Germline risk of clonal haematopoiesis

Alexander J. Silver1,2, Alexander G. Bick1,3,4,5 and Michael R. Savona1,2,4,5✉

Abstract | Clonal haematopoiesis (CH) is a common, age-related expansion of blood cells with somatic mutations that is associated with an increased risk of haematological malignancies, cardiovascular disease and all-cause mortality. CH may be caused by point mutations in genes associated with myeloid neoplasms, chromosomal copy number changes and loss of heterozygosity events. How inherited and environmental factors shape the incidence of CH is incompletely understood. Even though the several varieties of CH may have distinct phenotypic consequences, recent research points to an underlying genetic architecture that is highly overlapping. Moreover, there are numerous commonalities between the inherited variation associated with CH and that which has been linked to age-associated biomarkers and diseases. In this Review, we synthesize what is currently known about how inherited variation shapes the risk of CH and how this genetic architecture intersects with the biology of diseases that occur with ageing.

Haematopoietic stem cells (HSCs). Cells that are responsible for the creation of all blood cells in the human body and are multipotent in that they may differentiate into any type of mature blood cell. They are found in the bone marrow in adult humans.

Haematopoiesis, the process by which blood cells are generated, begins in embryogenesis and continues throughout an individual’s lifespan1. Haematopoietic stem cells (HSCs) are responsible for the creation of all mature blood cells, including red blood cells, platelets, and the numerous myeloid cells (such as monocytes and neutrophils) and lymphoid cells (such as T cells and B cells), that comprise the innate and adaptive immune systems. Roughly 50,000–200,000 HSCs in an adult human produce an estimated 1010–1012 progeny blood cells every day2. Over the course of repeated cell divisions during a lifetime, HSCs accumulate unique patterns of acquired DNA mutations. Each HSC gains approximately one new exonic variant per decade3, although somatic changes can and do occur throughout the non-coding genome as well. Although the majority of these acquired mutations involves genetic loci that do not lead to phenotypic consequences, mutations can occur in portions of the genome that may confer a relative fitness advantage to affected HSCs. Such a fitness advantage can take several forms, including an increased proliferative drive, a more durable capacity for self-renewal that counteracts ageing-related drop-out from the HSC pool, or an improved ability to evade death from cellular damage4. Over time, the relative fitness advantage of these mutated HSCs can result in the clonal production of a large number of progeny that bear the same somatic alterations.

This expansion of haematopoietic cells with the same acquired mutation is referred to as clonal haematopoiesis (CH). The first experimental evidence suggestive of widespread age-related clonality in the blood dates back to the mid-1990s5, but the genetic characterization of the acquired clonal mutations has only been possible over the past several years owing to three parallel developments. First, advances in next-generation sequencing technologies have enabled the identification of mutations with high resolution (that is, single base-pair changes) even when these lesions are present in just a fraction of sampled cells. Second, bioinformatic innovations in analyses of big data have allowed for the detection of mutations in meaningfully large datasets. Third, many simultaneous efforts to build institutional and national cohorts consisting of tens to hundreds of thousands of individuals have begun to come to fruition, providing ample substrate in which to look for acquired mutations as well as their associations with inherited variation and clinical phenotypes. The convergence of these trends has led to the identification of several distinct types of CH that are common and hold important implications for human health. Specifically, it is now known that CH is linked to a heightened risk of mortality and multiple common diseases of ageing, including blood cancers and cardiovascular disease (CVD).

Moreover, this recent work has shown that germline variation influences the risk of developing CH and the type of acquired mutation that a clone will have.

In this Review, we aim to provide a complete synthesis of available research of how inherited genetic variation influences the incidence of CH. We detail the current evidence from twin studies and large-scale genetic association studies regarding the heritable risk of CH and consider how this genetic architecture intersects the biology of ageing.

Clonal haematopoiesis
Somatic variation giving rise to CH. Here, we use ‘CH’ as an umbrella term that refers to the presence of an expanded mutant clone of any sort within the blood, excluding the reactive expansion of immune cells within
lymphoid organs and frank malignancy. CH is a common phenomenon among the general population and its prevalence increases significantly with age. The blood is not unique in the accumulation of mutations with age, a trend which has also been observed in the solid organs, although CH involves a distinct set of recurrently mutated genes and comes from a readily available tissue source. To date, the literature has largely classified CH by the type of somatic variation that can be observed within the clone: gain, loss and copy-neutral loss of heterozygosity (CN-LOH) events involving a large portion of a chromosome or single-nucleotide variation and short insertions/deletions (indels).

By far the most common genetic lesion seen in CH is mosaic loss of the Y chromosome (mLOY) in men. Additionally, mosaic chromosomal alterations (mCAs), single-nucleotide variants (SNVs) and indels in genes associated with myeloid malignancies, and putative incidental mutations/genomic drift (CH with unknown drivers) have all been documented. In the absence of a haematological malignancy, these SNVs/indels are known as CH of indeterminate potential (CHIP) when the mutations are present at ≥2% variant allele fraction (VAF). This classification scheme is largely a by-product of how CH is identified in existing studies. Chromosomal abnormalities, including mLOY and mCAs, can be interrogated using genome-wide genotyping arrays (such as those used for genome-wide association studies (GWAS)), whereas whole-exome or whole-genome sequencing (WGS) data can identify SNVs or indels but is suboptimal for identifying mCA events.

Of these subtypes of CH, the vast majority of studies of inherited risk have examined mLOY, mCAs or CHIP; therefore, the remainder of this Review focuses on these three entities (Fig. 1).

Epidemiology of CH. All types of CH are strongly age associated. It has been postulated that all adults have some CH mutations at extremely low clonal fractions, but prevalence estimates for CH in the population are typically based on the identification of clones with VAF of at least ~2%, which is approximately the limit of detection for many commonly used assays. An estimated 1.7–20% of men have some amount of mLOY, with the prevalence increasing to >40% of individuals by age 70 in the largest epidemiological study to date. The X chromosome seems to have a lower rate of mCA acquisition. Approximately 8% of women over the age of 65 years have detectable X chromosome mosaicism, whereas men rarely have X chromosome mosaicism in the blood at any age. Autosomal mCA events are the least commonly observed type of CH, affecting 1–5% of the population older than 70 years of age, whereas CHIP is estimated to affect >10% of individuals older than 70 years of age. An individual can have both mCAs and
**Mosaic loss of the Y chromosome (mLOY).** A common type of clonal haematopoiesis affecting XY individuals which is associated with the complete absence of the Y chromosome in the clonal cell population.

**Mosaic chromosomal alterations (mCAs).** A type of clonal haematopoiesis in which a portion of the genome has been duplicated (copy number gain), lost (copy number loss) or replaced with DNA from another allele (copy-neutral loss of heterozygosity).

**CH with unknown drivers**

This is CH in which there is no evidence of a genetic mutation that is known to cause clonal expansion. In such cases, clonality may be ascertained by the presence of a unique pattern of somatic mutations throughout the genome that is only present in a subset of cells.

**CH of indeterminate potential (CHIP).** A clinical term for a type of clonal haematopoiesis defined by single-nucleotide variants or small insertions/deletions (indels) in genes associated with myeloid malignancies in the presence of normal blood counts and at ≥2% variant allele fraction.

**Variant allele fraction (VAF).** The percentage of measured DNA alleles that contain a specified variant. For example, if half of a population of diploid cells each harbour a single copy of JAK2V617F, the variant allele fraction for this mutation would be 25%.

**Acute myeloid leukaemia (AML).** A blood cancer characterized by improper maturation of the myeloid lineage (monocytes and granulocytes), resulting in the production of many dysfunctional immature cells called blasts.

CHIP simultaneously, which occurs frequently with point mutations in JAK2 (a tyrosine kinase involved in multiple cytokine signalling pathways) and mCA events at the same locus. However, aside from the JAK2 locus, the co-occurrence of CHIP and mCAs as identified by bulk sequencing/genotyping from the same individual appears to be a rare event, although the prevalence is higher among patients treated for solid tumours.

The prevalence of CH varies across several demographic features. There is a sex bias for specific mCA lesions, with most of these having higher prevalence in men. Although some studies have suggested a male-bias for CHIP, other studies have found this association does not persist after controlling for potential confounders. Groups with different ancestries also have different prevalence. For instance, mLOY is less commonly observed in individuals of African ancestry than of European ancestry (0.4% versus 1.8%). Meanwhile, CHIP mutations are less frequently observed in individuals identifying as Hispanic or East Asian.

