Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2026 January 29.

Published in final edited form as:

Science. 2026 January ; 391(6780): 52–58. doi:10.1126/science.adx4174.

Inherited resilience to clonal hematopoiesis by modifying stem cell RNA regulation

Gaurav Agarwal[1,2,3,4], **Mateusz Antoszewski**[1,2,3,4], **Xueqin Xie**[5], **Yash Pershad**[6], **Uma P. Arora**[1,2,3,4], **Chi-Lam Poon**[7], **Peng Lyu**[1,2,3,4], **Andrew J. Lee**[1,2,3,4], **Chun-Jie Guo**[1,2,3,4], **Tianyi Ye**[1,2,3,4], **Laila Barakat Norford**[1,2,3,4,8], **Anna-Lena Neehus**[1,2,3,4], **Lucrezia della Volpe**[1,2,3,4], **Lara Wahlster**[1,2,3,4], **Diyanath Ranasinghe**[9], **Tzu-Chieh Ho**[5,†], **Trevor S. Barlowe**[5], **Arthur Chow**[5], **Alexandra Schurer**[5,‡], **James Taggart**[5,§], **Benjamin H. Durham**[5,¶], **Omar Abdel-Wahab**[5], **Kathy L. McGraw**[10,11,12], **James M. Allan**[9], **Ruslan Soldatov**[7], **Alexander G. Bick**[6], **Michael G. Kharas**[5,*], **Vijay G. Sankaran**[1,2,3,4,13,*]

[1]Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA.

[2]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA.

[3]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA.

[4]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[5]Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

This work is licensed under a Creative Commons Attribution 4.0 International License, which allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

*Corresponding author. sankaran@broadinstitute.org (V.G.S.); kharasm@mskcc.org (M.G.K.).
†Present address: Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, USA.
‡Present address: Albert Einstein College of Medicine, Bronx, NY, USA.
§Present address: Department of Systems Biology, Harvard Medical School, Boston, MA, USA.
¶Present address: Departments of Pediatrics and Pathology and Laboratory Medicine, Rutgers Cancer Institute, New Brunswick, NJ, USA.

Author contributions: Conceptualization: G.A., M.G.K., V.G.S.; Methodology: G.A., M.G.K., V.G.S.; Experiments: G.A., M.A., X.X., P.L., L.d.V., T.-C.H., T.S.B., A.C., A.S., J.T.; Computational analyses: G.A., Y.P., C.-L.P., A.J.L., L.B.N., D.R., K.L.M.; Resources and feedback: C.-J.G., T.Y., A.-L.N., L.d.V., L.W., B.H.D., O.A.-W., J.M.A., R.S., A.G.B.; Visualization: G.A., Y.P., U.P.A., A.J.L., M.G.K., V.G.S.; Funding acquisition: V.G.S., M.G.K.; Supervision: V.G.S., M.G.K.; Writing – original draft: G.A., V.G.S.; Writing – review & editing: G.A., V.G.S., with input from all authors.

Competing interests: O.A.-W. is a founder and scientific advisor of Codify Therapeutics and holds equity and receives research funding from this company. O.A.-W. has served as a consultant for Amphista Therapeutics and MagnetBio and is on scientific advisory boards of Envisagenics Inc. and Harmonic Discovery Inc.; O.A.-W. received research funding from Nurix Therapeutics, Minovia Therapeutics, and LOXO Oncology, unrelated to this work. M.G.K. serves as an advisor to 858 Therapeutics, Inc., and receives funding from AstraZeneca and Transition Bio, Inc., all unrelated to this work. V.G.S. serves as an advisor to Ensoma and Cellarity, unrelated to this work. All other authors declare that they have no competing interests.

Data, code and materials availability: All raw data for $CD34^+$ MSI2-HyperTRIBE are available from European Nucleotide Archive (ENA) under study accession no. ERP182264. All raw data for scRNA-seq (GSE291294) and related code are available in our GitHub repository (https://github.com/sankaranlab/chip_resilience) and Zenodo (67). Bulk RNA-seq (GSE291293) experiments are available in the Gene Expression Omnibus (GEO) repository. Bulk RNA-seq data of *ASXL1* mutant or wild-type $CD34^+$ cells from chronic myelomonocytic leukemia patients (GSE159543) were previously reported (48). No original code or materials were generated in this study.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[6]Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA.

