-type C-type lectin 2 (MGL2/CD301b): efficient uptake and presentation of glycosylated antigens by dendritic cells. *J Biol Chem* (2010) 285(25):19193–204. doi: 10.1074/jbc.M110.113613
41. Shin H, Kumamoto Y, Gopinath S, Iwasaki A. CD301b+ dendritic cells stimulate tissue-resident memory CD8+ T cells to protect against genital HSV-2. *Nat Commun* (2016) 7:13346. doi: 10.1038/ncomms13346
42. Linehan JL, Dileepan T, Kashem SW, Kaplan DH, Cleary P, Jenkins MK. Generation of Th17 cells in response to intranasal infection requires TGF-beta1 from dendritic cells and IL-6 from CD301b+ dendritic cells. *Proc Natl Acad Sci U.S.A.* (2015) 112(41):12782–7. doi: 10.1073/pnas.1513532112
43. Shook B, Xiao E, Kumamoto Y, Iwasaki A, Horsley V. CD301b+ Macrophages Are Essential for Effective Skin Wound Healing. *J Invest Dermatol* (2016) 136 (9):1885–91. doi: 10.1016/j.jid.2016.05.107
44. Kumamoto Y, Camporez JPG, Jurczak MJ, Shanabrough M, Horvath T, Shulman GII, et al. CD301b(+) Mononuclear Phagocytes Maintain Positive Energy Balance through Secretion of Resistin-like Molecule Alpha. *Immunity* (2016) 45(3):583–96. doi: 10.1016/j.immuni.2016.08.002
45. Dupasquier M, Stoitzner P, Wan H, Cerqueira D, van Oudenaren A, Voerman JS, et al. The dermal microenvironment induces the expression of the alternative activation marker CD301/mMGL in mononuclear phagocytes, independent of IL-4/IL-13 signaling. *J Leukoc Biol* (2006) 80(4):838–49. doi: 10.1189/jlb.1005564
46. Xu WD, Firestein GS, Taetle R, Kaushansky K, Zvaifler NJ. Cytokines in chronic inflammatory arthritis. II. Granulocyte-macrophage colony-stimulating factor in rheumatoid synovial effusions. *J Clin Invest* (1989) 83 (3):876–82. doi: 10.1172/JCI113971
47. Croizat H. Circulating cytokines in sickle cell patients during steady state. *Br J Haematol* (1994) 87(3):592–7. doi: 10.1111/j.1365-2141.1994.tb08318.x
48. Singhal R, Chawla S, Rathore DK, Bhasym A, Annarapu GK, Sharma V, et al. Development of pro-inflammatory phenotype in monocytes after engulfing Hb-activated platelets in hemolytic disorders. *Clin Immunol* (2017) 175:133–42. doi: 10.1016/j.clim.2016.12.007
49. Sloma I, Zilber MT, Charron D, Girot R, Tamouza R, Gelin C. Upregulation and atypical expression of the CD1 molecules on monocytes in sickle cell disease. *Hum Immunol* (2004) 65(11):1370–6. doi: 10.1016/j.humimm.2004.09.009
50. Thomas R, Lipsky PE. Human peripheral blood dendritic cell subsets. Isolation and characterization of precursor and mature antigen-presenting cells. *J Immunol* (1994) 153(9):4016–28.
51. MacDonald KP, Munster DJ, Clark GJ, Dzionek A, Schmitz J, Hart DN. Characterization of human blood dendritic cell subsets. *Blood* (2002) 100 (13):4512–20. doi: 10.1182/blood-2001-11-0097
52. van Ee TJ, Van Acker HH, van Oorschot TG, Van Tendeloo VF, Smits EL, Bakdash G, et al. BDCA1+CD14+ Immunosuppressive Cells in Cancer, a Potential Target? *Vaccines (Basel)* (2018) 6(3). doi: 10.3390/vaccines6030065
53. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* (2017) 356(6335). doi: 10.1126/science.aah4573
54. Schroder M, Melum GR, Landsverk OJ, Bujko A, Yaqub S, Gran E, et al. CD1c-Expression by Monocytes - Implications for the Use of Commercial CD1c+ Dendritic Cell Isolation Kits. *PloS One* (2016) 11(6):e0157387. doi: 10.1371/journal.pone.0157387
55. Wood GS, Freudenthal PS. CD5 monoclonal antibodies react with human peripheral blood dendritic cells. *Am J Pathol* (1992) 141(4):789–95.
56. Strobl H, Scheinecker C, Riedl E, Csmarits B, Bello-Fernandez C, Pickl WF, et al. Identification of CD68+lin- peripheral blood cells with dendritic precursor characteristics. *J Immunol* (1998) 161(2):740–8.