There are substantial differences across age in the distribution of mutated CHIP genes. In particular, mutations in the de novo DNA methyltransferase DNMT3A and in JAK2 can be observed with some regularity beginning in the third and fourth decade of life, whereas clones carrying mutations in spliceosome genes are generally detected no earlier than the fifth and sixth decades of life. The extent to which this distribution is shaped by differences in DNA sequence mutability, relative fitness advantage or interactions with an ageing microenvironment is still an area of active investigation.

Environmental exposures that increase somatic variant acquisition are significantly correlated with CH prevalence. In particular, smoking is robustly associated with CHIP and mLOY and mLOY is associated with CHIP. In the case of cytotoxic chemotherapy and radiation therapy, the mutational spectrum exhibits a marked enrichment of mutations in DNA damage response pathway genes. The outgrowth of CH clones following anticancer therapy is partly due to the expansion of pre-existing clones with a selective advantage but may also be from the introduction of new mutations by the anticancer agents themselves or due to stochastic effects from a bottleneck event for HSCs.

**Health consequences of CH.** Although most individuals with CH have normal haematological parameters, CH is associated with significant health consequences. With respect to larger chromosomal abnormalities, epidemiological studies have demonstrated associations between the mLOY and a broad range of health outcomes in men, including all-cause mortality, numerous types of cancer, cardiovascular events, Alzheimer disease, schizophrenia, autoimmune disease and age-related macular degeneration. Autosomal mCA events have been associated with an increased risk of haematological malignancies and diabetes, as well as with all-cause mortality only partially explained by excess cancer deaths. Additionally, even as mCAs are independently associated with a heightened risk of myeloid malignancies, a retrospective analysis of patients with solid tumours found relatively increased rates of haematological malignancies in patients with both mCAs and CHIP compared to those with either alone. Whether the presence of dual mCAs/CHIP is an indicator of individuals with particularly unstable genomes or whether the combination of these lesions cooperatively leads to malignancy risk remains to be determined. The heightened risk of infection and serious infectious complications may account for a portion of the excess mortality seen in patients with mCAs: a recent multinational study found that mCA events are moderately associated with risk for a wide range of infections (odds ratio (OR) = 1.06), including the risk of hospitalization for COVID-19 (OR = 1.6). Somatic mutations in recurrently mutated CHIP genes have been studied in both natural epidemiological contexts and in experimental models, which have revealed strong associations with mortality, malignancy and CVD. All-cause mortality is greater in individuals with CHIP compared to without CHIP; this is partly due to an increased risk of haematological malignancies, which has been observed across many studies. However, individuals with CHIP mutations have excess mortality compared with those who do not harbour such mutations even after controlling for blood cancer deaths.

This may be partly explained by an association between CHIP and CVD. On a population level, CHIP has been linked to a greater burden of atherosclerotic vessel disease and a heightened risk of myocardial infarction as well as to higher blood levels of the inflammatory marker C-reactive protein. Mouse models of CHIP have demonstrated mechanistic ties between certain common CHIP mutations and accelerated atherosclerosis as well as heart failure.

Despite the fact that numerous genes affected by somatic CHIP mutations have been associated with increased cancer and CVD risk, there are early indications of important functional differences in how each mutant gene might contribute to that risk. For instance, mutations in splicing factor U2AF1 are associated with a higher risk of acute myeloid leukaemia (AML) and with a shorter latency to disease than mutations in DNMT3A and TET2, encoding a dioxygenase that opposes the action of DNMT3A by promoting DNA demethylation, and in JAK2 are associated with coronary artery disease but may differ in how they contribute to blood cell dysfunction. In mouse models, mutations in Tet2, whose gene product recruits HDAC2 for the resolution of IL-6-mediated inflammation, are associated with increased expression of Il1b, Il6, Cxcl1, Cxcl2 and Cxcl3. While mutations in Jak2 also lead to higher Il1b expression, they additionally lead to plaque-promoting erythropoagocytosis, secretion of arterial spasm-inducing erythrocyte-derived microvesicles and thrombotic neutrophil extracellular traps. Yet, much remains to be learned about the relative risk of disease outcomes with specific CHIP genes, let alone how disease risk might vary across different protein-altering variants within each gene. The curation of large CHIP cohorts with inherited genotype and deep phenotype data will enable further investigation of how germline variation affects CHIP-to-disease risk.
There is accumulating evidence that CHIP mutations may interact with human illnesses beyond cancer and CVD. Somatic CHIP mutations have been associated with several diseases in which inflammation features prominently, including chronic obstructive pulmonary disease,10,11 adult-onset haemophagocytic lymphohistiocytosis,12 anti-neutrophil cytoplasmic antibody-associated vasculitis,13 CHIP also appears to be associated with several types of infections and with potentially severe disease manifestations among those infected with SARS-CoV-2 (REF 68), perhaps as a result of CHIP-exacerbated inflammatory signalling.74 Several recent analyses have also found high rates of somatic mutations in CHIP genes in people with immunodeficiency from HIV, which might be a consequence of a pro-inflammatory disease state but might equally well be due to the impaired clearance of CH clones by T cells.75,76 Furthermore, CHIP may have dynamic interactions with certain therapeutic interventions. Mutations in CHIP genes involved in the DNA damage response pathway, such as TP53 and PPM1D, are highly enriched following radiation treatment or treatment with a select few cytotoxic chemotherapies.77 Additionally, CHIP has been associated with significantly increased mortality following transcatheter aortic valve implantation,78 which is the first indication that CHIP might have an impact on surgical/procedural outcomes. Emerging research suggests that CHIP may impact patient outcomes following HSC transplantation (HSCT). Transplanted HSCs face several sizable and unique challenges, including the high replicative demand in order to reconstitute the entire population of blood cells as well as the exposure to immunosuppressive and cytotoxic therapies. The current evidence (nicely summarized in REFs.79,80) suggests that donor-derived CHIP is not uncommon in both allogeneic HSCT and autologous HSCT recipients and may increase risks of graft-versus-host disease, donor-derived leukaemia and overall mortality, although the interactions appear to be complex and may depend on both patient characteristics and the CHIP gene in question.

The disparate genomic lesions seen in CH and their associations with a broad range of consequential health outcomes has spurred research into how germline genetics influences the acquisition and outgrowth of specific somatic changes. In the next section, we discuss the associations between inherited variants and mLOY, mCAs and CHIP that have been described to date (FIG. 2).

**Early evidence for inherited risk of CH**

Although much of the knowledge about germline risk of CH has come from recent large-scale genetic association studies, some of the foundational insights in the field came from smaller studies relying on shared lineage to identify inherited risk factors. Starting even further back, research into inherited risk for haematological cancers provided signals that have informed the thinking around germline risk for CH, highlighting both the commonalities and differences between CH and malignancy.

**Insights from haematological cancers.** As CHIP mutations are also found in myeloid neoplasia, work on the genetic predispositions to haematological malignancies provided key initial insights linking the germline variation and expansion of somatic haematopoietic mutations. CHIP and myeloid malignancies such as myeloproliferative neoplasms (MPNs), myelodysplastic syndromes (MDS) and AML arise from similar origins in haematopoietic stem and progenitor cells (HSPCs).81,82 Despite the shared origins and patterns of acquired mutations, the vast majority of individuals with CHIP never develop a myeloid malignancy. Indeed, individuals with CHIP have normal counts of normal-appearing cells, whereas those with malignancy have abnormal numbers of blood cells and/or visibly dysmorphic cells. Given that CHIP is a potential precursor state to haematological cancer, many of the known germline risk factors for myeloid disease may also predispose to CHIP in a similar manner. Future study of the differences between the sets of germline variants predisposing more to CHIP versus the set predisposing more to malignancy may prove

---

**HSC transplantation (HSCT).** The provision of donor HSCs to patients with various types of malignant and non-malignant conditions. Autologous or allogeneic donor HSCs are provided to recipients after conditioning regimens of radiation and/or chemotherapy.

**Autologous HSCT**

HSC involving the harvest of patient HSCs to be provided later as a stem cell rescue after high-dose chemotherapy.

**Allogeneic HSCT**

HSC involving the transplantation of histocompatible HSCs from a non-self donor to a recipient.

