bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integrative analysis reveals aged clonal B cells, microenvironment and c-Myc activation in the origin of age-related lymphoma

Anastasia V. Shindyapina[1*], José P. Castro[1*$], Alessandro Barbieri[2], Olga S. Strelkova[3,4], João A. Paulo[5], Csaba Kerepesi[1], Anna P. Petrashen[6], Marco Mariotti[1], Margarita Meer[1], Yan Hu[1], Alexander Karamyshev[7], Grigoriy Losyev[1], Artur A. Indzhykulian[3,4], Steven P. Gygi[5], John M. Sedivy[6], John P. Manis[2], Vadim N. Gladyshev[1#]

[1] Brigham and Women's Hospital, Harvard Medical School, Boston, 02115, USA,
[2] Boston Children's Hospital, Harvard Medical School, Boston, 02115, USA,
[3] Eaton-Peabody Laboratories, Massachusetts Eye and Ear Infirmary, Boston 02114 USA,
[4] Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, 02114, USA,
[5] Department of Cell Biology, Harvard Medical School, Boston, 02115, USA,
[6] Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA
[7] UT College of Natural Sciences, The University of Texas at Austin, Austin, 78712, USA

* These authors contributed equally
[$] Current address: i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135, Porto, Portugal and Aging and Aneuploidy Laboratory, IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, 4200-135, Porto, Portugal.

[#] Corresponding author: vgladyshev@rics.bwh.harvard.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

While cancer is an age-related disease, most studies focus on genetically engineered younger mouse models. Here, we uncover how cancer develops as a consequence of the naturally aged immune system in mice. B-cell lymphoma frequently occurs in aged mice and is associated with increased cell size, splenomegaly, and a novel clonal B-cell population. Age-emergent B cells clonally expand outside of germinal centers driven by somatic mutations, activated c-Myc and hypermethylated promoters, and both genetically and epigenetically recapitulate human follicular and diffuse-large B-cell lymphomas. Mechanistically, mouse cancerous B cells originate from age-associated B cells, which are atypical memory B cells. Age-associated B cells secrete a spectrum of proinflammatory cytokines and activate paracrinally the expression of c-Myc in surrounding B cells. Although clonal B cells are a product of an aging microenvironment, they evolve being self-sufficient and support malignancy when transferred into young mice. Inhibition of mTOR and c-Myc attenuates premalignant changes in B cells during aging and emerges as a therapeutic strategy to delay the onset of age-related lymphoma. Together, we uncover how aging generates cancerous B cells, characterize a model that captures the origin of spontaneous cancer during aging and identify interventions that may postpone age-associated lymphoma.

Keywords

Aging; cancer; B-cell lymphoma; age-associated B cells; c-Myc; DNA methylation; DLBCL; follicular lymphoma; clonal hematopoiesis; clonal expansion; immune system aging; rapamycin.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The risk of developing invasive cancer grows exponentially with age starting from midlife[1]. Although cancer occasionally develops at a young age, the median age of its diagnosis in the US is 66 years old[2]. Mortality rates for cancer and aging follow the same trend consistent with the idea that cancer is a disease of aging[3]. It is also well established that both normal tissues and cancers accumulate age-related mutations during aging[4–6]. In particular, clonal hematopoiesis of indeterminate potential (CHIP) was described as a link between aging and cancer of immune cells[7,8].

However, aging has a potential to initiate cancer through a variety of non-genetic mechanisms: from epigenetic changes to microenvironment. One of the most potent oncogenes that is also implicated in the rate of aging is *c-Myc*[9]. Mice overexpressing *c-Myc* develop cancer[10–12], whereas *c-Myc* deficient mice are protected from lymphoma and are longer lived[13]. *c-Myc* also drives cell competition in healthy tissues[14–16]. Therefore, *c-Myc* may orchestrate age-related clonal expansion, constituting an early step in the transition to cancer. Aberrant DNA methylation is another suspect to initiate age-related cancer. Mouse knockout models of DNA methylation regulators, *Tet1/2/3* and *Dnmt3a*, develop myeloid leukemias or lymphomas[17–19]. CHIP in humans is predominantly driven by mutations in methylation regulators *TET2*, *DNMT3A*, and *ASXL1*[7,8]. Human cancers often exhibit global changes in DNA methylation[20]. At the same time, many human and mouse tissues accumulate recurrent DNA methylation changes with age[21,22], raising a possibility that aging promotes cancer through epigenetic changes alongside genetic aberrations.

Aging microenvironment is another major driver of cancer transformation[23], e.g. acting through chronic inflammation[24,25]. Age-associated B cells (ABC) are a prominent suspect to create a pro-inflammatory environment in lymphoid tissues as they overexpress pro-inflammatory cytokines such as IFNɣ, IL-6, and IL-4[26]. ABC are atypical memory B cells that accumulate in mammalian aging and disease[26–31] and express T-box transcription factor *TBX21*[28]. However, it remains unclear whether ABC contribute to age-related clonal selection and cancer.

Available information on mechanisms of age-related clonal expansion and cancer is currently largely limited to genetics due to lack of appropriate mouse models of spontaneous cancer. Commonly used C57BL/6 and BALB/c mice spontaneously develop and frequently die from B-cell lymphoma[32,33] offering a tool to investigate the full complexity of the origin of age-related cancer. Here, we extensively characterized spontaneous mouse age-related lymphoma and identified potential therapeutics against it. We further demonstrate that these systems can serve as translatable models to study mechanisms by which aging favors cancer in humans.

Results

B cells increase in size and undergo clonal expansions during mouse aging

We first examined age-related changes in the spleen and blood of young (6 months old) and old (27 months old) mice. Aged mice developed splenomegaly (Fig. 1a, Fig. S1a) due to a prominent increase in cell numbers (Fig. S1b). Additionally, we found that B cells (Fig. 1b, Fig.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S1f) and T cells (Fig. S1d), but not myeloid cells (Fig. S1e), profoundly increase in size with age.

To examine the etiology of enlarged B cells, we analyzed eight common populations of B cells (Fig. S1g,j,k). The proportions of follicular (FO) and marginal zone (MZ) B cells significantly decreased with age, whereas $CD21^{-}CD23^{-}$ B cells increased to 80% of B cells (Fig. 1f,h, Fig. S1i), consistent with previous findings[26]. Of all populations, splenic and blood $CD21^{-}CD23^{-}$ B cells showed the most robust increase in size (Fig. 1c), whereas the size and proportions of plasma cells and germinal center (Fig. S1h,i) B cells remained unchanged with age. Thus, accumulation of cells and increased $CD21^{-}CD23^{-}$ B-cell size largely account for the observed phenotype of enlarged B cells during aging.

Analysis of progenitor B cells (Fig. S1l) confirmed an age-related increase in the ratio of Pro- and IgM^{+} immature B cells to $B220^{+}$ cells and a drop in Pre-B cells (Fig. S1m)[34]. However, there was no increase in cell size (Fig. S1n) suggesting that B cells enlarge in the secondary lymphoid organs, e.g. the spleen. In addition to C57BL/6 mice, we examined BALB/c mice and found a similar spleen enlargement, an increase in average B-cell size and $CD21^{-}CD23^{-}$ B-cell size, and an increase in splenic cell number with age (Fig. S2a-f). Thus, increased B-cell size, in particular of $CD21^{-}CD23^{-}$ B cells, and splenomegaly are common features of mouse aging.

We next tested if enlarged B cells are premalignant. We sorted regular-sized B cells from spleens of young animals and regular-sized and enlarged B cells from old animals and sequenced variable CDR3 regions of Ig heavy chains by ImmunoSeq. We defined enlarged B cells as cells exceeding the size of young B cells based on forward scatter (FSC) in flow cytometry (Fig. 1d). Strikingly, we found that B cells develop large single clones with age representing on average 38% (Fig. S3a), and that clone size positively correlates with B-cell size ($R=0.74$, $p=0.0015$, Fig. 1e). Accordingly, CDR3 diversity profoundly decreases with age and with B-cell size increase (Fig. S3b). Clonal B cells collected from the same animal belonged to the same CDR3 clones independent of their cell size (Fig. 1f), suggesting that enlarged B cells originate from regular-sized B cells and clonally expand after VDJ rearrangement. Top clones did not carry somatic hypermutations within CDR3 regions (Fig. S3c). Together, these findings demonstrate that mouse B cells clonally expand and increase in size outside of germinal centers in the spleen during aging.

To investigate the molecular driver of B-cell clonal expansion we sequenced exomes, RNA and methylomes of the same set of samples. We first found that B cells isolated from old mice carry hundreds of somatic mutations across exomes (Fig. S3g) revealing that age-related B-cell clonal expansion relies on somatic mutations (Supplementary table 1). Only 5-10% of mutations were shared between regular and enlarged B cells, suggesting an independent path for clonal selection for enlarged and regular-sized B cells. Enlarged B cells formed larger clones as measured by variant allele fraction (VAF) (Fig. 1g). Driver somatic mutations were located in the genes commonly mutated in human lymphomas, e.g. *Trp53*, *Pim1* and *Myh11* (Fig. 1h), indicating conserved genetic mechanisms of clonal selection between human and mouse B cells. The top clones determined by CDR3 and exome sequencing were almost identical in size (Fig. S3d), consistent with the idea that the clones carrying driver mutations expand after VDJ rearrangements. B-cell size again significantly correlated with top genomic clone size (Fig. 1i), suggesting that B-cell size is a convenient marker of clonal B cells, especially when surface markers are unknown. Mutational signature deconvolution (Fig. S3f) revealed accumulation of

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SBS29 (73.8% of mutations), SBS42, SBS46, SBS15 and SBS5 (FIg. S3h, i) signatures in clonal B cells, as well as ID5 (Fig. S3j), DBS7, DBS11, and DBS5 (Fig. S3k), suggesting aging, impaired DNA repair, APOBEC and nucleotide modifications as sources of mutations in aging B cells. Despite large clonal expansion found in 12 mice, necropsy analysis revealed that only 3 mice had lymphoma (Fig. S4a). Thus, our analysis covered the premalignant state of B cells in addition to cancer itself.

We further assessed whether mouse premalignant B cells molecularly resemble human lymphomas. We examined gene expression patterns, promoter DNA methylation and somatic mutations in mouse clonal B cells against publicly available datasets of human blood cancer (Supplementary table 2) and aging. Clonal B cells from aged mice acquire highly similar epigenetic and genetic changes to those of human BCLs, and most strongly resemble diffuse large B-cell (DLBCL) and follicular lymphoma (FL) (Fig. 1j-l). Furthermore, analyses of spleen and blood sequences from The Genotype-Tissue Expression (GTEx)[35] project (Fig. S3i) revealed a decrease in Ig diversity with age (Fig. S3m,n) and with CHIP (Fig. S3o), as well as correlation between Ig and genomic clone sizes (Fig. S3p), similar to that in mice. Overall, the data suggest conservation of molecular mechanisms of B-cell malignancy during aging between mice and humans; thus, aged mice emerge as a first translational model to study age-related origin of cancer.

Age-associated clonal B cells (ACBC) represent a novel B-cell population with a characteristic premalignant epigenetic program

PCA analyses of gene expression, promoter methylation and protein levels (Fig. 2a) revealed that most of the variation between B-cell samples is explained by clonal expansion and enlargement of B cells. The biological age of clonal B cells[36], quantified by epigenetic aging clocks, significantly correlated with both clone size and B-cell size (Fig. 2b, Fig. S5b-c). Thus, clonal age-emergent B cells are substantially divergent at the molecular level and are epigenetically older than healthy age-matched B cells.

Clonal B cells further upregulated the expression of targets of *c-Myc*, central metabolic pathways, and G2/M checkpoint genes (Fig. S5f), as well as ribosome biogenesis and processing pathways (Fig. S5g-h), and rRNA was elevated in enlarged B cells (Fig. S5j). Upregulation of *c-Myc* targets and G2/M checkpoint was also validated at the protein level (Fig. S5f, Supplementary table 3). Targets of *c-Myc* were also consistently up-regulated in enlarged B cells (Fig. 2c), which were further validated in qRT-PCR (Fig. S5i). Altogether, our results strongly implicate *c-Myc* in driving clonal expansion and cell size increase of B cells with age.

While many genes became up-regulated by *c-Myc* in clonal B cells, many other genes were suppressed, and this pattern was consistent with the patterns of DNA methylation (Fig. S5d). In line with what is known about human cancer[20,37,38], we observed a genome-wide increase in promoter methylation in clonal mouse B cells from old mice (Fig. 2d). Some of the silenced genes were tumor suppressors, like *Id3* and *Mitf* (Fig. 2e). On the other hand, oncogene *Ptpn1* was up-regulated consistent with the loss of promoter methylation (Fig. S5e), as well as oxidative phosphorylation pathway genes (Fig. S5f). Our results strongly suggest that DNA methylation contributes to clonal selection and cancer transformation in aged tissues.

To understand the origin of clonal B cells we analyzed publicly available single-cell RNA-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seq datasets that previously unraveled two major clusters of B cells unique to aged spleens[39,40]. Both clusters were negative for CD21 (*Cr2*) and CD23 (*Fcer2a*) (Fig. S6a-b), and thus represented $CD21^{-}CD23^{-}$ B cells (Fig. 1c). We identified one cluster as ABC, based on the expression of the *Tbx21* and CD11c (*Itgax*)[28] (Fig. 2f, Fig. S6c). To test which B-cell population was clonal, we estimated the clonality signature score for each cell based on bulk RNA-seq of clonal B cells (Supplementary Table 3). The other cluster of age-related B cells had the highest scores among aged B cells (Fig. 2f,g). Reconstruction of CDR3 regions of Ig *kappa* chains in individual cells confirmed that clonal B cells belonged to one of the age-emergent B-cell clusters (Fig. 2h). We named these cells as age-associated clonal B cells (ACBC). We further confirmed significant up-regulation of *c-Myc* targets in clonal B cells at the single-cell level (Fig. 2j). Notably, the clonality signature score significantly correlated with Igk clone size in both datasets (Fig. S6d).

We observed that both age-related clusters of B cells shared many deregulated genes compared to the rest of splenic B cells (FO/MZ) (Fig. 2i). Moreover, ACBC and ABC shared CDR3 Igk sequences (Supplementary Table 4). Because ABC always co-occur with the ACBC population, but not the other way around, ACBC appear to originate from ABC, providing a mechanistic link between aging and cancer.

