# Shared and distinct genetic etiologies for different types of clonal hematopoiesis

Derek W. Brown,[1,2]* Liam D. Cato,[3,4]* Yajie Zhao,[5]* Satish K. Nandakumar,[3,4,6] Erik L. Bao,[3,4] Thomas Rehling,[1] Lei Song,[1] Kai Yu,[1] Stephen J. Chanock,[1] John R. B. Perry,[5,7]^ Vijay G. Sankaran,[3,4]^ Mitchell J. Machiela[1]^

*These authors contributed equally
^These authors jointly supervised this work

**Author Affiliations**

1. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA
2. Cancer Prevention Fellowship Program, Division of Cancer Prevention, National Cancer Institute, Rockville, MD, USA
3. Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA
4. Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
5. MRC Epidemiology Unit, Wellcome-MRC Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK
6. Department of Cell Biology, Albert Einstein College of Medicine, Albert Einstein Cancer Center, Ruth L. and David S. Gottesman Institute for Stem Cell Research and Regenerative Medicine, Bronx, NY 10461, USA
7. Metabolic Research Laboratory, Wellcome-MRC Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK

**Keywords**: clonal hematopoiesis, CHIP, mosaic chromosomal alterations, telomere length

**Corresponding Authors**:
John R. B. Perry, Ph.D.: john.perry@mrc-epid.cam.ac.uk
Vijay G. Sankaran, M.D., Ph.D.: sankaran@broadinstitute.org
Mitchell J. Machiela, Sc.D., M.P.H.: mitchell.machiela@nih.gov

**Abstract**

Clonal hematopoiesis (CH) – age-related expansion of mutated hematopoietic clones – can differ in frequency and cellular fitness. Descriptive studies have identified a spectrum of events (coding mutations in driver genes (CHIP), gains/losses and copy-neutral loss of chromosomal segments (mCAs) and loss of sex chromosomes), some of which have been associated with disease risk. Co-existence of different CH events raises key questions as to their origin, selection, and impact. Analyses of sequence and genotype array data in up to 482,378 individuals from UK Biobank revealed shared genetic architecture across different types of CH. We observed co-occurrence of CHIP and mCAs with co-localization at *TET2*, *DNMT3A* and *JAK2*, in which CHIP precedes mCA acquisition. CH types are associated with cellular differentiation, DNA repair, and cell cycle regulation pathways, which can drive clonal expansion and reduce telomere length. Insights into the shared architecture of CH could improve detection and prevention of CH-related outcomes.

## Main

Recent studies have reported the frequent occurrence of clonal expansion of post-zygotic mutations in the hematopoietic system, now seen in all human tissues but at different attained frequencies.[1–6] Initially, clonal expansion was recognized by the presence of skewed X chromosome inactivation.[7,8] Subsequent studies have revealed the presence of mosaic chromosomal alterations (mCAs), including frequent loss of chromosomes X and Y, in a subset of hematopoietic cells. Most recently, clonal expansion of recurrent somatic driver mutations observed in hematologic malignancies have been identified in individuals with otherwise normal hematologic parameters, a condition known as clonal hematopoiesis of indeterminate potential (CHIP). These somatic alterations can predispose to either myeloid or lymphoid malignancies but do not necessarily progress; in other words, many otherwise healthy individuals are observed to have CHIP.[9] Moreover, recent studies have shown the additive impact of mCAs and CHIP mutations on predisposition to blood cancers, with respect to overall risk for a primary hematologic malignancy and also in the setting of therapy-associated myeloid malignancies.[10,11]

Large studies have begun to reveal how germline genetic variants can increase risk for acquisition of CH but have yet to systematically investigate the co-existence of events as well as determine a comprehensive catalog of germline susceptibility variants. Moreover, the biologic mechanisms of hematopoiesis that confer risk have not been well-characterized. Here, we utilized a range of genotyping and sequencing data from large-scale studies of genetic susceptibility of types of clonal hematopoiesis (*e.g.*, mCAs, CHIP), hematologic malignancies (*e.g.*, myeloproliferative neoplasms (MPNs)), and hematopoietic phenotypes to assess genetic

and phenotypic relationships between these distinct but potentially related phenotypes to gain new insights into the underlying mechanisms and consequences of the range of CH events.

## Results

*CH states display shared genetic and phenotypic relationships*

We began by investigating the co-existence of different types of CH: loss of chromosome Y (LOY) in men, loss of chromosome X (LOX) in women, autosomal mCAs including gains, losses, and copy neutral loss of heterozygosity (CNLOH), CHIP, and MPNs (**Figure 1**) and subsequently examined associations with germline susceptibility variants, in anticipation of discovery of shared elements. Genome-wide association study (GWAS) summary statistics for each type of CH were analyzed and pairwise genetic correlations between traits were computed (**Online Methods**). Positive genetic correlations for LOY with both LOX ($\rho$= 0.25, P= $5.76\times10^{-5}$) and MPN ($\rho$= 0.21, P= $1.67\times10^{-2}$) (**Figure 2**) were observed. No other statistically significant genetic correlations were observed between types of CH (**Figure 2**). The lack of genetic correlations with autosomal mCAs and CHIP could arise from our approach of aggregating events into overall categories and thus underestimate the impact of distinct genomic regions on clonal expansions, potentially resulting in low estimated genetic heritabilities ($h^2_{AUTO_mCA}$= 0.0004, $se_{AUTO_mCA}$= 0.0007; $h^2_{CHIP}$= 0.013, $se_{CHIP}$= 0.0074; **Supplemental Figure 1**, **Supplemental Table 1**).

In an analysis of 482,378 subjects from the UK Biobank, we investigated adjusted phenotypic associations between types of CH (**Online Methods**, **Supplemental Table 2**). Consistent with previous studies of CH, each type of CH investigated demonstrated a strong positive association

with age (**Supplemental Figure 2**). We observed an inverse phenotypic association between LOY and MPN (T-statistic= -4.76, P= $1.96\times10^{-6}$) (**Figure 3**), which is opposite in direction from the genetic correlation. We were unable to evaluate the phenotypic association between LOY and LOX, as these are sex specific traits.

There were positive phenotypic associations of autosomal mCAs with LOY (T-statistic= 4.31, P= $1.61\times10^{-5}$), LOX (T-statistic= 12.44, P= $1.54\times10^{-35}$), CHIP (T-statistic= 9.95, P= $2.47\times10^{-23}$), and MPN (T-statistic= 41.00, P< $5\times10^{-324}$) (**Figure 3**). Sensitivity analyses removed individuals with mCAs spanning the *JAK2* region (N=550), a region frequently impacted by mCAs in MPNs,[12–15] and still observed a positive phenotypic association between autosomal mCAs and MPN, though the association was attenuated (T-statistic= 15.53, P= $2.39\times10^{-54}$). CHIP was also positively associated with MPN (T-statistic= 12.50, P= $8.11\times10^{-36}$) and inversely associated with LOY (T-statistic= -3.81, P= $1.37\times10^{-4}$) (**Figure 3**).

We examined the cellular fraction of individuals with autosomal mCA events and the variant allele frequency (VAF) of individuals with CHIP mutations and observed that individuals with higher cellular fractions of autosomal mCA events (*i.e.*, greater proportion of cells carrying the somatic event) were more likely to have LOX (T-statistic= 3.08, P= $2.09\times10^{-3}$), CHIP (T-statistic= 4.58, P= $4.67\times10^{-6}$), and MPN (T-statistic= 17.83, P= $3.17\times10^{-70}$) (**Supplemental Figure 3**, **Supplemental Table 3**). Higher autosomal mCA cellular fraction was inversely associated with LOY (T-statistic= -7.76, P= $9.76\times10^{-15}$) (**Supplemental Figure 3**, **Supplemental Table 3**). Individuals with higher VAF of CHIP mutations (*i.e.*, higher clonal

fractions) were more likely to have detectable autosomal mCAs (T-statistic= 5.20, P= $2.09\times10^{-7}$) and MPNs (T-statistic= 6.88, P= $6.71\times10^{-12}$) (**Supplemental Figure 4**, **Supplemental Table 4**).