**Known associations of germline variation with diseases of ageing:**

1. Malignancy
2. Cardiovascular disease
3. Dementia

**Fig. 2 | CH subtypes have shared and unique risk variants.** Many germline risk loci have been linked to the development of clonal haematopoiesis (CH). The three subtypes of CH that have received the greatest scrutiny in this area are mosaic loss of the Y chromosome (mLOY), mosaic chromosomal alterations (mCAs) and clonal haematopoiesis of indeterminate potential (CHIP). These three subtypes are enriched for several of the same germline variants such as those affecting DNA damage response genes CHEK2 and ATM, proliferation factor TCL1A, and telomerase component TERT. However, each of these entities also retains risk loci unique to it alone. Considering the spectrum of variants, one notable pattern is a rarity of mitosis-specific genes in CHIP compared to their relative abundance in mLOY and mCAs. Another broad theme is the high prevalence of previously identified associations between these germline loci and diseases of ageing, including malignancies, cardiovascular disease and dementia. The degree to which CH is involved in these known links to disease remains to be determined. Note: over 150 loci have been associated with mLOY, only a small number of which are depicted in this figure.

**mLOY**

- SPD1L, CENPU', CENPN, MAD2L1', PMF1, NAPAT, TP53', BCL2', BAX' and many others

**mCAs**

- FRA10B, MPL', TM2D3', DXZ1, DXZ4, FN', NEDD4'-TINF2', CTU2, NBN', MRE11', SP140', SP140L and TERC'

**CHIP**

- TET2
- RUNX1', KPNA4'-TRIM59', PINT (JAK2-CH) and GFI1B' (JAK2-CH)
- 46/1' (JAK2-CH) and SH2B3' (JAK2-CH)

**CHIP subtypes**

- (i) Malignancy
- (ii) Cardiovascular disease
- (iii) Dementia

---

www.nature.com/nrg
informative as to why only a minority of individuals ever progress from one to the other.

Many of the same germline variants predisposing to JAK2-mutated malignancies have also been associated with JAK2-CH. JAK2 is the most commonly mutated gene in MPNs and the JAK2 p.V617F mutation (JAK2v617F) is a characteristic feature of MPNs incorporated into the World Health Organization diagnostic criteria for over a decade. Consequently, some studies looking to define MPN germline risk have used cohorts formed exclusively of diagnosed myeloid disease/MPNs, whereas other studies have augmented cohorts of diagnosed MPNs with the addition of any individuals with a molecularly detectable JAK2v617F mutation (which may include undiagnosed MPNs as well as JAK2v617F-CH).

Box 1 | Germline variants can associate with the same somatic lesion but different levels of phenotypic risk

The somatic genetic lesions seen in clonal haematopoiisis of (CH) are commonly seen both in the context of malignancy and in otherwise healthy individuals. However, only a small minority of individuals with CH ever progress to malignancy. In a similar vein, evidence suggests that some CH may contribute to the risk of cardiovascular disease but only a fraction of people with CH ever experience a heart attack. This presents a unique challenge for researchers and clinicians: can we predict whether a given somatic mutant clone will follow a benign path or a more pathological one? In the future, examining an individual’s germline genotype may provide clues to this answer.

Increasing evidence suggests that there may be different degrees of risk for downstream phenotypes (for example, myeloproliferative neoplasms (MPNs)) even among germline polymorphisms that are all associated with the outgrowth of cell clones harbouring the same somatic mutation (for example, JAK2v617F) (see the figure). For instance, inherited variants in both intron 2 (SNPs rs2736100, rs2853677 and rs7705526) and intron 3 (SNP rs7726159) of TERT are associated with the risk of developing somatic JAK2V617F clones but only the germline variants in intron 2 have demonstrated a significant association with MPNs. Meanwhile, inherited variants affecting MECOM, HBS1L, MYB, RUNX1, HMGA1, FOXO1 and GATA2 are significantly enriched in cohorts of JAK2-mutated MPNs but have yet to be identified as significant signals in the much larger population who have JAK2-CH. One possible explanation for this could be that, once individuals with these germline variants develop a JAK2v617F clone, they have a short or non-existent CH phase and progress very quickly to MPNs. Future studies, including genetic association studies contrasting disease-negative and disease-positive CH cohorts, will hopefully provide greater information regarding the extent to which germline variation predisposing to clonal expansion also influences the risk of malignancy or other health outcomes.

The first inherited variation linked to JAK2v617F-mutated MPNs was the 46/1 or GGCC haplotype, a collection of single-nucleotide polymorphisms (SNPs) stretching across several hundred kilobases of DNA that includes the JAK2 gene itself. In several studies of patients with MPNs, the JAK2v617F somatic variant was identified in cis with the inherited 46/1 risk haplotype more often than would be predicted by chance, which might suggest that the haplotype provides a hypermutable substrate for somatic alterations. Furthermore, this haplotype may increase the rate of JAK2v617F clonal expansion. In a study of clonal dynamics preceding MPN diagnosis in 12 patients, the homozygosity for 46/1 was enriched in those patients with the highest average clonal growth rate, an intriguing finding which should be followed up in larger cohorts. Inherited polymorphisms in the telomerase reverse transcriptase (TERT) locus have also been linked to all varieties of MPNs in several studies and to JAK2v617F-mutated disease in several others. While most tissues in the human body lack the expression of TERT (a key enzyme in telomere maintenance), haematopoietic stem cells have the constitutive expression of this protein. The precise mechanism of how telomere regulation might influence the expansion of JAK2v617F or other CH clones is just beginning to be understood (see Overlap with biomarkers of ageing, below).

Two dozen additional loci imparting potential risk of MPNs have recently been identified. Similar to JAK2 and TERT, many of these are genes implicated in the functional regulation of HSCs (including SH2B3, TET2, ATM, GFI1B and RUNX1) among several others) although some of the loci with strong signals, such as PINT, have no known role in HSC biology. While the location of lead SNPs in or near key HSC regulators is strongly suggestive of mechanisms that disrupt normal HSC biology, variant-to-function analyses have provided added evidence in the case of GFI1B and CHEK2. In the first case, the lead SNP was located in a putative enhancer region downstream of GFI1B and was experimentally determined to lead to GFI1B expression, which, in turn, was shown to increase HSPC self-renewal. The second case involves a rare missense variant in CHEK2 and similarly demonstrated increased HSPC self-renewal following the knockdown of gene expression.

The genetic associations identified for JAK2-mutated malignancy and JAK2-CH are not completely overlapping. The examination of a JAK2v617F-CH cohort replicated associations with the 46/1 haplotype, TERT, SH2B3 and TET2 with nominally significant signals for CHEK2, ATM, PINT and GFI1B; additionally, KPN4 has been associated with a greater risk of all CHIP, inclusive of JAK2-CH. The present lack of replication of other MPN-associated loci in CHIP cohorts leads to the question of whether and how inherited variation might shape the convergent somatic mutational landscapes yet differ in the magnitude or type of attendant phenotypic risk.

Compared to the literature on JAK2, studies of haematological malignancies have been less revealing with respect to what germline factors may increase the risk
of somatic mutation in other CHIP genes. Family-based studies of inherited risk of MDS and AML have noted a high prevalence of non-disease CHIP in carriers of rare inherited variants affecting RUNX1, a member of the core binding factor family of transcription factors and a key regulator of definitive haematopoiesis. Aside from RUNX1, there are several other germline variants recognized to predispose to myeloid, lymphoid or plasma-cell neoplasms that could presumably also predispose to asymptomatic CHIP. Genetic association studies of CHIP-only cohorts (that is, only individuals without haematological disease; discussed in detail in the ‘Results from genetic association studies’ section below) have seen a significant signal with just one of these genes: TERT. Although the remainder are strong candidates for genes likely to predispose to CHIP, concrete evidence of this in asymptomatic individuals is currently lacking.

Evidence from sibling studies. Although no groups have conducted sibling studies of chromosomal mosaicism, several have looked at CHIP in siblings. The first study to examine the heritability of CHIP mutations using siblings looked only at the two most commonly mutated CHIP genes, DNMT3A and TET2 (Ref. 102). The authors looked at the risk-recurrence ratio (λs) for mutations within these genes among a set of 391 female sib-ships of French-Canadian ancestry and found no familial risk for DNMT3A mutations but a significantly increased risk for TET2 (λs = 2.24 for those ≥25 years of age, λs = 2.65 for those ≥65 years of age). One sib-ship consisting of seven sisters was notable for having TET2 mutations in 4/7 and a DNMT3A mutation in 1/7 sisters, raising the provocative but unanswered question of whether germline genetics or common environmental exposures did more to shape such a pedigree.

