Supplementary Information for "Monogenic and polygenic inheritance become instruments for clonal selection"

Po-Ru Loh, Giulio Genovese, Steven A McCarroll

Contents

Supplementary Notes

Additional supplementary figures showing detailed distributions of event calls on each chromosome are available at https://doi.org/10.1101/653691.

1 QC of mosaic chromosomal alteration calls

We called mCAs using the same approach we previously applied to the UK Biobank interim release. A full description of the method and a detailed exploration of its statistical properties is presented in the Supplementary Notes of ref. [9]. Below we describe the QC procedure we applied to mCA calls in the present analysis of the full UK Biobank data set, which included a few additional filters affecting $<1\%$ of the call set.

1.1 Identification of samples with possible DNA contamination

In our previous analysis of the UK Biobank N=150K interim release, we observed that a small fraction of samples ($<1\%$) exhibited evidence of possible DNA contamination based on apparent short interstitial CN-LOH calls in specific genomic regions of long-range linkage disequilibrium; we therefore used these likely-artifactual calls to flag samples for exclusion [9]. We applied the same QC approach to the full UK Biobank data set, identifying a total of 4,074 individuals to exclude based on short interstitial CN-LOH calls in the five regions we previously identified (chr3:$\sim$45Mb, chr6:$\sim$30Mb, chr8:$\sim$45Mb, chr10:$\sim$80Mb, chr17:$\sim$40Mb). As in our previous analysis, we also excluded individuals with three or more interstitial CN-LOH calls (a mostly-overlapping set of 534 individuals, bringing the total for exclusion to 4,100), and we excluded an additional 7 individuals with three or more calls with high implied switch error rates [9]. Finally, we added a filter excluding an additional 4 individuals based on having eight or more calls all within a very narrow BAF and LRR range (max $|\Delta\text{BAF}|<0.03$, LRR range <0.04), again indicating possible DNA contamination [6]. Together, these criteria resulted in 4,111 exclusions.

1.2 Additional filtering of mosaic event calls

Beyond the sample exclusions described above, we also performed QC on our mosaic event call set to filter calls that were unlikely to be true mosaic events (but did not suggest sample contamination, and hence did not require excluding samples from analysis). As in our previous analysis of the N=150K interim data set, our main post-processing step excluded events that might be constitutional (rather than mosaic) duplications [9]. As before, we filtered subchromosomal events of length $>$10Mb with LRR$>$0.35 or with LRR$>$0.2 and $|\Delta\text{BAF}|>0.16$, and we filtered events of length $<$10Mb with LRR$>$0.2 or with LRR$>$0.1 and $|\Delta\text{BAF}|>0.1$. We chose these thresholds

conservatively based on visual inspection of LRR and BAF distributions, in which likely constitutional duplications formed well-defined clusters (Extended Data Fig. 2e). (Most constitutional duplications were already masked in a pre-processing step involving a separate HMM [9].) We also added a filter for possible constitutional deletions based on LRR<–0.5 and heterozygosity rate <1/3 the expected rate within called event regions. (Constitutional deletions generally lead to genomic segments devoid of heterozygous sites, but occasionally heterozygous calls are erroneously made within the deletion regions, leading to the appearance of a large allelic imbalance. This behavior is easy to detect as it causes very low LRR and very low het rate within an event call.)

Additionally, we added filters for 29 event calls with LRR>0.2 and het rate >1.2x expected, LRR>0.1 and het rate >1.5x expected, or LRR>–0.05 and het rate >2x expected. Such event calls with elevated heterozygosity can arise from segmental duplications involving three distinct haploytpes or from genotype calling errors in which rare homozygotes are called as heterozygotes. Finally, we added a filter for 7 short interstitial events called at the *SNRPN* locus on chr15 between 24–25.5Mb. This locus is imprinted and exhibits differential replication timing between the paternal and the maternal haplotype [55]. Because a blood sample contains a fraction of replicating cells, an imbalance between maternal and paternal allelic fractions at this locus can sometimes be observed in genotype array data (without actual mosaicism).

1.3 Estimation of true false discovery rate

Our procedure for calling the existence of a mosaic event involved identifying significant autocorrelation in phased BAF deviations using a likelihood ratio test statistic [9]. We calibrated these test statistics empirically using a permutation-based procedure (phase randomization) to obtain a nominal 5% false discovery rate (FDR) threshold. However, this permutation-based 5% FDR threshold assumed that the only source of autocorrelation in phased BAF is a true mosaic event. In reality, other sources of autocorrelation exist; in particular, we found that sample contamination produced autocorrelation in regions of long-range LD (resulting in unusual false positive calls that we subsequently filtered). While we believe that our filtering eliminated most samples affected by spurious autocorrelation, our true FDR is likely to be slightly larger than 5% due to residual artifacts.

Fortunately, we can estimate our true FDR by leveraging the fact that true-positive events should be observed more frequently in the genomes of older people, while false-positive calls (which have no relation to age) should be observed in individuals whose age distribution matches that of the study population. This observation allows us to estimate FDR by comparing the age distributions of the highest-confidence calls (17,061 calls passing a permutation-based FDR of 1%) vs. medium-confidence calls (2,571 additional calls passing a permutation-based FDR of 5% when combined with the high-confidence calls, but failing the 1% threshold). The medium-confidence call set is expected to have a false positive rate of ≈32% based on the permutation-based FDRs—

meaning that its age distribution is expected to be an 68:32 mixture of (i) the age distribution of high-confidence calls and (ii) the age distribution of the study population. That is, the age distribution of medium-confidence calls should relax toward the age distribution of the overall study due to the inclusion of false positives—which is precisely what we see (Extended Data Fig. 2e). (The figure also includes low-confidence calls at FDR 10% for additional context, although we did not analyze these calls.)

Upon fitting the age distribution of medium-confidence calls as a mixture of the age distribution of high-confidence calls and the overall study distribution, the regression fit gives mixture proportions of ≈56:44 rather than 68:32, implying a true FDR of 6.6% (4.5–8.6%, 95% CI) when combined with the high-confidence calls—slightly higher than the permutation-based FDR of 5%, as expected. We note that this estimate is contingent on two assumptions: (i) the high-confidence call set predominantly contains true positives (which is supported by the observation that changing the high-confidence FDR threshold from 1% to 0.1% results in a near-identical "gold standard" age distribution; and (ii) the true positives in the high-confidence and medium-confidence call set have the same age distribution. While we acknowledge that these assumptions are imperfect, this analysis gives good evidence that our FDR is well-controlled. (We also note that while we cannot completely rule out the possibility that our FDR is higher than we estimated, the key results of our paper are robust to higher FDRs than estimated; e.g., we would only expect a higher-than-estimated FDR to weaken GWAS associations and decrease effect sizes.)

2 Population structure among UK Biobank participants

The large majority of UK Biobank participants are of European ancestry: 94% reported "White" ethnic background (88% British, 3% Irish, and 3% other White). Self-reported ethnic background in UK Biobank was previously demonstrated to very closely match genetically-defined ancestry based on principal components [10]; e.g., in a plot of the first two principal components, which separate European, African, and East Asian ancestries, nearly all self-reported White individuals cluster in the European upper-left corner of the plot (Extended Data Fig. 3). Consistent with this observation, we previously demonstrated that restricting to self-reported White individuals adequately addressed population stratification in standard association analysis pipelines (specifically, linear mixed model analysis or linear regression with principal component covariates on unrelated individuals [47]). However, given that our analyses here focus on rare variants, we took additional care to ensure that our results were not confounded by residual population structure.

Three lines of evidence indicated that residual population structure had not produced false positives in our results:

1. Quantile-quantile plots for our association results exhibited no deviation from the null distribution outside of the seven monogenic risk loci we identified (Extended Data Fig. 4).

2. Every risk variant (52 out of 52) that associated with acquisition of CN-LOH events in cis exhibited the expected direction of allelic bias toward removal of the risk allele (for *MPL*) or duplication of the risk allele (for *FH*, *NBN*, *MRE11*, *ATM*, *SH2B3*, and *TM2D3*) (Supplementary Table 7). In contrast, variants spuriously associated with mosaic CN-LOH events (e.g., due to uncorrected population stratification) would be randomly deleted or duplicated by CN-LOH events.

3. Individuals identified in our exome analyses (of self-reported White individuals with mosaic CN-LOH events) as carriers of rare coding or splice variants in frequently-targeted genes all clustered in the European corner of the PC plot (Extended Data Fig. 3).

Based on the above analyses, we concluded that self-reported White ethnic background was a sufficiently accurate indication of European ancestry for our analyses.

3 Phasing and imputation of the UK Biobank cohort

3.1 Phasing

The UK Biobank cohort was previously phased using SHAPEIT3 [10, 56]; however, to improve phasing accuracy [17, 18] and to expand the set of phased variants, we rephased the data set using Eagle2 [18], employing a multiple-run voting strategy [57] to optimize accuracy. The SHAPEIT3-phasing performed by UK Biobank included 670,739 autosomal markers present on both the BiLEVE and Biobank arrays that passed the following filters: (a) failed QC in at most 1 genotyping batch, (b) missingness <0.05, (c) MAF<0.0001 [10]. We performed five runs of phasing using Eagle2 on five distinct marker sets:

1. The same set of autosomal markers previously phased using SHAPEIT3.

2. A subset of 650,084 autosomal markers obtained by further excluding MAF<0.001 variants.

3. A separate set of 706,877 autosomal markers present on the Biobank array (but not necessarily the BiLEVE array) passing the following four filters: (i) allele frequency deviation <0.02 between the Biobank and BiLEVE arrays (for markers present on both the Biobank and BiLEVE array); (ii) missingness <0.15; (iii) either (iii-a) MAF>0.001 and passing filters used in SHAPEIT3 phasing, or (iii-b) MAF<0.001; (iv) Hardy-Weinberg disequilibrium $P>10^{-100}$ (according to `plink` [38]).

4. A larger set of 712,138 autosomal markers obtained using the same four filters as above, but relaxing the Hardy-Weinberg threshold to $P>10^{-200}$.

5. A larger set of 714,468 autosomal markers obtained using the same four filters as above, but relaxing the Hardy-Weinberg threshold to $P>4.9\times10^{-324}$ (the smallest representable double-precision floating point value).

In each phasing run, we ran Eagle2 (v2.3.5) on 105 overlapping chunks of ∼10,000 markers (with overlaps of at least 2,000 markers between consecutive chunks on the same autosome); on very large data sets, this partition-ligation approach improves Eagle2's accuracy because Eagle2 conditions on a fixed set of `--Kpbwt` haplotypes per individual within an input region. We set `--Kpbwt` to 100,000 for the first two runs and 80,000 for the remaining three runs, and we used the `--pbwtOnly` option to only use PBWT iterations [18].

We combined phasing results from the five Eagle2 runs (covering a total of 716,197 unique markers) using a voting approach [57], reasoning that phase switch errors incurred during different Eagle2 runs were likely to be partially independent. Specifically, we scanned through the phased haplotypes in order of genomic position, and at each successive variant, we set the phase of each heterozygous genotype by giving 7 votes to each phasing run containing the variant. These 7 votes were distributed among the most recently processed 7 hets from the run: each het voted according to the relative phase (estimated by Eagle2) between that het and the het currently under consideration. We implemented this approach to improve robustness against short 1–2 SNP "blips" that constitute a large fraction of phase switch errors in large data sets [17]; our goal was to maximize long-range phasing accuracy.

3.2 Benchmarking phasing accuracy using mosaic chromosomal alterations

We benchmarked the long-range phasing accuracy of our Eagle2-phased haplotypes as well as the SHAPEIT3-phased haplotypes provided by UK Biobank using a novel benchmarking method based on mosaic chromosomal alterations. The standard approach to benchmarking phasing accuracy is to analyze a data set that contains trios, comparing gold-standard trio phase to statistical phase estimated on trio children after removing trio parents. However, in this case, we were interested in comparing our accuracy to the accuracy of the existing SHAPEIT3 phasing [10], which had been already been performed on all individuals together (such that trio phase accuracy should be very high [58] and uninformative of phasing accuracy on unrelated individuals).

To overcome this challenge, we instead obtained gold standard phasing information from individuals who carried mosaic chr12 trisomies (the most common whole-chromosome mosaic event). Mosaic events produce allelic imbalances that provide information about phase within genotyping intensity data (i.e., BAF deviations) that is invisible within the genotype calls available to phasing algorithms. To apply this approach, we ran our hidden Markov model-based mCA detection algorithm using either Eagle2-estimated phase or SHAPEIT3-estimated phase, and we compared the numbers of phase switch errors in each data set detected by the HMM. We restricted our attention to 79 individuals with chr12 trisomies with $|\Delta\text{BAF}|$ in the range 0.02–0.05 (high enough that

switch errors are easily detectable, but low enough for the mosaicism not to compromise genotype calling accuracy).

Across these benchmark chromosomes, we observed that the SHAPEIT3 phasing achieved a long-range switch error rate of 0.060% (s.e.m. 0.005%), whereas our Eagle2 phasing (voting across five runs) achieved a long-range switch error rate of 0.027% (s.e.m. 0.004%), an improvement of $>$2x. Applying the same approach to benchmark the individual Eagle2 phasing runs showed that the voting approach achieved a $\sim$20% improvement in accuracy compared to each of the individual runs. We note that these benchmarks ignore small-scale phase switch errors (e.g., 1–2 SNP blips) that will be attributed by the HMM to measurement noise in BAF.

3.3 Imputation

The UK Biobank genetic data was previously imputed to $\sim$93 million autosomal variants [10] using the Haplotype Reference Consortium (HRC) panel [59] and a merge of the the UK10K and 1000 Genomes Phase 3 reference panels [60]. We augmented this imputed data set by further imputing very rare coding or splice variants contained on the BiLEVE array (used to genotyped 49,950 individuals [36]) but not on the Biobank array (used to genotype the remaining $\sim$90% of the cohort). To impute these variants to the remainder of the cohort, we first pre-phased the BiLEVE cohort using Eagle2 with `--Kpbwt=20000` using the same partition-ligation scheme we applied to the full cohort (105 overlapping chunks of $\sim$10,000 autosomal markers; Sec. 3.1). We imputed the BiLEVE-only variants into the full cohort using Minimac3 v2.0.1 [37].

We adopted a similar strategy to impute very rare variants from genic regions captured in exome sequencing of 49,960 UK Biobank participants [22] into the full cohort. We performed phasing and imputation on genotype calls from the SPB exome sequencing pipeline. We dropped singleton variants, phased the remaining exome-sequencing-derived variants together with variants genotyped on the UK Biobank array (using Eagle2 with `--Kpbwt=20000`), and imputed into the full cohort (using Minimac4 v1.0.1, with noncoding variants from the UK Biobank array used as an imputation scaffold).

We also re-imputed chromosomes with detected mCAs in order to obtain phase information at imputed variants, as the imputed data supplied in the UK Biobank imputation v3 release did not contain phase information. We performed this imputation using Minimac3 with the merged UK10K and 1000 Genomes Phase 3 reference panels.

4 Detection and calling of an inherited deletion variant in *MPL*

We discovered the inherited 454bp deletion variant associated with 1p CN-LOH by exploring genotyping intensities of carriers of the rs144279563 tag SNP [9]. We observed an unusual deviation in the total allelic intensities (LRR) for these carriers at four sites typed on the UK BiLEVE chip

(used to genotype $\sim$10% of the cohort): an increase in LRR for a probe at chromosome 1 base position 43,814,653 (hg19) and a decrease in LRR at 43,814,938, 43,814,963, and 43,814,979 (Extended Data Fig. 5a,b). Based on LRR at these four probes, we called 27 likely carriers of the structural variant among the 49,950 UK BiLEVE participants (Extended Data Fig. 5c).

Only the first probe (at base position 43,814,653) was included on the Biobank chip (used to genotype the remaining $\sim$90% of the cohort), so to call the structural variant in the remaining samples, we employed a hybrid approach using both LRR at 43,814,653 and imputation. First, we phased the structural variant in the UK BiLEVE cohort using Eagle2 [18] and imputed it into the remainder of the cohort using Minimac3 [37]. We then re-weighted the imputed allele probabilities for each individual according to the odds of observing the measured LRR at 43,814,653 assuming the individual was a carrier vs. non-carrier of the structural variant. We estimated these odds based on the empirical distribution of LRR among high-confidence carriers (imputed probability $>$0.99) vs. high-confidence non-carriers (imputed probability $<$0.01). This re-weighting modified the calls of 16 individuals and produced a final call set of 203 likely carriers of the structural variant in the full cohort.

Upon the release of exome sequencing data for 49,960 UK Biobank participants [22], we examined exome sequencing reads aligning to the region of *MPL* affected by the structural variant in individuals we had predicted to be carriers of the variant. We observed clear read support for a 454bp deletion removing base pairs 43,814,729 through 43,815,182 (spanning *MPL* exon 10): read pairs spanning this region exhibited unusually long insert sizes and clipped alignments (Extended Data Fig. 5d), and read depth in exon 10 was unusually low in all 32 predicted carriers of the structural variant who had been exome-sequenced (Extended Data Fig. 5e,f). Interestingly, we found no evidence of a duplication around 43,814,653 in exon 9 (Extended Data Fig. 5d–f, despite our earlier observation that carriers of the structural variant exhibited consistently high genotyping intensities (LRR) at this site (Extended Data Fig. 5a,b)). Based on the lack of read support for a duplication in exon 9, we believe the structural variant consists only of the 454bp deletion; the increase in LRR from genotyping at 43,814,653 could be a technical artifact arising from the probe being only 76bp away from the deletion.

5 Common variants influencing mCAs in *cis*

Our initial genotype–phenotype association analyses were well-powered to detect rare variants with large effects on *cis* CN-LOH mosaicism but were underpowered to detect weaker effects of more-common variants. To maximize power to detect common variants associated with CN-LOH mosaicism in *cis*, we performed a second genome-wide association analysis using a combined test for (i) association with CN-LOH events and (ii) allelic bias of CN-LOH directionality (i.e., tendency of CN-LOH events in hets to consistently duplicate vs. delete the risk allele; Methods). For

common variants, test (ii) can provide greater signal than (i): while a small fraction of individuals have CN-LOH events on a given chromosome arm (limiting the contribution of (i)), a large fraction of cases are heterozygous (allowing substantial signal from (ii)).

This test revealed two novel associations between common variants at the *TCL1A* and *DLK1* loci on 14q and acquired 14q CN-LOH mutations (Fisher's combined P=4.2×10^{-9} and 3.6×10^{-9}; Supplementary Table 12). Intriguingly, the reference alleles at both loci were recently observed to increase risk of mosaic Y chromosome loss in elderly males, a trait related to cell proliferation and cell cycle regulation [27, 61]. Here, 14q CN-LOH mutations in heterozygous carriers of these variants preferentially duplicate the same alleles (OR=1.91 (1.50–2.43) and 1.56 (1.28–1.90)), corroborating a pro-proliferative effect. (The *DLK1* locus also lies within an imprinted region that has previously been observed to be the target of parental bias in 14q CN-LOH mutations [31], raising the possibility of interaction between allelic and parental effects; however, familial data will be needed to investigate further.)