[7]Computational Oncology Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[8]PhD Program in Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

[9]Translational and Clinical Research Institute, Newcastle University Centre for Cancer, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK.

[10]Immune Deficiency Cellular Therapy Program, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[11]Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[12]Myeloid Malignancies Program, National Institutes of Health, Bethesda, MD, USA.

[13]Harvard Stem Cell Institute, Cambridge, MA, USA.

Abstract

Somatic mutations that increase the fitness of hematopoietic stem cells (HSCs) drive their expansion in clonal hematopoiesis (CH) and predispose individuals to blood cancers. Population variation in the growth rate and potential of mutant clones suggests that genetic factors may confer resilience against CH. Here, we identified a noncoding regulatory variant, rs17834140-T, that protects against CH and myeloid malignancies by selectively down-regulating the RNA-binding protein MSI2 in HSCs. By modeling variant effects and mapping MSI2 binding targets, we uncovered an RNA network that maintains human HSCs and influences CH risk. Variant rs17834140-T was associated with slower CH expansion, and stem cell MSI2 levels modified *ASXL1*-mutant HSC clonal dominance. These findings leverage natural resilience to illuminate posttranscriptional regulation in human HSCs, suggesting that inhibition of MSI2 or its targets could be rational strategies for blood cancer prevention.

Healthy aging tissues harbor a substantial burden of cancer driver mutations, underscoring the widespread nature of somatic mosaicism (1). In the hematopoietic system, myeloid malignancy (MyM)–associated mutations are commonly detected in healthy individuals, a phenomenon termed clonal hematopoiesis (CH). Some mutations in hematopoietic stem and progenitor cells (HSPCs) confer a fitness advantage (2), causing pervasive positive selection and expansion with age (3, 4). Individuals with detectable CH of indeterminate potential (CHIP) above ~2% in peripheral blood have an elevated lifetime risk of developing MyMs (5) and a range of chronic disorders (6-8), including cardiovascular disease (6, 9). However, although there has been marked progress in describing the natural history and consequences of CH, the mechanisms that drive the dominance of stem cell clones remain poorly understood.

Recently, mechanistic investigations based on genome-wide association studies (GWASs) (10-14) have identified genetic regulators of stem cell clonality. For example, functional follow-up of inherited MyM risk variants has revealed that increased HSC self-renewal

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

drives MyM predisposition (11, 14-17). By contrast, the germline mechanisms that protect HSCs from CH and MyMs remain poorly understood. Although CH driver mutations occur ubiquitously in aging adults (18), most never progress to a large clone size or overt MyM. Moreover, lineage tracing has demonstrated that driver mutations may be acquired decades before disease, suggesting that there are mechanisms that restrain expansion (19, 20). Indeed, longitudinal profiling (3, 21-23) and inference of clonal expansion rates (17) show that the growth rate and potential of mutant HSC clones vary greatly. Although chronic inflammation (24) and chemical exposures (25) could contribute to this variable expansion, these observations suggest that germline genetic variation might also promote resilience to CH and MyMs.

Here, we leveraged human genetic variation to identify an inherited mechanism that protects HSCs from CHIP, through down-regulation of the RNA-binding protein MSI2. By modeling variant effects in primary human hematopoietic cells and conducting mechanistic studies, we uncovered an MSI2-regulated RNA network that maintains human HSCs and that is attenuated through genetic variation to resist stem cell expansion and progression to MyMs.