The heritability of CHIP has also been examined in twin pairs in two recently published studies. One study consisted of 299 twin pairs from Denmark, whereas the other was comprised of 79 twin pairs from the UK. Neither study found a higher concordance among MZ pairs for CHIP mutations specifically in DNMT3A or TET2 (Ref. 104). Of note, these studies each identified sets of MZ twins that shared identical CH mutations (KDM6A p.Q692X and DNMT3A p.R598X in the UK cohort and SRSF2 p.P95H and c.912_916delCTGGT in DNMT3A in the Denmark cohort), suggesting these mutations occurred in utero; several subsequent studies of patients with MPNs have identified JAK2 V617F and DNMT3A mutations that similarly arose during embryogenesis or childhood. Taken together, these twin studies provide no evidence for common, strong germline effects on the development of CHIP in the populations studied. However, the moderate power afforded by the size of the study cohorts precludes the detection of more modest effects. Additional twin studies on diverse populations, with the potential for subsequent meta-analysis, could supplement the existing work in this area. Future twin studies would also be warranted for mLOY and mCAs, the present lack of which is a notable gap in the field.

Results from genetic association studies

The bulk of the data regarding the inherited risk for CH comes from genetic association studies. These studies identify the enrichment of genetic variants in people with CH across large, unrelated and (more-or-less) diverse samples. Such analyses are well suited to finding common germline variants with modest effects that are noticeable in the aggregate. The sheer size of newly usable national cohorts (on the scale of 100,000–500,000 individuals) has further enabled the detection of effects from rare inherited variants present in a tiny fraction of the overall population.

Mosaic loss of Y. A substantial fraction of risk for mLOY appears to be genetically determined, with estimates of mLOY heritability ranging from 9% to 34%. The first germline association with mLOY to be uncovered was with a common SNP (rs2887399) near the 5’ end of TCL1A, which encodes the protein T cell leukemia/lymphoma 1A (TCL1A). The TCL1A protein is a co-activator of AKT and it participates in B and T cell malignancies, largely through chromosomal rearrangements that place TCL1A near TCR-A (the gene for the T cell antigen receptor). This strong association between rs2887399 and mLOY has been replicated in subsequent studies with larger cohorts; notably, single-cell RNA-sequencing of B lymphocytes has demonstrated that TCL1A gene expression is significantly higher in the setting of mLOY, suggesting that such clonal outgrowth in mLOY could be partly driven by supra-normal TCL1A expression. GWAS projects have found over 150 additional loci significantly associated with mLOY, many of which functionally regulate various aspects of the cell cycle, including the formation of mitotic structures (for example, SPDL1, CENPN and CENP, MAD1LI and MAD2LI, and PMF1), the replication and stability of DNA (for example, ATM and NPAT), and cell arrest and apoptosis (for example, TP53, BCL2 and BAX). The implicated genes highlight three complementary processes influencing mLOY: increasing rates of functional mistakes during mitosis, a lack of ability to detect such DNA abnormalities and escape from normal apoptotic regulation in the face of recognized DNA damage.

Autosomal and X chromosome variation. As with mLOY, autosomal and X chromosome mCAs are associated with germline variants that increase risk of mutagenesis. Unlike mLOY, which only involves the unpaired Y chromosome, these mCAs may also be associated with variants that provide a strong selection pressure towards CN-LOH events. Studies conducted in population-scale biobanks in the UK (UK Biobank (UKB)) and Japan (BioBank Japan (BBJ)) have demonstrated significant germline associations with mCAs. These associations occur both in cis and in trans with the inherited variant. In both populations, the trans associations involve common alleles with modest odds ratios. Variants in TERT and the related TERC...
(encoding telomerase RNA component\(^{(11)}\)) as well as variants in \(\text{SPI}40\) (encoding a lymphoid-restricted nuclear body protein involved in B cell antigen response\(^{(12)}\)) are associated with mCAs occurring anywhere in the genome\(^{(1)}\), whereas the remaining inherited variants have only been associated with trans mCAs on a particular chromosome\(^{(4,7,26)}\) (Table 1). Apart from common variation in \(\text{TCL}1\text{A}\) and \(\text{DLK}1\) (a negative regulator of HSPC differentiation\(^{(13)}\)) that is linked to 14q CN-LOH\(^{(17)}\) and a known association between the \(\text{JAK}2\) 46/1 haplotype and 9p CN-LOH\(^{(17,22-24)}\), the identified cis mCA associations are predominantly rare variants. Many of these rare germline variants are missense or nonsense mutations predicted to damage protein function. The cis mCA lesions associated with these disruptive variants demonstrate a strong preferential CN-LOH duplication of either the risk or the non-risk allele. In \(\text{ATM}\), \(\text{NBN}\) and \(\text{MRE}11\), all of which are genes involved in maintaining genomic integrity, it is their damaged germline allele that is more commonly propagated\(^{(17)}\). Conversely, the presence of damaging germline variants in the \(\text{MPL}\) gene, which encodes the thrombopoietin receptor important for HSC self-renewal, are associated with the duplication of the non-damaged allele\(^{(17)}\). Preferential CN-LOH duplication arising from germline alleles that confer a relative fitness advantage may also extend to polygenic risk. When the group studying the UKB cohort constructed blood cell-proliferation polygenic risk scores consisting of signals within individual chromosomal arms, they found that these are often associated with CN-LOH events on the same arm\(^{(17)}\). This finding raises the possibility that a main driver of these common CN-LOH somatic events is the replacement of inherited DNA segments with homologous segments that impart a greater fitness advantage\(^{(17)}\).

The specific inherited variants associated with mCAs and the spectrum of mCAs themselves may differ significantly across populations. Several of the rare variants associated with cis mCAs in the UKB cohort (\(\text{ATM}\), \(\text{MPL}\), \(\text{FRA}10\text{B}\) and \(\text{TM}2\text{D}3–\text{TARS}L2\)) were absent in the BBJ cohort, whereas variants in several other genes (\(\text{MRE}11\), \(\text{NBN}\), \(\text{NEDD}8\	ext{TINF}2\) and \(\text{CTU}2\)) were present at higher frequencies\(^{(17)}\). These population-specific differences may shape not only the relative frequencies of observed mCAs but also patterns of downstream disease. For example, the incidences of chromosome 12 gain, 13q loss and 13q CN-LOH are between twofold to sixfold less in the BBJ cohort\(^{(17)}\); these mCAs are often seen in chronic lymphocytic leukaemia\(^{(17,26,27)}\), a malignancy that is four to five times more common among Europeans than among Japanese individuals\(^{(17)}\). Collectively, these studies highlight the importance of including diverse populations in genomics research\(^{(17)}\).

### Small variants: SNPs and indels.
Several recent large genomic studies have focused on CH identified with WGS data, using short-read sequencing to simultaneously identify germline and somatic SNPs and indels\(^{(13,14,27)}\). Mirroring one of the main signals found with \(\text{JAK}2\), one study using the deCODE cohort from Iceland found that variation in the \(\text{TERT}\) locus (lead SNP rs34002450) was associated with CH (OR = 1.37, minor allele frequency (MAF) = 0.41) as defined by an outlier status on WGS\(^{(15)}\). Meanwhile, in the same study, individuals with CH were found to have a shorter average telomere length than individuals without CH\(^{(15)}\).