To differentiate ABC and ACBC populations with flow cytometry, we established CD29 as a positive marker for both populations and CD24a as a specific positive marker for ACBC (Fig. 4g). qRT-PCR confirmed that our panel differentiate ABC from ACBC, as FACS-sorted ABC cells ($CD19^{+}CD21^{-}CD23^{-}CD29^{+}CD24^{-}$) overexpress *Tbx21*, and ACBC overexpress *c-Myc* (Fig. S6h), which was further validated with immunofluorescence microscopy (Fig. 2m). Staining for new markers (Fig. S6e) confirmed that ACBC ($CD19^{+}CD21^{-}CD23^{-}CD29^{+}CD24^{+}$) accumulated with age in the spleen (Fig. S6f), but also in blood (data not provided), and that they are larger than follicular B cells (Fig.S6g), which were further validated with immunofluorescence microscopy (Fig. 2k,l). Circulation of ABC and ACBC can explain their infiltration in peripheral tissues (Fig. S4b-e).

We conclude that clonal B cells in old mice are predominantly represented by a novel ACBC population, and that *c-Myc* is a potential driver of B-cell clonal expansion and cell size increase with age. We further established a panel of markers to differentiate clonal ACBC from ABC using flow cytometry. Finally, we found a significant enrichment for the ACBC signature in GTEx subjects[47] with clonal B cells and enrichment for both mouse and human ABC signatures with aging (Fig. 2n). Thus, human subjects accumulate clonal B cells that are transcriptionally similar to the newly discovered mouse ACBC.

Clonal B cells are predictors of lifespan, originate with the support of an aging microenvironment and become self-sufficient over time

Longitudinal blood analysis (Fig. 3a) established that enlarged B cells accumulate in the mouse blood starting approximately at the age of 28 months and that myeloid bias started at about 25 months of age (Fig. S7a,b). The trajectory of blood age-related changes became more consistent when the mice were aligned by the death rather than age (Fig. 3b,c). This suggests a prominent role of the age-related blood composition changes in lifespan control. Indeed, mice with the higher CD21-CD23- and FSC scores had a 69-day shorter maximum lifespan (Fig. 3d).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To capture how cell composition changes during aging related to each other, we developed custom scores (see Methods) for B-cell enlargement (FSC Score), myeloid bias (Myeloid score) and CD21$^-$CD23$^-$ B-cell bias (CD21$^-$CD23$^-$ Score). The CD21$^-$CD23$^-$ score correlated with the myeloid score at early ages, and correlated with the FSC score at later ages (Fig. S7d). This suggests that myeloid cells antedate the formation of CD21$^-$CD23$^-$ B cells, and later CD21$^-$CD23$^-$ B cells predate the enlargement of B cells.

We next established causality of age-related blood changes *in vitro* (Fig. 3e). In agreement with longitudinal data, follicular B cells convert into CD21$^-$CD23$^-$ B cells in the presence of the cells from the aging spleen microenvironment (Fig. S7f), and ABC triggered *c-Myc* expression and cell size increase in FO B cells *in vitro* (Fig. 3f, S7e). Concentration of seventeen cytokines were increased in the media of ABC cells (Fig. 3g), five of which were confirmed by single-cell RNA-seq (Fig. 3h, Fig. S7g), and we named the combination of these cytokines a cancer-associated secretory phenotype (CASP) by analogy to SASP. In addition, our phosphoproteome profiling of clonal and enlarged B cells revealed activation of MAPK1 and p38a kinases (Supplementary Table 5), confirming that the extracellular stimuli from aging microenvironment are involved in age-related clonal selection and B-cell enlargement.

To test whether cancerous ACBC can survive outside of the aging environment we injected spleen extracts from mice diagnosed with lymphoma (and containing ACBC) into *Rag2*$^{-/-}$ mice (Fig. 3i). Most recipients of extracts containing cancerous ACBC experienced B-cell expansion with time, which were mostly represented by ACBC (Fig. 3j). Six mice perished as a result of the B-cell expansion (Fig. 3k).

Altogether, our data suggest that old follicular B cells turn into ABC cells under the influence of the aging microenvironment of the spleen. ABC secrete a spectrum of cytokines and activate *c-Myc* expression in surrounding B cells thus selecting out somatic mutations that support *c-Myc* overexpression. Indeed, the same genes with driver mutations in mouse lymphoma are commonly mutated in human lymphomas with *MYC* activation (Supplementary Table 6). That process gives rise to clonal *Myc*-overexpressing B cells, or ACBC (Fig. 3i), in line with the Adaptive Oncogenesis Hypothesis[41,42]. Over time ACBC evolve into self-sufficient malignant cells that thrive outside of the aging microenvironment.

c-Myc deficiency and rapamycin attenuate aging and clonal expansion of B cells

To test a causal role of *c-Myc* in the age-related clonal expansion, we analyzed *c-Myc* heterozygous mice (*Myc*$^{+/-}$, Fig. 4a)[13] and mice overexpressing *c-Myc* specifically in B cells (Eμ-*Myc*)[11]. We found that *c-Myc* haploinsufficiency (Fig. S8a) attenuated many age-related phenotypes of B cells: splenomegaly (Fig. S8b), increase in total B-cell size and the ACBC size (Fig. 4b,c) but not other B-cell populations (Fig, S8e,f), the accumulation of ACBC (Fig. 4d) and RNA accumulation (Fig. S8c), compared to age-matched wild type (Wt) siblings. Moreover, *Myc*$^{+/-}$ B cells preserved a more diverse Ig repertoire with age (Fig. 4e). Other immune cell types were not affected by *c-Myc* deficiency (Fig. S8d,j) suggesting a cell-type specific role of *c-Myc* for B-cell aging and malignancy. At the same time, B cells from Eμ-*Myc* mice had higher levels of RNA (Fig. S8k), were enlarged in the spleen (Fig. S8l) and bone marrow (Fig. S8m) and exhibited a reduced diversity of the Ig repertoire than non-carriers (Fig. S8n). Together, these data strongly suggest that *c-Myc* regulates cell size and fosters clonal expansion of B cells at

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

young and old ages. We further subjected mice to rapamycin as it was recently reported to inhibit *c-Myc* activity[43] and found that rapamycin attenuated cell size increase of ABC and ACBC (Fig. 4g) and accumulation of ABC (Fig. 4h) in 28-month-old mice.

RNA-seq demonstrated that clonal B cells from aged mice and Eμ-*Myc* B cells changed the same pathways, whereas *c-Myc*-deficient B cells were opposite to both (Fig. 4i). Genes that belong to *c-Myc* targets, metabolic and inflammation pathways cooperatively respond to changes in *c-Myc* level, and thus the majority of changes in the clonal B-cell transcriptome can be attributed to *c-Myc* overexpression. These data highlight the therapeutic potential of rapamycin and *c-Myc* deficiency against malignant transformation of aged B cells.

Discussion

To understand how cancer develops in the context of aging we studied commonly used mouse strains, which spontaneously develop B-cell lymphoma with age[32,33]. We found that mouse clonal B cells acquire genomic, gene expression and DNA methylation patterns of human lymphomas, in particular follicular and diffuse-large B-cell lymphomas. This unexpected result points to evolutionary conserved mechanisms of age-related B-cell malignancy at the genetic and epigenetic levels. Further studies may investigate if other spontaneous mouse cancers evolve similarly to human cancers with age.

We further discovered a population of age-associated clonal B cells - ACBC - that over time evolve into cancerous B cells. *c-Myc* emerged as a major driver of the B-cell clonal expansion with age as confirmed by transcriptome studies and mouse models. Our DNA methylation analyses revealed hypermethylated promoters in clonal B cells that silenced gene expression similarly to those in human cancer[20,37,38]. One of the silenced genes was *Id3,* which is a human and mouse tumor suppressor[44–47], providing an example of how age-related changes in DNA methylation can contribute to B-cell malignancy. At the same time, ACBC lack somatic hypermutations in CDR3 regions and thus do not undergo Ig selection in germinal centers. Together, these results suggest that the age-related DNA methylome, somatic mutations and activated *c-Myc* provide a selective advantage to B-cell clones rather than the canonical Ig-selection.

ABC shared the majority of deregulated genes, secretory phenotype and Ig clones with ACBC, suggesting that age-associated B cell is the cell of origin for clonal B cells in mice. ABC accumulate in elderly patients that are at higher risk of B-cell lymphomas, in HIV-positive patients susceptible to Burkitt lymphoma[48,49], and in patients with autoimmune diseases[29,31] susceptible to diffuse-large B-cell lymphoma[50]. Thus, it is likely that ABC give rise to ACBC and, consequently, B-cell lymphoma in humans too.

Although activation of *c-Myc* is a consequence of genetic perturbation in many cancers[51], there were no genetic alterations that can explain up-regulation of *c-Myc* in B cells with aging. Instead, exposure of follicular B cells to ABC was sufficient to elevate *c-Myc* expression, and we revealed a spectrum of cytokines associated with that. Future studies may test which component of the secretome of aging microenvironment triggers *c-Myc* expression in B cells thereby contributing to their malignancy. Despite ACBC arising from the aging microenvironment, some ACBC are transplantable into young immunodeficient mice, demonstrating that surrounding signals are essential to initiate rather than to support cancer.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, we demonstrated that two interventions, genetic and pharmacological, that potently extend mouse lifespan (*c-Myc* deficiency and rapamycin)[13,52] attenuate malignant phenotypes of B cells. Our longitudinal study further revealed that accumulation of clonal B cells stratifies aged mice into long- and short-lived groups. It is possible that *c-Myc* deficiency and rapamycin (and possibly other longevity interventions) extend mouse lifespan, at least in part, by delaying age-related lymphoma. Based on our data of conserved mechanisms of B-cell malignancy between human and mouse, mouse longevity interventions emerge as therapeutic strategies to delay B-cell malignancy in elderly patients.

Our work demonstrates the relevance of spontaneous age-related mouse cancers to human cancer and shows that aged mice is a translational model to study how aging favors cancer.

Experimental Procedures

Animals

Balb/c and C57BL6/JNia mice were obtained from the National Institute on Aging Aged Rodent Colony. Only female mice were used in the present study. 6-wk-old female Eμ-*Myc* mice and control non-carriers were purchased from The Jackson Laboratory (cat. #002728). Animals were euthanized with CO_2. Spleens were harvested and stored in cold PBS until analysis, and liver samples were immediately frozen in liquid nitrogen and stored at -80°C. For bone marrow analyses, bones were stripped of the muscle in cold PBS, cut from both ends, and cells were aspirated with 5 ml of cold PBS. The aspirates were filtered through 40 μm Falcon Cell Strainers (Corning). Spleens from 24-27 months old Myc haploinsufficient (*Myc* $^{+/-}$) mice and their wild type siblings were collected at Brown University and transferred on cold PBS to Harvard Medical School for further analysis. Mice were subjected to encapsulated rapamycin (Rapamycin Holdings. 126 ppm of active compound) or to encapsulating material of the same concentration (Rapamycin Holdings) in 5053 diet (TestDiet) and were fed *ad libitum*. For transplanted experiments, total splenocytes were cryopreserved in 10% DMSO FBS. Recovered splenocytes were counted and 2-6 mln of cells in 100 ul of sterile saline solution (Sigma) were i.p. Injected into 2-3 months old *Rag2*$^{-/-}$ mice. *Rag2*$^{-/-}$ mice were provided and housed by Manis laboratory at Karp Family Research Building of Boston Children's Hospital. All experiments using mice were performed in accordance with institutional guidelines for the use of laboratory animals and were approved by the Brigham and Women's Hospital and Harvard Medical School Institutional Animal Care and Use Committees.

Flow cytometry and sorting

mAbs used for staining included: anti-CD19 [6D5], anti-CD43 [S11], anti-IgM [RMM-1], anti-CD11b [M1/70], anti-B220, anti-CD80 [16-10A1], anti-CD3 [17A2], anti-Fas [SA367H8], anti-CD138 [281-2], anti-CD23 [B3B4], anti-CD21 [7E9], anti-CD24 [M1/69], anti-CD29 [HMβ1-1], anti-CD93, and anti-CD45 [30-F11] (all from Biolegend) with fluorophores as in Supplementary table 7. Dead cells were excluded by DAPI staining. Data was collected on a Cytek DXP11 and analyzed by FlowJo software (BD). Cells were sorted on BD Aria Fusion. Spleens were gently pressed between microscopy slides to get single-cell suspensions. One ml of cell suspensions were 1) incubated with 14 ml of red blood lysis buffer for 10 min on ice, 2) centrifuged at 4°C,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

250 g for 10 min, 3) washed once with 1 ml of FACS buffer (1% FBS in PBS), 4) stained in 100 ul of AB solution (2 ng/ul of each AB) at 4°C for 20 min protected from light, 5) washed again, 6) filtered into tubes with cell strainer snap cap (Corning), and 7) analyzed with flow cytometry or FACS-sorted into 1 ml of FACS buffer. Surface markers of ABCs and ACBC were selected based on: 1) differential up-regulation in both populations in scRNA-seq, 2) annotation as surface protein in GO, 3) differentially up-regulated in clonal B cells in bulk RNA-seq data (Supplementary table 3). It yielded a few candidates, of which we chose CD29 (*Itgb1*) as a positive marker for both age-related populations, and CD24 to distinguish ACBC from ABC.

RNA sequencing and qRT-PCR

Total RNA, DNA and proteins were extracted from fresh or snap-frozen FACS-sorted 2-5 million splenic B cells using AllPrep DNA/RNA/Protein Mini Kit (Qiagen) following the manufacturer's instructions. RNA was eluted with 42 ul of RNAse-free water. RNA concentration was measured with Qubit using the RNA HS Assay kit. Libraries were prepared with TruSeq Stranded mRNA LT Sample Prep Kit according to TruSeq Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E. Libraries had been quantified using the Bioanalyzer (Agilent), and were sequenced with Illumina NovaSeq6000 S4 (2x150bp) (reads trimmed to 2x100bp) to get 20M read depth coverage per sample. The BCL (base calls) binary were converted into FASTQ using the Illumina package bcl2fastq. Fastq files were mapped to mm10 (GRCm38.p6) mouse genome. and gene counts were obtained with STAR v2.7.2b[53]. Myc targets for Figure 2c were taken from the hallmark set of genes 'MYC_TARGETS_V1' from msigdb database. For quantitative RT-PCR, RNA samples were normalized by DNA concentration isolated from the same sample, then 2 ul of normalized RNA were mixed with iTaq Universal SYBR (Bio-Rad) and primers for *c-Myc* (F: TTCCTTTGGGCGTTGGAAAC, R: GCTGTACGGAGTCGTAGTCG), *Actb* (F: GGCTGTATTCCCCTCCATCG, R: CCAGTTGGTAACAATGCCATGT) or *Gapdh* (F: AGGTCGGTGTGAACGGATTTG, R: TGTAGACCATGTAGTTGAGGTCA) for the final volume of 10 ul, and loaded into Multiplate 96-Well PCR Plates (Bio-rad). Data was acquired for 40 cycles on Bio-Rad C1000, CFX96 Thermal Cycler. Each sample was loaded in duplicate.