In an analysis of co-existence of types of CH, CHIP and autosomal mCAs significantly co-occurred in the same individual (hypergeometric P = $4.16\times10^{-28}$; **Supplemental Figure 5a**) with 365 individuals (6.4% of individuals with CHIP, 5.3% of individuals with autosomal mCAs) carrying both (**Supplemental Table 2**). Individuals with autosomal mCAs displayed a distinct pattern of CHIP gene mutations compared to individuals without autosomal mCAs (**Supplemental Figure 5b**, **Supplemental Table 5**). 10 CHIP gene mutations were significantly enriched in individuals with autosomal mCAs (*DNMT3A*, *TET2*, *TP53*, *SF3B1*, *JAK2*, *KDM6A*, *SRSF2*, *MPL*, *ETV6*, and *IDH1*); **Supplemental Figure 5b**), a similar pattern of co-occurrence as previously observed in cancer patients.[11] Additionally, individuals with CHIP mutations were more likely to acquire autosomal mCAs across 18 chromosomes (**Supplemental Figure 5c**, **Supplemental Table 6**), with enrichment for several chromosome-specific copy number states (*e.g.*, CNLOH in chromosomes 1, 4, and 9; **Supplemental Figure 5c**, **Supplemental Table 6**).

An evaluation of the 365 individuals with both a CHIP mutation and autosomal mCA revealed that 53 (14.5%) had events spanning the same genomic region (binomial P= $1.36\times10^{-13}$). 8 CHIP genes co-localized with autosomal mCAs, with *TET2* mutations accounting for 24 (42.9%) of the observed overlapping mutations (**Supplemental Figure 6**). CNLOH was the most frequently observed autosomal mCA event (N= 39 (69.6%)) among all overlapping mutations (**Supplemental Figure 6**). We examined the clonal fractions of both somatic mutations to provide a window into the clonal evolution of CHIP mutations and autosomal mCAs and found

higher estimated CHIP VAF than estimated mCA cellular fraction in a majority of co-localizing mutations, suggesting the acquisition of the CHIP mutation preceded the acquisition of autosomal mCAs (binomial $P=3.80\times10^{-3}$; **Supplemental Figure 7**); this finding is consistent with a multi-hit hypothesis in driving clonal evolution. This is particularly evident in loss-of-heterozygosity of chromosome 9 alterations after acquisition of a *JAK2 V617F* mutation, as has been seen in individuals with MPNs.[16] Subsequent autosomal mCA-induced loss of heterozygosity or amplification of CHIP driver mutations could confer strong selective advantages promoting rapid cellular expansion. The co-occurrence and co-localization of CHIP and autosomal mCAs motivates future studies that jointly assess both CH traits to better understand CH interactions that could confer increased clonal expansion and potentially elevated disease and mortality risk.[9]

Pathway-based analyses using GWAS summary statistics (**Online Methods**) utilized 6,290 curated gene sets and canonical pathways from Gene Set Enrichment Analysis (GSEA) and revealed significant associations between several biological pathways and types of CH (**Supplemental Figure 8**) with all types of CH associated with gene sets related to apoptosis, IL-2 signaling, DNA methylation, promyelocytic leukemia gene product (PML) targets, and cancer-related gene sets (**Supplemental Tables 7-11**). LOY, LOX, and MPN were significantly associated with hematopoietic progenitor cells, hematopoietic cell lineage and differentiation gene sets, and DNA damage response (**Supplemental Tables 7**, **8**, and **11**). LOY, autosomal mCAs, CHIP, and MPN were associated with telomere extension by telomerase, with LOY and MPN also associated with telomere stress induced senescence (**Supplemental Tables 7**, **9-11**). Additionally, the 12 genetic pathways significantly associated with autosomal mCAs were also

associated with CHIP, providing further evidence that these types of CH are interrelated. Overall, pathway analyses suggest core shared pathogenic mechanisms related to cellular differentiation, DNA damage repair, and cell cycle regulation that are critical for the development and clonal expansion of most types of CH.

*Correlation of types of CH with myeloid and lymphoid cell traits*

We examined genetic and phenotypic correlations between types of CH and 19 blood cell traits (**Figures 2-3**). LOY and MPN demonstrated genetic correlations enriched for myeloid lineage traits, namely, positive correlations with red blood cell, platelet, total white blood cell, and eosinophil counts ($P<0.025$; **Figure 2**). MPN was also positively correlated with other myeloid lineage traits including hematocrit, hemoglobin, and neutrophil count ($P< 0.005$; **Figure 2**), as previously reported.[17] LOY was additionally inversely correlated with mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), and mean reticulocyte volume ($P< 0.015$; **Figure 2**). In support of the genetic correlations, we observed strong phenotypic associations of LOY and MPN with myeloid traits that closely mirror the magnitude and significance of the genetic correlation results (**Figure 3**), and previously reported phenotypic associations.[18–20] Both LOY and MPN were positively associated with platelet count, monocyte count, and neutrophil count ($P< 6\times10^{-45}$; **Figure 3**). These findings suggest shared mechanisms for maintenance of myeloid lineage hematopoiesis likely also govern susceptibility to LOY and MPN.

We performed bidirectional two-sample Mendelian randomization (MR) analyses to investigate the directionality between LOY and the 19 blood cell traits. Using LOY GWAS genetic variants

as the instrumental variable (IV), MR inverse-variance weighted (MR-IVW) models showed a positive association between the LOY IV and platelet (*e.g.*, platelet count $Z_{filtered}$= 6.15, P= $1.13\times10^{-7}$), red blood cell (*e.g.*, red blood cell count $Z_{filtered}$= 5.51, P= $7.37\times10^{-7}$), and white blood cell counts (*e.g.*, white blood cell count $Z_{filtered}$= 7.98, P= $1.69\times10^{-11}$) (**Supplemental Figure 9**, **Supplemental Table 12**). The LOY IV was negatively associated with red blood cell traits including MCH, MCV, and mean reticulocyte volume (P< $5\times10^{-9}$). Similarly, IVs based on blood cell trait GWAS variants showed associations with LOY (*e.g.*, plateletcrit $Z_{filtered}$= 8.75, P= $4.50\times10^{-17}$; white blood cell count $Z_{filtered}$= 9.18, P= $2.21\times10^{-18}$) (**Supplemental Figure 10**, **Supplemental Table 13**). The bi-directional MR findings indicate shared genetic mechanisms between LOY and several blood cell traits (*e.g.*, plateletcrit, platelet count, red blood cell distribution width, white blood cell count); moreover, the effect of LOY on red blood cell traits (*e.g.*, red blood cell count, MCH, MCHC, and MCV) is notable.

LOX was the only type of CH to display a significant genetic correlation with lymphocyte count ($\rho$= 0.10, P= $4.15\times10^{-2}$; **Figure 2**), and a positive phenotypic association between LOX and lymphocyte count was also observed (T-statistic= 23.96, P= $9.11\times10^{-127}$) (**Figure 3**). LOX had a positive genetic correlation with myeloid traits such as red blood cell distribution width, basophil count, and eosinophil count, whereas an inverse correlation with hematocrit, hemoglobin, MCH, MCHC, and MCV (P< 0.05, **Figure 2**). LOX also had positive phenotypic associations with MCH, MCHC, MCV, and monocyte count (P< 0.015, **Figure 3**), and inverse associations with hematocrit, red blood cell count, and neutrophil count (P< 0.03, **Figure 3**).

We did not observe genetic correlations between autosomal mCAs or CHIP with blood cell traits (**Figure 2**). In the case of combined autosomal mCAs, there was evidence for positive phenotypic associations with both lymphocyte count (T-statistic= 60.33, $P< 5\times10^{-324}$) and total white blood cell count (T-statistic= 34.48, $P= 3.89\times10^{-260}$) (**Figure 3**). Combined CHIP was positively associated with platelet count (T-statistic= 6.59, $P= 4.44\times10^{-11}$), monocyte count (T-statistic= 6.15, $P=7.94\times10^{-10}$), and neutrophil count (T-statistic= 4.80, $P= 1.56\times10^{-6}$) (**Figure 3**), all of which are myeloid origin traits. The CHIP phenotypic association findings support recent evidence suggesting CHIP primarily results in myeloid-related disruptions, although select distinct CHIP events could increase risk for disruptions in the lymphoid lineage.[9] Together our results support lineage-specific effects that differ by type of CH, suggesting shared etiology, specifically shared genetic etiology for LOY and myeloid traits, as well as ample phenotypic associations that detail early downstream phenotypic disruptions in hematologic traits that could alter disease risk.