An analysis of the NHLBI Trans-Omics for Precision Medicine (TOPMed)\(^{(16)}\) cohort in the USA recapitulated the association between rs34002450 and CHIP (OR = 1.3), although this study identified a different lead SNP (rs7705526; MAF = 0.29; \(r^2 = 0.55\) with rs34002450) as well as a second SNP in \(\text{TERT}\) that was independently associated with CHIP (rs13167280; OR = 1.3; MAF = 0.11; \(r^2 = 0.2\) with rs7705526)\(^{(27)}\). Additionally, an analysis of the UKB similarly identified associations with CHIP for an SNP in linkage disequilibrium with rs34002450 (rs7726159; OR = 1.33; MAF = 0.33; \(r^2 = 0.70\) with rs34002450) and for a second independent SNP in \(\text{TERT}\) (rs2853677; OR = 1.32; MAF = 0.42)\(^{(14)}\).
Within the TOPMed cohort, two additional SNPs achieved genome-wide significant associations with CHIP. One variant (rs1210060191) is quite common (risk allele frequency = 0.54) and lies in the intronic region of TRIM59 but has a relatively weaker association with CHIP (OR = 1.16) than the TERT SNPs. The second is a variant in an intergenic region near TET2 (rs144418061), which is specific to individuals with African ancestry (MAF = 0.035 in African ancestry, not present in samples without African ancestry), that is strongly associated with CHIP (OR = 2.4). A subsequent variant-to-function analysis of this second locus revealed a variant (rs79901204) that is predicted to disrupt a GATA/E-box in an enhancer element. The risk allele for this variant indeed reduced luciferase activation fourfold in an in vitro experiment and had a dose-dependent association with decreased TET2 gene expression in whole-blood samples from patients. Thus, it appears this variant increases the self-renewal and proliferation capacity of haematopoietic stem cells via reduced TET2 expression, which might create a selective pressure for CHIP clone expansion in one of several ways. Increased rates of cell division may increase DNA replication strain and increase the likelihood of acquiring a lesion in a CHIP gene in the first place and/or the germline TERT variant might have a synergistic cooperativity with any subsequent incidental CHIP mutations to increase the relative fitness of the HSC.

The TOPMed study was also powered to investigate germline associations specifically in DNMT3A-CH and TET2-CH. Although there were no significant associations with TET2, there was a significant association for DNMT3A with variant rs2887399 (OR = 1.23; MAF = 0.23). Of note, this is the same variant near TCL1A that is associated with mLOY11,14. Each of these germline SNPs associated with CHIP are noted in Fig. 3.

Overlap with processes of ageing

Inherited and somatic variation at many CH risk loci have also been linked to diseases and processes of ageing (Table 2). Here, we specifically focus on the overlap of CH germline risk loci with germline variants associated with malignancy, CVD and several biomarkers of ageing. We also consider the challenges in distinguishing whether such overlap is independent or partially mediated by the presence of CH itself.

**Fig. 3 | CHIP has polygenic risk.** Genetic association studies have demonstrated that the inherited risk landscape for clonal haematopoiesis of indeterminate potential (CHIP) is characterized by numerous common variants with modest effect sizes and several rare variants associated with strong effects. Associations with the TERT locus have been replicated among numerous studies of individuals with CHIP as well as clonal haematopoiesis (CH) identified by high somatic mutational burden in whole-genome sequencing (WGS-outlier). Furthermore, CH by WGS-outlier is strongly correlated with CHIP and mosaic loss of the Y chromosome, highlighting a robust association between telomere biology and CHIP. Where studies have examined single genes affected by a somatic CHIP mutation, the results point to heterogeneity in their germline associations, both in terms of associated variants (for example, germline TCL1A variation is associated with DNMT3A-CH but not JAK2-CH) and the degree of association (for example, a stronger association of germline TERT variants with JAK2-CH than with CH overall).
Overlap with malignancy. In addition to the associations with MPNs described above, many of the inherited risk variants for CH also predispose to haematological and non-haematological cancers. This is particularly true of the genes involved in the DNA damage response: CHEK2, TP53, NBN, MRE11 and ATM. Inherited putative loss-of-function variants in CHEK2 (REFS 120–124) and TP53 (REFS 125–127) have long been known to be a cause of autosomal dominant familial cancer syndromes, while mutations in NBN (causing the autosomal recessive Nijmegen Breakage Syndrome)128,129 and MRE11 (REFS 130,131) confer an increased susceptibility to the development of a malignancy. Similarly, mutations in ATM, the aetiological agent of the autosomal recessive ataxia telangiectasia syndrome132, are associated with an increased risk of numerous types of cancer, including leukaemia and lymphoma133,134, breast cancer135,136, and prostate cancer137,138, among many others. Germline mutations in NPAT (nuclear protein, ataxia telangiectasia locus), whose gene product has been implicated in the transcriptional regulation of histone genes as well as ATM139, has been reported as a risk factor for

| Inherited risk locus candidate gene | Locus associated with CHIP? | Locus associated with mCAs? | Locus associated with mLOY? | Associations between germline variation in gene and diseases of ageing | Associations between somatic gene changes and diseases of ageing |
|-----------------------------------|----------------------------|-----------------------------|-----------------------------|-------------------------------------------------|-------------------------------------------------|
| TERT                              | Yes                        | Yes                         | Yes                         | Cancer164; CVD165–167; dementia168                | Cancer179; CVD180; dementia181,182               |
| CHEK2                             | Yes                        | Yes                         | Yes                         | Cancer120–124                                    | Cancer183                                      |
| ATM                               | Yes                        | Yes                         | Yes                         | Cancer135–138                                    | Cancer184                                      |
| TCL1A                             | Yes                        | Yes                         | No                          | NA                                               | Cancer185                                      |
| 46/1 haplotype                    | Yes                        | Yes                         | No                          | Cancer12–44                                      | NA                                              |
| SH2B3                             | Yes                        | Yes                         | No                          | Cancer144; CVD159–164; CVD165–167                | Cancer184                                      |
| TET2                              | Yes                        | No                          | Yes                         | Cancer161–164; CVD165–167                        | Cancer184                                      |
| MAD1L1                            | No                         | Yes                         | Yes                         | NA                                               | Cancer187                                      |
| RUNX1                             | No                         | Yes                         | No                          | Cancer103,104                                    | Cancer188                                      |
| KPNA4–TRIM59                      | Yes                        | No                          | No                          | NA                                               | KPNA4: cancer189; TRIM59: cancer190             |
| GFI1B                             | Yes                        | No                          | No                          | NA                                               | Cancer191                                      |
| MPL                               | No                         | Yes                         | No                          | Cancer192                                      | Cancer193                                      |
| TM2D3                             | No                         | Yes                         | No                          | Dementia194                                     | NA                                              |
| FN                                | No                         | Yes                         | No                          | NA                                               | Cancer195                                      |
| NEDD8                             | No                         | Yes                         | No                          | NA                                               | Cancer196                                      |
| TINF2                             | No                         | Yes                         | No                          | Cancer107                                       | Cancer196                                      |
| NBN                               | No                         | Yes                         | No                          | Cancer128,129                                   | NA                                              |
| MRE11                             | No                         | Yes                         | No                          | Cancer103,111                                   | Cancer196                                      |
| SP140                             | No                         | Yes                         | No                          | NA                                               | Cancer199                                      |
| HLA                               | No                         | Yes                         | Yes                         | Dementia190–192                                  | NA                                              |
| TERC                              | No                         | Yes                         | No                          | Cancer165; CVD167; dementia168                   | Cancer179; CVD180; dementia181                  |
| TP53                              | No                         | No                          | Yes                         | Cancer125–127                                   | Cancer201                                      |
| BCL2                              | No                         | No                          | Yes                         | NA                                               | Cancer204                                      |
| BAX                               | No                         | No                          | Yes                         | Cancer105                                       | Cancer206                                      |
| NPAT                              | No                         | No                          | Yes                         | Cancer140                                       | Cancer207                                      |
| CENPU                             | No                         | No                          | Yes                         | NA                                               | Cancer208                                      |
| MAD2L1                            | No                         | Yes                         | No                          | NA                                               | Cancer209                                      |

CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; CVD, cardiovascular disease; mCAs, mosaic chromosomal alterations; mLOY, mosaic loss of the Y chromosome; NA, not applicable.
Hodgkin lymphoma. The most plausible mechanism of action for the contribution of these inherited variants is similar to their role in cancer — establishing a cellular context that is permissive for DNA mutation — rather than the direct effects on clonal proliferation. By contrast, other inherited variants may directly influence proliferation or augment the rapidity of proliferation by later CH mutations. Mutation or experimental deletion of TET2, which is often mutated in familial myeloid and lymphoid malignancies, leads to increased HSC proliferation and secretion of pro-inflammatory cytokines, are associated with malignancies, including in the blood, breast, lung and colon.