Whole exome sequencing

Total DNA was extracted with AllPrep DNA/RNA/Protein Mini Kit (Qiagen) following the manufacturer's instructions. DNA was eluted with 100 ul of EB. DNA concentration was measured with Qubit using DNA BR Assay kit. Libraries were prepared with Agilent SureSelect XT Mouse All Exon according to SureSelectXT Target Enrichment System for Illumina Version B.2, April 2015. Libraries had been quantified using the Bioanalyzer (Agilent), and were sequenced with Illumina NovaSeq6000 S4 (2x150bp) (reads trimmed to 2x100bp) to get 100X throughput depth (roughly 50X on-target) coverage per sample. The BCL (base calls) binary were converted into FASTQ using the Illumina package bcl2fastq. Genomic reads were mapped to the GRCm38.p2 mouse genome assembly using BWA-MEM v0.7.15-r1140 (Li, 2013) and sorted using Samtools v1.6[54]. Somatic mutations were called with Mutect2 from the GATK package v4.1.8.0 using liver as matched control; high quality variants were selected with GATK FilterMutectCalls[55]. Variants were annotated with the Ensembl Variant Effect Predictor v100.2 (with option --everything)[56]. Only the variants that 1) were supported by 5 reads or more (DP.ALT>4), and 2) were in positions covered by 30 reads or more (DP.ALT+DP.REF>29),

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were taken for VAF analysis. Signatures of somatic mutations were extracted and deconvoluted using Sigproextractor v.1.0.20 with 100 nmf replicates.

Proteomics

Tandem mass tag (TMTpro) isobaric reagents were from ThermoFisher Scientific (Waltham, MA). Trypsin was purchased from Pierce Biotechnology (Rockford, IL) and LysC from Wako Chemicals (Richmond, VA). Samples were prepared as described previously[57,58]. Briefly, cell pellets were syringe-lysed in 8M urea complemented with protease and phosphatase inhibitors. Samples were reduced using 5mM TCEP for 30 min and alkylated with 10 mM iodoacetamide for 30 min. The excess of iodoacetamide was quenched with 10 mM DTT for 15 min. Protein was quantified using the BCA protein assay. Approximately 50 µg of protein were chloroform-methanol precipitated and reconstituted in 100 µL of 200 mM EPPS (pH 8.5). Protein was digested using Lys-C overnight at room temperature followed by trypsin for 6h at 37°C, both at a 100:1 protein:protease ratio. After digestion, the samples were labeled using the TMTpro16 reagents for 90 mins, the reactions were quenched using hydroxylamine (final concentration of 0.3% v/v). The samples were combined equally and subsequently desalted.

We enriched phosphopeptides from the pooled TMT-labeled mixtures using the Pierce High-Select Fe-NTA Phosphopeptide Enrichment kit ("mini-phos")[58,59] following manufacturer's instructions. The unbound fraction was retained and fractionated using basic pH reversed-phase (BPRP) HPLC. Ninety-six fractions were collected and then consolidated into 12 which were analyzed by LC-MS3[60].

All data were collected on an Orbitrap Fusion Lumos mass spectrometer coupled to a Proxeon NanoLC-1000 UHPLC. The peptides were separated using a 100 µm capillary column packed with ≈35 cm of Accucore 150 resin (2.6 µm, 150 Å; ThermoFisher Scientific). The mobile phase was 5% acetonitrile, 0.125% formic acid (A) and 95% acetonitrile, 0.125% formic acid (B). For BPRP fractions, the data were collected using a DDA-SPS-MS3 method with online real-time database searching (RTS)[61] to reduce ion interference[62,63]. Each fraction was eluted using a 90 min method over a gradient from 6% to 30% B. Peptides were ionized with a spray voltage of 2,600 kV. The instrument method included Orbitrap MS1 scans (resolution of 120,000; mass range 400−1400 m/z; automatic gain control (AGC) target 2x105, max injection time of 50 ms and ion trap MS2 scans (CID collision energy of 35%; AGC target 1x104; rapid scan mode; max injection time of 120 ms). RTS was enabled and quantitative SPS-MS3 scans (resolution of 50,000; AGC target 2.5x105; max injection time of 250 ms).

Phosphoproteomics

Phosphopeptides were analyzed with FAIMS/hrMS2 using our optimized workflow for multiplexed phosphorylation analysis[61,64,65]. Briefly, the Thermo FAIMSpro device was operated with default parameters (inner and outer electrodes were set at 100°C, yielding a FWHM between 10 V to 15 V and dispersion voltage (DV) was set at -5000 V). Each "mini-phos" was analyzed twice by the mass spectrometer using a 2.5h method having a gradient of 6% to 30%.

Raw files were first converted to mzXML. Database searching included all mouse entries from UniProt (downloaded March 2020). The database was concatenated with one composed of

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all protein sequences in the reversed order. Sequences of common contaminant proteins were also included. Searches were performed using a 50ppm precursor ion tolerance and 0.9 Da (low-resolution MS2) or 0.03 Da (high-resolution MS2) product ion tolerance. TMTpro on lysine residues and peptide N termini (+304.2071 Da) and carbamidomethylation of cysteine residues (+57.0215 Da) were set as static modifications, and oxidation of methionine residues (+15.9949 Da) was set as a variable modification. For phosphopeptide analysis, +79.9663 Da was set as a variable modification on serine, threonine, and tyrosine residues.

PSMs (peptide spectrum matches) were adjusted to a 1% false discovery rate (FDR)[66,67]. PSM filtering was performed using linear discriminant analysis (LDA) as described previously[68], while considering the following parameters: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy. Protein-level FDR was subsequently estimated. Phosphorylation site localization was determined using the AScore algorithm[69]. A threshold of 13 corresponded to 95% confidence that a given phosphorylation site was localized.

For reporter ion quantification, a 0.003 Da window around the theoretical m/z of each reporter ion was scanned, and the most intense m/z was used. Peptides were filtered to include only those with a summed signal-to-noise ratio ≥100 across all channels. For each protein, the filtered signal-to-noise values were summed to generate protein quantification values. To control for different total protein loading within an experiment, the summed protein quantities of each channel were adjusted to be equal in the experiment. For each protein in a TMTpro experiment, the signal-to-noise was scaled to sum to 100 to facilitate comparisons across experiments.

Spectral counts values were analyzed with R in Rstudio. Proteome and phosphoproteome data were normalized using the RLE method and log transformed using the *edgeR* package[71]. Values for phospho sites were normalized to corresponding protein level and differentially changed sites were calculated with the limma package. Ranked phospho sites were then assessed for enrichment for targets of mouse kinases using PTM-SEA resources[70] and kinact software[72].

Single-cell RNA sequencing analysis

To identify genes differentially expressed in this newly identified B cell cluster and other B cells, we downloaded the single cell RNA seq data from Calico's murine aging cell atlas (https://mca.research.calicolabs.com/data, spleen single-cell count data, filtered)[39]. Preprocessing of the downloaded data was performed using scanpy[73]. We first removed cells with a high (>0.05) percentage of mitochondrial reads. Cells not annotated as B cells were also removed. We then normalized the read counts by total reads number per cell and multiplied by a rescaling factor of 10000. Normalized reads were log transformed after adding a pseudo-count of 1. We scaled the log-transformed data to unit variance and zero mean and clipped maximum value to 10. After the above data preprocessing, we selected the cells corresponding to the age-related B-cell cluster, which was C130026I21Rik^{+}Apoe^{+}Cr2^{-}Fcer2a^{-}. For each cell, we used a linear combination of the RNA level of these four marker genes to calculate a score (score = C130026I21Rik + Apoe - Cr2 - Fcer2a), and within this cluster we selected the ABC cluster that is Tbx21^{+} and another one that is Myc^{+} using the same linear system. We then used

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a score threshold of 2.5 to select cells in the cluster of interest. Differential expression analysis was performed between the cell cluster of interest and all other remaining cells using the rank_genes_groups function in scanpy (Wilcoxon test). Details of the analysis can be found in our jupyter notebook B_cell_scRNAseq.ipynb. To reconstruct CDR3 regions of single cells, we demultiplexed bulk fastq files into single cell fastq files with *scruff* package in R[74], mapped each fastq to GRCm38.p6 genome and obtained gene counts with STAR v2.7.2b[53], and reconstructed CDR3 regions of Ig kappa chain from individual cells using mixcr[75]. Clone sizes were calculated as percent of templates supporting the current Ig kappa chain to the total number of reconstructed templates of Ig kappa chain for the sample. Cells with fewer than 200 gene counts were removed. Raw gene counts were log-transformed and normalized using the *edgeR* package in R[71]. Clonality signature score was calculated for each cell as transcript level of top 50 genes minus transcript level of bottom 50 genes ranked by p-value and taken from our bulk RNA-seq regressed against clone sizes (Supplementary Table 3). tSNEs were calculated using the *M3C* package in R[76].

Reconstruction of Ig CDR3 regions

Genomic CDR3 regions of Ig heavy chains were analyzed with Immunoseq (Adaptive Biotechnologies). CDR3 regions were reconstructed from RNA sequencing raw data using mixcr software[75] with recommended settings for transcriptome data. Filtering of reconstructed regions and diversity analysis was done with VDJtools software[77].

Immunofluorescence microscopy

B cells were FACS-sorted at 500 thousand cells per well and incubated with poly l-lysine treated coverslips for 1 hour in 24 well plates. Cells were permeabilized with 0.1% Triton X-100 2 times for 30 seconds, fixed in 3.7% PFA in PBS for 10 minutes and washed three times with PBS, incubated with the blocking buffer until further analysis (1% BSA, 0.1% Triton X-100 in PBS). Samples were incubated with primary antibodies overnight at 4 ℃ (1:100, Abcam #ab32072), then washed with PBS five times and incubated overnight with Alexa Fluor 568 - conjugated anti-rabbit secondary antibodies (1:500, Biotium cat#20098) and DAPI dye. Cells were washed with PBS and mounted with ProLong Diamond antifade (Thermo Fisher Scientific). Samples were imaged using Leica SP8 confocal microscope. Images were analyzed with ImageJ[78]. Raw Z-stacks were converted to the maximum intensity projection images. Nuclei and cell borders were detected using manual thresholding and “Analyze Particle” function. All crowded groups and not-round shaped cells were manually removed from the analysis.

Reduced representation bisulfite sequencing (RRBS)

Libraries were prepared and sequenced as in [22]. Bisulfite sequence reads were trimmed by TrimGalore v0.4.1 and mapped to the mouse genome sequence (mm10/GRCm38.p6) with Bismark v0.15.0 [79]. We kept CpG sites that were covered by five reads or more. Promoter regions were determined as the [-1500, +500] bp from the transcription start site (following the direction of the transcription) taken from Ensembl annotation file (Mus_musculus.GRCm38.100.chr.gtf). The start and end positions of gene bodies were taken from the Ensembl gene predictions (Mus_musculus.GRCm38.cds.all.fa). The mean methylation levels were calculated for regions that have at least 5 covered CpG sites with average

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methylation level above 1%. To determine ribosomal DNA methylation (rDNAm) age, we developed a blood rDNAm clock in a similar way as described in [80]. Briefly, we applied ElasticNet regression on ribosomal DNA (BK000964.3) CpG methylation levels of 153 control fed C57BL/6 blood samples with an age range from 0.67 to 35 months (GSE80672).

Longitudinal blood collection and analysis

Mice were anesthetized with isoflurane and then locally with topical anesthetic, restrained, and approximately 100 ul of blood was collected from mouse tails into EDTA-coated tubes (BD). Blood was incubated on ice until further analysis (2-3 hours), then mixed with 1 ml of red blood cell lysis buffer and centrifuged at 250g for 10 minutes at 4°C. Pellets were washed once with a FACS buffer (PBS with 1% FBS), split equally into 2 tubes and incubated for 20 minutes at 4°C with antibodies against B cell, T cell and myeloid cell markers, or follicular, marginal zone, and plasma cell markers. Stained cells were washed again, resuspended in 200 ul of FACS buffer and analyzed with FACS, with 20,000 events being recorded. Dead cells were gated by DAPI staining. Cell size was measured with forward scatter.

Longitudinal blood scores

We defined the FSC score as the difference between the last measurement of mean B cell size prior to death (if mouse died) or B cell size at the given round (if mouse was alive) and the mean B cell size of young mice at the same round. CD21.score was calculated as delta(CD21-CD23-) - delta(Follicular). Delta(CD21-CD23-) was calculated as percentage of $CD21^{-}CD23^{-}$ B cells of total B cells before death (if mouse died) or percentage of $CD21^{-}CD23^{-}$ B cells at the given round (if mouse was alive) minus percentage of $CD21^{-}CD23^{-}$ B cells in young mice at the same round. delta(Follicular) was calculated the same way for the percentage of follicular B cells of B cells. The myeloid score was calculated as delta(Myeloid) - delta(B cells). delta(Myeloid) was calculated as percentage of myeloid cells of $CD45^{+}$ cells before death (if mouse died) or percentage at the given round (if mouse was alive) minus the percentage of myeloid cells in young mice at the same round. delta(B cells) was calculated the same way for the percentage of B cells of CD45+ cells. A higher CD21.score indicates a higher proportion of $CD21^{-}CD23^{-}$ B cells to total B cells and/or lower proportion of follicular B cells to total B cells. A higher myeloid score indicates a higher proportion of myeloid cells to CD45+ cells and/or lower proportion of B cells to CD45+. A higher FSC score indicates a greater increase of B-cell size.

Cell culture

Freshly FACS-sorted cells were plated into 96 wells at 400,000 cells per well for each cell type and cultured for 24-48 hours in 200 ul of RPMI medium with 10% FCS (Gibco), 2 mM glutamine (ThermoFisher), 1% oxaloacetic acid (15 mg/ml), 5 mg/ml sodium pyruvate (ThermoFisher), 1% non-essential amino acids (ThermoFisher), and 50 µM 2-ME (Sigma). Where follicular B cells were incubated alone, 800,000 cells were plated. After co-incubation, cells were centrifuged at 250 g for 10 minutes at 4°C with slow deceleration. The media were carefully removed leaving ~50 ul, cells were washed once in 100 ul of FACS buffer, then resuspended in 100 ul of AB solution (1:100) and incubated at 4°C for 20 minutes protected from light, washed again and resuspended in 200 ul of FACS buffer, and filtered through 40 µm Falcon Cell Strainers (Corning). Follicular B cells (20-100 thousand) were gated as $CD19^{+}CD21^{int}CD23^{+}$ and FACS-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sorted into 300 ul of Trizol. RNA was extracted with Direct-zol RNA Microprep (Zymo Research).

Cytokine profiling

Thawed or fresh supernatants from cell culture experiments (see bove) were analyzed with Proteome Profiler Mouse Cytokine Array Kit, Panel A (R&D Systems) following manufacturer's protocol, 400-800 ul of supernatants pooled from three wells were analyzed. The exposure times were 1, 10 or 60 minutes depending on the intensity of the signal. Intensity of the signals was quantified using ImageJ. Individual intensities were normalized to mean intensity of all measures within the experiment to allow combining results from three independent experiments.

