Germline risk of clonal haematopoiesis

Alexander J. Silver[1,2], *Alexander G. Bick*[1,3,4,5] *and Michael R. Savona*[1,2,4,5] ✉

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.

[1]*Program in Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN, USA.*

[2]*Division of Hematology and Oncology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA.*

[3]*Division of Genetic Medicine, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA.*

[4]*Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.*

[5]*Center for Immunobiology, Vanderbilt University School of Medicine, Nashville, TN, USA.*

✉*e-mail: michael.savona@vanderbilt.edu*

https://doi.org/10.1038/s41576-021-00356-6

Haematopoiesis, the process by which blood cells are generated, begins in embryogenesis and continues throughout an individual's lifespan[1]. 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[2] produce an estimated 10^{10}–10^{12} progeny blood cells every day[3]. 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 decade[4], 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 damage[5]. Over time, the relative fitness advantage of these mutated HSCs can result in the clonal production of a large number of progeny that all 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-1990s[6], 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

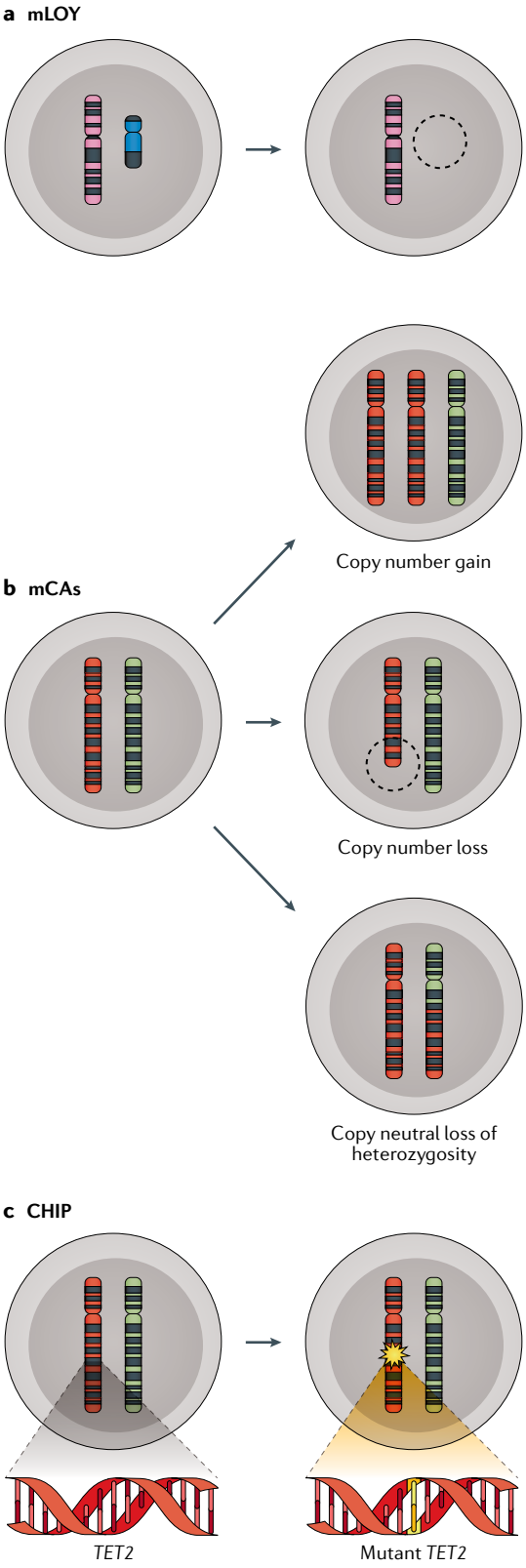


Fig. 1 | **Types of clonal haematopoiesis.** Clonal haematopoiesis refers to a clonal expansion of blood cells that are often identified based on shared genetic mutations. **a** | An especially common type of clonal haematopoiesis is the mosaic loss of the Y chromosome (mLOY), an entity that is often studied separately from other large chromosomal events. **b** | When large segments of one or more chromosomes are gained, lost or recombine resulting in the loss of heterozygosity, this may result in clonal haematopoiesis with mosaic chromosomal alterations (mCAs). **c** | Clonal haematopoiesis may also occur through mutations in myeloid-associated genes, termed clonal haematopoiesis of indeterminate potential (CHIP).

lymphoid organs and frank malignancy. CH is a common phenomenon among the general population and its prevalence increases significantly with age[7–11]. The blood is not unique in the accumulation of mutations with age, a trend which has also been observed in the solid organs[12], 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[11,13–16]. Additionally, mosaic chromosomal alterations (mCAs)[9,17,18], single-nucleotide variants (SNVs) and indels in genes associated with myeloid malignancies, and putative incidental mutations/genomic drift (CH with unknown drivers)[8,10] 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)[7,8]. 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[19,20] 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[11,13–16,21], with the prevalence increasing to >40% of individuals by age 70 in the largest epidemiological study to date[11]. 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[9], whereas men rarely have X chromosome mosaicism in the blood at any age[22]. Autosomal mCA events are the least commonly observed type of CH, affecting ~1–5% of the population older than 70 years of age[9,17,23–26], whereas CHIP is estimated to affect >10% of individuals older than 70 years of age[7,8,27]. An individual can have both mCAs and

Somatic
Variation in the genome arising from changes happening over the lifespan of an organism.

Clonal haematopoiesis
(CH). The presence of many blood cells originating from a single mutant stem or progenitor cell.

Germline
Variation in the genome arising from DNA that is passed from the parent to the offspring.

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 harboured a single copy of *JAK2*^{V617F}, 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[28]) and mCA events at the same locus[17,29]. 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[18,27], although the prevalence is higher among patients treated for solid tumours[30].

The prevalence of CH varies across several demographic features. There is a sex bias for specific mCA lesions, with most of these having greater prevalence in men[17,26]. Although some studies have suggested a male-bias for CHIP[7], other studies have found this association does not persist after controlling for potential confounders[27]. 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%)[15]. Meanwhile, CHIP mutations are less frequently observed in individuals identifying as Hispanic[7,27] or East Asian[27].

There are substantial differences across age in the distribution of mutated CHIP genes[27,31]. 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[27,31]. The extent to which this distribution is shaped by differences in DNA sequence mutability[20], relative fitness advantage[20] or interactions with an ageing microenvironment[31] 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[8,27,32] and mLOY[11,13–16,33,34]. In the case of cytotoxic chemotherapy and radiation therapy, the mutational spectrum exhibits a marked enrichment of mutations in DNA damage response pathway genes[32,35,36]. The outgrowth of CH clones following anticancer therapy is partly due to the expansion of pre-existing clones with a selective advantage[35] but may also be from the introduction of new mutations by the anticancer agents themselves[37] 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[15,21], numerous types of cancer[11,14,21,33,38–40], cardiovascular events[41], Alzheimer disease[42], schizophrenia[43], autoimmune disease[44,45], diabetes[15] and age-related macular degeneration[46]. Autosomal mCA events have been associated with an increased risk of haematological malignancies[17,18,26,30] as well as with all-cause mortality only partially explained by excess cancer deaths[9]. 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[30]. 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)[47]. 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[7,10]; this is partly due to an increased risk of haematological malignancies, which has been observed across many studies[7,8,10,48,49]. However, individuals with CHIP mutations have excess mortality compared with those who do not harbour such mutations even after controlling for blood cancer deaths[7,10]. 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[50–52] as well as to higher blood levels of the inflammatory marker C-reactive protein[53]. Mouse models of CHIP have demonstrated mechanistic ties between certain common CHIP mutations and accelerated atherosclerosis[50,54,55] as well as heart failure[56,57].