Overlap with CVD. CVD is a major source of morbidity in ageing. Early epidemiological and functional studies of CHIP identified strong links between CHIP mutations and CVD, raising the question of whether these entities exhibit shared germline predispositions. In addition to the risk of a haematological malignancy, germline TET2 mutations have also been associated with pulmonary arterial hypertension, which is a lethal vasculopathy. In contrast to the role of this epigenetic regulator in tumorigenesis, which is thought to rest on increased HSPC self-renewal, lineage skewing, and an increased tendency towards mutation, the contribution of mutant TET2 to pulmonary arterial hypertension may stem from overproduction of inflammatory cytokines (for example, IL-1β) in differentiated immune cells. Meanwhile, genetic variation in the gene SH2B3 has been linked to numerous aspects of cardiovascular dysfunction, including hypertension, aortic dissection, atherosclerosis and stroke. However, for at least one well-studied variant, there appears to be a trade-off between CVD and cancer risk: the C allele of rs3184504, which encodes SH2B3 p.R262W, is associated with a reduced risk of CVD (OR = 0.95) but a heightened risk of cancer (OR = 1.03). If such risk trade-offs persist more generally for SH2B3, this could limit the utility of targeting the gene itself for disease prevention, although future work may find distinct downstream effectors that could be targeted to limit either CVD risk or cancer risk.

Overlap with biomarkers of ageing. The links between telomere biology and CH are robust but complicated. Across tissues, telomere length is inversely correlated with ageing. Inherited genetic variation influences telomere length, which is also tightly linked to the somatic expression of telomerase genes. The risk variants associated with CH have substantial overlap with multiple portions of the cellular machinery responsible for telomere maintenance: TERT has been implicated in the risk for all CH subtypes, while TERC and TINFB (encoding the TIN2 protein, part of the shelterin complex) are associated with mCAs. However, even though CH is strongly associated with ageing, the germline variation in telomere genes that predisposes to CH tends to associate with longer telomeres not shorter. The TERT intron 2 SNPs rs7705526 (REF.149) (the lead variant for increased risk for CHIP and global mCA events) and rs2853677 (REF.150) (associated with 14q CN-LOH) associate with longer telomeres and greater telomere length, as predicted by germline variation, is positively associated with mCA events. Recent work using Mendelian randomization has suggested that the telomere–CH relationship is actually bidirectional: longer telomeres may partly cause CHIP (perhaps through an increased propensity for mutation) whereas CHIP, once acquired, may contribute to telomere shortening (possibly via increased rates of cell cycling). Importantly, genetic association studies of telomere maintenance genes have revealed links to a broad spectrum of diseases, including strong ties to cancer and CVD.

Future work investigating links between telomere length and these diseases (or CH and these diseases) may need to account for mediating effects through the telomere–CH axis.

DNA methylation at CpG sites is a promising biomarker that has been used to generate highly accurate estimates of chronological age, such as with the Horvath epigenetic clock. Many diseases that disproportionately affect the elderly, such as cancer and dementia, are linked to accelerated epigenetic ageing, in which an individual’s DNA methylation profile suggests an older chronological age than is true. Likewise, individuals with CHIP (also an age-associated feature) have epigenetic age acceleration in blood cells. The rate of epigenetic ageing has a heritable component, including variation at loci associated with CH risk: TERT, TET2, TRIM59 and KPNA4. Paradoxically, faster epigenetic ageing is linked to TERT variants associated with longer telomeres, matching the directionality of the CH risk variants at this locus. It is also worth noting that several of the genes that are most often affected by somatic CHIP mutations are epigenetic regulators whose (impaired) performance could plausibly shape an individual’s rate of epigenetic ageing. The top CHIP genes DNM3A and TET2 directly modulate CpG methylation and dictate global methylation patterns within HSPCs. Less common CHIP mutations in IDH1 and IDH2 lead to the production of the metabolite 2-hydroxylglutarate, which interferes with the function of TET2. Additionally, interestingly, many of the CpG sites used in the Horvath epigenetic clock are near target genes of Polycomb repressive complex 2 (PRC2), a protein complex whose function is impaired by CHIP mutations in ASXL1. Yet, the extent to which CHIP, or CH more broadly, might cause alterations in epigenetic ageing remains to be determined.

Determining causality: MR approaches. As described above, CH is associated with many diseases of ageing, which naturally begets the question: does CH contribute to these phenotypes? Although potential CH–phenotype relationships will be studied by conducting future natural or laboratory experiments like those that have demonstrated ties between CH and haematological malignancies or heart disease, there is a wealth of already generated genetic and phenotypic data that may provide insights on a shorter horizon.
Mendelian randomization (MR) is a statistical technique that utilizes inherited variation to test causation between an exposure (here, CH) and outcome (disease of interest)\textsuperscript{175}. MR relies on a quasi-experimental setup in which individuals have a higher or lower probability of experiencing the exposure based on the alleles they were randomly assigned at birth. Using this random germline-determined variation in exposure allows for the estimation of a causal relationship between the exposure and outcome (Fig. 4a). MR analysis has already been used to demonstrate that a higher risk of prostate, testicular, breast, glioma cell and renal cell cancers is predicted by the inherited risk for mLOY\textsuperscript{174}. However, attempts to characterize the contribution of CH to diseases of ageing via MR approaches are likely to face several hurdles. The first is that, as described here, many germline variants associated with CH have also been previously associated with the diseases of ageing, leading to issues of horizontal pleiotropy that confound the estimation of a causal effect by MR\textsuperscript{175} (Fig. 4b). The risk of this can be minimized but not eliminated by only using variants with no described relation to the disease being studied. A second challenge is that data on CH is obtained through the sequencing of blood cells, which is rarely done in routine clinical practice and, when performed for research studies, is often done only once. As a result, data on CH is often a cross-sectional snapshot lacking information on the evolution and temporal duration of a clone. This will present difficulties for effect size estimation but may be ameliorated by future longitudinal studies and as sequencing costs drop this test becomes more widely deployed in clinical settings. These potential methodological challenges aside, we anticipate that, with the increased power derived from larger CH GWAS sample sizes, MR will become an increasingly useful tool in answering questions about the health consequences of CH.

**Conclusions and perspectives**

How germline genetics contributes to CH risk is an emerging field with a rapidly growing body of work. By simultaneously analysing germline and somatic genetic variation on a population scale, research in this area in just the past 5 years has made dramatic contributions to our understanding of HSC biology and disease risk.

To date, the patterns of germline susceptibility to mLOY, mCAs and CHIP have largely been studied in isolation from one another. However, the comparison of the inherited risk landscape for each of these phenomena reveals that these entities share many genetic signals (Fig. 2). In particular, the DNA damage response and telomere maintenance pathway genes are commonly implicated in genetic association studies with these CH subtypes. The substantial overlap in germline risk suggests that there may be common mechanisms that predispose individuals to mLOY, mCAs and CHIP. Therefore, there is likely to be a benefit to studying these phenomena jointly. Additionally, the existence of shared risk loci raises the important question of what additional factors may influence the likelihood that an HSC will acquire one type of CH over another (BOX 2). It also remains to be fully explored whether and to what degree inherited variants contribute to the co-occurrence (or co-interaction, if one somatic change influences the next) of acquired CH mutations of different varieties, especially CHIP mutations and focal deletions or loss-of-heterozygosity events. These questions are important for our understanding of how HSCs adapt to the
Box 2 | Open questions in understanding inherent risk of CH

- Why does the same inherent variant predispose to multiple subtypes of clonal haematopoiesis (CH)? To what extent is the relationship between inherent genotype and subsequent CH phenotype determined by genetic interactions, environmental exposures and random chance?
- Is the germline risk of CH due to the result of accumulated constitutive effects across the entire lifespan of haematopoietic stem cells or is it the product of heightened probability of dysfunction of aged haematopoietic stem cells?
- To what extent do inherited variants impart sex-specific risk for CH?
- Do inherited variants affect downstream CH consequences?
- What mechanisms underlie the strong association between telomere-regulating genes and CH? Are these associations entirely a function of these genes’ canonical role in maintaining telomere length or is there a contribution from non-canonical activity?

stresses of ageing and to improve our ability to assess the risk of disease for individuals carrying predisposing germline variants.

Moving forward, we expect studies in this field to focus not just on how inherited variation influences the risk of somatic mutations but also on how inherited variation interacts with these acquired mutations to influence disease phenotypes and biological ageing. For example, although CHIP mutations are associated with an increased risk of leukaemia and myocardial infarction, these outcomes are observed only in a minority of CHIP carriers. Recent work has identified an inherited polymorphism in the IL-6 receptor that reduces the likelihood of heart disease in individuals with CHIP; however, the full extent to which germline factors mitigate or contribute to disease manifestations in individuals with CHIP is still to be explored.