Necropsy analysis

Mice were euthanized with CO_2 followed by cervical dislocation. The chest and abdomen were opened, and the body was immersed into formalin solution and stored at 4°C until further analysis. For necropsy analysis all organs, including small endocrine organs, were dissected, trimmed at 5 mm thickness and embedded in paraffin blocks. Paraffin blocks were sectioned at 5 µm and stained with hematoxylin and eosin. The slides were examined blindly by a pathologist. Lymphoma was diagnosed when multiple solid tissues contained large uniform sheets of atypical lymphocytes with large nuclei. Lymphocytic hyperplasia was diagnosed when any solid tissue had small infiltrates of atypical lymphocytes with large nuclei.

Data analysis and availability

All data were analyzed and plotted with R in Rstudio. RNA sequencing, DNA methylation and proteomics data were preprocessed and analyzed for differential changes and GSEA with *limma*[81], *edgeR*[71] and *clusterprofiler*[82] packages. All p-values for group means comparisons were calculated with two-tailed Student t-test, unless otherwise specified. Correlations between two variables were evaluated using Pearson's correlation coefficient. PCA analysis was done with the factoextra package. Color schemes are from the ggsci package. Raw reads for RNA-sequencing, whole exome sequencing, and RRBS are available at SRA (PRJNA694093).

Acknowledgments

The authors thank members of the Gladyshev laboratory for discussion. Supported by Max Kade Foundation to JPC and by NIH grants to VNG.

Author Contributions

AVS and JPC conceived the project, designed and performed experiments, analyzed the data and drafted the manuscript. AB, OSS, JAP, CK, APP, MM, MM, and YH performed experiments, analyzed the data and revised the manuscript. GL, SPG, and JMS provided research materials, assisted with experimental design and revised the manuscript. JPM and VNG interpreted the data, designed experiments, provided research materials, and revised the manuscript. VNG supervised the overall project.

Declaration of Interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Literature cited

1. Niccoli, T. & Partridge, L. Ageing as a Risk Factor for Disease. *Curr. Biol.* **22**, R741–R752 (2012).
2. Howlader, N. *et al.* SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. *Bethesda MD April* (2020).
3. Zenin, A. *et al.* Identification of 12 genetic loci associated with human healthspan. *Commun. Biol.* **2**, 41 (2019).
4. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
5. Yizhak, K. *et al.* RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* **364**, (2019).
6. Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
7. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
8. Jaiswal, S. Clonal hematopoiesis and non-hematologic disorders. *Blood* blood.2019000989 (2020) doi:10.1182/blood.2019000989.
9. Gabay, M., Li, Y. & Felsher, D. W. MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb. Perspect. Med.* **4**, (2014).
10. Adams, J. M. *et al.* The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* **318**, 533–538 (1985).
11. Harris, A. W. *et al.* The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells. *J. Exp. Med.* **167**, 353–371 (1988).
12. Lefebure, M. *et al.* Genomic characterisation of Eμ-Myc mouse lymphomas identifies Bcor as a Myc co-operative tumour-suppressor gene. *Nat. Commun.* **8**, 14581 (2017).
13. Hofmann, J. W. *et al.* Reduced expression of MYC increases longevity and enhances healthspan. *Cell* **160**, 477–488 (2015).
14. Simpson, P. & Morata, G. Differential mitotic rates and patterns of growth in compartments in the Drosophila wing. *Dev. Biol.* **85**, 299–308 (1981).
15. Clavería, C., Giovinazzo, G., Sierra, R. & Torres, M. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500**, 39–44 (2013).
16. Muñoz-Martín, N., Sierra, R., Schimmang, T., Villa Del Campo, C. & Torres, M. Myc is dispensable for cardiomyocyte development but rescues Mycn-deficient hearts through functional replacement and cell competition. *Dev. Camb. Engl.* **146**, (2019).
17. Mayle, A. *et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* **125**, 629–638 (2015).
18. An, J. *et al.* Acute loss of TET function results in aggressive myeloid cancer in mice. *Nat. Commun.* **6**, 10071 (2015).
19. Wang, Y. *et al.* Tet2-mediated clonal hematopoiesis in nonconditioned mice accelerates age-associated cardiac dysfunction. *JCI Insight* **5**, (2020).
20. Saghafinia, S., Mina, M., Riggi, N., Hanahan, D. & Ciriello, G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* **25**, 1066-1080.e8 (2018).
21. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
22. Meer, M. V., Podolskiy, D. I., Tyshkovskiy, A. & Gladyshev, V. N. A whole lifespan mouse multi-tissue DNA methylation clock. *eLife* **7**, (2018).
23. Fane, M. & Weeraratna, A. T. How the ageing microenvironment influences tumour progression. *Nat. Rev. Cancer* **20**, 89–106 (2020).
24. Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899 (2010).
26. Hao, Y., O'Neill, P., Naradikian, M. S., Scholz, J. L. & Cancro, M. P. A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood* **118**, 1294–1304 (2011).
27. Zheng, Y. *et al.* A human circulating immune cell landscape in aging and COVID-19. *Protein Cell* **11**, 740–770 (2020).
28. Rubtsova, K., Rubtsov, A. V., van Dyk, L. F., Kappler, J. W. & Marrack, P. T-box transcription factor T-bet, a key player in a unique type of B-cell activation essential for effective viral clearance. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3216-3224 (2013).
29. Rubtsov, A. V. *et al.* Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c$^+$ B-cell population is important for the development of autoimmunity. *Blood* **118**, 1305–1315 (2011).
30. Moir, S. *et al.* Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. *J. Exp. Med.* **205**, 1797–1805 (2008).
31. Wu, C. *et al.* Lupus-associated atypical memory B cells are mTORC1-hyperactivated and functionally dysregulated. *Ann. Rheum. Dis.* **78**, 1090–1100 (2019).
32. Brayton, C. F., Treuting, P. M. & Ward, J. M. Pathobiology of aging mice and GEM: background strains and experimental design. *Vet. Pathol.* **49**, 85–105 (2012).
33. Turturro, A., Duffy, P., Hass, B., Kodell, R. & Hart, R. Survival characteristics and age-adjusted disease incidences in C57BL/6 mice fed a commonly used cereal-based diet modulated by dietary restriction. *J. Gerontol. A. Biol. Sci. Med. Sci.* **57**, B379-389 (2002).
34. Labrie, J. E., Sah, A. P., Allman, D. M., Cancro, M. P. & Gerstein, R. M. Bone marrow microenvironmental changes underlie reduced RAG-mediated recombination and B cell generation in aged mice. *J. Exp. Med.* **200**, 411–423 (2004).
35. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
36. Petkovich, D. A. *et al.* Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab.* **25**, 954-960.e6 (2017).
37. Esteller, M., Corn, P. G., Baylin, S. B. & Herman, J. G. A gene hypermethylation profile of human cancer. *Cancer Res.* **61**, 3225–3229 (2001).
38. Herman, J. G. & Baylin, S. B. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* **349**, 2042–2054 (2003).
39. Kimmel, J. C. *et al.* Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res.* **29**, 2088–2103 (2019).
40. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
41. Henry, C. J., Marusyk, A. & DeGregori, J. Aging-associated changes in hematopoiesis and leukemogenesis: what's the connection? *Aging* **3**, 643–656 (2011).
42. Marusyk, A. & DeGregori, J. Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. *Biochim. Biophys. Acta* **1785**, 1–11 (2008).
43. Sun, L. *et al.* Rapamycin targets STAT3 and impacts c-Myc to suppress tumor growth. *Cell Chem. Biol.* S2451945621004402 (2021) doi:10.1016/j.chembiol.2021.10.006.
44. Greenough, A. & Dave, S. S. New clues to the molecular pathogenesis of Burkitt lymphoma revealed through next-generation sequencing. *Curr. Opin. Hematol.* **21**, 326–332 (2014).
45. Richter, J. *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* **44**, 1316–1320 (2012).
46. Love, C., Remple, R., Li, G. & Dave, S. ID3 Loss In Vivo Synergizes with MYC-Driven Oncogenesis in Burkitt Lymphoma. *Blood* **126**, 3906–3906 (2015).
47. Chen, S. *et al.* Id3 Orchestrates Germinal Center B Cell Development. *Mol. Cell. Biol.* **36**,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2543–2552 (2016).
48. Atallah-Yunes, S. A., Murphy, D. J. & Noy, A. HIV-associated Burkitt lymphoma. *Lancet Haematol.* **7**, e594–e600 (2020).
49. Knox, J. J. *et al.* T-bet+ B cells are induced by human viral infections and dominate the HIV gp140 response. *JCI Insight* **2**, (2017).
50. Zintzaras, E., Voulgarelis, M. & Moutsopoulos, H. M. The risk of lymphoma development in autoimmune diseases: a meta-analysis. *Arch. Intern. Med.* **165**, 2337–2344 (2005).
51. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8**, 976–990 (2008).
52. Miller, R. A. *et al.* Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. *Aging Cell* **13**, 468–477 (2014).
53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
55. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).
56. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
57. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, 399–404 (2020).
58. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. *J. Proteome Res.* **17**, 2226–2236 (2018).
59. Paulo, J. A., Navarrete-Perea, J., Erickson, A. R., Knott, J. & Gygi, S. P. An Internal Standard for Assessing Phosphopeptide Recovery from Metal Ion/Oxide Enrichment Strategies. *J. Am. Soc. Mass Spectrom.* **29**, 1505–1511 (2018).
60. Paulo, J. A. *et al.* Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 S. cerevisiae proteins across 10 carbon sources. *J. Proteomics* **148**, 85–93 (2016).
61. Schweppe, D. K., Rusin, S. F., Gygi, S. P. & Paulo, J. A. Optimized Workflow for Multiplexed Phosphorylation Analysis of TMT-Labeled Peptides Using High-Field Asymmetric Waveform Ion Mobility Spectrometry. *J. Proteome Res.* **19**, 554–560 (2020).
62. Gygi, J. P. *et al.* Web-Based Search Tool for Visualizing Instrument Performance Using the Triple Knockout (TKO) Proteome Standard. *J. Proteome Res.* **18**, 687–693 (2019).
63. Paulo, J. A., O'Connell, J. D. & Gygi, S. P. A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments. *J. Am. Soc. Mass Spectrom.* **27**, 1620–1625 (2016).
64. Pfammatter, S., Bonneil, E., McManus, F. P. & Thibault, P. Accurate Quantitative Proteomic Analyses Using Metabolic Labeling and High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS). *J. Proteome Res.* **18**, 2129–2138 (2019).
65. Saba, J., Bonneil, E., Pomiès, C., Eng, K. & Thibault, P. Enhanced sensitivity in proteomics experiments using FAIMS coupled with a hybrid linear ion trap/Orbitrap mass spectrometer. *J. Proteome Res.* **8**, 3355–3366 (2009).
66. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol. Clifton NJ* **604**, 55–71 (2010).
67. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).
68. McAlister, G. C. *et al.* Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

69. Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292 (2006).
70. Krug, K. *et al.* A Curated Resource for Phosphosite-specific Signature Analysis. *Mol. Cell. Proteomics MCP* **18**, 576–593 (2019).
71. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
72. Wirbel, J., Cutillas, P. & Saez-Rodriguez, J. Phosphoproteomics-Based Profiling of Kinase Activities in Cancer Cells. *Methods Mol. Biol. Clifton NJ* **1711**, 103–132 (2018).
73. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
74. Wang, Z., Hu, J., Johnson, W. E. & Campbell, J. D. scruff: an R/Bioconductor package for preprocessing single-cell RNA-sequencing data. *BMC Bioinformatics* **20**, 222 (2019).
75. Bolotin, D. A. *et al.* Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* **35**, 908–911 (2017).
76. John, C. R. *et al.* M3C: Monte Carlo reference-based consensus clustering. *bioRxiv* 377002 (2019) doi:10.1101/377002.
77. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
78. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).
79. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinforma. Oxf. Engl.* **27**, 1571–1572 (2011).
80. Wang, M. & Lemos, B. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res.* **29**, 325–333 (2019).
81. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
82. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1. B cells increase in size and undergo clonal expansion driven by somatic mutations during aging. (a) Spleens from young and old C57BL/6 mice. (b) Mean B-cell size measured by FACS in young and old mice. (c) Mean $CD21^-CD23^-$ B-cell size in young and old mice measured by FACS. (d) Scheme explaining how young, old regular sized and old enlarged B cells were gated and analyzed. (e) Correlation between B-cell size and size of the top B-cell CDR3 clone. (f) Circular plot with line width corresponding to frequency of top clones. Lines connect CDR3 clones with the same amino acid sequences. (g) VAF of somatic mutations detected from whole exome sequences of old and old enlarged B cells (n=7 each). Lines connect variants that were present in old and old enlarged B cells from the same animal. (h) Manhattan plot for somatic mutations. Labeled genes are these that carry mutations with VAF over 0.4. (i) Correlation between B-cell size and top clone size estimated by VAF. (j) Similarity between the transcriptomes of human blood cancers and mouse clonal B cells. (k) GSEA of genes with differentially methylated promoters in mouse clonal B cells against these in human blood cancers. (l) GSEA of genes with somatic mutations in mouse clonal B cells (ranked by VAF) against lists of genes mutated in human blood cancers.

Figure 2. Aged B cells mimic the transcriptional and epigenetic programs of malignant cells, and single-cell RNA-seq reveals them as a novel B-cell population. (a) PCA plots of omics of young, old and old enlarged B cells. (b) Correlation between top Ig clone size estimated by ImmunoSeq and the rDNA methylation age of B cells. (c) Gene expression levels of *c-Myc* targets in young, old and old enlarged B cells. (d) Correlation of top Ig clone size and mean global DNA methylation of promoters and gene bodies in B cells of old mice. (e) Correlation between mean promoter DNA methylation and gene expression of tumor suppressor genes, *Id3* and *Mitf*. (f) Single-cell RNA-seq of 24-month-old mouse spleen colored according to clonality signature score. Dot size is the size of the Ig clone. (g) Clonality signature scores for ABCs, ACBC and the rest of B cells (FO/MZ). (h) Clone sizes to which each B cell (FO/MZ), ABC or ACBC belong. P-value was calculated with the Wilcoxon signed-rank test. (i) GSEA for genes that are differentially expressed in B cells from the top Ig clone compared to the B cells that belong to smallest clones. (j) Overlap between up-regulated and down-regulated genes (FDR<0.05) in ABC and ACBC compared to B cells (FO/MZ) in scRNA-seq dataset. (k) Representative images of follicular and ACBC cells immunolabeled for *c-Myc* and for DNA with DAPI. Scale bar, 10 µm. (l) Cell sizes and (m) mean *c-Myc* intensity for follicular and ACBC estimated from confocal images for two independent experiments. (n) GSEA for mouse ACBC and ABC, and human ABC signatures in blood transcriptomes of GTEx subjects with clonal B cells or T cells, and age-related gene expression changes. **** p-value < 0.0001.