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*[48,58]. Somatic mutations in *TET2*, encoding a dioxygenase that opposes the action of DNMT3A by promoting DNA demethylation[58], and in *JAK2* are associated with coronary artery disease[50] 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[59], are associated with increased expression of *Il1b*, *Il6*, *Cxcl1*, *Cxcl2* and *Cxcl3* (REFS[50,54]). While mutations in *Jak2* also lead to higher *Il1b* expression, they additionally lead to plaque-promoting erythrophagocytosis[55], secretion of arterial spasm-inducing erythrocyte-derived microvesicles[60] and thrombotic neutrophil extracellular traps[61]. 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.

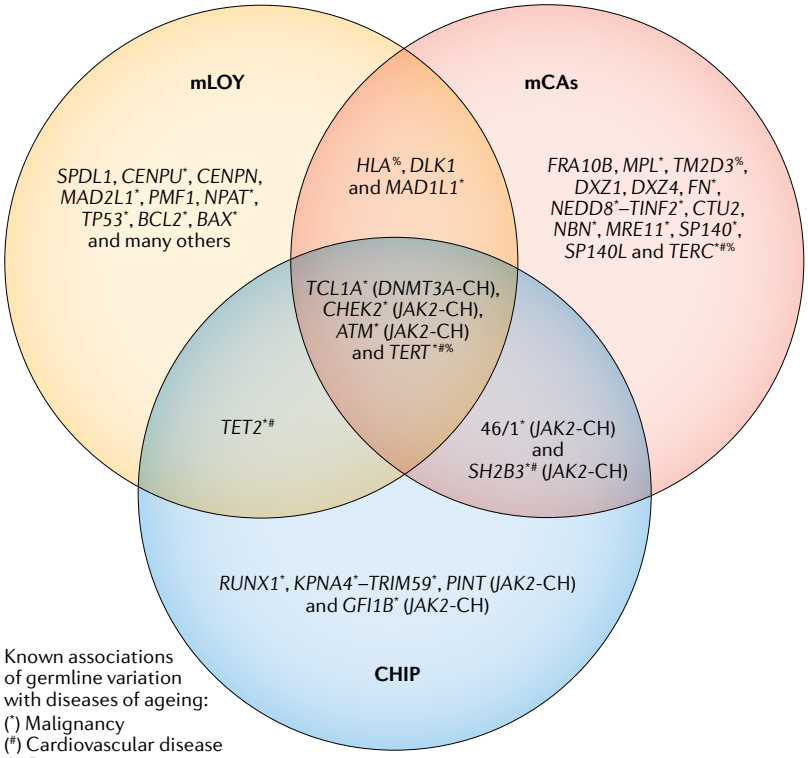


Fig. 2 | **CH subtypes have shared and unique risk variants.** Many germline risk loci have been linked to the development of clonal haematopoiesis (CH). The three subtypes of CH that have received the greatest scrutiny in this area are mosaic loss of the Y chromosome (mLOY), mosaic chromosomal alterations (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.

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[18,62], adult-onset haemophagocytic lymphohistiocytosis[63] and anti-neutrophil cytoplasmic antibody-associated vasculitis[64]. 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.[65]), perhaps as a result of CHIP-exacerbated inflammatory signalling[66]. 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[67,68]. 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[32,35,69–72]. Additionally, CHIP has been associated with significantly increased mortality following transcatheter aortic valve implantation[73], 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[74,75]) 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)[76–78]. 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.

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

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

Myeloproliferative neoplasms
(MPNs). Haematological malignancies generally defined by the overproduction of myeloid cells, red blood cells or platelets.

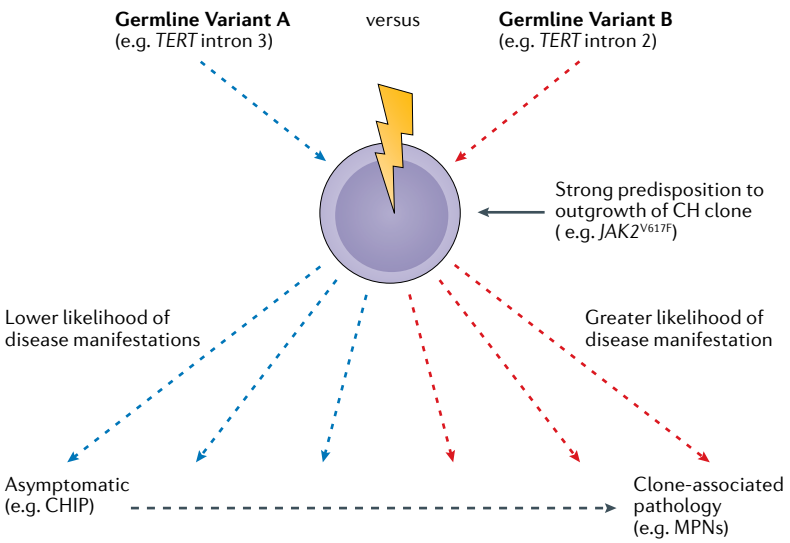
Myelodysplastic syndromes
(MDS). A set of haematological malignancies with abnormal appearing cells within the bone marrow and reduced production of mature blood cells without increased levels of immature blasts.

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[79]. *JAK2* is the most commonly mutated gene in MPNs[29] and the *JAK2* p.Val617Phe mutation (*JAK2*^{V617F}) is a characteristic feature of MPNs incorporated into the World Health Organization diagnostic criteria for over a decade[80,81]. Consequently, some studies looking to define MPN germline risk have used cohorts formed exclusively of diagnosed myeloid disease/MPNs[82–89], whereas other studies have augmented cohorts of diagnosed MPNs with the addition of any individuals with a molecularly detectable *JAK2*^{V617F} mutation[78,79] (which may include undiagnosed MPNs as well as *JAK2*^{V617F}-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 haematopoiesis (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, *JAK2*^{V617F}) (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 *JAK2*^{V617F} clones but only the germline variants in intron 2 have demonstrated a significant association with MPNs[18]. Meanwhile, inherited variants affecting *MECOM*[78,87,88], *HBS1L–MYB*[87,88], *RUNX1* (REF.[78]), *HMGA1* (REF.[78]), *FOXO1* (REF.[78]) and *GATA2* (REF.[78]) 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 *JAK2*^{V617F} 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.



CHIP, clonal haematopoiesis of indeterminate potential.

The first inherited variation linked to *JAK2*^{V617F}-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[82–88]. In several studies of patients with MPNs, the *JAK2*^{V617F} somatic variant was identified in *cis* with the inherited 46/1 risk haplotype more often than would be predicted by chance[82–84], which might suggest that the haplotype provides a hypermutable substrate for somatic alterations. Furthermore, this haplotype may increase the rate of *JAK2*^{V617F} 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[90], 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[27,78,87,89] and to *JAK2*^{V617F}-mutated disease in several others[79,86,88]. 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[91,92]. The precise mechanism of how telomere regulation might influence the expansion of *JAK2*^{V617F} 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[78,79,87]. Similar to *JAK2* (REF.[93]) and *TERT*[94], many of these are genes implicated in the functional regulation of HSCs (including *SH2B3* (REF.[95]), *TET2* (REF.[96]), *ATM*[97], *GFI1B*[98] and *RUNX1* (REF.[99]), 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 lower *GFI1B* expression, which, in turn, was shown to increase HSPC self-renewal[78]. The second case involves a rare missense variant in *CHEK2* and similarly demonstrated increased HSPC self-renewal following the knockdown of gene expression[78].

The genetic associations identified for *JAK2*-mutated malignancy and *JAK2*-CH are not completely overlapping. The examination of a *JAK2*^{V617F}-CH cohort replicated associations with the 46/1 haplotype, *TERT*, *SH2B3* and *TET2* with nominally significant signals for *CHEK2*, *ATM*, *PINT* and *GFI1B*[79]; additionally, *KPNA4* has been associated with a greater risk of all CHIP, inclusive of *JAK2*-CH[27]. 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 (BOX 1).