We also searched for associations between inherited variants and copy-number-altering mCAs (i.e., loss and gain events) in *cis* but did not find any associations aside from the *FRA10B* locus, at which we previously observed that fragile alleles confer risk of mosaic 10q deletions [9]. This association replicated here, with the common tag SNP rs11595735 (which tags rare fragile alleles) exhibiting P=5.2×10^{-169} association with 10q deletions. The fact that no other genotyped or imputed variants associated with losses or gains in *cis* indicates that no other fragile sites in Europeans influence mCA formation in the same way (or at least to the same extent) as *FRA10B*.

6 Inherited variants associated with mCAs in *trans*

A common haplotype in *TERT* broadly increases risk of clonal hematopoiesis involving any mosaic point mutation [7], and common variants also exert *trans* effects on the likelihood of mosaic *JAK2* V617F mutation [62] and sex chromosome loss [9, 27, 61, 63]. To identify more inherited haplotypes that similarly modify risk of autosomal mosaic chromosomal alterations in general, we conducted a genome-wide association analysis between common variants and presence of any detectable autosomal mCA (Methods). Three loci reached significance (P<5×10^{-8}): *TERT* (P=6.9×10^{-18}, OR=1.11 (1.08–1.14) for rs7705526), *TERC* (P=2.9×10^{-8}, OR=0.93 (0.91–0.96) for rs12638862), and *SP140* (P=9.4×10^{-9}, OR=1.08 (1.05–1.10) for rs62191195) (Supplementary Table 13).

The associations at *TERT* and *TERC* suggest that genetic differences in telomere maintenance make some individuals more susceptible to clonal expansions than others. Consistent with this hypothesis, we further observed that two additional telomere-length-associated SNPs [64] at *OBFC1* and *RTEL1* also associated with mCA susceptibility at nominal P<0.05 significance (Supplementary Table 14). For 6 out of 7 SNPs previously associated with telomere length [64], the telomere

length-increasing allele exhibited a risk-increasing effect sign for mCAs (Supplementary Table 14).

The lead associated variant in *SP140* matches the second-strongest association for chronic lymphocytic leukemia (CLL) [65], and we further observed that most CLL risk alleles increase risk of clonal hematopoiesis involving CLL-related +12 or 13q LOH events (Supplementary Note 8). These results demonstrate that common variants play a modest, quantitative role in altering processes that facilitate clonal expansion (in contrast to and in addition to the larger *cis* effects of variants on which mosaic CN-LOH mutations directly act by changing allele dosage).

7 Action of CN-LOH events on risk alleles for blood cancers

Genome-wide association studies have previously identified many inherited variants associated with increased risk of developing hematological malignancies such as chronic lymphocytic leukemia (CLL) and myeloproliferative neoplasms (MPN). We examined whether risk alleles identified by the largest GWAS conducted to date for CLL [65] and MPN [66] tended to be made homozygous by clonally expanded CN-LOH events. This hypothesis was very plausible for MPN GWAS hits given that three top MPN risk loci (*JAK2*, *ATM*, and *SH2B3*) also confer risk for mCA-associated CH.

We observed that CN-LOH clones in individuals heterozygous for MPN risk alleles did indeed tend to make these risk alleles homozygous (Supplementary Table 17). This effect was clearest at *JAK2*, *ATM*, and *SH2B3* ($P<6\times10^{-6}$ at each locus) but appeared to extend broadly to other MPN risk loci: among 24 risk alleles originally heterozygous in at least one CN-LOH clone, 19 were made homozygous more often than they were removed by CN-LOH mutations, whereas only 3 were made homozygous less often than they were removed (P=0.0004, one-sided binomial sign test); the remaining 2 alleles were made homozygous and removed in equal numbers of clones. Four MPN risk alleles exhibited directional biases significant at FDR$<$0.05 on their own: the aforementioned *JAK2*, *ATM*, and *SH2B3* alleles and a *TET2* allele.

In contrast to the results for MPN, none of the 46 CLL risk alleles exhibited a significant association with CN-LOH directionality ($P>0.01$ for all alleles), and we also did not observe a significant sign test across risk alleles (26 alleles were made homozygous more often than they were removed, while 19 alleles were removed more often than they were made homozygous; P=0.19, one-sided binomial sign test).

8 Shared genetic risk of CLL and mosaic +12 and 13q LOH

Chronic lymphocytic leukemia (CLL) is a highly heritable hematological malignancy, with 42 risk loci identified to date by GWAS on up to 6,200 cases [65,67–74]. Given the relatively large number of carriers of CLL-associated mosaic events in the UK Biobank data set ($\sim$2,000), we sought

to investigate the extent to which CLL risk alleles also influence risk of clonal hematopoiesis involving CLL-associated chromosomal alterations. We examined the two types of mCAs most strongly associated with CLL: mosaic trisomy 12 ("+12") and mosaic 13q LOH spanning *DLEU2* (including both del(13q) and 13q CN-LOH to maximize power).

For each of 46 independent lead variants at the 42 previously identified CLL risk loci [65], we performed association tests with three case-control phenotypes: mosaic +12, mosaic 13q LOH, and (as a check) CLL in UK Biobank. We restricted each association test to individuals who reported European ancestry, and we pruned to unrelated subsets of samples as in our *cis* GWAS analyses (Methods). For mosaic +12, we tested 634 cases, defined as individuals with a whole-chromosome 12 mosaic event (including unclassified events as well as events confidently classified as gains), and 378,107 controls, defined as individuals with no chr12 mosaic event. For 13q LOH, we tested 914 cases, defined as individuals with loss or CN-LOH events spanning *DLEU2*, and 378,048 controls, defined as individuals with no chr13 mosaic event. For CLL, we tested 656 cases, defined as individuals with any reported CLL (prevalent or incident), and 378,606 controls. These case sets contained modest overlap: of the 634 mosaic +12 cases, 78 had prevalent or incident CLL, and of the 914 mosaic 13q LOH cases, 243 had prevalent or incident CLL. Thirty-five individuals had both mosaic +12 and mosaic 13q LOH; 14 of the 35 also had prevalent or incident CLL. We performed association tests on imputed genotypes using logistic regression in `plink` [38] adjusting for age and sex as covariates. We verified that previously reported CLL risk variants replicated well in UK Biobank, with a regression coefficient of 0.85 (0.75–0.96) for observed vs. expected log(OR) and consistent effect directions for 43 of 46 variants (Supplementary Table 21, plotted in Supplementary Fig. 37 of `https://doi.org/10.1101/653691`).

We observed that most CLL risk variants conferred risk for either mosaic +12, mosaic 13q LOH, or both (Supplementary Table 21). Of the 46 CLL risk alleles tested, 40 had risk-increasing effect directions on mosaic +12, with 21 reaching nominal significance ($P<0.05$). For mosaic 13q LOH, 42 of 46 CLL risk alleles had risk-increasing effects, with 29 reaching nominal significance (Supplementary Table 21). Broadly, the estimated effects of these 46 risk alleles on mosaic +12 and 13q LOH risk were moderately smaller than their reported effects on CLL (log(OR) regression coefficients of 0.52 (0.40–0.64) and 0.63 (0.53–0.73), respectively).

Interestingly, we observed heterogeneity in the effects of variants on mosaic +12 vs. mosaic 13q LOH risk: some variants appeared to influence one form of mosaicism much more than the other (Supplementary Table 21). We quantified this genetic heterogeneity by performing bivariate BOLT-REML analysis [75] and estimated a genetic correlation of 0.79 (0.09)—significantly less than 1—between mosaic +12 and mosaic 13q LOH risk. These results suggest that different genetic mechanisms may lead to different subtypes of CLL that manifest different chromosomal alterations.

9 Chromosome arms with multiple overlapping CN-LOH events

In our previous analysis of the UK Biobank interim data, we observed that in a small fraction of carriers of mosaic CN-LOH mutations, we could detect multiple clonal expansions of CN-LOH mutations with different breakpoints, often attributable to high-risk inherited alleles made homozygous or removed from the genome by those mutations [9]. Upon extending this analysis to the full cohort, we identified 110 examples of multiple overlapping CN-LOH mutations occurring on the same chromosome arm (among a total of 8,185 CN-LOH events). For 68 of these 110 cases, the events could be attributed to a high-risk inherited or acquired variant on the affected arm (Supplementary Table 23). Specifically, most cases of multiple overlapping CN-LOH clones appeared to be explained by action on rare inherited *TM2D3* variants (found in 13 of 17 events on 15q), rare inherited *MPL* variants (found in 10 of 20 events on 1p), inherited *JAK2* 46/1 risk haplotypes (found in 32 of 37 events on 9p) that had presumably caused somatic *JAK2* V617F mutation [50–53]), and somatic deletions of the *DLEU* region (found in 13 of 14 events on 13q).

The remaining events were scattered across 18 different chromosome arms, raising the possibility that inherited allelic configurations on other chromosomal arms might in rare instances impart a very high risk of CN-LOH (e.g., by "lining up" a set of proliferative alleles on one homologous chromosome). To further explore this question, we examined CN-LOH events called in 356 individuals for whom a monozygotic twin was also present in the data set (178 twin-pairs). Six CN-LOH events were ascertained among these 356 individuals; in two of these cases (one twin-pair), the twin also had acquired a clone with a CN-LOH event; and in this family, the twins' acquired mutations affected the same chromosome arm (13q) in the same direction (i.e., the same parental haplotypes were gained and lost in both twins). These observations suggest that the alleles inherited on 13q in this pair of twins imparted very high risk of CN-LOH; however, we were unable to determine whether this high risk (if real) was contributed by a polygenic effect or by an ultra-rare, strong-effect variant we have not yet identified (e.g., in the *DLEU* locus commonly targeted by mCAs on 13q).

Taken together, these results suggest that while certain chromosome arms could in theory impart a very high risk of CN-LOH because they "line up" a set of proliferative alleles, this scenario probably happens quite rarely—which is reasonable given that alleles distributed across a chromosome arm assort approximately independently in the population, so the probability of lining up many proliferative alleles decreases exponentially with the number of loci.

10 Estimation of mortality risk conferred by mCAs

UK death registry data provided by UK Biobank reported 10,498 deaths on or before December 31, 2015 (the censoring date suggested by UK Biobank) from among 415,867 individuals of self-

reported European ancestry with no evidence of potential undiagnosed blood cancer based on anomalous blood counts (Methods). This censoring date corresponded to a median follow-up time of 6.9 years (range 5–10 years).

While the relatively small number of deaths in the cohort limited our power to identify links between mCAs and mortality, we observed that clones with gain of chromosome 8 (which contains the *MYC* oncogene) associated with increased mortality even in individuals with normal blood cell counts (P=3.5$\times$10^{-5}), reaching Bonferroni significance among the 78 events tested, presumably due to large effect size (OR=5.10 (2.41–9.85)). In this analysis we applied statistical tests analogous to our cancer outcome analyses, using Cochran-Mantel-Haenszel (CMH) tests to adjust for sex and age while accounting for case-control imbalance.

11 Additional discussion of clonal proliferation in the hematopoietic system

Hematopoiesis in adult humans involves a hierarchy of progressive cell division and differentiation that enables a relatively small pool of $\sim$10^4 hematopoietic stem cells (HSCs) to ultimately replenish a much larger supply of mature, terminally-differentiated blood cells that turns over at the rate of $\sim$10^{12} cells per day [76–78]. At the top of the hierarchy, HSCs have the unique ability to self-renew as well as differentiate; in contrast, hematopoietic progenitor cells in the middle tiers of the hierarchy can differentiate into increasingly large pools of increasingly differentiated cells downstream but cannot self-renew. The self-renewal property of HSCs enables maintenance of the HSC pool over the course of an individual's lifetime and gives rise to clonal lineages of HSCs and their progeny in the blood system.

The hierarchical nature of the hematopoietic system enables very different dynamics of cell division during self-renewal of the HSC pool at the top vs. during differentiation toward mature blood cells at the bottom. In lieu of direct observation of HSC division dynamics in vivo, previous studies have inferred an average rate of $\sim$1–2 HSC self-renewal divisions per year based on the rate of leukocyte telomere shortening with age [79, 80] and on X-chromosome inactivation patterns in females [81]. Recent work has further established heterogeneity among HSCs, with a rare population of dormant HSCs experiencing only four self-renewal divisions within an individual's lifetime [82, 83]. In contrast, the differentiation process that replenishes $\sim$10^{12} cells per day requires a few dozen cell divisions to reach this population size from a pool of $\sim$10^4 HSCs.

Recent studies have begun obtaining insights into the points in the hematopoietic hierarchy at which the somatic mutations observed in clonal hematopoiesis occur and enjoy proliferative advantage. Young et al. [84] used flow cytometry to sort peripheral blood cells from individuals with clonal hematopoiesis into myeloid and lymphoid compartments and observed the same clonal SNVs in both compartments, frequently with similar variant allele frequencies, and concluded that the mutations likely originated in HSCs. More recently, Arends et al. [85] performed flow-

sorting of both peripheral blood and bone marrow samples and traced clonal mutations to Lin– CD34+CD38– HSCs in bone marrow with subsequent expansion to myeloid progenitors; Jann et al. [86] obtained confirmatory results in similar analyses.

Interestingly, Arends et al.'s flow-sorting analyses revealed differing dynamics of clonal expansion in different individuals, with some exhibiting roughly constant expansion rates from HSC to progenitor to mature cells and others exhibiting greatest expansion during early differentiation. Additionally, mutant allelic fractions differed across different cell types.

The above studies have shed light on the manner in which clonal hematopoiesis mutations rise to detectable allelic fractions in peripheral blood, yet many details of this progression remain unclear. Clonality at different cell fractions in different lineages could be explained by either 1) a mutation conferring cell-type-specific survival or proliferative advantages or disadvantages; or 2) the mutation biasing differentiation toward one or more lineages and away from others. Future work will be necessary to determine, for each mutation, which of these mechanisms shape hematopoetic population dynamics and outcomes.

References

1. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nature Genetics* **44**, 651–658 (2012).

2. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature Genetics* **44**, 642–650 (2012).

3. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *New England Journal of Medicine* **371**, 2477–2487 (2014).

4. Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine* **371**, 2488–2498 (2014).

5. Machiela, M. J. *et al.* Characterization of large structural genetic mosaicism in human autosomes. *American Journal of Human Genetics* **96**, 487–497 (2015).

6. Vattathil, S. & Scheet, P. Extensive hidden genomic mosaicism revealed in normal tissue. *American Journal of Human Genetics* **98**, 571–578 (2016).

7. Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742–752 (2017).

8. Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400–404 (2018).

9. Loh, P.-R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).

10. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

11. Uziel, T. *et al.* Requirement of the MRN complex for ATM activation by DNA damage. *The EMBO Journal* **22**, 5612–5621 (2003).

12. Lee, J.-H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **308**, 551–554 (2005).

13. Deng, Y., Guo, X., Ferguson, D. O. & Chang, S. Multiple roles for MRE11 at uncapped telomeres. *Nature* **460**, 914 (2009).

14. Kimura, S., Roberts, A. W., Metcalf, D. & Alexander, W. S. Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. *Proceedings of the National Academy of Sciences* **95**, 1195–1200 (1998).

15. Solar, G. P. *et al.* Role of c-mpl in early hematopoiesis. *Blood* **92**, 4–10 (1998).

16. Seita, J. *et al.* Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction. *Proceedings of the National Academy of Sciences* **104**, 2349–2354 (2007).

17. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nature Genetics* **48**, 811–816 (2016).

18. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* **48**, 1443–1448 (2016).

19. Auer, P. L. *et al.* Rare and low-frequency coding variants in *CXCR2* and other genes are associated with hematological traits. *Nature Genetics* **46**, 629 (2014).

20. Schultz, K. A. P. *et al. PTEN*, *DICER1*, *FH*, and their associated tumor susceptibility syndromes: clinical features, genetics, and surveillance recommendations in childhood. *Clinical Cancer Research* **23**, e76–e82 (2017).

21. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* **46**, D1062–D1067 (2018).

22. Van Hout, C. V. *et al.* Whole exome sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank. *bioRxiv* (2019).

23. Meuwissen, T., Hayes, B. & Goddard, M. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).

24. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

25. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219 (2018).

26. Loh, P.-R. *et al.* Efficient Bayesian mixed model analysis increases association power in large cohorts. *Nature Genetics* **47**, 284–290 (2015).

27. Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).

28. Jaiswal, S. *et al.* Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *New England Journal of Medicine* **377**, 111–121 (2017).

29. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).

30. O'Keefe, C., McDevitt, M. A. & Maciejewski, J. P. Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood* **115**, 2731–2739 (2010).

31. Chase, A. *et al.* Profound parental bias associated with chromosome 14 acquired uniparental disomy indicates targeting of an imprinted locus. *Leukemia* **29**, 2069–2074 (2015).

32. Choate, K. A. *et al.* Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in *KRT10*. *Science* **330**, 94–97 (2010).

33. Tesi, B. *et al.* Gain-of-function *SAMD9L* mutations cause a syndrome of cytopenia, immunodeficiency, MDS and neurological symptoms. *Blood* **129**, 2266–2279 (2017).

34. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

35. 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 Medicine* **12**, 1–10 (2015).

36. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respiratory Medicine* **3**, 769–781 (2015).

37. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* **48**, 1284–1287 (2016).

38. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 1–16 (2015).

39. Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Research* **16**, 1136–1148 (2006).

40. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Research* **36**, e126 (2008).

41. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

42. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research* **47**, D886–D894 (2019).

43. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

44. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).

45. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2017).

46. Regier, A. A. *et al.* Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nature Communications* **9**, 4038 (2018).

47. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906–908 (2018).

48. Turner, J. J. *et al.* InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood* **116**, e90–e98 (2010).

49. Arber, D. A. *et al.* The 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

50. Jones, A. V. *et al. JAK2* haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nature Genetics* **41**, 446–449 (2009).

51. Kilpivaara, O. *et al.* A germline *JAK2* SNP is associated with predisposition to the development of *JAK2*V617F-positive myeloproliferative neoplasms. *Nature Genetics* **41**, 455–459 (2009).

52. Olcaydu, D. *et al.* A common *JAK2* haplotype confers susceptibility to myeloproliferative neoplasms. *Nature Genetics* **41**, 450–454 (2009).

53. Koren, A. *et al.* Genetic variation in human DNA replication timing. *Cell* **159**, 1015–1026 (2014).

54. Gusev, A. *et al.* Whole population, genome-wide mapping of hidden relatedness. *Genome Research* **19**, 318–326 (2009).

55. Kitsberg, D. *et al.* Allele-specific replication timing of imprinted gene regions. *Nature* **364**, 459 (1993).

56. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nature Genetics* **48**, 817–820 (2016).

57. Williams, A. L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. *American Journal of Human Genetics* **91**, 238–251 (2012).

58. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLOS Genetics* **10**, e1004234 (2014).

59. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279–1283 (2016).

60. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nature Communications* **6** (2015).

61. Wright, D. J. *et al.* Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nature Genetics* **49**, 674–679 (2017).

62. Hinds, D. A. *et al.* Germ line variants predispose to both *JAK2* V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121–1128 (2016).

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

64. Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics* **45**, 422–427 (2013).

