

# Co-occurring clonal hematopoiesis exhibits strong selection and high leukemia risk

Received: 30 January 2026

Accepted: 5 May 2026

Cite this article as: Barnao, K.M., Hubbard, A.K., Chan, I.C. *et al.* Co-occurring clonal hematopoiesis exhibits strong selection and high leukemia risk. *Nat Commun* (2026). <https://doi.org/10.1038/s41467-026-73302-x>

Kara M. Barnao, Aubrey K. Hubbard, Irenaeus C. C. Chan, Weiyin Zhou, Yasminka A. Jakubek, Giulio Genovese, Wendy S. W. Wong, Rebecca L. Kelly, Corey D. Young, Derek W. Brown, Wen-Yi Huang, Neal D. Freedman, Kristine Jones, Amy Hutchinson, Belynda Hicks, Duc Tran, Donna Arnett, Kathleen C. Barnes, Joshua C. Bis, Eric Boerwinkle, Jennifer A. Brody, April P. Carson, Daniel I. Chasman, Michael H. Cho, Pinkal Desai, Margaret F. Doyle, Myriam Fornage, Xiuqing Guo, Nancy Heard-Costa, Marguerite Ryan Irvin, Andrew D. Johnson, Sharon L. R. Kardia, Charles Kooperberg, Daniel Levy, Joshua P. Lewis, Yun Li, Ruth J. F. Loos, Taralynn M. Mack, Rasika A. Mathias, Braxton D. Mitchell, Kari E. North, Nathan Pankratz, Patricia A. Peyser, Michael H. Preuss, Bruce M. Psaty, Laura M. Raffield, Susan Redline, Stephen S. Rich, Jerome I. Rotter, Edwin K. Silverman, Albert V. Smith, Jennifer A. Smith, Adrienne Stilp, Yin Cao, Paul Scheet, Alexander P. Reiner, Alexander G. Bick, Stephen J. Chanock, Paul L. Auer, Kelly L. Bolton & Mitchell J. Machiela

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

**Co-occurring clonal hematopoiesis exhibits strong selection and high leukemia risk**

Kara M. Barnao<sup>1,\*</sup>, Aubrey K. Hubbard<sup>1</sup>, Irenaeus C. C. Chan<sup>2</sup>, Weiyin Zhou<sup>1,3</sup>, Yasminka A. Jakubek<sup>4</sup>, Giulio Genovese<sup>5</sup>, Wendy S. W. Wong<sup>1</sup>, Rebecca L. Kelly<sup>1</sup>, Corey D. Young<sup>1</sup>, Derek W. Brown<sup>1</sup>, Wen-Yi Huang<sup>1</sup>, Neal D. Freedman<sup>1</sup>, Kristine Jones<sup>1,3</sup>, Amy Hutchinson<sup>1,3</sup>, Belynda Hicks<sup>1,3</sup>, Duc Tran<sup>2</sup>, Donna Arnett<sup>6</sup>, Kathleen C. Barnes<sup>7</sup>, Joshua C. Bis<sup>8</sup>, Eric Boerwinkle<sup>9</sup>, Jennifer A. Brody<sup>8</sup>, April P. Carson<sup>10</sup>, Daniel I. Chasman<sup>11,12</sup>, Michael H. Cho<sup>13</sup>, Pinkal Desai<sup>14</sup>, Margaret F. Doyle<sup>15</sup>, Myriam Fornage<sup>9,16</sup>, Xiuqing Guo<sup>17</sup>, Nancy Heard-Costa<sup>18</sup>, Marguerite Ryan Irvin<sup>19</sup>, Andrew D. Johnson<sup>20</sup>, Sharon L. R. Kardia<sup>21</sup>, Charles Kooperberg<sup>22</sup>, Daniel Levy<sup>20</sup>, Joshua P. Lewis<sup>23</sup>, Yun Li<sup>24</sup>, Ruth J. F. Loos<sup>25,26</sup>, Taralynn M. Mack<sup>27</sup>, Rasika A. Mathias<sup>28</sup>, Braxton D. Mitchell<sup>23</sup>, Kari E. North<sup>29</sup>, Nathan Pankratz<sup>30</sup>, Patricia A. Peyser<sup>21</sup>, Michael H. Preuss<sup>25</sup>, Bruce M. Psaty<sup>8,31,32</sup>, Laura M. Raffield<sup>33</sup>, Susan Redline<sup>34</sup>, Stephen S. Rich<sup>35</sup>, Jerome I. Rotter<sup>36</sup>, Edwin K. Silverman<sup>13</sup>, Albert V. Smith<sup>37</sup>, Jennifer A. Smith<sup>21,38</sup>, Adrienne Stilp<sup>39</sup>, Yin Cao<sup>40,41,42</sup>, Paul Scheet<sup>43</sup>, Alexander P. Reiner<sup>44</sup>, Alexander G. Bick<sup>27</sup>, Stephen J. Chanock<sup>1</sup>, Paul L. Auer<sup>45</sup>, Kelly L. Bolton<sup>2</sup>, Mitchell J. Machiela<sup>1,\*</sup>

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD, USA

<sup>2</sup>Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

<sup>3</sup>Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc., Frederick, MD, USA

<sup>4</sup>Department of Internal Medicine, University of Kentucky, Lexington, KY, USA

<sup>5</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>6</sup>University of South Carolina, Columbia, SC, USA

<sup>7</sup>Galatea Bio, Inc., Miami, FL, USA

<sup>8</sup>Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

<sup>9</sup>Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

<sup>10</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

<sup>11</sup>Harvard Medical School, Boston, MA, USA

<sup>12</sup>Division of Preventative Medicine, Brigham and Women's Hospital, Boston, MA, USA

<sup>13</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA

<sup>14</sup>Division of Hematology and Oncology, Weill Cornell Medical College, New York, NY, USA

<sup>15</sup>Department of Pathology and Laboratory Medicine, The University of Vermont Lerner College of Medicine, Colchester, VT, USA

<sup>16</sup>Institute of Molecular Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

<sup>17</sup>The Institute for Translational Genomics and Population Sciences, Department of Medicine, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

<sup>18</sup>Department of Medicine, Boston University School of Medicine, Boston, MA, USA

<sup>19</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA

- <sup>20</sup>Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
- <sup>21</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA
- <sup>22</sup>Division of Public Health Science, Fred Hutchinson Cancer Center, Seattle, WA, USA
- <sup>23</sup>Department of Medicine, University of Maryland Baltimore, Baltimore, MD, USA
- <sup>24</sup>Department of Genetics, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>25</sup>The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- <sup>26</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
- <sup>27</sup>Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
- <sup>28</sup>Laboratory of Allergic Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA
- <sup>29</sup>Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>30</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA
- <sup>31</sup>Department of Medicine, University of Washington, Seattle, WA, USA
- <sup>32</sup>Department of Health Systems and Population Health, University of Washington, Seattle, WA, USA
- <sup>33</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
- <sup>34</sup>Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA
- <sup>35</sup>Department of Genome Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA
- <sup>36</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA
- <sup>37</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA
- <sup>38</sup>Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA
- <sup>39</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA
- <sup>40</sup>Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA
- <sup>41</sup>Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA
- <sup>42</sup>Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
- <sup>43</sup>Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA
- <sup>44</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA
- <sup>45</sup>Division of Biostatistics, Data Science Institute, and Cancer Center, Medical College of Wisconsin, Milwaukee, WI, USA

\*Corresponding Authors: Kara M. Barnao ([kara.barnao@nih.gov](mailto:kara.barnao@nih.gov)) and Mitchell J. Machiela ([mitchell.machiela@nih.gov](mailto:mitchell.machiela@nih.gov))

ARTICLE IN PRESS

**Abstract**

Clonal hematopoiesis of indeterminate potential (CHIP) and mosaic chromosomal alterations (mCAs) are two types of clonal hematopoiesis (CH) associated with hematological parameters and malignancy risk. Here we show, in genomic data from 546,090 biobank participants, that co-occurring CH ( $\geq 2$  CH mutations detected) is present in 1.6% of cancer-free individuals and shows strong evidence for selection (up to 804x enrichment). Co-occurrence is more frequent in those with a prior cancer (3.6%), suggesting treatment-induced selection. Acquisition of CHIP usually precedes mCAs with co-occurrences manifesting stronger phenotypic disruptions in telomere attrition and hematologic parameters than component CH events. Individuals with co-occurring CH have pronounced elevations in risk of myeloid and lymphoid malignancies (HRs $>40$ ), particularly when CHIP and mCAs overlap genomically. Our findings indicate CH co-occurrences are selected for in the aging population and identify CH clones with notable implications for future malignancy risk.

**(140/ 150 words)**

ARTICLE IN PRESS

## Introduction

The accumulation of post-zygotic mutations from environmental exposures and DNA replication errors provides the foundation for molecular evolution. While most somatic mutations are inconsequential, some promote dysregulation of core cellular pathways, leading to the acquisition or selection of additional mutations favoring progression from premalignant states to cancer<sup>1</sup>. Tissue-specific knowledge of commonly acquired mutations, their co-occurrence, and interactions can provide novel etiologic insights critical for identifying precancerous clones with high potential for malignant transformation.

Clonal hematopoiesis (CH) is the expansion of hematopoietic stem and progenitor cells containing post-zygotic mutations in individuals with normal hematologic parameters. Clonal proliferation can be driven by single nucleotide variants (SNVs) or small insertions and deletions (indels) in leukemia driver genes, termed clonal hematopoiesis of indeterminate potential (CHIP). Previous studies have observed *DNMT3A*, *TET2*, *ASXL1*, *PPM1D*, and *JAK2* as the most frequent genes mutated in CHIP<sup>2-6</sup>. CH can also be driven by large (typically megabase-scale) structural chromosomal changes (e.g., chromosomal gains, losses or copy-neutral loss of heterozygosity, CNLOH), collectively termed mosaic chromosomal alterations (mCAs). The most frequently detected mCAs are loss of the sex chromosomes: mosaic loss of the Y chromosome (mLOY) in males and mosaic loss of the X chromosome (mLOX) in females. CH has increased population frequencies as age advances, demonstrating the erosive associations of aging on genomic integrity.

A sizeable fraction of cancer-free individuals have some type of detectable CH that varies in frequency based on population characteristics (e.g., estimated population frequencies CHIP=6%<sup>7</sup>, autosomal mCAs=3.5%<sup>8</sup>, mLOY=20%<sup>9</sup> and mLOX=12%<sup>10</sup>). While both CHIP and mCAs are risk factors for developing a subsequent hematologic malignancy, most individuals with CH do not develop a malignancy motivating the need to identify high risk CH subtypes<sup>2,4-6,8,11-14</sup>. Specifically, few studies have evaluated the potential synergistic role played by the co-occurrence or overlap of CH types, although preliminary observations suggest elevated risk of cancer and cancer-related phenotypes<sup>2,3,15,16</sup>. In the present study, we characterized the frequency of co-occurring CH in a large, population-based analysis comprising three studies: the UK Biobank (UKBB, N=478,441) for our initial discovery analyses, and the Trans-Omics for Precision Medicine Program (TOPMed, N=67,390) and the NCI's Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO, N=259) as independent replication sets. We identified 210 CH co-occurrences enriched over what is expected by chance, characterized associations of 28 CHIP-mCA co-occurrences with cancer-related phenotypes, and nominated high-risk clones associated with hematologic cancer risk.

## Results

### *Characterization of CHIP in cancer-free UKBB participants*

A total of 21,737 CHIP mutations were detected in 20,354 UKBB participants with no prior cancer diagnosis (4.5%; **Supplementary Data 1**). CHIP was detected in 533 genes, with *DNMT3A*, *TET2*, and *ASXL1* variants present in 70.2% of cancer-free participants with detectable CHIP (**Fig. 1a**). All subsequent co-occurrence analyses were restricted to CHIP genes

harbored by at least 30 UKBB participants to ensure adequate sample size, resulting in 43 genes for investigation. This criterion included 18,693 CHIP events detected in 17,865 UKBB participants. The number of detected mutations in individuals with CHIP ranged from 1 to 7.

#### *Characterization of mCAs in cancer-free UKBB participants*

A total of 71,246 mCAs were detected in 67,081 UKBB participants with no prior cancer diagnosis (14.8%; **Supplementary Data 1**). The number of detectable mCAs per individual with mCAs ranged from 1 to 14. mLOY was the most frequent mCA detected in males (N=41,432, 19.8% of males) and mLOX was most common in females (N=14,839, 6.1% of females). A total of 14,975 autosomal mCAs were detected in 13,422 (3.0%) UKBB participants, most frequently on chromosomes 1, 11, and 14 (**Fig. 1b**).