57. Yin X, Yu H, Jin X, Li J, Guo H, Shi Q, et al. Human Blood CD1c+ Dendritic Cells Encompass CD5high and CD5low Subsets That Differ Significantly in Phenotype, Gene Expression, and Functions. *J Immunol* (2017) 198(4):1553–64. doi: 10.4049/jimmunol.1600193
58. Meyerson HJ, Osei E, Schweitzer K, Blidaru G, Edinger A, Schlegelmilch J, et al. CD1c(+) myeloid dendritic cells in myeloid neoplasia. *Cytometry B Clin Cytom* (2016) 90(4):337–48. doi: 10.1002/cyto.b.21332
59. Dutertre CA, Becht E, Irac SE, Khalilnezhad A, Narang V, Khalilnezhad S, et al. Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* (2019) 51(3):573–89 e8. doi: 10.1016/j.immuni.2019.08.008
60. Alcantara-Hernandez M, Leylek R, Wagar LE, Engleman EG, Keler T, Marinkovich MP, et al. High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization. *Immunity* (2017) 47(6):1037–50.e6. doi: 10.1016/j.immuni.2017.11.001
61. Bourdely P, Anselmi G, Vaivode K, Ramos RN, Missolo-Koussou Y, Hidalgo S, et al. Transcriptional and Functional Analysis of CD1c(+) Human Dendritic Cells Identifies a CD163(+) Subset Priming CD8(+)CD103(+) T Cells. *Immunity* (2020) 53(2):335–52. doi: 10.1016/j.immuni.2020.06.002
62. Erlich Z, Shlomovitz I, Edry-Botzer L, Cohen H, Frank D, Wang H, et al. Macrophages, rather than DCs, are responsible for inflammasome activity in the GM-CSF BMDC model. *Nat Immunol* (2019) 20(4):397–406. doi: 10.1038/s41590-019-0313-5
63. Cytlak U, Resteu A, Pagan S, Green K, Milne P, Maisuria S, et al. Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic

Cell Development in Humans. *Immunity* (2020) 53(2):353–70.e8. doi: 10.1016/j.immuni.2020.07.003
64. Proietto AII, Mittag D, Roberts AW, Sprigg N, Wu L. The equivalents of human blood and spleen dendritic cell subtypes can be generated in vitro from human CD34(+) stem cells in the presence of fms-like tyrosine kinase 3 ligand and thrombopoietin. *Cell Mol Immunol* (2012) 9(6):446–54. doi: 10.1038/cmi.2012.48
65. Helft J, Anjos-Afonso F, van der Veen AG, Chakravarty P, Bonnet D, Reis e Sousa C. Dendritic Cell Lineage Potential in Human Early Hematopoietic Progenitors. *Cell Rep* (2017) 20(3):529–37. doi: 10.1016/j.celrep.2017.06.075
66. Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, et al. IRF8 mutations and human dendritic-cell immunodeficiency. *N Engl J Med* (2011) 365(2):127–38. doi: 10.1056/NEJMoa1100066
67. Bigley V, Maisuria S, Cytlak U, Jardine L, Care MA, Green K, et al. Biallelic interferon regulatory factor 8 mutation: A complex immunodeficiency syndrome with dendritic cell deficiency, monocytopenia, and immune dysregulation. *J Allergy Clin Immunol* (2018) 141(6):2234–48. doi: 10.1016/j.jaci.2017.08.044
68. Kong XF, Martinez-Barricarte R, Kennedy J, Mele F, Lazarov T, Deenick EK, et al. Disruption of an antimycobacterial circuit between dendritic and helper T cells in human SPPL2a deficiency. *Nat Immunol* (2018) 19(9):973–85. doi: 10.1038/s41590-018-0178-z
69. Sontag S, Forster M, Qin J, Wanek P, Mitzka S, Schuler HM, et al. Modelling IRF8 Deficient Human Hematopoiesis and Dendritic Cell Development with Engineered iPS Cells. *Stem Cells* (2017) 35(4):898–908. doi: 10.1002/stem.2565
70. Borriello F, Iannone R, Di Somma S, Vastolo V, Petrosino G, Visconte F, et al. Lipopolysaccharide-Elicited TSLPR Expression Enriches a Functionally Discrete Subset of Human CD14(+) CD1c(+) Monocytes. *J Immunol* (2017) 198(9):3426–35. doi: 10.4049/jimmunol.1601497
71. Corren J, Ziegler SF. TSLP: from allergy to cancer. *Nat Immunol* (2019) 20 (12):1603–9. doi: 10.1038/s41590-019-0524-9
72. Watchmaker PB, Lahl K, Lee M, Baumjohann D, Morton J, Kim SJ, et al. Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. *Nat Immunol* (2014) 15(1):98–108. doi: 10.1038/ni.2768
73. Lindstedt M, Lundberg K, Borrebaeck CA. Gene family clustering identifies functionally associated subsets of human in vivo blood and tonsillar dendritic cells. *J Immunol* (2005) 175(8):4839–46. doi: 10.4049/jimmunol.175.8.4839