Compared to the literature on *JAK2*, studies of haematological malignancies have been less revealing with respect to what germline factors may increase the risk

Haematopoietic stem and progenitor cells
(HSPCs). A population of blood cells that includes multipotent HSCs but also more differentiated progenitor cells that are capable of producing many daughter cells but only within restricted lineages.

cis
mCAs in *cis* refers to an mCA lesion that affects the portion of the genome containing the risk allele.

trans
mCAs in *trans* refers to an mCA lesion affecting a different part of the genome than the DNA on which the risk allele is located.

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[100,101]. 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[102]. 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*[27]. 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 CH mutations using siblings looked only at the two most commonly mutated CHIP genes, *DNMT3A* and *TET2* (REF.[62]). The authors looked at the risk-recurrence ratio (λ_s) 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* ($\lambda_s = 2.24$ for those ≥55 years of age, $\lambda_s = 2.65$ for those ≥65 years of age)[62]. 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[62].

The heritability of CHIP has also been examined in twin pairs in two recently published studies[103,104]. One study consisted of 299 twin pairs from Denmark[104], whereas the other was comprised of 79 twin pairs from the UK[103]. Neither study found a higher concordance for the incidence of CHIP among monozygotic (MZ) twins than among dizygotic pairs. The larger of the two studies additionally found no increased 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[103] and *SRSF2* p.P95H and c.912_916delCTGGT in *DNMT3A* in the Denmark cohort[104]), suggesting these mutations occurred in utero[103]; several subsequent studies of patients with MPNs have identified *JAK2*^{V617F} and *DNMT3A* mutations that similarly arose during embryogenesis or childhood[105,106]. 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%[11,13,107]. 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 leukaemia/lymphoma 1A (TCL1A)[14]. The TCL1A protein is a co-activator of AKT and it participates in B and T cell malignancies[108], largely through chromosomal rearrangements that place *TCL1A* near *TCR-A* (the gene for the T cell antigen receptor)[109]. This strong association between rs2887399 and mLOY has been replicated in subsequent studies with larger cohorts[11,13]; 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[11]. 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*, *CENPU* and *CENPN*, *MAD1L1* and *MAD2L1*, and *PMF1*), the replication and stability of DNA (for example, *ATM* and *NPAT*), and cell arrest and apoptosis (for example, *TP53*, *BCL2* and *BAX*)[11,13,107]. 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[9,17]. Studies conducted in population-scale biobanks in the UK (UK Biobank (UKB))[9,17,110] and Japan (BioBank Japan (BBJ))[26,111] 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[91]) as well as variants in *SP140* (encoding a lymphoid-restricted nuclear body protein involved in B cell antigen response[112]) are associated with mCAs occurring anywhere in the genome[17], whereas the remaining inherited variants have only been associated with *trans* mCAs on a particular chromosome[9,17,26] (TABLE 1). Apart from common variation in *TCL1A* and *DLK1* (a negative regulator of HSPC differentiation[113]) that is linked to 14q CN-LOH[17] and a known association between the *JAK2* 46/1 haplotype and 9p CN-LOH[9,17,82–84,114], 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 *ATM*, *NBN* and *MRE11*, all of which are genes involved in maintaining genomic integrity, it is their damaged germline allele that is more commonly propagated[9,17]. Conversely, the presence of damaging germline variants in the *MPL* gene, which encodes the thrombopoietin receptor important for HSC self-renewal, are associated with the duplication of the non-damaged allele[9,17]. 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].

Table 1 | **Cis-acting and *trans*-acting risk variants for mCAs**

Inherited risk locus (chromosome)	Risk variant frequency	Effect type	Associated mCAs	Reported odds ratio (95% CI)	Ref.
FH (1q)	Rare	*Cis*	CN-LOH	28 (14–55)	17
NBN (8q)	Rare	*Cis*	CN-LOH	210 (92–484)[17]; 91 (52–159)[26]	17,26
MRE11 (11q)	Rare	*Cis*	CN-LOH	130 (50–338)[17]; 37 (17–84)[26]	17,26
SH2B3 (12q)	Rare	*Cis*	CN-LOH	11 (5.8–20)	17
MPL (1p)	Rare	*Cis*	CN-LOH	142 (111–184)[17]; 54 (30–100)[26]	17,26
ATM (11q)	Rare	*Cis*	CN-LOH	96 (52–177)	17
TM2D3 (15q)	Rare	*Cis*	CN-LOH	555 (425–724)	17
TCL1A (14q)	Common	*Cis*	CN-LOH	0.84 (0.75–0.94)[17]; 0.88 (0.79–0.98)[26]	17,26
DLK1 (14q)	Common	*Cis*	CN-LOH	1.24 (1.13–1.37)[17]; 1.38 (1.31–1.44)[26]	17,26
JAK2 (9p)	Common	*Cis*	CN-LOH	2.29 (1.99–2.63)[17]	17
SP140 (2q)	Common	*Trans*	Any autosomal mCAs	1.08 (1.05–1.10)[17]	17
TERC (3q)	Common	*Trans*	Any autosomal mCAs	0.93 (0.91–0.96)[17]	17
TERT (5p)	Common	*Trans*	Any autosomal mCA 14q CN-LOH	1.11 (1.08–1.14)[17]; 1.27 (1.21–1.33)[26]	17,26
FRA10B (10q)	Common	*Cis*	Loss of 10q	18 (12–26)	9
DXZ1 (X)	Common	*Cis*	Loss of X	1.09 (1.04–1.15)	9
DXZ4 (X)	Common	*Cis*	Loss of X	1.10 (1.04–1.17)	9
HLA (6p)	Common	*Trans*	Loss of X	1.18 (1.12–1.25)	9
SP140L (2q)	Common	*Trans*	Loss of X	1.17 (1.12–1.24)	9
NEDD8–TINF2 (14q)	Common	*Cis*	CN-LOH	1.62 (1.42–1.85)	26
CTU2 (16q)	Rare	*Cis*	CN-LOH	28 (17–45)	26
MAD1L1 (7p)	Common	*Trans*	Gain of 15	1.61 (1.46–1.77)	26

Odds ratios given for the variant with the most significant *p* value. CI, confidence interval; CN-LOH, copy-neutral loss of heterozygosity; mCAs, mosaic chromosomal alterations.

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 (*ATM*, *MPL*, *FRA10B* and *TM2D3–TARSL2*) were absent in the BBJ cohort, whereas variants in several other genes (*MRE11*, *NBN*, *NEDD8–TINF2* and *CTU2*) were present at higher frequencies[26]. 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[26]; these mCAs are often seen in chronic lymphocytic leukaemia[115,116], a malignancy that is four to five times more common among Europeans than among Japanese individuals[117]. Collectively, these studies highlight the importance of including diverse populations in genomics research[118].

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[10,18,27]. Mirroring one of the main signals found with *JAK2*, one study using the deCODE cohort from Iceland found that variation in the *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[10]. Meanwhile, in the same study, individuals with CH were found to have a shorter average telomere length than individuals without CH[10]. An analysis of the NHLBI Trans-Omics for Precision Medicine (TOPMed)[119] 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 *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 *TERT* (rs2853677; OR = 1.32; MAF = 0.42)[18].