#### *Characterization of co-occurring CH in cancer-free UKBB participants*

We first examined co-occurrences within CH types (e.g., CHIP-CHIP or mCA-mCA co-occurrences) in the UKBB population. Of the 18,693 individuals with a detectable CHIP mutation in the 43 genes investigated in this study (N $\geq$ 30), we observed 957 (5.1%) individuals with CHIP in more than one gene. Most common co-occurrences were *DNMT3A-TET2* (N=336), *TET2-SRSF2* (N=65), and *TET2-ASXL1* (N=60) (**Supplementary Data 2**). These co-occurrences were largely attributable to high frequencies of the individual CHIP mutations resulting in joint acquisition of both events which align with anticipated frequencies. To identify enrichment of specific co-occurrences, we tested if the frequency of the co-occurrences differed from expected frequencies derived from marginal frequencies, adjusting for age, age<sup>2</sup>, sex, smoking status, and genetic similarity to reference populations. We observed 14 CHIP-CHIP co-occurrences at higher frequencies than expected ( $P \leq 2.7 \times 10^{-5}$ ; **Fig. 2a, Supplementary Data 2**) with the strongest enrichment observed for *SRSF2-IDH2* (Odds ratio (OR)=804.1, 95% Confidence Interval (CI)=(430.8–1500.8),  $P=5.2 \times 10^{-98}$ ), followed by *JAK2-NFE2* (OR=182.7, 95% CI=(76.6–435.8),  $P=7.6 \times 10^{-32}$ ) and *SRSF2-TET2* (OR=28.8, 95% CI=(21.7–38.4),  $P=5.4 \times 10^{-117}$ ). Six genes (*DNMT3A*, *TET2*, *ASXL1*, *SRSF2*, *SF3B1*, *JAK2*) were found to have multiple significantly elevated CHIP-CHIP co-occurrences; for example, *DNMT3A* co-occurred at higher frequencies than expected with mutations in *TET2*, *YLMP1*, *ZBTB33*, *SF3B1*, and *ATM* (**Fig. 2a**). For *TET2* co-occurrences, many participants had more than one detectable *TET2* point mutation, ranging from 2 to 4, within the gene. Comparison of variant allele fractions (VAFs) among enriched CHIP-CHIP component events revealed no consistent patterns suggesting the order of mutational acquisition did not favor more frequently observed CHIP mutations (e.g., *DNMT3A*, *TET2*, *ASXL1*). CHIP-CHIP co-occurrences were not observed significantly less frequently than expected, indicating no evidence for negative selection.

Of the 67,081 individuals with a detectable mCA, 3,425 (5.1%) individuals harbored an mCA on more than one chromosome. Most common co-occurrences were chr14-mLOY and chr15-mLOY in males and chr22-mLOX and chr12-mLOX in females, largely attributable to the high frequencies of the component events (**Supplementary Data 3**). Out of the 286 unique mCA-mCA combinations, 168 (58.7%) had higher frequencies than expected ( $P \leq 8.7 \times 10^{-5}$ ) (**Fig. 2b**

**and Supplementary Data 3).** Chr3-chr18 (OR=119.7, 95% CI=(89.4–160.3),  $P=1.9\times 10^{-226}$ ), chr18-chr19 (OR=102.8, 95% CI=(74.0–142.8),  $P=2.6\times 10^{-168}$ ), and chr12-chr18 (OR=94.5, 95% CI=(73.0–122.3),  $P=4.3\times 10^{-261}$ ) had the strongest effects (**Supplementary Data 3**) and often included whole chromosome amplifications of both chromosomes. Interestingly, other frequent mCA-mCA co-occurrences, such as chr12-chr19 ( $N=74$ ) and chr3-chr12 ( $N=60$ ), were also often comprised of whole chromosome gains. We then further refined our analysis to mCAs defined by chromosomal region and event type (e.g., chr9 p-arm CNLOH) and found that enriched co-occurrence patterns were largely consistent with increased magnitude of association. For example, co-occurrence of whole chromosome gains of chr3 and chr18 showed substantially greater enrichment (OR=1014.6, 95% CI=(641.6–1604.5),  $P=1.4\times 10^{-192}$ ) (**Supplementary Fig. 1 and Supplementary Data 4**). As with CHIP-CHIP co-occurrences, no mCA-mCA co-occurrences occurred less frequently than expected.

In an investigation of CHIP-mCA co-occurrences, we observed 3,618 (0.8%) UKBB participants with both detectable CHIP and mCAs. Most frequent co-occurrences were *DNMT3A* with mLOY ( $N=869$ ) or mLOX ( $N=444$ ) as well as *TET2* with mLOY ( $N=340$ ), again attributable to the high frequencies of the component events (**Supplementary Data 5**). Interestingly, co-occurrences including mLOY occurred less frequently than expected (OR<sub>*DNMT3A*</sub>=0.7, 95% CI=(0.7–0.8),  $P=9.5\times 10^{-16}$ ; OR<sub>*TET2*</sub>=0.7, 95% CI=(0.6–0.8),  $P=4.5\times 10^{-9}$ ), suggesting possible exclusivity between these CH types (**Supplementary Data 5**). CHIP and autosomal mCAs were detected in 1,165 (0.3%) UKBB study participants, representing approximately 2-fold more co-occurrences than expected ( $N_{\text{expected}}=602$ , binomial  $P=5.1\times 10^{-92}$ ). Twenty-eight (28) CHIP-autosomal mCA co-occurrences emerged from enrichment analyses ( $P\leq 4.8\times 10^{-5}$ ) (**Fig. 2c and Supplementary Data 5**).

Co-occurrences were grouped into two categories for further analysis: overlapping and non-overlapping<sup>3</sup>. Overlapping CHIP-mCA events, where a CHIP mutation was detected within the bounds of a co-occurring mCA, were more frequent and observed in 314 participants (27.0% of UKBB participants with CHIP-autosomal mCA co-occurrences). This accounted for 8 enriched co-occurrences (*ATM*-chr11, *DNMT3A*-chr2, *JAK2*-chr9, *KMT2D*-chr12, *MYD88*-chr3, *TET2*-chr4, *TP53*-chr17, *YLP1M1*-chr14) (**Fig. 2c**). Of these, *JAK2*-chr9 was detected most frequently ( $N=98$ ) and exhibited the greatest enrichment (OR=481.4, 95% CI=(360.5–642.8),  $P<2.2\times 10^{-308}$ ), followed by *MYD88*-chr3 (OR=133.6, 95% CI=(78.9–226.0),  $P=2.5\times 10^{-74}$ ) (**Fig. 2c and Supplementary Data 5**). We noted high consistency in the location and type of the CHIP mutation and mCA in participants with these two co-occurrences. For example, *JAK2*-chr9 co-occurrences were primarily *JAK2* V617F-chr9pCNLOH, and *MYD88*-chr3 co-occurrences were generally *MYD88* L252P-whole chromosome 3 amplifications. In other cases, we observed instances where the CHIP mutation within a gene varied but the mCA was the same (e.g., *ATM* missense variants-chr11qCNLOH, *DNMT3A* missense variants-chr2pCNLOH). Similar to *TET2* CHIP-CHIP co-occurrences, individuals with *TET2*-chr4 co-occurrences often had more than one *TET2* point mutation. In analyses refined to include mCA chromosomal location and event types, we observed increases in magnitude of association for many CHIP-mCA co-occurrences. For example, *JAK2*-chr9pCNLOH showed a stronger association (OR=1741.1, 95% CI=(1261.3–2404.3),  $p<2.2\times 10^{-308}$ ) (**Supplementary Fig. 2 and Supplementary Data 6**).

In UKBB participants with non-overlapping CHIP-mCA co-occurrences, we noted 20 enriched co-occurrences among 223 participants (19.3% of participants with both CHIP and autosomal mCAs), where the CHIP mutation was not detected within the bounds of a co-occurring mCA. The most notable co-occurrence observed was *MYD88*-chr18 (OR=165.6, 95% CI=(96.5–284.3),  $P=1.1\times 10^{-76}$ ), where *MYD88* L252P-whole chr18 gains predominated the specific events composing this co-occurrence (**Fig. 2c and Supplementary Data 5**). In refined analyses of mCA co-occurrence, *MYD88*-whole chr18 gains were further enriched (OR=267.8, 95% CI=(141.1-508.3),  $P=1.5\times 10^{-65}$ ); though CIs overlapped (**Supplementary Fig. 2 and Supplementary Data 6**). *MYD88*-chr13 also frequently co-occurred (OR=105.6, 95% CI=(64.9–171.6),  $P=7.7\times 10^{-79}$ ), primarily involving chr13q loss (**Fig 2c and Supplementary Data 5**). 27 (93.1%) participants with non-overlapping enriched *JAK2*-mCA co-occurrences (*JAK2*-chr1, *JAK2*-chr8, *JAK2*-chr14, *JAK2*-chr20) had more than 1 mCA. Many participants (58.3%) with non-overlapping *ASXL1*-chr4qCNLOH also had overlapping *TET2*-chr4qCNLOH mutations, emphasizing the relationship between these two epigenetic regulators. No overlapping or co-occurring CHIP-autosomal mCA co-occurrences occurred less frequently than expected.

For enriched CHIP-mCA co-occurrences, we examined the clonal fractions of component CHIP and mCA events to provide insight into the clonal history of CH mutations. In general, estimated CHIP clonal fraction was higher than estimated mCA clonal fraction in the case of both overlapping (92.6%) and non-overlapping CH (86.7%) (**Supplementary Fig. 3**). The relationship was most pronounced in co-occurrences with *JAK2* CHIP mutations, suggesting that the acquisition of this CHIP mutation likely preceded the acquisition of the associated autosomal mCA; although these analyses were performed in the absence of longitudinal data and assume clones with higher observed fraction develop earlier.

#### *TOPMed replication of CH co-occurrences*

A total of 3,300 CHIP mutations were detected in 3,019 (5.8%) TOPMed participants. Variant calling identified CHIP in 49 genes, as compared to 533 in the UKBB, reflecting the different CHIP calling and sequencing methods used for TOPMed (Online Methods). Similar to UKBB, CHIP mutations were most frequently detected in *DNMT3A*, *TET2*, and *ASXL1* (**Supplementary Fig. 4a**). In addition, a total of 9,963 mCAs were detected in 9,008 (13.4%) TOPMed participants. mCAs were frequently detected on the sex chromosomes (mLOY and mLOX) and chromosomes 11, 1, and 9 (**Supplementary Fig. 4b**).

Of the 3,019 TOPMed participants with detectable CHIP, 245 (8.1%) had more than one detectable CHIP mutation. Most CHIP-CHIP co-occurrences were comparable to those observed in the UKBB with 6 significantly enriched ( $P\leq 8.7\times 10^{-5}$ ): *U2AF1-JAK2*, *SRSF2-TET2*, *JAK2-TET2*, *JAK2-ASXL1*, *ASXL1-TET2*, and *DNMT3A-TET2* (**Supplementary Fig. 5a, Supplementary Fig. 6a, and Supplementary Data 7**). *U2AF1-JAK2* had the greatest magnitude of enrichment (OR=237.0, 95% CI=68.8–816.8,  $P=4.6\times 10^{-18}$ ) (**Supplementary Fig. 5a and Supplementary Table 7**). *TET2* mutations were again distributed across the gene and frequently observed as concurrent point mutations within individuals.

Investigating mCA-mCA co-occurrences in TOPMed, we noted that 754 individuals had more than one mCA, 414 of which had multiple autosomal mCAs. Enrichment testing identified 41 mCA-mCA co-occurrences that occurred more frequently than expected ( $P\leq 8.7\times 10^{-5}$ )

(**Supplementary Fig. 5b and Supplementary Data 8**). This represents fewer enriched co-occurrences than observed in the UKBB, likely attributable to the difference in sample size of each cohort. Of these, 39 (95.1%) were also identified in the UKBB with similar magnitude of enrichment (**Supplementary Fig. 6b**). The strongest enrichment was observed for co-occurrences that often involved whole chromosome gains, such as chr18-chr3 (OR=70.5, 95% CI=36.3–136.8,  $P=3.0\times 10^{-36}$ ) and chr19-chr18 (OR=67.6, 95% CI=29.6–154.6,  $P=1.6\times 10^{-23}$ ) (**Supplementary Data 8**). No mCA-mCA co-occurrences occurred less frequently than expected.

We examined CHIP-mCA co-occurrences, which were detected in 716 (1.4%) TOPMed participants. In agreement with UKBB results, we noted that *DNMT3A*-mLOY was observed less frequently than expected (OR=0.6, 95% CI=(0.5-0.8),  $P=2.7\times 10^{-4}$ ), though this association was no longer significant when corrected for multiple comparisons (**Supplementary Fig. 5c and Supplementary Data 9**). For CHIP and autosomal mCAs, 4 overlapping co-occurrences emerged from enrichment analyses ( $P\leq 8.7\times 10^{-5}$ ): *DNMT3A*-chr2, *TET2*-chr4, *JAK2*-chr9, and *TP53*-chr17 (**Supplementary Fig. 5c and Supplementary Data 9**). Non-overlapping CH co-occurrences were also observed for *TET2*-chr14, *JAK2*-chr11 and *TET2*-chr9; the last two of which were not observed in UKBB. The greatest overall enrichment among CHIP-mCA co-occurrences in TOPMed was observed for *JAK2*-chr9 (OR=193.6, 95% CI=(119.7–313.1),  $P=3.4\times 10^{-102}$ ) (**Supplementary Fig. 5c and Supplementary Data 9**), replicating the striking enrichment observed in the UKBB (OR=481.4); albeit with attenuated effect (**Supplementary Fig. 6c**). Most genomically overlapping co-occurrences consisted of partial arm CNLOH where the CHIP mutation was located (e.g., *DNMT3A*-2pCNLOH, *JAK2*-9pCNLOH, *TET2*-4qCNLOH). No CHIP-autosomal mCA co-occurrences were detected less frequently than expected in TOPMed.