Identification of a CHIP and MyM resilience haplotype at the 17q22 locus

Given population variability in the expansion of mutant CH clones (3, 21-23), we hypothesized that there may be inherited mechanisms protecting individuals from developing a large clone size. To explore this, we conducted a GWAS meta-analysis for CHIP. We replicated associations at 24 loci reported previously (in UK Biobank and the Geisinger Health Study) (11) in the All of Us cohort and conducted a meta-analysis of 43,619 CHIP-carriers and 598,761 controls (Fig. 1A and table S1). This analysis identified a CHIP-protective haplotype at the 17q22 locus—that is common across populations (fig. S1, A and B)—as the most protective effect [additive odds ratio (OR) = 0.84, 95% confidence interval (CI) = 0.82 to 0.87, $P = 9.6 \times 10^{-22}$] (Fig. 1B), with homozygotes carrying an ~30% reduced risk of CHIP (Fig. 1C, **top**). An independent GWAS (26) had replicated this CHIP-resilience haplotype (fig. S1, C and D), but the underlying mechanisms have remained unexplored.

In a meta-analysis, the CHIP-protective haplotype was associated with global protection from CHIP subtypes and MyMs. In contrast to variation at other loci with opposing and mutation-selective effects [such as at *TCL1A* (17) and *CD164* (11)], the 17q22 haplotype was associated with consistent resilience to canonical CHIP driver mutations, including *DNMT3A*-CHIP (OR = 0.81), *TET2*-CHIP (OR = 0.87), *ASXL1*-CHIP (OR = 0.74), and *JAK2*-CHIP (OR = 0.48) (Fig. 1C, **middle, and** table S2). Furthermore, carriers had a reduced risk of MyMs (OR = 0.80, 95% CI = 0.70 to 0.90), including myeloproliferative neoplasms (OR = 0.63, 95% CI = 0.50 to 0.81) and acute myeloid leukemia (AML) (OR = 0.83, 95% CI = 0.72 to 0.95) (Fig. 1C, **bottom**), and a phenome-wide association study strongly associated this haplotype with reduced peripheral counts of several blood cell lineages (Fig. 1D) (11). No other phenotypic associations were identified (fig. S1E), suggesting that the CHIP-resilience 17q22 haplotype might selectively alter hematopoiesis to confer a pan-protective effect against CHIP and MyMs.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

rs17834140-T protects from CHIP through HSC-selective down-regulation of MSI2

To investigate the mechanisms underlying inherited CHIP-resilience, we fine-mapped the 17q22 haplotype. The sentinel single-nucleotide polymorphism (SNP) rs80093687 tagged six fine-mapped SNPs in perfect linkage disequilibrium ($r^2 = 1.0$) (Fig. 2A, **top**). Although no candidate variants mapped to coding regions or were predicted to alter splicing (table S3), we identified rs17834140-T as a likely causal variant, which was compelling for several reasons. First, rs17834140-T resided within a putative regulatory element bound by a complex of transcription factors critical for maintaining human HSCs (Fig. 2A, **bottom**) (27, 28). Second, the element was selectively accessible in HSCs (Fig. 2B) but not in more differentiated progenitors or mature blood cells, or in hematopoietic cell lines (fig. S2A), suggesting a stem cell-specific regulatory function. Third, H3K27ac-HiChIP (acetylated histone H3 lysine 27–chromatin precipitation with Hi-C) in human HSPCs (29) showed that the element could physically contact the promoter of its proximal gene, *MSI2* (Fig. 2C), which encodes MSI2, an RNA-binding protein that is critical for the maintenance of HSCs (30, 31) and that promotes leukemia and other cancers (32). *MSI2* was the only gene appreciably expressed within 1 Mb of rs17834140 in human HSCs (fig. S2B), and there was high correlation between variant accessibility [by ATAC-seq (assay for transposable-accessible chromatin-sequencing)] and *MSI2* expression across human hematopoietic cells (ATAC-RNA $r = 0.65$) (fig. S2C). CRISPR interference with dCas9-KRAB–mediated repression of the regulatory element reduced *MSI2* expression in $CD34^+CD45RA^-CD90^+$ HSC-enriched populations (Fig. 2D). These findings suggest that the rs17834140-harboring regulatory element acts as an endogenous enhancer of *MSI2*.