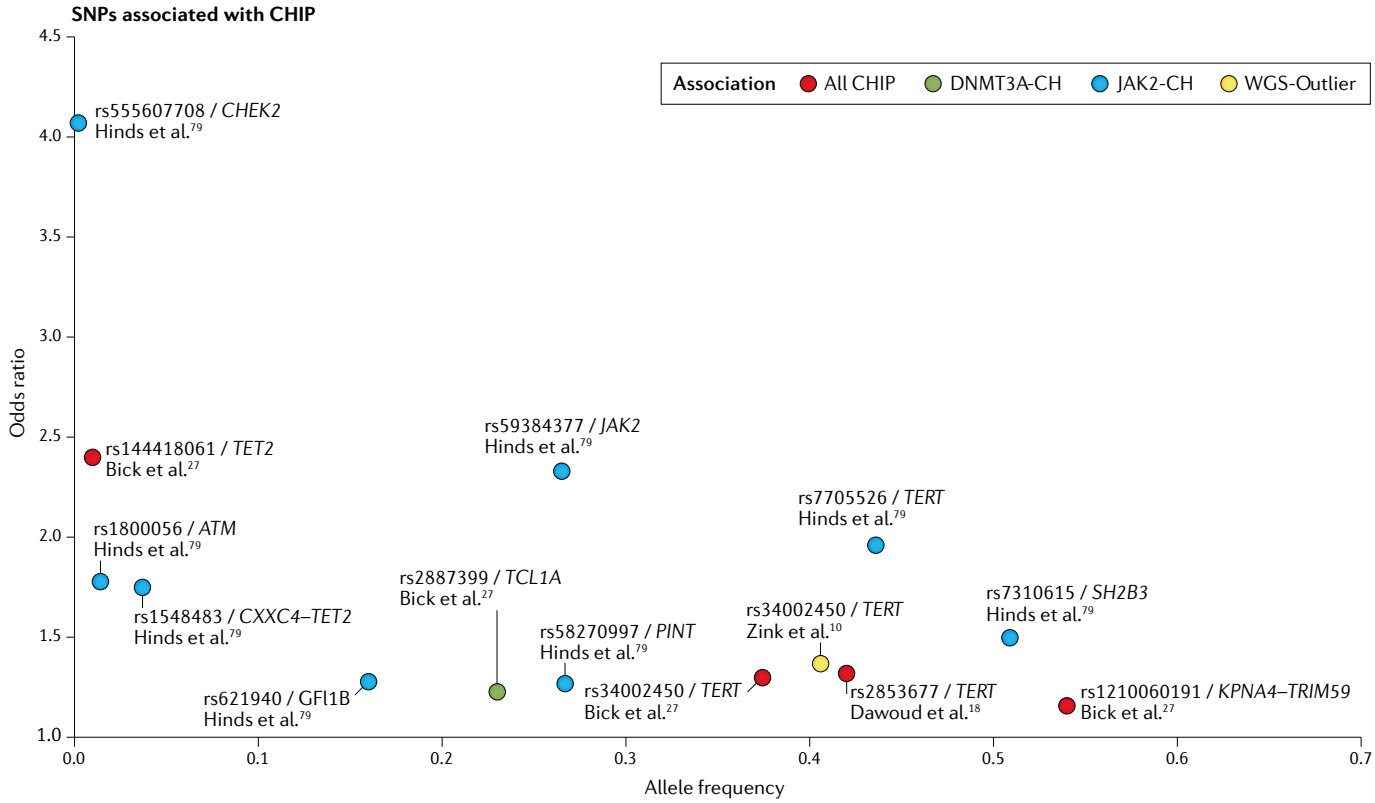


Fig. 3 | **CHIP has polygenic risk.** Genetic association studies have demonstrated that the inherited risk landscape for clonal haematopoiesis of indeterminate potential (CHIP) is characterized by numerous common variants with modest effect sizes and several rare variants associated with strong effects. 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).

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[27]. 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)[27]. 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[27]. 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 *TET2* SNP 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)[27]. Of note, this is the same variant near *TCL1A* that is associated with mLOY[11,14]. 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.

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 syndrome[132], are associated with an increased risk of numerous types of cancer, including leukaemia and lymphoma[133,134], breast cancer[135,136], and prostate cancer[137,138], 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 *ATM*[139], has been reported as a risk factor for

Table 2 | **Select germline risk loci associated with CH and diseases of ageing**

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	Cancer[164]; CVD[165–167]; dementia[178]	Cancer[179]; CVD[180]; dementia[181,182]
CHEK2	Yes	Yes	Yes	Cancer[120–124]	Cancer[183]
ATM	Yes	Yes	Yes	Cancer[133–138]	Cancer[184]
TCL1A	Yes	Yes	Yes	NA	Cancer[185]
46/1 haplotype	Yes	Yes	No	Cancer[82–84]	NA
SH2B3	Yes	Yes	No	Cancer[144]; CVD[153–157]	Cancer[144]
TET2	Yes	No	Yes	Cancer[141–143]; CVD[148]	Cancer[186]; CVD[50,51,54,56,147]
MAD1L1	No	Yes	Yes	NA	Cancer[187]
RUNX1	Yes	No	No	Cancer[100,101]	Cancer[188]
KPNA4–TRIM59	Yes	No	No	NA	*KPNA4*: cancer[189]; *TRIM59*: cancer[190]
GFI1B	Yes	No	No	NA	Cancer[191]
MPL	No	Yes	No	Cancer[192]	Cancer[193]
TM2D3	No	Yes	No	Dementia[194]	NA
FN	No	Yes	No	NA	Cancer[195]
NEDD8	No	Yes	No	NA	Cancer[196]
TINF2	No	Yes	No	Cancer[197]	Cancer[159]
NBN	No	Yes	No	Cancer[128,129]	NA
MRE11	No	Yes	No	Cancer[130,131]	Cancer[198]
SP140	No	Yes	No	NA	Cancer[199]
HLA	No	Yes	Yes	Dementia[200–202]	NA
TERC	No	Yes	No	Cancer[164]; CVD[167,168]; dementia[178]	Cancer[179]; CVD[180]; dementia[181]
TP53	No	No	Yes	Cancer[125–127]	Cancer[203]
BCL2	No	No	Yes	NA	Cancer[204]
BAX	No	No	Yes	Cancer[205]	Cancer[206]
NPAT	No	No	Yes	Cancer[140]	Cancer[207]
CENPU	No	No	Yes	NA	Cancer[208]
MAD2L1	No	No	Yes	NA	Cancer[209]

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.

DNA methylation
The presence of methyl groups added to the DNA base, often occurring on cytosine adjacent to a guanine (CpG). The methylation of DNA in gene promoter sequences may reduce gene expression.

Accelerated epigenetic ageing
Contrary to an individual's chronological age defined by the passage of time, an individual's epigenetic age is a measure of how the pattern of DNA methylation compares to the average pattern observed in the population across all ages. Accelerated epigenetic age means one's DNA methylation pattern is more similar to the average for individuals who are chronologically older than oneself.

Hodgkin lymphoma[140]. 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 of 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[141–143], leads to increased HSC proliferation[96] and secretion of pro-inflammatory cytokines[50,54,56,66]. Lastly, variants in *SH2B3*, a negative regulator of the pro-proliferative JAK–STAT signalling pathway in haematopoietic cells[144], are associated with malignancies[145], including in the blood[144], breast[78,146], lung[146] and colon[78,146].

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[50,51,54,56,147], 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[148]. In contrast to the role of this epigenetic regulator in tumorigenesis, which is thought to rest on increased HSPC self-renewal[96,149–151], lineage skewing[96,149–151] and an increased tendency towards mutation[152], the contribution of mutant *TET2* to pulmonary arterial hypertension may stem from overproduction of inflammatory cytokines (for example, IL-1β) in differentiated immune cells[148]. Meanwhile, genetic variation in the gene *SH2B3* has been linked to numerous aspects of cardiovascular dysfunction, including hypertension[153,154], aortic dissection[155], atherosclerosis[156] and stroke[157]. However, for at least one well-studied variant, there appears to be a trade-off between CVD risk 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[158]. Inherited genetic variation influences telomere length, which is also tightly linked to the somatic expression of telomerase genes[158]. 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 *TINF2* (encoding the TIN2 protein, part of the shelterin complex[159]) 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.[160]) (the lead variant for increased risk for CHIP[27] and global mCA events[17]) and rs2853677 (REF.[161]) (associated with 14q CN-LOH[26]) associate with longer telomeres and greater telomere length, as predicted by germline variation, is positively associated with mCA events[162]. 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)[163]. Importantly, genetic association studies of telomere maintenance genes have revealed links to a broad spectrum of diseases, including strong ties to cancer[164] and CVD[165–168]. 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[169]. 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[169]. Likewise, individuals with CHIP (also an age-associated feature) have epigenetic age acceleration in blood cells[170]. The rate of epigenetic ageing has a heritable component[169], including variation at loci associated with CH risk: *TERT*, *TET2*, *TRIM59* and *KPNA4* (REF.[171]). Paradoxically, faster epigenetic ageing is linked to *TERT* variants associated with longer telomeres[172], 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 *DNMT3A* and *TET2* directly modulate CpG methylation and dictate global methylation patterns within HSPCs[173]. Less common CHIP mutations in *IDH1* and *IDH2* lead to the production of the metabolite 2-hydroxyglutarate, which interferes with the function of TET2 (REF.[174]). Additionally, interestingly, many of the CpG sites used in the Horvath epigenetic clock are near target genes of Polycomb repressive complex 2 (PRC2)[169], a protein complex whose function is impaired by CHIP mutations in *ASXL1* (REF.[174]). 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.