65. Law, P. J. *et al.* Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nature Communications* **8**, 14175 (2017).

66. Bao, E. L. *et al.* Genetic predisposition to myeloproliferative neoplasms implicates hematopoietic stem cell biology. *bioRxiv* 790626 (2019).

67. Di Bernardo, M. C. *et al.* A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nature Genetics* **40**, 1204–1210 (2008).

68. Crowther-Swanepoel, D. *et al.* Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nature Genetics* **42**, 132 (2010).

69. Slager, S. L. *et al.* Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood* **117**, 1911–1916 (2011).

70. Slager, S. L. *et al.* Common variation at 6p21.31 (*BAK1*) influences the risk of chronic lymphocytic leukemia. *Blood* **120**, 843–846 (2012).

71. Berndt, S. I. *et al.* Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nature Genetics* **45**, 868–876 (2013).

72. Speedy, H. E. *et al.* A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nature Genetics* **46**, 56–60 (2014).

73. Sava, G. *et al.* Common variation at 12q24.13 (*OAS3*) influences chronic lymphocytic leukemia risk. *Leukemia* **29**, 748 (2015).

74. Berndt, S. I. *et al.* Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nature Communications* **7** (2016).

75. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance components analysis. *Nature Genetics* **47**, 1385–1392 (2015).

76. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).

77. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120–136 (2012).

78. Abkowitz, J. L., Catlin, S. N., McCallie, M. T. & Guttorp, P. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood* **100**, 2665–2667 (2002).

79. Shepherd, B. E., Guttorp, P., Lansdorp, P. M. & Abkowitz, J. L. Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. *Experimental Hematology* **32**, 1040–1050 (2004).

80. Sidorov, I., Kimura, M., Yashin, A. & Aviv, A. Leukocyte telomere dynamics and human hematopoietic stem cell kinetics during somatic growth. *Experimental Hematology* **37**, 514–524 (2009).

81. Catlin, S. N., Busque, L., Gale, R. E., Guttorp, P. & Abkowitz, J. L. The replication rate of human hematopoietic stem cells in vivo. *Blood* **117**, 4460–4466 (2011).

82. Foudi, A. *et al.* Analysis of histone 2b-gfp retention reveals slowly cycling hematopoietic stem cells. *Nature Biotechnology* **27**, 84 (2009).

83. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic stem cells count and remember self-renewal divisions. *Cell* **167**, 1296–1309 (2016).

84. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature Communications* **7** (2016).

85. Arends, C. M. *et al.* Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32**, 1908 (2018).

86. Jann, J.-C. *et al.* Comparative analysis of clonal hematopoiesis of multipotent stem cells in healthy elderly in blood and bone marrow. *Leukemia Research* **82**, 15 (2019).

Supplementary Table 1. Number of mosaic chromosomal alterations detected per chromosome.

Chromosome	N_{loss}	$N_{\text{CN-LOH}}$	N_{gain}	$N_{\text{undetermined}}$	N_{total}
chr1	97	1179	68	522	1866
chr2	214	264	26	159	663
chr3	74	225	174	185	658
chr4	138	245	24	116	523
chr5	172	131	88	110	501
chr6	119	311	33	194	657
chr7	176	176	23	118	493
chr8	69	127	156	141	493
chr9	52	706	112	258	1128
chr10	290	129	11	131	561
chr11	296	875	4	366	1541
chr12	81	251	530	342	1204
chr13	596	444	12	237	1289
chr14	161*	704	162	377	1404
chr15	44	407	223	332	1006
chr16	180	470	8	202	860
chr17	231	435	135	290	1091
chr18	58	100	200	153	511
chr19	17	339	48	275	679
chr20	458	204	8	133	803
chr21	58	138	153	228	577
chr22	137*	325	191	471	1124
All autosomes	3718	8185	2389	5340	19632

*Deletions on chr14 and chr22 include V(D)J recombination events (61 events on chr14 and 80 events on chr22).

Supplementary Table 2. Distribution of the number of detected somatic autosomal mCAs per individual.

mCA count	Frequency
0	465678
1	15520
2	1084
3	329
4	95
5	40
6	18
7	6
8	6
9	1
10	3
11	0
12	1
13	1
14	2
15	1
16	1
17	2
18	0
19	0
20	0
21	0
22	1

Most individuals with several detected mCAs have prevalent or incident blood cancers.

Supplementary Table 3. Fraction of individuals with detected mCAs as a function of age.

Age range	% of males with autosomal event (s.e.)	% of females with autosomal event (s.e.)
<45	1.8% (0.1%)	1.8% (0.1%)
45-50	2.1% (0.1%)	2.0% (0.1%)
50-55	2.6% (0.1%)	2.4% (0.1%)
55-60	3.5% (0.1%)	3.1% (0.1%)
60-65	4.7% (0.1%)	4.0% (0.1%)
>65	6.0% (0.1%)	4.9% (0.1%)

This table provides numerical data plotted in Extended Data Fig. 2f. Consistent with previous work [1, 2, 5, 6, 9], mosaic chromosomal alterations are detected more frequently with increasing age and in males.

Supplementary Table 4. Age and sex distributions of individuals with detected mCAs on each chromosome.

	Loss events				CN-LOH events				Gain events	
	p-arm		q-arm		p-arm		q-arm			
chr	Mean age	Frac. male	Mean age	Frac. male	Mean age	Frac. male	Mean age	Frac. male	Mean age	Frac. male
1	60.6 (1.1)	0.40 (0.08)	61.2 (0.9)	0.62 (0.08)	59.4 (0.3)	0.46 (0.02)	58.8 (0.4)	0.48 (0.02)	60.4 (0.8)	0.45 (0.06)
2	60.9 (0.6)	0.39 (0.04)	61.3 (0.9)	0.50 (0.07)	59.9 (0.7)	0.44 (0.05)	57.7 (0.6)	0.44 (0.04)	58.6 (1.4)	0.54 (0.10)
3	60.9 (1.2)	0.57 (0.08)	61.5 (1.6)	0.36 (0.10)	59.3 (0.7)	0.54 (0.05)	60.2 (0.7)	0.53 (0.05)	61.9 (0.5)	0.56 (0.04)
4	63.5 (1.1)	0.25 (0.13)	61.5 (0.7)	0.50 (0.05)	55.5 (1.3)	0.49 (0.08)	62.9 (0.5)	0.46 (0.04)	61.2 (1.9)	0.52 (0.11)
5	62.0 (2.7)	0.40 (0.16)	59.6 (0.6)	0.35 (0.04)	57.9 (1.7)	0.57 (0.14)	58.4 (0.8)	0.50 (0.05)	60.0 (0.7)	0.59 (0.05)
6	61.8 (1.3)	0.39 (0.09)	61.8 (0.7)	0.55 (0.06)	58.7 (0.5)	0.47 (0.03)	59.8 (1.0)	0.49 (0.06)	59.3 (1.5)	0.55 (0.09)
7	59.5 (1.2)	0.30 (0.08)	61.7 (0.5)	0.52 (0.04)	59.8 (0.9)	0.43 (0.06)	59.5 (0.8)	0.50 (0.05)	59.0 (1.8)	0.43 (0.11)
8	62.0 (0.8)	0.51 (0.07)	62.0 (1.3)	0.56 (0.13)	56.3 (1.4)	0.41 (0.09)	59.4 (0.8)	0.50 (0.05)	60.4 (0.6)	0.40 (0.04)
9	66.6 (1.1)	0.43 (0.20)	61.1 (1.4)	0.47 (0.08)	60.6 (0.4)	0.55 (0.03)	59.9 (0.4)	0.44 (0.03)	61.1 (0.6)	0.54 (0.05)
10	63.7 (2.1)	0.50 (0.22)	57.4 (0.5)	0.24 (0.03)	58.5 (1.3)	0.51 (0.08)	58.2 (0.9)	0.37 (0.05)	58.8 (2.8)	0.36 (0.15)
11	59.0 (1.3)	0.57 (0.08)	61.0 (0.4)	0.64 (0.03)	58.8 (0.3)	0.50 (0.02)	60.5 (0.4)	0.55 (0.03)	–	–
12	62.3 (1.1)	0.57 (0.09)	60.7 (1.1)	0.52 (0.08)	57.8 (1.2)	0.34 (0.07)	58.9 (0.5)	0.48 (0.04)	62.3 (0.3)	0.55 (0.02)
13	–	–	62.0 (0.3)	0.60 (0.02)	–	–	60.3 (0.4)	0.54 (0.02)	56.7 (2.9)	0.75 (0.13)
14	–	–	61.3 (0.5)	0.62 (0.04)	–	–	60.0 (0.3)	0.47 (0.02)	63.6 (0.4)	0.59 (0.04)
15	–	–	62.2 (1.1)	0.46 (0.08)	–	–	59.6 (0.4)	0.48 (0.03)	65.1 (0.3)	0.79 (0.03)
16	59.1 (0.6)	0.30 (0.04)	61.7 (1.1)	0.61 (0.08)	59.2 (0.5)	0.47 (0.03)	59.4 (0.5)	0.48 (0.03)	55.0 (3.8)	0.50 (0.19)
17	61.7 (0.5)	0.51 (0.04)	60.9 (0.9)	0.40 (0.07)	59.6 (0.8)	0.54 (0.05)	58.6 (0.4)	0.46 (0.03)	61.3 (0.6)	0.49 (0.04)
18	60.6 (1.1)	0.67 (0.09)	61.5 (1.6)	0.32 (0.10)	60.3 (1.8)	0.71 (0.13)	59.7 (0.9)	0.34 (0.05)	61.7 (0.5)	0.60 (0.03)
19	56.8 (2.3)	0.70 (0.15)	61.7 (2.5)	0.71 (0.18)	58.8 (0.7)	0.42 (0.04)	59.9 (0.5)	0.47 (0.04)	59.2 (1.1)	0.69 (0.07)
20	62.8 (1.4)	0.50 (0.15)	61.7 (0.3)	0.66 (0.02)	57.4 (1.3)	0.44 (0.08)	58.8 (0.6)	0.49 (0.04)	57.4 (1.3)	0.25 (0.16)
21	–	–	60.7 (0.9)	0.34 (0.06)	–	–	58.1 (0.7)	0.49 (0.04)	61.4 (0.5)	0.68 (0.04)
22	–	–	62.9 (0.5)	0.50 (0.04)	–	–	60.6 (0.4)	0.46 (0.03)	60.8 (0.5)	0.48 (0.04)

This table provides numerical data plotted in Extended Data Fig. 1b. (Events detected in fewer than 50 individuals were excluded from Extended Data Fig. 1b for clarity, and events detected in fewer than 5 individuals are excluded here. Chromosomes 13, 14, 15, 21, and 22 are acrocentric and have little or no p-arm genotyping.)

Supplementary Table 5. Enrichment of mCAs in individuals with anomalous (top 1%) blood indices.

mCA	Blood index	*P*-value	*q*-value	OR (95% CI)
1p–	Lymphocyte #	1.5e-7	3.4e-6	31.3 (12.5–78.4)
1p–	Lymphocyte %	4.1e-9	1.1e-7	38.6 (16.1–92.4)
1p–	RBC dist. width	0.0019	0.022	13.5 (4.0–45.2)
1q–	Lymphocyte #	0.00015	0.0021	17.2 (6.0–49.9)
1q–	Lymphocyte %	0.00015	0.0021	17.2 (6.0–49.9)
1p=	Platelet #	8.9e-5	0.0014	3.0 (1.9–4.9)
1p=	Platelet crit	8.9e-5	0.0014	3.0 (1.9–4.9)
1+	Monocyte #	0.002	0.022	8.1 (2.9–22.4)
1+	Monocyte %	0.002	0.022	8.1 (2.9–22.4)
1+	RBC dist. width	0.002	0.022	8.1 (2.9–22.4)
1+	Platelet dist. width	0.002	0.022	8.1 (2.9–22.4)
2q–	Lymphocyte #	2e-7	4.4e-6	19.8 (8.8–44.7)
2q–	Lymphocyte %	3.8e-6	7.8e-5	16.5 (7.0–39.2)
2+	Lymphocyte #	0.00051	0.0066	22.9 (6.5–80.3)
3p–	Neutrophil %	0.003	0.031	11.4 (3.5–37.8)
3p–	Platelet #	9.7e-6	0.00017	20.6 (7.9–54.1)
3p–	Platelet crit	0.00019	0.0028	15.9 (5.5–45.6)
3+	Lymphocyte #	1.9e-12	7.4e-11	16.2 (9.2–28.5)
3+	Lymphocyte %	6.2e-9	1.7e-7	12.3 (6.6–23.0)
4q–	Monocyte #	6.2e-5	0.00098	7.6 (3.5–16.5)
4q–	Monocyte %	7e-6	0.00013	8.8 (4.3–18.2)
4q=	Monocyte #	0.00025	0.0036	5.2 (2.5–10.5)
4q=	Monocyte %	1.1e-12	4.2e-11	11.6 (7.0–19.3)
5q–	Lymphocyte #	0.0018	0.021	5.0 (2.2–11.3)
5q–	Monocyte %	4.4e-5	0.0007	6.7 (3.3–13.8)
5q–	RBC dist. width	0.0018	0.021	5.0 (2.2–11.3)
5+	Lymphocyte #	4.6e-6	9e-5	11.8 (5.4–25.8)
5+	RBC dist. width	0.0044	0.043	6.4 (2.3–17.6)
6q–	Lymphocyte #	4.4e-6	8.8e-5	16.1 (6.8–38.1)
6q–	Lymphocyte %	0.0009	0.012	10.2 (3.6–28.5)
7q–	Lymphocyte #	7.6e-11	2.7e-9	16.1 (8.7–29.6)
7q–	Lymphocyte %	1.2e-9	3.9e-8	14.6 (7.7–27.4)
8p–	Lymphocyte #	0.00051	0.0066	12.0 (4.3–33.9)
8p–	Lymphocyte %	3.3e-5	0.00055	15.5 (6.0–39.8)
9p=	Basophil #	0.001	0.013	3.4 (1.8–6.4)
9p=	Monocyte #	0.0036	0.037	3.1 (1.6–6.0)
9p=	Neutrophil #	6.7e-12	2.5e-10	7.5 (4.8–11.7)
9p=	Neutrophil %	3.6e-10	1.2e-8	6.7 (4.2–10.7)
9p=	Red #	6.9e-24	5.3e-22	12.3 (8.6–17.7)
9p=	Hematocrit	1.3e-15	6.7e-14	9.0 (6.0–13.6)
9p=	RBC dist. width	1.5e-29	1.2e-27	14.5 (10.3–20.4)
9p=	Platelet #	3.2e-97	1.1e-94	40.6 (31.5–52.2)
9p=	Platelet crit	5.1e-94	1.4e-91	39.2 (30.4–50.6)
9p=	Platelet dist. width	1e-13	4.8e-12	8.3 (5.4–12.6)
9q=	Lymphocyte #	0.0029	0.031	3.2 (1.6–6.1)
9+	Lymphocyte #	3.9e-5	0.00063	8.3 (3.8–17.9)
9+	Basophil #	0.0023	0.025	5.8 (2.3–14.2)
9+	Monocyte #	2.2e-9	6.3e-8	13.6 (7.3–25.6)
9+	Neutrophil #	3e-8	7.3e-7	12.2 (6.3–23.6)
9+	Neutrophil %	2.2e-9	6.3e-8	13.6 (7.3–25.6)
9+	RBC dist. width	5.1e-13	2.1e-11	18.1 (10.2–31.9)
9+	Platelet #	1.5e-10	5.2e-9	15.1 (8.2–27.7)
9+	Platelet crit	2.2e-9	6.3e-8	13.6 (7.3–25.6)
9+	Platelet dist. width	2.2e-9	6.3e-8	13.6 (7.3–25.6)
11q–	Lymphocyte #	7.3e-17	4.3e-15	12.9 (8.3–20.2)
11q–	Lymphocyte %	1.3e-9	4.2e-8	8.5 (5.0–14.4)
11p=	Monocyte %	0.0017	0.021	2.8 (1.6–5.0)
11q=	Lymphocyte %	0.0024	0.026	3.0 (1.6–5.7)
11q=	Basophil #	0.0024	0.026	3.0 (1.6–5.7)
12q–	Lymphocyte %	0.0044	0.043	9.9 (3.0–32.5)
12q=	Lymphocyte #	0.0019	0.022	4.2 (2.0–9.0)
12+	Lymphocyte #	1.6e-77	3.4e-75	25.0 (19.6–32.0)
12+	Lymphocyte %	1.3e-58	1.6e-56	19.8 (15.2–25.8)
12+	Basophil #	5e-8	1.2e-6	4.9 (3.1–7.7)
12+	Monocyte #	5e-6	9.6e-5	4.1 (2.5–6.7)
13q–	Lymphocyte #	1.9e-293	2.1e-290	106.9 (87.8–130.3)
13q–	Lymphocyte %	3.6e-242	1.9e-239	82.1 (67.3–100.1)
13q–	Basophil #	3.8e-16	2e-14	7.7 (5.3–11.2)
13q–	Monocyte #	1.6e-13	7e-12	6.8 (4.6–10.2)
13q–	Platelet dist. width	0.0029	0.031	2.8 (1.5–5.1)
13q=	Lymphocyte #	8.2e-73	1.5e-70	25.9 (20.0–33.4)
13q=	Lymphocyte %	3.2e-67	5e-65	24.1 (18.6–31.3)
13q=	Monocyte #	0.0004	0.0055	3.4 (1.9–6.0)
13q=	RBC dist. width	0.0046	0.044	2.8 (1.5–5.2)
14q–	Lymphocyte #	1.8e-53	1.8e-51	63.0 (42.6–93.3)
14q–	Lymphocyte %	1.8e-53	1.8e-51	63.0 (42.6–93.3)
14q–	Basophil #	0.0044	0.043	4.9 (2.0–12.0)
14q=	Monocyte %	0.00049	0.0066	2.7 (1.6–4.4)
15q–	Lymphocyte #	3.6e-6	7.5e-5	26.1 (9.7–69.9)
15q–	Lymphocyte %	9e-5	0.0014	19.8 (6.8–58.0)
15q–	Basophil #	0.0017	0.021	14.2 (4.2–47.5)
16q–	Lymphocyte #	7.2e-7	1.5e-5	22.9 (9.4–55.6)
16q–	Lymphocyte %	1.6e-5	0.00028	18.4 (7.1–47.7)
16q–	Monocyte #	0.004	0.04	10.2 (3.1–33.7)
17p–	Lymphocyte #	4.8e-18	3.2e-16	18.2 (11.3–29.4)
17p–	Lymphocyte %	8.8e-9	2.3e-7	10.2 (5.6–18.5)
18p–	Lymphocyte %	0.001	0.013	17.5 (5.1–59.7)
18+	Lymphocyte #	2.2e-17	1.4e-15	16.7 (10.4–26.9)
18+	Lymphocyte %	2e-8	5.1e-7	9.4 (5.2–17.0)
19+	Lymphocyte #	3.6e-13	1.6e-11	47.3 (22.2–100.4)
19+	Lymphocyte %	2.1e-8	5.2e-7	28.9 (12.5–67.2)
19+	Platelet dist. width	0.0036	0.037	10.6 (3.2–34.9)
20q–	Neutrophil %	3.3e-7	7.2e-6	4.8 (2.9–7.7)
20q–	RBC dist. width	7.3e-6	0.00013	4.2 (2.5–7.0)
20q–	Platelet #	0.0015	0.019	3.0 (1.7–5.5)
20q–	Platelet crit	0.00045	0.0061	3.3 (1.9–5.9)
20q–	Platelet dist. width	3.3e-7	7.2e-6	4.8 (2.9–7.7)
20q=	Lymphocyte %	0.0032	0.033	4.4 (1.9–9.9)
21q–	Lymphocyte %	0.0036	0.037	10.6 (3.2–34.9)
21q–	Platelet dist. width	0.00025	0.0036	14.7 (5.1–42.0)
22q–	Lymphocyte #	1.4e-63	1.9e-61	90.0 (60.1–134.8)
22q–	Lymphocyte %	1.6e-48	1.4e-46	63.7 (42.1–96.3)
22+	Lymphocyte #	6.1e-6	0.00011	6.6 (3.5–12.5)
22+	Lymphocyte %	1.2e-8	3e-7	8.7 (4.9–15.4)

This table provides numerical data plotted in Extended Data Fig. 1c. Mosaic chromosomal alterations significantly enriched (at an FDR threshold of 0.05; one-sided Fisher's exact test) in individuals with anomalous blood indices (top 1% among *N*=455,009 self-reported white individuals) are reported. Events were grouped by chromosome and copy number, with loss and CN-LOH events subdivided by p-arm vs. q-arm. (We did not subdivide gain events by arm because most gain events are whole-chromosome trisomies; e.g., "3+" combines all gains—partial or complete—on chromosome 3.)

