Disease-biased and shared characteristics of the immunoglobulin gene repertoires in marginal zone B cell lymphoproliferations

Xochelli, A., Bikos, V., Polychronidou, E., Galigalidou, C., Agathangelidis, A., Charlotte, F., Moschonas, P., Davis, Z., Colombo, M., Roumelioti, M., Sutton, L.-A., Groenen, P., van den Brand, M., Boudjoghra, M., Algara, P., Traverse-Glehen, A., Ferrer, A., Stalika, E., Karypidou, M., ... Stamatopoulos, K. (2018). Disease-biased and shared characteristics of the immunoglobulin gene repertoires in marginal zone B cell lymphoproliferations. *Journal of Pathology*. Advance online publication. https://doi.org/10.1002/path.5209

Published in:
Journal of Pathology

Document Version:
Peer reviewed version

Disease-biased and shared characteristics of the immunoglobulin gene repertoires in marginal zone B cell lymphoproliferations

Aliki Xochelli[1,2], Vasilis Bikos[3], Eleftheria Polychronidou[4,5], Chrysi Galigalidou[1], Andreas Agathangelidis[6], Frédéric Charlotte[7], Panagiotis Moschonas[4], Zadie Davis[8], Monica Colombo[9], Maria Roumelioti[10], Lesley-Ann Sutton[11], Patricia Groenen[12], Michiel van den Brand[12], Myriam Boudjoghra[13], Patricia Algara[14], Alexandra Traverse-Glehen[15], Ana Ferrer[16], Evangelia Stalika[1], Maria Karypidou[17], George Kanellis[18], Christina Kalpadakis[19], Manuella Mollejo[14], Gerasimos Pangalis[20], Panayiotis Vlamos[5], Rose-Marie Amini[2], Sarka Pospisilova[3], David Gonzalez[21], Maurilio Ponzoni[22], Achilles Anagnostopoulos[17], Véronique Giudicelli[23], Marie-Paule Lefranc[23], Blanca Espinet[16], Panagiotis Panagiotidis[10], Miguel Angel Piris[24], Ming-Qing Du[25], Richard Rosenquist[11], Theodora Papadaki[18], Chrysoula Belessi[26], Manlio Ferrarini[27], David Oscier[8], Dimitrios Tzovaras[4], Paolo Ghia[6], Frederic Davi[13], Anastasia Hadzidimitriou[1], Kostas Stamatopoulos[1,2]

1. Institute of Applied Biosciences, CERTH, Thessaloniki, Greece
2. Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
3. Central European Institute of Technology, Masaryk University, Brno, Czech Republic
4. Information Technologies Institute, CERTH, Thessaloniki, Greece
5. Department of Informatics, Ionian University, Corfu, Greece
6. Division of Experimental Oncology and Department of Onco-Hematology, IRCCS San Raffaele Scientific Institute and Università Vita-Salute San Raffaele, Milan, Italy

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/path.5209

7. Department of Pathology, Hopital Pitie-Salpetriere and Sorbonne University, Paris, France
8. Department of Haematology, Royal Bournemouth Hospital, Bournemouth, United Kingdom
9. Molecular Pathology, Ospedale Policlinico SanMartino, Genova, Italy
10. First Department of Propaedeutic Medicine, University of Athens, Athens, Greece
11. Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
12. Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands
13. Department of Hematology, Hopital Pitie-Salpetriere and Sorbonne University, Paris, France
14. Hospital Virgen de la Salud, Toledo, Spain
15. Department of Pathology and Hematology, Hospices Civils de Lyon, Universite Lyon 1, Lyon, France
16. Laboratori de Citologia Hematològica i Citogenètica Molecular, Servei de Patologia, Hospital del Mar, Barcelona, Spain
17. Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece
18. Hematopathology Department, Evangelismos Hospital, Athens, Greece
19. Department of Haematology, University of Crete, Heraklion, Greece
20. Department of Haematology, Athens Medical Center, Athens, Greece
21. Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, United Kingdom
22. Pathology Unit, San Raffaele Scientific Institute, Milan, Italy

23. IMGT®, the international ImMunoGeneTics information system®, Université de Montpellier, LIGM, Institut de Génétique Humaine IGH, UMR CNRS UM, Montpellier, France
24. Pathology Department, IIS ”Fundacion Jimenez Diaz”, Madrid, Spain
25. Division of Cellular and Molecular Pathology, Department of Pathology, University of Cambridge, Cambridge, UK
26. Hematology Department, Nikea General Hospital, Piraeus, Greece
27. Direzione Scientifica, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Azienda Ospedaliera Universitaria (AOU) San Martino-IST, Genoa, Italy