*Telomere length displays a bi-directional association with types of CH*

Telomere length in leukocytes provides a metric of hematopoietic stem cell activity and can provide insights into how variation in hematopoietic stem cells interact with risk for acquiring CH.[17,21] The genetic relationship between each type of CH with leukocyte telomere length (TL) was evaluated to determine whether this could also contribute to predisposition to CH and no significant genetic correlations between each type of CH with telomere length were evident (**Figure 2**). To further test for a genetic association, we conducted one-directional MR between TL and each CH type using 130 previously published TL-associated variants.[22] Based on MR-IVW models, we observed positive associations between the TL IV and autosomal mCAs

($Z_{filtered}$= 5.65, P= $1.21\times10^{-7}$), CHIP ($Z_{filtered}$= 4.05, P= $9.68\times10^{-5}$), and MPNs ($Z_{filtered}$= 5.61, P= $1.88\times10^{-7}$), and observed a negative association between the TL IV and LOY ($Z_{filtered}$= -5.41, P= $5.60\times10^{-7}$) and did not identify evidence for a genetic relationship between telomere length and LOX (**Supplemental Figure 11**, **Supplemental Table 14**), providing additional support to previous reports of associations between inherited telomere length and select CH traits.[17,23,24] The intercept from MR-Egger regression was significant ($p< 0.05$) for both autosomal mCAs and MPN (**Supplemental Table 14**), so we performed additional MR weighted median (MR-WM) analyses which displayed the same positive association between the TL IV and autosomal mCAs ($Z_{filtered}$= 4.16, P= $6.19\times10^{-5}$), and MPNs ($Z_{filtered}$= 4.34, P= $3.53\times10^{-5}$) (**Supplemental Table 14**). These MR associations are supported by our pathway analyses, which demonstrate telomere pathways are significantly associated with LOY, autosomal mCAs, CHIP, and MPN (**Supplemental Tables 7**, **9-11**); based on these data it is plausible that inherited variation in telomere length maintenance contributes to clonal expansion of mutated hematopoietic stem cells, or alternatively confers greater risk for mutation acquisition in hematopoietic stem cells.

Once clonal expansion ensues, measured telomere length is a metric of hematopoietic stem cell growth and clonal expansion. Using available measured telomere length data from UK Biobank, we observed inverse phenotypic associations between each type of CH and measured telomere length (**Figure 3**). CHIP, which presents with the smallest fraction of mutated clones, had the weakest phenotypic association with measured TL (T-statistic= -2.32, P= $2.01\times10^{-2}$) (**Figure 3**). To further examine this relationship, we conducted analyses between CHIP VAF and measured TL and observed individuals with higher VAF, *i.e.*, higher CHIP cellular fraction, had a more inverse association with measured TL (T-statistic= -6.16, P= $7.84\times10^{-10}$, **Supplemental Figure**

**12a**). Additionally, individuals with higher autosomal mCA cellular fraction also demonstrated a stronger inverse association with measured TL (T-statistic= -9.02, P=$2.16\times10^{-19}$, **Supplemental Figure 12b**). The number of mutations present in an individual were also inversely associated with TL for both increasing CHIP mutation count (T-statistic= -2.60, P= $9.38\times10^{-3}$) and increasing autosomal mCA count (T-statistic= -10.01, P= $1.35\times10^{-23}$). Individuals with both CHIP and autosomal mCAs also demonstrated an inverse association with TL (T-statistic= -3.06, P= $2.23\times10^{-3}$) with individuals carrying co-localizing CHIP and autosomal mCAs displaying a stronger inverse association with TL (T-statistic= -3.94, P= $8.06\times10^{-5}$) compared to individuals without CHIP or autosomal mCAs (**Supplemental Figure 13**). These inverse TL associations indicate increased clonal expansion leads to reduced measured telomere length and suggest reductions in telomere length from the expansion of mutated clones could lead to further genomic instability and the acquisition of additional CH mutations.

## Discussion

Understanding the underlying molecular mechanisms of types of CH is critical for disentangling age-related clonal evolution and its possible impact on subsequent disease risk. Herein, our analysis of CH using large-scale genetic data highlights both similarities in the underlying mechanisms and key differences, particularly with respect to events related to lymphoid and myeloid lineages. Common to the types of CH are core pathways, namely, cellular differentiation, DNA repair, and cell cycle regulation, that contribute to generation and clonal expansion. Together, these findings detail specific characteristics of CH that should be investigated to improve the utility of detectable CH for disease risk and possible intervention or prevention.

We provide evidence for genome-wide genetic correlations between LOY and LOX as well as between LOY and MPN, suggesting shared biologic mechanisms promoting or predisposing to the development and clonal expansion. Phenotypic associations between types of CH further support similar genetic etiology, but could also indicate shared environmental factors that drive CH growth and expansion (*e.g.*, medication[25,26] or tobacco usage[24,27–31]). A notable discordance in directionality between the genetic correlation and the phenotypic relationship between LOY and MPN supports a shared genetic etiology, but also suggests a mutual exclusivity of some CH types such that once one type of CH develops, the occurrence of others may be suppressed. Individuals with CHIP and higher cellular fraction autosomal mCA events also demonstrated an inverse phenotypic association with LOY indicating a similar mutually exclusive relationship. These findings further support common genetic factors, and raise the importance of pursuing shared environmental contributors beyond smoking and air-pollution.[24,27–32] Our findings reveal that there is mutual exclusivity of CH, most likely due to hematopoietic stem cells that cannot tolerate multiple independent somatic drivers of CH. However, we also observe strong evidence for the co-occurrence of CHIP and autosomal mCAs in the same individual, and in many instances, co-localizing within known CH driver mutations (*e.g.*, *TET2*, *DNMT3A*, *JAK2*). Cross sectional observations of cellular fraction indicate these CHIP mutations often precede autosomal mCAs, which can lead to preferential clonal expansion of mCAs containing CHIP mutations, as has been mechanistically examined in specific cases.[33]

Rapid clonal expansion afforded by each type of CH leads to marked reductions in measured telomere length. These reductions in telomere length can lead to increased genomic instability in

individuals with CH and could increase the likelihood of acquiring additional types of CH. Individuals who acquired overlapping CHIP and autosomal mCAs were found to have greater reductions in telomere length, which is a marker of past clonal expansion. Detection of these highly clonal, co-localizing CH events, especially at *TET2*, *DNMT3A* and *JAK2*, could be helpful in identifying individuals at increased risk of developing hematological malignancies. Future studies should focus on investigating the co-occurrence and co-localization of CHIP and autosomal mCAs to further evaluate associations with environmental factors, elevated disease, and mortality.

## Online Methods

*Hematopoietic Phenotypes*

We used genome-wide association study (GWAS) summary statistics to investigate germline similarities and differences of 25 hematopoietic related phenotypes. These included loss of chromosome Y (LOY) in men,[34] loss of chromosome X (LOX) in women, autosomal mCAs including gains, losses and copy neutral loss of heterozygosity,[35] CHIP,[24] MPNs,[17] leukocyte telomere length (TL),[22] and 19 blood cell traits[17] (**Supplemental Table 15**). For LOX, we used previously generated data on copy number variation,[20,35] and performed GWAS on 243,106 women in the UK Biobank using a linear mixed model implemented in BOLT-LMM,[36] to account for cryptic population structure and relatedness, a similar methodology was used to conduct the LOY GWAS.[34]

Data on 482,378 subjects from the UK Biobank were used for phenotypic association analyses, after removing individuals with sex discordance or whose DNA failed genotyping QC. We used

previously generated data for each of the hematopoietic related phenotypes,[17,22,34,35] with the exception of CHIP. The UK Biobank 200K whole exome sequencing data[37] were somatically called for CHIP mutations using Mutect2. A QCed set of N=198,178 individuals were analyzed using a panel of normals created from UKBB participants with age <40, and included as part of the QC process. Only variants passing standard Mutect2 filters, with a variant allele fraction (VAF) of >=2% and a DP>20 were included. CHIP was identified in these individuals using a curated list of CHIP mutated variants as previously described.[24] Individuals with a diagnosis of myeloid malignancy (AML, MDS, MPN) before blood draw were excluded from the CHIP phenotype while those that went on to develop myeloid malignancy post-blood draw were retained. In total, 5,689 (2.9%) individuals were found to have at least one CHIP curated variant.