#### *PLCO replication of CH co-occurrences*

We performed exome sequencing on 259 PLCO participants with detectable mCAs. We detected CHIP mutations in 83 (32.0%) individuals. Observed CHIP-mCA co-occurrences included *ASXL1*-chr14, *DNMT3A*-chr2, *DNMT3A*-chr4, *MYD88*-chr3, *MYD88*-chr12, *MYD88*-chr13, *TET2*-chr11, and *TET2*-chr14, further replicating previously identified CH co-occurrences (**Supplementary Data 10**).

#### *Characterization of co-occurring CH in UKBB participants with prior cancer*

We also characterized CH in 24,634 UKBB participants with a prior cancer diagnosis (**Supplementary Data 11**). A total of 1,876 (7.6%) participants harbored 2,115 CHIP events which is 1.7 times the frequency of cancer-free individuals (**Fig. 1c**); this elevation remained significant in multivariable models adjusted for potential confounders (OR=1.4, 95% CI=(1.3-1.5),  $P=3.4\times 10^{-40}$ ). *CALR*, *JAK2*, *ATM*, *MYD88*, *PPM1D*, *TP53*, *KMT2D*, *NF1*, *ASXL1*, *TET2*, and *DNMT3A* CHIP mutations were more frequently observed in prior cancer cases (**Supplementary Data 12**). Autosomal mCAs were also detected at a higher frequency with 1,289 participants (5.2%) harboring 1,685 events (OR=1.5 compared to participants with no prior cancer) (**Fig. 1d**). mCAs on 20 autosomes were more frequent, with chr13 mCAs, composed

primarily of losses, showing the greatest increase in frequency (OR=3.5, 95% CI=2.9–4.0,  $P=8.9\times 10^{-54}$ ) (**Supplementary Data 12**).

Co-occurring CH was also more common in prior cancer cases (3.6%) with patterns similar to those without prior cancer; although sample size was reduced for identifying enriched co-occurrences. All previously identified co-occurring CHIP-CHIP mutations (*DNMT3A-TET2*, *TET2-SRSF2*, *TET2-JAK2*) were also observed in the prior cancer population (**Supplementary Fig. 7a and Supplementary Data 13**). 88 (30.8%) of 286 unique mCA-mCA co-occurrences reached statistical significance ( $P\leq 8.7\times 10^{-5}$ ) (**Supplementary Fig. 6b and Supplementary Data 14**). Of the 28 CHIP and autosomal mCA co-occurrences previously described, 8 were enriched in prior cancer cases (**Supplementary Fig. 7c and Supplementary Data 15**). *TP53*-chr17 mCAs were more frequent in the prior cancer population (OR=60.8, 95% CI=(27.2–135.8),  $P=1.2\times 10^{-23}$ ) compared to those with no prior cancer (OR=26.6, 95% CI=(20.8–33.0),  $P=5.4\times 10^{-28}$ ); although the difference was not statistically significant (**Supplementary Fig. 8**). In contrast, *JAK2*-chr9 co-occurrences were detected less frequently in those with prior cancer (OR=209.3, 95% CI=(121.2–361.5,  $P=6.8\times 10^{-82}$ , **Supplementary Fig. 8**). 2 enriched CHIP-mCA co-occurrences emerged in the prior cancer population that were not previously identified (*SF3B1*-chr13 and *TP53*-chr14) ( $P\leq 4.8\times 10^{-5}$ ) (**Supplementary Fig. 7c and Supplementary Data 15**).

#### *Risk factors for CH co-occurrences*

To investigate whether participant characteristics associated with CHIP or autosomal mCAs (e.g., age, sex, smoking status) were associated with CHIP-autosomal mCA co-occurrence, we performed meta-analyses of UKBB and TOPMed data. Increasing age exhibited the most significant association (OR = 1.14 per one year increase in age, 95%CI=(1.13-1.15),  $P=2.5\times 10^{-174}$ ), though the effect size was comparable to that observed for CHIP only (**Fig. 3 and Supplementary Data 16**). In contrast, ever smoking (OR = 1.3, 95% CI=(1.2–1.5),  $P=3.6\times 10^{-8}$ ) and male sex (OR=1.7, 95% CI=(1.5–1.9),  $P=5.8\times 10^{-26}$ ) were more strongly associated with CHIP-autosomal mCA co-occurrence than either CH type alone (**Fig. 3 and Supplementary Data 16**). BMI was modestly inversely associated with co-occurrence (OR=0.98, 95% CI=(0.97-1.00),  $P=6.4\times 10^{-3}$ ) (**Fig. 3 and Supplementary Data 16**).

#### *The effect of co-occurring CH on telomere length*

We evaluated the combined effects of CH co-occurrences relative to their individual component events and to individuals without detectable CH on measured leukocyte telomere length (LTL), an indicator of cellular replication, in UKBB participants without prior cancer. We observed evidence for telomere degradation in individuals with overlapping CHIP-mCA events ( $\beta = -0.46$ , 95% CI=(-0.56– -0.35),  $P=2.0\times 10^{-17}$ ), with 11 enriched CHIP-mCA co-occurrences associated with shorter telomeres (**Fig. 4a and Supplementary Data 17**). Individuals with *JAK2*-chr20 harbored the shortest telomeres ( $\beta = -1.9$ , 95% CI=(-2.5– -1.3),  $P=5.1\times 10^{-10}$ ) (**Fig. 4a and Supplementary Data 17**).

We next compared the effect of co-occurrences to their component events (e.g., *JAK2*-chr9 vs. *JAK2* CHIP only or chr9 mCA only) to identify potential compounded effects of CH co-

occurrences. For example, telomere shortening was observed in participants with *JAK2* only ( $\beta = -0.4$ , 95% CI= $-0.6$ – $-0.2$ ,  $P=7.5 \times 10^{-4}$ ) and chr9 mCAs only ( $\beta = -0.1$ , 95% CI= $-0.2$ – $-0.02$ ),  $P=1.4 \times 10^{-2}$ ) (**Supplementary Data 17**); however, those with overlapping *JAK2*-chr9 co-occurrences had substantially shorter telomeres ( $\beta = -0.8$ , 95% CI= $-0.9$ – $-0.5$ ),  $P=1.1 \times 10^{-12}$ ) (**Supplementary Table 17**).

#### *Hematopoietic-related phenotypes in individuals with CHIP and mCAs*

Based on established reference ranges, we determined the percentage of UKBB participants without a prior cancer who had blood counts outside normal hematologic ranges to identify potential biological impacts of CHIP-mCA co-occurrences relative to component events. Among individuals without CH, 21% had abnormal white blood cell, red blood cell, or platelet counts. Similar elevated proportions were observed in those with CHIP only (23%) or mCAs only (23%) (**Supplementary Fig. 8**). In contrast, 52% of participants with overlapping co-occurrences had cell counts outside these ranges, including 66% of individuals with *JAK2*-chr9 (**Supplementary Fig. 9**).

For myeloid lineage traits, the greatest effects were observed for CH co-occurrences including *ASXL1*, *JAK2*, *NFE2*, *SRSF2*, or *TET2*. *JAK2* co-occurrences (*JAK2*-chr1, *JAK2*-chr9, *JAK2*-chr14, *JAK2*-chr20) had significantly elevated platelet counts, with *JAK2*-chr9 participants having the most elevation ( $\beta = 2.7$ , 95% CI= $(2.5$ – $2.9)$ ,  $P=1.0 \times 10^{-159}$ ) (**Fig. 4b and Supplementary Data 17**). Participants with *JAK2* co-occurrences also had notable increases in monocyte, neutrophil, eosinophil, and basophil count. In contrast, participants with *TET2* co-occurrences had decreased platelet, neutrophil, and eosinophil count (**Fig. 4b and Supplementary Data 17**). When comparing the effect of a co-occurrence compared to their component CHIP and mCA events, co-occurrences impacted myeloid lineage traits to a greater degree (**Supplementary Data 17**). Furthermore, participants harboring co-occurrences with high mCA cellular fraction ( $\geq 0.1$ ) generally had greater alterations in myeloid lineage traits. For example, participants with high cellular fraction *JAK2*-chr14 co-occurrences had evidence for higher platelet counts than those with low cellular fraction *JAK2*-chr14 co-occurrences ( $\beta_{\text{high\_cf}} = 1.5$ , 95% CI= $(0.8$ – $2.3)$ ,  $P=8.5 \times 10^{-5}$ ;  $\beta_{\text{low\_cf}} = 1.1$ , 95% CI= $(0.4$ – $1.8)$ ,  $P=2.0 \times 10^{-3}$ ) (**Supplementary Data 17**).

For lymphoid lineage traits, we noted 13 CH co-occurrences with significant effects ( $P \leq 3.33 \times 10^{-3}$ ) (**Fig. 4b and Supplementary Data 17**). *MYD88* co-occurrences were associated with increased lymphocyte counts with standardized effect sizes ranging from 1.2 to 6.0 (**Supplementary Data 17**). In comparing these effect sizes to component events, *MYD88* CHIP only and chr13 mCAs had associations with increased lymphocytes ( $\beta=1.8$ , 95% CI= $(1.6$ – $1.9)$ ,  $P=9.8 \times 10^{-85}$ ,  $\beta=1.4$ , 95% CI= $(1.4$ – $1.5)$ ,  $P < 2.2 \times 10^{-308}$ , respectively), but the effect was profoundly stronger for *MYD88*-chr13 co-occurrences ( $\beta=8.4$ , 95% CI= $(8.1$ – $8.6)$ ,  $P < 2.2 \times 10^{-308}$ ) (**Supplementary Data 17**). Furthermore, participants with high cellular fraction *MYD88*-chr13 had higher elevations in lymphocyte counts compared to those with low cellular fraction *MYD88* ( $\beta_{\text{high\_cf}}=10.9$ , 95% CI= $(10.6$ – $11.2)$ ,  $P < 2.2 \times 10^{-308}$ ,  $\beta_{\text{low\_cf}}=1.6$ , 95% CI= $(1.1$ – $2.1)$ ,  $P=1.1 \times 10^{-10}$ ) (**Supplementary Data 17**). Interestingly, participants with 3 of the 5 *JAK2* co-occurrences also had significant associations with respect to lymphocyte count; however, in this case, lymphocyte

count decreased (e.g.,  $JAK2$ -chr9: $\beta$ =-0.39, 95% CI=(-0.51–0.27),  $P=9.5\times 10^{-11}$ ) (**Supplementary Data 15**).

### *Myeloid malignancy risk*

During follow-up, 1,135 (0.3%) UKBB participants without prior cancer developed an incident myeloid malignancy. Of these, 531 (47%) individuals had detectable CH at baseline. As anticipated, participants with CHIP only had the greatest risk of myeloid malignancy compared to those with autosomal mCAs only (hazard ratio (HR)<sub>CHIP\_only</sub>=5.5, 95% CI=(4.8–6.4),  $P=6.6\times 10^{-113}$ ; HR<sub>mCA\_only</sub>=3.0, 95% CI=(2.3–3.8),  $p=2.2\times 10^{-18}$ ) (**Supplementary Data 18**). For general CH co-occurrences, risk was notably elevated in those with CHIP-CHIP co-occurrences compared to those with mCA-mCA co-occurrences (HR<sub>CHIP-CHIP</sub>=36.4, 95%CI=(29.3–45.2),  $P=6.4\times 10^{-230}$ ). Overlapping CH conferred the greatest risk (HR=80.6, 95% CI=(61.7–105.4),  $P=9.78\times 10^{-226}$ ), exceeding that of those with CHIP and any autosomal mCA (HR=38.0, 95% CI=(31.8–47.3),  $P=6.3\times 10^{-288}$ ) (**Supplementary Data 18**).

Of the 28 enriched CHIP-mCA co-occurrences, 6 were shown to have elevated risk of myeloid malignancies ( $P\leq 8.8\times 10^{-4}$ ): *ASXL1*-chr4, *JAK2*-chr9, *JAK2*-chr14, *JAK2*-chr20, *NFE2*-chr9, and *TET2*-chr4. Individuals with *NFE2*-chr9 co-occurrences ( $N_{\text{cases}}=6$ ) had the most extreme risk (HR=649.0, 95% CI=(290.4–1,450.4),  $P=4.09\times 10^{-56}$ ), with a median time-to-event (TTE) of 5 years (**Fig. 5b**). *JAK2* co-occurrences had the next strongest risk, with *JAK2*-chr14 co-occurrences ( $N=10$ ) with the shortest median TTE of 2.9 years (HR=573.0, 95% CI=(306.3–1,071.8),  $P=6.5\times 10^{-88}$ ) (**Fig. 5a and Supplementary Data 18**). Overlapping *JAK2*-chr9 ( $N_{\text{cases}}=40$ ) had a median TTE of 4.2 years (HR=233.6, 95% CI=(170.3–320.4),  $P=6.0\times 10^{-251}$ ) (**Fig. 5b and Supplementary Data 18**).

When comparing risk of co-occurrences to component events, it was noted that risk is further potentiated by co-occurring CH. For example, participants with *JAK2* V617F only ( $N = 75$ ) started with high levels of risk (HR=225.0, 95% CI=(154.75–327.14),  $P=5.8\times 10^{-177}$ ); however, individuals with *JAK2*-mCA co-occurrences had higher estimates of risk (**Supplementary Data 18**). In comparison, participants with only chr9 or chr14 mCAs had substantially lower HRs of 27.2 and 5.0, respectively (**Supplementary Data 18**).