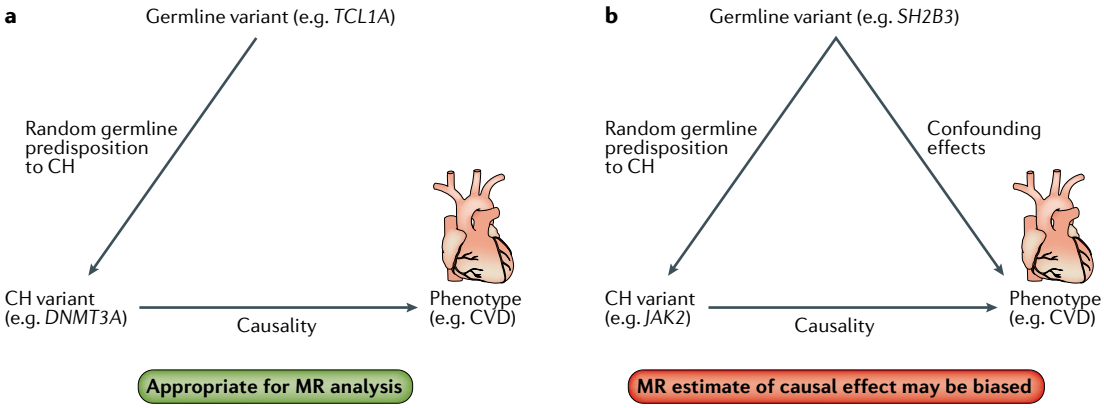


Fig. 4 | **Using germline variation to study causal associations of CH.** Certain situations in which germline variants affect the risk of developing clonal haematopoiesis (CH) can be used to determine whether CH has a causal contribution to a given phenotype. A Mendelian randomization (MR) approach takes advantage of the fact that individuals acquire their germline allelic composition by chance. Any change in CH frequency or clone size due to germline variation can then be treated as the result of a random genetic assignment and used to estimate a causal association between CH and the phenotype. Current evidence suggests that some germline variants may affect CH but do not independently influence the associated phenotypes — such variants are the most likely to be appropriate for MR analyses. For example, *TCL1A* is known to increase the risk of *DNMT3A*-CH but is not known to have direct effects on cardiovascular disease (CVD) (part **a**). However, there is substantial overlap in the genetic architecture of CH and diseases of ageing, so some inherited variants may affect both CH and observed phenotypes. In such cases, MR estimates of causality are confounded by the direct association of the germline variant with the outcome. A good example of this is germline variation in *SH2B3*, which is associated with both the risk of *JAK2*-CH and CVD (part **b**). Unless this latter association can be properly accounted for, a typical MR approach would overestimate the effect of *JAK2* on CVD.

Mendelian randomization (MR) is a statistical technique that utilizes inherited variation to test causation between an exposure (here, CH) and outcome (disease of interest)[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[11]. 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[176] (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 and 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.

Horizontal pleiotropy
In Mendelian randomization, when a genetic variant exerts its effects on a measured outcome through mechanisms separate from the pathway being studied.

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 inherited risk of CH**

- Why does the same inherited variant predispose to multiple subtypes of clonal haematopoiesis (CH)? To what extent is the relationship between inherited 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?
- How much of the relatively higher prevalence of CH in certain populations compared to others may be explained by inherited population-specific variation?
- Can we identify population-specific risk variants for CH?
- Do inherited variants affect downstream CH consequences?
- How do inherited variants affect the rate of clonal expansion?
- 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[7,8,50,51]. Recent work has identified an inherited polymorphism in the IL-6 receptor that reduces the likelihood of heart disease in individuals with CHIP[51]; 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 germline 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[177], 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. Hematopoiesis: 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 acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
5. Jan, M., Ebert, B. L. & Jaiswal, S. Clonal hematopoiesis. *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).
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.
7. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
This work was the first to find associations between non-cancer mortality and CHIP.
8. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
This study delineated CH from unknown drivers and the strong association between CHIP and leukaemia.
9. Loh, P.-R. et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**, 350–355 (2018).
10. Zink, F. et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**, 742–752 (2017).
11. Thompson, D. J. et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).
This study on the genetic determinants of mLOY is the largest to date and used multiple modalities to validate associations.
12. Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat. Rev. Cancer* **21**, 239–256 (2021).
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.
13. 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).
14. Zhou, W. et al. Mosaic loss of chromosome Y is associated with common variation near TCL1A. *Nat. Genet.* **48**, 563–568 (2016).
15. Loftfield, E. et al. Predictors of mosaic chromosome Y loss and associations with mortality in the UK Biobank. *Sci. Rep.* **8**, 12316 (2018).
16. Dumanski, J. P. et al. Smoking is associated with mosaic loss of chromosome Y. *Science* **347**, 81–83 (2015).
17. Loh, P.-R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* **584**, 136–141 (2020).
This large study of mCA genetics found that both monogenic and polygenic risk variants can predict mCA events.
18. Dawoud, A. A. Z., Tapper, W. J. & Cross, N. C. P. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *Leukemia* **34**, 2660–2672 (2020).
19. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7**, 12484 (2016).
This paper measuring somatic mutations down to very low allele burden found acquired variants in the blood of 95% (19/20) of sampled adults.
20. Watson, C. J. et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science* **367**, 1449–1454 (2020).
21. 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).
22. Zhou, W. et al. Detectable chromosome X mosaicism in males is rarely tolerated in peripheral leukocytes. *Sci. Rep.* **11**, 1193 (2021).
23. Jacobs, K. B. et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, 651–658 (2012).
24. Machiela, M. J. et al. Characterization of large structural genetic mosaicism in human autosomes. *Am. J. Hum. Genet.* **96**, 487–497 (2015).
25. Laurie, C. C. et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–650 (2012).
26. Terao, C. et al. Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature* **584**, 130–135 (2020).
27. Bick, A. G. et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**, 763–768 (2020).
This study is the largest GWAS of CHIP done to date.
28. Perner, F., Perner, C., Ernst, T. & Heidel, H. F. Roles of JAK2 in aging, inflammation, hematopoiesis and malignant transformation. *Cells* **8**, 854 (2019).
29. Kralovics, R. et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* **352**, 1779–1790 (2005).
30. Gao, T. et al. Interplay between chromosomal alterations and gene mutations shapes the evolutionary trajectory of clonal hematopoiesis. *Nat. Commun.* **12**, 338 (2021).
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.
31. McKerrell, T. et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep.* **10**, 1239–1245 (2015).
32. Coombs, C. C. et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. *Cell Stem Cell* **21**, 374–382.e4 (2017).
33. Loftfield, E. et al. Mosaic Y loss is moderately associated with solid tumor risk. *Cancer Res.* **79**, 461 (2019).
34. 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).
35. Bolton, K. L. et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**, 1219–1226 (2020).
This paper offers a detailed look at how radiotherapy, chemotherapy, targeted agents and immune therapies are associated with CHIP.
36. Boucai, L. et al. Radioactive iodine-related clonal hematopoiesis in thyroid cancer is common and associated with decreased survival. *J. Clin. Endocrinol. Metab.* **103**, 4216–4223 (2018).
37. Pich, O. et al. The evolution of hematopoietic cells under cancer therapy. *bioRxiv* https://doi.org/10.1101/2020.10.29.360230 (2020).