*Genetic Correlation*

We used linkage disequilibrium score regression (LDSC) to compute pairwise genetic correlation between traits. We utilized an LD score reference panel generated on 6,285 European ancestry individuals combined from the 1000 Genomes Phase 3 and UK10K cohorts, with a total of 17,478,437 available variants. For pairwise genetic correlation estimation, GWAS summary statistics were filtered to include overlapping variants with > 1% MAF and >90% imputation quality score. As several of the included GWAS summary statistics utilized UK Biobank data, we constrained the LDSC intercept by accounting for both the known sample overlap and phenotypic correlation between traits (**Supplemental Table 16**). This was performed using the --intercept-gencov command within LDSC.

*Phenotypic Associations*

Pairwise phenotypic associations between all hematopoietic phenotypes were generated using linear regression adjusting for age, age-squared, sex (in non sex-specific traits), and a 25-level smoking variable to reduce the potential for confounding variables driving associations.[29] To ensure compatibility between binary and continuous phenotypic association results, association T-statistics were generated and reported to measure strength and direction of phenotypic associations.

*Mendelian Randomization Analyses*

We performed bidirectional two-sample Mendelian randomization (MR) analyses between LOY and 19 blood cell phenotypes, and one-directional MR between TL and LOY, LOX, autosomal mCAs, CHIP, and MPN. Briefly, MR analyses utilize genetic variants from GWAS as instrumental variables (IVs) to assess the directional association between an exposure and outcome, which can mimic the biological link between exposure and outcome.[38] Each variant used in a MR analysis must satisfy three assumptions: 1) it is associated with the risk factor, 2) it is not associated with any confounder of the risk factor–outcome association, 3) it is conditionally independent of the outcome given the risk factor and confounders.[39,40]

For our analyses, we extracted summary statistics of reported independent genome-wide significant signals ($p < 5x10^{-8}$) and used them as IVs for each phenotype. LOY signals were extracted from a previous LOY GWAS,[34] and blood cell phenotype signals were downloaded from the PGS Catalog (https://www.pgscatalog.org/publication/PGP000078/).[41] For TL, we used summary statistics and 130 significant signals of the largest TL GWAS to date.[22] We removed X chromosome variants from all MR analyses due to sex-specific nature of some types of CH.

After creating each phenotypic IV, we extracted the same set of signals from summary statistics for each outcome. If any signals were missing in the outcome summary statistics, we collected proxies for these signals using GCTA[42] with European UK Biobank individuals as reference (within 1 MB of reported signals and $R^2 > 0.4$). We chose the proxy of each missing signal with the largest $R^2$ value as the replacement IV, which was contained in both GWAS summary statistics of the exposure and outcome. All signals of exposure were aligned to increasing allele and alleles for outcome were realigned accordingly.

The MR inverse-variance weighted (MR-IVW) model, which can provide high statistical power,[43] was used as our primary analysis. As some signals may have a stronger association with the outcome than the exposure, which may induce reverse causality, we applied Steiger filtering to each IV in order to remove these variants using the "TwoSampleMR" R package.[44] We then applied Radial filtering to exclude signals that were identified as outliers according to Rücker's Q′ statistic.[45]

The sensitivity of MR models was checked by the degree of heterogeneity ($I^2$ statistics and Cochran's Q-derived P-value), horizontal pleiotropy (MR-Egger $p_{intercept} < 0.05$), and funnel and dosage plots. To account for potential horizontal pleiotropy and heterogeneity, three additional MR models were performed: MR-Egger,[46] weighted median (MR-WM),[47] and penalized weighted median (MR-PWM).[47]

*Pathway and Gene Set Analyses*

We performed agnostic pathway-based analyses using the summary data-based adaptive rank truncated product (sARTP) method, which combines GWAS summary statistics across SNPs in a gene or a pathway,[48] to identify gene sets and pathways associated with each type of CH. A total of 6,290 curated gene sets and canonical pathways from GSEA (https://www.gsea-msigdb.org/gsea/msigdb/) were used for the analyses. For each type of CH, the signals from up to two of the most associated SNPs in a gene were accumulated. We adjusted for the number of SNPs in a gene and the number of genes in a pathway through a resampling procedure that controls for false positives.[48] The P values of gene- and pathway-level associations were estimated from the resampled null distribution generated from 100 million permutations. Linkage disequilibrium between SNPs was computed from European individuals within 1000 Genomes Project data.[49] To reduce the potential for population stratification to bias the results, we rescaled the marginal SNP results for each CH trait to set the genomic control inflation factor to 1. A Bonferroni corrected level of significance of $7.95\times10^{-6}$ (0.05/6,290 GSEA pathways) was used to assess statistical significance.

**Acknowledgments**: This work was conducted using the UK Biobank resource (applications 9905, 21552, and 31063). The UK Biobank was established by the Wellcome Trust, the Medical Research Council, the United Kingdom Department of Health, and the Scottish Government. The UK Biobank has also received funding from the Welsh Assembly Government, the British Heart Foundation, and Diabetes UK. The clonal hematopoiesis infographic was created with BioRender.com. We thank Dr. Jacqueline B. Vo for her graphical support. The opinions expressed by the authors are their own and this material should not be interpreted as representing

the official viewpoint of the U.S. Department of Health and Human Services, the National Institutes of Health or the National Cancer Institute.

**Author contributions**: J.R.B.P, V.G.S., and M.J.M. conceived the study. L.D.C. and V.G.S. carried out the CHIP calls. D.W.B., L.D.C., Y.Z., T.R., L.S., K.Y. performed computational and statistical analyses. J.R.B.P, V.G.S., and M.J.M. supervised the study. D.W.B., S.K.N., S.J.C., J.R.B.P, V.G.S., and M.J.M. drafted the manuscript with input from all authors. All authors critically read and approved the final version of the manuscript.

**Competing interests**: V.G.S. serves as an advisor to and/or has equity in Branch Biosciences, Ensoma, Novartis, Forma, and Cellarity, all unrelated to the present work. All other authors declare no relevant competing interests.

**Financial support**: This work was supported by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health (D.W.B, T.R., L.S., K.Y., S.J.C., M.J.M.). J.R.B.P and Y.Z. are supported by the Medical Research Council (Unit programs: MC_UU_12015/2, MC_UU_00006/2). SKN is a Scholar of the American Society of Hematology. This work was also supported by the New York Stem Cell Foundation (V.G.S.), a gift from the Lodish Family to Boston Children's Hospital (V.G.S.), and National Institutes of Health Grants R01 DK103794 and R01 HL146500 (V.G.S.). V.G.S. is a New York Stem Cell-Robertson Investigator.

**Data Availability**: All data used in analyses is available through application to the UK Biobank.

## References


1. Liggett, L. A. & Sankaran, V. G. Unraveling Hematopoiesis through the Lens of Genomics. *Cell* **182**, 1384–1400 (2020).
2. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
3. Jaiswal, S. & Ebert, B. L. Clonal hematopoiesis in human aging and disease. *Science* **366**, (2019).
4. Mustjoki, S. & Young, N. S. Somatic Mutations in 'Benign' Disease. *N. Engl. J. Med.* **384**, 2039–2052 (2021).
5. Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
6. Lawson, A. R. J. *et al.* Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* **370**, 75–82 (2020).
7. Busque, L. *et al.* Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. *Blood* **113**, 3472–3474 (2009).
8. Ayachi, S., Buscarlet, M. & Busque, L. 60 Years of clonal hematopoiesis research: From X-chromosome inactivation studies to the identification of driver mutations. *Exp. Hematol.* **83**, 2–11 (2020).
9. Niroula, A. *et al.* Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat. Med.* 1–7 (2021).
10. Saiki, R. *et al.* Combined landscape of single-nucleotide variants and copy number alterations in clonal hematopoiesis. *Nat. Med.* **27**, 1239–1249 (2021).