We additionally evaluated whether myeloid malignancy risk differed by cellular fraction in individuals with *JAK2*-chr9, the only co-occurrence with sufficient sample size to support this analysis. Although the hazard ratio was higher for individuals with high cellular fraction *JAK2*-chr9 events ( $N_{\text{cases}}=73$ ; HR=230.1, 95% CI=(160.0–331.0),  $P=5.8\times 10^{-189}$ ), the confidence intervals directly overlapped with those of the low cellular fraction group ( $N_{\text{cases}}=25$ ; HR=218.1, 95% CI=(117.1–406.4),  $P=1.6\times 10^{-64}$ ), suggesting similar risk estimates (**Supplementary Data 18**).

Lastly, we conducted sensitivity analyses excluding participants with white blood cell or platelet counts that deviated substantially from normal reference ranges to account for potential undiagnosed disease at baseline (Online Methods). Co-occurrences previously nominated as high risk remained significantly associated with similar estimates of risk (**Supplementary Data 18**).

### *Lymphoid malignancy risk*

A total of 4,319 (1.0%) UKBB participants who were cancer-free at baseline developed an incident diagnosis of a lymphoid malignancy; 1,614 (37.3%) of which had detectable CH at

baseline. As anticipated, individuals with autosomal mCAs only had greater risk of lymphoid malignancies compared to those with CHIP only ( $HR_{mCA\_only}=4.5$ , 95% CI=(4.1–5.0),  $P=1.0\times 10^{-185}$ ;  $HR_{CHIP\_only}=1.4$ , 95% CI=(1.3–1.6),  $P=3.4\times 10^{-8}$ ) (**Supplementary Data 19**). Of general CH co-occurrences, risk was greater for those with mCA-mCA co-occurrences compared to those with CHIP-CHIP co-occurrences ( $HR_{mCA\_only}=8.5$  vs.  $HR_{CHIP\_only}=3.4$ ) (**Fig. 5d**). Risk of lymphoid malignancies was again further elevated in individuals with overlapping CH, compared to those with CHIP and any autosomal mCA ( $HR=11.8$ , 95% CI=(8.7–16.1),  $P=7.2\times 10^{-56}$ ) (**Fig. 5d**).

Of the 28 enriched CH co-occurrences, 4 were shown to have elevated risk of lymphoid malignancies: *MYD88*-chr3, *MYD88*-chr13, *MYD88*-chr18, and *TP53*-chr13. Participants with *TP53*-chr13 ( $N_{cases} = 5$ ) had the shortest median TTE of 2.2 years, as well as the highest risk ( $HR=124.7$ , 95% CI=(51.9–299.8),  $P=4.1\times 10^{-27}$ ) (**Fig. 5c,d**). *MYD88*-chr13, *MYD88*-chr18, and *MYD88*-chr3 had the next greatest risk associations ( $HRs=43.4$ -62.3;  $N_{cases}=13, 8, 8$ , respectively) (**Fig. 5d**). Of these, *MYD88*-chr18 participants had the shortest median TTE of 5.2 years (**Fig. 5c**).

In comparing the risk of these co-occurrences to their component events, we noted that participants with *TP53* only had negligible risk of lymphoid malignancy ( $HR=1.6$ , 95% CI=(0.6–4.3),  $P=0.3$ ) and participants with only chr13 mCAs had moderate levels of risk ( $HR=16.5$ , 95% CI=(13.1–20.7),  $P=6.2\times 10^{-127}$ ) compared to those with *TP53*-chr13 (**Supplementary Data 19**). As such, it was unanticipated that these co-occurrences had such amplified risk ( $HR=124.7$ ). We were unable to evaluate lymphoid malignancy risk by mCA cellular fraction due to insufficient sample size. Similar to myeloid malignancy risk estimates, those co-occurrences associated with increased risk of lymphoid malignancy remained associated with high risk in sensitivity analyses that removed individuals with elevated white blood cell and platelet counts (**Methods, Supplementary Data 19**).

## Discussion

Understanding the progression of detectable CH clones to malignant states is crucial in light of the prevalence of CH for aging adult populations. Most prior studies of CH have examined either CHIP or mCAs, limiting ability to assess the combined effects of co-occurring or overlapping CH<sup>4,5,17–19</sup>. Our integrative analysis of detectable CHIP and mCAs in UKBB, TOPMed and PLCO identified co-occurring CH in approximately 1.6% of adults in the general population. We characterized previously unreported co-occurrences and observed strong evidence for enrichment of co-occurrence, with some clones detected orders of magnitude more frequently than anticipated. Our findings detail the persistence of highly selected co-occurring clones. We identified alterations in telomere length and hematological parameters, as well as pronounced elevations in risk of myeloid and lymphoid malignancies for specific CH co-occurrences.

Our results build on prior observations of CH co-occurrence<sup>2,3,15,16,20</sup>. First, our findings demonstrate that the co-occurrence of multiple CH types within a single clone is not only tolerated but may also confer a selective advantage, demonstrated by the substantially higher frequencies of certain CH co-occurrences relative to the marginal frequencies of individual events. Overlapping CH could be part of a “two-hit” model at the genomic locus in which normal copies of a gene are lost<sup>21</sup>. Clonal fraction data indicates that CHIP events are present at higher cell fraction and likely occur first at the locus. This first coding driver mutation could confer an initial selective advantage of the clone. Subsequently acquired mCAs at the locus can delete the normal copy of the gene or amplify the mutated copy, resulting in a clone with amplified selective potential. Genetic mechanisms selective for non-overlapping CH events are less clear but potentially impact multiple interrelated genes within a pathway to confer a selective advantage. Some CH co-occurrences, however, are observed less frequently than expected, as is the case for *DNMT3A*-mLOY. The reduction in frequency of such co-occurrences suggests potential clonal competition or that the presence of both events is not tolerated within a single clone.

Second, CH co-occurrence has higher rates of cellular expansion than either component CHIP or mCA events. This is observed not only in higher cell fractions of co-occurring clones but also through greater reductions in telomere length, an indicator of cellular replication. This finding indicates cells with co-occurring CH can support the necessary cellular replication to expand and maintain this clonal expansion. Such a predilection for clonal expansion could lead to concerning growth trajectories as well as promote increased genomic instability that predisposes to the acquisition of additional mutations that further drive growth and expansion.

Third, the clonal expansion of co-occurring CH leads to marked phenotypic manifestations. We noted key disruptions in the balance of hematopoietic compartment in individuals with co-occurring CH. Many imbalances were lineage-specific such that some co-occurrences lead to more pronounced alterations in the myeloid compartment whereas others were more associated with alterations in lymphoid lineage traits. Several investigated co-occurrences conferred greater hematopoietic alterations than the cumulative effects of their component events, further confirming the importance of assessing CHIP and mCAs simultaneously.

Finally, our analysis of a large prospective cohort enabled the compilation of a list of CH co-occurrences that are associated with high risk of progression to hematologic malignancies

with hazard ratios exceeding 40. The six CH co-occurrences that confer such risk of myeloid malignancy progression include *ASXL1*-chr4, *JAK2*-chr9, *JAK2*-chr14, *JAK2*-chr20, *NFE2*-chr9<sup>22,23</sup>, and *TET2*-chr4. The CH co-occurrences that confer such risk of lymphoid malignancy progression include *MYD88*-chr3, *MYD88*-chr13, *MYD88*-chr18, and *TP53*-chr13. Surveillance for the presence and expansion of these clones could enable earlier detection and intervention in patients at risk of hematologic malignancies, especially in light of the short time to events observed for some co-occurrences.

Although the co-occurrence of CHIP and mCAs is suggestive of shared clonal origin, particularly for events detected at overlapping genomic loci, our analyses were based on bulk sequencing and genotyping data and therefore cannot definitively resolve clonality. In instances where CHIP mutations and mCAs are detected at low cellular fractions, it remains possible that these alterations arose in independent clones within the same individual. Nevertheless, the magnitude of enrichment observed for many of the identified CH co-occurrences is more consistent with a shared clonal origin than with independent clonal events occurring by chance. Accordingly, our findings should be interpreted as population-level evidence of non-random co-occurrence that is consistent with, but does not conclusively establish, shared clonal architecture.

A few limitations of this study warrant consideration. Differences in CH detection methods across UKBB, TOPMed, and PLCO could introduce variability in the sensitivity for low frequency events and alterations in specific genomic regions not captured well in the library preparation. Despite this being the largest study of its kind, limited sample sizes for many co-occurrences may reduce our power to identify associations and affect the precision of risk measurements. The cross-sectional assessment of CH limits our ability to examine the temporality in the acquisition of CH event types and interpret patterns of clonal evolution. In comparing VAF and cellular fraction, we did not account for local ploidy, which may further confound interpretations of mutational timing. However, the prospective nature of the UK Biobank was a key strength for identifying potential associations with future malignancy risk. Although the inclusion of TOPMed improved representation of individuals similar to non-European ancestry reference groups, the generalizability of our findings to these populations remains limited. Lastly, we acknowledge the possibility of undiagnosed hematologic malignancies at baseline, which could bias hazards ratios due to the inclusion of individuals with pre-existing but clinically unrecognized disease. Sensitivity analyses were performed that removed individuals with elevated white blood cell and platelet counts to address this potential bias.

Our findings highlight the importance to remain vigilant to the complexities of multiple concurrent CH events, the forces governing selection and clonal expansion of these clones, and the multitude of potential phenotypic consequences. Future population-level, multi-omic, and clinical research is crucial for further disentangling the clinical relevance of co-occurring CH clones.

## Methods

All studies included in this work were conducted in accordance with relevant ethical regulations and were approved by the appropriate institutional review boards or ethics committees. All participants provided written informed consent. Study-specific details regarding ethical approvals and consent procedures are provided in the corresponding cohort descriptions and references.

### UK Biobank

#### *Ethics statement*

The research conducted in this study complies with all relevant ethical regulations. UK Biobank (UKBB) has approval from the Northwest Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. All participants provided informed consent to participate in the UK Biobank.

#### *Study participants*

The UK Biobank is a prospective study with more than half a million participants between 37-73 years of age. Participants were recruited between 2006 and 2010 in 22 assessment centers throughout the UK. Baseline assessment visits were comprised of electronic consent, a touch-screen questionnaire, computer-assisted interview, physical and functional measurements, and collection of specimens including blood, urine, and saliva. A more in-depth description on the UK Biobank study design and cohort is provided by Sudlow et al., as well as at <https://www.ukbiobank.ac.uk><sup>24</sup>. Within the present study, analyses were performed on 453,807 UKBB subjects, after removing those who had withdrawn from the study, failed genotyping QC or exhibited sex discordance. In addition, those with reported cancer (ICD-10 C codes) within inpatient and cancer registry records prior to study enrollment were removed from primary analyses but were later evaluated in analyses restricted to cancer survivors. Genetic similarity to 1000 Genomes reference populations (CEU, YRI, ASN) was estimated in SNPweights to adjust for confounding effects by ancestry<sup>25</sup>.

#### *Detection of CH*

For CHIP, somatic mutations were called within whole-exome sequencing (WES) data using the Artifact filtering Clonal Hematopoiesis (ArCH) variant calling pipeline for detecting single nucleotide variants (SNVs) and short insertions/deletions (INDELs)<sup>26</sup>. Mutect2 (v4.2.1.0)<sup>27</sup>, VarDict (v1.6.0) were used for variant calling, following base quality score recalibration with GATK's BQSRPipelineSpark (v4.2.0.0)<sup>27-29</sup>. Variants supported by both variant callers, with a variant allele frequency (VAF) >2 % and at least 2 supporting reads from forward and reverse strands were retained. To remove potential false positive calls due to sequencing error and artifact, we performed extensive filtering. Fisher's exact tests were performed to test that mutant read depths were significantly higher relative to panel of normal samples (PON) (N=50) under the age of 21 without any prior known CH hotspot mutations. P-values were required to reach stringent Bonferroni-correction thresholds based on the number of bases in the exome capture design to be considered significant. A recurrent filter was applied such that variants appearing in >10% of the samples were flagged as potential sequencing errors. Pre-compiled lists of CH driver genes were used to rescue potential variants flagged by the

recurrent filter. Additionally, known germline mutations as annotated in the gnomAD database with allele frequency greater than 0.0005 were filtered out as potential germline variants. Further filters removed low variant allele fraction (VAF) Mutect2 calls, synonymous mutations, variants with high strand bias, variants not reaching minimum alternative allele counts, and homopolymers regions. Only variants passing all filters and classified as putative drivers were retained in the final analytic set of CHIP variants. Variants were classified as putative drivers based on knowledge from COSMIC, oncoKB, and ClinVar databases as well as utilizing a list of previously identified putative driver mutations and rules in Bolton et al and Bick et al<sup>30,31</sup>. Finally, all CHIP analyses were restricted to genes mutated in least 30 participants (N=43) as power was limited to assess genes less frequently impacted by CHIP. For clonal fraction comparison, CHIP cellular fraction was calculated by  $2 \times \text{VAF}$ , with a ceiling of 1.0; although, this assumes no chromosomal losses or gains spanning the variant.

Previous studies have described the detection of mCAs in the UKBB<sup>8,17</sup>. DNA was extracted from baseline peripheral blood samples followed by genotyping on the UK BiLEVE or UK Biobank Axiom array. Genotype intensities were transformed to log<sub>2</sub>R ratio (LRR) and B-allele frequency (BAF) values to estimate relative allelic intensities and allelic imbalances. Long-range phase information was utilized for assigning phase for BAF deviations. MoChA, an mCA detection algorithm which employs a three-state hidden Markov model to analyze LRR and BAF data, in conjunction with genetic phase, was utilized to detect mCAs (<https://github.com/freeseek/mochawdl>), assign copy number state and estimate cellular fraction.