38. Ganster, C. et al. New data shed light on Y-loss-related pathogenesis in myelodysplastic syndromes. *Genes Chromosomes Cancer* **54**, 717–724 (2015).
39. Noveski, P. et al. Loss of Y chromosome in peripheral blood of colorectal and prostate cancer patients. *PLoS ONE* **11**, e0146264 (2016).
40. Machiela, M. J. et al. Mosaic chromosome Y loss and testicular germ cell tumor risk. *J. Hum. Genet.* **62**, 637–640 (2017).
41. Haitjema, S. et al. Loss of Y chromosome in blood is associated with major cardiovascular events during follow-up in men after carotid endarterectomy. *Circ. Cardiovasc. Genet.* **10**, e001544 (2017).
42. Dumanski, J. P. et al. Mosaic loss of chromosome Y in blood is associated with Alzheimer disease. *Am. J. Hum. Genet.* **98**, 1208–1219 (2016).
43. Hirata, T. et al. Investigation of chromosome Y loss in men with schizophrenia. *Neuropsychiatr. Dis. Treat.* **14**, 2115–2122 (2018).
44. 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).
45. Lleo, A. et al. Y chromosome loss in male patients with primary biliary cirrhosis. *J. Autoimmun.* **41**, 87–91 (2013).
46. Grassmann, F. et al. Y chromosome mosaicism is associated with age-related macular degeneration. *Eur. J. Hum. Genet.* **27**, 36–41 (2019).
47. Zekavat, S. M. et al. Hematopoietic mosaic chromosomal alterations and risk for infection among 767,891 individuals without blood cancer. *medRxiv* https://doi.org/10.1101/2020.11.12.20230821 (2020).
48. Abelson, S. et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400–404 (2018).
49. Desai, P. et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat. Med.* **24**, 1015–1023 (2018).
50. Jaiswal, S. et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **337**, 111–121 (2017).
51. Bick, A. G. et al. Genetic interleukin 6 signaling deficiency attenuates cardiovascular risk in clonal hematopoiesis. *Circulation* **141**, 124–131 (2020). **This paper was the first to examine how inherited germline variation can alter the clinical consequences of CHIP, finding lower rates of cardiovascular events in individuals with CHIP and a germline polymorphism attenuating IL-6 receptor signalling compared to those without the polymorphism**.
52. Honigberg, M. C. et al. Premature menopause, clonal hematopoiesis, and coronary artery disease in postmenopausal women. *Circulation* **143**, 410–423 (2021).
53. Busque, L. et al. High-sensitivity C-reactive protein is associated with clonal hematopoiesis of indeterminate potential. *Blood Adv.* **4**, 2430–2438 (2020).
54. Fuster, J. J. et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842–847 (2017).
55. Wang, W. et al. Macrophage inflammation, erythrophagocytosis and accelerated atherosclerosis in Jak2V617F Mice. *Circ. Res.* **11**, e35–e47 (2018).
56. Sano, S. et al. Tet2-mediated clonal hematopoiesis accelerates heart failure through a mechanism involving the IL-1β/NLRP3 inflammasome. *J. Am. Coll. Cardiol.* **71**, 875–886 (2018).
57. Sano, S. et al. JAK2V617F-mediated clonal hematopoiesis accelerates pathological remodeling in murine heart failure. *JACC* **4**, 684–697 (2019).
58. Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* **19**, 81–92 (2018).
59. Zhang, Q. et al. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**, 389–393 (2015).
60. Poisson, J. et al. Erythrocyte-derived microvesicles induce arterial spasms in JAK2V617F myeloproliferative neoplasm. *J. Clin. Invest.* **130**, 2630–2643 (2020).
61. Wolach, O. et al. Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. *Sci. Transl. Med.* **10**, eaan8292 (2018).
62. Buscarlet, M. et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130**, 753–762 (2017).
63. Miller, P. et al. Contribution of clonal hematopoiesis to adult-onset hemophagocytic lymphohistiocytosis. *Blood* **136**, 3051–3055 (2020).
64. Arends, C. M. et al. Clonal hematopoiesis in patients with anti-neutrophil cytoplasmic antibody-associated vasculitis. *Haematologica* **105**, e264–e267 (2020).
65. Bolton, K. L. et al. Clonal hematopoiesis is associated with risk of severe Covid-19. *medRxiv* https://doi.org/10.1101/2020.11.25.20233163 (2020).
66. Abplanalp, W. T. et al. Association of clonal hematopoiesis of indeterminate potential with inflammatory gene expression in patients with severe degenerative aortic valve stenosis or chronic postischemic heart failure. *JAMA Cardiol.* **5**, 1170–1175 (2020).
67. Bick, A. G. et al. Increased CHIP Prevalence Amongst People Living with HIV. *medRxiv* https://doi.org/10.1101/2020.11.06.20225607 (2020).
68. Dharan, N. J. et al. Age-related clonal haematopoiesis is more prevalent in older adults with HIV: the ARCHIVE study. *medRxiv* https://doi.org/10.1101/2020.11.19.20235069 (2020).
69. Takahashi, K. et al. Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet Oncol.* **18**, 100–111 (2017).
70. Lindsley, R. C. et al. Prognostic mutations in myelodysplastic syndrome after stem-cell transplantation. *N. Engl. J. Med.* **376**, 536–547 (2017).
71. Kahn, J. D. et al. PPM1D truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. *Blood* **132**, 1095–1105 (2018).
72. Hsu, J. I. et al. PPM1D mutations drive clonal hematopoiesis in response to cytotoxic chemotherapy. *Cell Stem Cell* **23**, 700–713.e706 (2018).
73. Mas-Peiro, S. et al. Clonal haematopoiesis in patients with degenerative aortic valve stenosis undergoing transcatheter aortic valve implantation. *Eur. Heart J.* **41**, 933–939 (2019).
74. Nawas, M. T. et al. The clinical implications of clonal hematopoiesis in hematopoietic cell transplantation. *Blood Rev.* **46**, 100744 (2020).
75. Wilk, C. M., Manz, M. G. & Boettcher, S. Clonal hematopoiesis in hematopoietic stem cell transplantation. *Curr. Opin. Hematol.* **28**, 94–100 (2021).
76. Steensma, D. P. et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).
77. Sperling, A. S., Gibson, C. J. & Ebert, B. L. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat. Rev. Cancer* **17**, 5–19 (2017).
78. Bao, E. L. et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. *Nature* **586**, 769–775 (2020).
79. Hinds, D. A. et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121–1128 (2016). **This large study of individuals with *JAK2* mutations found several strong germline associations**.
80. Vardiman, J. W. et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937 (2009).
81. Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).
82. Jones, A. V. et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat. Genet.* **41**, 446–449 (2009).
83. Kilpivaara, O. et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. *Nat. Genet.* **41**, 455–459 (2009).
84. Olcaydu, D. et al. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat. Genet.* **41**, 450–454 (2009). **Jones et al.[82], Kilpivaara et al.[83] and Olcaydu et al.[84] identified variation in the 46/1 haplotype as linked to MPNs**.
85. Trifa, A. P. et al. The G allele of the JAK2 rs10974944 SNP, part of JAK2 46/1 haplotype, is strongly associated with JAK2 V617F-positive myeloproliferative neoplasms. *Ann. Hematol.* **89**, 979–983 (2010).
86. Trifa, A. P. et al. TERT rs2736100 A>C SNP and JAK2 46/1 haplotype significantly contribute to the occurrence of JAK2 V617F and CALR mutated myeloproliferative neoplasms – a multicentric study on 529 patients. *Br. J. Haematol.* **174**, 218–226 (2016).
87. Trifa, A. P. et al. MECOM, HBS1L-MYB, THRB-RARB, JAK2, and TERT polymorphisms defining the genetic predisposition to myeloproliferative neoplasms: a study on 939 patients. *Am. J. Hematol.* **93**, 100–106 (2018).
88. Tapper, W. et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat. Commun.* **6**, 6691 (2015).
89. Oddsson, A. et al. The germline sequence variant rs2736100_C in TERT associates with myeloproliferative neoplasms. *Leukemia* **28**, 1371–1374 (2014).
90. McKerrell, T. et al. *JAK2* V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion kinetics. *Blood Adv.* **1**, 968–971 (2017).
91. Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* **20**, 299–309 (2019).
92. Zimmermann, S. & Martens, U. M. Telomeres, senescence, and hematopoietic stem cells. *Cell Tissue Res.* **331**, 79–90 (2008).
93. Akada, H. et al. Critical role of Jak2 in the maintenance and function of adult hematopoietic stem cells. *Stem Cells* **32**, 1878–1889 (2014).
94. Allsopp, R. C., Morin, G. B., DePinho, R., Harley, C. B. & Weissman, I. L. Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. *Blood* **102**, 517–520 (2003).
95. Gery, S. & Koeffler, H. P. Role of the adaptor protein LNK in normal and malignant hematopoiesis. *Oncogene* **32**, 3111–3118 (2013).
96. Moran-Crusio, K. et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
97. Ito, K. et al. Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature* **431**, 997–1002 (2004).
98. van der Meer, L. T., Jansen, J. H. & van der Reijden, B. A. Gfi1 and Gfi1b: key regulators of hematopoiesis. *Leukemia* **24**, 1834–1843 (2010).
99. Ichikawa, M. et al. AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis. *J. Immunol.* **180**, 4402–4408 (2008).
100. Churpek, J. E. et al. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood* **126**, 2484–2490 (2015).
101. Brown, A. L. et al. RUNX1-mutated families show phenotype heterogeneity and a somatic mutation profile unique to germline predisposed AML. *Blood Adv.* **4**, 1131–1144 (2020).
102. Mangaonkar, A. A. & Patnaik, M. M. Hereditary predisposition to hematopoietic neoplasms: when bloodline matters for blood cancers. *Mayo Clin. Proc.* **95**, 1482–1498 (2020).
103. Fabre, M. A. et al. Concordance for clonal hematopoiesis is limited in elderly twins. *Blood* **135**, 269–273 (2020).
104. Hansen, J. W. et al. Clonal hematopoiesis in elderly twins: concordance, discordance, and mortality. *Blood* **135**, 261–268 (2020). **Fabre et al.[103] and Hansen et al.[104] are the first studies examining the prevalence of CHIP among twin pairs; these papers also provided the first evidence suggesting some CHIP mutations may be acquired in utero**.
105. Williams, N. et al. Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. *bioRxiv* https://doi.org/10.1101/2020.11.09.374710 (2020).
106. Van Egeren, D. et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell* **28**, 514–523.e9 (2021).
107. Terao, C. et al. GWAS of mosaic loss of chromosome Y highlights genetic effects on blood cell differentiation. *Nat. Commun.* **10**, 4719 (2019).
108. Laine, J., Künstle, G., Obata, T., Sha, M. & Noguchi, M. The protooncogene *TCL1* Is an Akt kinase coactivator. *Mol. Cell* **6**, 395–407 (2000).
109. Virgilio, L. et al. Deregulated expression of *TCL1* causes T cell leukemia in mice. *Proc. Natl Acad. Sci. USA* **95**, 3885–3889 (1998).
110. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