11. Gao, T. *et al.* Interplay between chromosomal alterations and gene mutations shapes the evolutionary trajectory of clonal hematopoiesis. *Nat. Commun.* **12**, 338 (2021).
12. Hahm, C. *et al.* Genomic aberrations of myeloproliferative and myelodysplastic/myeloproliferative neoplasms in chronic phase and during disease progression. *Int. J. Lab. Hematol.* **37**, 181–189 (2015).
13. Klampfl, T. *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* **118**, 167–176 (2011).
14. Rumi, E. *et al.* Identification of genomic aberrations associated with disease transformation by means of high-resolution SNP array analysis in patients with myeloproliferative neoplasm. *Am. J. Hematol.* **86**, 974–979 (2011).
15. Kawamata, N. *et al.* Genetic profiling of myeloproliferative disorders by single-nucleotide polymorphism oligonucleotide microarray. *Exp. Hematol.* **36**, 1471–1479 (2008).
16. Grinfeld, J. *et al.* Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *N. Engl. J. Med.* **379**, 1416–1430 (2018).
17. Bao, E. L. *et al.* Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).
18. Lin, S.-H. *et al.* Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK Biobank men. *Sci. Rep.* **10**, 1–10 (2020).
19. Terao, C. *et al.* GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nat. Commun.* **10**, 4719 (2019).
20. Loh, P.-R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).

21. Werner, B. *et al.* Reconstructing the in vivo dynamics of hematopoietic stem cells from telomere length distributions. *eLife* **4**, e08687 (2015).

22. Codd, V. *et al.* Polygenic basis and biomedical consequences of telomere length variation. *Nat. Genet.* **53**, 1425–1433 (2021).

23. Brown, D. W. *et al.* Genetically predicted telomere length is associated with clonal somatic copy number alterations in peripheral leukocytes. *PLOS Genet.* **16**, e1009078 (2020).

24. Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).

25. 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**, 143 (2021).

26. Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).

27. Dumanski, J. P. *et al.* Smoking is associated with mosaic loss of chromosome Y. *Science* **347**, 81–83 (2015).

28. Zhou, W. *et al.* Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat. Genet.* **48**, 563–568 (2016).

29. Loftfield, E. *et al.* Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci. Rep.* **8**, 12316–12316 (2018).

30. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).

31. Coombs, C. C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**, 374-382.e4 (2017).
32. Wong, J. Y. Y. *et al.* Outdoor air pollution and mosaic loss of chromosome Y in older men from the Cardiovascular Health Study. *Environ. Int.* **116**, 239–247 (2018).
33. Godfrey, A. L. *et al.* JAK2V617F homozygosity arises commonly and recurrently in PV and ET, but PV is characterized by expansion of a dominant homozygous subclone. *Blood* **120**, 2704–2707 (2012).
34. Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).
35. Loh, P.-R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* **584**, 136–141 (2020).
36. Loh, P.-R. *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
37. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 1–10 (2021).
38. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
39. Didelez, V. & Sheehan, N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16**, 309–330 (2007).

40. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).

41. Vuckovic, D. *et al.* The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231 (2020).

42. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

43. Slob, E. A. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44**, 313–329 (2020).

44. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship between imprecisely measured traits using genetic instruments. *BiorXiv* 117101 (2017).

45. Bowden, J. *et al.* Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int. J. Epidemiol.* **48**, 728–742 (2019).

46. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).

47. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).

48. Zhang, H. *et al.* A Powerful Procedure for Pathway-Based Meta-analysis Using Summary Statistics Identifies 43 Pathways Associated with Type II Diabetes in European Populations. *PLOS Genet.* **12**, e1006122 (2016).

49. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

**Figure Legends**:

**Figure 1**. Description of each type of clonal hematopoiesis.

**Figure 2**. Pairwise genetic correlations between each type of clonal hematopoiesis, telomere length, and 19 blood cell traits. The upper diagonal gives the genetic correlation, and the lower diagonal gives the -log10(P) value derived using linkage disequilibrium score regression. The genetic correlation display color reflects the magnitude of association.

**Figure 3**. Pairwise phenotypic associations between each type of clonal hematopoiesis, telomere length, and 19 blood cell traits. The upper diagonal gives the T-statistic, and the lower diagonal gives the -log10(P) value derived using linear regression adjusted for age, age-squared, 25-level smoking status, and sex (in non LOY or LOX comparisons). The T-statistic display color reflects the magnitude of association.

**Supplemental material**:

**Supplemental Table 1.** Observed heritability for each type of CH, telomere length, and 19 blood cell traits

**Supplemental Table 2**. Overlapping counts of each CH type in UK Biobank

**Supplemental Table 3**. Proportion of individuals with each type of CH by autosomal mCA CF categories

**Supplemental Table 4**. Proportion of individuals with each type of CH by CHIP VAF categories

**Supplemental Table 5**. Frequency of CHIP gene mutations by autosomal mCA status

**Supplemental Table 6**. Frequency of autosomal mCA mutations by CHIP status

**Supplemental Table 7**. Pathways and gene sets significantly associated with LOY

**Supplemental Table 8**. Pathways and gene sets significantly associated with LOX

**Supplemental Table 9**. Pathways and gene sets significantly associated with autosomal mCAs

**Supplemental Table 10**. Pathways and gene sets significantly associated with CHIP

**Supplemental Table 11**. Pathways and gene sets significantly associated with MPN

**Supplemental Table 12**. Mendelian randomization results for LOY and 19 blood cell traits using LOY instrumental SNPs with Steiger and Radial filters

**Supplemental Table 13**. Mendelian randomization results for 19 blood cell traits and LOY using blood cell phenotype signals downloaded from the PGS Catalog (https://www.pgscatalog.org/publication/PGP000078/) with Steiger and Radial filters

**Supplemental Table 14**. Mendelian randomization results between telomere length and each type of CH using telomere length instrumental SNPs with Steiger and Radial filters

**Supplemental Table 15**. Previously published hematopoietic phenotypes

**Supplemental Table 16**. Pairwise hematopoietic phenotype correlation and intercept value used to constrain the LD score regression intercept. The sample size for each phenotype is number of participants from each GWAS study. Pearson correlations were calculated between phenotypes using 482,378 UK Biobank participants

**Supplemental Figure 1**. Observed heritability for each type of CH, telomere length, and 19 blood cell traits.

**Supplemental Figure 2**. Age distribution of each type of CH in the UK Biobank.

**Supplemental Figure 3**. Proportion of individuals with each type of CH by autosomal mCA cellular fraction. 95% confidence intervals (gray fill) are also plotted. Detailed plot data is given in **Supplemental Table 3**.

**Supplemental Figure 4**. Proportion of individuals with each type of CH by CHIP VAF. 95% confidence intervals (gray fill) are also plotted. Detailed plot data is given in **Supplemental Table 4**.

**Supplemental Figure 5**. (**a**) Overlap between autosomal mCAs and CHIP. The number of subjects from UK Biobank are given. (**b**) Proportion of CHIP gene mutations significantly enriched in subjects with autosomal mCAs. Frequencies for each CHIP mutation are given in Supplemental Table 5. (**c**) Circos plots of autosomal mCAs by CHIP mutation status. Frequency of autosomal mCAs by chromosome are given in **Supplemental Table 6**.

**Supplemental Figure 6**. Frequency of overlapping CHIP mutations and autosomal mCA events. 56 co-localizing mutations were observed within 53 individuals. The y-axis depicts the frequency of each CHIP mutation. The frequency of each autosomal mCA copy number is labeled in each bar.

**Supplemental Figure 7**. Clonal evolution of CHIP and autosomal mCA mutations. The y-axis depicts the estimated CHIP VAF. The x-axis depicts estimated autosomal mCA cellular fraction. Autosomal mCA cellular fraction was missing for 7 events, due to unknown copy number state. CHIP VAF was larger than mCA cellular fraction for 35 (71.4%) co-localizing mutations.

**Supplemental Figure 8**. Overlap of significant pathways and gene sets by type of CH. Detailed results are given in **Supplemental Tables 7-11**.

**Supplemental Figure 9**. MR inverse-variance weighted results between LOY and 19 blood cell traits using LOY instrumental SNPs with Steiger and Radial filters. Detailed results are given in **Supplemental Table 12**.