For all subsequent analyses, the presence of both CHIP and mCAs was treated as binary. Individuals were classified as either having (1) or not having (0) each event type, regardless of the number of co-occurrences. For CHIP, this was at the gene level; for mCAs, this was at the chromosome level. Additionally, all p-values from subsequent statistical analyses were derived from two-sided tests.

#### *Characterization of co-occurring CH*

Upon detection of both mCAs and CHIP within the population, data were merged to determine co-occurrence and overlap. Logistic regression models were used to examine if co-occurrences were more frequent than expected based on individual characteristics. Models were fit using the glm() function in R (v4.3.1) and were adjusted for age, age<sup>2</sup>, sex, smoking status (categorized as never, former, or current) and genetic similarity. ORs were estimated by exponentiating the regression coefficient from these models. Corresponding 95% CIs were calculated by exponentiating the coefficient  $\pm 1.96 \times$  its standard error (SE). Co-occurring CH was categorized in three groups: CHIP-CHIP (co-occurrence of CHIP mutations in two different genes), mCA-mCA (co-occurrence of mCAs on two different chromosomes), and CHIP-mCA (co-occurrence of a specific CHIP mutation and an mCA). For analyses of CHIP-CHIP and mCA-mCA co-occurrence, regression models were constructed such that CHIP mutations in each gene and mCAs on each chromosome were tested as both a predictor and outcome variable. For CHIP-mCA analyses, CHIP was tested as the predictor of the mCA outcome variable. Given that our analyses were based on bulk sequencing and genotyping data, co-occurrence was defined as the presence of multiple CH types within the same individual. Overlapping CH was defined as a CHIP mutation falling within the bounds of a co-occurring mCA.

To address multiple comparisons, a Bonferroni threshold of significance for each co-occurring CH group (CHIP x CHIP, mCA x mCA, CHIP x mCA) was determined. This was

calculated by  $0.05 / (\text{number of groups in CH type 1} \times \text{number of groups in CH type 2})$ . For example, within the co-occurring CHIP and mCA group, this was calculated by  $0.05 / (24 \times 43)$ . Any enriched CHIP-mCA co-occurrences with  $P \leq 4.8 \times 10^{-5}$  that was detected in at least 5 participants were selected for inclusion in subsequent phenotypic association analyses.

For refined mCA co-occurrence analyses, mCAs were additionally grouped by chromosome region (p, q, both, or whole chromosome) and event type (gain, loss, CNLOH). Logistic regression models were constructed analogously to the chromosome-level analyses, testing mCAs as both predictors and outcomes and adjusting for age, age<sup>2</sup>, sex, smoking status (never, former, current), and genetic similarity. Bonferroni-adjusted significance thresholds were determined based on the number of arm-level comparisons in which  $\geq 5$  participants exhibited co-occurrence. For mCA-mCA co-occurrences, this threshold was  $0.05 / 102 = 4.9 \times 10^{-4}$ ; for CHIP-mCA co-occurrences, this threshold was  $0.05 / 66 = 7.6 \times 10^{-4}$ .

## Replication in TOPMed

### *Study participants*

A total of 67,390 participants from 19 studies within the National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) program were included in our analyses: Genetics of Cardiometabolic Health in the Amish (Amish; N=1,109)<sup>32</sup>, Atherosclerosis Risk in Communities Study (ARIC, N=3,870)<sup>33</sup>, Barbados Genetics Asthma Study (BAGS, N=980), Mount Sinai BioMe Biobank (BioMe, N=9,392)<sup>34</sup>, Coronary Artery Risk Development in Young Adults (CARDIA, N=3,293)<sup>35</sup>, Cleveland Family Study (CFS, N=1,281), Cardiovascular Health Study (CHS, N=3,517)<sup>36</sup>, Genetic Epidemiology of COPD Study (COPDGene, N=10,050)<sup>37</sup>, Framingham Heart Study (N=4,007)<sup>38</sup>, Genetic Studies of Atherosclerosis Risk (GeneSTAR, N=1,733)<sup>39</sup>, Genetic Epidemiology Network of Arteriopathy (GENOA, N=1,157), Genetics of Lipid Lowering Drugs and Diet Network (GOLDN, N = 942), Hispanic Community Health Study – Study of Latinos (HCHS/SOL, N=3,857)<sup>40</sup>, Hypertension Genetic Epidemiology Network (HyperGEN, N=1,865), Jackson Heart Study (JHS, N=3,317)<sup>41</sup>, Multi-Ethnic Study of Atherosclerosis (MESA, N=5,222)<sup>42</sup>, The Vanderbilt Atrial Fibrillation Registry (VU\_AF, N=1,085), Women’s Genome Health Study (WGHS, N=108), and Women’s Health Initiative (WHI, N=10,695)<sup>43</sup>. Genetic similarity proportions for the percentage of African, Asian, and European ancestry were inferred for each participant using GrafPop<sup>44</sup>. Cancer status of these participants was not available for all contributing studies. All studies received institutional review board (IRB) approval, and informed consent was obtained from all participants.

### *Detection of CH*

Whole genome sequencing was generated as part of the NHLBI TOPMed program<sup>45</sup>. Detection of of CHIP and mCAs in TOPMed has been reported in prior studies<sup>16,31</sup>. A total of 25,051 males were interrogated for mLOY in TOPMed. mLOX was called as ChrX events, defined as losses by MoChA, that spanned at least 100 Mb in length of ChrX.

### *Characterization of co-occurring CH*

We replicated the co-occurrence analysis in TOPMed using the same methodology as described for UKBB participants. CHIP and mCA data were merged and logistic regression models, adjusted for age, age<sup>2</sup>, sex, smoking status (never or ever), genetic similarity, and

TOPMed study, were fit using the `glm` function in R (v4.4.2). CHIP-CHIP, mCA-mCA, and CHIP-mCA co-occurrences were evaluated. For those including CHIP, 24 specific CHIP genes were chosen for inclusion, listed in **Supplementary Data 20**<sup>16</sup>. To address multiple testing, a Bonferroni correction was again applied using a uniform threshold across co-occurrence types. This was calculated as  $0.05 / (24 \times 24) = 8.7 \times 10^{-5}$ .

### PLCO replication

#### *Study participants*

The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) was a controlled, prospective study conducted across multiple sites in the United States<sup>46</sup>. Between 1993 and 2001, 154,807 individuals between the ages of 55 of 74 were enrolled at 10 screening centers. Participants provided biospecimens and completed detailed surveys on demographic information, family and medical history, and lifestyle and dietary risk factors. For those in the screening arm, blood samples were collected at baseline and during annual follow-up visits for up to five years. Written, informed consent was provided by all participants, and the study received approval from all relevant Institutional Review Boards.

#### *DNA Preparation*

Genomic DNA (200ng per sample) was purified using Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA, USA) following the manufacturer's instructions. Adapter-ligated libraries were then prepared using the KAPA HyperPlus Kit (KAPA Biosystems, Wilmington, MA) in conjunction with Bioo Scientific NEXTflex™ DNA Barcoded Adapters (Bioo Scientific, Austin, TX, USA), according to the KAPA-provided protocol.

#### *Pre-Hybridization LM-PCR*

Prior to hybridization, genomic DNA sample libraries were amplified via ligation-mediated PCR (LM-PCR). Each reaction included 20μL of library DNA, 25μL of 2x KAPA HiFi HotStart ReadyMix, and 5μL of 10x Library Amplification Primer Mix, containing primers with the sequences: 5'-AATGATACGGCGACCACCGA-3' and 5'-CAAGCAGAAGACGGCATACGA-3'. PCR conditions consisted of an initial denaturation at 98°C for 45 seconds, followed by 5 cycles of 98°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 1 minute. Reactions were maintained at 4°C until further processing. Amplified sample libraries were cleaned using the Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA, USA), according to the KAPA-provided protocol, and quantified using Quant-iT™ PicoGreen dsDNA Reagent (Life Technologies, Carlsbad, CA, USA).

#### *Liquid Phase Sequence Capture*

Amplified sample libraries with unique barcodes were pooled in equal amounts to form 1.1 μg input for multiplexed sequence capture. Exome sequence capture was performed using NimbleGen's SeqCap EZ Human Exome Library, Exome+UTR with 96Mb targeted (Roche NimbleGen, Inc., Madison, WI, USA). Prior to hybridization, the following components were added to the 1.1 μg pooled sample library, 4 μL of NEXTflex HE Universal Oligo 1, 250 μM (5'-AATGATACGGCGACCACCGAGATCTACTCTTTCCCTACACGACGCTCTTCCGATCT-3'), 40 μL total of 25 μM NEXTflex INV-HE blocking oligos, equal volumes of each blocking oligo complementary to the barcodes in the pool (5'-

CAAGCAGAAGACGGCATAACGAGATXGTGACT

GGAGTTCAGACGTGTGCTCTTCCGATCT/C3 Spacer/-3', where X is 8-bases of sequence specific to adapter barcode used for library construction), and 5  $\mu$ L of 1 mg/mL COT-1 DNA (Invitrogen, Inc., Carlsbad, CA, USA). Samples were dried in an Eppendorf 5301 Vacuum Concentrator (Eppendorf, Hauppauge, NY, USA) at 60°C for approximately 1 hour. The dried pool was resuspended in 7.5  $\mu$ L of NimbleGen Hybridization Buffer and 3.0  $\mu$ L of NimbleGen Hybridization Component A, incubated at 95°C for 10 minutes, and then hybridized with 4.5  $\mu$ L of EZ Exome Probe Library at 47°C for 64 to 72 hours. Post-hybridization washing and DNA recovery were performed according to the NimbleGen SeqCap EZ Library SR Protocol.

#### *Post-Hybridization LM-PCR*

Captured DNA pools were amplified by in LM-PCR reactions consisting of 20 $\mu$ L captured library DNA, 25  $\mu$ L of 2x KAPA HiFi HotStart ReadyMix, and 5 $\mu$ L of 10x Library Amplification Primer Mix, containing the same two primers previously described. PCR cycling conditions were: 98°C for 45 seconds, followed by 8 cycles of 98°C for 15 seconds, 60°C for 30 seconds, 72°C for 30 seconds, with a final step of an extension at 72°C for 1 minute. Reactions were held at 4°C until further processing. Amplified material was purified with Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA, USA) in accordance with the NimbleGen SeqCap EZ Library SR Protocol. Quantification of amplified, captured libraries were performed using Kapa's Library Quantification Kit for Illumina (Kapa Biosystems, Woburn, MA, USA) on the LightCycler 480 (Roche, Indianapolis, IN, USA).

#### *Sequencing*

Resulting post-capture, enriched, and multiplexed libraries were used for cluster generation on an Illumina cBOT (Illumina, San Diego, CA, USA). Paired-end sequencing (2 x 150 bp) was performed on the Illumina HiSeq 4000 following Illumina-provided protocols.

#### *Detection of CH*

CHIP variant calling, annotation, and filtering for somatic mutations was performed within whole exome sequencing (WES) data from 259 participants using Version 2.2.0 of the ArCH workflow<sup>26</sup>. Somatic mutations were detected using Mutect2<sup>27</sup>, VarDict<sup>29</sup>, and Lofreq<sup>47</sup>, whereas Pindel<sup>48</sup> was used for long and complex indels. Variant annotation was performed through VEP<sup>49</sup> using data from various resources, such as TOPMed<sup>45</sup>, MSK-IMPACT<sup>50</sup>, COSMIC<sup>51</sup>, and OncoKB<sup>52</sup>. Fisher's exact tests were used to determine significantly higher mutant read depth to total depth in study participants relative to a PON (N=18) composed of samples  $\leq$ 60 years of age. We applied a pre-compiled list of CHIP driver genes to recover any putative variants that may have been inadvertently flagged by the recurrent filter<sup>26</sup>. Any germline mutations identified in the gnomAD<sup>53</sup> database with an allele frequency greater than 0.0005 (0.05%) were filtered as likely germline variants. Additional filters were applied to exclude low VAF calls from Mutect2, synonymous mutations, variants exhibiting high strand bias, those with insufficient alternative allele counts, and variants located in homopolymer regions. The final analytic set included only high-confidence CHIP variants that passed all filtering criteria and were classified as putative driver mutations.

For mCAs, 110,562 participants were scanned with one of four genotyping arrays (Illumina Global Screening Array, OncoArray, Omni2.5 M, and OmniExpress). These samples

met quality thresholds for harmonization and imputation<sup>54</sup>. mCAs were again identified using MoChA with an mCA calling threshold  $\geq 0.97$  and a BAF  $\leq 0.03$ .

#### *Characterization of co-occurring CH in PLCO*

Following detection of CHIP and mCAs, data were merged to determine co-occurrences. All participants with prior cancers were removed from these analyses. We then summarized the co-occurrence of CHIP and autosomal mCA events by counting the number of individuals with each possible combination. Given limited sample size with CHIP calls, statistical evaluation of enrichment was not performed.

#### *Characterization of CH in UKBB participants with prior cancer*

Following the characterization of co-occurring CH in UKBB individuals with no prior cancer, similar analyses were run on UKBB participants with prior cancer. After excluding participants with non-melanoma neoplasms of the skin (N=9,705), this population consisted of 24,634 participants. Logistic regression models, adjusted for age, age<sup>2</sup>, sex, smoking status (never, former, current) and genetic similarity to reference populations, were used to compare the frequencies of specific CHIP mutations and mCAs by chromosome in both the prior and no prior cancer populations. Analyses comparing mLOX and mLOY frequencies were restricted to females and males, respectively, and sex was removed as a covariate.