Supplementary Table 6. Numbers of cases and controls for association tests with CN-LOH mutations in *cis*.

Arm	Locus	N_{case}	N_{control}
1p	*MPL*	633	377674
1q	*FH*	666	377674
8q	*NBN*	76	379049
9p	*JAK2*	394	378410
11q	*MRE11*	520	378073
11q	*ATM*	581	378073
12q	*SH2B3*	250	378874
14q	*TCL1A*	1021	378180
14q	*DLK1*	1052	378180
15q	*TM2D3*	605	378617

Sample sets were determined by first filtering on ancestry and relatedness, and then at each locus, defining cases to be individuals with a mosaic event spanning the locus (or within 4Mb of the locus) likely to be a CN-LOH event (Methods). Individuals with likely CN-LOH events on the same chromosome but not within 4Mb of the locus were excluded from association analyses.

Supplementary Table 7. Rare coding or splice variants associated at FDR<0.05 significance with mosaic CN-LOH mutations in *cis*.

								GWAS		Allelic shift in hets		
Chr	Position[a]	Variant	Effect[b]	Alleles[c]	AF[d]	Source	INFO/R2	*P*	OR (95% CI)	N_{REF}[e]	N_{ALT}	*P*
***MPL*: 28/61 tested variants significant at FDR<0.05**												
1	43803600	rs146249964	splice donor	T/A	0.0001	HRC imp	0.685	2.8×10^{-23}	97 (55–171)	12	0	0.00049
1	43803817	rs148434485	stop gained	C/T	2×10^{-5}	BB array	–	1.6×10^{-6}	128 (37–446)	2	0	0.5
1	43803824	rs145714475	missense	T/C	2×10^{-5}	HRC imp	0.394	1.9×10^{-6}	120 (35–414)	3	0	0.25
1	43803835	rs764333753	missense	A/G	3×10^{-5}	WES imp	0.782	1.7×10^{-2}	31 (4–235)	1	0	1
1	43803877	rs766172846	missense	T/C	4×10^{-5}	WES imp	0.849	7.7×10^{-4}	37 (9–156)	2	0	0.5
1	43803903	rs142565191	splice donor	G/A	4×10^{-5}	WES imp	0.914	7.5×10^{-6}	72 (22–238)	3	0	0.25
1	43804234	rs587778514	frameshift	CCT/C	1×10^{-5}	BB array	–	3.9×10^{-5}	199 (40–987)	2	0	0.5
1	43804305	rs28928907	missense	G/C	0.0006	BB array	–	1.9×10^{-130}	142 (111–184)	70	0	1.7×10^{-21}
1	43804375	rs587778515	frameshift	CT/C	0.0002	BB array	–	7.0×10^{-41}	105 (68–161)	24	0	1.2×10^{-7}
1	43804396	rs752453717	splice modifier	G/C	0.0003	BB array	–	5.8×10^{-36}	74 (48–113)	24	0	1.2×10^{-7}
1	43804957	rs764904424	missense	C/G	0.0001	WES imp	0.750	2.1×10^{-8}	35 (15–79)	6	0	0.031
1	43805052	rs6088	missense	G/A	9×10^{-5}	WES imp	0.825	8.3×10^{-10}	61 (26–141)	6	0	0.031
1	43805059	rs769867913	missense	G/A	2×10^{-5}	WES imp	0.439	1.4×10^{-2}	37 (5–282)	1	0	1
1	43805656	rs144210383	missense	G/T	0.0001	WES imp	0.676	5.3×10^{-9}	44 (19–101)	6	0	0.031
1	43805686	rs587778518	frameshift	C/CCTGG	3×10^{-5}	WES imp	0.825	1.6×10^{-2}	33 (4–249)	1	0	1
1	43805713	rs121913611	missense	C/T	0.0002	BB array	–	3.3×10^{-28}	102 (61–171)	17	0	1.5×10^{-5}
1	43806073	1:43806073	missense	A/C	2×10^{-5}	WES imp	0.853	1.5×10^{-4}	92 (21–408)	2	0	0.5
1	43812115	rs769297582	splice acceptor	G/C	2×10^{-5}	WES imp	0.760	5.1×10^{-7}	199 (54–737)	3	0	0.25
1	43812574	rs200454070	missense	G/A	3×10^{-5}	WES imp	0.568	2.0×10^{-2}	26 (4–192)	1	0	1
1	43814551	rs765671565	missense	T/A	9×10^{-6}	WES imp	0.771	5.9×10^{-3}	100 (12–827)	1	0	1
1	43814590	rs1175548872	missense	G/C	1×10^{-5}	WES imp	0.898	7.5×10^{-3}	75 (9–597)	1	0	1
1	43814627	rs754859909	stop gained	G/A	7×10^{-5}	WES imp	0.939	1.7×10^{-16}	126 (61–258)	9	0	0.0039
1	43814673	rs923814653	missense	G/T	3×10^{-5}	WES imp	0.966	4.1×10^{-4}	52 (12–221)	3	0	0.25
1	43814729	454bp del[f]	exon 10 deletion	ref/del	0.0002	array LRR	–	3.6×10^{-58}	153 (104–225)	31	0	9.3×10^{-10}
1	43815009	rs121913615	missense	G/T	2×10^{-5}	WES imp	0.936	1.4×10^{-2}	37 (5–282)	1	0	1
1	43817942	rs369156948	stop gained	C/T	3×10^{-5}	HRC imp	0.225	4.8×10^{-8}	114 (39–333)	4	0	0.12
1	43817973	rs971379181	frameshift	CG/C	3×10^{-5}	BB array	–	5.8×10^{-13}	240 (93–618)	6	0	0.031
1	43818435	rs1366403560	stop gained	C/T	2×10^{-5}	WES imp	0.866	1.3×10^{-4}	100 (22–446)	2	0	0.5
***FH*: 1/41 tested variants significant at FDR<0.05**												
1	241675301	rs199822819	missense	G/C	0.0003	WES imp	0.869	4.9×10^{-11}	28 (14–55)	1	8	0.039
***NBN*: 2/48 tested variants significant at FDR<0.05**												
8	90983420	rs777460725	missense	A/C	0.0001	WES imp	0.794	8.1×10^{-5}	114 (28–465)	0	2	0.5
8	90983441	rs1187082186	frameshift	ATTTGT/A	0.0002	WES imp	0.844	4.8×10^{-13}	210 (92–484)	0	6	0.031
***MRE11*: 1/42 tested variants significant at FDR<0.05**												
11	94189489	rs587781384	stop gained	C/A	4×10^{-5}	WES imp	0.945	5.6×10^{-10}	130 (50–338)	0	5	0.062
***ATM*: 13/352 tested variants significant at FDR<0.05**												
11	108127067	rs1137887	splice modifier	G/A	4×10^{-5}	WES imp	0.768	9.6×10^{-6}	65 (20–214)	0	2	0.5
11	108141801	rs786203054	missense	T/G	7×10^{-6}	BB array	–	1.2×10^{-5}	437 (73–2618)	0	2	0.5
11	108155007	rs781357995	frameshift	AG/A	0.0001	WES imp	0.888	3.0×10^{-9}	48 (21–111)	0	6	0.031
11	108172425	rs587779844	missense	C/T	0.0001	BB array	–	3.5×10^{-20}	96 (52–177)	0	12	0.00049
11	108175420	rs786204751	stop gained	C/T	2×10^{-5}	BLVE imp	0.467	1.6×10^{-4}	87 (20–380)	0	1	1
11	108175528	rs376603775	stop gained	C/T	6×10^{-5}	BLVE imp	0.848	2.8×10^{-5}	44 (14–143)	0	4	0.12
11	108179837	rs774925473	splice modifier	A/G	8×10^{-5}	BLVE imp	0.677	6.8×10^{-5}	33 (10–104)	0	3	0.25
11	108181006	rs56399311	missense	A/G	8×10^{-5}	WES imp	0.957	1.7×10^{-6}	44 (16–120)	0	4	0.12
11	108201108	rs56399857	missense	T/G	0.0002	WES imp	0.921	4.9×10^{-5}	18 (6.6–48)	0	4	0.12
11	108202611	rs587776547	inframe deletion	CTCTAGAATT/C	7×10^{-5}	WES imp	0.754	8.5×10^{-9}	73 (29–183)	0	5	0.062
11	108206686	rs371638537	stop gained	A/T	6×10^{-5}	WES imp	0.877	9.9×10^{-4}	33 (8–135)	0	2	0.5
11	108216545	rs587779872	missense	C/T	2×10^{-5}	WES imp	0.594	3.6×10^{-11}	251 (89–706)	0	5	0.062
11	108224608	rs17174393	splice donor	G/A	4×10^{-5}	WES imp	0.824	6.5×10^{-4}	41 (10–170)	0	2	0.5
***SH2B3*: 2/57 tested variants significant at FDR<0.05**												
12	111885295	rs148636776	missense	G/A	0.0004	WES imp	0.861	4.0×10^{-5}	19 (7–50)	0	5	0.062
12	111885310	rs72650673	missense	G/A	0.002	WES imp	0.882	3.1×10^{-8}	11 (5.8–20)	1	8	0.039
***TM2D3*: 5/15 tested variants significant at FDR<0.05**												
15	102151467	70kb del[g]	gene deletion	ref/del	0.0003	array LRR	–	9.8×10^{-224}	555 (425–724)	2	110	2.4×10^{-30}
15	102182739	rs113189685	missense	G/T	3×10^{-5}	WES imp	0.544	2.8×10^{-8}	132 (45–389)	1	3	0.62
15	102182749	rs754640606	missense	G/C	5×10^{-5}	WES imp	0.769	1.2×10^{-40}	544 (289–1025)	0	19	3.8×10^{-6}
15	102182761	rs976377433	missense	A/G	3×10^{-5}	WES imp	0.761	2.3×10^{-8}	140 (47–413)	0	4	0.12
15	102190214	rs768556490	frameshift	G/GT	3×10^{-5}	WES imp	0.850	8.2×10^{-29}	758 (327–1759)	1	11	0.0063

See next page for caption.

Caption for Supplementary Table 7.

P-values from two independent statistical tests are reported: (i) a two-sided Fisher's exact test treating individuals with a mosaic CN-LOH mutation in *cis* as cases ($N \geq 378{,}307$ individuals; Supplementary Table 6); and (ii) a binomial test for biased allelic imbalance in heterozygous cases. Loci reaching FDR<0.05 significance in the first test (based on the number of variants tested in each gene) are reported. Note that both tests are two-sided (to retain consistency with Extended Data Table 1); *P*-values for biased allelic imbalance in the expected directions (removing rare alleles in *MPL* and duplicating rare alleles in other genes) would be half the values reported in the last column of this table. For full details of statistical tests, see Methods.

[a]Base pair position in hg19 coordinates.
[b]Variant effects according to Ensembl VEP [43] (coding variants) or ClinVar [21] (splice variants).
[c]Reference/alternate allele.
[d]Alternate allele frequency (in UK Biobank European-ancestry individuals).
[e]Number of mosaic individuals heterozygous for the variant in which the somatic event shifted the allelic balance in favor of the reference allele (by duplication of its chromosomal segment and loss of the homologous segment).
[f]This 454bp deletion spans 1:43,814,729-43,815,182, deleting *MPL* exon 10 (Extended Data Fig. 5).
[g]This $\sim$70kb deletion spans 15:102.15–102.22Mb, deleting *TM2D3* and part of *TARSL2* [9].

Supplementary Table 8. Previously reported CN-LOH risk variants at *MPL* and *ATM* tag likely causal coding variants.

Locus	Previously reported variant [9]	Likely causal coding variant in Extended Data Table 1	R^2
MPL	rs182971382	rs28928907 (missense)	0.83
MPL	rs144279563	454bp del (exon 10 deletion)	0.36
MPL	rs369156948	rs369156948 (stop gained)	1
ATM	rs532198118	rs587779844 (missense)	0.19

Note that R^2 reported above may be underestimated due to imputation error.

Supplementary Table 9. Numbers of distinct coding or splice variants at each risk locus likely to causally drive associations with mosaic CN-LOH events in *cis*.

Gene	Mosaic event	Variants associated at Bonferroni significance	Variants associated at FDR<0.05 significance	Additional coding or splice variants contributing to ultra-rare burden
MPL	1p CN-LOH	17	28	+4
FH	1q CN-LOH	1	1	+2
NBN	8q CN-LOH	2	2	+0
MRE11	11q CN-LOH	1	1	+1
ATM	11q CN-LOH	10	13	+6
SH2B3	12q CN-LOH	2	2	+5
TM2D3	15q CN-LOH	5	5	+3
Total		38	52	+21

This table summarizes the numbers of likely-causal variants at each CN-LOH risk locus identified by our association analyses (Extended Data Table 1 and Supplementary Table 7) and burden analyses (Supplementary Table 10).

Supplementary Table 10. Burden of additional ultra-rare coding or splice variants in genes frequently targeted by CN-LOH events in *cis*.

Gene	Mosaic event	Exome-sequenced mosaic individuals not carrying FDR-significant variant (Supplementary Table 7)	Carriers of ultra-rare coding or splice variants among these individuals		
			Observed	Expected	*P*
MPL	1p CN-LOH	37	5	0.067	8.4×10^{-9}
FH	1q CN-LOH	51	2	0.132	0.0079
NBN	8q CN-LOH	5	0	0.017	1
MRE11	11q CN-LOH	52	1	0.222	0.20
ATM	11q CN-LOH	57	6	0.682	6.3×10^{-5}
SH2B3	12q CN-LOH	25	5	0.110	8.3×10^{-8}
TM2D3	15q CN-LOH	44	3	0.051	2.0×10^{-5}
DNMT3A	2p CN-LOH	16	4	0.206	4.4×10^{-5}
TET2	4q CN-LOH	21	6	0.197	3.3×10^{-8}
JAK2	9p CN-LOH	33	15	0.217	1.7×10^{-24}

For each gene, we examined individuals with CN-LOH events spanning the gene (not already explained by any of the 52 variants identified in our association analyses) and tabulated the number of such individuals who carried a rare coding or splice variant under consideration (see Methods). We then computed a burden P-value using a one-sided binomial test comparing the observed count to expectation. Note that the counts of observed carriers include two 1p CN-LOH individuals with the same *MPL* variant (rs1362911656, 1:43814994_T_C) and fifteen 9p CN-LOH individuals with *JAK2* V617F. An additional five 9p CN-LOH individuals had at least one read supporting *JAK2* V617F (but did not have *JAK2* V617F genotype calls). Allelic read depth analyses indicated that all or most of the rare variant burden in the seven inherited risk loci arose from inherited variants, while all or most of the burden in *DNMT3A*, *TET2*, and *JAK2* arose from somatic point mutations (Extended Data Fig. 8).

Supplementary Table 11. Rare coding or splice variants carried by exome-sequenced individuals with mosaic CN-LOH events spanning frequently-targeted genes.

Gene	Mosaic event	Fraction of mosaic individuals with a rare coding/splice variant	Count	Variant	Effect
MPL	1p CN-LOH	39 / 71	8	1:43804305_G_C	missense
		(expected: 0.525 / 71)	5	1:43804396_G_C	splice modifier
			4	454bp del	exon 10 deletion
			3	1:43805713_C_T	missense
			2	1:43804375_CT_C	frameshift
			2	1:43814994_T_C	missense
			1	1:43803600_T_A	splice donor
			1	1:43803903_G_A	splice donor
			1	1:43804268_C_T	stop gained
			1	1:43804957_C_G	missense
			1	1:43805052_G_A	missense
			1	1:43805059_G_A	missense
			1	1:43805656_G_T	missense
			1	1:43805686_C_CCTGG	frameshift
			1	1:43812115_G_C	splice acceptor
			1	1:43812574_G_A	missense
			1	1:43814563_C_G	missense
			1	1:43814627_G_A	stop gained
			1	1:43814673_G_T	missense
			1	1:43815009_G_T	missense
			1	1:43818405_C_G	missense
FH	1q CN-LOH	3 / 52	1	1:241675301_G_C	missense
		(expected: 0.163 / 52)	1	1:241675313_C_T	missense
			1	1:241675443_C_A	missense
NBN	8q CN-LOH	2 / 7	1	8:90983420_A_C	missense
		(expected: 0.027 / 7)	1	8:90983441_ATTTGT_A	frameshift
MRE11	11q CN-LOH	2 / 57	1	11:94189447_G_A	missense
		(expected: 0.247 / 57)	1	11:94189489_C_A	stop gained
ATM	11q CN-LOH	12 / 64	2	11:108155007_AG_A	frameshift
		(expected: 0.880 / 64)	1	11:108115595_G_T	missense
			1	11:108121479_CTG_C	frameshift
			1	11:108121546_AC_A	frameshift
			1	11:108127067_G_A	splice modifier
			1	11:108159805_T_C	missense
			1	11:108172383_T_C	missense
			1	11:108179837_A_G	splice modifier
			1	11:108199833_G_A	missense
			1	11:108202611_CTCTAGAATT_C	inframe deletion
			1	11:108216545_C_T	missense
SH2B3	12q CN-LOH	6 / 26	1	12:111856537_G_GT	frameshift
		(expected: 0.226 / 26)	1	12:111856620_T_TGC	frameshift
			1	12:111856623_G_GCCGGGCC	frameshift
			1	12:111884838_G_A	splice donor
			1	12:111885295_G_A	missense
			1	12:111885497_G_A	missense
TM2D3	15q CN-LOH	20 / 61	8	70kb del	gene deletion
		(expected: 0.131 / 61)	4	15:102190214_G_GT	frameshift
			3	15:102182749_G_C	missense
			2	15:102182739_G_T	missense
			1	15:102187018_A_G	missense
			1	15:102192520_A_T	stop gained
			1	15:102192558_C_A	stop gained
DNMT3A	2p CN-LOH	4 / 16	1	2:25463194_T_C	missense
		(expected: 0.206 / 16)	1	2:25463218_C_T	missense
			1	2:25463248_G_A	missense
			1	2:25463536_C_T	missense
TET2	4q CN-LOH	6 / 21	1	4:106157029_C_T	stop gained
		(expected: 0.197 / 21)	1	4:106157446_G_T, 4:106157983_C_T, 4:106193937_AG_A	stop gained, frameshift
			1	4:106164061_C_T	stop gained
			1	4:106164085_G_GT	splice donor
			1	4:106180896_G_GT	frameshift
			1	4:106197285_T_C	missense
JAK2	9p CN-LOH	15 / 33	15	9:5073770_G_T	missense
		(expected: 0.217 / 33)			

See next page for caption.