*Correspondence to: Anastasia Hadzidimitriou, Institute of Applied Biosciences, Center for Research and Technology Hellas, 57001 Thermi, Thessaloniki, Greece. Phone: +30 2310 498474. Fax: +30 2310 498270. E-mail: achatzidimitriou@certh.gr

Conflict of interest: The authors declare no conflict of interest

Running title: MZ lymphoproliferations: shared and unique characteristics

Word count: 1500

ABSTRACT

The B cell receptor immunoglobulin (BcR IG) gene repertoires of marginal zone (MZ) lymphoproliferations were analyzed in order to obtain insight into their ontogenetic relationships. Our cohort included cases with MZ lymphomas (n=488) i.e. splenic (SMZL), nodal (NMZL) and extranodal (ENMZL) as well as provisional entities (n=76) according to the World Health Organization classification. The most striking IG gene repertoire skewing was observed in SMZL. However, restrictions were also identified in all other MZ lymphomas studied, particularly ENMZL, with significantly different IG gene distributions depending on the primary site of involvement. Cross-entity comparisons of the MZ IG sequence dataset with a large dataset of IG sequences (MZ-related or not; n=65,837) revealed four major clusters of cases sharing homologous ('public') heavy variable complementarity-determining region 3. These clusters included rearrangements from SMZL, ENMZL (gastric, salivary gland, ocular adnexa), chronic lymphocytic leukemia but also rheumatoid factors and non-malignant spleen MZ cells. In conclusion, different MZ lymphomas display biased immunogenetic signatures indicating distinct antigen exposure histories. The existence of rare public stereotypes raises the intriguing possibility that common, pathogen-triggered, immune-mediated mechanisms, may result in diverse B lymphoproliferations due to targeting versatile progenitor B cells and/or operating in particular microenvironments.

Keywords (3-10): Marginal zone lymphoma, ontogeny, immunoglobulin gene, antigen

INTRODUCTION

The 2016 update of the World Health Organization (WHO) Classification of tumors of the hematopoietic and lymphoid tissues [1] recognizes three distinct types of marginal zone (MZ) lymphomas: splenic marginal zone lymphoma (SMZL), nodal MZ lymphoma (NMZL), and extranodal MZ lymphoma (ENMZL). Provisional entities pertain to splenic B cell lymphoma leukemia unclassifiable [splenic diffuse red pulp lymphoma (SDRL); hairy cell leukemia variant (HCL-v)] and pediatric nodal marginal zone lymphoma. In addition, non-CLL type monoclonal B lymphocytosis potentially related to SMZL (also termed clonal lymphocytosis of MZ origin, CBL-MZ)[2,3] has been recognized for the first time.

(Auto)antigenic stimulation and (micro)environmental interactions have been implicated in MZ lymphomagenesis. This claim is supported by the (i) pronounced skewing of the immunoglobulin (IG) gene repertoire of MZ lymphomas[4,5]; (ii) the strong association between chronic inflammation (due to infection or autoimmunity) and extranodal MZL lymphomas [6]; and, (iii) the increased incidence of genetic aberrations (recurrent translocations or gene mutations) affecting key components of B cell signaling pathways, further underscoring their significance in the natural history of these entities [7,8]. However, despite significant advances in characterizing the biological landscape of MZ lymphomas, questions abound regarding their immune pathogenesis and precise ontogenetic relationships.

Here, we aimed to obtain insight into MZ lymphomagenesis through: (i) systematic characterization of the IG gene repertoires in MZ lymphomas and (ii) comparison of IG sequences from different MZ lymphoma subtypes, newly recognized clonal entities (CBL-MZ) and non-malignant entities [persistent polyclonal B-cell lymphocytosis, PPBL] for which a marginal zone origin is speculated, against a large IG gene sequence dataset.

MATERIALS AND METHODS

Characterization of the studied IG gene sequence dataset

Included in the study were 66,401 IGHV-IGHD-IGHJ gene rearrangement sequences from (i) patients with MZ lymphoproliferations (n=1339) obtained from both the collaborating institutions and public databases (ii) CLL (n=20,451) cases deposited into the IMGT/CLL-DB; and, (iii) a dataset of 44,611 IGHV-IGHD-IGHJ gene rearrangement sequences from the IMGT/LIGM-DB.