#### *Risk factors for CH co-occurrences*

To investigate risk factors associated with CH, we used logistic regression models to evaluate associations between participant characteristics and the presence of CH in three mutually exclusive groups: (1) CHIP only, (2) autosomal mCAs only, and (3) co-occurring CHIP and autosomal mCAs. Analyses were conducted separately in TOPMed and UK Biobank participants. We then performed a random effects meta-analysis to combine results using `metafor()` in R (v4.4.2). The primary predictors of interest were participant characteristics including age, sex, and smoking status. BMI was also included but could only be investigated in UKBB participants. All sex-related estimates compared males to females. All smoking estimates compared ever smokers to never smokers. Age and BMI were treated as continuous variables, whereas sex and smoking status were treated as categorical variables. Models were adjusted for the following covariates, unless they were the variable under investigation: age, age<sup>2</sup>, sex, smoking status, and genetic similarity.

#### Phenotypic Associations in UKBB participants without prior cancer

##### *Hematopoietic phenotypes*

All data for hematopoietic phenotypes were extracted from the UK Biobank (Category 100081: Blood count).

We first determined whether individuals with each enriched CHIP-mCA co-occurrence had blood cell counts within normal ranges. To do this, we utilized the reference ranges quoted by the manufacturer in Table 1 of the UK Biobank Haematology Data Companion Document (<https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/haematology.pdf>). Some parameters of interest (nucleated red blood cell count, high light scatter reticulocyte count, mean reticulocyte volume, mean spheroid cell volume, platelet crit, platelet distribution width) did not have reference ranges provided and were excluded from this portion of our analyses.

All resulting blood count measures were standardized in R (v4.3.1). Multivariable linear regression adjusted for age, age<sup>2</sup>, sex, genetic similarity and smoking status was used to evaluate the association between co-occurring CH groups and 22 hematopoietic-related phenotypes.

**Supplementary Data 21** outlines all phenotypes examined as well as the associated UK Biobank Field IDs. The Z-adjusted T/S log (UKBB Field 22191) was used within this model for associations with telomere length; full details of the measures and adjustments are found in Codd et al<sup>21</sup>.

To assess the impact of co-occurring CH groups on hematopoietic phenotypes, we employed a stepwise breakdown within co-occurrence groups. The objective was twofold: first, to evaluate the effect of enriched co-occurrences, and second, to understand how the co-occurrence alters the hematopoietic landscape in comparison to individual instances of CH. By comparing the effect of co-occurrence to individual occurrences, we examined if co-occurring CH had super-additive effects on blood cell counts relative to the individual additive effects of each contributing CH.

To determine the impact of high vs. low mCA cellular fraction on hematopoietic phenotypes, some *JAK2* and *MYD88* co-occurrences (*JAK2*-chr9, *JAK2*-chr14, *MYD88*-chr3, *MYD88*-chr13, *MYD88*-chr18) were broken down into two groups. mCA cellular fraction was denoted “high” if it was  $\geq 0.1$ .

To address multiple comparisons, we assessed the number of effective independent tests amongst the hematopoietic phenotypes using `meff()` in R (v4.3.1). 15 independent tests were determined. The significance threshold was then calculated the significance threshold by  $0.05/n_{\text{eff}}$ , or  $0.05/15$ , yielding  $3.33 \times 10^{-3}$ .

#### *Hematological malignancy association*

Using available data within the UK Biobank, we extracted cancer information from both inpatient and cancer registry data. Incident hematological cancers were defined as occurring after study enrollment. **Supplementary Data 22** details hematologic disease-associated ICD-10 codes as lymphoid or myeloid cancers. We performed Cox proportional hazards regression using the `coxph` package in R (v4.3.1) to assess the risk of hematological malignancies (any hematological malignancy, lymphoid malignancy, myeloid malignancy) across CH groups adjusting for age, age<sup>2</sup>, sex, genetic similarity and smoking status (Never, Former, Current). Hazards ratios and 95% confidence intervals were estimated. The stepwise and cellular fraction evaluations of CH groups described in the analyses of hematopoietic phenotypes was likewise utilized within the scope of hematological malignancy associations. Co-occurrences were only stratified by cellular fraction if the resulting groups had at least 5 remaining cancer cases.

To address multiple comparisons, we determined a Bonferroni threshold of significance based on the number of co-occurrences, component events, and general CH types within the model. This calculation determined a *P*-value threshold of  $7.90 \times 10^{-4}$ .

#### *Statistics and reproducibility*

This study utilized data from large, previously established cohorts (UKBB, TOPMed studies, PLCO). Sample sizes are based on data availability within each cohort; no statistical method was used to predetermine sample size. Participants were excluded based on quality control criteria as described above. This study was based on observational data and did not involve randomization. All statistical analyses were performed in R. All statistical tests were two-sided, and multiple testing was addressed using Bonferroni correction or estimation of the

effective number of independent tests, as described above. Key findings were evaluated in independent cohorts where data was available.

### Data Availability

The data used in the present study were obtained from UKBB under application numbers 92005 and 55288 and are available upon application to the UKBB (<https://www.ukbiobank.ac.uk>).

Data for each participating TOPMed study can be accessed through dbGaP with their corresponding TOPMed accession numbers. All genomic datasets that can be shared will be deposited in dbGaP under controlled access. Controlled access and data use restrictions are based on individual contributing TOPMed study approval. The relevant studies and dbGaP accession numbers are for each participating study are: Amish (phs000956.v5.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000956.v5.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000956.v5.p1)], ARIC (phs001211.v5.p4) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001211.v5.p4](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001211.v5.p4)], BAGS (phs001143.v4.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001143.v4.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001143.v4.p1)], BioMe (phs001644.v3.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001644.v3.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001644.v3.p2)], CARDIA (phs001612.v3.p3) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001612.v3.p3](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001612.v3.p3)], CFS (phs000954.v4.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000954.v4.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000954.v4.p2)], CHS (phs001368.v4.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001368.v4.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001368.v4.p2)], COPDGene (phs00951.v6.p5) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000951.v6.p5](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000951.v6.p5)], FHS (phs000974.v6.p5) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000974.v6.p5](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000974.v6.p5)], GeneSTAR (phs001219.v3.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001219.v3.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001219.v3.p1)], GENOA (phs001345.v3.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001345.v3.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001345.v3.p1)], GOLDN (phs001359.v3.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001359.v3.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001359.v3.p1)], HCHS/SOL (phs001395.v3.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001395.v3.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001395.v3.p2)], HyperGEN (phs001293.v3.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001293.v3.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001293.v3.p1)], JHS (phs000964.v5.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000964.v5.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000964.v5.p1)], MESA (phs001416.v4.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001416.v4.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001416.v4.p1)], VU\_AF (phs001032.v6.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001032.v6.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001032.v6.p2)], WGHS (phs001040.v6.p1) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001040.v6.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001040.v6.p1)], and WHI (phs001237.v4.p2) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001237.v4.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001237.v4.p2)].

For PLCO, data can be accessed through dbGaP with the following accession numbers: phs001286.v4.p2 (genotyping) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001286.v4.p2](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001286.v4.p2)] and phs001286 (sequencing) [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001286.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001286.v1.p1)].

### Code Availability

Mosaic chromosomal alterations were detected using MoChA pipelines (mocha.wdl), which is publicly available at <https://doi.org/10.5281/zenodo.10892520><sup>55</sup>. For the most updated and detailed version, visit <https://github.com/freeseek/mochawdl>. All other analyses were performed using standard and previously published methods as described in the **Methods**.

### References

1. Merlo, L. M. F., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* **6**, 924–935 (2006).
2. Brown, D. W. *et al.* Shared and distinct genetic etiologies for different types of clonal hematopoiesis. *Nat. Commun.* **14**, (2023).
3. Gao, T. *et al.* Interplay between chromosomal alterations and gene mutations shapes the evolutionary trajectory of clonal hematopoiesis. *Nat. Commun.* **12**, (2021).
4. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* **371**, (2014).
5. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* **371**, (2014).
6. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, (2012).
7. Kessler, M. D. *et al.* Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature* **612**, (2022).
8. Loh, P. R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* **584**, (2020).
9. International Lung Cancer Consortium (INTEGRAL-ILCCO) *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).

10. Liu, A. *et al.* Genetic drivers and cellular selection of female mosaic X chromosome loss. *Nature* **631**, 134–141 (2024).
11. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, (2012).
12. Lin, S. H. *et al.* Incident disease associations with mosaic chromosomal alterations on autosomes, X and Y chromosomes: insights from a phenome-wide association study in the UK Biobank. *Cell Biosci.* **11**, (2021).
13. Schick, U. M. *et al.* Confirmation of the Reported Association of Clonal Chromosomal Mosaicism with an Increased Risk of Incident Hematologic Cancer. *PLoS ONE* **8**, (2013).
14. Terao, C. *et al.* Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature* **584**, (2020).
15. Saiki, R. *et al.* Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. *Nat. Med.* **27**, (2021).
16. Jakubek, Y. A. *et al.* Mosaic chromosomal alterations in blood across ancestries using whole-genome sequencing. *Nat. Genet.* **55**, 1912–1919 (2023).
17. Loh, P. R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, (2018).
18. Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742–752 (2017).
19. Machiela, M. J. *et al.* Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am. J. Hum. Genet.* **96**, 487–497 (2015).

20. Brown, D. W. *et al.* Germline-somatic JAK2 interactions are associated with clonal expansion in myelofibrosis. *Nat. Commun.* **13**, 5284 (2022).
21. Knudson, A. G. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* **68**, 820–823 (1971).
22. Jutzi, J. S. *et al.* MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. *J. Exp. Med.* **210**, 1003–1019 (2013).
23. Jutzi, J. S. *et al.* Altered NFE2 activity predisposes to leukemic transformation and myelosarcoma with AML-specific aberrations. *Blood* **133**, 1766–1777 (2019).
24. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, (2015).
25. Chen, C. Y. *et al.* Improved ancestry inference using weights from external reference panels. *Bioinformatics* **29**, (2013).
26. Chan, I. C. C. *et al.* ArCH: improving the performance of clonal hematopoiesis variant calling and interpretation. *Bioinformatics* **40**, btae121 (2024).
27. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. Preprint at <https://doi.org/10.1101/861054> (2019).
28. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
29. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **44**, e108–e108 (2016).
30. Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).

31. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
32. Mitchell, B. D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: Rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *Am. Heart J.* **155**, 823–828 (2008).
33. THE ARIC INVESTIGATORS. THE ATHEROSCLEROSIS RISK IN COMMUNIT (ARIC) STUDY: DESIGN AND OBJECTIVES. *Am. J. Epidemiol.* **129**, 687–702 (1989).
34. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).
35. Hughes, G. H. *et al.* Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) study. *Control. Clin. Trials* **8**, 68–73 (1987).
36. Fried, L. P. *et al.* The cardiovascular health study: Design and rationale. *Ann. Epidemiol.* **1**, 263–276 (1991).
37. Regan, E. A. *et al.* Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J. Chronic Obstr. Pulm. Dis.* **7**, 32–43 (2011).
38. Splansky, G. L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, Recruitment, and Initial Examination. *Am. J. Epidemiol.* **165**, 1328–1335 (2007).
39. Becker, D. M. *et al.* Sex Differences in Platelet Reactivity and Response to Low-Dose Aspirin Therapy. *JAMA* **295**, 1420 (2006).
40. Sorlie, P. D. *et al.* Design and Implementation of the Hispanic Community Health Study/Study of Latinos. *Ann. Epidemiol.* **20**, 629–641 (2010).

41. Taylor, H. A. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn. Dis.* **15**, S6-4–17 (2005).
42. Bild, D. E. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *Am. J. Epidemiol.* **156**, 871–881 (2002).
43. Design of the Women’s Health Initiative Clinical Trial and Observational Study. *Control. Clin. Trials* **19**, 61–109 (1998).
44. Jin, Y., Schaffer, A. A., Feolo, M., Holmes, J. B. & Kattman, B. L. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components Analysis. *G3 Bethesda Md* **9**, 2447–2461 (2019).
45. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
46. Hayes, R. B. *et al.* Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control. Clin. Trials* **21**, 349S-355S (2000).
47. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
48. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
49. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. Preprint at <https://doi.org/10.1101/042374> (2016).

50. Cheng, D. T. *et al.* Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT). *J. Mol. Diagn.* **17**, 251–264 (2015).
51. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
52. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* 1–16 (2017) doi:10.1200/PO.17.00011.
53. Chen, S. *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
54. Machiela, M. J. *et al.* GWAS Explorer: an open-source tool to explore, visualize, and access GWAS summary statistics in the PLCO Atlas. *Sci. Data* **10**, 25 (2023).
55. Genovese, G. MoChA WDL pipelines 2022-12-21. Preprint at <https://doi.org/10.5281/ZENODO.10892520> (2022).

### Acknowledgements

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. This work was also supported in part by the National Institutes of Health (Grants: R01HL146500 and R01HL165061). Acknowledgements for all contributing TOPMed studies are provided as supplementary information.