Caption for Supplementary Table 11.

This table lists variants found in exome-sequenced mosaic CN-LOH individuals at each locus at which inherited or somatic variants are targets of clonal CN-LOH events. Variants identified by our association analyses (Extended Data Table 1 and Supplementary Table 7) and burden analyses (Supplementary Table 10) are both included. Allelic read depth analyses indicated that all or most of the variants found in the seven inherited risk loci arose from inherited variants, while all or most of the variants found in *DNMT3A*, *TET2*, and *JAK2* arose from somatic point mutations (Extended Data Fig. 8). We note that while this table indicates that fifteen individuals with 9p CN-LOH events were carriers of *JAK2* V617F (9:5073770_G_T), an additional five 9p CN-LOH individuals had at least one read supporting *JAK2* V617F (but did not have *JAK2* V617F genotype calls). One individual with 4q CN-LOH appeared to have three distinct somatic mutations in *TET2*.

Supplementary Table 12. Associations of mosaic CN-LOH mutations with inherited common variants in *cis*.

						GWAS		Allelic shift in hets			
Arm	Locus	Position[a]	Variant	Alleles[b]	AF[c]	*P*	OR (95% CI)	N_{REF}[d]	N_{ALT}	*P*	$P_{combined}$
Novel common variant associations with CN-LOH in *cis*											
14q	*TCL1A*	96180695	rs2887399	G/T	0.20	0.0024	0.84 (0.75–0.94)	195	102	7.4×10^{-8}	4.2×10^{-9}
14q	*DLK1*	101172227	rs7141110	G/C	0.22	1.4×10^{-5}	1.24 (1.13–1.37)	252	162	1.1×10^{-5}	3.6×10^{-9}
Previously reported common variant associations with CN-LOH in *cis*											
9p	*JAK2*	5037393	rs75032480[e]	A/C	0.26	2.6×10^{-29}	2.29 (1.99–2.63)	31	170	2×10^{-24}	6.3×10^{-51}

P-values from two independent statistical tests are reported: (i) a two-sided Fisher's exact test treating individuals with a mosaic CN-LOH mutation in *cis* as cases ($N\geq$379,201 individuals; Supplementary Table 6); and (ii) a binomial test for biased allelic imbalance in heterozygous cases. Loci reaching genome-wide significance in the combination of the tests (Fisher's combined *P*) are reported. For full details of statistical tests, see Methods.

[a]Base pair position in hg19 coordinates.

[b]Reference/alternate allele.

[c]Alternate allele frequency (in UK Biobank European-ancestry individuals).

[d]Number of mosaic individuals heterozygous for the variant in which the somatic event shifted the allelic balance in favor of the reference allele (by duplication of its chromosomal segment and loss of the homologous segment).

[e]rs75032480 belongs to the *JAK2* 46/1 haplotype [50–52].

Supplementary Table 13. Common variants associated with detectable mosaic chromosomal alterations on any autosome.

Locus	Variant	Chr	Position	REF/ALT	AAF	OR (95% CI)	*P*
SP140	rs13023767	2	231122057	T/G	0.25	1.07 (1.05–1.10)	2.3×10^{-8}
	rs776205558	2	231122089	CAGTA/C	0.25	1.07 (1.05–1.10)	3.2×10^{-8}
	rs55657711	2	231122210	A/G	0.30	1.07 (1.05–1.10)	2.0×10^{-8}
	rs62191185	2	231122290	G/A	0.42	1.06 (1.04–1.09)	4.0×10^{-8}
	rs1356532206	2	231124230	TA/T	0.26	1.07 (1.05–1.10)	3.3×10^{-8}
	rs6755306	2	231126528	G/A	0.25	1.08 (1.05–1.10)	1.2×10^{-8}
	rs1582833	2	231129729	C/G	0.31	1.07 (1.05–1.10)	1.7×10^{-8}
	rs62191195	2	231129794	C/T	0.25	1.08 (1.05–1.10)	9.4×10^{-9}
	rs34790921	2	231130508	G/T	0.25	1.08 (1.05–1.10)	1.2×10^{-8}
	rs890581	2	231131387	G/A	0.25	1.08 (1.05–1.10)	9.7×10^{-9}
	rs767031837	2	231134078	AGCGTG/A	0.25	1.08 (1.05–1.10)	1.2×10^{-8}
	rs62191198	2	231141196	C/G	0.25	1.08 (1.05–1.10)	9.5×10^{-9}
	rs12694846	2	231148128	A/G	0.27	1.07 (1.05–1.10)	2.1×10^{-8}
	rs34004493	2	231154012	A/G	0.27	1.07 (1.05–1.10)	2.6×10^{-8}
	rs6710297	2	231157512	A/G	0.27	1.07 (1.05–1.10)	2.2×10^{-8}
	rs35256947	2	231161026	T/C	0.27	1.07 (1.05–1.10)	2.0×10^{-8}
	rs13007094	2	231171194	C/T	0.25	1.08 (1.05–1.10)	1.1×10^{-8}
	rs2396742	2	231171423	C/T	0.25	1.08 (1.05–1.10)	1.9×10^{-8}
TERC	rs12638862	3	169477506	A/G	0.26	0.93 (0.91–0.96)	2.9×10^{-8}
	rs9811216	3	169487501	T/C	0.26	0.93 (0.91–0.96)	3.4×10^{-8}
TERT	rs33961405	5	1277577	G/A	0.52	0.93 (0.91–0.96)	6.4×10^{-9}
	rs10054203	5	1279964	G/C	0.40	1.07 (1.05–1.10)	3.0×10^{-9}
	rs7734992	5	1280128	T/C	0.42	1.09 (1.06–1.11)	4.2×10^{-13}
	rs4975538	5	1280830	G/C	0.36	1.08 (1.05–1.10)	1.2×10^{-10}
	rs6897196	5	1280938	A/G	0.39	1.08 (1.06–1.10)	5.6×10^{-11}
	rs749685059	5	1280940	GAGCCCACC/G	0.38	1.08 (1.06–1.11)	8.0×10^{-12}
	rs7726159	5	1282319	C/A	0.33	1.10 (1.07–1.12)	2.8×10^{-14}
	rs7725218	5	1282414	G/A	0.34	1.09 (1.06–1.11)	2.0×10^{-12}
	rs4449583	5	1284135	C/T	0.33	1.10 (1.07–1.12)	2.3×10^{-14}
	rs7705526	5	1285974	C/A	0.33	1.11 (1.08–1.14)	6.9×10^{-18}
	rs2736100	5	1286516	C/A	0.50	0.92 (0.90–0.94)	1.2×10^{-12}
	rs2853677	5	1287194	G/A	0.58	0.94 (0.92–0.96)	9.5×10^{-9}

Results from BOLT-LMM [26, 47] analysis of the "any autosomal mCA" phenotype in N=452,469 individuals are reported for all common variants (MAF>0.05) passing a significance threshold of $P<5\times10^{-8}$. AAF = ALT allele frequency; the ALT allele is the effect allele for reported odds ratios.

Supplementary Table 14. Associations of telomere length SNPs with mosaic chromosomal alterations on any autosome.

Locus	SNP	Chr	Position	Alleles	EAF	β_{telo} (s.e.)	P_{telo}	β_{mCA} (s.e.)	P_{mCA}
TERC	rs10936599	3	169492101	T/C	0.252	–0.097 (0.008)	2.5×10^{-31}	–0.0024 (0.0005)	1.1×10^{-7}
TERT	rs2736100	5	1286516	A/C	0.514	–0.078 (0.009)	4.4×10^{-19}	–0.0028 (0.0004)	1.2×10^{-12}
NAF1	rs7675998	4	164007820	A/G	0.217	–0.074 (0.009)	4.4×10^{-16}	–0.0006 (0.0005)	2.0×10^{-1}
OBFC1	rs9420907	10	105676465	A/C	0.865	–0.069 (0.010)	6.9×10^{-11}	–0.0019 (0.0006)	1.1×10^{-3}
ZNF208	rs8105767	19	22215441	A/G	0.709	–0.048 (0.008)	1.1×10^{-9}	–0.0005 (0.0004)	2.8×10^{-1}
RTEL1	rs755017	20	62421622	A/G	0.869	–0.062 (0.011)	6.7×10^{-9}	–0.0014 (0.0006)	1.5×10^{-2}
ACYP2	rs11125529	2	54475866	C/A	0.858	–0.056 (0.010)	4.5×10^{-8}	0.0002 (0.0006)	7.9×10^{-1}

Results from BOLT-LMM [26,47] analysis of the "any autosomal mCA" phenotype in N=452,469 individuals are reported for variants previously associated with telomere length [64]. Alleles, effect allele / other allele. EAF, effect allele frequency as reported by ref. [64]. β_{telo} (s.e.) and P_{telo}, effect size and association P-value for telomere length reported by ref. [64]. β_{mCA} (s.e.) and P_{mCA}, effect size and association P-value for presence of an mCA on any autosome.

Supplementary Table 15. Mean changes in polygenic scores for blood count and Y loss traits produced by CN-LOH mutations.

Arm	*N*	Platelet #	Red cell #	Basophil #	Neutrophil #	Eosinophil #	Monocyte #	Lymphocyte #	Y loss risk
1p	927	**–0.0632 (0.0055)**	0.0032 (0.0022)	–0.0003 (0.0006)	0.0005 (0.0022)	0.0013 (0.0021)	0.0022 (0.0022)	–0.0011 (0.0020)	0.0003 (0.0004)
1q	694	0.0056 (0.0040)	–0.0029 (0.0030)	0.0003 (0.0013)	0.0034 (0.0027)	–0.0045 (0.0023)	–0.0037 (0.0031)	0.0002 (0.0022)	0.0011 (0.0006)
2p	169	0.0026 (0.0053)	0.0035 (0.0056)	0.0020 (0.0011)	0.0061 (0.0039)	0.0071 (0.0038)	–0.0023 (0.0035)	0.0035 (0.0042)	–0.0006 (0.0010)
2q	205	**0.0192 (0.0059)**	0.0055 (0.0046)	0.0014 (0.0015)	**0.0153 (0.0063)**	0.0048 (0.0048)	0.0094 (0.0068)	0.0043 (0.0043)	**–0.0014 (0.0007)**
3p	164	0.0007 (0.0057)	0.0053 (0.0042)	0.0013 (0.0017)	**0.0082 (0.0030)**	–0.0066 (0.0043)	0.0045 (0.0059)	0.0016 (0.0042)	0.0011 (0.0009)
3q	195	0.0084 (0.0062)	–0.0080 (0.0047)	0.0016 (0.0014)	**0.0087 (0.0036)**	0.0051 (0.0045)	**0.0093 (0.0047)**	0.0015 (0.0035)	0.0015 (0.0013)
4p	61	0.0018 (0.0060)	0.0037 (0.0034)	0.0000 (0.0009)	0.0054 (0.0043)	0.0001 (0.0043)	**0.0107 (0.0040)**	0.0028 (0.0043)	0.0010 (0.0009)
4q	237	0.0004 (0.0038)	–0.0012 (0.0040)	0.0000 (0.0010)	0.0008 (0.0037)	0.0014 (0.0030)	0.0006 (0.0034)	–0.0011 (0.0035)	0.0005 (0.0006)
5p	71	0.0004 (0.0049)	0.0003 (0.0044)	–0.0002 (0.0009)	0.0061 (0.0041)	–0.0069 (0.0038)	**0.0099 (0.0025)**	–0.0022 (0.0035)	–0.0001 (0.0009)
5q	162	0.0094 (0.0052)	–0.0001 (0.0037)	0.0018 (0.0011)	0.0000 (0.0042)	0.0011 (0.0064)	**0.0104 (0.0047)**	0.0006 (0.0037)	0.0004 (0.0009)
6p	327	**–0.0090 (0.0043)**	–0.0050 (0.0036)	0.0003 (0.0008)	–0.0006 (0.0027)	0.0045 (0.0036)	0.0035 (0.0035)	0.0004 (0.0034)	0.0008 (0.0005)
6q	116	–0.0037 (0.0086)	–0.0011 (0.0106)	0.0005 (0.0013)	0.0004 (0.0035)	0.0036 (0.0050)	0.0028 (0.0053)	0.0031 (0.0048)	0.0021 (0.0018)
7p	90	0.0039 (0.0044)	0.0024 (0.0047)	–0.0012 (0.0018)	–0.0047 (0.0050)	0.0009 (0.0052)	0.0046 (0.0052)	0.0002 (0.0047)	0.0007 (0.0015)
7q	167	0.0007 (0.0062)	–0.0030 (0.0050)	–0.0018 (0.0017)	0.0024 (0.0031)	0.0039 (0.0043)	0.0037 (0.0035)	0.0006 (0.0034)	**0.0030 (0.0013)**
8p	80	0.0058 (0.0035)	0.0013 (0.0038)	–0.0032 (0.0031)	–0.0037 (0.0042)	–0.0030 (0.0037)	0.0018 (0.0052)	–0.0029 (0.0042)	0.0010 (0.0008)
8q	134	–0.0025 (0.0057)	–0.0002 (0.0038)	0.0002 (0.0012)	0.0001 (0.0041)	0.0010 (0.0045)	–0.0011 (0.0068)	–0.0005 (0.0057)	0.0004 (0.0008)
9p	386	–0.0081 (0.0054)	0.0043 (0.0023)	0.0004 (0.0004)	**0.0053 (0.0019)**	0.0067 (0.0034)	–0.0028 (0.0019)	–0.0035 (0.0019)	0.0001 (0.0003)
9q	472	0.0050 (0.0039)	–0.0029 (0.0033)	–0.0002 (0.0006)	0.0004 (0.0024)	0.0024 (0.0026)	–0.0085 (0.0044)	0.0003 (0.0021)	–0.0002 (0.0006)
10p	50	0.0059 (0.0059)	–0.0005 (0.0062)	–0.0003 (0.0037)	0.0053 (0.0041)	0.0050 (0.0075)	0.0011 (0.0054)	–0.0031 (0.0053)	–0.0009 (0.0009)
10q	85	–0.0029 (0.0077)	–0.0024 (0.0054)	–0.0005 (0.0021)	0.0086 (0.0055)	0.0105 (0.0053)	0.0017 (0.0065)	0.0020 (0.0050)	–0.0001 (0.0008)
11p	564	0.0012 (0.0031)	0.0014 (0.0022)	0.0007 (0.0004)	0.0006 (0.0019)	0.0032 (0.0017)	–0.0015 (0.0018)	0.0002 (0.0014)	–0.0001 (0.0004)
11q	647	0.0049 (0.0032)	**0.0084 (0.0019)**	–0.0004 (0.0011)	0.0039 (0.0021)	0.0024 (0.0023)	0.0007 (0.0023)	0.0014 (0.0020)	**0.0032 (0.0005)**
12p	88	0.0012 (0.0079)	**–0.0122 (0.0055)**	–0.0000 (0.0016)	0.0016 (0.0036)	–0.0012 (0.0060)	–0.0072 (0.0052)	0.0154 (0.0081)	–0.0001 (0.0022)
12q	302	**0.0358 (0.0090)**	**0.0105 (0.0041)**	0.0015 (0.0011)	**0.0112 (0.0034)**	**0.0180 (0.0042)**	**0.0115 (0.0035)**	**0.0154 (0.0041)**	0.0003 (0.0007)
13q	623	–0.0002 (0.0028)	–0.0024 (0.0020)	0.0000 (0.0005)	–0.0005 (0.0017)	–0.0006 (0.0019)	–0.0068 (0.0036)	–0.0002 (0.0023)	0.0004 (0.0005)
14q	956	**0.0097 (0.0029)**	0.0023 (0.0017)	–0.0008 (0.0005)	**0.0036 (0.0015)**	**0.0038 (0.0018)**	0.0024 (0.0019)	0.0003 (0.0015)	**0.0039 (0.0007)**
15q	638	0.0042 (0.0025)	–0.0016 (0.0024)	0.0003 (0.0009)	–0.0028 (0.0018)	0.0016 (0.0021)	–0.0001 (0.0025)	0.0031 (0.0020)	–0.0005 (0.0003)
16p	318	–0.0019 (0.0025)	0.0040 (0.0029)	**0.0011 (0.0004)**	0.0010 (0.0016)	0.0018 (0.0029)	0.0008 (0.0020)	0.0008 (0.0020)	0.0002 (0.0003)
16q	280	–0.0015 (0.0034)	**0.0071 (0.0034)**	0.0002 (0.0007)	0.0015 (0.0027)	0.0003 (0.0028)	–0.0034 (0.0068)	0.0008 (0.0026)	–0.0004 (0.0007)
17p	224	0.0037 (0.0055)	–0.0030 (0.0033)	–0.0015 (0.0015)	–0.0039 (0.0049)	0.0023 (0.0043)	–0.0038 (0.0046)	–0.0036 (0.0045)	–0.0001 (0.0008)
17q	521	0.0012 (0.0034)	–0.0015 (0.0028)	0.0007 (0.0011)	0.0005 (0.0034)	–0.0019 (0.0026)	–0.0007 (0.0035)	0.0022 (0.0026)	0.0001 (0.0005)
18p	36	–0.0011 (0.0046)	–0.0036 (0.0046)	–0.0006 (0.0010)	–0.0029 (0.0039)	0.0034 (0.0039)	0.0020 (0.0037)	–0.0028 (0.0025)	–0.0006 (0.0011)
18q	127	0.0039 (0.0042)	0.0049 (0.0037)	**0.0044 (0.0014)**	**0.0074 (0.0035)**	**0.0136 (0.0040)**	0.0058 (0.0043)	–0.0003 (0.0031)	**0.0097 (0.0027)**
19p	189	0.0001 (0.0037)	0.0027 (0.0044)	0.0026 (0.0013)	0.0019 (0.0034)	–0.0027 (0.0035)	–0.0029 (0.0047)	–0.0033 (0.0050)	0.0006 (0.0006)
19q	267	0.0021 (0.0035)	0.0014 (0.0030)	0.0020 (0.0012)	0.0023 (0.0029)	–0.0011 (0.0029)	–0.0010 (0.0043)	–0.0005 (0.0024)	0.0002 (0.0005)
20p	62	–0.0018 (0.0049)	–0.0011 (0.0031)	0.0002 (0.0013)	–0.0027 (0.0032)	–0.0035 (0.0033)	–0.0035 (0.0048)	–0.0039 (0.0049)	0.0003 (0.0007)
20q	210	0.0070 (0.0039)	0.0023 (0.0036)	0.0005 (0.0009)	0.0013 (0.0022)	0.0009 (0.0021)	0.0052 (0.0043)	0.0001 (0.0029)	0.0005 (0.0006)
21q	244	**0.0058 (0.0027)**	0.0011 (0.0028)	–0.0004 (0.0007)	**0.0035 (0.0017)**	0.0066 (0.0034)	0.0020 (0.0020)	0.0021 (0.0018)	0.0004 (0.0004)
22q	550	0.0052 (0.0030)	0.0044 (0.0026)	0.0009 (0.0006)	0.0013 (0.0013)	–0.0007 (0.0021)	0.0005 (0.0019)	0.0011 (0.0020)	**0.0008 (0.0004)**
Any	11,638	–0.0014 (0.0008)	**0.0012 (0.0006)**	0.0003 (0.0002)	**0.0022 (0.0005)**	**0.0020 (0.0005)**	0.0004 (0.0006)	0.0009 (0.0005)	**0.0009 (0.0001)**

This table provides numerical data plotted in Fig. 2b. Units for polygenic scores are standard deviations for blood count traits; for Y loss, polygenic scores were computed on a 0/1 binary trait (modeled additively). Mean changes in polygenic scores reaching nominal significance ($P<0.05$ before multiple hypothesis correction) are indicated in bold; those that reached significance at FDR 0.05 or after Bonferroni correction are indicated in Fig. 2b.