To decipher variant effects, we introduced the rs17834140-harboring regulatory element upstream of a *MSI2* promoter–driven green fluorescent protein (GFP) reporter in primary human HSPCs (fig. S2D). Within transduced $CD34^+CD45RA^-CD90^+$ cells, addition of the wild-type (rs17834140-C) element augmented reporter expression, validating that this element acts as an enhancer in HSCs (Fig. 2E). Mutagenesis to recreate the CHIP-resilience variant (rs17834140-T) alone in otherwise matched constructs reduced the additional enhancer activity relative to the *MSI2* promoter alone by ~2-fold. We observed that rs17834140-T was predicted to break a highly conserved GATA-binding site [combined annotation dependent depletion (CADD) score: 15.6] (Fig. 2F, **left**), which was notable given that GATA2 binds strongly at this enhancer (fig. S2E). In heterozygous donors, GATA2 chromatin immunoprecipitation (ChIP)–polymerase chain reaction (PCR) at rs17834140 demonstrated ~1.6-fold enrichment of GATA2 binding at the wild-type (C) allele over the CHIP-resilience (T) allele (61 vs. 39% reads) (Fig. 2F, **right**). Consistent with this, we observed a depletion of T alleles in ATAC-sequencing reads from HSPCs of heterozygote donors (Fig. 2G). These data demonstrate that rs17834140-T disrupts GATA2 binding and reduces chromatin accessibility, reducing *MSI2* enhancer function in human HSCs.

We next sought to faithfully model resilience variant effects in primary $CD34^+$ human HSPCs. We used CRISPR-Cas9 to minimally excise the putative GATA2-binding site with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

~100-base pair (bp) (ENH-1) or ~300-bp (ENH-2) microdeletions harboring rs17834140 (Fig. 2H). We benchmarked phenotypes against edits targeting the RNA-recognition motif 1 (RRM-1) domain of MSI2 resulting in inactivation (KO). Three days after editing, we observed a high efficiency of microdeletions (~70 to 80%) and MSI2-KO (~90%) (fig. S2, F to I). Regulatory element microdeletions reduced *MSI2* mRNA expression in $CD34^+CD45RA^-CD90^+$ HSPCs relative to safe-harbor *AAVS1*-edited controls (Fig. 2I), selectively within molecularly defined HSCs (fig. S2J) with no impact on other genes at the locus (fig. S2K). Intracellular flow cytometry showed concordant decreases in MSI2 protein abundance in $CD34^+CD45RA^-CD90^+$ HSPCs after ENH-1 (by ~20% versus AAVS1) and ENH-2 (by ~30% versus AAVS1) microdeletions (Fig. 2J). Thus, our functional analyses suggest that rs17834140-T protects from CHIP through HSC-selective downregulation of MSI2.

Genetic variation–driven loss of *MSI2* enhancer reduces human HSC fitness

We next sought to understand how inherited CHIP resilience affects human HSCs (Fig. 3A). Six days after editing, MSI2-KO did not affect the proportion of $CD34^+CD45RA^-$ cells in culture, but reduced phenotypic long-term HSCs (LT-HSCs) ($CD34^+CD45RA^-CD90^+CD133^+EPCR^+ITGA3^+$) principally through loss of primitive $CD90^+$ HSPCs (Fig. 3B), consistent with the role of MSI2 in HSC maintenance (30-33). ENH-1 and ENH-2 microdeletions of the rs17834140-site partially phenocopied this effect, in both proportion (Fig. 3C) and total number (Fig. 3D and fig. S3A) of LT-HSCs. In an orthogonal model, we used cytokine-free and chemically defined cultures of human cord blood $CD34^+$ HSCs (34). In these conditions, after 14 days, we further observed loss of LT-HSC maintenance (Fig. 3E), with total LT-HSC numbers reduced after ENH-1 (by 44% versus AAVS1) or MSI2-KO (by 85% versus AAVS1) editing (Fig. 3F and fig. S3B). Collectively, these data suggest that loss of *MSI2* enhancer function through genetic variation might alter LT-HSC maintenance.