111. Nagai, A. et al. Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27** (Suppl. 3), S2–S8 (2017).
112. Madani, N. et al. Implication of the lymphocyte-specific nuclear body protein Sp140 in an innate response to human immunodeficiency virus type 1. *J. Virol.* **76**, 11133–11138 (2002).
113. Li, L., Forman, S. J. & Bhatia, R. Expression of DLK1 in hematopoietic cells results in inhibition of differentiation and proliferation. *Oncogene* **24**, 4472–4476 (2005).
114. Koren, A. et al. Genetic variation in human DNA replication timing. *Cell* **159**, 1015–1026 (2014).
115. Landau, D. A. et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
116. Puente, X. S. et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
117. Tamura, K. et al. Chronic lymphocytic leukemia (CLL) is rare, but the proportion of T-CLL is high in Japan. *Eur. J. Haematol.* **67**, 152–157 (2001).
118. Ben-Eghan, C. et al. Don't ignore genetic data from minority populations. *Nature* **585**, 184–186 (2020).
119. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
120. Bell, D. W. et al. Heterozygous germ line *hCHK2* mutations in Li-Fraumeni syndrome. *Science* **286**, 2528 (1999).
121. Vahteristo, P. et al. p53, CHK2, and CHK1 genes in Finnish families with Li-Fraumeni syndrome: further evidence of CHK2 in inherited cancer predisposition. *Cancer Res.* **61**, 5718–5722 (2001).
122. Siddiqui, R. et al. The TP53 mutational spectrum and frequency of CHEK2*1100delC in Li-Fraumeni-like kindreds. *Fam. Cancer* **4**, 177–181 (2005).
123. Ruijs, M. W. et al. The contribution of CHEK2 to the TP53-negative Li-Fraumeni phenotype. *Hered. Cancer Clin. Pract.* **7**, 4 (2009).
124. Manoukian, S. et al. Two new CHEK2 germ-line variants detected in breast cancer/sarcoma families negative for BRCA1, BRCA2, and TP53 gene mutations. *Breast Cancer Res. Treat.* **130**, 207–215 (2011).
125. Malkin, D. et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233–1238 (1990).
126. Chompret, A. et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *Br. J. Cancer* **82**, 1932–1937 (2000).
127. Gonzalez, K. D. et al. Beyond Li fraumeni syndrome: clinical characteristics of families With p53 germline mutations. *J. Clin. Oncol.* **27**, 1250–1256 (2009).
128. Carney, J. P. et al. The hMre11/hRad50 protein complex and nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93**, 477–486 (1998).
129. Varon, R. et al. Nibrin, a novel DNA double-strand break repair protein, is mutated in nijmegen breakage syndrome. *Cell* **93**, 467–476 (1998).
130. Bartkova, J. et al. Aberrations of the MRE11–RAD50–NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. *Mol. Oncol.* **2**, 296–316 (2008).
131. Belhadj, S. et al. Candidate genes for hereditary colorectal cancer: mutational screening and systematic review. *Hum. Mutat.* **41**, 1563–1576 (2020).
132. Savitsky, K. et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
133. Suarez, F. et al. Incidence, presentation, and prognosis of malignancies in ataxia-telangiectasia: a report from the French national registry of primary immune deficiencies. *J. Clin. Oncol.* **33**, 202–208 (2015).
134. Reiman, A. et al. Lymphoid tumours and breast cancer in ataxia telangiectasia; substantial protective effect of residual ATM kinase activity against childhood tumours. *Br. J. Cancer* **105**, 586–591 (2011).
135. Swift, M., Morrell, D., Massey, R. B. & Chase, C. L. Incidence of cancer in 161 families affected by ataxia–telangiectasia. *N. Engl. J. Med.* **325**, 1831–1836 (1991).
136. Thompson, D. et al. Cancer risks and mortality in heterozygous ATM mutation carriers. *J. Natl Cancer Inst.* **97**, 813–822 (2005).
137. Lu, C. et al. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat. Commun.* **6**, 10086 (2015).
138. Angèle, S. et al. ATM polymorphisms as risk factors for prostate cancer development. *Br. J. Cancer* **91**, 783–787 (2004).
139. Medina, R. et al. The HiNF-P/p220^{NPAT} cell cycle signaling pathway controls nonhistone target genes. *Cancer Res.* **67**, 10334–10342 (2007).
140. Saarinen, S. et al. Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. *Blood* **118**, 493–498 (2011).
141. Duployez, N. et al. Familial myeloid malignancies with germline TET2 mutation. *Leukemia* **34**, 1450–1453 (2020).
142. Stremenova Spegarova, J. et al. Germline TET2 loss of function causes childhood immunodeficiency and lymphoma. *Blood* **136**, 1055–1066 (2020).
143. Kaasinen, E. et al. Impact of constitutional TET2 haploinsufficiency on molecular and clinical phenotype in humans. *Nat. Commun.* **10**, 1252 (2019).
144. Maslah, N., Cassinat, B., Verger, E., Kiladjian, J. J. & Velazquez, L. The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. *Leukemia* **31**, 1661–1670 (2017).
145. Kuo, C.-L. et al. The longevity-associated SH2B3 (LNK) genetic variant: selected aging phenotypes in 379,758 subjects. *J. Gerontol.* **75**, 1656–1662 (2020).
146. Hung, R. J. et al. Cross cancer genomic investigation of inflammation pathway for five common cancers: lung, ovary, prostate, breast, and colorectal cancer. *J. Natl Cancer Inst.* **107**, djv246 (2015).
147. Wang, Y. et al. Tet2-mediated clonal hematopoiesis in nonconditioned mice accelerates age-associated cardiac dysfunction. *JCI Insight* **5**, e135204 (2020).
148. Potus, F. et al. Novel mutations and decreased expression of the epigenetic regulator TET2 in pulmonary arterial hypertension. *Circulation* **141**, 1986–2000 (2020).
149. Ko, M. et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl Acad. Sci. USA* **108**, 14566–14571 (2011).
150. Li, Z. et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509–4518 (2011).
151. Quivoron, C. et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25–38 (2011).
152. Pan, F. et al. Tet2 loss leads to hypermutagenicity in haematopoietic stem/progenitor cells. *Nat. Commun.* **8**, 15102 (2017).
153. Levy, D. et al. Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).
154. Newton-Cheh, C. et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).