The MZ dataset included: (i) Definite entities of MZ origin, namely SMZL, n=353 cases; NMZL, n=37 cases; ENMZL, n=98 cases; (ii) provisional entities of postulated MZ origin, including SDRL, n=16 cases; and, CBL-MZ, n=60 cases, (iii) PPBL, n=2 cases/286 sequences; and (iv) non malignant MZ cells isolated from 6 spleen specimens free of neoplastic cells at histological inspection (non-malignant MZ), obtained at surgery for non-hematologic cancer (n=489 sequences) (supplementary material, Supplementary materials and methods).

IG gene repertoire analysis was performed only for rearrangement sequences from definite clonal entities of MZ origin. Sequences from non-clonal MZ lymphoproliferations and/or non-malignant splenic MZ cells were used for cross-comparison purposes and the identification of stereotyped sequences; repertoire findings from these latter subgroups are summarized in (supplementary material, Supplementary materials and methods). The study was approved by local Ethics Review Committees.

PCR amplification and immunoinformatic analysis of IGHV-IGHD-IGHJ gene rearrangements

PCR amplification and sequence analysis of IGHV-IGHD-IGHJ gene rearrangements was performed as described [9] and stereotyped rearrangements were identified through the use of a validated bioinformatics algorithm [9].

In order to ensure consistency, all sequences underwent a multi-level evaluation aimed at (i) removing unproductive, incomplete and/or IGHV-IGHD-IGHJ gene rearrangement sequences with ambiguous characters and, (ii) eliminating redundancy. Rearrangements were classified as either unique (present only once in the investigated dataset), or recurrent, (belonging to the same clonal family) sharing the same IGHV/IGHD/IGHJ gene combinations and a VH CDR3 of identical length and identical or highly similar (≥60%) amino acid sequence. In the case of groups of recurrent sequences, only a single representative sequence was considered for repertoire characterization and comparisons [10].

Statistics

Comparisons and associations between different subgroups were performed with the use of descriptive statistics assessed using Chi-square or Fisher's exact test for independence. The significance level for all tests was set to $p<0.05$. Statistical analyses were performed with the statistical package, SPSS version 22.0 (SPSS, Chicago, IL, USA).

RESULTS AND DISCUSSION

Capitalizing on a large dataset of IGHV-IGHD-IGHJ gene rearrangement sequences available through our multi-institutional collaboration, we assessed the immunogenetic characteristics of MZ lymphomas and searched for stereotyped BcR IG sequences across MZ and MZ-related entities.

IG gene repertoire analysis revealed restrictions in the use of particular IGHV genes in all examined MZ lymphomas indicating disease-biased patterns (Figure 1A). IGHV gene repertoire skewing was more pronounced in SMZL, where the IGHV1-2 gene was used by 27.8% of all cases, in keeping with published studies by us and others [4]; no particular bias to any IGHV gene was noted amongst the HCV (+) SMZL cases of our cohort (supplementary material, Supplementary materials and methods). In NMZL, the most frequently expressed IGHV genes were IGHV4-34 and IGHV1-69 (13.5% each). Finally, in ENMZL the IGHV1-69 gene predominated (16.3%) albeit with significantly different distribution depending on the primary site of involvement (34.8% in salivary ENMZL, 19.2% in gastric ENMZL, 6.1% in ocular adnexa ENMZL, OAMZL, p=0.008 for all comparisons) (Figure 1B). Information regarding *Chlamydia psittaci* (*C. psittaci)* infection status was available for 38 OAMZL [11] of whom 16 (42%) were positive. Even though not statistically significant the IGHV3-23 gene predominated amongst *C. psittaci* (-) cases (6/22 versus 2/16 in positive cases; p=0.27), whereas IGHV3-7 predominated amongst *C. psittaci* (+) cases (3/16 versus 1/22; p=0.15). *H. pylori* status in gastric ENMZL cases was not available due to the retrospective nature of the study. Overall, these findings confirm and significantly extend previous observations about the existence of IG gene repertoire biases in MZ lymphomas [12–14].

Turning to somatic hypermutation (SHM) and following previously proposed definitions [4], IGHV-IGHD-IGHJ gene rearrangements carrying IGHV genes with no SHM (100% V-REGION

identity to that ot the closest germline gene, (germline identity (GI)) were assigned to a "truly unmutated" (TU) subgroup whereas sequences with 97% to 99.9% GI were classified as "borderline/minimally mutated"(BM) and those with <97% identity as "mutated" (M).