Supplementary Table 16. Mean changes in polygenic scores for six non-blood-related control traits produced by CN-LOH mutations.

Arm	*N*	Height	BMI	Bone mineral density	FEV1/FVC	Blood pressure (systolic)	Blood pressure (diastolic)
1p	927	–0.0027 (0.0031)	–0.0033 (0.0019)	–0.0012 (0.0028)	*–0.0055 (0.0021)*	–0.0022 (0.0021)	–0.0009 (0.0019)
1q	694	0.0015 (0.0039)	–0.0001 (0.0022)	0.0017 (0.0030)	–0.0006 (0.0025)	*0.0041 (0.0019)*	0.0038 (0.0021)
2p	169	0.0047 (0.0072)	–0.0065 (0.0048)	0.0035 (0.0057)	–0.0070 (0.0046)	–0.0005 (0.0038)	–0.0017 (0.0040)
2q	205	0.0042 (0.0075)	0.0017 (0.0043)	–0.0065 (0.0060)	–0.0054 (0.0051)	*0.0092 (0.0039)*	*0.0082 (0.0039)*
3p	164	–0.0063 (0.0056)	0.0023 (0.0039)	0.0057 (0.0045)	–0.0010 (0.0043)	*0.0081 (0.0037)*	0.0047 (0.0034)
3q	195	0.0082 (0.0066)	*0.0083 (0.0042)*	–0.0026 (0.0041)	–0.0061 (0.0039)	–0.0036 (0.0034)	–0.0017 (0.0036)
4p	61	0.0105 (0.0090)	0.0049 (0.0052)	0.0010 (0.0069)	–0.0036 (0.0039)	–0.0041 (0.0044)	–0.0011 (0.0041)
4q	237	0.0049 (0.0065)	0.0007 (0.0034)	–0.0035 (0.0044)	–0.0007 (0.0053)	0.0027 (0.0039)	0.0019 (0.0038)
5p	71	–0.0036 (0.0061)	0.0021 (0.0049)	–0.0010 (0.0047)	–0.0044 (0.0046)	0.0003 (0.0045)	0.0034 (0.0038)
5q	162	0.0107 (0.0081)	0.0004 (0.0044)	–0.0028 (0.0048)	–0.0074 (0.0059)	–0.0057 (0.0036)	–0.0023 (0.0038)
6p	327	–0.0026 (0.0046)	–0.0016 (0.0025)	–0.0035 (0.0030)	*0.0071 (0.0034)*	0.0027 (0.0021)	0.0011 (0.0022)
6q	116	0.0087 (0.0095)	0.0005 (0.0044)	0.0077 (0.0104)	0.0001 (0.0059)	0.0075 (0.0040)	*0.0111 (0.0041)*
7p	90	–0.0065 (0.0089)	*0.0109 (0.0043)*	–0.0013 (0.0075)	–0.0038 (0.0043)	–0.0038 (0.0049)	–0.0028 (0.0042)
7q	167	0.0015 (0.0057)	–0.0068 (0.0037)	–0.0008 (0.0088)	–0.0016 (0.0035)	–0.0015 (0.0036)	0.0016 (0.0036)
8p	80	0.0017 (0.0066)	–0.0022 (0.0039)	–0.0017 (0.0047)	0.0009 (0.0038)	0.0039 (0.0046)	0.0032 (0.0045)
8q	134	0.0096 (0.0079)	–0.0045 (0.0040)	0.0009 (0.0049)	*–0.0085 (0.0037)*	0.0070 (0.0044)	0.0071 (0.0040)
9p	386	0.0039 (0.0025)	–0.0010 (0.0020)	–0.0014 (0.0021)	–0.0012 (0.0019)	–0.0017 (0.0016)	–0.0014 (0.0014)
9q	472	0.0035 (0.0050)	0.0041 (0.0024)	–0.0051 (0.0028)	0.0004 (0.0025)	0.0001 (0.0019)	–0.0003 (0.0021)
10p	50	0.0014 (0.0073)	–0.0007 (0.0041)	0.0149 (0.0080)	0.0019 (0.0066)	0.0084 (0.0048)	0.0076 (0.0053)
10q	85	–0.0105 (0.0073)	0.0028 (0.0056)	0.0117 (0.0078)	–0.0031 (0.0054)	–0.0045 (0.0058)	–0.0051 (0.0060)
11p	564	–0.0002 (0.0028)	–0.0016 (0.0019)	–0.0037 (0.0024)	*–0.0035 (0.0016)*	–0.0009 (0.0021)	–0.0014 (0.0021)
11q	647	0.0020 (0.0027)	–0.0017 (0.0021)	–0.0017 (0.0033)	0.0016 (0.0020)	–0.0002 (0.0021)	*0.0050 (0.0020)*
12p	88	0.0033 (0.0083)	–0.0019 (0.0046)	0.0003 (0.0061)	0.0059 (0.0052)	–0.0014 (0.0041)	0.0027 (0.0044)
12q	302	0.0022 (0.0059)	0.0009 (0.0031)	–0.0001 (0.0037)	0.0017 (0.0029)	0.0026 (0.0031)	*0.0067 (0.0032)*
13q	623	–0.0016 (0.0031)	0.0016 (0.0022)	–0.0048 (0.0028)	–0.0011 (0.0021)	–0.0007 (0.0017)	–0.0024 (0.0018)
14q	956	0.0005 (0.0028)	–0.0006 (0.0017)	0.0019 (0.0020)	–0.0023 (0.0017)	0.0008 (0.0015)	0.0004 (0.0014)
15q	638	–0.0079 (0.0047)	0.0021 (0.0020)	–0.0036 (0.0024)	–0.0025 (0.0029)	0.0006 (0.0020)	0.0002 (0.0023)
16p	318	0.0013 (0.0035)	*–0.0064 (0.0025)*	0.0005 (0.0026)	–0.0035 (0.0019)	–0.0034 (0.0020)	–0.0026 (0.0020)
16q	280	0.0015 (0.0043)	0.0028 (0.0033)	*0.0123 (0.0035)*	*0.0086 (0.0029)*	0.0014 (0.0022)	*0.0053 (0.0022)*
17p	224	0.0034 (0.0059)	0.0024 (0.0030)	0.0004 (0.0049)	0.0009 (0.0034)	–0.0050 (0.0031)	0.0008 (0.0028)
17q	521	0.0053 (0.0047)	–0.0027 (0.0021)	0.0000 (0.0036)	–0.0020 (0.0026)	–0.0041 (0.0023)	–0.0028 (0.0020)
18p	36	–0.0076 (0.0058)	0.0087 (0.0048)	0.0025 (0.0077)	–0.0045 (0.0052)	0.0003 (0.0051)	–0.0032 (0.0039)
18q	127	–0.0042 (0.0049)	–0.0034 (0.0046)	–0.0019 (0.0036)	–0.0014 (0.0032)	–0.0012 (0.0034)	–0.0026 (0.0034)
19p	189	0.0005 (0.0048)	–0.0026 (0.0025)	–0.0044 (0.0038)	–0.0016 (0.0023)	–0.0042 (0.0030)	–0.0021 (0.0028)
19q	267	*0.0091 (0.0034)*	*–0.0046 (0.0022)*	0.0012 (0.0027)	0.0003 (0.0024)	0.0007 (0.0017)	–0.0000 (0.0019)
20p	62	0.0019 (0.0069)	*0.0085 (0.0035)*	–0.0023 (0.0073)	–0.0053 (0.0038)	0.0016 (0.0040)	0.0018 (0.0046)
20q	210	0.0030 (0.0043)	–0.0009 (0.0026)	0.0014 (0.0028)	–0.0027 (0.0026)	–0.0031 (0.0028)	–0.0034 (0.0031)
21q	244	–0.0035 (0.0032)	0.0016 (0.0020)	–0.0041 (0.0031)	*0.0038 (0.0019)*	0.0020 (0.0019)	0.0039 (0.0021)
22q	550	0.0004 (0.0023)	–0.0025 (0.0014)	–0.0005 (0.0024)	0.0007 (0.0017)	0.0004 (0.0012)	–0.0005 (0.0011)
Any	11638	0.0009 (0.0008)	–0.0005 (0.0005)	–0.0007 (0.0006)	*–0.0013 (0.0005)*	0.0001 (0.0004)	0.0007 (0.0004)

This table is the analog of Supplementary Table 15 for polygenic scores computed for six highly heritable, polygenic non-blood-cell traits [47]. These traits serve as controls, as variants influencing these traits are not typically expected to affect cell proliferation. Units are standard deviations. Mean changes in polygenic scores reaching nominal significance ($P<0.05$ before multiple hypothesis correction) are indicated in italics; no changes were significant after Bonferroni correction.

Supplementary Table 17. Action of CN-LOH events on risk alleles for myeloproliferative neoplasms.

						GWAS		Allelic shift in hets		
Locus	Gene	Position[a]	Variant	Alleles[b]	RAF[c]	P_{MPN}[d]	$P_{CN\text{-}LOH}$[e]	N_{risk}[f]	$N_{nonrisk}$	P
2p21	*PRKCE*	45956545	rs12616536	A/G	0.0042	5.5×10^{-7}	0.59	0	0	1
2q34	*CPS1*	211478366	rs13415932	A/C	0.9604	6.1×10^{-7}	0.56	9	5	0.42
3q13.33	*STXBP5L*	120972200	rs75405916	C/T	0.0004	7.4×10^{-7}	1	1	0	1
3q21.3	*GATA2*	128316939	rs9864772	G/A	0.6050	2×10^{-7}	0.36	18	19	1
3q25.33	*SCHIP1*	159633461	rs77249081	G/C	0.0075	1.6×10^{-7}	1	0	0	1
3q25.33	*KPNA4*	160284736	rs74676712	T/C	0.1062	3.3×10^{-9}	1	20	11	0.15
3q26.2	*MECOM*	168846701	rs12491785	C/T	0.3938	4.5×10^{-9}	0.96	46	35	0.27
4q24	*TET2*	105749895	rs1548483	T/C	0.0391	4×10^{-21}	1	14	2	0.0042
5p15.33	*SLC12A7*	1100831	rs60833263	G/A	0.4538	1.1×10^{-6}	0.3	18	9	0.12
5p15.33	*SLC12A7*	1138335	rs4131149	G/T	0.4279	3×10^{-7}	0.13	20	11	0.15
5p15.33	*TERT*	1285974	rs7705526	A/C	0.3376	1.9×10^{-48}	0.3	10	10	1
5p15.33	*TERT*	1287194	rs2853677	G/A	0.4227	8.4×10^{-39}	0.58	15	11	0.56
5q22.1	*NREP*	111061883	rs56084922	A/G	0.9271	9.6×10^{-7}	0.78	4	7	0.55
6p21.32	*TAP2*	32668411	rs9275373	A/G	0.1110	3.8×10^{-7}	0.52	14	14	1
6p21.31	*NUDT3*	34235378	rs116466979	C/T	0.0453	3.3×10^{-9}	0.63	5	4	1
7p22.3	*MAD1L1*	2112506	rs1860826	A/G	0.3545	3.9×10^{-7}	0.85	24	19	0.54
7q32.3	*MKLN1*	130746955	rs62471615	C/A	0.2953	6.1×10^{-16}	0.3	34	29	0.61
9p24.1	*JAK2*	4998401	rs7868130	T/C	0.2695	9×10^{-115}	4.5×10^{-26}	167	33	9.3×10^{-23}
9p24.1	*INSL6*	5149250	rs75035022	T/C	0.0311	2.7×10^{-24}	0.017	20	6	0.0094
9q34.13	*GFI1B*	135870130	rs621940	G/C	0.1572	4.1×10^{-9}	0.29	61	68	0.6
11q22.3	*ATM*	108143456	rs1800057	G/C	0.0262	2.6×10^{-9}	0.29	25	2	5.6×10^{-6}
12q24.12	*SH2B3*	111865049	rs7310615	C/G	0.4850	2.1×10^{-20}	0.079	85	35	5.7×10^{-6}
13q14.11	*FOXO1*	41204015	rs7323267	C/T	0.2031	1.1×10^{-7}	0.25	78	74	0.81
13q31.1	*SPRY2*	82102166	rs9545761	T/C	0.3977	7.3×10^{-7}	0.19	132	121	0.53
14q12	*FOXG1*	28341915	rs144202762	A/T	0.0060	5.6×10^{-7}	0.92	0	0	1
20q13.33	*PRPF6*	62651978	rs816925	G/A	0.5644	7.8×10^{-7}	0.44	63	44	0.081
21q22.12	*RUNX1*	36347627	rs55857134	C/T	0.3339	1×10^{-8}	0.18	42	37	0.65
22q12.1	*CHEK2*	29121087	rs17879961	G/A	0.0174	9.1×10^{-7}	1	0	0	1

Risk alleles for myeloproliferative neoplasms (identified by Bao et al. [66]) tended to be made homozygous by CN-LOH events in UK Biobank. The first seven columns of this table are from Supplementary Table 2 of ref. [66] (which provided data for independent variants reaching $P<1\times10^{-6}$); the last four columns show, for each MPN risk allele, its P-value for association with CN-LOH events in *cis*, and the directionality of CN-LOH events in heterozygous carriers of the risk allele. For details of statistical tests and sample sizes, see Supplementary Table 7.

[a]Base pair position in hg19 coordinates.

[b]Risk/nonrisk allele for myeloproliferative neoplasms [66].

[c]Risk allele frequency.

[d]P-value for association with myeloproliferative neoplasms [66].

[e]P-value for association with likely-CN-LOH events in *cis* in UK Biobank.

[f]Number of mosaic individuals heterozygous for the variant in which the somatic event shifted the allelic balance in favor of the risk allele (by duplication of its chromosomal segment and loss of the homologous segment).

Supplementary Table 18. Accuracy of predicting CN-LOH directionality using genetic risk.

		CN-LOH-associated alleles only			Polygenic scores (for blood traits)			Both		
Arm	*N*	Pred. acc. (s.e.)	Pred. *R* (95% CI)	*P*	Pred. acc. (s.e.)	Pred. *R* (95% CI)	*P*	Pred. acc. (s.e.)	Pred. *R* (95% CI)	*P*
1p	927	0.639 (0.007)	0.523 (0.474,0.568)	2×10^{-66}	0.590 (0.016)	0.338 (0.279,0.394)	1.8×10^{-26}	0.636 (0.010)	0.518 (0.470,0.564)	3.7×10^{-65}
1q	694	0.505 (0.002)	0.080 (0.006,0.154)	0.017	0.535 (0.018)	0.023 (-0.051,0.097)	0.27	0.550 (0.017)	0.088 (0.013,0.161)	0.01
2p	169	–	–	–	0.556 (0.034)	0.123 (-0.028,0.269)	0.055	0.556 (0.034)	0.123 (-0.028,0.269)	0.055
2q	205	–	–	–	0.605 (0.034)	0.227 (0.093,0.353)	0.00053	0.605 (0.034)	0.227 (0.093,0.353)	0.00053
3p	164	–	–	–	0.497 (0.033)	0.002 (-0.151,0.155)	0.49	0.497 (0.033)	0.002 (-0.151,0.155)	0.49
5p	71	–	–	–	0.704 (0.055)	0.417 (0.203,0.592)	0.00015	0.704 (0.055)	0.417 (0.203,0.592)	0.00015
5q	162	–	–	–	0.574 (0.039)	0.178 (0.025,0.324)	0.012	0.574 (0.039)	0.178 (0.025,0.324)	0.012
8q	134	0.526 (0.010)	0.229 (0.061,0.383)	0.0039	–	–	–	0.526 (0.010)	0.229 (0.061,0.383)	0.0039
9p	386	0.680 (0.016)	0.501 (0.422,0.572)	3.6×10^{-26}	0.674 (0.016)	0.472 (0.391,0.546)	3.8×10^{-23}	0.680 (0.016)	0.501 (0.422,0.572)	3.6×10^{-26}
11q	647	0.543 (0.006)	0.294 (0.222,0.363)	1.1×10^{-14}	0.577 (0.019)	0.200 (0.125,0.273)	1.4×10^{-7}	0.614 (0.019)	0.352 (0.283,0.418)	1.3×10^{-20}
12q	302	0.523 (0.006)	0.216 (0.105,0.321)	7.9×10^{-5}	0.603 (0.028)	0.269 (0.161,0.371)	1×10^{-6}	0.608 (0.018)	0.338 (0.235,0.435)	7.9×10^{-10}
14q	956	0.569 (0.012)	0.194 (0.133,0.255)	6.8×10^{-10}	0.544 (0.016)	0.155 (0.092,0.216)	7.7×10^{-7}	0.559 (0.014)	0.171 (0.109,0.232)	4.8×10^{-8}
15q	638	0.612 (0.009)	0.460 (0.397,0.519)	4.3×10^{-35}	0.494 (0.011)	-0.045 (-0.123,0.032)	0.87	0.601 (0.010)	0.453 (0.389,0.513)	6.3×10^{-34}
18q	127	–	–	–	0.535 (0.044)	0.170 (-0.005,0.334)	0.028	0.535 (0.044)	0.170 (-0.005,0.334)	0.028

This table provides numerical data plotted in Fig. 2c. CN-LOH directions were predicted using: (i) only CN-LOH-associated alleles on affected chromosomal segments (for chromosome arms containing at least one association; Extended Data Table 1); (ii) polygenic score differentials on affected chromosomal segments; (iii) both CN-LOH-associated alleles and polygenic scores.

For each chromosome arm with at least one available predictor (Methods), prediction accuracy was computed as the fraction of predicted CN-LOH directions (hard-called) that matched observed CN-LOH directions. Prediction *R* was computed as the correlation between predicted CN-LOH directions (continuous-valued, as output by the linear predictor) and observed CN-LOH directions. Predictive performance was assessed using 10-fold cross-validation, and both accuracy and *R* metrics (and standard errors and 95% CIs) were computed over a merge of all held-out folds. *P*-values for Pearson correlation $R>0$ were computed using a one-sided *t*-test (on transformed correlations).