**Supplemental Figure 10**. MR inverse-variance weighted results between 19 blood cell traits and LOY using blood cell phenotype signals downloaded from the PGS Catalog (https://www.pgscatalog.org/publication/PGP000078/) with Steiger and Radial filters. Detailed results are given in **Supplemental Table 13**.

**Supplemental Figure 11**. MR inverse-variance weighted results between telomere length and each type of CH using telomere length instrumental SNPs with Steiger and Radial filters. Detailed results are given in **Supplemental Table 14**.

**Supplemental Figure 12**. (**a**) Average telomere length and 95% confidence intervals by CHIP VAF. (**b**) Average telomere length and 95% confidence intervals by autosomal mCA cellular fraction.

**Supplemental Figure 13**. Telomere length association (β) with 95% confidence intervals by CHIP and autosomal mCA status. The number of subjects within each group from UK Biobank are given. Associations are adjusted for age, age-squared, sex, and detailed smoking status.

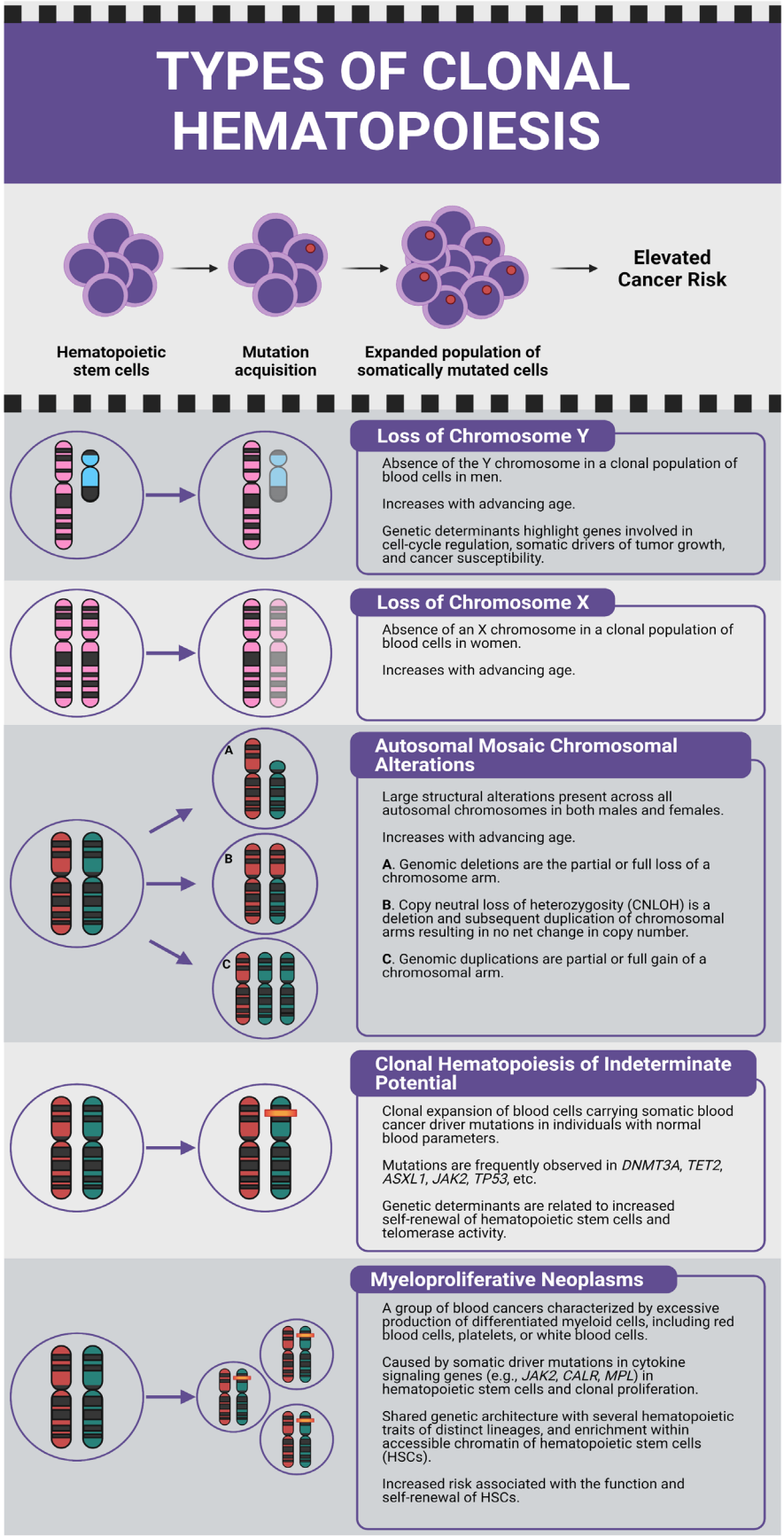


**Figure 1**. Description of each type of clonal hematopoiesis.

| *Correlation (above diagonal)*<br>*-log $_{10}$ p-value (lower diagonal)* | LOY | LOX | AUTO_mCA | CHIP | MPN | TL | MPV | PLT_CRIT | PDW | PLT_COUNT | HCT | HGB | MCH | MCHC | MCV | MEAN_RETIC_VOL | RBC_COUNT | RDW | RETIC_COUNT | BASO_COUNT | EO_COUNT | LYMPH_COUNT | MONO_COUNT | NEUTRO_COUNT | WBC_COUNT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **LOY** | NA | 0.2 | -0.1 | 0.2 | 0.2 | 0.1 | 0.0 | 0.2 | -0.1 | 0.2 | 0.0 | 0.0 | -0.1 | -0.1 | -0.1 | -0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 |
| **LOX** | 4.2 | NA | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1 | -0.1 | -0.1 | -0.1 | -0.1 | -0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.0 |
| **AUTO_mCA** | 0.2 | 0.0 | NA | 0.2 | 0.1 | 0.7 | 0.1 | 0.2 | 0.2 | 0.2 | -0.1 | -0.1 | -0.3 | -0.2 | -0.3 | -0.2 | 0.1 | 0.1 | 0.2 | -0.1 | 0.0 | 0.0 | -0.2 | -0.2 | -0.1 |
| **CHIP** | 0.8 | 0.3 | 0.1 | NA | 0.503 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | -0.1 | 0.0 | -0.1 | -0.1 | 0.1 | 0.2 | 0.1 | -0.1 | 0.0 | -0.1 | 0.1 | 0.0 | 0.0 |
| **MPN** | 1.8 | 0.3 | 0.2 | 0.863 | NA | 0.1 | 0.0 | 0.3 | 0.0 | 0.2 | 0.1 | 0.2 | -0.1 | 0.0 | -0.1 | -0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.2 | 0.2 |
| **TL** | 1.2 | 0.7 | 0.4 | 0.2 | 0.5 | NA | 0.0 | -0.1 | 0.1 | -0.1 | 0.2 | 0.2 | 0.0 | 0.0 | -0.1 | -0.1 | 0.2 | 0.0 | 0.0 | -0.2 | -0.1 | -0.1 | 0.0 | -0.1 | -0.1 |
| **MPV** | 0.7 | 0.5 | 0.1 | 0.0 | 0.2 | 0.2 | NA | 0.0 | 0.5 | -0.7 | 0.1 | 0.1 | 0.0 | -0.2 | 0.0 | -0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 |
| **PLT_CRIT** | 7.5 | 1.1 | 0.4 | 0.0 | 1.9 | 2.2 | 0.5 | NA | -0.3 | 1.1 | 0.1 | 0.1 | -0.1 | -0.2 | 0.0 | 0.0 | 0.1 | 0.1 | 0.0 | 0.3 | 0.3 | 0.3 | 0.5 | 0.4 | 0.5 |
| **PDW** | 2.8 | 0.3 | 0.3 | 0.8 | 0.0 | 7.8 | 39.1 | 16.7 | NA | -0.5 | -0.1 | 0.0 | -0.1 | 0.1 | -0.2 | -0.2 | 0.1 | 0.0 | 0.2 | 0.0 | -0.1 | -0.1 | -0.2 | 0.0 | -0.1 |