As we understand more about how inherited germ-line genetic variation interacts with CH, there will be increasing motivation to develop and deploy precision medicine applications that incorporate knowledge of the germline genome to precisely estimate the risk for CH and for developing associated disease sequelae. Given that most individuals with CH do not display overt symptoms of the condition, in time, these approaches may enable more precise CH screening regimens. In the more immediate future, the recent creation of specialty CH clinics, well suited to capturing CH carriers in populations whose at-risk status warrants more extensive screenings (such as cancer patients), may afford opportunities for the rapid translation of new research insights in this space into impactful patient care.

Published online 13 May 2021

1. Orkin, S. H. & Zon, L. I. Haematopoiesis: an evolving paradigm for stem cell biology. Cell 132, 631–644 (2008).
2. Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. Nature 561, 473–478 (2018).
3. Beerman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc. Natl Acad. Sci. USA 107, 5465–5470 (2010).
4. Welch, J. S. et al. The origin and evolution of mutations in early human blood inferred from somatic mutations. Cell 150, 264–278 (2012).
5. Jan, M., Ebert, B. L. & Jaiswal, S. Clonal haematopoiesis. Semin. Hematol. 54, 43–50 (2017).
6. Busque, L. et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood 88, 59–65 (1996).
7. This paper, showing age-associated chromosome X inactivation skew, provided the initial evidence that was suggestive of clonality in the blood of individuals without haematological disease.
8. Jaiswal, S. et al. Age-related clonal haematopoiesis associated with adverse outcomes. N. Engl. J. Med. 371, 2488–2498 (2014).
9. This work was the first to find associations between non-cancer mortality and CHIP.
10. Genovese, G. et al. Clonal haematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. 371, 2477–2487 (2014).
11. This study delineated CH from unknown drivers and the strong association between CHIP and leukaemia.
12. Loh, P.-R. et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. Nature 559, 350–355 (2018).
13. Zink, F. et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood 130, 762–752 (2017).
14. Thompson, D. J. et al. Genetic predisposition to mosaic chromosome Y loss in blood. Nature 572, 652–657 (2019).
15. This study on the genetic determinants of mLOY is the largest to date and used multiple modalities to validate associations.
16. Kakuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. Nat. Rev. Cancer 21, 259–256 (2021).
17. This recent review presents the current knowledge regarding clonality throughout the human body and is a good entry point for those seeking to learn more about the field.
18. Wright, D. J. et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. Nat. Genet. 49, 674–679 (2017).
19. Zhou, W. et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. Nat. Genet. 48, 565–568 (2016).
20. Lofffeld, E. et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. Sci. Rep. 8, 12516 (2018).
21. Dumanski, J. P. et al. Smoking is associated with mosaic loss of chromosome Y. Science 347, 81–85 (2015).
22. Loh, P.-R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. Nature 586, 136–141 (2020).
23. This large study of mCA genetics found that both monogenic and polygenic risk variants can predict mCA events.
24. Dawoud, A. A. Z., Tapper, W. J. J. & Cross, N. C. P. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. Leukemia 34, 2660–2672 (2020).
25. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbours AML-associated mutations in ubiquitous in healthy adults. Nat. Commun. 7, 12484 (2016).
26. This paper measuring somatic mutations down to very low allele burden found acquired variants in the blood of 95% (19/20) of sampled adults.
27. Watson, C. J. et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. Science 367, 1449–1454 (2020).
28. Forsberg, L. A. et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat. Genet. 46, 624–628 (2014).
29. Zhou, W. et al. Detectable chromosome X mosaicism in males is rarely tolerated in peripheral leukocytes. Sci. Rep. 11, 1153 (2021).
30. Jacob, K. B. et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat. Genet. 44, 651–658 (2012).
31. Machiela, M. J. et al. Characterization of large structural genetic mosaicism in human automeses. J. Hum. Genet. 103, 421–426 (2018).
32. Laurie, C. C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat. Genet. 44, 642–650 (2012).
33. Terao, C. et al. Chromosomal alterations among age-related hematopoietic malignancies in Japan. Nature 584, 130–135 (2020).
34. Bick, A. G. et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature 586, 763–768 (2020).
35. This study is the largest GWAS of CHIP done to date.
36. Perner, F., Perner, C., Ernst, T. & Heidel, H. F. Roles of JAK2 in aging, inflammation, hematopoiesis and malignant transformation. Cells 8, 854 (2019).
37. Kraalivis, R. et al. A pair-of-function mutation of JAK2 in myeloproliferative disorders. N. Engl. J. Med. 352, 1779–1790 (2005).
38. Gao, T. et al. Interplay between chromosomal alterations and gene mutations shapes the evolutionary trajectory of clonal hematopoiesis. Nat. Commun. 12, 538 (2021).
39. This study was the first to examine the combined effect of CHIP and mCAs on patient outcomes, finding increased mortality in individuals with both types of lesions.
40. McKerrill, T. et al. Leukaemia-associated somatic mutations drive distinct patterns of age-related clonal hematopoiesis. Cell Rep. 19, 1295–1245 (2015).
41. Cooimbs, C. C. et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. Cell Stem Cell 21, 374–382 e4 (2017).
42. Lofffeld, E. et al. Mosaic Y loss is moderately associated with solid tumor risk. Cancer Res. 79, 461 (2019).
43. Lin, S.-H. et al. Mosaic chromosome Y loss is associated with alterations in blood cell counts in UK Biobank men. Sci. Rep. 10, 3655 (2020).
44. Bolton, K. L. et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Cell 178, 1219–1226 (2020).
45. This paper offers a detailed look at how radiotherapy, chemotherapy, targeted agents and immune therapies are associated with CHIP.
46. Boucai, L. et al. Radioactive iodine-related clonal hematopoiesis in thyroid cancer is common and associated with decreased survival. J. Clin. Endocrinol. Metab. 105, 6216–6221 (2020).
47. Pich, O. et al. The evolution of hematopoietic cells under cancer therapy. bioRxiv https://doi.org/10.1101/2020.10.30.360230 (2020).
58. Gardiner, C. et al. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. Genes Chromosomes Cancer 54, 717–724 (2015).

59. Naveska, P. et al. Loss of Y chromosome in peripheral blood of colorectal cancer patients. PLoS ONE 11, e0142664 (2016).

60. Machiela, M. J. et al. Mosaic chromosome Y loss and testicular cancer risk. J. Hum. Genet. 62, 637–640 (2017).

61. Haijema, S. et al. Loss of Y chromosome in blood is associated with colorectal cancer in Barrett’s esophagus and Barrett’s associated adenocarcinoma. Cancer Epidemiol. Biomarkers Prev. 26, 170–175 (2017).

62. Dumanpasa, J. et al. Loss of chromosome Y in blood is associated with late-onset Alzheimer disease. Am. J. Hum. Genet. 98, 1208–1219 (2016).

63. Hirata, T. et al. Investigation of chromosome Y loss in male patients with autoimmune thyroid disease. J. Neuropsychiatr. Dis. Treat. 14, 2115–2122 (2018).

64. Persani, L. et al. Increased loss of the Y chromosome in peripheral blood cells in male patients with autoimmune thyroiditis. J. Autoimmun. 38, J193–J196 (2012).

65. Liu, A. et al. Y chromosome loss in male patients with primary biliary cirrhosis. J. Autoimmun. 41, 87–91 (2013).

66. Grassmann, F. et al. Y chromosome mosaicism is associated with age-related macular degeneration. Eur. J. Hum. Genet. 27, 36–41 (2019).

67. Zekavat, S. M. et al. Hematopoietic mosaicism and risk for infection among 767,891 individuals without blood cancer. medRxiv. https://doi.org/10.1101/2020.11.25.20233163 (2020).

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

69. Desai, P. et al. Mosaic mutations precede acute myeloid leukemia years before diagnosis. Nat. Med. 24, 1015–1023 (2018).

70. Jaiswal, S. et al. De novo haploinsufficiency. J. Am. Coll. Cardiol. 71, 875–886 (2018).

71. Busque, L. et al. High-sensitivity C-reactive protein is associated with clonal hematopoiesis of indeterminate potential to predict cardiovascular events in twins. Blood 141, 124–131 (2020).

72. Fuster, J. J. et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. Science 355, 842–847 (2017).

73. Wang, W. et al. Macrophage infiltration, erythropoiesis dysregulation, and accelerated atherosclerosis in JAK2V617F mice. Circ. Res. 113, e55–e67 (2018).