Supplementary Table 19. Logistic models used for predicting directionality of clonally expanded CN-LOH mutations.

Arm	Logistic regression model
1p	5.9 × (CN-LOH allele count)
1q	1.9 × (CN-LOH allele count) + 2.8 × (platelet crit PRS)
2p	4.8 × (eosinophil# PRS) – 11.9 × (monocyte% PRS)
2q	7.5 × (hemoglobin PRS) + 7.9 × (platelet crit PRS)
3p	5.8 × (neutrophil# PRS)
5p	47.6 × (monocyte# PRS)
5q	10.0 × (platelet crit PRS)
8q	102.6 × (CN-LOH allele count)
9p	1.8 × (CN-LOH allele count)
11q	104.7 × (CN-LOH allele count) + 1.7 × (MPN PRS) + 36.4 × (mLOY PRS) + 4.2 × (platelet distribution width PRS)
12q	123.7 × (CN-LOH allele count) + 3.7 × (MPN PRS)
14q	0.5 × (CN-LOH allele count)
15q	3.5 × (CN-LOH allele count)
18q	24.0 × (mLOY PRS)

We ran logistic regression independently on each chromosome arm (using stepwise forward selection for variable selection) to enable the logistic model to pick up PRS signals concentrated on specific arms that might wash out genome-wide. To guard against overfitting, we ran logistic regression within 10-fold cross-validation; above we report median coefficients for logistic models across the 10 cross-validation folds (for each of 14 arms for which stepwise forward selection found at least one predictor (on average across folds). All chromosome arms either contained at most one CN-LOH-associated locus or contained two loci with similar effects (large-effect *MRE11* and *ATM* alleles on 11q; small-effect *TCL1A* and *DLK1* alleles on 14q), so we aggregated the effects of all CN-LOH-associated alleles on an arm into a single "CN-LOH allele count" variable in the logistic regression models. For each locus, the effect allele in these models was the allele that tends to be made homozygous by CN-LOH events (which differs from the risk allele for *MPL* and *DLK1*).

Supplementary Table 20. Risk increase for incident cancers conferred by mCAs.

(a) Analyses restricted to individuals with normal blood counts at assessment

	CLL		MPN		MDS		Any blood cancer	
	N_{case}=107, $N_{control}$=361,850		N_{case}=67, $N_{control}$=358,820		N_{case}=56, $N_{control}$=358,807		N_{case}=1,055, $N_{control}$=346,965	
mCA	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)
1+	1	0 (0–280)	1	0 (0–568)	1	0 (0–606)	0.0073	16.8 (1.93–66.6)
1p–	1	0 (0–1.1e+03)	1	0 (0–1.62e+03)	1	0 (0–1.74e+03)	0.037	30 (0.67–224)
1q–	1	0 (0–621)	1	0 (0–1.28e+03)	1	0 (0–1.27e+03)	1	0 (0–70.3)
1p=	1	0 (0–25.7)	0.084	11.6 (0.29–67.6)	1	0 (0–48.6)	0.17	2.15 (0.44–6.35)
1q=	1	0 (0–32.3)	1	0 (0–53.1)	1	0 (0–59.9)	0.32	1.76 (0.21–6.44)
2p–	1	0 (0–164)	1	0 (0–236)	1	0 (0–261)	1	0 (0–16.1)
2q–	1	0 (0–478)	1	0 (0–740)	1	0 (0–736)	1	0 (0–46)
2p=	1	0 (0–158)	1	0 (0–280)	1	0 (0–304)	1	0 (0–16.6)
2q=	1	0 (0–123)	1	0 (0–217)	1	0 (0–246)	1	0 (0–12.6)
3+	0.00029	85.8 (9.96–337)	1	0 (0–249)	1	0 (0–245)	0.00013	17.1 (4.47–46.9)
3p–	1	0 (0–681)	1	0 (0–1.41e+03)	1	0 (0–1.4e+03)	1	0 (0–73.4)
3p=	1	0 (0–118)	1	0 (0–202)	1	0 (0–215)	1	0 (0–12.2)
3q=	1	0 (0–167)	1	0 (0–294)	1	0 (0–307)	0.21	4.32 (0.11–25.2)
4q–	1	0 (0–125)	1	0 (0–224)	0.018	58.5 (1.42–355)	0.032	7.42 (0.88–28.1)
4p=	1	0 (0–453)	1	0 (0–622)	1	0 (0–829)	0.088	11.3 (0.27–70.3)
4q=	1	0 (0–88.2)	1	0 (0–158)	0.00035	79.2 (9.14–312)	0.0011	9.4 (2.5–25)
5+	1	0 (0–205)	1	0 (0–336)	1	0 (0–358)	0.16	5.86 (0.14–34.7)
5q–	1	0 (0–134)	1	0 (0–265)	0.013	77.4 (1.87–480)	0.003	11 (2.21–33.6)
5q=	1	0 (0–155)	1	0 (0–261)	1	0 (0–293)	1	0 (0–16.1)
6+	1	0 (0–1.1e+03)	1	0 (0–1.41e+03)	1	0 (0–1.96e+03)	1	0 (0–111)
6q–	1	0 (0–367)	1	0 (0–570)	1	0 (0–589)	0.1	9.56 (0.23–58.7)
6p=	1	0 (0–74.8)	1	0 (0–118)	1	0 (0–134)	0.094	3.96 (0.47–14.7)
6q=	1	0 (0–227)	1	0 (0–377)	1	0 (0–400)	1	0 (0–23.5)
7p–	1	0 (0–546)	1	0 (0–1.23e+03)	1	0 (0–1.41e+03)	1	0 (0–57)
7q–	1	0 (0–190)	1	0 (0–298)	1	0 (0–298)	0.019	9.77 (1.15–37.5)
7p=	1	0 (0–237)	1	0 (0–369)	1	0 (0–373)	1	0 (0–23.3)
7q=	1	0 (0–170)	0.013	81.3 (1.99–497)	3.1×10^{-7}	267 (51–910)	9.8×10^{-5}	18.2 (4.76–49.9)
8+	0.023	44.4 (1.09–263)	1	0 (0–322)	0.011	93.2 (2.25–588)	2.3×10^{-6}	26.7 (8.24–67)
8p–	0.0072	147 (3.49–979)	1	0 (0–808)	1	0 (0–1.06e+03)	0.066	15.4 (0.37–98)
8p=	1	0 (0–566)	1	0 (0–1.16e+03)	1	0 (0–1.41e+03)	1	0 (0–54.4)
8q=	1	0 (0–175)	1	0 (0–283)	1	0 (0–311)	1	0 (0–17.9)
9+	0.019	54.3 (1.33–329)	1	0 (0–402)	1	0 (0–390)	0.15	6.3 (0.15–37.8)
9q–	1	0 (0–561)	1	0 (0–1.08e+03)	1	0 (0–1.17e+03)	1	0 (0–71)
9p=	1	0 (0–57.8)	3.6×10^{-13}	260 (89.4–631)	1	0 (0–158)	8.2×10^{-6}	13.8 (4.91–31.1)
9q=	1	0 (0–46.8)	1	0 (0–83.7)	1	0 (0–85.1)	1	0 (0–4.81)
10q–	1	0 (0–75.2)	1	0 (0–122)	1	0 (0–162)	1	0 (0–7.28)
10p=	1	0 (0–393)	1	0 (0–726)	1	0 (0–822)	1	0 (0–39)
10q=	1	0 (0–199)	1	0 (0–338)	1	0 (0–380)	1	0 (0–20.1)
11p–	1	0 (0–394)	1	0 (0–698)	1	0 (0–687)	1	0 (0–48)
11q–	0.045	22.1 (0.55–130)	1	0 (0–147)	1	0 (0–144)	9.2×10^{-5}	11.9 (3.75–28.8)
11p=	1	0 (0–32.4)	1	0 (0–53.3)	1	0 (0–60.9)	0.63	0 (0–3.32)
11q=	0.092	10.6 (0.26–60.9)	0.056	17.7 (0.44–104)	0.055	17.9 (0.44–106)	0.0004	6.6 (2.39–14.7)
12+	5.1×10^{-22}	149 (72.9–278)	0.059	16.8 (0.41–98.6)	0.061	16.2 (0.4–95.7)	2.9×10^{-20}	23.3 (13.9–37.3)
12p–	1	0 (0–788)	1	0 (0–1.71e+03)	1	0 (0–1.89e+03)	1	0 (0–89.4)
12q–	1	0 (0–601)	1	0 (0–821)	1	0 (0–913)	1	0 (0–56.5)
12p=	1	0 (0–455)	1	0 (0–793)	1	0 (0–938)	1	0 (0–47.5)
12q=	1	0 (0–80.2)	1	0 (0–141)	1	0 (0–161)	1	0 (0–8.3)
13q–	5.1×10^{-13}	127 (48.3–280)	1	0 (0–77.3)	1	0 (0–73.6)	1.8×10^{-7}	14.4 (6.08–29.5)
13q=	8.2×10^{-5}	38.4 (7.7–117)	1	0 (0–68.6)	1	0 (0–72.2)	0.047	3.81 (0.78–11.4)
14+	1	0 (0–73.1)	1	0 (0–135)	0.033	30.7 (0.75–184)	0.39	2.07 (0.05–11.9)
14q–	0.00017	115 (13.2–456)	1	0 (0–294)	1	0 (0–280)	0.00074	18.5 (3.63–59.2)
14q=	1	0 (0–23.5)	1	0 (0–40)	1	0 (0–42.9)	1	0.65 (0.02–3.64)
15+	1	0 (0–46.8)	1	0 (0–93)	1	0 (0–71.5)	0.55	1.27 (0.03–7.22)
15q–	1	0 (0–943)	1	0 (0–1.59e+03)	1	0 (0–1.71e+03)	1	0 (0–106)
15q=	1	0 (0–37.1)	1	0 (0–66.2)	1	0 (0–71.8)	0.62	1.03 (0.03–5.83)
16p–	1	0 (0–142)	1	0 (0–259)	1	0 (0–303)	1	0 (0–13.9)
16q–	1	0 (0–438)	1	0 (0–850)	1	0 (0–942)	1	0 (0–50.8)
16p=	1	0 (0–63.9)	1	0 (0–110)	1	0 (0–116)	0.11	3.52 (0.42–13)
16q=	1	0 (0–80.4)	0.029	35 (0.86–207)	1	0 (0–144)	0.38	2.13 (0.05–12.2)
17p–	1	0 (0–125)	1	0 (0–172)	1	0 (0–168)	0.27	3.16 (0.08–18.3)
17q–	1	0 (0–288)	1	0 (0–544)	1	0 (0–537)	0.13	7.51 (0.18–45.4)
17p=	1	0 (0–197)	1	0 (0–348)	1	0 (0–347)	1	0 (0–20.2)
17q=	1	0 (0–52.6)	1	0 (0–90.3)	1	0 (0–101)	0.51	1.42 (0.04–8.04)
18+	6.7×10^{-6}	91 (18–288)	1	0 (0–186)	1	0 (0–181)	2.9×10^{-5}	15.3 (4.8–37.6)
18p–	1	0 (0–665)	1	0 (0–1.41e+03)	1	0 (0–1.28e+03)	1	0 (0–77.6)
18q=	1	0 (0–233)	1	0 (0–454)	1	0 (0–493)	1	0 (0–25.3)
19+	1	0 (0–687)	1	0 (0–982)	1	0 (0–1.26e+03)	1	0 (0–72.9)
19p=	1	0 (0–116)	1	0 (0–209)	0.018	58 (1.41–351)	0.28	3.13 (0.08–18.1)
19q=	1	0 (0–93)	0.023	45.1 (1.11–265)	0.02	51.1 (1.25–308)	0.058	5.28 (0.63–19.7)
20q–	1	0 (0–35.2)	1	0 (0–61.1)	0.064	15.5 (0.38–91.8)	0.021	3.94 (1.06–10.3)
20p=	1	0 (0–503)	1	0 (0–796)	1	0 (0–953)	1	0 (0–48.4)
20q=	1	0 (0–104)	1	0 (0–165)	1	0 (0–192)	0.31	2.78 (0.07–15.9)
21+	1	0 (0–100)	1	0 (0–170)	0.023	43.8 (1.07–263)	0.31	2.75 (0.07–15.8)
21q–	1	0 (0–808)	1	0 (0–1.55e+03)	1	0 (0–1.73e+03)	0.048	22.3 (0.51–154)
21q=	1	0 (0–127)	1	0 (0–210)	1	0 (0–223)	0.26	3.32 (0.08–19.2)
22+	1	0 (0–88.3)	1	0 (0–142)	1	0 (0–139)	0.073	4.61 (0.55–17.2)
22q–	1	0 (0–256)	1	0 (0–305)	1	0 (0–296)	0.13	7.19 (0.18–43.4)
22q=	1	0 (0–48.9)	1	0 (0–81.5)	1	0 (0–82.8)	0.54	1.29 (0.03–7.28)

This table provides numerical data plotted in Fig. 3a. Events were grouped by chromosome and copy number, with loss and CN-LOH events subdivided by p-arm vs. q-arm; events observed in $\geq$30 individuals were tested for association with incident blood cancers (diagnosed $>$1 year after DNA collection in individuals with no previous cancer) using a two-sided CMH test.

(b) Analyses with no restrictions on blood counts at assessment

	CLL		MPN		MDS		Any blood cancer	
	N_{case}=199, $N_{control}$=375,954		N_{case}=138, $N_{control}$=375,893		N_{case}=70, $N_{control}$=375,818		N_{case}=1,383, $N_{control}$=377,192	
mCA	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)
1+	1	0 (0–151)	1	0 (0–250)	0.0091	116 (2.76–760)	0.00081	18 (3.52–57.8)
1p–	1	0 (0–441)	1	0 (0–666)	1	0 (0–989)	0.0027	29.6 (3.21–132)
1q–	1	0 (0–327)	1	0 (0–597)	1	0 (0–921)	0.092	11 (0.26–71.5)
1p=	0.24	3.59 (0.09–20.4)	0.00087	16.9 (3.41–50.8)	0.098	9.89 (0.25–57.5)	0.00079	4.27 (1.83–8.53)
1q=	1	0 (0–17.1)	1	0 (0–25.6)	1	0 (0–47.8)	0.66	1.35 (0.16–4.91)
2p–	0.053	18.8 (0.47–110)	1	0 (0–117)	1	0 (0–204)	0.31	2.76 (0.07–15.9)
2q–	1	0 (0–220)	1	0 (0–365)	1	0 (0–556)	1	0 (0–30.5)
2p=	1	0 (0–80.9)	1	0 (0–126)	1	0 (0–225)	1	0 (0–11.8)
2q=	1	0 (0–65.8)	1	0 (0–98.4)	1	0 (0–194)	1	0 (0–9.35)
3+	1.9×10^{-5}	63.6 (12.6–200)	1	0 (0–127)	1	0 (0–196)	1.1×10^{-7}	20.9 (8.02–45.9)
3p–	1	0 (0–377)	1	0 (0–614)	1	0 (0–940)	1	0 (0–53.6)
3p=	1	0 (0–63.7)	1	0 (0–99.4)	1	0 (0–180)	1	0 (0–9.61)
3q=	1	0 (0–89.6)	1	0 (0–141)	1	0 (0–247)	0.26	3.4 (0.08–19.8)
4q–	0.059	16.8 (0.42–98.2)	1	0 (0–105)	0.026	39.2 (0.96–235)	0.0086	7.47 (1.5–22.8)
4p=	1	0 (0–239)	1	0 (0–295)	1	0 (0–618)	0.12	8.01 (0.19–49.1)
4q=	1	0 (0–44.8)	1	0 (0–77.3)	6.2×10^{-6}	94.3 (18.5–299)	0.00039	8.62 (2.73–20.9)
5+	1	0 (0–93.6)	1	0 (0–157)	1	0 (0–265)	0.24	3.72 (0.09–21.8)
5q–	0.054	18.2 (0.45–107)	0.034	29.5 (0.73–172)	0.00015	122 (14–482)	3.1×10^{-6}	16.5 (5.84–37.6)
5q=	1	0 (0–80.3)	1	0 (0–118)	1	0 (0–213)	0.29	2.96 (0.07–17.1)
6+	1	0 (0–514)	1	0 (0–670)	1	0 (0–1.64e+03)	1	0 (0–71.7)
6q–	1	0 (0–168)	1	0 (0–294)	1	0 (0–495)	0.15	6.45 (0.16–39)
6p=	1	0 (0–39.3)	1	0 (0–57.8)	1	0 (0–108)	0.15	2.97 (0.36–11)
6q=	1	0 (0–124)	1	0 (0–193)	1	0 (0–316)	1	0 (0–18.2)
7p–	1	0 (0–282)	1	0 (0–499)	1	0 (0–1.28e+03)	1	0 (0–46.8)
7q–	1	0 (0–87.1)	1	0 (0–154)	1	0 (0–237)	0.00028	13.7 (3.6–37.4)
7p=	1	0 (0–124)	1	0 (0–195)	1	0 (0–313)	1	0 (0–17.9)
7q=	1	0 (0–94.7)	0.027	37.8 (0.93–221)	6.8×10^{-7}	203 (39–664)	0.00027	13.9 (3.66–37.9)
8+	0.046	21.7 (0.54–127)	1	0 (0–129)	0.018	58.9 (1.43–356)	1.3×10^{-6}	19.4 (6.84–45)
8p–	0.016	64.9 (1.56–406)	1	0 (0–415)	1	0 (0–847)	0.1	9.91 (0.24–61.4)
8p=	1	0 (0–297)	1	0 (0–473)	1	0 (0–1.16e+03)	1	0 (0–42)
8q=	1	0 (0–93)	1	0 (0–138)	1	0 (0–238)	1	0 (0–13.6)
9+	0.036	27.8 (0.68–166)	0.00026	90 (10.5–354)	1	0 (0–270)	0.0024	12.1 (2.4–37.8)
9q–	1	0 (0–271)	1	0 (0–399)	1	0 (0–661)	1	0 (0–37.8)
9p=	1	0 (0–31.3)	6.3×10^{-52}	402 (239–671)	1	0 (0–85.8)	1.9×10^{-29}	33.5 (21.1–51.2)
9q=	1	0 (0–24)	1	0 (0–38.8)	1	0 (0–66)	0.63	0 (0–3.56)
10q–	1	0 (0–38.4)	1	0 (0–54.4)	1	0 (0–121)	1	0 (0–5.4)
10p=	1	0 (0–227)	1	0 (0–341)	1	0 (0–701)	1	0 (0–32.4)
10q=	1	0 (0–110)	1	0 (0–161)	1	0 (0–312)	1	0 (0–16)
11p–	1	0 (0–199)	1	0 (0–297)	1	0 (0–516)	1	0 (0–30.3)
11q–	0.00012	33.8 (6.79–103)	1	0 (0–70.4)	1	0 (0–112)	2.3×10^{-8}	15.1 (6.71–29.9)
11p=	1	0 (0–17.4)	1	0 (0–26.1)	1	0 (0–48.1)	0.41	0 (0–2.53)
11q=	0.17	5.52 (0.14–31.5)	0.11	8.79 (0.22–50.5)	0.07	14.1 (0.35–82.3)	0.00032	5.71 (2.26–12)
12+	8.9×10^{-42}	142 (86.5–225)	0.11	8.46 (0.21–48.6)	0.077	12.8 (0.32–74.9)	4.8×10^{-35}	26.9 (18.1–38.9)
12p–	1	0 (0–431)	1	0 (0–769)	1	0 (0–1.27e+03)	1	0 (0–66.2)
12q–	1	0 (0–263)	0.011	98.6 (2.37–628)	1	0 (0–680)	0.11	9.2 (0.22–57.4)
12p=	1	0 (0–227)	1	0 (0–339)	1	0 (0–709)	1	0 (0–32.8)
12q=	1	0 (0–41.9)	1	0 (0–66.9)	1	0 (0–128)	1	0 (0–6.27)
13q–	1.2×10^{-54}	212 (133–327)	1	0 (0–38.4)	1	0 (0–56.5)	6.1×10^{-33}	28.9 (19–42.6)
13q=	1.4×10^{-11}	49.5 (20.8–102)	1	0 (0–34.4)	1	0 (0–55.7)	4.7×10^{-6}	7.71 (3.47–14.9)
14+	1	0 (0–37.5)	1	0 (0–65.2)	0.041	24.3 (0.6–144)	0.49	1.5 (0.04–8.55)
14q–	2.3×10^{-9}	109 (33.5–276)	1	0 (0–139)	1	0 (0–215)	8.7×10^{-8}	21.7 (8.31–47.9)
14q=	1	0 (0–12.2)	0.18	5.14 (0.13–29.3)	1	0 (0–33.6)	0.062	2.4 (0.77–5.67)
15+	0.14	6.8 (0.17–39.4)	1	0 (0–47.1)	1	0 (0–56.2)	0.086	2.95 (0.6–8.84)
15q–	0.011	94.6 (2.21–640)	1	0 (0–700)	1	0 (0–1.19e+03)	0.074	13.9 (0.33–93.1)
15q=	1	0 (0–20.1)	1	0 (0–31.5)	1	0 (0–58)	1	0.81 (0.02–4.54)
16p–	1	0 (0–73.1)	1	0 (0–115)	1	0 (0–235)	1	0 (0–10.6)
16q–	0.00018	115 (12.9–472)	1	0 (0–362)	1	0 (0–595)	0.0074	16.9 (1.91–69.5)
16p=	1	0 (0–32.7)	1	0 (0–53)	1	0 (0–91.1)	0.18	2.6 (0.31–9.57)
16q=	1	0 (0–42.6)	0.057	17.4 (0.43–101)	1	0 (0–111)	0.46	1.62 (0.04–9.25)
17p–	1	0 (0–56.2)	1	0 (0–91.5)	1	0 (0–135)	0.083	4.27 (0.51–16)
17q–	1	0 (0–154)	1	0 (0–258)	1	0 (0–403)	0.16	5.88 (0.14–35.5)
17p=	1	0 (0–107)	1	0 (0–164)	1	0 (0–277)	1	0 (0–15.3)
17q=	1	0 (0–28.6)	1	0 (0–43.8)	1	0 (0–80.5)	0.6	1.09 (0.03–6.19)
18+	1×10^{-6}	58.9 (15.5–159)	1	0 (0–92.3)	1	0 (0–145)	8.1×10^{-7}	15.2 (5.91–32.9)
18p–	1	0 (0–330)	1	0 (0–581)	1	0 (0–974)	0.085	12.1 (0.28–80.1)
18q=	1	0 (0–121)	1	0 (0–188)	1	0 (0–347)	1	0 (0–17.6)
19+	1	0 (0–301)	1	0 (0–463)	1	0 (0–842)	1	0 (0–42.6)
19p=	1	0 (0–61.9)	1	0 (0–97)	0.023	45.3 (1.11–271)	0.35	2.36 (0.06–13.5)
19q=	1	0 (0–50.4)	0.047	21.1 (0.53–123)	0.00034	79.7 (9.25–310)	0.002	7.88 (2.1–20.9)
20q–	1	0 (0–18.7)	1	0 (0–31)	0.084	11.7 (0.29–68.4)	0.053	2.9 (0.78–7.54)
20p=	1	0 (0–235)	1	0 (0–361)	1	0 (0–812)	1	0 (0–33.6)
20q=	0.069	14.2 (0.35–82)	1	0 (0–81.8)	1	0 (0–158)	0.085	4.19 (0.5–15.6)
21+	1	0 (0–52.9)	1	0 (0–87.5)	0.028	36.4 (0.89–217)	0.38	2.09 (0.05–12)
21q–	1	0 (0–343)	1	0 (0–540)	1	0 (0–1.09e+03)	0.08	12.8 (0.3–82.9)
21q=	1	0 (0–66.2)	1	0 (0–101)	1	0 (0–173)	0.34	2.46 (0.06–14.2)
22+	0.088	11 (0.27–63.2)	1	0 (0–71.1)	1	0 (0–108)	0.00045	8.31 (2.64–20.1)
22q–	6.4×10^{-16}	188 (75.5–404)	0.027	37.8 (0.93–224)	1	0 (0–233)	2.2×10^{-12}	34.9 (15.7–69.7)
22q=	1	0 (0–25.1)	1	0 (0–40.5)	1	0 (0–63.1)	1	0.96 (0.02–5.4)