To determine whether the rs17834140-T-harboring regulatory element could modulate in vivo self-renewal and HSC function, we performed xenotransplantation experiments involving NBSGW (Kit-mutant and immunodeficient NOD.Cg-$Kit^{W41J}Tyr^+$ $Prkdc^{scid}Il2rg^{tm1Wjl}$/ThomJ) mice (35, 36). Mice transplanted with ENH-1–edited human cord blood–derived $CD34^+$ cells had reduced total human bone marrow (BM) cellularity (Fig. 3G) and multilineage engraftment of $CD34^+$ HSPCs, $CD19^+$ B cells, and $CD33^+$ myeloid cells (Fig. 3H and fig. S3C) at 16 weeks after transplantation, partially phenocopying a more severe defect in BM reconstitution with MSI2-KO–edited cells. Among engrafted human cells, ENH-1– or MSI2-KO–edited cells had reduced proportions of $CD34^+$ HSPCs and $CD19^+$ B cells in BM and spleen (fig. S3D), representing a loss of HSCs with multilineage potential. There was a greater percentage of $CD33^+$ myeloid cells (fig. S3D) and myeloid colony formation from human $CD34^+$ cells harvested from BM (fig. S3, E and F), suggestive of increased myeloid lineage commitment in vivo, and consistent with the role of MSI2 in promoting symmetric self-renewal over asymmetric differentiative divisions (32, 37). Additionally, ENH-1–edited HSPCs had loss of serial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

replating potential (Fig. 3I), consistent with reduced self-renewal. These data support the idea that CHIP-resilient HSCs have a greater propensity for differentiation at the expense of self-renewal capacity.

MSI2 stem cell RNA network protects HSCs from CHIP and MyMs

Our functional studies led us to investigate mechanisms by which genetic variation–driven down-regulation of MSI2 reduces HSC fitness and protects from CHIP. Because MSI2 is an RNA-binding protein that regulates stability and translation (38) of mRNAs, we sought to define its direct binding targets. Because prior studies were conducted in cell lines (33) or mouse cells (38), the MSI2-regulated network in human HSCs has remained undefined. We therefore used MSI2-HyperTRIBE (38) (Fig. 4A), in which a fusion MSI2-ADAR (RNA-editing enzyme) protein catalyzes A>G edits on MSI2-bound RNA targets, in human cord blood–derived $CD34^+$ HSPCs. MSI2-HyperTRIBE identified 12,928 edits across 3614 genes (fig. S4, A to C, and table S4). MSI2-bound mRNAs were enriched for the UAG motif in their 3' untranslated regions (UTRs) (Fig. 4B), consistent with prior reports (33, 38), and there was strong concordance in bound mRNAs with orthogonal cross-linking and immunoprecipitation sequencing (CLIP-seq) targets of MSI2 (33) (fig. S4, D and E).

We then investigated how reduced MSI2 levels affect its network to confer HSC resilience to CHIP. Single-cell RNA sequencing (scRNA-seq) was performed in human $CD34^+CD45RA^-CD90^+$ cells with AAVS1, ENH-1, or MSI2-KO edits, and analyses were focused on molecularly defined HSCs (fig. S4, F and G). CHIP-resilient HSCs (ENH-1 versus AAVS1) had a 24% loss of *MSI2* expression and 1005 differentially expressed genes (Fig. 4C and table S5) with high concordance to the MSI2-KO signature (Pearson's $r = 0.75$) (fig. S4H), representing pathways that include down-regulation of proliferation, MYC targets, and cholesterol biosynthesis genes (Fig. 4D). There was strong negative enrichment of MSI2-HyperTRIBE targets in ENH-1–microdeleted HSCs with reduced MSI2, suggesting that most MSI2-bound mRNA targets were down-regulated in CHIP-resilient HSCs (Fig. 4E).