### Author Contributions

M.J.M. conceived the study. M.J.M., K.L.B., and P.L.A. supervised the study. K.M.B., A.K.H., R.L.K., C.D.Y., D.W.B., and S.J.C. contributed to or performed statistical analyses and interpretation. I.C.C.C., Y.C., and D.T. performed CHIP calling and data aggregation in UKBB. G.G. performed mCA calling in UKBB. W.Z. and W.S.W.W performed CH event calling in PLCO. W.Y. H., N.D.F., K.J., A.H., and B. H. contributed to PLCO design and conduct.

Y.A.J., D.A., K.C.B., J.C.B., E.B., J.A.B., A.P.C., D.I.C., M.H.C., P.D., M.F.D., M.F., X.G., N.H.C., M.R.I., A.D.J., S.L.R.K., C.K., D.L., J.P.L., Y.L., R.J.F.L., T.M.M., R.A.M., B.D.M., K.E.N., N.P., P.A.P., M.H.P., B.M.P., L.M.R., S.R., S.S.R., J.I.R., E.K.S., A.V.S., J.A.S., A.S., P.S., A.P.R., A.G.B., P.L.A. contributed to the design, conduct and CH calling of the contributing TOPMed studies. K.M.B. and M.J.M drafted the initial version of the manuscript. All authors reviewed and approved the paper.

### Competing Interests Statement

The authors declare the following competing interests: S.R. has consulted for Amgen, Inc. E.K.S. received institutional grant support from Bayer and Northpond laboratories over the past three years. Y.C. provided consultancy services to Need and Bayer for work outside the scope of the manuscript. M.H.C. has received grant support from Bayer and Genentech, and consulting fees from Apogee Therapeutics and BMS, unrelated to the current work. L.M.R. is a consultant for the TOPMed Administrative Coordinating Center (through Westat). S.S.R. is a consultant for Westat, the Administrative Coordinating Center for the NHLBI TOPMed program. The remaining Authors declare no competing interests.

### Figure Legends

#### Fig. 1. Frequencies of CHIP and autosomal mCA events among UK Biobank participants with and without prior cancer at baseline.

- Percentage of participants without prior cancer with detectable clonal hematopoiesis of indeterminate potential (CHIP), stratified by sex (female: orange; male: blue).
- Percentage of participants without prior cancer with autosomal mCAs
- Percentage of participants with prior cancer with detectable CHIP.
- Percentage of participants with prior cancer with detectable autosomal mCAs. Percentages indicate the fraction of participants within each group (n=453,807 without prior cancer; n=24,634 with prior cancer).

#### Fig. 2. Enriched CH co-occurrences in UKBB participants without prior cancer at baseline.

- Co-occurrence of mutations in common CHIP genes.
- Co-occurrence of mCAs.
- Co-occurrences of mutations in common CHIP genes and mCAs. Associations were tested using logistic regression adjusted for age, age<sup>2</sup>, sex, smoking status, and genetic ancestry; odds ratios (ORs) are shown for associations with  $P < 0.05$  (two-sided tests). Colored cells indicate associations meet Bonferroni-corrected thresholds (**a**,  $P \leq 2.7 \times 10^{-5}$ ; **b**,  $P \leq 8.7 \times 10^{-5}$ ; **c**,  $P \leq 4.8 \times 10^{-5}$ ). Blue denotes  $OR > 1$  with intensity proportional to effect size and red  $OR < 1$ . Bold boxes indicate the chromosomal location of each CHIP gene. Full results are provided in **Supplementary Data 2, 3 and 5**.

#### Fig. 3. Risk factors associated with CHIP-autosomal mCA co-occurrence.

Associations between participant characteristics (age, sex, and smoking status) and three CH categories: CHIP only (light blue), autosomal mCA only (dark blue), and CHIP and autosomal mCA co-occurrences (green). Points represent odds ratios (ORs) and error bars represent 95%

confidence intervals (CIs) from multivariable logistic regression models adjusted for age, age<sup>2</sup>, sex, smoking status and genetic similarity, where appropriate. Sex-related estimates use females as the reference group and smoking estimates compare ever versus never smokers. Analyses were performed at the level of individual participants. Estimates are from meta-analyses of UKBB (n=453,807) and TOPMed (n=67,390); all tests are two-sided. Full results are provided in **Supplementary Data 16**.

**Fig. 4. Associations between clonal hematopoiesis (CH) types, leukocyte telomere length, and hematopoietic-related phenotypes.**

Analyses were performed in UK Biobank participants without prior cancer (n=453,807). Sample sizes for each CH category vary.

**a**, Association between CH types and leukocyte telomere length (LTL). Points show  $\beta$  estimates and error bars represent 95% confidence intervals (CIs) from multivariable linear regression models adjusted for age, age<sup>2</sup>, sex, smoking status, and genetic similarity.

**b**, Associations between CH types and 22 hematopoietic traits. Colors represent standardized (mean=0, s.d.=1)  $\beta$  estimates from multivariable linear regression models adjusted as in **a**.  $\beta$  estimates are capped at  $|\beta| = 2$  for display. Asterisks indicate statistical significance (\*\* $P \leq 3.33 \times 10^{-3}$ , multiple-testing threshold; \* $P > 3.33 \times 10^{-3}$  and  $P \leq 0.05$ ; two-sided tests). Full results are provided in **Supplementary Data 17**.

**Fig. 5. Risk of myeloid and lymphoid malignancy associated with CH types.**

CHIP-mCA co-occurrences shown had  $\geq 5$  cases, hazards ratio (HR) $\geq 40$ , and  $P \leq 7.9 \times 10^{-4}$ .

**a**, Kaplan-Meier curves of time to myeloid malignancy by CH type

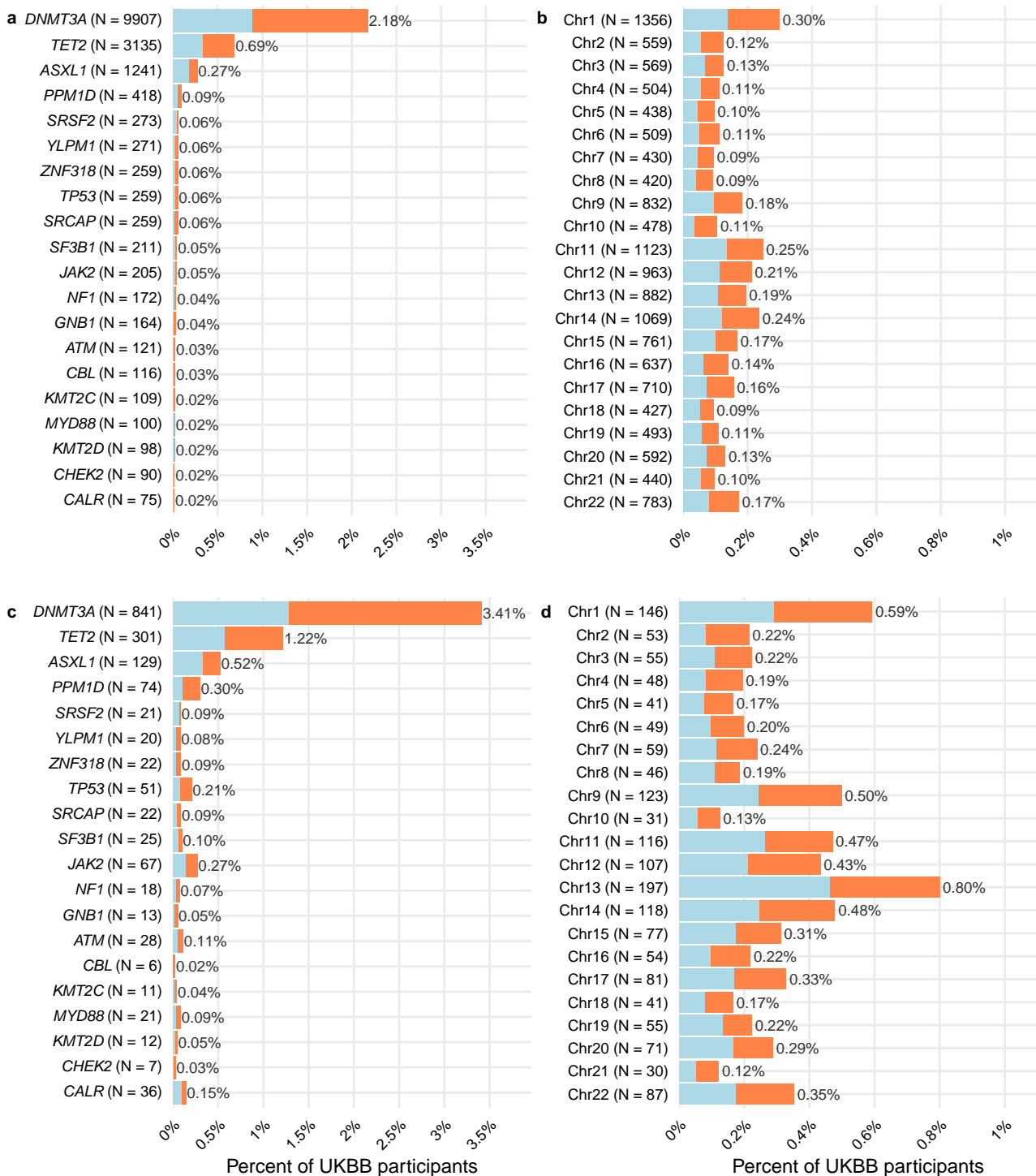
**b**, HRs and 95% confidence intervals (CIs) for myeloid malignancy from Cox proportional hazards models adjusted for age, age<sup>2</sup>, sex, smoking status and genetic similarity; two-sided Wald test  $P$ -values are shown.

**c**, Kaplan-Meier curves of time to lymphoid malignancy by CH type.

**d**, HRs and 95% CIs for lymphoid malignancy from Cox proportional hazards models adjusted as in **b**; two-sided Wald test  $P$ -values are shown. Full results are provided in **Supplementary Data 18 and 19**.

Results from integrative population-based investigations indicate co-occurring types of clonal hematopoiesis are highly enriched and markedly increase blood cancer risk, highlighting new opportunities for early detection and targeted surveillance of high-risk individuals.

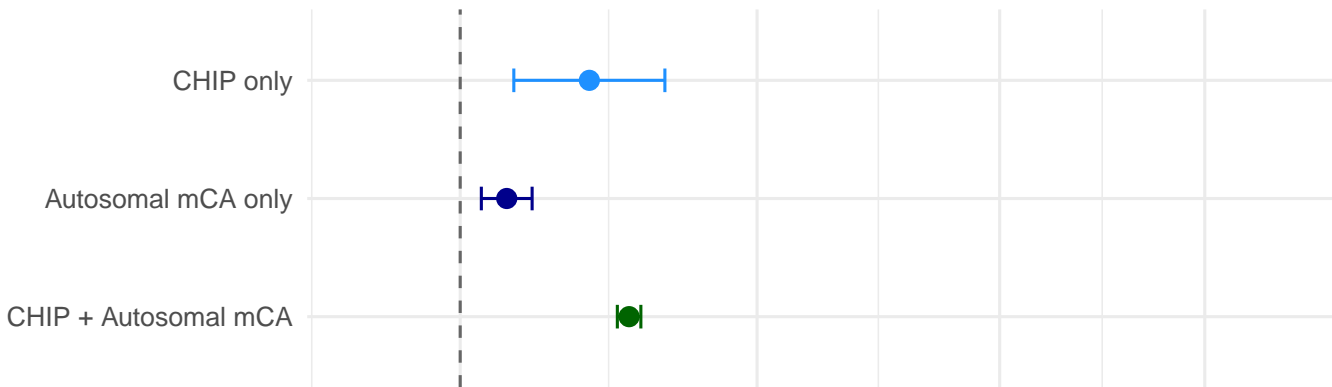
ARTICLE IN PRESS



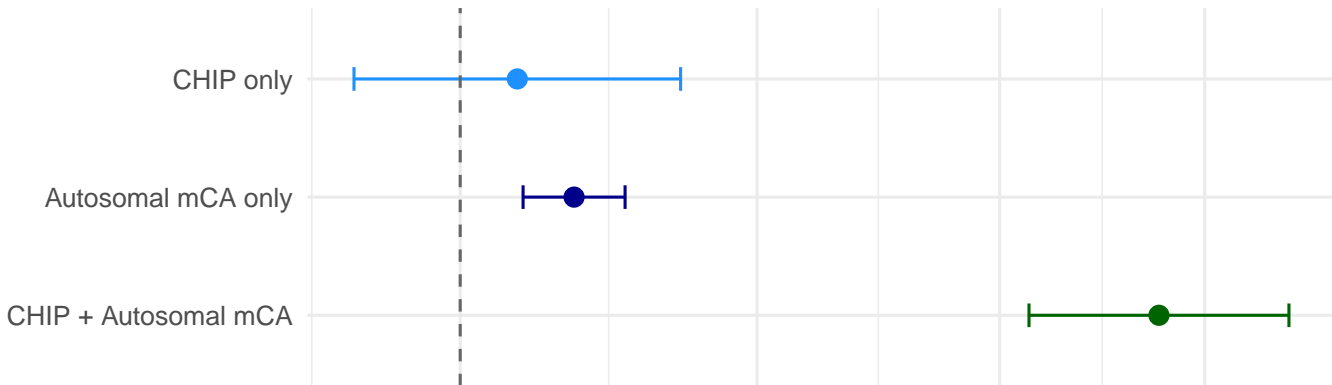
Sex ■ Female participants ■ Male participants