This table provides results of analogous analyses removing the restrictions we imposed on blood counts (relevant to the cancer(s) analyzed) in our primary analyses (lymphocyte count 1–3.5×10^9/L, red cell count $<6.1\times10^{12}$/L for males and $<5.4\times10^{12}$/L for females, platelet count $<450\times10^9$/L, RDW $<15\%$).

Supplementary Table 21. Effects of known CLL GWAS variants on mosaic +12 and 13q LOH risk.

			CLL, Law et al. [65]		CLL, UK Biobank		Mosaic +12		Mosaic 13q LOH	
Locus	hg19 bp	Risk allele	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
2p22.2	37603801	rs888096:A	1.15 (1.09–1.21)	5.2e-08	1.02 (0.91–1.14)	0.74	1.04 (0.93–1.16)	0.54	1.08 (0.98–1.18)	0.12
2q13	111616619	rs1002015:C	1.30 (1.23–1.37)	2.2e-23	1.18 (1.06–1.32)	0.003	1.15 (1.03–1.29)	0.016	1.15 (1.05–1.27)	0.003
2q13	111831793	rs58055674:C	1.41 (1.32–1.50)	2e-27	1.17 (1.02–1.33)	0.027	1.17 (1.02–1.34)	0.025	1.16 (1.03–1.30)	0.011
2q13	111927379	rs6708784:G	1.30 (1.24–1.37)	2.7e-25	1.40 (1.25–1.56)	2.8e-09	1.22 (1.10–1.37)	0.00033	1.26 (1.15–1.38)	1.3e-06
2q33.1	202023949	rs7558911:A	1.18 (1.12–1.24)	5.1e-11	1.26 (1.13–1.41)	3.1e-05	1.09 (0.98–1.22)	0.12	1.06 (0.97–1.16)	0.2
2q37.1	231154012	rs34004493:G	1.39 (1.31–1.47)	3.7e-32	1.40 (1.25–1.57)	9.9e-09	1.30 (1.16–1.46)	1.1e-05	1.33 (1.21–1.47)	1.2e-08
2q37.3	242294913	rs3755397:G	1.32 (1.22–1.43)	9.5e-12	1.32 (1.12–1.56)	0.00081	1.14 (0.96–1.36)	0.14	1.12 (0.96–1.29)	0.14
3p24.1	27777779	rs9880772:A	1.16 (1.11–1.22)	1.9e-09	1.16 (1.04–1.30)	0.006	1.10 (0.98–1.23)	0.096	1.12 (1.02–1.23)	0.014
3q26.2	169497585	rs1317082:A	1.19 (1.12–1.26)	5.8e-09	1.21 (1.06–1.38)	0.0049	1.12 (0.98–1.28)	0.09	1.40 (1.25–1.58)	1.8e-08
3q28	188128794	rs73192661:C	1.13 (1.07–1.19)	1.7e-06	1.03 (0.93–1.15)	0.56	1.04 (0.93–1.16)	0.54	1.07 (0.97–1.17)	0.18
4q25	109025865	rs7690934:C	1.16 (1.11–1.22)	6.1e-09	1.19 (1.06–1.33)	0.0024	1.19 (1.06–1.33)	0.0027	1.00 (0.91–1.10)	1
4q26	114698696	rs1476569:G	1.14 (1.08–1.20)	4.5e-06	1.04 (0.93–1.18)	0.48	1.11 (0.99–1.26)	0.083	1.08 (0.98–1.20)	0.14
5p15.33	1285974	rs7705526:A	1.18 (1.12–1.25)	5.9e-10	1.19 (1.06–1.33)	0.0028	1.11 (0.99–1.24)	0.081	1.20 (1.09–1.31)	0.00025
5p15.33	1321873	rs10073340:T	1.13 (1.06–1.20)	0.00028	1.21 (1.06–1.39)	0.006	1.15 (1.00–1.33)	0.045	1.13 (1.00–1.27)	0.044
6p25.3	412802	rs9392504:A	1.33 (1.26–1.40)	9.8e-29	1.34 (1.20–1.50)	1.8e-07	1.07 (0.96–1.19)	0.25	1.19 (1.08–1.30)	0.0003
6p25.2	2969278	rs73718779:T	1.14 (1.06–1.23)	0.0007	1.18 (1.00–1.39)	0.054	1.08 (0.91–1.28)	0.4	1.00 (0.86–1.16)	0.97
6p21.32	32578127	rs9271176:G	1.29 (1.22–1.36)	3.2e-20	1.33 (1.18–1.50)	4.4e-06	1.10 (0.98–1.24)	0.11	1.18 (1.07–1.31)	0.0012
6p21.31	33546930	rs210143:C	1.26 (1.19–1.33)	5.8e-16	1.01 (0.90–1.14)	0.81	1.22 (1.07–1.38)	0.0021	1.00 (0.91–1.11)	0.99
6q25.2	154471225	rs4869818:G	1.15 (1.09–1.21)	4.1e-08	1.16 (1.04–1.29)	0.0093	1.13 (1.01–1.26)	0.029	0.95 (0.87–1.04)	0.3
7q31.33	124392512	rs2267708:T	1.16 (1.10–1.22)	8.6e-09	1.12 (1.00–1.24)	0.047	0.98 (0.88–1.10)	0.74	1.12 (1.02–1.23)	0.015
8q22.3	103577865	rs2511713:G	1.17 (1.10–1.23)	6e-08	1.14 (1.01–1.28)	0.032	1.16 (1.03–1.31)	0.016	1.11 (1.00–1.23)	0.046
8q24.21	128200971	rs2466029:G	1.23 (1.17–1.30)	7.5e-16	1.16 (1.03–1.29)	0.0098	1.10 (0.98–1.23)	0.091	1.09 (0.99–1.19)	0.081
9p21.3	22206987	rs1679013:C	1.16 (1.10–1.22)	2.2e-08	1.18 (1.05–1.31)	0.0038	1.12 (1.00–1.25)	0.05	1.21 (1.10–1.33)	7.4e-05
10q23.31	90752018	rs6586163:A	1.23 (1.17–1.29)	1.1e-15	1.31 (1.18–1.47)	9.2e-07	1.30 (1.17–1.46)	3.2e-06	1.20 (1.10–1.32)	8.5e-05
11p15.5	2321650	rs2651823:A	1.18 (1.13–1.25)	5.2e-11	1.20 (1.07–1.33)	0.0012	0.95 (0.85–1.06)	0.38	1.26 (1.15–1.38)	6.4e-07
11q24.1	123355391	rs35923643:G	1.63 (1.53–1.72)	4.3e-58	1.58 (1.40–1.78)	1.3e-13	1.17 (1.02–1.33)	0.024	1.44 (1.30–1.60)	8e-12
12q24.13	113381376	rs6489882:G	1.16 (1.10–1.22)	4.8e-08	1.10 (0.98–1.23)	0.099	1.13 (1.01–1.27)	0.031	1.06 (0.97–1.17)	0.2
15q15.1	40403657	rs8024033:C	1.26 (1.20–1.32)	7.1e-19	1.32 (1.18–1.47)	6.9e-07	1.19 (1.07–1.33)	0.0016	1.12 (1.02–1.23)	0.013
15q21.3	56777691	rs142215530:G	1.39 (1.29–1.50)	2.5e-18	1.35 (1.16–1.57)	0.0001	1.36 (1.17–1.59)	7.9e-05	1.26 (1.10–1.43)	0.00075
15q23	70020525	rs11637565:G	1.35 (1.28–1.42)	2e-31	1.34 (1.20–1.49)	1.4e-07	1.03 (0.92–1.15)	0.64	1.16 (1.06–1.27)	0.0019
15q25.2	83237899	rs17356118:A	1.12 (1.05–1.19)	0.00025	1.06 (0.93–1.21)	0.38	1.10 (0.96–1.26)	0.16	1.08 (0.97–1.21)	0.18
16q24.1	85973866	rs305065:C	1.16 (1.10–1.22)	7.6e-08	0.98 (0.88–1.10)	0.76	1.00 (0.89–1.12)	0.96	1.17 (1.06–1.29)	0.0017
16q24.1	85928621	rs391855:A	1.34 (1.27–1.41)	1.3e-28	1.28 (1.14–1.43)	1.7e-05	1.17 (1.05–1.31)	0.0058	1.13 (1.03–1.24)	0.012
18q21.32	57622287	rs4368253:C	1.17 (1.11–1.24)	1.3e-08	1.11 (0.98–1.25)	0.091	1.21 (1.07–1.37)	0.002	1.00 (0.91–1.10)	1
18q21.33	60788745	rs77551289:A	1.37 (1.25–1.50)	1.8e-11	1.26 (1.03–1.53)	0.026	1.15 (0.94–1.40)	0.17	1.18 (1.00–1.39)	0.049
18q21.33	60793921	rs4987852:C	1.32 (1.20–1.44)	4.7e-09	1.20 (0.98–1.46)	0.076	1.29 (1.06–1.57)	0.0099	1.13 (0.95–1.34)	0.17
19q13.3	47176752	rs874460:C	1.24 (1.15–1.34)	3.4e-08	1.29 (1.08–1.55)	0.0042	0.89 (0.76–1.04)	0.14	1.16 (1.00–1.34)	0.047
1p36.11	23943735	rs34676223:C	1.19 (1.14–1.25)	5e-13	1.09 (0.97–1.24)	0.14	0.96 (0.85–1.08)	0.47	1.14 (1.03–1.26)	0.013
1q42.13	228880296	rs41271473:G	1.19 (1.13–1.26)	1.1e-10	1.04 (0.91–1.19)	0.6	1.19 (1.03–1.38)	0.018	1.02 (0.91–1.15)	0.71
4q24	102741002	rs71597109:C	1.17 (1.11–1.22)	1.4e-10	1.26 (1.11–1.42)	0.00023	1.00 (0.89–1.13)	0.98	1.20 (1.08–1.32)	0.00065
4q35.1	185254772	rs57214277:T	1.13 (1.08–1.18)	3.7e-08	1.06 (0.95–1.18)	0.33	1.04 (0.93–1.16)	0.51	1.10 (1.00–1.21)	0.038
6p21.31	34616322	rs3800461:C	1.20 (1.13–1.28)	2e-08	1.29 (1.10–1.51)	0.0013	1.27 (1.08–1.49)	0.0033	1.21 (1.06–1.38)	0.0053
11q23.2	113517203	rs61904987:T	1.24 (1.16–1.32)	2.5e-11	1.19 (1.02–1.38)	0.025	1.03 (0.88–1.21)	0.73	1.22 (1.07–1.38)	0.0026
18q21.1	47843534	rs1036935:A	1.15 (1.10–1.21)	3.3e-08	0.98 (0.86–1.12)	0.77	1.14 (1.00–1.30)	0.046	0.99 (0.89–1.11)	0.91
19p13.3	4069119	rs7254272:A	1.17 (1.10–1.23)	4.7e-08	0.96 (0.83–1.11)	0.58	1.00 (0.86–1.15)	0.97	1.10 (0.98–1.24)	0.099
22q13.33	50971266	rs140522:T	1.15 (1.10–1.20)	2.7e-09	1.21 (1.08–1.35)	0.001	0.95 (0.84–1.06)	0.35	1.18 (1.08–1.30)	0.0005

For 46 lead CLL-associated variants reported by Law et al. [65], we computed CLL effect size, mosaic +12 effect size, and mosaic 13q LOH (i.e., del(13q) or 13q CN-LOH) effect size in UK Biobank and compared these effect sizes to reported CLL effect size in Law et al. [65]. We computed effect sizes and *P*-values (two-sided) using logistic regression. Details of analyses are provided in Supplementary Note 8.

Supplementary Table 22. Risk increase for cardiovascular disease (MI/stroke) during 5–10-year follow-up conferred by mCAs.

Event	# carriers	# matched controls	MI/stroke OR (95% CI)	*P*
Any loss	2700	62100	1.16 (0.92–1.47)	0.22
Any CN-LOH	7020	140400	0.96 (0.81–1.13)	0.68
Any gain	1822	45550	0.72 (0.51–1.02)	0.068
Any mCA	15015	180180	0.97 (0.87–1.09)	0.67
DNMT3A loss	107	3638	1.47 (0.46–4.75)	0.46
TET2 loss	68	1428	1.24 (0.29–5.28)	0.68
JAK2 CN-LOH	291	9312	**2.49 (1.47–4.19)**	**0.0024**

This table provides numerical data plotted in Fig. 3b. The number of controls varies across mosaic events because cases and controls for each event were matched for assessment year, age, sex, smoking, hypertension, BMI, and type 2 diabetes status. The case-control ratio was chosen independently for each event to optimize statistical power. *P*-values, two-sided Fisher's exact test.

Supplementary Table 23. Numbers of individuals carrying multiple overlapping CN-LOH mutations on the same chromosome arm.

Arm	# of multi-CN-LOH individuals	# carrying high-risk inherited or acquired variants
1p	20	10 (rare inherited *MPL* variants): rs146249964 splice donor (1), rs28928907 missense (5), rs144210383 missense (1), 454bp exon 10 deletion (3)
2q	1	0
3p	1	0
3q	1	0
4q	1	0
6p	4	0
9p	37	32 (common inherited *JAK2* 46/1 haplotype)
11q	2	0
13q	14	13 (somatic deletions of *DLEU* region)
14q	7	0
15q	17	13 (rare inherited *TM2D3* variants) 70kb whole-gene deletion (12), rs754640606 missense (1)
16p	1	0
17p	1	0
17q	2	0
19q	1	0
20q	1	0
21q	1	0
22q	3	0

We identified 110 examples of multiple overlapping CN-LOH mutations occurring on the same chromosome arm using the modified hidden Markov model we previously developed and applied to the UK Biobank interim release (described in Supplementary Note 1.8 of ref. [9]). For each affected chromosome arm, we determined which events could be attributed to a high-risk inherited or acquired variant on the affected arm (listed in the third column of this table).