155. Laroumanie, F. et al. LNK deficiency promotes acute aortic dissection and rupture. *JCI Insight* **3**, e122558 (2018).
156. Safarova Maya, S. et al. Targeted sequencing study to uncover shared genetic susceptibility between peripheral artery disease and coronary heart disease — brief report. *Arterioscler. Thromb. Vasc. Biol.* **39**, 1227–1233 (2019).
157. Dichgans, M. et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24–36 (2014).
158. Demanelis, K. et al. Determinants of telomere length across human tissues. *Science* **369**, eaaz6876 (2020).
159. Hosokawa, K. & Arai, F. The role of telomere binding molecules for normal and abnormal hematopoiesis. *Int. J. Hematol.* **107**, 646–655 (2018).
160. Bojesen, S. E. et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.* **45**, 371–384 (2013).
161. Melin, B. S., Nordfjäll, K., Andersson, U. & Roos, G. hTERT cancer risk genotypes are associated with telomere length. *Genet. Epidemiol.* **36**, 368–372 (2012).
162. Brown, D. W. et al. Genetically predicted telomere length is associated with clonal somatic copy number alterations in peripheral leukocytes. *PLoS Genet.* **16**, e1009078 (2020).
163. Nakao, T. et al. Bidirectional Mendelian randomization supports bidirectional causality between telomere length and clonal hematopoiesis of intermediate potential. *medRxiv* https://doi.org/10.1101/2021.02.26.21252199 (2021).
164. McNally, E. J., Luncsford, P. J. & Armanios, M. Long telomeres and cancer risk: the price of cellular immortality. *J. Clin. Invest.* **129**, 3474–3481 (2019).
165. Han, H. et al. Association of TERT polymorphisms and risk of coronary heart disease in a Chinese Han population. *Oncotarget* **8**, 67519–67525 (2017).
166. Bressler, J. et al. Sequence variation in telomerase reverse transcriptase (TERT) as a determinant of risk of cardiovascular disease: the Atherosclerosis Risk in Communities (ARIC) study. *BMC Med. Genet.* **16**, 52 (2015).
167. Codd, V. et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **45**, 422–427 (2013).
168. Maubaret, C. G. et al. Association of TERC and OBFC1 haplotypes with mean leukocyte telomere length and risk for coronary heart disease. *PLoS ONE* **8**, e83122 (2013).
169. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
This review provides a nuanced overview of epigenetic clocks.
170. Robertson, N. A. et al. Age-related clonal haemopoiesis is associated with increased epigenetic age. *Curr. Biol.* **29**, R786–R787 (2019).
171. McCartney, D. L. et al. Genome-wide association studies identify 137 loci for DNA methylation biomarkers of ageing. *bioRxiv* https://doi.org/10.1101/2020.06.29.133702 (2020).
172. Lu, A. T. et al. GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nat. Commun.* **9**, 387 (2018).
173. Izzo, F. et al. DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **52**, 378–387 (2020).
174. Bowman, R. L., Busque, L. & Levine, R. L. Clonal hematopoiesis and evolution to hematopoietic malignancies. *Cell Stem Cell* **22**, 157–170 (2018).
175. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, R89–R98 (2014).
176. Holmes, M. V., Ala-Korpela, M. & Smith, G. D. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat. Rev. Cardiol.* **14**, 577–590 (2017).
177. Bolton, K. L. et al. The clinical management of clonal hematopoiesis: creation of a clonal hematopoiesis clinic. *Hematol. Oncol. Clin. North Am.* **34**, 357–367 (2020).
This paper describes a new type of clinic that focuses on patients with CH.
178. Scheller Madrid, A. et al. Observational and genetic studies of short telomeres and Alzheimer's disease in 67,000 and 152,000 individuals: a Mendelian randomization study. *Eur. J. Epidemiol.* **35**, 147–156 (2020).
179. Roake, C. M. & Artandi, S. E. Regulation of human telomerase in homeostasis and disease. *Nat. Rev. Mol. Cell Biol.* **21**, 384–397 (2020).
180. Yeh, J.-K. & Wang, C.-Y. Telomeres and telomerase in cardiovascular diseases. *Genes* **7**, 58 (2016).
181. Boccardi, V., Pelini, L., Ercolani, S., Ruggiero, C. & Mecocci, P. From cellular senescence to Alzheimer's disease: the role of telomere shortening. *Ageing Res. Rev.* **22**, 1–8 (2015).
182. Herrmann, M., Pusceddu, I., März, W. & Herrmann, W. Telomere biology and age-related diseases. *Clin. Chem. Lab. Med.* **56**, 1210–1222 (2018).
183. Wu, X., Dong, X., Liu, W. & Chen, J. Characterization of CHEK2 mutations in prostate cancer. *Hum. Mutat.* **27**, 742–747 (2006).
184. Nanda, N. & Roberts, N. J. ATM serine/threonine kinase and its role in pancreatic risk. *Genes* **11**, 108 (2020).
185. Teitell, M. A. The TCL1 family of oncoproteins: co-activators of transformation. *Nat. Rev. Cancer* **5**, 640–648 (2005).
186. Lio, C.-W. J., Yuita, H. & Rao, A. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood* **134**, 1487–1497 (2019).
187. Ryan, S. D. et al. Up-regulation of the mitotic checkpoint component Mad1 causes chromosomal instability and resistance to microtubule poisons. *Proc. Natl Acad. Sci. USA* **109**, E2205–E2214 (2012).
188. Klco, J. M. & Mullighan, C. G. Advances in germline predisposition to acute leukaemias and myeloid neoplasms. *Nat. Rev. Cancer* **21**, 122–137 (2021).

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. Aniththa, T. et al. Gfi1b: 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 TM2D3 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, sine quibus 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. & Latham, M. P. A survey of reported disease-related mutations in the MRE11-RAD50-NBS1 complex. *Cells* **9**, 1678 (2020).
199. Bolli, 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**, 1372–1375 (2013).
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 Institute 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 Burroughs Wellcome Foundation career award for medical scientists and an NIH Director's Early Independence Award from the National Institute of Health Common Fund (DP5 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. Vassiliou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.