The vast majority of MZL cases bore some impact of SHM. However, differences were identified between various MZ lymphoma subtypes regarding their SHM profiles. In particular, the proportion of TU cases ranged from 0% in ENMZL gastric cases to 12.2% in SMZL. Regarding the remaining cases bearing SHM, 37.3% of IGHV-IGHD-IGHJ gene rearrangements in SMZL were BM versus only 18.9% in NMZL and 18.3% in ENMZL ($p=0.02$ and $p=0.0004$, respectively). Finally, only 50.4% of SMZL cases exhibited a significant SHM load versus 72.9% and 75.5% of NMZL and ENMZL cases ($p=0.008$ and $p<0.0001$, respectively) (Figure 2).

Following established bioinformatics approaches[9], we searched for stereotyped BcR IG sequences i.e. IGHV-IGHD-IGHJ gene rearrangements with restricted antigen-binding site sequence motifs. Of the 66,401 IGHV-IGHD-IGHJ rearrangements included in the analysis, 40,103 (60.4%) were assigned to at least one stereotyped subset. Overall, 26,143 different clusters were recognized. Of these 2499 (9.5%) included IGHV-IGHD-IGHJ gene rearrangements from a single entity and were considered as 'disease-specific', whereas the remainder included rearrangements from different entities and were, thus, deemed 'public'.

Most clusters (17,844/26,143; 68.3%) contained only two sequences; of the remaining 8299 clusters with 3 or more members, 1653 (6.3% of all clusters) included six or more members. The vast majority of disease-specific clusters (2468/2499, 98.7%) concerned CLL cases, which is not paradoxical, considering that the incidence of BcR IG stereotypy is much higher in CLL versus all other B cell lymphomas thus far analyzed [15]. A much smaller proportion of 'disease-specific' clusters (31/2499, 1.2%) concerned MZ lymphomas, mostly SMZL.

Public clusters were mostly relatively small in size with only 4 being populated by more than 20 members (supplementary material, Table S1). All 4 large public clusters included somatically hypermutated IGHV-IGHD-IGHJ gene rearrangements that derived from MZ lymphomas, non-malignant spleen MZ cells and other conditions. Two of 4 subsets utilized the IGHV1-69 gene, while the remaining 2 utilized the IGHV3-7 and IGHV4-59 gene, respectively (Figure 3). They included IGHV-IGHD-IGHJ gene rearrangements from various entities, including SMZL, ENMZL (gastric, salivary gland, ocular adnexa), CLL, hepatitis C virus-associated diffuse large B cell lymphoma (DLBCL), and non-malignant spleen MZ cells and rearrangements rheumatoid factors [with regards to the latter, recalling previous reports [11,12,16,17]. Notably, shared (recurrent) amino acid changes introduced by SHM (i.e. the same amino acid replacement at the same position) were identified in each large public cluster (supplementary material, Tables S2–5).

As the story of antigen-driven lymphoma unfolds, multidisciplinary investigation will be paramount to the identification of the selecting agent(s) and immune processes underlying the development of MZ lymphomas. Our present findings highlight biased immunogenetic signatures for distinct MZ lymphomas that pertain to both distinctive IGHV gene repertoires and differential SHM imprints. Although highlighting (auto)antigen selection pressure as a major driver in MZ lymphomagenesis, these results also suggest that unique immune pathways lead to distinct MZ lymphomas.

That said, the existence of rare public stereotypes raises the intriguing possibility that, occasionally, common patterns of immune stimulation triggered by distinct and/or cross-reactive (auto)antigens may result in diverse B lymphoproliferations due to targeting versatile progenitor B cells and/or operating in particular microenvironments. Future immunogenetic and functional studies are clearly warranted for testing this hypothesis but also for offering answers as to how normal MZ cells, the presumed normal counterparts for

several MZ lymphomas, are co-opted into pathogenic immune responses in these tumors; where does affinity maturation take place; and, what is the significance of various bystander cells, particularly T cells.

ACKNOWLEDGMENTS

This work was supported in part by H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe", by the European Union (EU); H2020 No. 692298 project "MEDGENET, Medical Genomics and Epigenomics Network" by the EU; grant AZV 15-30015A from the Ministry of Health of the Czech Republic, and the project CEITEC2020 LQ1601 from the Ministry of Education, Youth, and Sports of the Czech Republic; Bloodwise Research Grant (15019); the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Stockholm, the Lion's Cancer Research Foundation, Uppsala, the Marcus Borgström Foundation and Selander's Foundation, Uppsala.

Statement of author contributions

AX designed the study, performed research and wrote the paper; VB performed research and wrote the paper; EP, CG, AA and PM, performed research; FC, ZD, MC, MR, LAS, PG, MB, PA, AT-G, AF, ES, MK, GK, CK, MM, GP, PV, RM A, SP, DG, MP, AA, VG, MP L, BE, PP, MA P, MD, RR, TP, CB, MF, provided data; DO, DT, PG, FD, AH, supervised research; KS designed the study, supervised research and wrote the paper.

References

1 Swerdlow SH, Campo E, Pileri SA, *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* 2016; **127**: 2375–2390.

2 Xochelli A, Kalpadakis C, Gardiner A, *et al.* Clonal B-cell lymphocytosis exhibiting immunophenotypic features consistent with a marginal-zone origin: is this a distinct entity? *Blood* 2014; **123:** 1199–1206.

3 Xochelli A, Oscier D, Stamatopoulos K. Clonal B-cell lymphocytosis of marginal zone origin. *Best Pract Res Clin Haematol* 2017; **30:** 77–83.

4 Bikos V, Darzentas N, Hadzidimitriou A, *et al.* Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: Ontogenetic implications. *Leukemia* 2012; **26:** 1638–1646

5 Bikos V, Karypidou M, Stalika E, *et al.* An immunogenetic signature of ongoing antigen interactions in splenic marginal zone lymphoma expressing IGHV1-2*04 receptors. *Clin Cancer Res* 2016; **22:** 2032–2040.

6 Thieblemont C, Bertoni F, Copie-Bergman C, *et al.* Chronic inflammation and extra-nodal marginal-zone lymphomas of MALT-type. *Semin Cancer Biol* 2014; **24:** 33–42

7 Parry M, Rose-Zerilli MJJ, Ljungström V, *et al.* Genetics and prognostication in splenic marginal zone lymphoma: Revelations from deep sequencing. *Clin Cancer Res* 2015; **21:** 4174–4183

8 Agathangelidis A, Xochelli A, Stamatopoulos K. A gene is known by the company it keeps: enrichment of TNFAIP3 gene aberrations in MALT lymphomas expressing IGHV4-34 antigen receptors. *J Pathol* 2017; **243:** 403–406.

9 Agathangelidis A, Darzentas N, Hadzidimitriou A, *et al.* Stereotyped B-cell receptors in one third of chronic lymphocytic leukemia: towards a molecular classification with implications for targeted therapeutic interventions. *Blood* 2012; **119:** 4467–4476.

10 Colombo M, Cutrona G. Expression of immunoglobulin receptors with distinctive features indicating antigen selection by marginal zone B cells from human spleen. *Mol Med* 2013; **19:** 294–302.

11 Dagklis A, Ponzoni M, Govi S, *et al.* Immunoglobulin gene repertoire in ocular adnexal lymphomas: hints on the nature of the antigenic stimulation. *Leukemia* 2012; **26:** 814–821.

12 Bende RJ, Aarts WM, Riedl RG, *et al.* Among B cell non-Hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. *J Exp Med* 2005; **201:** 1229–1241.

13 Moody S, Escudero-Ibarz L, Wang M, *et al.* Significant association between *TNFAIP3* inactivation and biased immunoglobulin heavy chain variable region 4-34 usage in mucosa-associated lymphoid tissue lymphoma. *J Pathol* 2017; **243:** 3–8.

14 Gachard N, Parrens M, Soubeyran I, *et al.* IGHV gene features and MYD88 L265P mutation separate the three marginal zone lymphoma entities and Waldenström macroglobulinemia/lymphoplasmacytic lymphomas. *Leukemia* 2013; **27:** 183–189.

15 Stamatopoulos K, Agathangelidis A, Rosenquist R, *et al.* Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia* 2017; **31:** 282–291.

16 Bende RJ, Slot LM, Hoogeboom R, *et al.* Stereotypic rheumatoid factors that are frequently expressed in mucosa-associated lymphoid tissue-type lymphomas are rare in the labial salivary glands of patients with Sjögren's syndrome. *Arthritis Rheumatol* 2015; **67:** 1074–1083.

17 Kostareli E, Gounari M, Janus A, *et al.* Antigen receptor stereotypy across B-cell lymphoproliferations: the case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity. *Leukemia* 2012; **26:** 1127–1131.

18 Pommié C, Levadoux S, Sabatier R, *et al.* IMGT standardized criteria for statistical

analysis of immunoglobulin V-Region amino acid properties. *J Mol Recognit* 2004; **17:** 17–32.

FIGURE LEGENDS

Figure 1.

IGHV gene repertoire of marginal zone lymphomas. (A) Immunogenetic analysis of marginal zone lymphomas (MZLs) revealed IGHV gene repertoire skewing that was most pronounced in SMZL. (B) Relative frequency of the IGHV1-69 gene in different extranodal MZ lymphomas: significant differences depending on the primary site of involvement.

Notes: (i) an asterisk is used to indicate the most frequent genes in each entity; (ii) in the repertoire analysis presented in this figure, every gene is represented by a unique color in all pie charts.

Figure 2.

Differential imprint of somatic hypermutation in marginal zone lymphomas. Most marginal zone lymphoma cases bore at least some imprint of somatic hypermutation, however different profiles were identified in distinct subtypes.

Figure 3.

Public clonotypic rearrangements. VH CDR3 amino acid alignment of the four largest public clusters i.e. clusters that were not disease-specific (comprising IGHV-IGHD-IGHJ gene rearrangements from a single entity) but rather included cases deriving from different MZ lymphoproliferations but also other lymphoma entities as well as healthy and autoreactive clones; additional relevant information is provided in supplementary material, Table S1. (A) cluster #3; (B) cluster #5; (C) cluster #12; (D) cluster #16. Construction of the logos was

performed with the use of WebLogo (http://weblogo.berkeley.edu/). Each letter represents a single amino acid whereas the size of the letter represents its frequency. Colors correspond to the physicochemical properties color code used by IMGT [18]. The numbering of the clusters is not indicative of their characteristics, nor does it correspond to the nomenclature for stereotypes CLL subsets, but only concerns their succession in cluster identification.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods YES

Supplementary figure legends NO SUPPL FIGURES

Table S1. Only four public clusters were populated with more than 20 sequences. These included IGHV-IGHD-IGHJ gene rearrangements of different origin such as MZ lymphomas, non-malignant spleen tissue and various other conditions, especially HCV-associated entities and autoimmune disorders. Minor public clusters (populated with less than 10 sequences) were also identified.

Table S2. Recurrent amino acid changes introduced by SHM in cluster #3

Table S3. Recurrent amino acid changes introduced by SHM in cluster #5

Table S4. Recurrent amino acid changes introduced by SHM in cluster #12

Table S5. Recurrent amino acid changes introduced by SHM in cluster #16

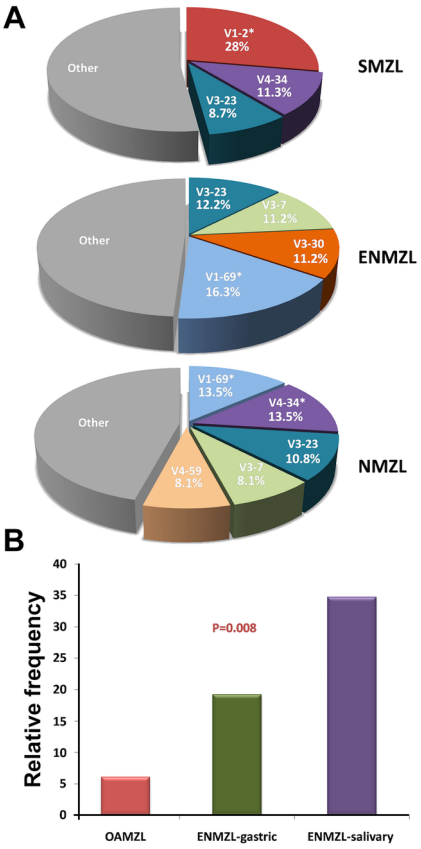

Figure_1 RP.tif

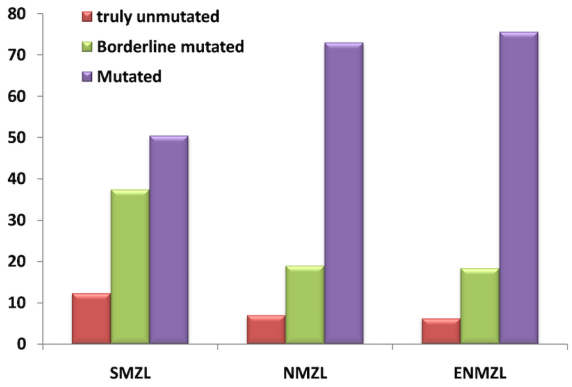

Figure_2 RP.tif

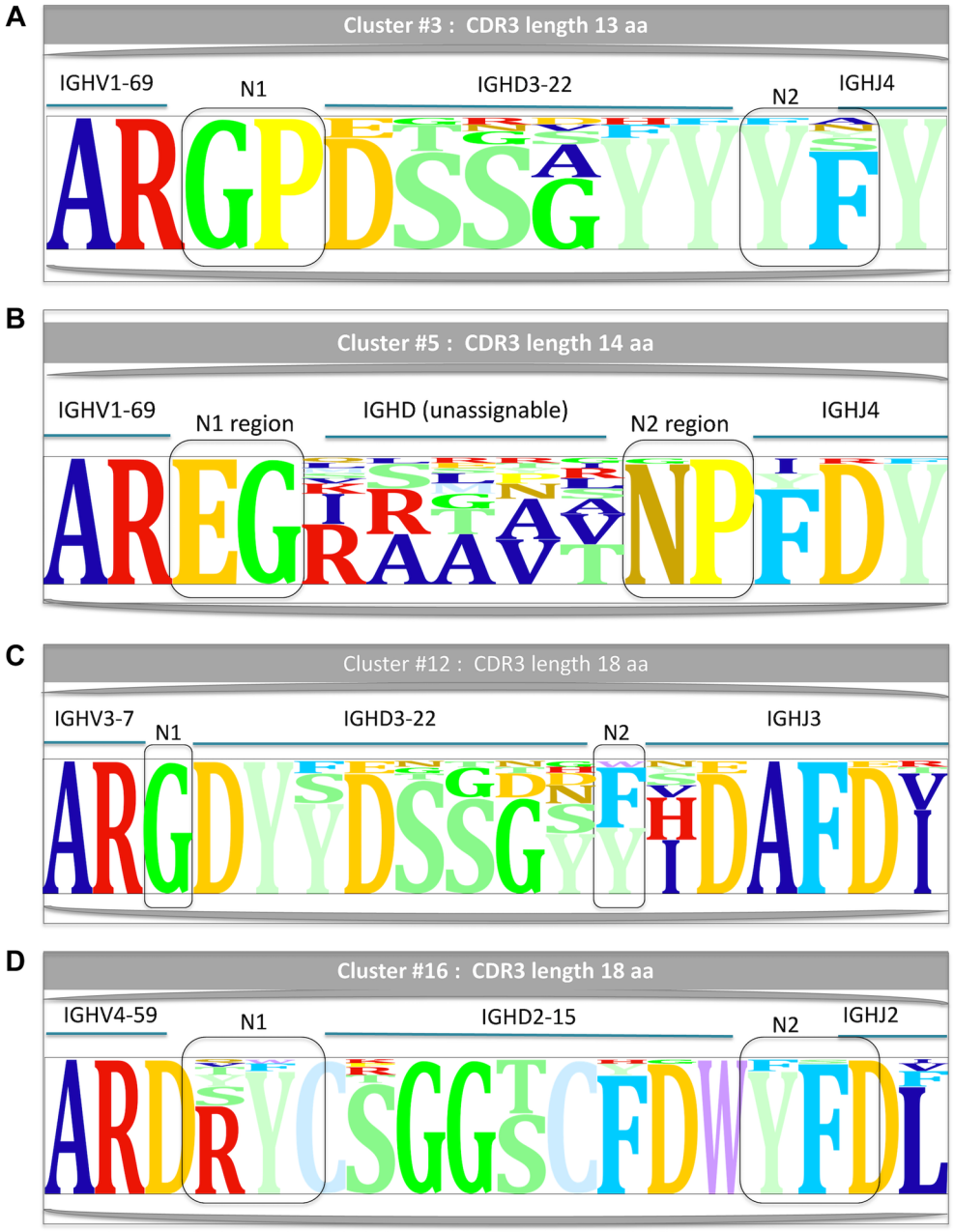

Figure_3 RP.tif