Figure 3. Longitudinal analysis reveals B-cell size and $CD21^-CD23^-$ B cells as predictors of lifespan, and co-culture experiments point to their origin. (a) Blood was collected from tails of C57BL/6 female mice across seven timepoints. (b) Changes in blood cell composition in old mice aligned to the last blood collection before animal death (e.g., -1 is one round before the last blood collection). (c) Same as (b) but for B-cell populations. (d) Survival of mice with higher and lower sum of $CD21^-CD23^-$ and FSC scores in 27-month-old mice. (e) Scheme for co-culture

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

experiment. (f) *c-Myc* fold change normalized to total DNA level. Results are from two independent experiments, 3-4 biological replicas per experiment, 2 technical replicas per biological replica. (g) Protein levels of cytokines measured from cell culture supernatants of old follicular B cells and ABC. Results are from three independent experiments, 2 biological replicas per experiment. (h) Single-cell gene expression of cytokines in B cells. Secretome is a mean gene expression level of five cytokines shown. P-values were calculated with the Wilcoxon test. (i) Injection of splenocytes from old mice diagnosed with lymphoma or healthy old mice (Old control) into *Rag2*$^{-/-}$ mice. (j) Percentage of B-cell populations to total live PBMC measured in blood of *Rag2*$^{-/-}$ mice post injection or left without injections (Negative Control). (k) Survival of *Rag2*$^{-/-}$ mice post injections. P-values for survival curves were calculated with a log-rank test.

Figure 4. Myc deficiency and rapamycin treatment inhibits cell size increase and clonal expansion. (a) Analysis of Wt vs *Myc*$^{+/-}$ mice. (b) Mean splenic B-cell size of Wt and *Myc*$^{+/-}$ mice. (c) Mean splenic ACBC cell size of Wt and *Myc*$^{+/-}$ mice. (d) Percentages of ACBC to total B cells (CD19^{+}) in 26-month-old Wt and *Myc*$^{+/-}$ mice. (e) Ig diversity (chaoE_mean) of splenic B cells from 26-27 months old Wt and *Myc*$^{+/-}$ mice. (f) Analysis of the effect of rapamycin on B-cell malignant phenotypes. (g) Mean ACBC cell size of Eudragit- or Rapamycin-treated mice. (h) Percentage of ABC and ACBC to total splenic B cells (CD19^{+}) in Eudragit- and Rapamycin-treated^{-} mice. (i) GSEA for differential gene expressions: 1) correlated with top B cell clone size in old C57BL/6 mice, 2) correlated with top clone size in B cells of young Eμ-*Myc* mice, and 3) changed in B cells of old *Myc*$^{+/-}$ mice compared to their wild type siblings. P-values were calculated with a one-sided Student t-test. (j) Proposed mechanism for the development of ACBC and age-related B-cell lymphoma. Wt - wild type.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure Legends

Supplementary Figure 1. Age-related changes in cell sizes and immune cell composition in C57BL/6 mice. (a) Quantification of spleen weights normalized to body weight of young (6 months old) and old (27 months old) C57BL/6 mice (b) Cell number (log10) from young and old mouse spleens excluding red blood cells. (c) Cell proportions of immune ($CD45^+$) cells in spleens of young and old mice measured by FACS. Myeloid cells were gated as $CD45^+CD3^-CD11b^+CD19^-$, B cells as $CD45^+CD11b^-CD3^-CD19^+$, T cells as $CD45^+CD3^+CD11b^-CD19^-$, and $CD11b^+CD19^+$ cells as $CD45^+CD11b^+CD3\text{-}CD19^+$. (d) Sizes of T cells and (e) myeloid cells measured by FACS and gated as described above. Each point represents the mean cell size in mice. (f) Gating strategy for splenic B cells ($CD45^+CD11b^-CD3^-CD19^+$) and representative FACS. (g) Gating strategy for marginal zone, follicular and $CD21^-CD23^-$ B cells and representative FACS. (h) Proportions of B-cell populations in spleens of young and old mice measured by FACS. $CD21^-CD23^-$ B cells were gated as $CD19^+CD21^-CD23^-$, follicular B cells as $CD19^+CD21^{int}CD23^+$, germinal center B cells as $CD19^+Fas^+$, marginal zone B cells as $CD19^+CD21^+CD23^{low}$, and plasma cells as $CD19^+CD138^+$. (i) Sizes of follicular, germinal center, marginal zone, and plasma B cells measured by FACS. Each point represents the mean cell size in mice. (j) Gating strategy for germinal center and (k) plasma cells and representative FACS. (l) Gating strategy for Pro-B cells, Pre-B cells and IgM^+ B cells in the bone marrow and representative FACS. (m) Percentage of Pro-B cells, Pre-B cells and IgM^+ B cells in the bone marrow from young and old mice. (n) Cell size (FSC, in thousands) of Pro-B cells, Pre-B cells and IgM^+ B cells in the bone marrow of young and old mice measured by FACS.

Supplementary Figure 2. Age-related changes in cell sizes and immune cell composition in BALB/c mice. (a) Spleen weight of young (6 months old) and old (27 months old) BALB/c female mice. (b) Cell composition of lymphoid (CD45+) cells in spleens of young and old BALB/c mice measured by FACS. Myeloid cells were gated as $CD45^+CD3^-CD11b^+CD19^-$, B cells as $CD45^+CD11b^-CD3^-CD19^+$, T cells as $CD45^+CD3^+CD11b^-CD19^-$, and $CD11b^+CD19^+$ cells as $CD45^+CD11b^+CD3\text{-}CD19^+$. (c) Size of B cells, T cells, myeloid cells and $CD19^+CD11b^+$ cells from BALB/c mice measured by FACS and gated as described above. Each point represents the mean cell size in mice. (d) Cell number (log10) from young and old mouse spleens of BALB/c mice excluding red blood cells. (e) Percentage of B-cell populations (out of total B cells) in spleens of young and old Balb/c mice measured by FACS. B-cell populations were gated as described above. (f) Cell size of B-cell populations measured by FACS in young and old BALB/c mice. Each point represents the mean cell size in mice. (g) Cell composition of immune ($CD45^+$) cells in spleens of young and old BALB/c mice measured by FACS.

Supplementary Figure 3. Mutational signatures of clonal mouse B cells, and analyses of Ig diversity in human subjects and mice. (a) Top Ig clone sizes of young, old and old enlarged B cells FACS-sorted from the spleens and measured by sequencing of CDR3 regions of Ig heavy chains. (b) Correlation between B-cell size measured by FACS and size of top B-cell clone measured by sequencing of CDR3 regions of Ig heavy chains in FACS-sorted B cells. (c) Proportion of somatic mutations per template within the CDR3 region in young, old and old

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enlarged B cells. (d) Correlation between top clone size (Ig) estimated from Immunoseq data and top clone size estimated by VAF from whole exome sequences. (d) Correlation between B-cell size measured by FACS and top clone size estimated by VAF from whole exome sequences. (e) Correlation of clone size for the same clones reconstructed from RNA-seq and Immunoseq. (f) Scheme with experimental layout of analysis of sorted splenic B cells from 27 months old mice by RNA-seq, Immunoseq and WES. Liver WES were used as a control to call somatic mutations. (g) Number of predicted somatic mutations per sample. (h), (i), (j) and (k) show the results of mutational signatures of total somatic mutations in B cells from old mice. (l) Scheme of the analysis of GTEx dataset. Data from spleen and blood was used to reconstruct CDR3 regions, estimate Ig diversity (chaoE mean), and blood WGS was used to predict somatic mutations. (m) Ig diversity (chaoE_mean) in spleens of young and old GTEx subjects calculated from RNA-seq data using mixcr and VDJtools software. P-value was calculated with a two-sided Student t-test. (n) Correlation between Ig diversity (chaoE_mean) in blood of GTEx subjects and age, in years. (o) Ig diversity in blood of GTEx subjects with CHIP (Clonal Hematopoiesis of Indeterminate Potential) and without it. CHIP was calculated from whole genome sequences of blood. (p) Correlation between top clone (WES) and Ig clone sizes (sum of top 5 CDR3 clones) in blood of GTEx subjects.

Supplementary Figure 4. Age-associated and clonal age-associated B cells infiltrate peripheral tissues. (a) Representative hematoxylin and eosin staining of old unaffected mice (upper row) and old mice with enlarged spleen and diagnosed with lymphoma (bottom row). Asterisks show infiltrates of lymphoid cells. (b) and (c) show two UMAPs displaying B-cell populations in lung and kidney, respectively, extracted from *Tabula Muris Senis* dataset[40]. (d) and (e) show two UMAPs of B cell populations from lung and kidney, respectively, extracted from Kimmel *et al* dataset[39]. (f-i) show the clonality signature scores applied to each dataset and each tissue independently.

Supplementary Figure 5. Clonal expansions of B cells are supported by *c-Myc* activation and DNA methylation alterations. (a) Mean DNA promoter methylation levels for genes encoding Myc targets whose expression increase or did not change in clonal B cells. (b) Correlation between mean B cell size and the rDNA methylation age. (c) Correlation between the DNA methylation age calculated by the blood epigenetic clock[36] and the size of top B cell clones. (d) Slope values from correlations between gene expressions and mean promoter DNA methylation or gene body DNA methylation. The dashed line indicates slope value without correlation. (e) Correlation between *Ptpn1* gene expression and mean DNA methylation of its promoter, dots are sized by size of top B cell clone. Correlations between two variables were evaluated using Pearson's correlation coefficients. (f) Gene set enrichment analysis of Msigdb pathways for genes ranked based on the following correlations (from left to right): 1) gene expression with the size of top B-cell clone, 2) gene expression with the mean size of B cells, 3) protein levels with the size of top B-cell clone, 4) mean promoter DNA methylation with the size of top B-cell clone, 5) mean gene body DNA methylation with the size of top B-cell clone. B cells were FACS-sorted from the spleens of 27 months old C57BL/6 female mice. Dots are sized based on -log10 of adjusted p-value and colored based on enrichment score. B-cell clone sizes were estimated from CDR3 sequencing results. (g) Same as in (f) but for KEGG pathways. (h)

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Same as in (f) but for GO pathways. Proteome changes did not reach significance for any of the GO pathways. (l) *c-Myc* fold change in young, old and old enlarged FACS-sorted B cells normalized for DNA concentration and measured by qRT-PCR, p-values were calculated with a two-sided Student t-test. (j) Tapestation image showing rRNA 28S and 18S abundance in young, old and old enlarged B cells. Prior to loading, RNA was normalized to the DNA content isolated from the same sample. (k) Representative images of old and old enlarged B cell nuclei stained for DNA with DAPI and the measured DAPI area for three independent experiments, and (l) quantification of nuclei size. **** p-value < 0.0001. Scale bar, 5 μm.

Supplementary Figure 6. Single-cell RNA sequencing uncovered clonal B-cell population and its relation to age-associated B cells (ABC). (a) tSNE plots of single-cell RNA-seq of spleen of one old mouse from the Kimmel dataset[39] (all upper plots) and *Tabula Muris Senis* dataset[40] (all bottom plots). (b) tSNE plot for single-cell RNA-seq of spleen of old mice. The size of each dot is a combined CPM of *Cr2* (CD21) and *Fcer2a* (CD23), and color of each dot is CPM of *Cd19*. (c) tSNE plot for single-cell RNA-seq of spleen from old mice. The size of each dot is a combined CPM of *Itgax* (Cd11c), and *Tbx21* (T-bet), and the color of each dot is CPM of *Cd19*. (d) Correlation between clonality signature score estimated for each cell and clone size (% of reconstructed IgK) to which the cell belongs. (e) Correlation of gene expression of ABCs and ACBC markers with size of top B cell clones (sum of the top 5 CDR3 clones) in GTEx subjects, colored by sex, 1 is male, and 2 is female. (f) Gating strategy to sort ABC and ACBC. (g) Percentage of splenic ABCs and ACBC of total B cells in young to old female C57BL/6 mice. (h) Cell sizes of follicular and ACBC in spleens of young (n=4) and old (n=10) mice. (i) *c-Myc* and *Tbx21* fold change to *Actnb* measured with qRT-PCR in splenic follicular B cells, ABC and ACBC FACS-sorted from old mice, p-value was calculated with a one-sided Student t-test.

Supplementary Figure 7. Longitudinal analysis reveals B-cell size and CD21$^-$CD23$^-$ B cells as predictors of lifespan, and co-culture experiments point to their origin. (a) Schematic representation of the longitudinal experiment. Blood was collected from tails of C57BL/6 female mice across seven timepoints for FACS analysis. (b) Mean chronological changes in blood cell composition in 25 old mice. Myeloid cells were gated as CD45$^+$CD3$^-$CD11b$^+$CD19$^-$, B cells as CD45$^+$CD11b$^-$CD3$^-$CD19$^+$, T cells as CD45$^+$CD3$^+$CD11b$^-$CD19$^-$. (c) Same as in (b) but aligned to the last blood collection before animal death, -1 is one round before last blood collection, -2 is two rounds before last blood collection. (d) Correlation between the cell size score (FSC.score), myeloid score and CD21-CD23- (CD21.score) calculated for 25, 27, 28 and 31 months old C57BL/6 female mice. Red squares are statistically significant (p<0.05) positive correlations. (e) Mean cell size measured by FACS in young (YF) and old follicular (OF) B cells incubated with i) ABCs from old spleens, ii) immune cells (CD45$^+$) from young spleens excluding B cells (yEE), or iii) immune cells (CD45$^+$) from old spleens excluding B cells (oEE). Results shown are from three independent experiments, 3-4 biological replicas per experiment. (f) Percentages of CD21$^-$ CD23$^-$ B cells after incubation with yEE or oEE measured by FACS. Due to big variability in cell viability each value is normalized to OF value within the experiment. Results are shown from three independent experiments, 3-4 biological replicas per experiment. (j) Single-cell CPM of cytokines in follicular/marginal zone B cells, ACBC and ABC from the *Tabula Muris Senis* [40] spleen of 21-month-old mice. Secretome is a mean gene expression level of five cytokines

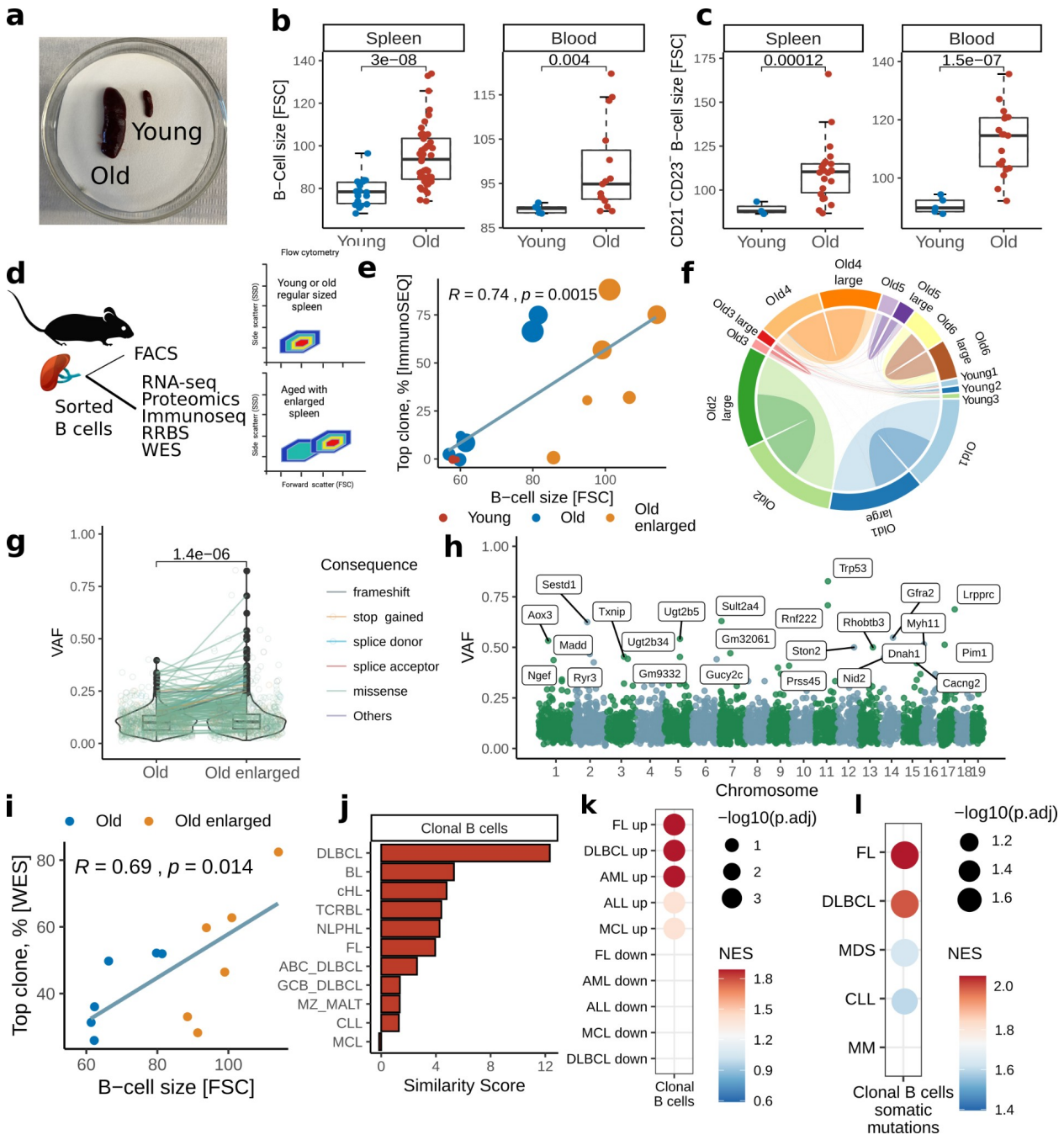
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shown.

Supplementary Figure 8. Effects of *c-Myc* deficiency and overexpression on immune cell composition and premalignant phenotypes of B cells. (a) *c-Myc* expression fold change normalized to DNA concentration in B cells, (b) spleen weight normalized by body weight, and (c) RNA to DNA concentration ratio in B cells, (d-f) splenic cell sizes from *Myc*$^{+/-}$ mice and their Wt siblings. (g) Cell composition of immune cells (CD45^{+}) and B-cell populations (h,i) in spleens from old Wt and *Myc*$^{+/-}$ mice, gated as in Fig. S1f-g. (j) Analysis of Emu-*Myc* mice. (k) Total RNA/DNA ratio in FACS-sorted splenic B cells from non-carrier (NC) controls and Emu-*Myc* mice. (l) Mean size of B and T cells measured by FACS in spleens of NC and Emu-*Myc* mice. (m) Mean cell size of B-cell populations measured by FACS in bone marrow of NC and Emu-*Myc* mice. (n) Ig diversity (chaoE_mean) of splenic B cells from NC and Emu-*Myc* mice, calculated from RNA-seq data using mixcr and VDJtools software.

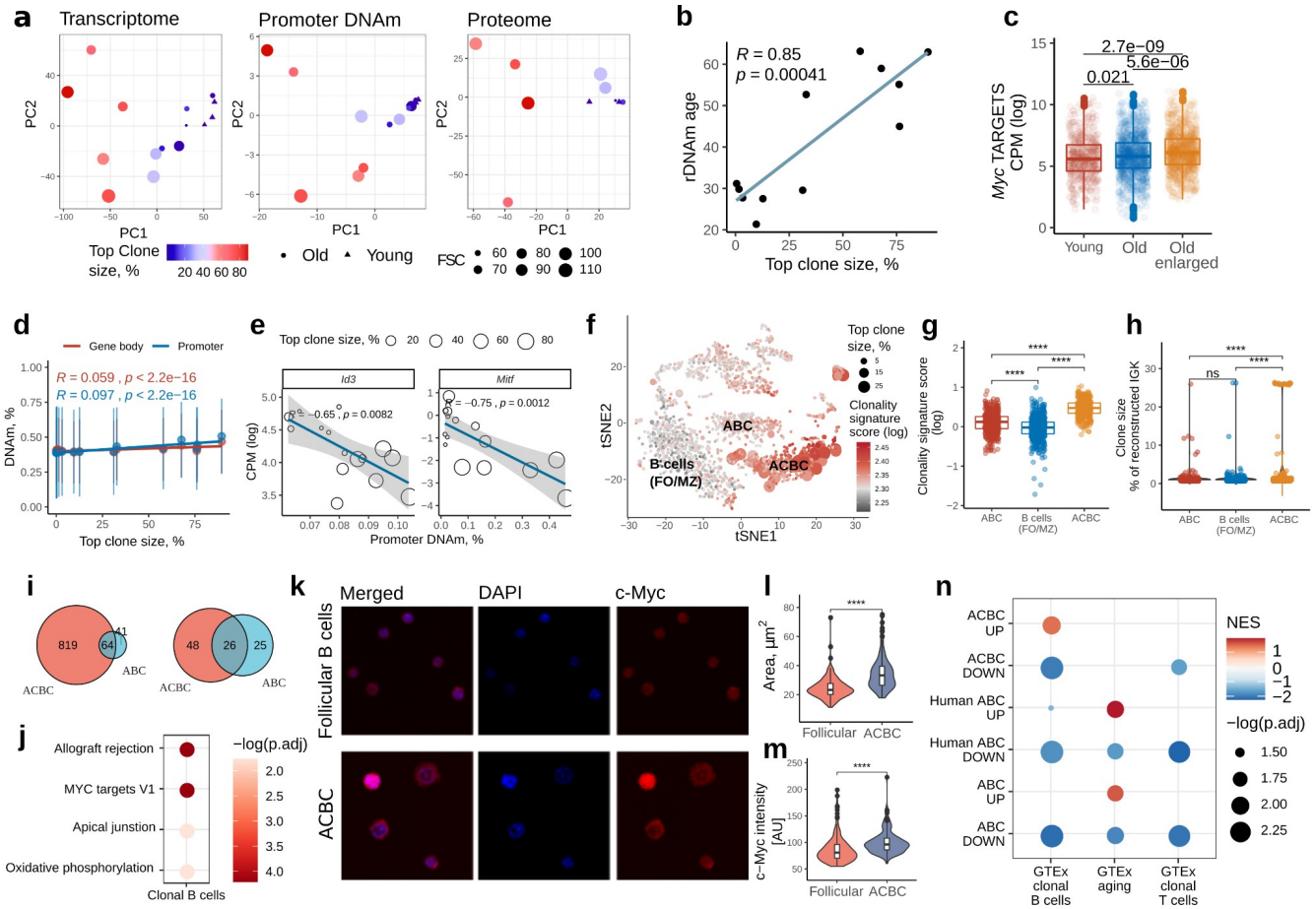
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



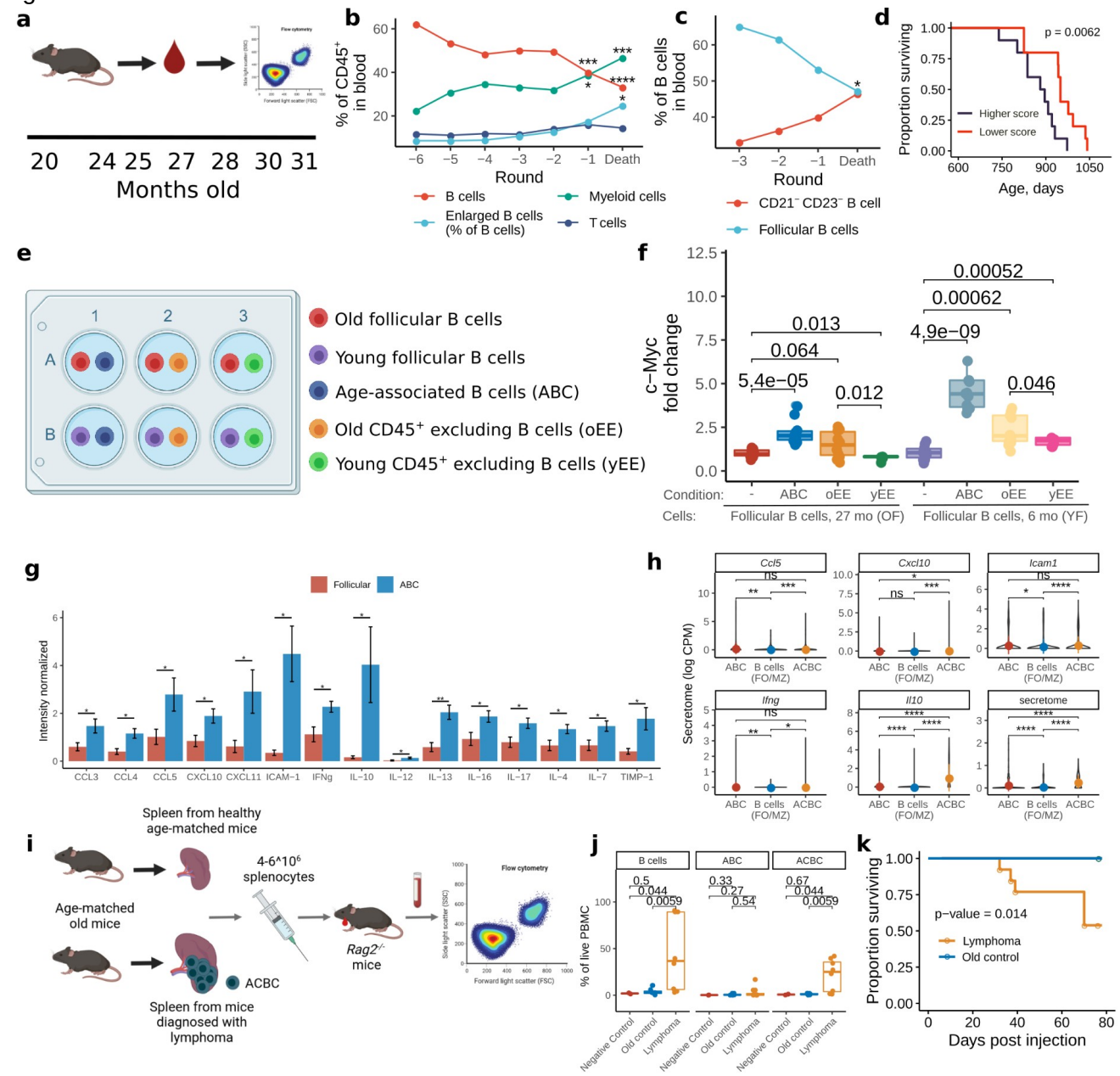
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



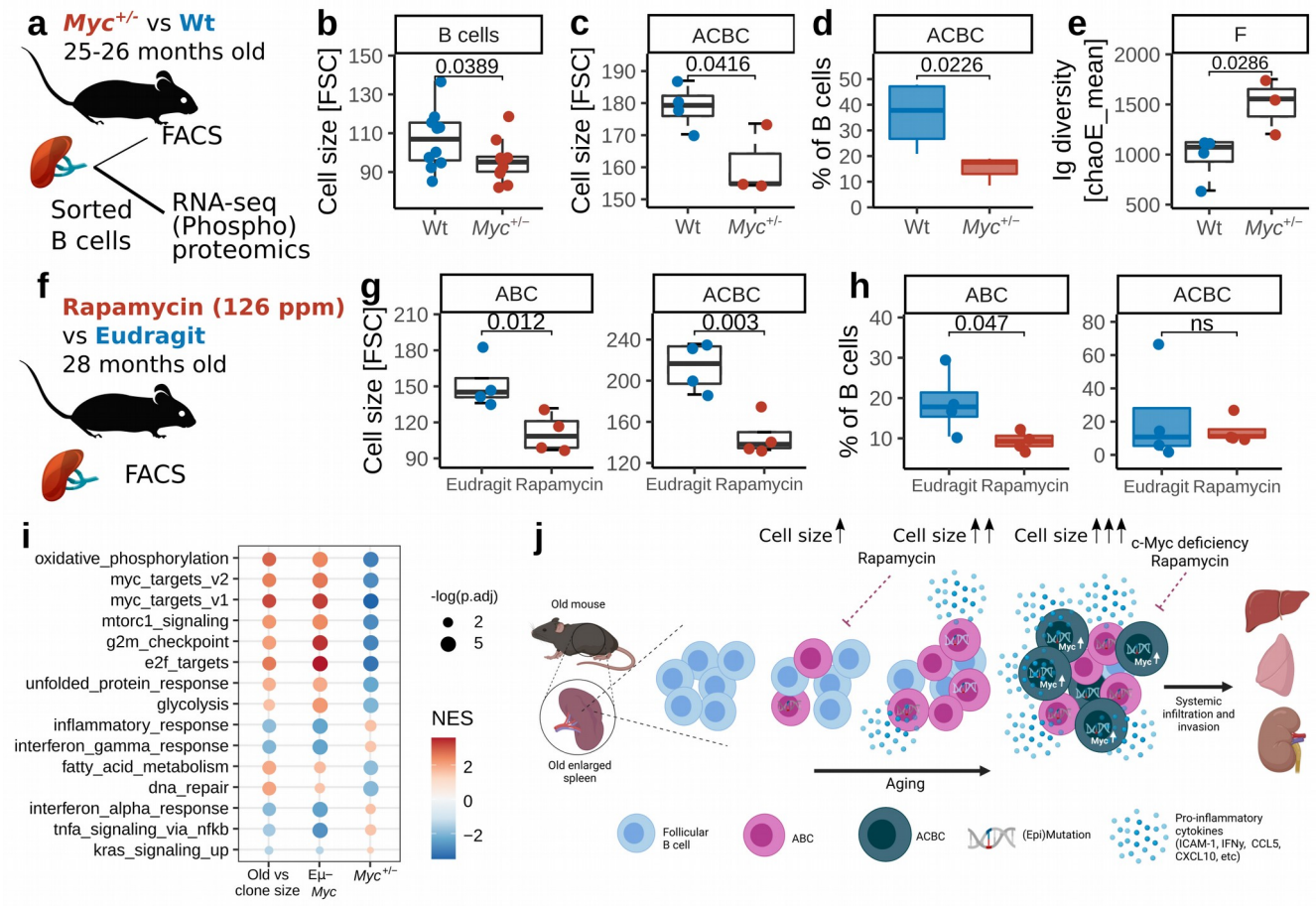
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



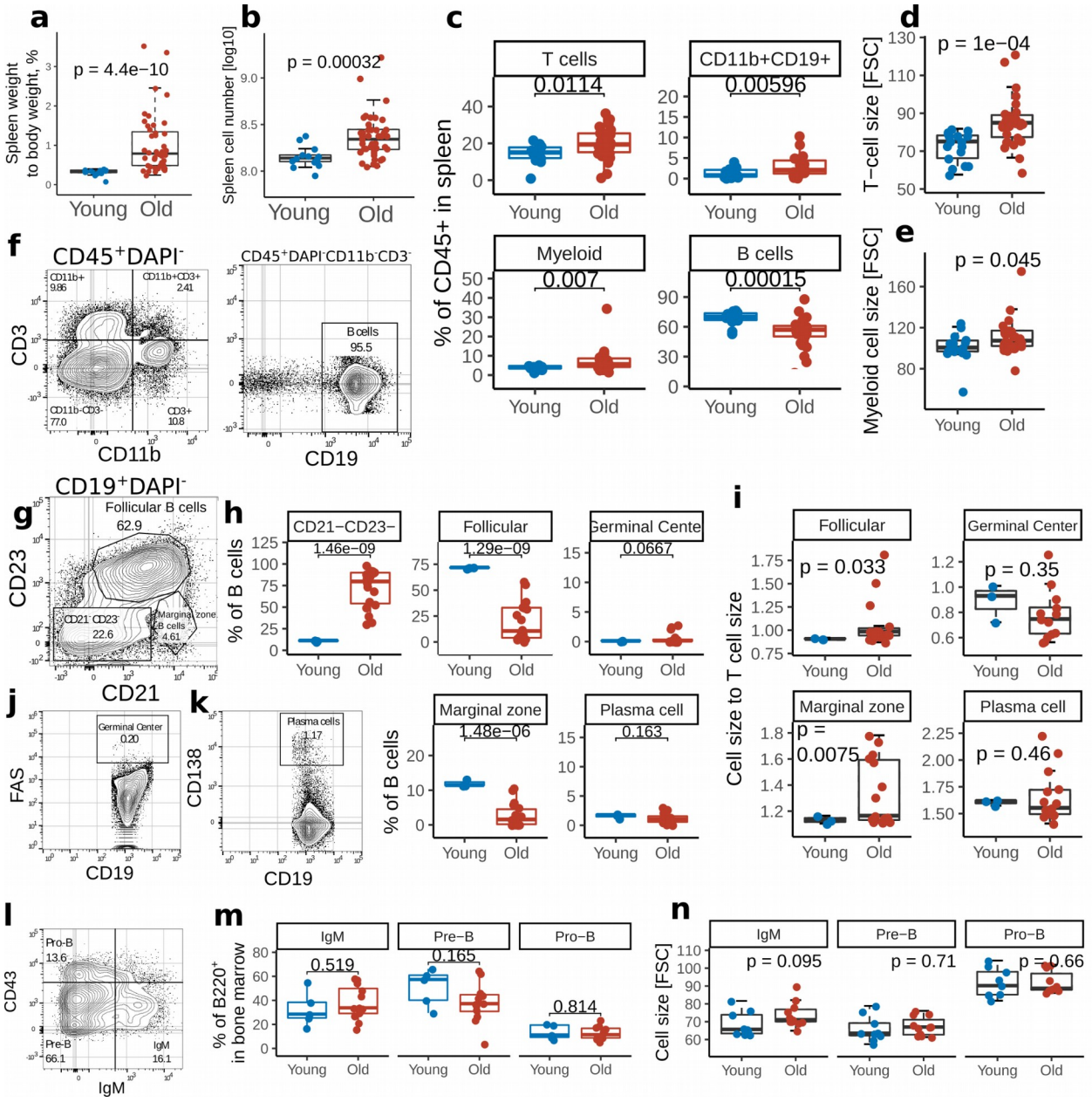
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4



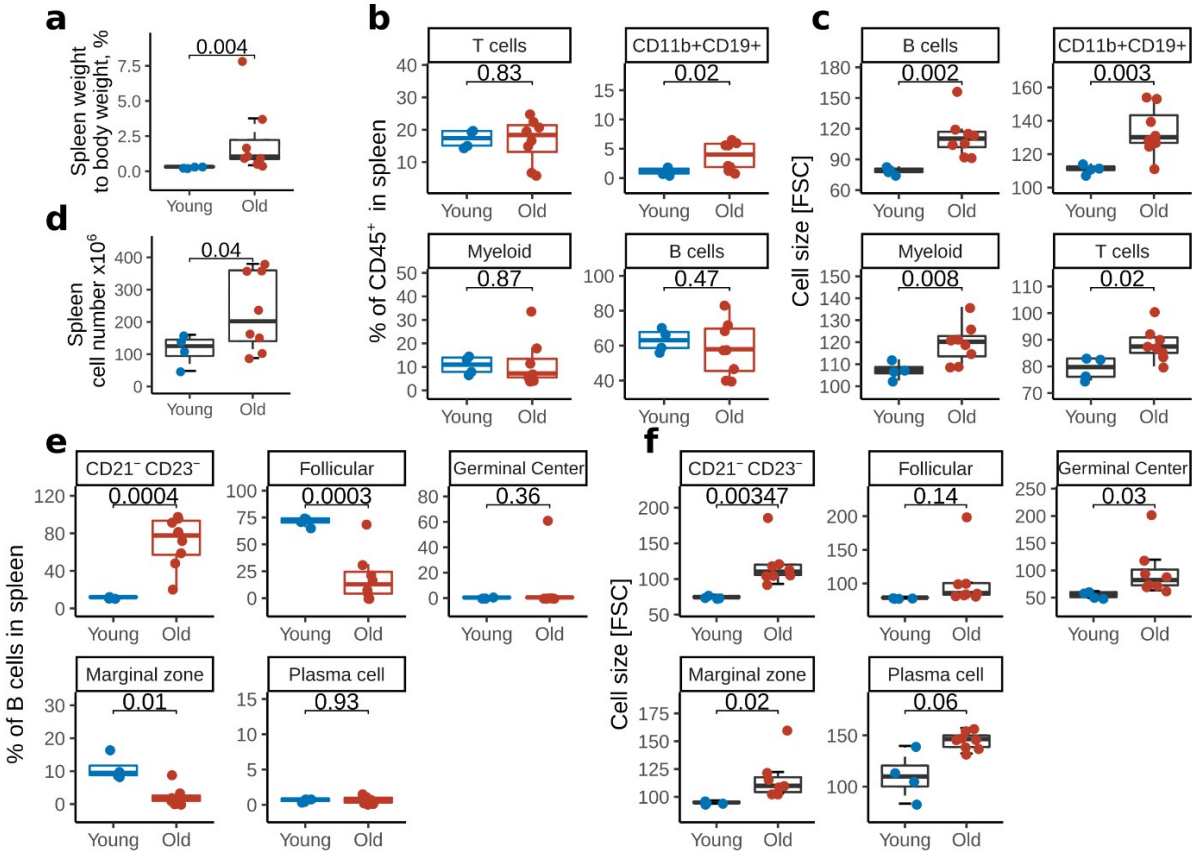
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S1



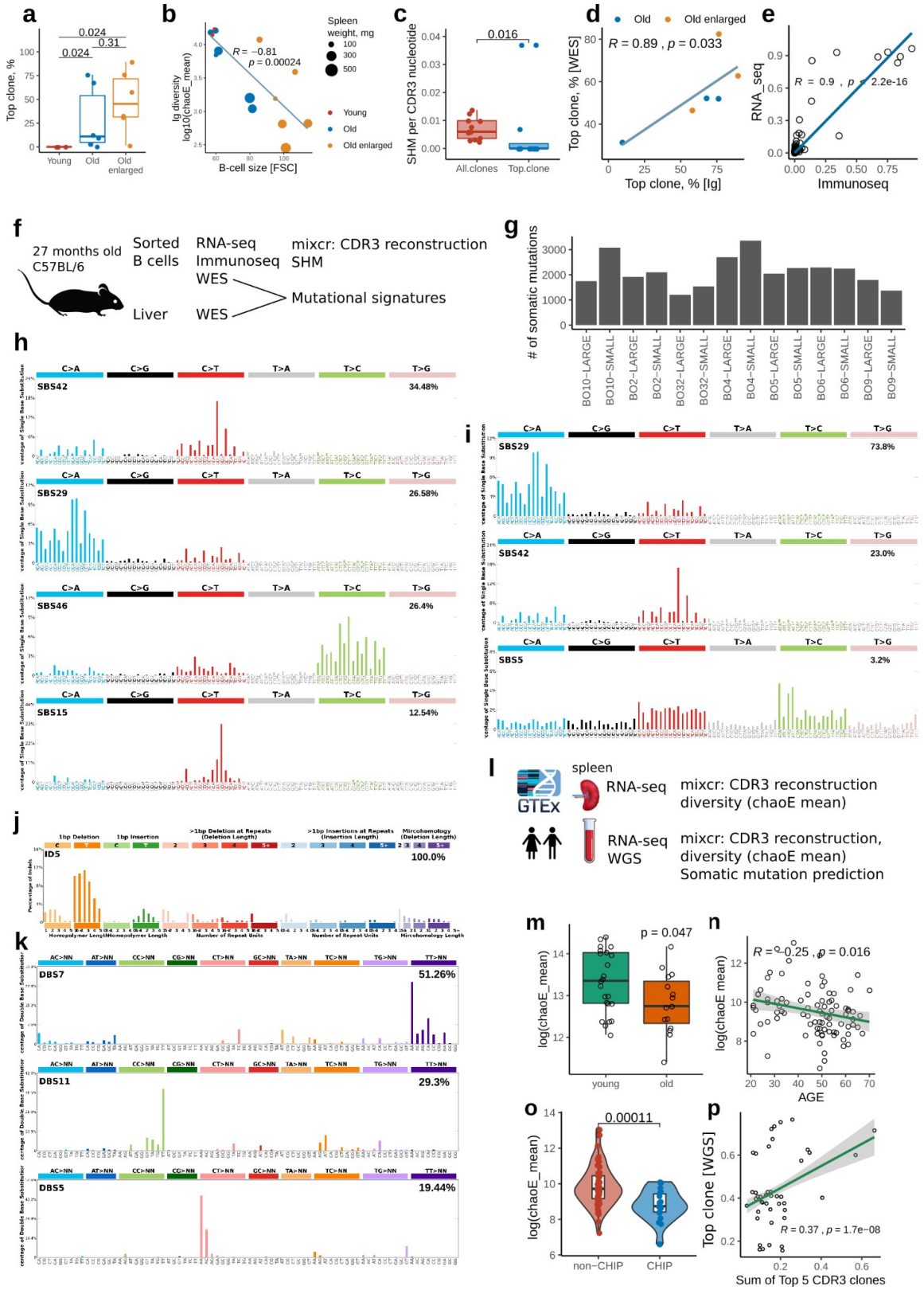
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S2



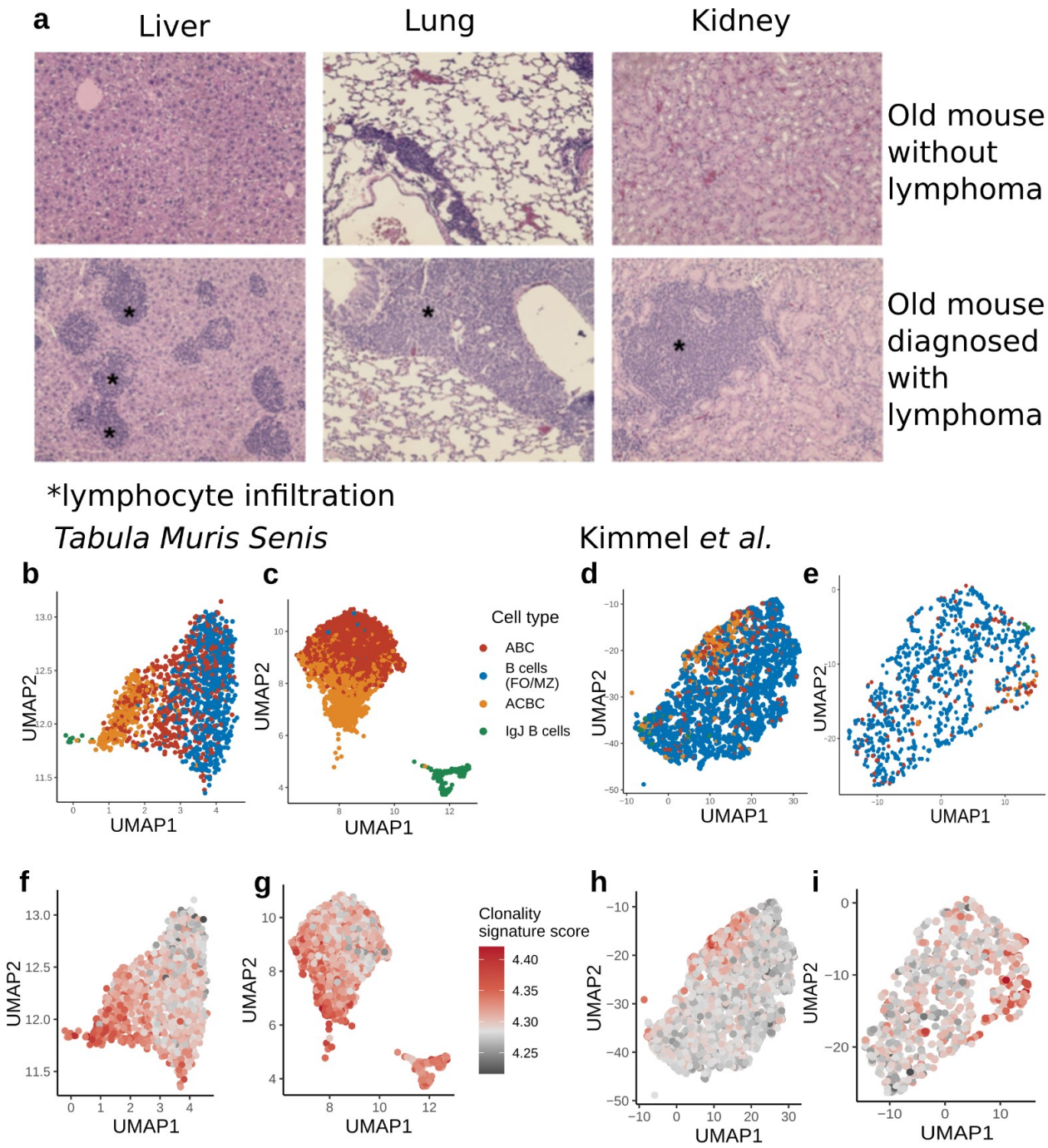
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S3



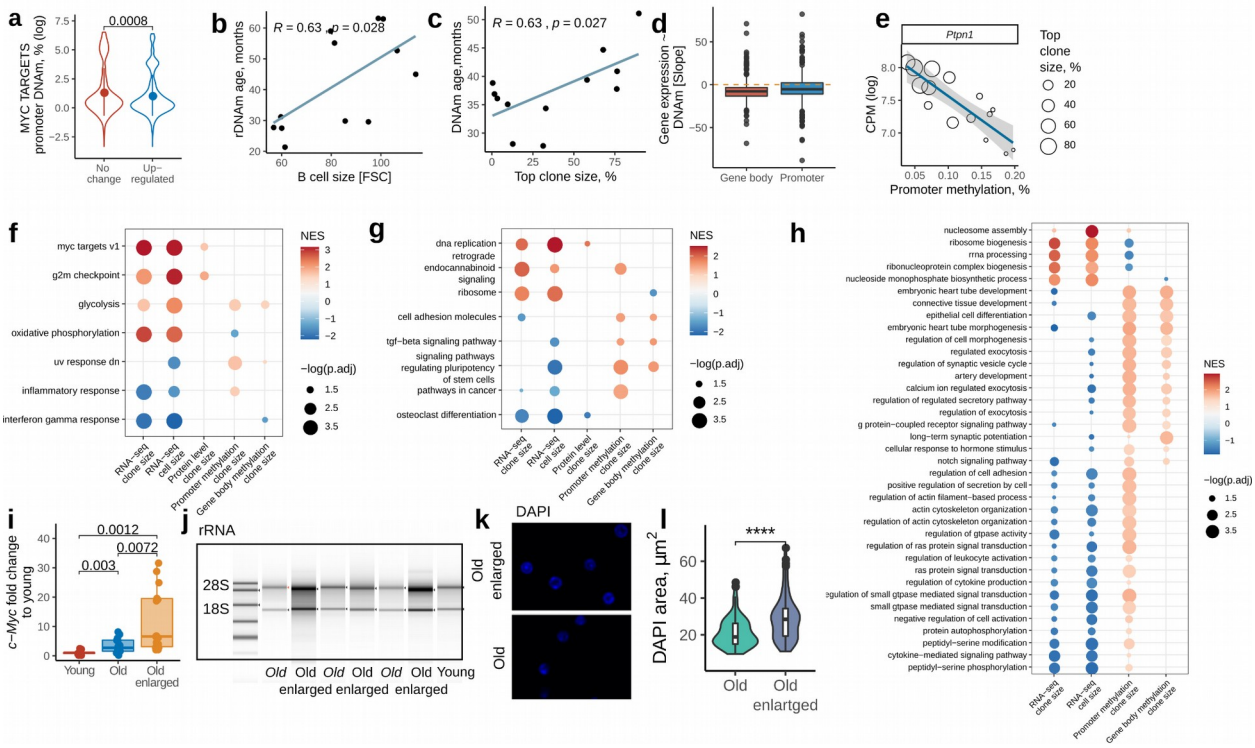
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S4



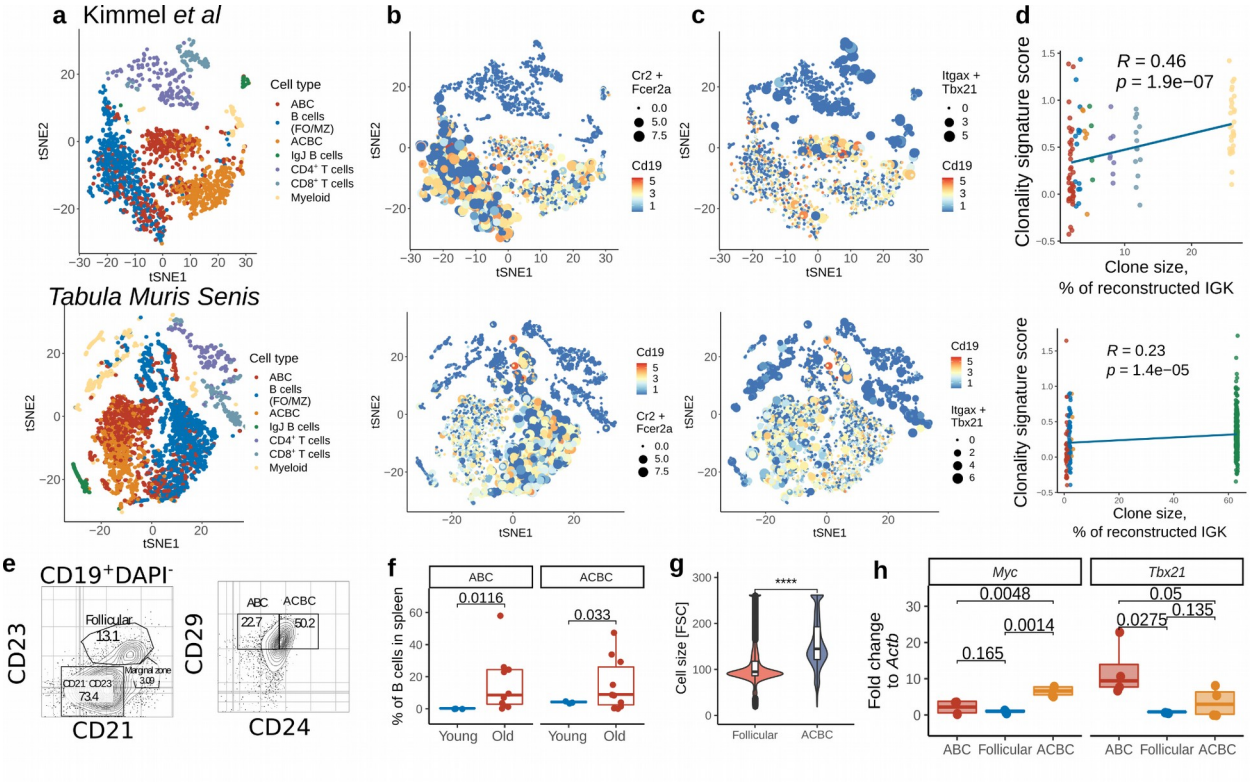
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S5



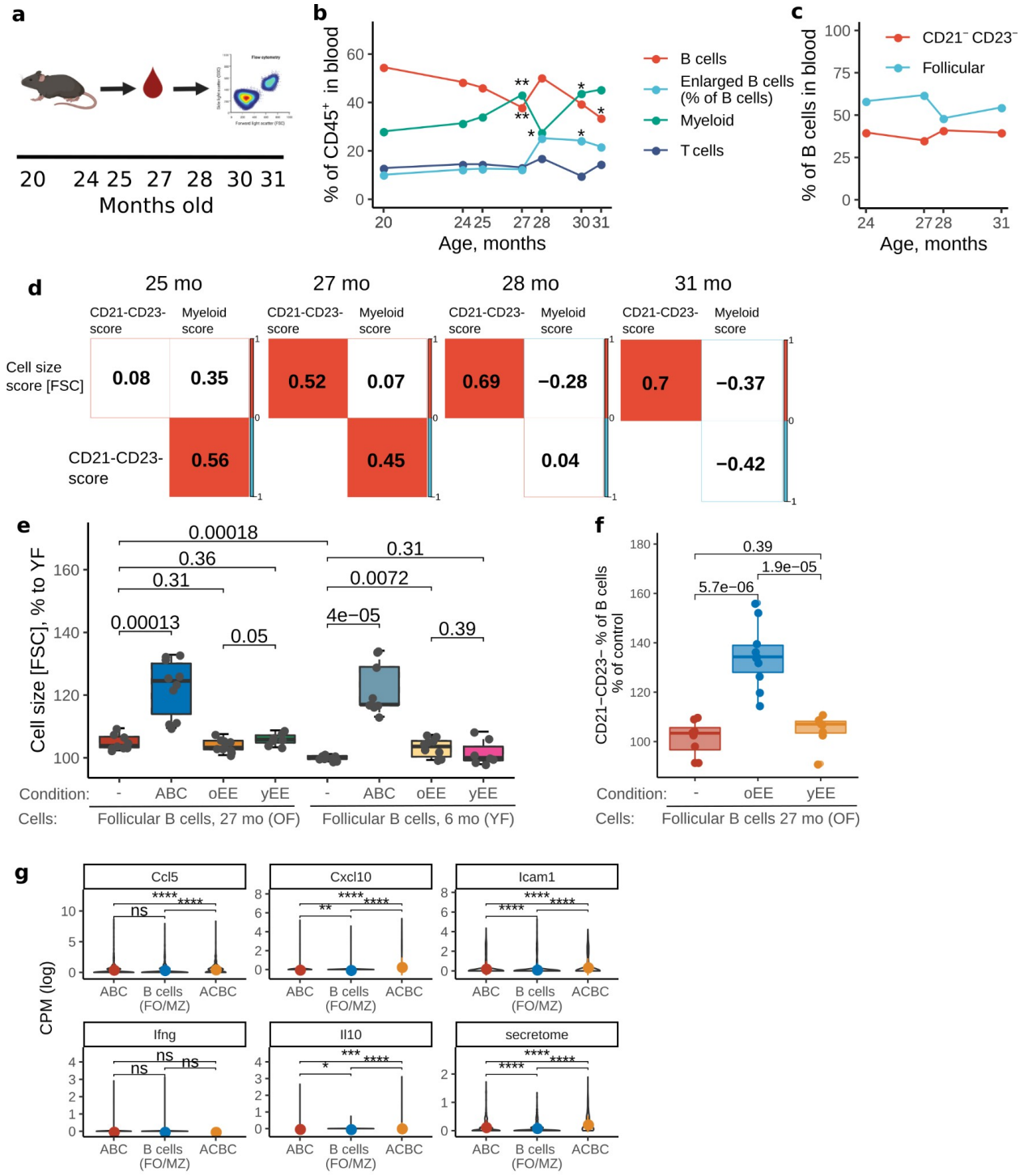
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S6



bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S7



bioRxiv preprint doi: https://doi.org/10.1101/2021.02.23.432500; this version posted December 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

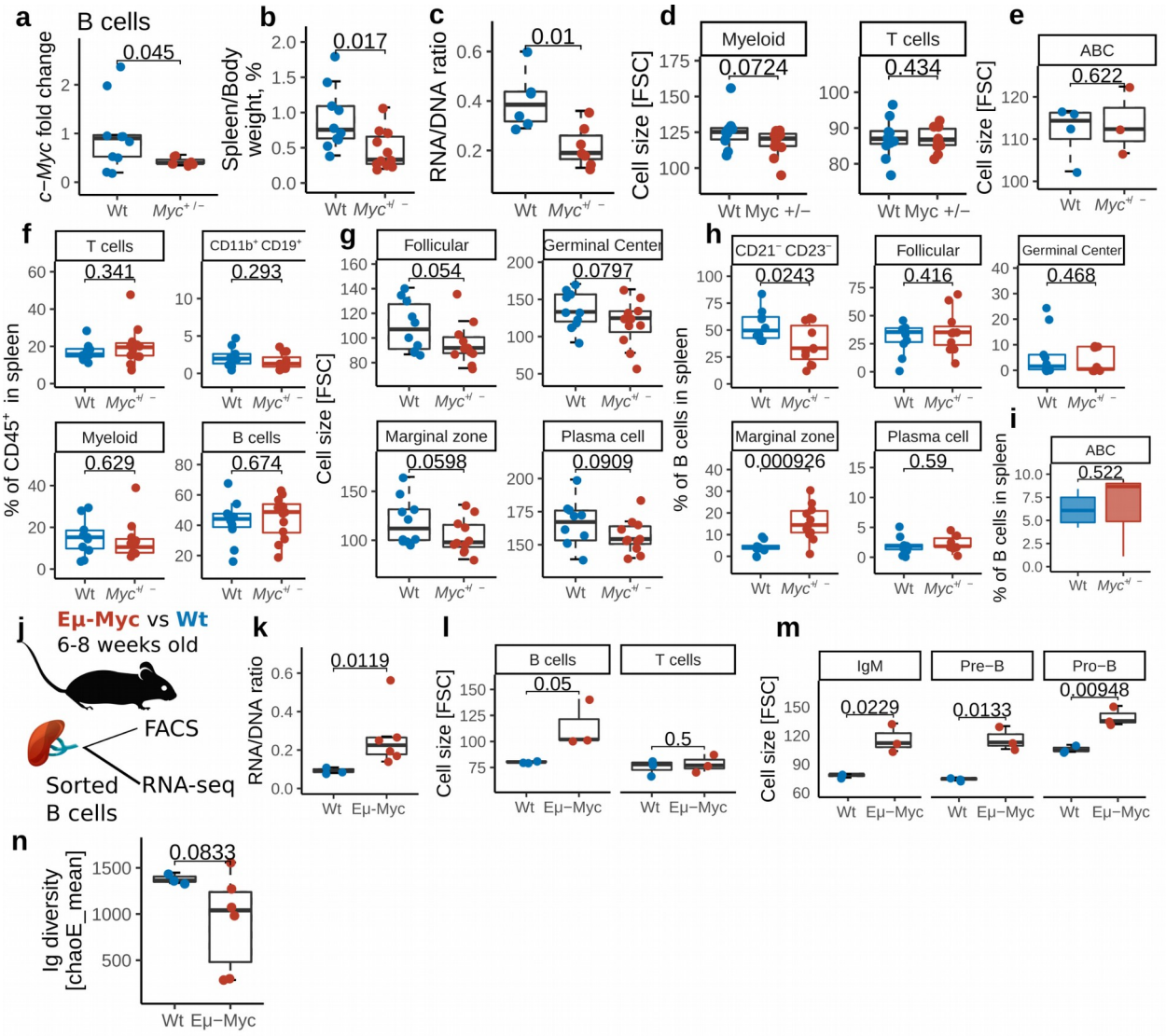



Figure S8