74. Sano, S. et al. Tet2-mediated clonal hematopoiesis accelerates heart failure through a mechanism involving the IL-1β/IL-6/NF-κB axis. J. Am. Coll. Cardiol. 71, 875–886 (2018).

75. Sano, S. et al. JAK2V617F-meditated clonal hematopoiesis pathophysiologic remodeling in murine heart failure. JACC. 68, 684–697 (2019).

76. Lyko, F. The DNA methyltransferase family: a versatile tool in epigenetic regulation. Nat. Rev. Genet. 19, 81–92 (2018).

77. Zhang, Q. et al. Tet2 is required to resolve infarct remodeling by regulating Hdc2 to specifically repress IL-6. Nature 525, 389–395 (2019).

78. Poisson, J. et al. Erythrocyte-derived microvesicles induce arterial spasms in JAK2V617F myeloproliferative neoplasms. Proc. Natl. Acad. Sci. U.S.A. 108, 329–334 (2011).

79. Polacek, J. et al. JAK2V617F mutant receptor signalling compared to those without JAK2 mutations. J. Med. Microbiol. 65, 186–194 (2016).

80. Nabavi, S. M. et al. JAK2V617F clonal hematopoiesis is associated with increased arterial stiffness and reduced vagal parasympathetic activity. Circ. Res. 127, 1271–1279 (2020).

81. Shih, H. et al. JAK2V617F mutation drives clonal hematopoiesis in response to cytotoxic chemotherapy. Blood 131, 629–637 (2018).

82. Timcik, T. et al. TET2 loss leads to increased risk of myeloid neoplasms – a case-control study. Lancet Oncol. 18, 100–111 (2017).

83. Lindsey, R. C. et al. Prognostic mutations in myelodysplastic syndrome after stem cell transplantation. N. Engl. J. Med. 376, 536–547 (2017).

84. Khan, J. D. et al. PMID1 truncating mutations confer resistance to chemotherapy and sensitivity to PMID1 inhibition in hematopoietic cells. Blood 132, 1095–1105 (2018).

85. Hitosugi, T. et al. Clonal hematopoiesis in patients with degenerative aortic valve stenosis undergoing transcatheter aortic valve implantation. Eur. Heart J. 41, 935–940 (2020).

86. Nawas, M. T. et al. The clinical implications of clonal hematopoiesis in hematopoietic cell transplantation. Blood 135, 299–309 (2020).

87. Wilk, C. M., Manz, M. G. & Boettcher, S. Clonal hematopoiesis in hematopoietic stem cell transplantation. Curr. Opin. Hematol. 28, 94–100 (2021).

88. Steensma, D. P. et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood 126, 9–16 (2015).

89. Sperling, A. S., Gibson, C. J. & Ebert, B. L. The genetics of myelodysplastic syndrome: from clonal haemopoiesis to clonal leukaemia. Nat. Rev. Cancer 17, 5–19 (2017).

90. Bao, E. L. et al. Inherited myeloproliferative neoplasm risk affects hematopoietic stem cells. Nature 586, 769–775 (2020).

91. Hinds, D. A. et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood Adv. 4, 1148–1149 (2020).

92. Mangankar, A. A. et al. Hereditary predisposition to hematopoietic neoplasms: when bloodline matters for blood cancers. Mayo Clin. Proc. 95, 1181–1189 (2020).

93. Fabre, M. A. & Patnaik, M. M. Concordance for clonal hematopoiesis is limited in elderly twins. Blood 135, 269–273 (2020).

94. Hansen, J. W. et al. Clonal hematopoiesis in elderly twins: concordance, discordance, and mortality. Blood 135, 261–268 (2020).
189. Hazawa, M. et al. Disease-specific alteration of karyopherin-α subtype establishes feed-forward oncogenic signaling in head and neck squamous cell carcinoma. Oncogene 39, 2212–2223 (2020).

190. Wang, M. et al. Prognostic significance of TRIM59 for cancer patient survival: A systematic review and meta-analysis. Medicine 98, e18024 (2019).

191. Anithitha, T. et al. G61b: a key player in the genesis and maintenance of acute myeloid leukemia and myelodysplastic syndrome. Haematologica 103, 614–625 (2018).

192. Plo, I. et al. Genetic alterations of the thrombopoietin/MPL/JAK2 axis impacting megakaryopoiesis. Front. Endocrinol. 8, 234 (2017).

193. Vainchenker, W. & Kralovics, R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood 129, 667–679 (2017).

194. Jakobsdottir, J. et al. Rare functional variant in TM2D5 is associated with late-onset Alzheimer’s disease. PLoS Genet. 12, e1006327 (2016).

195. Rick, J. W. et al. Fibronectin in malignancy: cancer-specific alterations, protumoral effects, and therapeutic implications. Semin. Oncol. 46, 284–290 (2019).

196. Watson, I. R., Irwin, M. S. & Ohh, M. NEDD8 pathways in cancer, senequitas non. Cancer Cell 19, 168–176 (2011).

197. He, H. et al. A truncating germline mutation of TINF2 in individuals with thyroid cancer or melanoma results in longer telomeres. Thyroid 30, 204–213 (2020).

198. Rahman, S., Canny, M. D., Buschmann, T . A. & Perucchini, M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. Proc. Natl Acad. Sci. USA 97, 10872–10877 (2000).

199. Bolti, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat. Commun. 5, 2997 (2014).

200. Huang, P. et al. Association of early-onset Alzheimer’s disease with germline-generated high affinity self-antigen load. Transl. Psychiatry 10, 146 (2020).

201. Wang, Z.-X. et al. Genetic association of HLA gene variants with MRI brain structure in Alzheimer’s disease. Mol. Neurobiol. 54, 3195–3204 (2017).

202. Steele, N. Z. R. et al. Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: a case–control study. PLoS Med. 14, e1002272 (2017).

203. Bykov, V. J. N., Eriksson, S. E., Bianchi, J. & Wiman, K. G. Targeting mutant p53 for efficient cancer therapy. Nat. Rev. Cancer 18, 89–102 (2018).

204. Singh, R., Letai, A. & Sarosiek, K. Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. Nat. Rev. Mol. Cell Biol. 20, 175–193 (2019).

205. Silva, A. G. et al. Germline BAX deletion in a patient with melanoma and gastrointestinal stromal tumor. Am. J. Gastroenterol. 108, 1572–1575 (2015).

206. Ionov, Y., Yamamoto, H., Krajewski, S., Reed, J. C. & Perucho, M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. Proc. Natl Acad. Sci. USA 97, 10872–10877 (2000).

207. Han, Y. et al. DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies. Nucleic Acids Res. 47, e45 (2019).

208. Wang, X. et al. Centromere protein U expression promotes non-small-cell lung cancer cell proliferation through FOXM1 and predicts poor survival. Cancer Manag. Res. 10, 6971–6984 (2018).

209. Ding, X., Duan, H. & Luo, H. Identification of core gene expression signature and key pathways in colorectal cancer. Front. Genet. 11, 45 (2020).

Acknowledgements

A.J.S. received financial support from the Ann Melly Summer Scholarship in Oncology and from the US National Institutes of Health (NIH) under Ruth L. Kirschstein National Research Service Award F30DK127699 from the US National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK) and grant T32GM007347 from the US National Institute of General Medical Sciences (NIGMS). A.G.B. is supported by a Burnworth Welcome Foundation career award for medical scientists and an NIH Director’s Early Independence Award from the National Institute of Health Common Fund (OD029586). M.R.S. is a Leukemia and Lymphoma Society Clinical Scholar and receives funding from the E.P. Evans Foundation, The Biff Ruttenberg Foundation, the Adventure Allie Fund, the Beverly and George Rawlings Directorship, and the NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

Each of the authors contributed to all aspects of the manuscript.

Competing interests

A.J.S. and A.G.B. declare no competing interests. M.R.S. receives research funding from ALX Oncology, Astex, Incyte, Takeda and TG Therapeutics, has equity with Karyopharm; and serves as an advisor or consultant to AbbVie, Astex, BMS, Geron, Incyte, Karyopharm, Ryvu, Sierra Oncology, Takeda, Taiho and TG Therapeutics.

Peer review information

Nature Reviews Genetics thanks G. Vassilouli and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher’s note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.