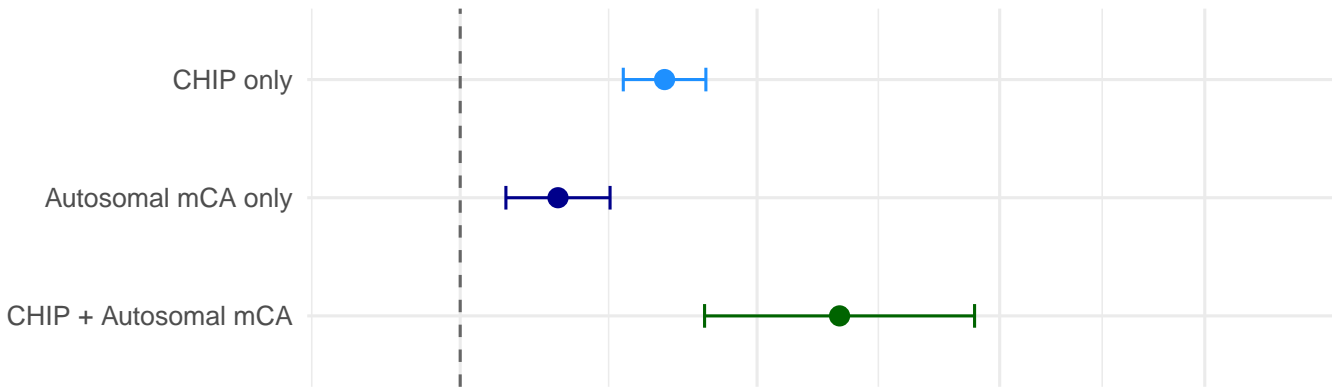
## Age



## Sex

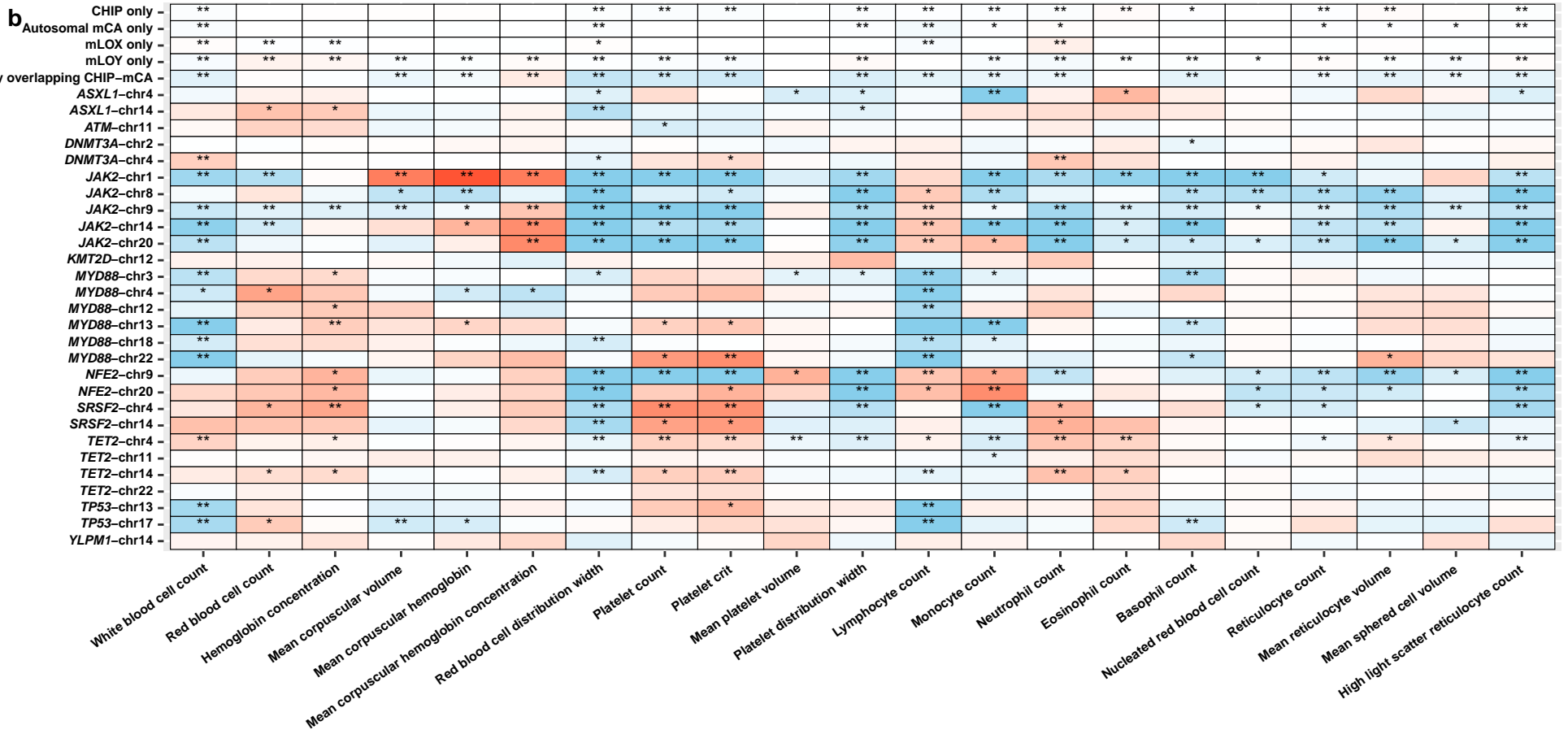
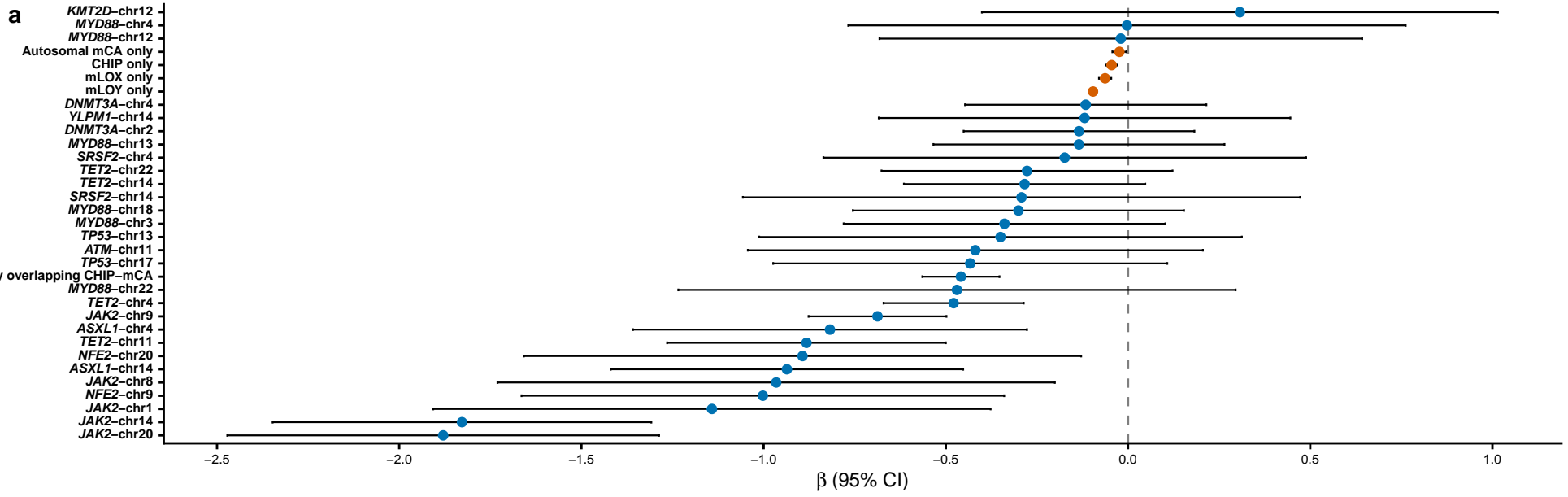


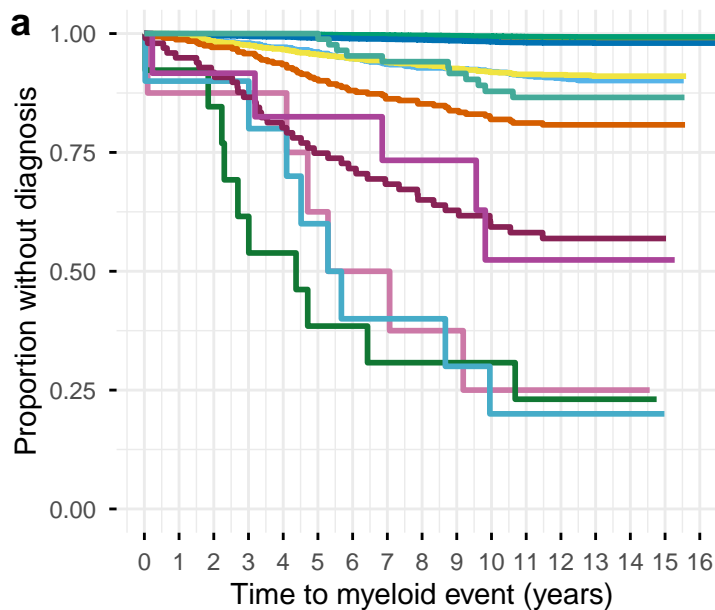
## Smoking



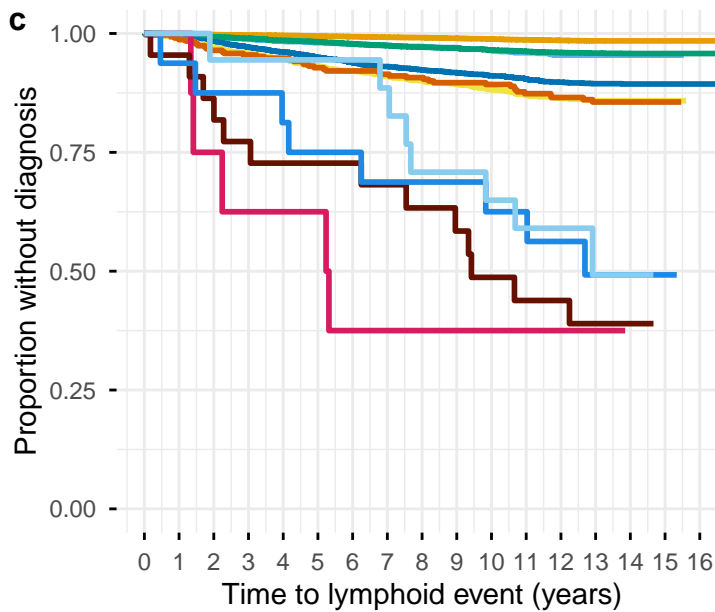
Odds Ratio (log scale, 95% CI)

Outcome ● CHIP only ● Autosomal mCA only ● CHIP + autosomal mCA



**b**

Risk of myeloid malignancy associated with CH types			
CH type	Cases	Hazards Ratio (95% CI)	p-value
CHIP only	201	5.5 (4.8-6.4)	$6.6 \times 10^{-113}$
CHIP-CHIP	88	36.4 (29.3-45.2)	$6.4 \times 10^{-230}$
Autosomal mCA only	66	3.0 (2.3-3.8)	$2.2 \times 10^{-18}$
mCA-mCA	65	7.0 (5.4-9.0)	$5.7 \times 10^{-52}$
CHIP-Autosomal mCA	108	38.8 (31.9-47.3)	$6.3 \times 10^{-288}$
Overlapping CHIP-mCA	56	80.6 (61.7-105.4)	$9.8 \times 10^{-266}$
NFE2-chr9	6	649.0 (290.4-1450.4)	$4.1 \times 10^{-56}$
JAK2-chr14	10	573.0 (306.3-1071.8)	$6.5 \times 10^{-88}$
JAK2-chr20	8	548.4 (272.7-1102.6)	$4.4 \times 10^{-70}$
JAK2-chr9	40	233.6 (170.3-320.4)	$6.0 \times 10^{-251}$
ASXL1-chr4	5	183.0 (75.9-441.4)	$4.1 \times 10^{-31}$
TET2-chr4	11	44.9 (24.8-81.3)	$4.0 \times 10^{-36}$

**d**

Risk of lymphoid malignancy associated with CH types			
CH type	Cases	Hazards Ratio (95% CI)	p-value
CHIP only	241	1.4 (1.3-1.6)	$3.4 \times 10^{-8}$
CHIP-CHIP	40	3.4 (2.5-4.7)	$1.0 \times 10^{-14}$
Autosomal mCA only	395	4.5 (4.1-5.0)	$1.0 \times 10^{-185}$
mCA-mCA	346	8.5 (7.6-9.5)	$1.6 \times 10^{-320}$
CHIP-Autosomal mCA	123	9.3 (7.8-11.2)	$3.2 \times 10^{-132}$
Overlapping CHIP-mCA	41	11.8 (8.7-16.1)	$7.2 \times 10^{-56}$
TP53-chr13	5	124.7 (51.8-299.8)	$4.1 \times 10^{-27}$
MYD88-chr13	13	62.3 (36.2-107.4)	$4.8 \times 10^{-50}$
MYD88-chr18	8	48.3 (24.1-96.6)	$6.3 \times 10^{-28}$
MYD88-chr3	8	43.4 (21.7-86.9)	$1.7 \times 10^{-26}$

CH type