74. Segura E, Valladeau-Guilemond J, Donnadieu MH, Sastre-Garau X, Soumelis V, Amigorena S. Characterization of resident and migratory dendritic cells in human lymph nodes. *J Exp Med* (2012) 209(4):653–60. doi: 10.1084/jem.20111457
75. Segura E, Durand M, Amigorena S. Similar antigen cross-presentation capacity and phagocytic functions in all freshly isolated human lymphoid organ-resident dendritic cells. *J Exp Med* (2013) 210(5):1035–47. doi: 10.1084/jem.20121103
76. Melum GR, Farkas L, Scheel C, Van Dieren B, Gran E, Liu YJ, et al. A thymic stromal lymphopoietin-responsive dendritic cell subset mediates allergic responses in the upper airway mucosa. *J Allergy Clin Immunol* (2014) 134 (3):613–621 e7. doi: 10.1016/j.jaci.2014.05.010
77. Patel VII, Booth JL, Duggan ES, Cate S, White VL, Hutchings D, et al. Transcriptional Classification and Functional Characterization of Human Airway Macrophage and Dendritic Cell Subsets. *J Immunol* (2017) 198 (3):1183–201. doi: 10.4049/jimmunol.1600777
78. Jardine L, Wiscombe S, Reynolds G, McDonald D, Fuller A, Green K, et al. Lipopolysaccharide inhalation recruits monocytes and dendritic cell subsets to the alveolar airspace. *Nat Commun* (2019) 10(1):1999. doi: 10.1038/s41467-019-09913-4
79. Vremec D, Shortman K. What's in a Name? Some Early and Current Issues in Dendritic Cell Nomenclature. *Front Immunol* (2015) 6:267. doi: 10.3389/fimmu.2015.00267
80. Durai V, Murphy KM. Functions of Murine Dendritic Cells. *Immunity* (2016) 45(4):719–36. doi: 10.1016/j.immuni.2016.10.010
81. Ginhoux F, Guilliams M, Naik SH. Editorial: Dendritic Cell and Macrophage Nomenclature and Classification. *Front Immunol* (2016) 7:168. doi: 10.3389/fimmu.2016.00168
82. Schraml BU, Reis e Sousa C. Defining dendritic cells. *Curr Opin Immunol* (2015) 32:13–20. doi: 10.1016/j.coi.2014.11.001
83. Kasahara S, Clark EA. Dendritic cell-associated lectin 2 (DCAL2) defines a distinct CD8alpha- dendritic cell subset. *J Leukoc Biol* (2012) 91(3):437–48. doi: 10.1189/jlb.0711384
84. Lewis KL, Caton ML, Bogunovic M, Greter M, Grajkowska LT, Ng D, et al. Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. *Immunity* (2011) 35(5):780–91. doi: 10.1016/j.immuni.2011.08.013
85. Kirkling ME, Cytlak U, Lau CM, Lewis KL, Resteu A, Khodadadi-Jamayran A, et al. Notch Signaling Facilitates In Vitro Generation of Cross-Presenting Classical Dendritic Cells. *Cell Rep* (2018) 23(12):3658–3672 e6. doi: 10.1016/j.celrep.2018.05.068
86. McDaniell R, Warthen DM, Sanchez-Lara PA, Pai A, Krantz ID, Piccoli DA, et al. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet* (2006) 79(1):169–73. doi: 10.1086/505332
87. Schiavoni G, Mattei F, Sestili P, Borghi P, Venditti M, Morse HC, et al. ICSBP is essential for the development of mouse type I interferon-producing cells and for the generation and activation of CD8alpha(+) dendritic cells. *J Exp Med* (2002) 196(11):1415–25. doi: 10.1084/jem.20021263
88. Aliberti J, Schulz O, Pennington DJ, Tsujimura H, Reis e Sousa C, Ozato K, et al. Essential role for ICSBP in the in vivo development of murine CD8alpha + dendritic cells. *Blood* (2003) 101(1):305–10. doi: 10.1182/blood-2002-04-1088
89. Brown CC, Gudjonson H, Pritykin Y, Deep D, Lavallee VP, Mendoza A, et al. Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* (2019) 179(4):846–63.e24. doi: 10.1016/j.cell.2019.09.035
90. Schlitzer A, McGovern N, Teo P, Zelante T, Atarashi K, Low D, et al. IRF4 transcription factor-dependent CD11b+ dendritic cells in human and mouse control mucosal IL-17 cytokine responses. *Immunity* (2013) 38(5):970–83. doi: 10.1016/j.immuni.2013.04.011
91. Bosteels C, Neyt K, Vanheerswynghels M, van Helden MJ, Sichien D, Debeuf N, et al. Inflammatory Type 2 cDCs Acquire Features of cDC1s and Macrophages to Orchestrate Immunity to Respiratory Virus Infection. *Immunity* (2020) 52 (6):1039–56. doi: 10.1016/j.immuni.2020.04.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership