



American Society of Hematology  
 2021 L Street NW, Suite 900,  
 Washington, DC 20036  
 Phone: 202-776-0544 | Fax 202-776-0545  
 editorial@hematology.org

## Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL

Tracking no: BLD-2020-007039R1

Andreas Agathangelidis (Centre for Research and Technology Hellas, Greece) Anastasia Chatzidimitriou (Karolinska Institute, Sweden) Katerina Gemenetzi (Democritus University of Thrace, Greece) Veronique Giudicelli (IMGT, France) Maria Karypidou (Centre for Research and Technology Hellas, Greece) Karla Plevova (Masaryk University, Czech Republic) Zadie Davis (Royal Bournemouth Hospital, United Kingdom) Xiao-Jie Yan (Feinstein Institute for Medical Research, Northwell Health, United States) Sabine Jeromin (MLL GmbH, Germany) Christof Schneider (University Hospital Medical Center, Germany) Lone Pedersen (Rigshospitalet, Denmark) Renee Tschumper (Mayo Clinic College of Medicine, United States) Lesley Sutton (Karolinska Institutet, ) Panagiotis Baliakas (Science for Life Laboratory, Uppsala University, Sweden) Lydia Scarfò (Laboratory of B cell Neoplasia, Division of Experimental Oncology, Istituto Scientifico San Raffaele, Italy) Ellen van Gastel (Erasmus MC, University Medical Center, Netherlands) Marine Armand (Sorbonne Université, UMR\_S 1138, Centre de Recherche des Cordeliers, France) Eugen Tausch (Ulm University, Germany) Bella Biderman (National Research Center for Hematology, Russian Federation) Constance Baer (Munich Leukemia Laboratory, Germany) Davide Bagnara (University of Genova, Italy) Alba Navarro (Institut d'Investigacions Biomediques August Pi I Sunyer, Spain) Anne de Septenville (Sorbonne Université, UMR\_S 1138, Centre de Recherche des Cordeliers, France) Valentina Guido (Niguarda Cancer Center, Niguarda Hospital, Italy) Gerlinde Mitterbauer-Hohendanner (Medical University of Vienna, Austria) Aleksandar Dimovski (University Ss Cyril and Methodius, Macedonia, Republic of) Christian Brieghel (Copenhagen University, Rigshospitalet, Denmark) Sarah Lawless (Belfast City Hospital, Belfast Health and Social Care Trust, United Kingdom) Manja Meggendorfer (MLL, Germany) Kamila Stranska (Masaryk University, Czech Republic) Matthias Ritgen (University Clinic of Schleswig-Holstein, Kiel, Germany) Monica Facco (Veneto Institute of Molecular Medicine, Italy) Cristina Tresoldi (ospedale San Raffaele, Italy) Andrea Visentin (Veneto Institute of Molecular Medicine, Italy) Andrea Patriarca (Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Italy) Mark Catherwood (Belfast City Hospital, United Kingdom) Lisa Bonello (CeRMS, Italy) Andrey Sudarikov (National Research Center for Hematology, Russian Federation) Katrina Vanura (Medical University of Vienna, Austria) Maria Roumelioti (1st Department of Propedeutics, University of Athens, Laikon General Hospital, Athens, Greece, Greece) Hana Skuhrova Francova (University Hospital Brno, Masaryk University, ) Theodoros Moysiadis (Centre for reasearch and Technology Hellas, ) Silvio Veronese (Azienda Ospedaliera Ospedale Niguarda Ca' Granda, Italy) Krzysztof Giannopoulos (Medical University of Lublin, Poland) Larry Mansouri (Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, Sweden) Teodora Karan-Djurasevic (Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia) Raphael Sandaltzopoulos (Department of Molecular Biology and Genetics, Democritus University of Thrace, Greece) Csaba Bödör (Institute of Cancer, United Kingdom) Franco Fais (U.O. Molecular Pathology, IRCCS Ospedale Policlinico San Martino, Italy) Arnon Kater (Dept Hematology, Cancer Center Amsterdam, Amsterdam, University of Amsterdam, Netherlands) Irina Panovska-Stavridis (University Clinic of Hematology-Skopje, Macedonia (The Former Yugoslav Republic of)) Davide Rossi (Oncology Institute of Southern Switzerland, Switzerland) Salem Alshemmari (Faculty of Medicine, Kuwait University, Kuwait) Panagiotis Panagiotidis (University of Athens Medical School, Greece) Paul Costeas (The Center for the Study of Haematological Malignancies, Cyprus) Blanca Espinet (Hospital del Mar, Spain) Darko Antic (Clinic for hematology, Clinical Center Serbia, Serbia) Letizia Foroni (Imperial College Hammersmith Campus, United Kingdom) Marco Montillo (Niguarda Cancer Center, Niguarda Hospital, Italy) Livio Trentin (Veneto Institute of Molecular Medicine, Italy) Niki Stavroyianni (G. Papanicolaou Hospital, Greece) Gianluca Gaidano (University of Eastern Piedmont, Italy) Paola Francia di Celle (AO Città della Salute e della Scienza, Italy) Carsten Niemann (Rigshospitalet, Denmark) Elias Campo (Institut d'Investigacions Biomediques August Pi I Sunyer, Spain) Achilles Anagnostopoulos (George Papanicolaou Hospital, Greece) Christiane Pott (University Hospital Schleswig-Holstein, Germany) Kirsten Fischer (University Hospital Cologne, Germany) Michael Hallek (University of Cologne, Germany) David Oscier (Royal Bournemouth Hospital, United Kingdom) Stephan Stilgenbauer (University of Ulm, Germany) Claudia Haferlach (MLL Munich Leukemia Laboratory, Germany) Diane Jelinek (Mayo Clinic, United States) Nicholas Chiorazzi (The Feinstein Institute for Medical Research, United States) Sarka Pospisilova (Central European Institute of Technology (CEITEC), Masaryk University, Czech Republic) Marie-Paule Lefranc (IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Moléculaire (IGH), CNRS, Université de Montpellier (UM), France) Sofia Kossida (IMGT, IGH, CNRS, University of Montpellier, France) Anton Langerak (Erasmus MC, Netherlands) Chrysoula Belessi (General Hospital of Nikea, Greece) Frederic DAVI (Department of Hematology, AP-HP La Pitie-Salpetriere, Paris, FR, France) Richard Rosenquist (Clinical Genetics, Karolinska University Laboratory, Karolinska University Hospital, Sweden) Paolo Ghia (Università Vita-Salute San Raffaele C.F 97187560152, Italy) Kostas Stamatopoulos (Institute of Applied Biosciences, Centre for Research and Technology

Hellas, Greece)

**Abstract:**

Chronic lymphocytic leukemia (CLL) is characterized by the existence of subsets of patients with (quasi)identical, stereotyped B cell receptor immunoglobulins (BcR IG). Patients in certain major stereotyped subsets often display remarkably consistent clinicobiological profiles, suggesting that the study of BcR IG stereotypy in CLL has important implications for understanding disease pathophysiology and refining clinical decision-making. Nevertheless, several issues remain open, especially pertaining to the actual frequency of BcR IG stereotypy and major subsets, as well as the existence of higher-order connections between individual subsets. In order to address these issues, we investigated clonotypic IGHV-IGHD-IGHJ gene rearrangements in a series of 29,856 patients with CLL, by far the largest series worldwide. We report that the stereotyped fraction of CLL peaks at 41% of the entire cohort and that all 19 previously identified major subsets retained their relative size and ranking, while 10 new ones emerged; overall, major stereotyped subsets had a cumulative frequency of 13.5%. Higher-level relationships were evident between subsets, particularly for major stereotyped subsets with unmutated IGHV genes (U-CLL), for which close relations with other subsets, termed 'satellites', were identified. Satellite subsets accounted for 3% of the entire cohort. These results confirm our previous notion that major subsets can be robustly identified and are consistent in relative size, hence representing distinct disease variants amenable to compartmentalized research with the potential of overcoming the pronounced heterogeneity of CLL. Furthermore, the existence of satellite subsets reveals a novel aspect of repertoire restriction with implications for refined molecular classification of CLL.

**Conflict of interest:** COI declared - see note

**COI notes:** Si.Ve. has received honoraria from Bayer, Astra-Zeneca and Janssen. Ar.Ka. has received research funding from Abbvie, Roche/Genentech, Janssen and Astra-Zeneca. Da.Ro. has received honoraria from Abbvie, Astra-Zeneca, Gilead, Janssen, Verastem, and research grants from Abbvie, Gilead, Janssen, Cellestia. Sa.Al. has received educational grants from J&J, AbbVie, and Roche. Kr.Gi. has received Honoraria from AbbVie, Janssen and Roche. Li.Te. has received honoraria from Abbvie, Roche, Janssen, Shire. An.Vi. has received honoraria from Janssen, Abbvie. Gi.Ga. has received honoraria from Abbvie, Janssen, Sunesis and Astra-Zeneca for advisory boards or speaker's bureau services. Ca.Ni. has received research support, consultancy fees and/or travel grants from Abbvie, Gilead, Janssen, Roche, CSL Behring, Genmab, Sunesis and Acerta/Astra-Zeneca outside this work. Ki.Fi. has received honoraria from Roche and Abbvie, and Roche Travel grants. St.St. has received honoraria and research support from AbbVie, Astra-Zeneca, Celgene, Gilead, GSK, Hoffmann La-Roche, Janssen, Novartis. Ri.Ro. has received honoraria from AbbVie, Illumina, Janssen and Roche. Pa.Gh. has received honoraria from AbbVie, Acerta, BeiGene, Gilead, Janssen, Sunesis and research funding from AbbVie, Gilead, Janssen, Novartis, Sunesis. Ko.St. has received honoraria and research support from Abbvie, Janssen, Astra-Zeneca and Gilead.

**Preprint server:** No;

**Author contributions and disclosures:** An.Ag. designed the research, performed the experiments, analyzed the data and wrote the manuscript; Ka.Ge., Ma.Ka., Pa.Ba. and Th.Mo. analyzed data; Ve.Gi., Ma.Le. and So.Ko. were responsible for the curation of the IGHV-IGHD-IGHJ sequences; Ka.Pl., Za.Da., Xi.Ya., Sa.Je., Ch.Sc., Lo.Br-Pe., Re.Ts., Le.Su., Ly.Sc., El.Ga., Ma.Ar., Eu.Ta., Be.Bi., Co.Ba., Da.Ba., Al.Na., An.La., Va.Gu., Ge.Mi., Al.Di., Ch.Br., Sa.La., Ma.Me., Ka.Br., Ma.Ri., Mo.Fa., Cr.Tr., An.Vi., An.Pa., Ma.Ca., Li.Bo., An.Su., Ka.Va., Ma.Ro., Ha.Sk., Si.Ve., Kr.Gi., La.Ma., Te.Ka., Ra.Sa., Cs.Bö., Fr.Fa., Ar.Ka., Ir.Pa., Da.Ro., Sa.Al., Pa.Pa., Pa.Co., Bl.Es., Da.An., Le.Fo., Ma.Mo., Li.Tr., Ni.St., Gi.Ga., Pa.Ce., Ca.Ni., El.Ca., Ac.An., Ch.Po., Ki.Fi., Mi.Ha., Da.Os., St.St., Cl.Ha., Di.Je., Ni.Ch. and Sa.Po., provided data; An.Ch., An.La., Ch.Be., Fr.Da. Ri.Ro. and Pa.Gh. supervised research; Ko.St. designed the study, supervised research, and wrote the manuscript.

**Non-author contributions and disclosures:** No;

**Agreement to Share Publication-Related Data and Data Sharing Statement:** emails to the corresponding author

**Clinical trial registration information (if any):**

## HIGHER-ORDER CONNECTIONS BETWEEN STEREOTYPED SUBSETS: IMPLICATIONS FOR IMPROVED PATIENT CLASSIFICATION IN CLL

Short title for the running head: BcR stereotypes in CLL

Andreas Agathangelidis<sup>1</sup>, Anastasia Chatzidimitriou<sup>1,2</sup>, Katerina Gemenetzi<sup>1,3</sup>, Veronique Giudicelli<sup>4</sup>, Maria Karypidou<sup>1</sup>, Karla Plevova<sup>5,6</sup>, Zadie Davis<sup>7</sup>, Xiao-Jie Yan<sup>8</sup>, Sabine Jeromin<sup>9</sup>, Christof Schneider<sup>10</sup>, Lone Bredo-Pedersen<sup>11</sup>, Renee C. Tschumper<sup>12</sup>, Lesley-Ann Sutton<sup>2</sup>, Panagiotis Baliakas<sup>13</sup>, Lydia Scarfò<sup>14</sup>, Ellen J. van Gastel<sup>15</sup>, Marine Armand<sup>16</sup>, Eugen Tausch<sup>17</sup>, Bella Biderman<sup>18</sup>, Constance Baer<sup>9</sup>, Davide Bagnara<sup>19</sup>, Alba Navarro<sup>20,21</sup>, Anne Langlois de Septenville<sup>16</sup>, Valentina Guido<sup>22</sup>, Gerlinde Mitterbauer-Hohendanner<sup>23</sup>, Aleksandar Dimovski<sup>24</sup>, Christian Brieghel<sup>11</sup>, Sarah Lawless<sup>25</sup>, Manja Meggendorfer<sup>9</sup>, Kamila Braazdilova<sup>5,6</sup>, Matthias Ritgen<sup>26</sup>, Monica Facco<sup>27,28</sup>, Cristina Tresoldi<sup>29</sup>, Andrea Visentin<sup>27,28</sup>, Andrea Patriarca<sup>29</sup>, Mark Catherwood<sup>25</sup>, Lisa Bonello<sup>31</sup>, Andrey Sudarikov<sup>18</sup>, Katrina Vanura<sup>23</sup>, Maria Roumelioti<sup>32</sup>, Hana Skuhrova-Francova<sup>5</sup>, Theodoros Moysiadis<sup>1</sup>, Silvio Veronese<sup>22</sup>, Krzysztof Giannopoulos<sup>33</sup>, Larry Mansouri<sup>2</sup>, Teodora Karan-Djurasevic<sup>34</sup>, Raphael Sandaltzopoulos<sup>3</sup>, Csaba Bödör<sup>35</sup>, Franco Fais<sup>19,36</sup>, Arnon Kater<sup>37</sup>, Irina Panovska<sup>38</sup>, Davide Rossi<sup>39</sup>, Salem Alshemmari<sup>40</sup>, Panagiotis Panagiotidis<sup>32</sup>, Paul Costeas<sup>41,42</sup>, Blanca Espinet Solà<sup>43</sup>, Darko Antic<sup>44</sup>, Letizia Foroni<sup>45</sup>, Marco Montillo<sup>22</sup>, Livio Trentin<sup>27,28</sup>, Niki Stavroyianni<sup>46</sup>, Gianluca Gaidano<sup>30</sup>, Paola Francia di Celle<sup>31</sup>, Carsten Niemann<sup>11</sup>, Elias Campo<sup>20,21,47</sup>, Achilles Anagnostopoulos<sup>45</sup>, Christiane Pott<sup>26</sup>, Kirsten Fischer<sup>48</sup>, Michael Hallek<sup>49</sup>, David Oscier<sup>7</sup>, Stephan Stilgenbauer<sup>17</sup>, Claudia Haferlach<sup>9</sup>, Diane Jelinek<sup>50</sup>, Nicholas Chiorazzi<sup>8</sup>, Sarka Pospisilova<sup>5,6</sup>, Marie-Paule Lefranc<sup>4</sup>, Sofia Kossida<sup>4</sup>, Anton W. Langerak<sup>15</sup>, Chrysoula Belessi<sup>51</sup>, Frederic Davi<sup>15</sup>, Richard Rosenquist<sup>2,52</sup>, Paolo Ghia<sup>14</sup>, Kostas Stamatopoulos<sup>1,2</sup>, on behalf of ERIC, The European Research Initiative on CLL

<sup>1</sup>Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

<sup>2</sup>Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

<sup>3</sup>Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

<sup>4</sup>IMGT®, The International ImMunoGeneTics Information System®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), UMR 9002 CNRS-UM, Université de Montpellier, Montpellier Cedex 5, France

<sup>5</sup>Department of Internal Medicine - Hematology and Oncology, Faculty of Medicine, Masaryk University and University Hospital Brno, Brno, Czech Republic

<sup>6</sup>Center of Molecular Medicine, Central European Institute of Technology, Masaryk University, Brno, Czech Republic

<sup>7</sup>Department of Haematology, Royal Bournemouth Hospital, Bournemouth, UK

<sup>8</sup>The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY, USA

<sup>9</sup>MLL Munich Leukemia Laboratory, Munich, Germany

<sup>10</sup>University Hospital Medical Center, Ulm, Germany

<sup>11</sup>Department of Hematology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

<sup>12</sup>Department of Immunology, Mayo Clinic, Rochester, MN, USA

<sup>13</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

<sup>14</sup>Strategic Research Program on CLL and B-cell neoplasia Unit, Division of Experimental Oncology, Università Vita-Salute San Raffaele and IRCCS Ospedale San Raffaele, Milan, Italy

<sup>15</sup>Department of Immunology, Laboratory Medical Immunology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

<sup>16</sup>AP-HP, Hôpital Pitié-Salpêtrière, Department of Biological Hematology, Sorbonne Université, UMR\_S 1138, Centre de Recherche des Cordeliers, Paris, France

<sup>17</sup>Department of Internal Medicine III, Ulm University, Ulm, Germany

<sup>18</sup>National Research Center for Hematology, Novy Zykovski lane 4a, 125167, Moscow, Russia

<sup>19</sup>Department of Experimental Medicine, University of Genoa, Genoa, Italy

<sup>20</sup>Centro de Investigacion Biomedica en Red en Oncologia (CIBERONC), Madrid, Spain

<sup>21</sup>Institut d'Investigacions Biomediques August Pi I Sunyer, Barcelona, Spain

<sup>22</sup>Molecular Pathology Unit and Haematology Department, Niguarda Cancer Center, Niguarda Hospital, Milan, Italy

<sup>23</sup>Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

<sup>24</sup>Faculty of Pharmacy, Ss Cyril and Methodius University of Skopje, Skopje, Republic of Northern Macedonia

<sup>25</sup>Clinical Haematology, Belfast City Hospital, Belfast Health and Social Care Trust, Belfast, Northern Ireland, UK

<sup>26</sup>Second Medical Department, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

<sup>27</sup>Department of Medicine-DIMED, Hematology and Clinical Immunology Unit, University of Padua, 35128 Padua, Italy

<sup>28</sup>Veneto Institute of Molecular Medicine, 35128 Padua, Italy

<sup>29</sup>Division of Immunology, Transplantation and Infectious, IRCCS San Raffaele Scientific Institute, Milan, Italy

<sup>30</sup>Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont and Ospedale Maggiore della Carità, Novara, Italy

<sup>31</sup>General Anatomopathology and Molecular Oncogenetics - AOU City of Health and Science of Turin, I-10126 Turin, Italy

<sup>32</sup>First Department of Propaedeutic Medicine, University of Athens, Athens, Greece

<sup>33</sup>Experimental Hematooncology Department, Medical University of Lublin, Lublin, Poland

<sup>34</sup>Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia

<sup>35</sup>MTA-SE Momentum Molecular Oncohematology Research Group, 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

<sup>36</sup>U.O. Molecular Pathology, IRCCS Ospedale Policlinico San Martino, Genoa, Italy

<sup>37</sup>Amsterdam UMC, University of Amsterdam, Department of Hematology, Amsterdam Infection and Immunity Institute, Cancer Center Amsterdam, Amsterdam, the Netherlands

<sup>38</sup>Department of Hematology, Faculty of Medicine, Ss Cyril and Methodius University of Skopje, Skopje, Republic of Northern Macedonia

<sup>39</sup>Division of Hematology, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland

<sup>40</sup>Department of Medicine, Faculty of Medicine, Kuwait University, Safat, Kuwait

<sup>41</sup>The Center for the Study of Haematological Malignancies, Nicosia, Cyprus

<sup>42</sup>Karaiskakio Foundation, Nicosia, Cyprus

<sup>43</sup>Laboratorio de Citogenètica Molecular, Laboratori de Citologia Hematològica, Servei de Patologia i Servei de Hematologia, Hospital del Mar, Barcelona, Spain

<sup>44</sup>Clinic for Hematology, Clinical Center of Serbia, Belgrade, Serbia

<sup>45</sup>Hammersmith Hospital, London, UK

<sup>46</sup>Hematology Department and HCT Unit, G. Papanicolaou Hospital, Thessaloniki, Greece

<sup>47</sup>Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain

<sup>48</sup>University Hospital Cologne, Cologne, Germany

<sup>49</sup>University of Cologne, Cologne, Germany

<sup>50</sup>Department of Immunology, Mayo Clinic, Scottsdale, AZ, USA

<sup>51</sup>Hematology Department, Nikea General Hospital, Pireaus, Greece

<sup>52</sup>Clinical Genetics, Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden

**Corresponding author:**

Kostas Stamatopoulos

Institute of Applied Biosciences

Centre for Research and Technology Hellas

6<sup>th</sup> km Charilaou – Thermis

57001 Thermi, Thessaloniki, Greece

Phone: +302310498271

e-mail: kostas.stamatopoulos@certh.gr

Abstract word count: 249

Text word count: 3994

Number of figures/tables: 7

Number of references: 80

## **Key points**

- In a series of 29,856 CLL patients, the incidence of B cell receptor stereotypy peaked at 41%.
- Higher-order relations exist between stereotyped subsets, particularly for those from U-CLL, for which ‘satellite’ subsets were identified.

## **ABSTRACT**

Chronic lymphocytic leukemia (CLL) is characterized by the existence of subsets of patients with (quasi)identical, stereotyped B cell receptor immunoglobulins (BcR IG). Patients in certain major stereotyped subsets often display remarkably consistent clinicobiological profiles, suggesting that the study of BcR IG stereotypy in CLL has important implications for understanding disease pathophysiology and refining clinical decision-making. Nevertheless, several issues remain open, especially pertaining to the actual frequency of BcR IG stereotypy and major subsets, as well as the existence of higher-order connections between individual subsets. In order to address these issues, we investigated clonotypic IGHV-IGHD-IGHJ gene rearrangements in a series of 29,856 patients with CLL, by far the largest series worldwide. We report that the stereotyped fraction of CLL peaks at 41% of the entire cohort and that all 19 previously identified major subsets retained their relative size and ranking, while 10 new ones emerged; overall, major stereotyped subsets had a cumulative frequency of 13.5%. Higher-level relationships were evident between subsets, particularly for major stereotyped subsets with unmutated IGHV genes (U-CLL), for which close relations with other subsets, termed ‘satellites’, were identified. Satellite subsets accounted for 3% of the entire cohort. These results confirm our previous notion that major subsets can be robustly identified and are consistent in relative size, hence representing distinct disease variants amenable to compartmentalized research with the potential of overcoming the pronounced heterogeneity of CLL. Furthermore, the existence of satellite subsets reveals a novel aspect of repertoire restriction with implications for refined molecular classification of CLL.



## INTRODUCTION

Antigen interactions mediated by the B cell receptor immunoglobulin (BcR IG) are critical for the survival and proliferation of chronic lymphocytic leukemia (CLL) malignant cells<sup>1,2</sup>.

From a clinical perspective, this notion was corroborated by the therapeutic efficacy of pharmacological agents inhibiting kinases downstream the BcR signaling pathway, even in relapsed/refractory CLL patients<sup>3-5</sup>. From a biological standpoint, support for the role of antigenic interactions in CLL was provided by pivotal studies from the 1990s reporting significant skewing in immunoglobulin (IG) gene usage, that was different from that of normal B cells<sup>6</sup>. Soon thereafter, the somatic hypermutation (SHM) status of the clonotypic rearranged immunoglobulin heavy variable (IGHV) genes was shown to strongly correlate with patient outcome<sup>7-9</sup>, a finding of seminal importance for both understanding and managing CLL<sup>10</sup>.

The relevance of the BcR IG in CLL was further reinforced through the identification of subsets of patients expressing stereotyped BcR IG<sup>11-15</sup>. In our previous study of 7,424 patients, we reported that ~30% of all CLL were assigned to stereotyped subsets, each defined by a distinctive amino acid pattern within the variable heavy complementarity determining-region 3 (VH CDR3) of the clonotypic BcR IG<sup>15</sup>. Moreover, we identified 19 sizeable subsets that were defined as major, collectively accounting for 12.4% of all CLL<sup>15</sup>.

Accumulating evidence indicates that patients assigned to the same stereotyped subset exhibit consistent biological background as reflected in similar profiles of antigen reactivity<sup>16-20</sup>, genomic aberrations<sup>21-23</sup>, gene expression<sup>24,25</sup>, epigenetic modifications<sup>26-28</sup>, Toll-like receptor signaling<sup>29,30</sup>, "classic"<sup>31</sup> and cell-autonomous<sup>32,33</sup> BcR signaling, amongst others. Moreover, BcR IG stereotypy defines subgroups with consistent clinical presentation and outcome, particularly exemplified by: (i) subset #2 that is associated with aggressive clinical course regardless of the SHM status and appears unresponsive to chemoimmunotherapy<sup>34,35</sup>, and (ii) subset #8 that displays the highest risk for Richter's transformation amongst all CLL. Hence, information regarding subset

membership can contribute to refined risk stratification of CLL patients<sup>36-42</sup>, at least for certain subsets.

Altogether, these findings support the notion that the study of BcR IG stereotypy in CLL has important implications for both understanding disease pathophysiology and refining clinical decision-making<sup>34,35,38,41,42</sup>. That said, several issues remain open, especially pertaining to: (i) the exact frequency of BcR IG stereotypy and of the major subsets; (ii) its relation to IGHV gene usage and SHM status; and (iii) the existence of high-order relationships between subsets.

To this end, we undertook an in-depth study of BcR IG stereotypy in a cohort of 29,856 patients with CLL, 4 times larger compared to our previous study<sup>15</sup>. We report a significant increase in the relative size of the stereotyped fraction, which now exceeds 41% of all CLL. Stereotyped subsets previously designated as major retained their relative size, while 10 novel major subsets emerged. Finally, high-order relationships were detected between subsets, with certain major U-CLL subsets displaying several closely related 'satellite' subsets, often of considerable size. This finding argues for a similar pathophysiology and clonal behavior of these subsets, with implications for refined risk stratification and improved clinical decision-making.



## METHODS

### ***Patient group***

Included in the study were 29,856 patients with CLL from 45 collaborating Institutes in Europe, 3 in the USA and 1 in Asia. A subgroup of 7,424 patients included in the present cohort have been reported in our previous study<sup>15</sup>. Furthermore, cases from the cohort contributed by the German CLL Study Group (GCLLSG) have been investigated for membership to stereotyped subsets #1, #2, #4 and #8 in the context of clinicobiological association studies<sup>35</sup>. Detailed information regarding the geography and the contribution of each Institute is given in **Supplementary Table 1**. All cases met the diagnostic criteria of the International Workshop on Chronic Lymphocytic Leukemia (iwCLL)<sup>43</sup>. The study was approved by the local ethics review committee of each participating Institute, and conducted with informed consent.

### ***PCR amplification and sequence analysis***

PCR amplification of IGHV-IGHD-IGHJ gene rearrangements was performed prior to the application of any treatment on either genomic DNA (gDNA) or complementary DNA (cDNA), as described<sup>15</sup>. PCR products were gel-excised, purified and subjected to bidirectional direct sequencing<sup>44</sup>. Sequence data were annotated with the IMGT/HighV-QUEST tool<sup>45</sup>. Only in-frame, productive rearrangements were subjected to further analysis.

### ***Metadata analysis***

IMGT/HighV-QUEST output data were processed with available in-house algorithms. Information regarding IGH gene repertoires, SHM patterns and VH CDR3 characteristics was extracted, as described<sup>15</sup>.

### ***Subset assignment of IGH sequences***

Stereotyped BcR IGHV-IGHD-IGHJ gene rearrangements were identified using our established bioinformatics method<sup>15,46</sup>. Briefly, the applied clustering criteria were: (i) utilization of IGHV genes from the same phylogenetic clan, (ii) at least 50% amino acid (aa) identity and 70% similarity within the VH CDR3, (iii) identical VH CDR3

length and, (iv) identical offset of the shared aa pattern. Common sequence patterns between sequences belonging to different clusters led to their grouping in clusters at progressively higher levels of hierarchy describing more distant, thus more relaxed, sequence relationships with more widely shared sequence patterns.

A subset was defined as major if it represented at least 0.2% of the cohort i.e. when containing at least 60 cases; a similar cutoff (0.26%) had been used for identifying major subsets in our previous study<sup>15</sup>.

Sequence relatives of major subsets, i.e. subsets with similar immunogenetic features, hereafter termed 'satellites', were identified using the following parameters: (i) utilization of phylogenetically related IGHV genes, (ii) maximum VH CDR3 length difference of 2 aa, and, (iii) presence of the 'subset-specific' VH CDR3 sequence motifs. The analysis was performed individually for each major subset.

### ***Subset nomenclature***

In order to avoid potential confusion, the original subset names were retained for stereotyped subsets reported in previous studies<sup>11,12,14</sup>. For novel subsets, a new nomenclature was adopted, whereby subset names consisted of 4 bits, separated by vertical lines: (i) IGHV gene(s) utilized, (ii) IGHJ gene utilized, (iii) VH CDR3 aa length, (iv) a numerical character in serial order (in the case of subsets with identical IGHV gene and CDR3 length, e.g., two distinct novel subsets utilizing the IGHV3-23 and IGHJ4 genes and carrying VH CDR3s of 13 aa would be termed V3-23|J4|13|1 and V3-23|J4|13|2). Detailed information is provided in the **Supplementary material**.

### ***Data visualization tools***

VH CDR3 aa patterns of different subsets were visualized using WebLogo (<http://weblogo.berkeley.edu/>). Each logo consists of stacks of symbols, one stack for each position of the sequence. Amino acid position numbering is according to the IMGT numbering for the V domain.

### ***Statistical analysis***

Descriptive statistics for discrete variables included counts and frequency distributions. For quantitative variables, statistical measures included the mean,

median, and min-max values. The significance of bivariate/multivariate relationships between variables was assessed using the  $\chi^2$  and unpaired t-tests. For all comparisons a significance level of  $p=0.05$  was set. All statistical analyses were performed with SPSS V-22 and R V-3.6.1.

## RESULTS

### ***IGH gene repertoire and somatic hypermutation analysis***

Overall, 30,413 IGHV-IGHD-IGHJ gene rearrangements were amplified from 29,856 CLL patients; in 29,319 cases (98.2%) a single productive rearrangement was identified, whereas two productive rearrangements were detected in 517 cases (1.7%); finally, 20 cases (0.1%) carried 3 distinct IGH rearrangements. The IGH gene repertoires were in line with our previous study<sup>15</sup> (**Figure 1; Supplementary material; Supplementary Tables 2 and 3**).

The classification into unmutated CLL (U-CLL) and mutated CLL (M-CLL) subgroups was based on the 98% germline identity (GI) cutoff value. Amongst 29,319 cases carrying a single productive rearrangement, 15,859 (54.1%) were classified as M-CLL (GI<98%), and 13,460 (45.9%) as U-CLL (GI≥98%) (**Table 1**). More information can be found in the **Supplementary Material**.

### ***The stereotyped fraction accounts for a large percentage of all CLL***

In total, 12,541/30,413 sequences, representing 41.2% of the entire cohort, were assigned to 5,157 distinct ground level clusters, including 2-160 sequences each. In many cases, the existence of multiple shared patterns within VH CDR3 sequences led to their concurrent assignment in different ground-level subsets. This overlap enabled the discovery of subsets at 4 successive hierarchically higher levels, characterized by more broadly shared sequence patterns, hence, greater individual subset size. When considering all clustering levels, we identified 2,195 distinct stereotyped subsets ranging in size from 2 to 764 sequences each. Interestingly, random set simulations indicated that the size of the cohort correlated with the number of stereotyped subsets but not their relative frequency (**Supplementary Material; Supplementary Figure 1**).

### ***Major stereotyped subsets***

Twenty-nine major subsets with an individual frequency of at least 0.2% of the entire cohort were identified, cumulatively accounting for 4,099/30,413 (13.5%) of all CLL and for almost one-third (32.7%) of the stereotyped fraction. These included the 19

previously identified major subsets and 10 novel ones. Subset #2 (comprising cases from both M-CLL and U-CLL)<sup>34</sup> was the largest (764 sequences, 2.5%), followed by subsets #1 (U-CLL, 651 sequences, 2.1%) and #4 (M-CLL, 265 sequences, 0.9%). Subset #148B (M-CLL, 152 sequences, 0.5%) was the largest of the novel major subsets (**Figure 2**). Most major subsets concerned U-CLL (18/29 subsets, 62%). Detailed information for all major stereotyped subsets is provided in **Supplementary Tables 4** and **5**, whereas a list of all minor subsets is given in **Supplementary Table 6**.

### ***IGHV gene usage in stereotyped CLL subsets***

A single IGHV gene clearly predominated in the stereotyped IGHV-IGHD-IGHJ gene rearrangements of most major subsets (23/29, 79.3%). Within this category of ‘single dominant IGHV’ subsets, IGHV1-69 predominated (6/23 subsets, 26%), followed by the IGHV4-34, IGHV4-39 and IGHV3-48 genes (each found in 3/23 subsets, 13%). ‘Single dominant IGHV’ major subsets exhibited concordant SHM status.

Three of the remaining 6 major subsets (#12, #59 and #77)<sup>15</sup> included rearrangements utilizing 2 different IGHV genes with comparable frequencies. In all 3 cases, the respective IGHV genes belonged to the same phylogenetic clan and, thus, exhibited high sequence similarity: IGHV1-2 and IGHV1-46 in subset #12 (U-CLL), IGHV1-58 and IGHV1-69 in subset #59 (U-CLL) and, finally, IGHV4-4 and IGHV4-59 in subset #77 (M-CLL), respectively (**Supplementary Figure 2**).

The remaining 3 major subsets included cases expressing several different yet phylogenetically-related IGHV genes; these included the well-established subset #1, the closely related subset #99 (both exhibiting rearrangements of clan I IGHV genes) and subset #202 (clan III IGHV genes). All 3 patterns of IGHV gene usage (i.e. ‘single dominant IGHV’, ‘two IGHV/same clan’, ‘multiple IGHV/same clan’) were also evident in minor subsets (**Supplementary Table 7**).

### ***Different BcR IG stereotypy profiles in U-CLL and M-CLL***

The incidence of stereotyped BcR IG differed significantly between U-CLL versus M-CLL. Specifically, 53% of U-CLL BcR IGs were assigned to stereotyped subsets, compared to 34% in M-CLL ( $p < 0.05$ ) (**Figure 3**). In our previous study<sup>15</sup>, the respective values were 41% for U-CLL and 22% for M-CLL: hence, we have now uncovered a



significant increase ( $p < 0.05$ ) in the frequency of BcR IG stereotypy in both mutational subgroups.

Most major U-CLL subsets exhibited high conservation across the entire VH CDR3. Among major subsets, the most widely conserved motif was identified in subset #10 (**Supplementary Figure 3A**), where the subset-defining motif comprised 20/22 VH CDR3 aa. Such remarkable conservation was also evident in minor subsets, e.g. minor subset #19A (U-CLL) where the shared motif comprised 17/19 VH CDR3 aa (**Supplementary Figure 4A**).

In contrast, M-CLL stereotyped subsets were often characterized by relatively degenerate sequence motifs with only few highly restricted ('landmark') aa residues. Indicatively, in major subset #148B the conserved residues were mostly located at either the 5' (IGHV-encoded) or the 3' (IGHJ-encoded) part of the VH CDR3. Most intermediate positions, corresponding to the tip of the VH CDR3 loop, were characterized by extreme degeneracy with two striking exceptions of 'landmark' residues: in particular, 150/152 cases (98.7%) carried a tryptophan (W) residue at VH CDR3 position 10, while 135/152 cases (88.8%) carried a glycine (G) residue at VH CDR3 position 13 (**Supplementary Figure 3B**).

Subsequently, we scanned M-CLL subsets for additional traces of stereotypy beyond the VH CDR3, in particular for "subset-specific" patterns of recurrent SHM as reflected in identical aa substitutions at the same positions across the rearranged IGHV gene. Restricted aa changes were documented in all major and several minor M-CLL subsets, often with degenerate VH CDR3 motifs (**Supplementary Tables 8-10; Supplementary Figures 4B-C**), supporting that they likely represent true CLL variants with shared immunogenetic features rather than mere algorithmically-derived sequence clusters.

A detailed account of recurrent SHMs in stereotyped subsets is provided in the **Supplementary Material**.

### ***Immunogenetically related stereotyped CLL subsets: introducing the new concept of satellite subsets***

Amongst major stereotyped subsets, 4 pairs of U-CLL subsets with highly similar VH CDR3 motifs were identified. The first concerned subsets #1 and #99, both

comprising cases expressing different IGHV genes of phylogenetic clan I (IGHV1/IGHV5/IGHV7 gene subgroups). The VH CDR3 of these subsets had similar composition but differed in length (13 aa in subset #1 versus 14 aa in subset #99). The second pair concerned subsets #8 and #8B, both utilizing the IGHV4-39 gene, whose major difference again involved the VH CDR3 length (19 aa in subset #8 versus 18 aa in subset #8B) (**Supplemental Figure 5A**). For the third pair, subsets #3C2 and #3C3 both used the IGHV1-69 gene: the identified difference involved the offset of the shared VH CDR3 motif by a single position (**Supplemental Figure 5B**). Finally, subsets #7C2 and #7D3, both expressing the IGHV1-69 gene, differed in both VH CDR3 length (23 versus 24 aa, respectively) and motif offset (by a single position) (**Supplemental Figure 5C**).

Next, we searched for immunogenetic relatives of major subsets across the entire stereotyped fraction and identified 156 distinct minor subsets that showed high similarity and, thus, were considered ‘satellites’ to 21 major subsets. Altogether, ‘satellite’ subsets (934 sequences) comprised 3% of the total cohort and 7.5% of the stereotyped fraction. A list of all BcR IG sequences belonging to ‘satellite’ subsets is given in **Supplementary Table 11**. Most ‘satellite’ subsets were small (median number of 3 sequences), albeit populated subsets were also evident (with up to 49 sequences).

Finally, we assessed whether the grouping of each major subset together with its satellite(s) would have a substantial impact on its relative size (**Figure 4**). Pronounced increases in subset size were seen in major U-CLL subsets #7D3, #169 and #7C2. In M-CLL, subset #148B exhibited the largest number of satellite subsets (collectively accounting for 61 sequences; **Supplementary Material; Supplementary Figure 6**).

A final noteworthy case concerned subset #2 that exhibited 8 immunogenetically related subsets all carrying the characteristic ‘landmark’ residue D at VH CDR3 position 3. One of these subsets was major, namely subset #169 (62 sequences), whereas the others were minor (**Figure 5**). Grouping together subset #2 (764 sequences) and its relatives (106 sequences) led to a cumulative frequency of 2.9% (870/30,413 sequences).

In order to obtain supportive evidence for the relevance of satellite subsets, particularly their clinical significance, if any, we focused on the paradigmatic examples concerning subsets #2 and #169 as well as subsets #1 and #99. In specific, time-to-first-treatment (TTFT) and overall survival analysis for all cases with available data belonging to subsets #2 (212 cases) and #169 (14 cases) showed consistent clinical profiles without any statistical significance ( $p=0.98$  and  $p=0.23$ , respectively). Similar observations were made when considering patients assigned to subsets #1 and #99 for TTFT (139 versus 17 cases, respectively) ( $p=0.45$ ) and OS (153 versus 19 cases, respectively) ( $p=0.65$ ) (**Figure 6**).

## DISCUSSION

BcR IG stereotypy has emerged as relevant for unraveling the complex disease biology of CLL, at least for specific patient subgroups,<sup>46,47</sup> while also improving patient stratification<sup>34,35,38,41,42,48,49</sup>. BcR IG stereotypy is not exclusive to CLL, being evident in other mature B cell malignancies as well<sup>40,50-52</sup>. However, BcR IG stereotypes are significantly less frequent in these malignancies and display fundamentally different immunogenetic properties compared to CLL. BcR IG stereotypy has been also identified in other contexts, like infections<sup>53,54</sup> and immune-mediated diseases e.g. celiac disease<sup>55</sup>. Furthermore, specific B cell subpopulations have been reported to show an inherently higher tendency to express restricted BcR IG<sup>56</sup>. These findings underscore the need for further analysis to fully understand the relevance of restricted immune profiles in health and disease.

In this context, we recently scanned the BcR IG repertoire of normal B cell subpopulations in healthy individuals<sup>57,58</sup> as well as individuals with monoclonal B cell lymphocytosis (MBL)<sup>59</sup> for the presence of “CLL-specific” stereotypes. We reported that “CLL-specific” stereotyped BcR IG represented a small fraction of the repertoire in all B cell subpopulations studied, thus, questioning the hypothesis for a single normal B cell progenitor of CLL. Moreover, a very low frequency or even absence of CLL stereotypes related to disease aggressiveness was evident in all settings. These findings highlight the unique immunogenetic signature of CLL, indicating that censoring may normally be exerted on cells expressing “CLL-like” BcR IG. Moreover, they emphasize the role of microenvironmental triggering, mediated through the BcR, in CLL ontogeny and evolution.

Our present analysis of BcR IG stereotypy in a series of 29,856 CLL patients, 4-times larger than our previous study<sup>15</sup>, aimed to refine the immunogenetic characterization of stereotyped subsets as a first, critical step towards identifying potential distinct variants of CLL. Overall, a significant increase in the size of the CLL stereotyped fraction was identified, compared to our previous estimate<sup>15</sup>. However, this did not affect the relative frequency of individual subsets, since those defined previously as major retained their status. Arguably, these may represent distinct

disease subgroups amenable to compartmentalized research with the potential of overcoming the pronounced heterogeneity of CLL.

To gain better insight into the impact of cohort size on BcR IG stereotypy, we performed random data subsampling and built cohorts of different size. In line with previous studies<sup>11-15</sup>, the frequency of stereotypy correlated with cohort size, accomplished through the formation of more subsets without any change in the average subset size. Still, the rate by which the frequency of stereotypy increased was not constant but gradually decreased as the analysis progressed to larger cohorts. This finding supports the notion that not all CLL cases are stereotyped, while still consistent with different ontogenetic trajectories for stereotyped and non-stereotyped cases<sup>14,60</sup>. Arguably, the stereotyped fraction of CLL might be largely dependent on the functional properties of the clonotypic BcR IG, whereas other factors, i.e. genomic aberrations, could drive disease pathogenesis in the non-stereotyped group.

Subsequently, we assessed the relationship between BcR IG stereotypy and other critical immunogenetic characteristics of the clonotypic BcR IG, such as IGHV gene usage and SHM status. We report that the usage of 2 or more different, albeit phylogenetically related, IGHV genes in cases assigned to the same subset is relatively frequent in CLL, recalling findings from other contexts, particularly infections, e.g. (i) CD4-mimetic broadly neutralizing antibodies against HIV-1 that utilize the IGHV1-2 and IGHV1-46 genes<sup>61-63</sup>; (ii) influenza-virus receptor binding site-directed antibodies that express the IGHV4-59, IGHV4-39, IGHV4-38-1 and IGHV4-30-2 genes<sup>64</sup>; (iii) anti-influenza antibodies against the hemagglutinin stalk that are preferentially encoded by the IGHV1-69 and the highly related IGHV1-18 gene<sup>65</sup>; (iv) potent neutralizing antibodies against SARS-CoV-2, which are encoded by the IGHV3-53 and IGHV3-66 genes<sup>66</sup>. A possible explanation of this finding, plausible for both CLL but other contexts as well, could be that related IGHV genes have either germline-encoded (U-CLL) or SHM-acquired (M-CLL) specific residues that are crucial for antigen recognition. Alternatively, different IGHV genes may adopt similar conformations in the three-dimensional (3D) space. This notion is supported by published evidence regarding shape-based clustering of 3D IG models, where subset

#1 cases utilizing different, yet phylogenetically related, IGHV genes displayed highly similar structures<sup>67</sup>.

The clinical importance of SHM status for clinical decision making in CLL is reflected in the guidelines for disease management provided by the iwCLL in 2018, which mandate that this information is essential before deciding about treatment in all CLL patients both in general practice as well as in clinical trials<sup>43</sup>. In our previous study<sup>15</sup>, we observed that stereotyped rearrangements were enriched in U-CLL and the same has held in the present analysis. This might reflect a distinct cellular derivation for U-CLL versus M-CLL, whereby the former may correspond to cell subpopulations with inherently restricted BcR IG repertoires, capable of recognizing a broad range of antigens due to the expression of certain IGHV genes, provided that these are retained in germline configuration i.e. devoid of SHM<sup>68-70</sup>. In that respect, the BcR IGs in U-CLL recall natural autoantibodies that are almost exclusively unmutated and display restricted immunogenetic features<sup>68-71</sup>.

U-CLL subsets displayed a high level of conservation across the entire VH CDR3, essentially deriving from the utilization of the same IGHV, IGHD and IGHJ genes. On the other hand, many M-CLL subsets were characterized by degenerate VH CDR3 motifs with few “landmark” (i.e. highly conserved) residues, albeit they also displayed remarkably restricted, “subset-biased” SHM patterns. Again, these findings are strongly reminiscent of the IG repertoires in viral infections: for instance, in many influenza-virus receptor binding site-directed antibodies, viral engagement appears to depend on the presence of “landmark” residues i.e. just a single dipeptide within the VH CDR3<sup>64</sup>.

Last, we searched for more distant sequence relatives of major BcR IG stereotypes in CLL. The rationale behind this analysis was largely based on accumulating evidence that BcR IG encoded by distinct yet phylogenetically related IGHV genes can recognize similar antigenic elements in various contexts, particularly infections, as mentioned above; and, preliminary evidence from our previous study<sup>15</sup> strongly alluding to the existence of immunogenetically related subsets: some of these were subsequently found to share similar landscapes of genomic aberrations (e.g. subset #2 and its satellite subset #169 have been found to exhibit a strikingly higher frequency of *SF3B1* gene mutations compared to the general CLL cohort)<sup>72,73</sup>.

Immunogenetic relations were identified between major subsets as well as between major and minor subsets, mostly concerning U-CLL. Indeed, many major U-CLL subsets had satellite subsets, often large in size, whereas most major M-CLL subsets exhibited few connections, perhaps due to the high level of specificity of their VH CDR3 shaped by SHM. Overall, the grouping of major subsets along with their satellites would lead, at least in certain cases, to a significant increase in their relative size.

Even though the status of 'satellite' subsets could be considered as rather 'provisional', we here provide clinical evidence supporting their relevance, as exemplified by the finding of similar TTFT and OS in stereotyped subset #2 and its satellite subset #169 as well as subset #1 and its satellite subset #99. Admittedly, further research is clearly warranted, however, this preliminary evidence indicates that the identification of 'satellite' subsets could have important implications for biological as well as clinical research in CLL aimed at dissecting disease heterogeneity towards the identification of distinct profiles that would assist in clinical decisions.

BcR stereotypy studies in CLL have focused mainly on the analysis of the VH CDR3, a choice justified by the critical role of this part of the BcR IG in antigen recognition, at least in the pre-immune repertoire<sup>74</sup>. That said, we and others have offered evidence that BcR IG light chains are also involved in antigen selection processes in CLL<sup>75-77</sup>. This was further supported by recent studies highlighting a unique biological background for CLL cases expressing the IGLV3-21 gene<sup>78</sup>, particularly those carrying distinctive imprints of SHM<sup>79</sup>. Our present analysis, although confined to the BcR IG heavy chain, appears to corroborate this notion, considering that subsets #2 and #169, the prime example of higher-order connections between stereotyped subsets, both utilize the IGLV3-21 gene and carry restricted VL CDR3. Evidently, further work is required, however, the possibility exists for antibody light-chain-restricted antigen recognition in CLL, similar to what has been reported in other contexts as well<sup>80</sup>.

In conclusion, we report a significant increase in the frequency of BcR IG stereotypy in by far the largest-to-date CLL cohort, while also offering evidence that not all CLL will end up being stereotyped. Importantly, on the basis of our present findings, major subsets emerge as distinct immunogenetic disease variants that can be robustly identified and further studied with the aim of better understanding the

natural history of CLL. Finally, we introduce the new concept of satellite subsets, which argues for even greater restriction in the CLL BcR IG repertoire than previously thought. This novel immunogenetic classification scheme has the potential to assist in assessing the clinical impact of BcR IG stereotypy through the characterization of the immunogenetic relationship between major subsets and their satellites with obvious implications for refined molecular classification of CLL.



## **Acknowledgments**

This project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 336 (Project CLLon); the Hellenic Precision Medicine Network in Oncology; the Swedish Cancer Society, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, Karolinska Institutet, Karolinska University Hospital, and Radiumhemmets Forskningsfonder, Stockholm; DFG subprojects B1 and B2 of SFB1074; the Bournemouth Leukaemia Fund; the Kuwait Foundation for Advancement of Sciences (KFAS research grant PR1713MM03); the Polish Scientific Centre (NCN 2018/29/B/NZ5/02706); the Swiss Cancer League, ID 3746, 4395 4660, and 4705, Bern, Switzerland; the European Research Council (ERC) Consolidator Grant CLLCLONE, ID: 772051; the Swiss National Science Foundation, ID 320030\_169670/1 and 310030\_192439, Berne, Switzerland; the Leukemia & Lymphoma Society, Translational Research Program, ID 6594-20, New York; the AIRC I.G. 15426; the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 794075; the Momentum grant (LP-95021) of the Hungarian Academy of Sciences and the NVKP\_16-1-2016-0004 grant of the Hungarian National Research, Development and Innovation Office; the Leukaemia and Lymphoma NI; the AIRC Milan project IG 15397; the grant MH-CZ AZV NV19-03-00091, project NPUII MEYSCZ (CEITEC2020 LQ1601), University Hospital Brno project (MH-CZ RVO 65269705); the Associazione Italiana per la Ricerca sul Cancro (AIRC), 5 x 1000 grant No. 21198, Milan, Italy; the Novo Nordisk Foundation grant NNF16OC0019302.

## **Authorship Contributions**

An.Ag. designed the research, performed the experiments, analyzed the data and wrote the manuscript; Ka.Ge., Ma.Ka., Pa.Ba. and Th.Mo. analyzed data; Ve.Gi., Ma.Le. and So.Ko. were responsible for the curation of the IGHV-IGHD-IGHJ

sequences; Ka.Pl., Za.Da., Xi.Ya., Sa.Je., Ch.Sc., Lo.Br-Pe., Re.Ts., Le.Su., Ly.Sc., El.Ga., Ma.Ar., Eu.Ta., Be.Bi., Co.Ba., Da.Ba., Al. Na., An.La., Va.Gu., Ge.Mi., Al.Di., Ch.Br., Sa.La., Ma.Me., Ka.Br., Ma.Ri., Mo.Fa., Cr.Tr., An.Vi., An.Pa., Ma.Ca., Li.Bo., An.Su., Ka.Va., Ma.Ro., Ha.Sk., Si.Ve., Kr.Gi., La.Ma., Te.Ka., Ra.Sa., Cs.Bö., Fr.Fa., Ar.Ka., Ir.Pa., Da.Ro., Sa.Al., Pa.Pa., Pa.Co., Bl.Es., Da.An., Le.Fo., Ma.Mo., Li.Tr., Ni.St., Gi.Ga., Pa.Ce., Ca.Ni., El.Ca., Ac.An., Ch.Po., Ki.Fi., Mi.Ha., Da.Os., St.St., Cl.Ha., Di.Je., Ni.Ch. and Sa.Po., provided data; An.Ch., An.La., Ch.Be., Fr.Da. Ri.Ro. and Pa.Gh. supervised research; Ko.St. designed the study, supervised research, and wrote the manuscript.

### **Disclosure of Conflicts of Interest**

Si.Ve. has received honoraria from Bayer, Astra-Zeneca and Janssen. Ar.Ka. has received research funding from Abbvie, Roche/Genentech, Janssen and Astra-Zeneca. Da.Ro. has received honoraria from Abbvie, Astra-Zeneca, Gilead, Janssen, Verastem, and research grants from Abbvie, Gilead, Janssen, Cellectia. Sa.Al. has received educational grants from J&J, AbbVie, and Roche. Kr.Gi. has received Honoraria from AbbVie, Janssen and Roche. Li.Te. has received honoraria from Abbvie, Roche, Janssen, Shire. An.Vi. has received honoraria from Janssen, Abbvie. Gi.Ga. has received honoraria from Abbvie, Janssen, Sunesys and Astra-Zeneca for advisory boards or speaker's bureau services. Ca.Ni. has received research support, consultancy fees and/or travel grants from Abbvie, Gilead, Janssen, Roche, CSL Behring, Genmab, Sunesis and Acerta/Astra-Zeneca outside this work. Ki.Fi. has received honoraria from Roche and Abbvie, and Roche Travel grants. St.St. has received honoraria and research support from AbbVie, Astra-Zeneca, Celgene, Gilead, GSK, Hoffmann La-Roche, Janssen, Novartis. Ri.Ro. has received honoraria from AbbVie, Illumina, Janssen and Roche. Pa.Gh. has received honoraria from AbbVie, Acerta, BeiGene, Gilead, Janssen, Sunesis and research funding from AbbVie, Gilead, Janssen, Novartis, Sunesis. Ko.St. has received honoraria and research support from Abbvie, Janssen, Astra-Zeneca and Gilead.

### **Data Sharing**

For original data, please email the corresponding author.

## References

1. Sutton LA, Rosenquist R. The complex interplay between cell-intrinsic and cell-extrinsic factors driving the evolution of chronic lymphocytic leukemia. *Semin Cancer Biol.* 2015;34:22-35.
2. Puente XS, Jares P, Campo E. Chronic lymphocytic leukemia and mantle cell lymphoma: crossroads of genetic and microenvironment interactions. *Blood.* 2018;131(21):2283-2296.
3. Byrd JC, Furman RR, Coutre SE, et al. Three-year follow-up of treatment-naive and previously treated patients with CLL and SLL receiving single-agent ibrutinib. *Blood.* 2015;125(16):2497-2506.
4. Jones JA, Robak T, Brown JR, et al. Efficacy and safety of idelalisib in combination with ofatumumab for previously treated chronic lymphocytic leukaemia: an open-label, randomised phase 3 trial. *Lancet Haematol.* 2017;4(3):e114-e126.
5. Brown JR, Hillmen P, O'Brien S, et al. Extended follow-up and impact of high-risk prognostic factors from the phase 3 RESONATE study in patients with previously treated CLL/SLL. *Leukemia.* 2018;32(1):83-91.
6. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J Clin Invest.* 1998;102(8):1515-1525.
7. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* 1999;94(6):1840-1847.
8. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood.* 1999;94(6):1848-1854.
9. Stevenson FK, Forconi F, Packham G. The meaning and relevance of B-cell receptor structure and function in chronic lymphocytic leukemia. *Semin Hematol.* 2014;51(3):158-167.
10. Langerak AW, Davi F, Stamatopoulos K. Immunoglobulin heavy variable somatic hyper mutation status in chronic lymphocytic leukaemia: on the threshold of a new era? *Br J Haematol.* 2020;189(5):809-810.
11. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. *Blood.* 2007;109(1):259-270.
12. Murray F, Darzentas N, Hadzidimitriou A, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. *Blood.* 2008;111(3):1524-1533.
13. Bomben R, Dal Bo M, Capello D, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. *Br J Haematol.* 2009;144(4):492-506.
14. Darzentas N, Hadzidimitriou A, Murray F, et al. A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. *Leukemia.* 2010;24(1):125-132.

15. Agathangelidis A, Darzentas N, Hadzidimitriou A, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119(19):4467-4475.
16. CATERA R, LIU Y, GAO C, et al. Binding of CLL subset 4 B-cell receptor immunoglobulins to viable human memory B lymphocytes requires a distinctive IGKV somatic mutation. *Mol Med*. 2017;23:1-12.
17. CHU CC, CATERA R, HATZI K, et al. Chronic lymphocytic leukemia antibodies with a common stereotypic rearrangement recognize nonmuscle myosin heavy chain IIA. *Blood*. 2008;112(13):5122-5129.
18. CHU CC, CATERA R, ZHANG L, et al. Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. *Blood*. 2010;115(19):3907-3915.
19. GOUNARI M, NTOUFA S, APOLLONIO B, et al. Excessive antigen reactivity may underlie the clinical aggressiveness of chronic lymphocytic leukemia stereotyped subset #8. *Blood*. 2015;125(23):3580-3587.
20. SEILER T, WOELFLE M, YANCOPOULOS S, et al. Characterization of structurally defined epitopes recognized by monoclonal antibodies produced by chronic lymphocytic leukemia B cells. *Blood*. 2009;114(17):3615-3624.
21. MALCIKOVA J, STALIKA E, DAVIS Z, et al. The frequency of TP53 gene defects differs between chronic lymphocytic leukaemia subgroups harbouring distinct antigen receptors. *Br J Haematol*. 2014;166(4):621-625.
22. NAVRKALOVA V, YOUNG E, BALIAKAS P, et al. ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset #2 is associated with markedly short telomeres. *Haematologica*. 2016;101(9):e369-373.
23. STREFFORD JC, SUTTON LA, BALIAKAS P, et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia*. 2013;27(11):2196-2199.
24. KANDURI M, MARINCEVIC M, HALDORSDOTTIR AM, et al. Distinct transcriptional control in major immunogenetic subsets of chronic lymphocytic leukemia exhibiting subset-biased global DNA methylation profiles. *Epigenetics*. 2012;7(12):1435-1442.
25. MARINCEVIC M, MANSOURI M, KANDURI M, et al. Distinct gene expression profiles in subsets of chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors. *Haematologica*. 2010;95(12):2072-2079.
26. PAPANIKOLAOU N, NTOUFA S, CHARTOMATSIDOU E, et al. Differential microRNA profiles and their functional implications in different immunogenetic subsets of chronic lymphocytic leukemia. *Mol Med*. 2013;19:115-123.
27. PAPANIKOLAOU N, NTOUFA S, TSAGIPOULOU M, et al. Integrated epigenomic and transcriptomic analysis reveals TP63 as a novel player in clinically aggressive chronic lymphocytic leukemia. *Int J Cancer*. 2018;144(11):2695-2706.
28. MAURA F, CUTRONA G, MOSCA L, et al. Association between gene and miRNA expression profiles and stereotyped subset #4 B-cell receptor in chronic lymphocytic leukemia. *Leuk Lymphoma*. 2015;56(11):3150-3158.
29. ARVANITI E, NTOUFA S, PAPANIKOLAOU N, et al. Toll-like receptor signaling pathway in chronic lymphocytic leukemia: distinct gene expression profiles of potential pathogenic significance in specific subsets of patients. *Haematologica*. 2011;96(11):1644-1652.
30. NTOUFA S, VARDI A, PAPANIKOLAOU N, et al. Distinct innate immunity pathways to activation and tolerance in subgroups of chronic lymphocytic leukemia with distinct immunoglobulin receptors. *Mol Med*. 2012;18:1281-1291.
31. NTOUFA S, PAPANIKOLAOU N, APOLLONIO B, et al. B Cell Energy Modulated by TLR1/2 and the miR-17 approximately 92 Cluster Underlies the Indolent Clinical Course of Chronic Lymphocytic Leukemia Stereotyped Subset #4. *J Immunol*. 2016;196(10):4410-4417.

32. Duhren-von Minden M, Ubelhart R, Schneider D, et al. Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. *Nature*. 2012;489(7415):309-312.
33. Minici C, Gounari M, Ubelhart R, et al. Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. *Nat Commun*. 2017;8:15746.
34. Baliakas P, Agathangelidis A, Hadzidimitriou A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood*. 2015;125(5):856-859.
35. Jaramillo S, Agathangelidis A, Schneider C, et al. Prognostic impact of prevalent chronic lymphocytic leukemia stereotyped subsets: analysis within prospective clinical trials of the German CLL Study Group (GCLLSG). *Haematologica*. 2019.
36. Rossi D, Spina V, Cerri M, et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. *Clin Cancer Res*. 2009;15(13):4415-4422.
37. Visco C, Maura F, Tuana G, et al. Immune thrombocytopenia in patients with chronic lymphocytic leukemia is associated with stereotyped B-cell receptors. *Clin Cancer Res*. 2012;18(7):1870-1878.
38. Baliakas P, Hadzidimitriou A, Sutton LA, et al. Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. *Lancet Haematol*. 2014;1(2):e74-84.
39. Del Giudice I, Chiaretti S, Santangelo S, et al. Stereotyped subset #1 chronic lymphocytic leukemia: a direct link between B-cell receptor structure, function, and patients' prognosis. *Am J Hematol*. 2014;89(1):74-82.
40. Xochelli A, Baliakas P, Kavakiotis I, et al. Chronic Lymphocytic Leukemia with Mutated IGHV4-34 Receptors: Shared and Distinct Immunogenetic Features and Clinical Outcomes. *Clin Cancer Res*. 2017;23(17):5292-5301.
41. Baliakas P, Mattsson M, Hadzidimitriou A, et al. No improvement in long-term survival over time for chronic lymphocytic leukemia patients in stereotyped subsets #1 and #2 treated with chemo(immuno)therapy. *Haematologica*. 2018;103(4):e158-e161.
42. Baliakas P, Moysiadis T, Hadzidimitriou A, et al. Tailored approaches grounded on immunogenetic features for refined prognostication in chronic lymphocytic leukemia. *Haematologica*. 2019;104(2):360-369.
43. Hallek M, Cheson BD, Catovsky D, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131(25):2745-2760.
44. Rosenquist R, Ghia P, Hadzidimitriou A, et al. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. *Leukemia*. 2017;31(7):1477-1481.
45. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol Biol*. 2012;882:569-604.
46. Agathangelidis A, Psomopoulos F, Stamatopoulos K. Stereotyped B Cell Receptor Immunoglobulins in B Cell Lymphomas. *Methods Mol Biol*. 2019;1956:139-155.
47. Ten Hacken E, Gounari M, Ghia P, Burger JA. The importance of B cell receptor isotypes and stereotypes in chronic lymphocytic leukemia. *Leukemia*. 2019;33(2):287-298.
48. Maura F, Cutrona G, Fabris S, et al. Relevance of stereotyped B-cell receptors in the context of the molecular, cytogenetic and clinical features of chronic lymphocytic leukemia. *PLoS One*. 2011;6(8):e24313.

49. Moysiadis T, Baliakas P, Rossi D, et al. Different time-dependent changes of risk for evolution in chronic lymphocytic leukemia with mutated or unmutated antigen B cell receptors. *Leukemia*. 2019;33(7):1801-1805.
50. Belhouachi N, Xochelli A, Boudjoghra M, et al. Primary vitreoretinal lymphomas display a remarkably restricted immunoglobulin gene repertoire. *Blood Adv*. 2020;in press.
51. 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(7):1638-1646.
52. Hadzidimitriou A, Agathangelidis A, Darzentas N, et al. Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. *Blood*. 2011;118(11):3088-3095.
53. Henry Dunand CJ, Wilson PC. Restricted, canonical, stereotyped and convergent immunoglobulin responses. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1676).
54. Joyce MG, Wheatley AK, Thomas PV, et al. Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses. *Cell*. 2016;166(3):609-623.
55. Snir O, Chen X, Gidoni M, et al. Stereotyped antibody responses target posttranslationally modified gluten in celiac disease. *JCI Insight*. 2017;2(17).
56. Giltiay NV, Giordano D, Clark EA. The Plasticity of Newly Formed B Cells. *J Immunol*. 2019;203(12):3095-3104.
57. Vergani S, Bagnara D, Mazzarello AN, et al. CLL Stereotyped IGHV-D-J Rearrangements Can Be Detected Throughout Normal B-Cell Developmental Stages in Aged People When Using Ultra-Deep, Next Generation Sequencing Techniques. 58th ASH annual meeting. San Diego, CA, USA; 2016:2028.
58. Colombo M, Bagnara D, Reverberi D, et al. Tracing CLL-biased stereotyped immunoglobulin gene rearrangements in normal B cell subsets using a high-throughput immunogenetic approach. *Mol Med*. 2020;26(1):25.
59. Agathangelidis A, Galigalidou C, Scarfo L, et al. Infrequent "chronic lymphocytic leukemia-specific" immunoglobulin stereotypes in aged individuals with or without low count monoclonal B cell lymphocytosis. *Haematologica*. 2020.
60. Stamatopoulos K, Agathangelidis A, Rosenquist R, Ghia P. Antigen receptor stereotypy in chronic lymphocytic leukemia. *Leukemia*. 2017;31(2):282-291.
61. Scheid JF, Mouquet H, Ueberheide B, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011;333(6049):1633-1637.
62. Wu X, Zhou T, Zhu J, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*. 2011;333(6049):1593-1602.
63. Zhou T, Zhu J, Wu X, et al. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity*. 2013;39(2):245-258.
64. Schmidt AG, Therkelsen MD, Stewart S, et al. Viral receptor-binding site antibodies with diverse germline origins. *Cell*. 2015;161(5):1026-1034.
65. Ekiert DC, Bhabha G, Elsliger MA, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science*. 2009;324(5924):246-251.
66. Cao Y, Su B, Guo X, et al. Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell*. 2020;182(1):73-84 e16.
67. Polychronidou E, Kalamaras I, Agathangelidis A, et al. Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia. *BMC Bioinformatics*. 2018;19(Suppl 14):414.
68. Forconi F, Potter KN, Sozzi E, et al. The IGHV1-69/IGHJ3 recombinations of unmutated CLL are distinct from those of normal B cells. *Blood*. 2012;119(9):2106-2109.

69. Forconi F, Potter KN, Wheatley I, et al. The normal IGHV1-69-derived B-cell repertoire contains stereotypic patterns characteristic of unmutated CLL. *Blood*. 2010;115(1):71-77.
70. Herve M, Xu K, Ng YS, et al. Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *J Clin Invest*. 2005;115(6):1636-1643.
71. Borche L, Lim A, Binet JL, Dighiero G. Evidence that chronic lymphocytic leukemia B lymphocytes are frequently committed to production of natural autoantibodies. *Blood*. 1990;76(3):562-569.
72. Baliakas P, Hadzidimitriou A, Sutton L-A, et al. Differential Distribution Of Recurrent Gene Mutations In Subsets Of Chronic Lymphocytic Leukemia Patients With Stereotyped B-Cell Receptors: Results From A Multicenter Project Of The European Research Initiative On CLL In A Series Of 2482 Cases. *56th ASH Annual Meeting*. 2013;122(21):4113.
73. Sutton LA, Young E, Baliakas P, et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica*. 2016;101(8):959-967.
74. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000;13(1):37-45.
75. Hadzidimitriou A, Darzentas N, Murray F, et al. Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. *Blood*. 2009;113(2):403-411.
76. Kostareli E, Sutton LA, Hadzidimitriou A, et al. Intraclonal diversification of immunoglobulin light chains in a subset of chronic lymphocytic leukemia alludes to antigen-driven clonal evolution. *Leukemia*. 2010;24(7):1317-1324.
77. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*. 2005;106(10):3575-3583.
78. Stamatopoulos B, Smith T, Crompton E, et al. The Light Chain IgLV3-21 Defines a New Poor Prognostic Subgroup in Chronic Lymphocytic Leukemia: Results of a Multicenter Study. *Clin Cancer Res*. 2018;24(20):5048-5057.
79. Maity PC, Bilal M, Koning MT, et al. IGLV3-21\*01 is an inherited risk factor for CLL through the acquisition of a single-point mutation enabling autonomous BCR signaling. *Proc Natl Acad Sci U S A*. 2020;117(8):4320-4327.
80. Wiehe K, Easterhoff D, Luo K, et al. Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity*. 2014;41(6):909-918.

**Tables**

**Table 1.** Somatic hypermutation status at cohort level.

<b>Mutational groups</b>	<b>No of BcR IG rearrangements</b>	<b>% of BcR IG rearrangements</b>
Truly unmutated (GI=100%)	9.585	32.7
Unmutated (98≤GI<99.9%)	3.875	13.2
Mutated (GI<98%)	15.859	54.1



## Figure Legends

**Figure 1. The IGHV gene repertoire of stereotyped CLL is clearly distinct from the general cohort.** Cases expressing the IGHV1-69 and IGHV3-21 genes were significantly more in the stereotyped fraction of CLL compared to the heterogeneous group (16.4% versus 12.5% and 7.9% versus 4.8%, respectively). On the other hand, cases carrying the IGHV3-23 showed the opposite trend (4.6% in the stereotyped fraction versus 7.6% in the general cohort). Finally, the frequency of IGHV4-34 expressing cases was high in both groups (8.8% versus 8.7%).

**Figure 2. Cases assigned to major stereotyped subsets represent a significant portion of CLL.** Twenty-nine different subsets were identified in the present study containing a minimum of 60 cases (0.2% of the cohort) and were defined as major. The relative size of each major subset is indicated in the graph. Their basic immunogenetic information is given in **Supplementary Table 4**, while all sequences assigned to each major subset are listed in **Supplementary Table 5**. Altogether, major subsets comprised in total 4,098/30,413 rearrangements that corresponds to the 13.5% of the cohort.

**Figure 3. Frequency of BcR IG stereotypy in the CLL mutational subgroups.** The majority of cases carrying BcR IG sequences belonging to U-CLL were assigned to stereotyped subsets compared to around one-third of cases from M-CLL.

**Figure 4. The impact of satellite subsets on the relative size of major stereotyped subsets.** More pronounced increases were evident in major subsets comprising U-CLL cases, whereas most M-CLL major subsets were not significantly affected.

Characteristic examples concern subsets #1, #2 and #8 that correlate with dismal prognosis. On the other hand, subset #148B showed the highest increase in the category of M-CLL subsets.

**Figure 5. Major subset #2 and its close immunogenetic relatives (satellite subsets).**

Stereotyped subset #2 had 8 satellite subsets with the most frequent being subset #169, also characterized as major. The remaining satellite subsets were minor and cumulatively accounted for 44 BcR IG sequences. Subset #2 satellite full names are the following: #164 (Sat-1), #2B (Sat-2), V3-30|9|1 (Sat-3), V3|11|42 (Sat-4), V3-74|8|1 (Sat-5), V3-48|7|1 (Sat-6) and V3|10|25 (Sat-7).

**Figure 6. Patients assigned to major stereotyped subsets and their satellites exhibit consistent clinical profiles.**

Consistent clinical outcomes were observed between patients belonging to subset #2 and subset #169, its main satellite, with no statistical difference regarding: **(A)** time-to-first-treatment (TTFT) ( $p=0.98$ ) and **(B)** overall survival (OS) ( $p=0.23$ ). Similarly, patients from stereotyped subset #1 and #99, its main satellite, showed highly similar **(C)** TTFT ( $p=0.45$ ), and **(D)** OS ( $p=0.65$ ).

Figure 1

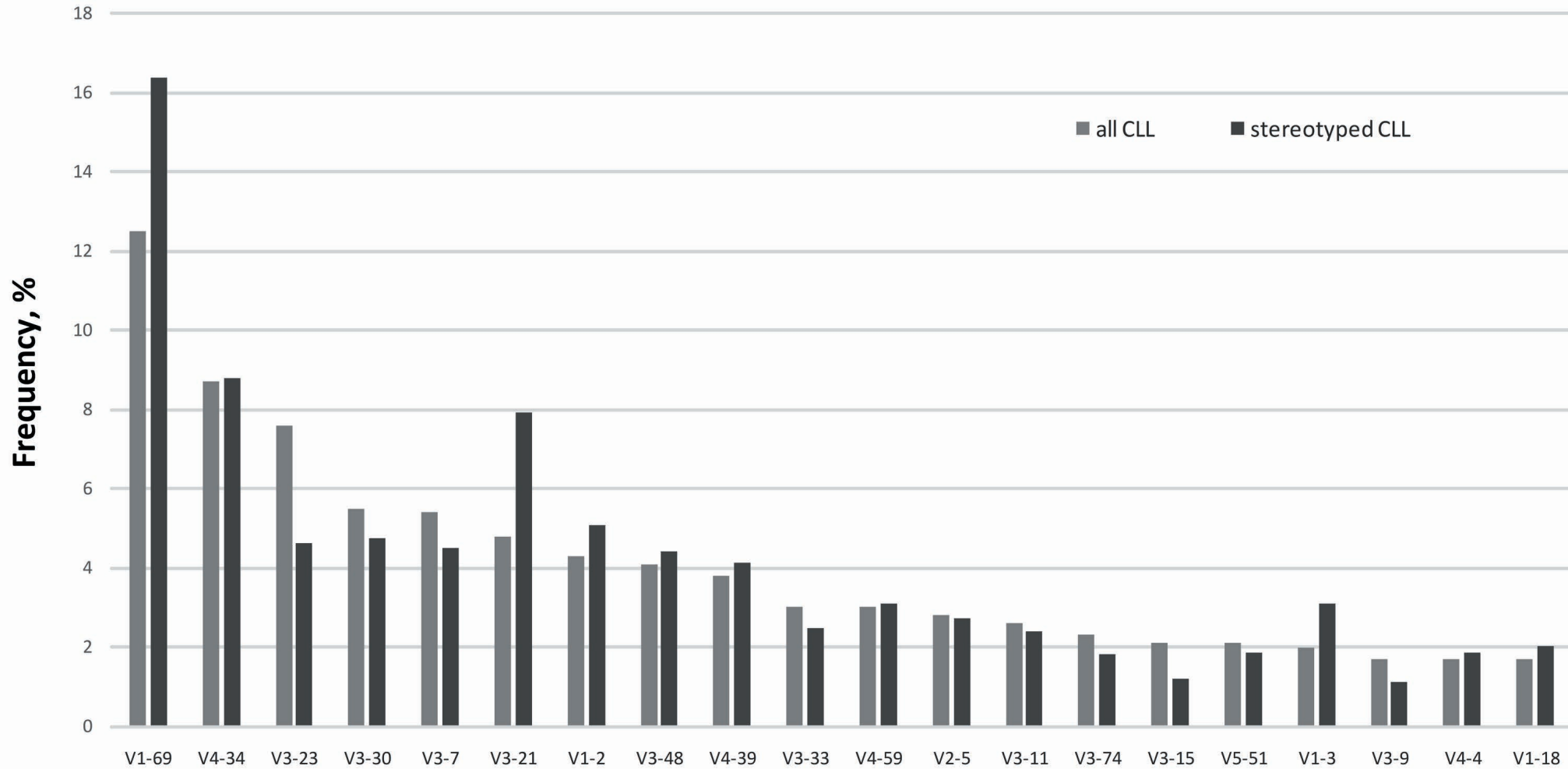
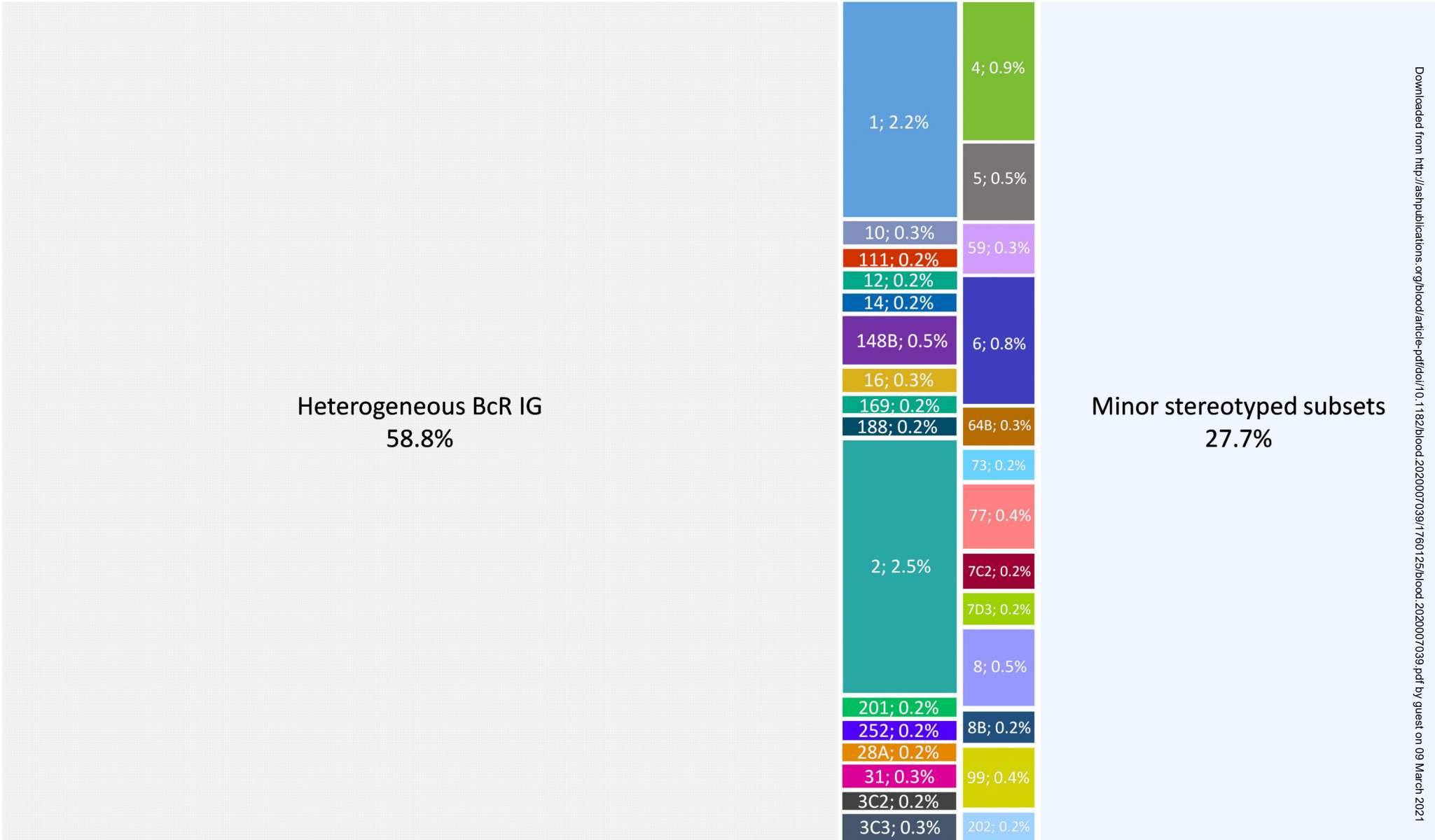
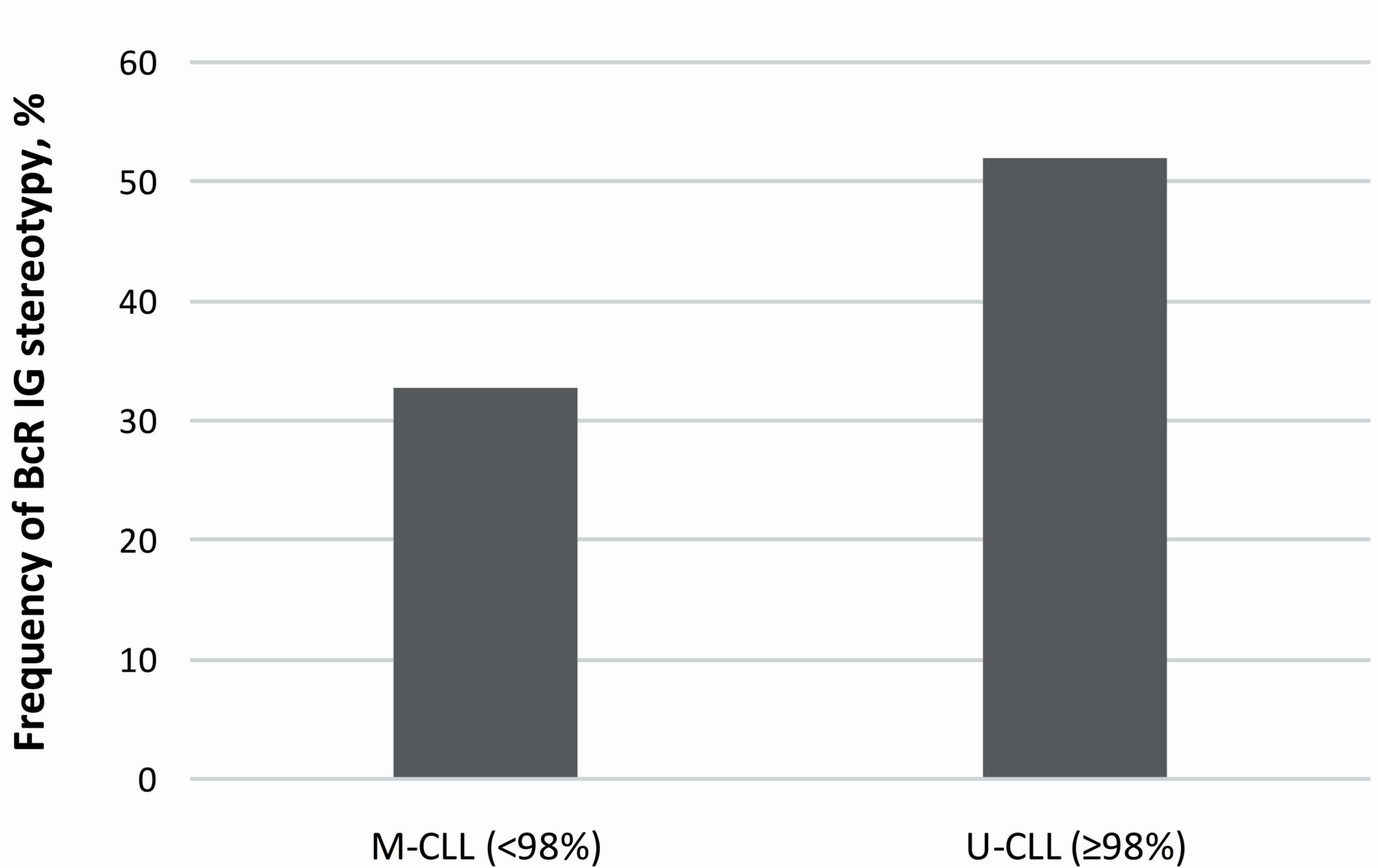


Figure 2



# Figure 3



# Figure 4

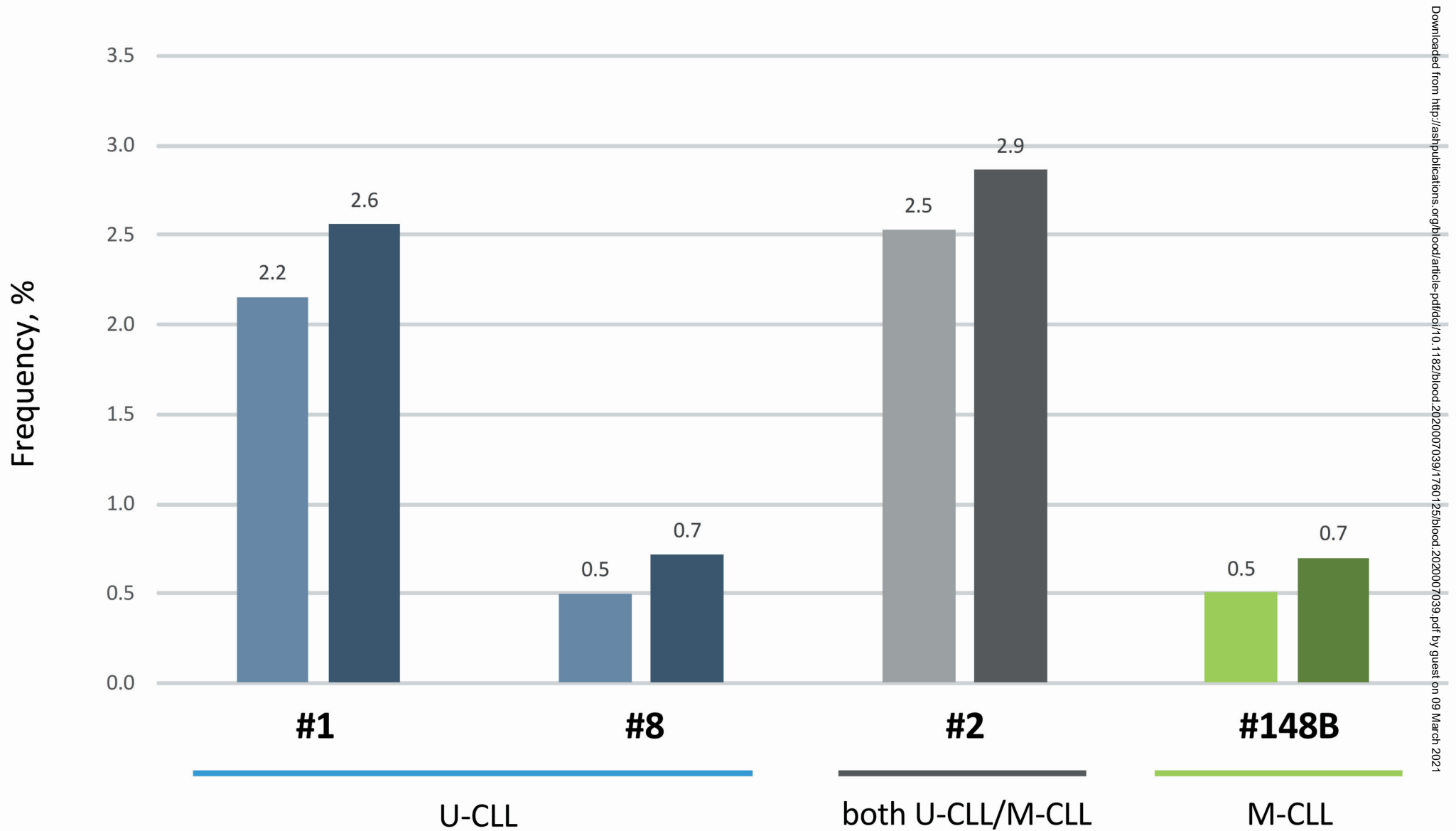
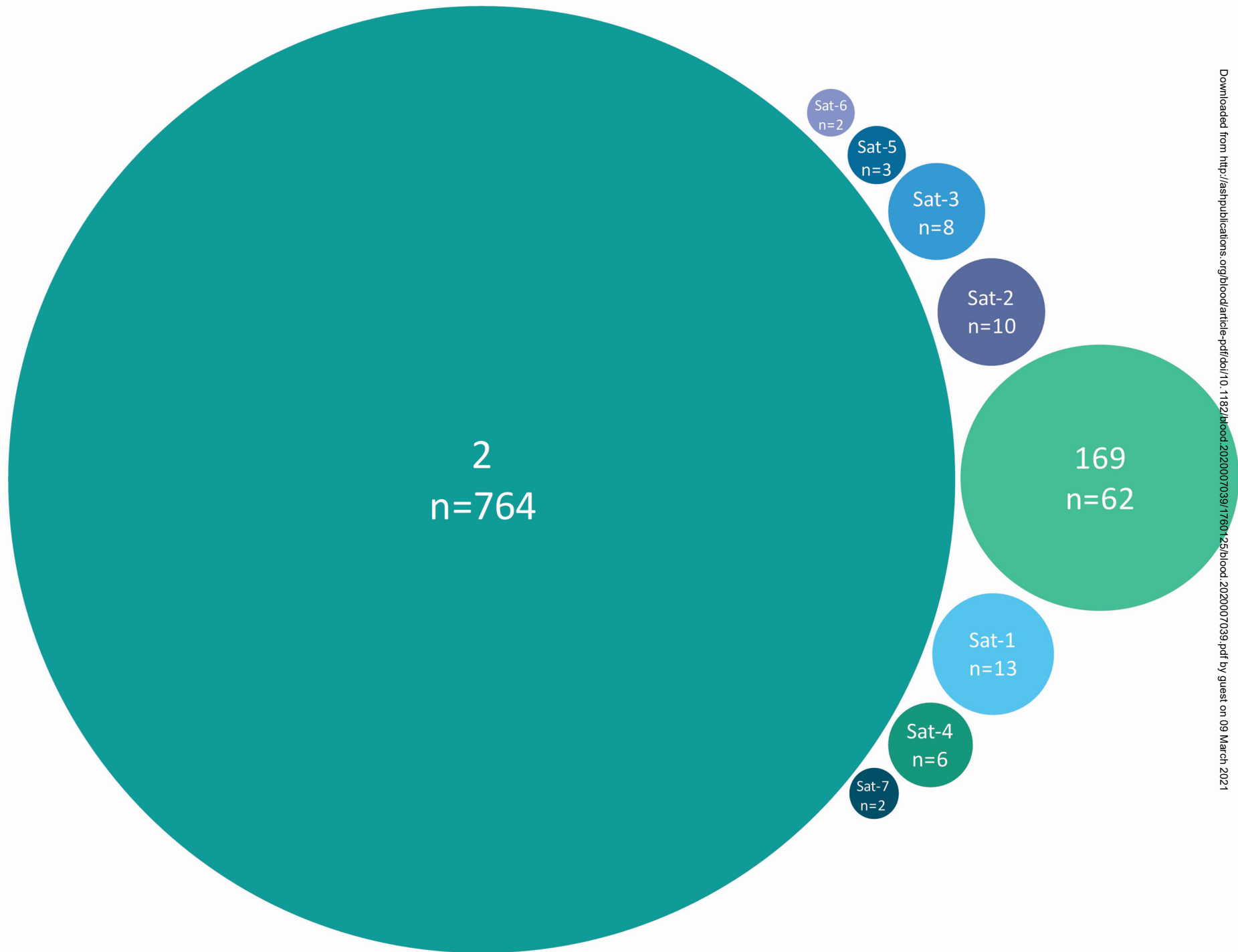
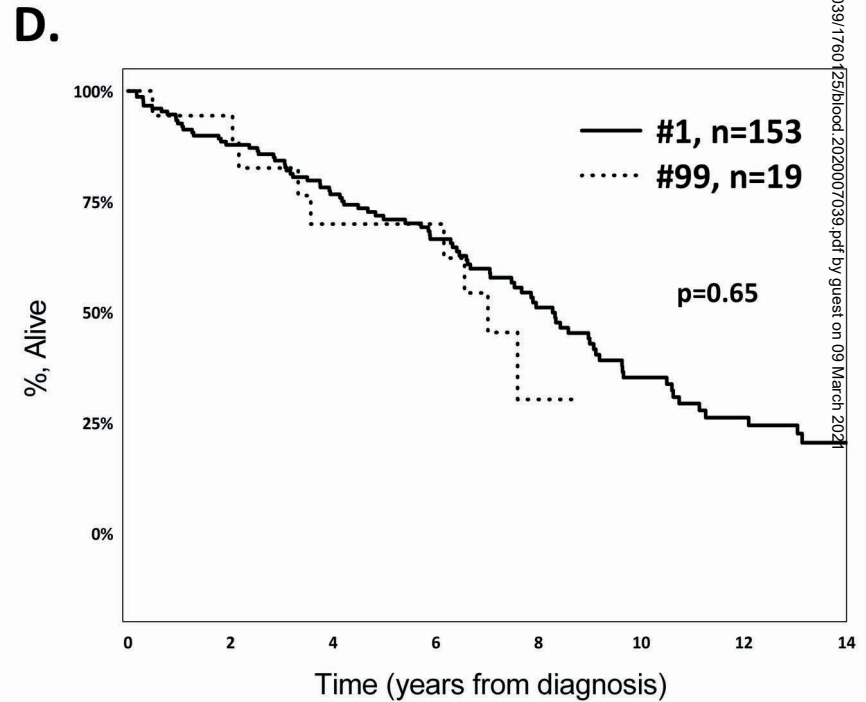
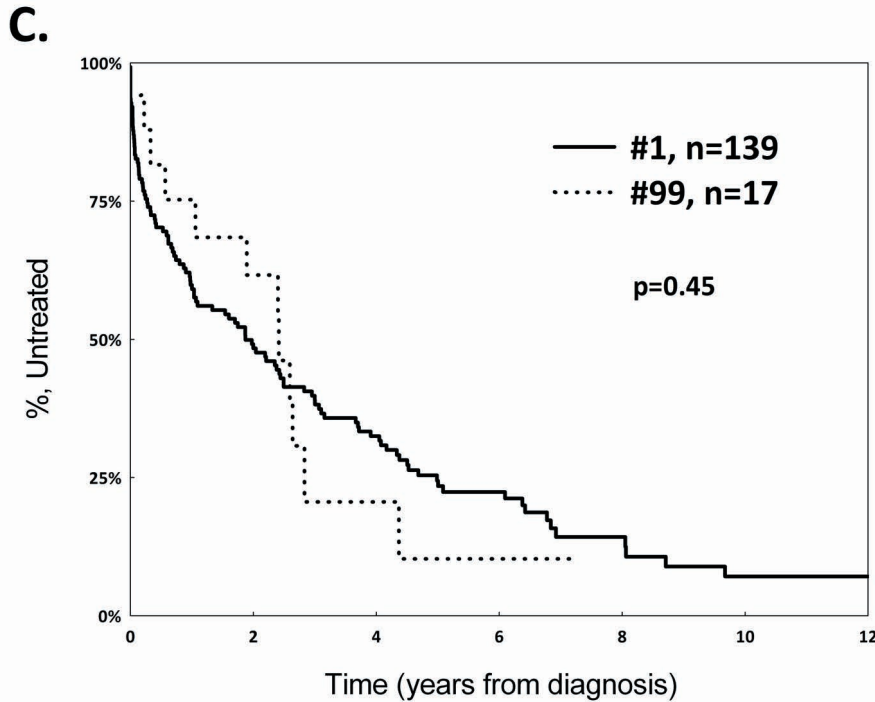
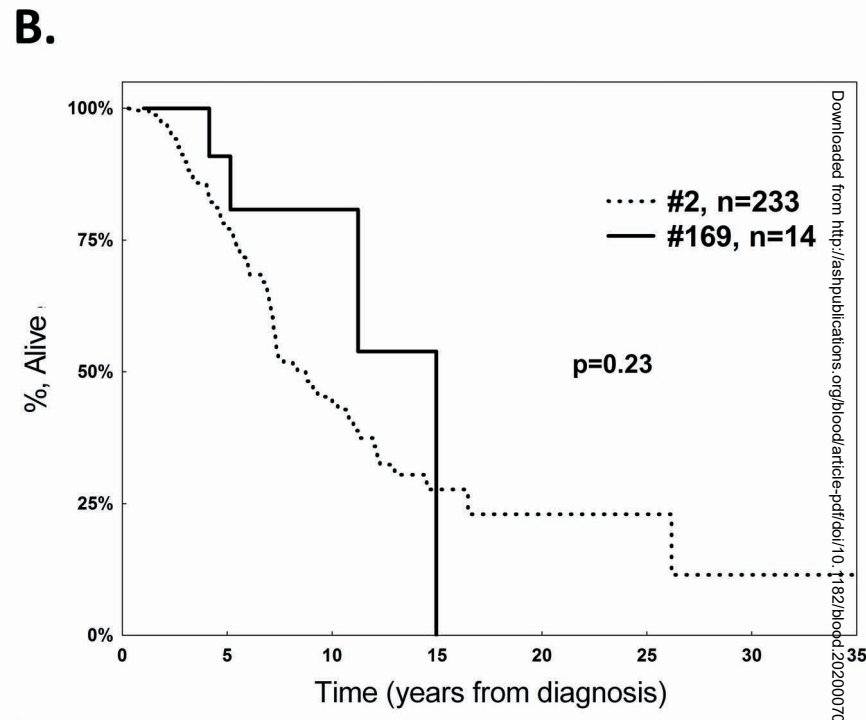
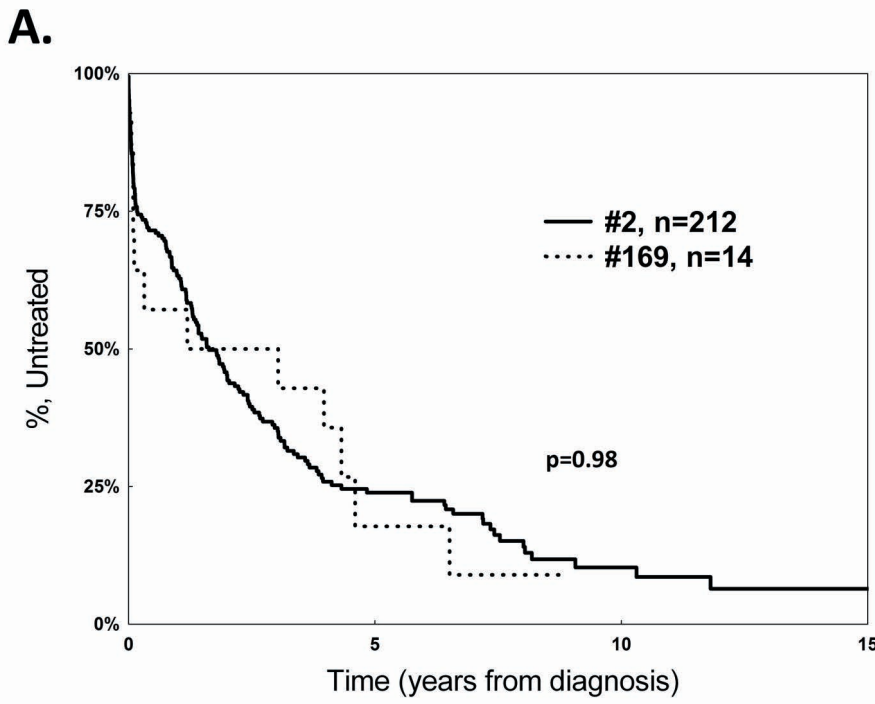


Figure 5



**Figure 6**



Downloaded from <http://ashpublications.org/blood/article-pdf/doi/10.1182/blood-2020070391760125> by guest on 09 March 2021