| **PLT_COUNT** | 6.3 | 0.8 | 0.4 | 0.0 | 1.7 | 2.4 | 86.1 | 163.9 | 57.9 | NA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.2 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 |
| **HCT** | 0.7 | 2.4 | 0.2 | 0.2 | 2.6 | 15.1 | 13.0 | 2.7 | 2.8 | 0.6 | NA | 1.3 | -0.2 | -0.5 | 0.0 | -0.1 | 1.0 | 0.0 | 0.2 | 0.2 | 0.0 | 0.2 | 0.0 | 0.3 | 0.3 |
| **HGB** | 0.2 | 3.2 | 0.3 | 0.4 | 2.4 | 16.0 | 5.0 | 1.1 | 0.3 | 0.2 | 89.1 | NA | 0.0 | 0.1 | 0.0 | -0.1 | 0.9 | -0.2 | 0.3 | 0.1 | 0.0 | 0.2 | 0.0 | 0.2 | 0.2 |
| **MCH** | 2.3 | 1.6 | 0.4 | 0.6 | 0.6 | 1.1 | 1.6 | 0.8 | 5.4 | 0.4 | 1.6 | 0.3 | NA | 0.6 | 1.2 | 0.8 | -0.9 | -0.4 | 0.0 | 0.0 | -0.1 | 0.0 | -0.1 | -0.1 | -0.1 |
| **MCHC** | 1.9 | 1.5 | 0.3 | 0.1 | 0.2 | 1.1 | 7.6 | 3.0 | 7.3 | 0.5 | 5.7 | 0.5 | 21.2 | NA | 0.2 | -0.2 | -0.5 | -0.6 | 0.4 | -0.2 | -0.1 | -0.1 | -0.3 | -0.2 | -0.2 |
| **MCV** | 2.2 | 1.3 | 0.4 | 0.8 | 0.8 | 1.7 | 0.1 | 0.3 | 12.0 | 0.2 | 0.1 | 0.2 | 55.9 | 1.2 | NA | 1.0 | -0.9 | -0.3 | -0.1 | 0.0 | -0.1 | 0.0 | 0.0 | -0.1 | -0.1 |
| **MEAN_RETIC_VOL** | 1.9 | 0.6 | 0.4 | 0.8 | 2.4 | 3.2 | 3.3 | 0.1 | 24.5 | 0.6 | 2.9 | 5.9 | 36.6 | 2.4 | 52.2 | NA | -0.7 | 0.1 | -0.4 | 0.0 | 0.0 | 0.0 | 0.0 | -0.1 | -0.1 |
| **RBC_COUNT** | 2.0 | 0.4 | 0.3 | 0.7 | 2.7 | 12.9 | 5.9 | 1.7 | 2.7 | 0.7 | 106.3 | 62.3 | 74.9 | 19.2 | 89.7 | 67.6 | NA | 0.2 | 0.3 | 0.1 | 0.0 | 0.1 | 0.0 | 0.2 | 0.2 |
| **RDW** | 0.1 | 1.4 | 0.2 | 1.2 | 0.5 | 0.9 | 3.2 | 2.3 | 0.9 | 0.8 | 0.1 | 3.4 | 11.9 | 28.3 | 5.6 | 1.3 | 7.4 | NA | -0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 |
| **RETIC_COUNT** | 1.8 | 0.2 | 0.3 | 0.4 | 0.4 | 0.8 | 0.2 | 1.1 | 5.9 | 0.5 | 17.4 | 35.7 | 0.5 | 21.2 | 2.8 | 20.5 | 15.7 | 2.1 | NA | 0.1 | 0.0 | 0.2 | 0.0 | 0.2 | 0.2 |
| **BASO_COUNT** | 0.3 | 1.8 | 0.2 | 0.2 | 0.5 | 7.6 | 5.0 | 27.6 | 1.0 | 9.0 | 12.3 | 5.1 | 1.3 | 7.4 | 0.1 | 0.2 | 7.5 | 4.5 | 3.4 | NA | 0.3 | 0.4 | 0.4 | 0.5 | 0.6 |
| **EO_COUNT** | 1.5 | 2.5 | 0.0 | 0.1 | 2.5 | 4.0 | 0.7 | 18.5 | 5.6 | 17.5 | 0.1 | 1.0 | 4.3 | 6.8 | 2.5 | 2.0 | 0.7 | 4.0 | 0.7 | 26.4 | NA | 0.3 | 0.3 | 0.2 | 0.4 |
| **LYMPH_COUNT** | 1.2 | 1.4 | 0.0 | 0.6 | 0.5 | 7.4 | 0.6 | 22.1 | 5.5 | 22.2 | 14.3 | 9.1 | 0.1 | 4.5 | 0.9 | 0.9 | 9.4 | 1.4 | 5.8 | 28.2 | 18.6 | NA | 0.5 | 0.4 | 0.9 |
| **MONO_COUNT** | 0.9 | 0.6 | 0.3 | 0.6 | 1.6 | 0.2 | 3.7 | 28.8 | 14.3 | 23.9 | 1.1 | 0.4 | 2.6 | 12.2 | 0.0 | 0.0 | 0.4 | 5.1 | 1.4 | 17.1 | 25.6 | 41.0 | NA | 0.5 | 0.7 |
| **NEUTRO_COUNT** | 0.9 | 0.7 | 0.4 | 0.2 | 3.2 | 3.7 | 3.2 | 40.1 | 1.7 | 23.7 | 34.9 | 24.3 | 6.5 | 7.3 | 4.4 | 4.7 | 24.9 | 6.4 | 21.8 | 44.4 | 26.1 | 75.3 | 62.7 | NA | 1.1 |
| **WBC_COUNT** | 1.7 | 0.6 | 0.2 | 0.0 | 2.7 | 7.1 | 2.1 | 44.7 | 5.0 | 30.6 | 34.5 | 21.7 | 5.9 | 9.7 | 2.5 | 2.1 | 26.1 | 6.9 | 19.3 | 60.1 | 62.8 | 150.9 | 87.0 | 238.1 | NA |

**Figure 2.** Pairwise genetic correlations between each type of clonal hematopoiesis, telomere length, and 19 blood cell traits. The upper diagonal gives the genetic correlation, and the lower diagonal gives the -$\log_{10}$(P) value derived using linkage disequilibrium score regression. The genetic correlation display color reflects the magnitude of association.

| *T-statistic (above diagnal)*<br>*-log $_{10}$ p-value (below diagnal)* | LOY | LOX | AUTO_mCA | CHIP | MPN | TL | MPV | PLT_CRIT | PDW | PLT_COUNT | HCT | HGB | MCH | MCHC | MCV | MEAN_RETIC_VOL | RBC_COUNT | RDW | RETIC_COUNT | BASO_COUNT | EO_COUNT | LYMPH_COUNT | MONO_COUNT | NEUTRO_COUNT | WBC_COUNT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **LOY** | NA | NA | 4.3 | -3.8 | -4.8 | -17.4 | -0.2 | 20.0 | -9.1 | 17.6 | 1.9 | 1.1 | 10.0 | -1.5 | 13.2 | 1.4 | -6.5 | -12.0 | -4.8 | 3.2 | 6.5 | -2.8 | 14.1 | 20.1 | 13.5 |
| **LOX** | NA | NA | 12.4 | 0.5 | 1.0 | -9.3 | -0.9 | -1.8 | 1.6 | -1.2 | -2.2 | -1.4 | 3.6 | 2.5 | 3.0 | 0.0 | -3.8 | 1.9 | -0.2 | 1.7 | 1.0 | 24.0 | 5.1 | -16.2 | 0.9 |
| **AUTO_mCA** | 4.8 | 34.8 | NA | 10.0 | 41.0 | -7.7 | -0.5 | 0.6 | 11.3 | 0.9 | -2.1 | -2.0 | 6.4 | 1.2 | 6.8 | 5.5 | -5.2 | 14.6 | 2.9 | 7.5 | -3.3 | 60.3 | 10.6 | -1.2 | 34.5 |
| **CHIP** | 3.9 | 0.2 | 22.6 | NA | 12.5 | -2.3 | -1.2 | 6.7 | 4.1 | 6.6 | -1.1 | -1.4 | 0.0 | -0.3 | 0.3 | -0.5 | -1.0 | 4.7 | 3.7 | 4.4 | -4.2 | 2.2 | 6.1 | 4.8 | 5.0 |
| **MPN** | 5.7 | 0.5 | >323 | 35.1 | NA | -12.2 | -5.6 | 56.3 | 23.8 | 55.6 | 6.2 | 2.0 | 11.2 | -9.0 | 18.7 | 32.2 | -1.3 | 48.5 | 10.3 | 10.5 | 3.3 | 1.1 | 23.7 | 20.8 | 18.1 |
| **TL** | 66.7 | 19.9 | 13.9 | 1.7 | 33.4 | NA | 10.8 | 8.0 | 8.7 | 1.9 | 7.9 | 7.0 | -36.2 | -1.9 | -43.0 | -30.8 | 33.8 | 2.2 | -2.6 | -11.3 | -20.3 | -20.8 | -24.0 | -13.0 | -25.2 |
| **MPV** | 0.1 | 0.5 | 0.2 | 0.6 | 7.6 | 26.4 | NA | 2.1 | 312.5 | -366.8 | 19.8 | 6.3 | -37.6 | -30.8 | -26.7 | -30.1 | 34.2 | 17.2 | 5.9 | 9.2 | 3.3 | -4.1 | 17.3 | 2.5 | 1.9 |
| **PLT_CRIT** | 88.2 | 1.2 | 0.3 | 10.8 | >323 | 14.8 | 1.4 | NA | -114.5 | 1205.5 | -39.7 | -68.1 | -105.4 | -67.3 | -87.2 | -42.4 | 18.1 | 95.0 | 11.2 | 56.3 | 83.1 | 48.2 | 100.1 | 194.3 | 181.3 |
| **PDW** | 18.9 | 0.9 | 28.8 | 4.4 | 124.5 | 17.7 | >323 | >323 | NA | -246.2 | -1.6 | 20.7 | -21.2 | 53.0 | -59.0 | -70.1 | 36.3 | 11.0 | 44.7 | -2.5 | -9.0 | -8.4 | -20.1 | -6.5 | -12.1 |
| **PLT_COUNT** | 68.2 | 0.7 | 0.4 | 10.4 | >323 | 1.2 | >323 | >323 | >323 | NA | -45.3 | -64.0 | -72.8 | -44.6 | -62.0 | -21.1 | -2.1 | 75.1 | 7.3 | 44.5 | 70.6 | 43.0 | 78.4 | 168.1 | 156.3 |
| **HCT** | 1.3 | 1.5 | 1.4 | 0.6 | 9.3 | 14.6 | 86.5 | >323 | 1.0 | >323 | NA | 1594.1 | 7.7 | -150.7 | 108.1 | -17.8 | 865.6 | -121.3 | 84.5 | 1.7 | 15.3 | 31.7 | 38.9 | 83.3 | 81.2 |
| **HGB** | 0.6 | 0.8 | 1.3 | 0.8 | 1.4 | 11.5 | 9.5 | >323 | 94.4 | >323 | >323 | NA | 159.3 | 122.3 | 120.4 | -35.5 | 674.9 | -221.9 | 91.9 | -10.5 | 6.5 | 22.6 | 34.5 | 69.6 | 65.2 |
| **MCH** | 22.7 | 3.4 | 9.9 | 0.0 | 28.3 | 285.3 | 307.5 | >323 | 99.4 | >323 | 13.9 | >323 | NA | 476.2 | 1106.8 | 273.8 | -411.5 | -303.9 | -15.9 | -26.5 | -39.7 | -20.2 | -8.8 | -41.9 | -44.8 |
| **MCHC** | 0.9 | 2.0 | 0.6 | 0.1 | 18.8 | 1.3 | 207.1 | >323 | >323 | >323 | >323 | >323 | >323 | NA | 42.2 | -39.3 | -157.8 | -230.9 | 11.6 | -29.0 | -20.8 | -21.5 | -11.1 | -32.6 | -37.8 |
| **MCV** | 39.1 | 2.6 | 11.1 | 0.1 | 77.5 | >323 | 156.7 | >323 | >323 | >323 | >323 | >323 | >323 | >323 | NA | 381.2 | -378.3 | -219.7 | -26.9 | -14.8 | -35.5 | -11.0 | -4.0 | -31.0 | -31.0 |
| **MEAN_RETIC_VOL** | 0.8 | 0.0 | 7.5 | 0.2 | 227.2 | 206.7 | 197.8 | >323 | >323 | 98.2 | 70.3 | 275.4 | >323 | >323 | >323 | NA | -231.7 | 79.3 | -163.9 | 4.0 | -23.7 | -15.0 | -7.2 | -29.5 | -31.4 |
| **RBC_COUNT** | 10.1 | 3.9 | 6.6 | 0.5 | 0.7 | 248.8 | 255.1 | 72.4 | 286.9 | 1.5 | >323 | >323 | >323 | >323 | >323 | >323 | NA | 30.2 | 92.3 | 10.8 | 35.6 | 35.4 | 37.0 | 92.9 | 91.4 |
| **RDW** | 32.4 | 1.2 | 47.7 | 5.7 | >323 | 1.6 | 65.6 | >323 | 27.2 | >323 | >323 | >323 | >323 | >323 | >323 | >323 | 199.2 | NA | 3.1 | 29.3 | 16.7 | 4.8 | 30.5 | 42.3 | 37.3 |
| **RETIC_COUNT** | 5.7 | 0.1 | 2.4 | 3.7 | 24.1 | 2.1 | 8.4 | 28.3 | >323 | 12.6 | >323 | >323 | 56.5 | 30.5 | 158.4 | >323 | >323 | 2.7 | NA | 10.6 | 18.1 | 21.3 | 28.8 | 67.1 | 63.3 |
| **BASO_COUNT** | 2.9 | 1.1 | 13.2 | 4.9 | 25.2 | 29.1 | 19.5 | >323 | 2.0 | >323 | 1.0 | 25.1 | 154.4 | 183.9 | 48.9 | 4.1 | 26.5 | 188.1 | 25.6 | NA | 57.4 | 111.2 | 102.9 | 86.7 | 156.0 |
| **EO_COUNT** | 10.1 | 0.5 | 3.1 | 4.5 | 3.0 | 91.1 | 3.0 | >323 | 18.7 | >323 | 52.0 | 10.1 | >323 | 94.9 | 274.3 | 123.5 | 276.7 | 61.9 | 72.5 | >323 | NA | 71.6 | 101.7 | 66.3 | 148.7 |
| **LYMPH_COUNT** | 2.2 | 126.0 | >323 | 1.6 | 0.6 | 95.5 | 4.4 | >323 | 16.5 | >323 | 219.5 | 112.3 | 90.4 | 102.1 | 27.4 | 50.1 | 273.3 | 5.7 | 99.8 | >323 | >323 | NA | 146.2 | 82.2 | 638.3 |
| **MONO_COUNT** | 44.2 | 6.4 | 25.4 | 9.1 | 123.8 | 127.0 | 66.0 | >323 | 88.8 | >323 | >323 | 259.2 | 17.7 | 27.9 | 4.1 | 12.3 | 298.1 | 203.7 | 182.0 | >323 | >323 | >323 | NA | 137.4 | 278.0 |
| **NEUTRO_COUNT** | 89.2 | 58.3 | 0.6 | 5.8 | 95.5 | 37.9 | 2.0 | >323 | 10.0 | >323 | >323 | >323 | >323 | 232.1 | 210.4 | 190.7 | >323 | >323 | >323 | >323 | >323 | >323 | >323 | NA | 893.3 |
| **WBC_COUNT** | 40.9 | 0.4 | 259.4 | 6.2 | 72.2 | 139.4 | 1.2 | >323 | 32.9 | >323 | >323 | >323 | >323 | 312.2 | 209.9 | 215.0 | >323 | 302.8 | >323 | >323 | >323 | >323 | >323 | >323 | NA |

**Figure 3**. Pairwise phenotypic associations between each type of clonal hematopoiesis, telomere length, and 19 blood cell traits. The upper diagonal gives the T-statistic, and the lower diagonal gives the -$\log_{10}$(P) value derived using linear regression adjusted for age, age-squared, 25-level smoking status, and sex (in non LOY or LOX comparisons). The T-statistic display color reflects the magnitude of association.