To understand direct MSI2-bound targets that are functionally altered in human HSCs, we next overlapped MSI2-HyperTRIBE and scRNA-seq datasets. More highly downregulated genes (by ≥5% in ENH-1 versus AAVS1) and edited MSI2 binding targets (mRNAs with ≥20% editing by MSI2-ADAR) were selected. This identified a network of 208 direct MSI2-binding targets that were also down-regulated in ENH-1– versus AAVS1–edited HSCs, hereafter termed MSI2-DOWN genes (Fig. 4F and table S6), whose expression was positively correlated with interindividual variation in *MSI2* levels in HSCs among 131 donors (fig. S4I) (39). This network was enriched for functional MSI2 targets [such as *TSPAN3* (40), *HSP90AA1* (33), *SYNCRIP* (41), and *CDK6* (42)], validating our approach. In addition, the MSI2 network revealed positive regulators of human HSCs [including *TFRC* (43, 44), *DKC1* (45), and *H3-3B* (46)] with no previously studied role in CHIP, nominating candidate regulators of stem cell clonality downstream of MSI2. We leveraged ribosome profiling in primary human $CD34^+CD45RA^-CD90^+$ HSCs (36) and found that the transcripts of MSI2-DOWN CHIP-resilience network genes were enriched for higher numbers of ribosome protected fragments in comparison to other transcripts (Fig. 4G). This

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

indicated that the RNA network positively regulated by MSI2 is among the more highly translated mRNAs in human HSCs, validating its functionality.

Having identified an RNA network underlying CHIP resilience in HSCs, we sought to understand its clinical relevance. We leveraged TARGET-seq (joint single-cell genotyping and RNA-seq) data of CHIP mutant versus wild-type HSCs from the same donors, enabling insight into the transcriptional impact of CHIP mutations in an isogenic setting (47). The MSI2-DOWN network was up-regulated in *TET2* mutant HSCs in comparison to their wild-type counterparts (Fig. 4H). Furthermore, the MSI2-DOWN network was up-regulated in *ASXL1* mutant versus nonmutant HSPCs of patients with chronic myelomonocytic leukemia (fig. S4J) (48), and ENH-1–edited HSCs were negatively enriched for functional dependencies of *Dnmt3a*$^{R882H/+}$ mutant HSPCs (fig. S4K) (49). To explore the prognostic value of the MSI2-DOWN network in MyMs, we interrogated the TARGET-AML and BEAT-AML cohorts (50, 51). Elevated expression of either *MSI2* or its downstream network at diagnosis predicted a poorer prognosis in pediatric AML (Fig. 4I and fig. S4L) and was further associated with a more primitive and therapy-resistant adult AML subtype (Fig. 4J and fig. S4M). Collectively, we show that the MSI2-DOWN CHIP-resilience network is enriched for positive regulators of HSC proliferation that may promote stem cell clonal advantage in CHIP and MyMs, suggesting that inherited down-regulation of this network might neutralize the impact of acquired CHIP mutations.

MSI2 levels modify clonal dominance of *ASXL1*-mutant HSCs

Finally, we sought to understand how altered MSI2 levels in human HSCs affect the progression of CHIP by analyzing a cohort with targeted sequencing of two serial blood samples taken ~6 years apart (Fig. 5A). Among 513 individuals with persistent CHIP [second variant allele frequency (VAF) also ≥2%], rs17834140-T carriers had a significantly slower measured growth rate of mutant CHIP clones ($P < 0.001$), with many clones showing regression or minimal growth (median growth rates: C/C = 0.12%/year, C/T = –0.31%/year) (Fig. 5B). Indeed, rs17834140-T carriers had greater odds of their CHIP mutation being transient (second VAF <2%) (OR = 1.82, 95% CI = 1.29 to 2.52, $P = 0.0007$) (Fig. 5C), which was most pronounced for *ASXL1*-CHIP (OR = 2.72, 95% CI = 1.00 to 7.33, $P = 0.057$). Together, these data suggest that rs17834140-T attenuates CHIP progression.

To validate MSI2 levels as a modifier of CHIP, we focused on *ASXL1*-mutant CHIP, a common driver in ~20% of myelodysplastic syndrome (MDS) and AML cases, and for which rs17834140-T carried among the most protective associations. We modeled patient *ASXL1* mutations (fig. S5A) in primary human CD34$^+$ HSPCs using an exon 12–targeting guide RNA that increases phenotypic LT-HSC maintenance in culture (fig. S5B) and recapitulates clonal expansion in a xenograft model (52). HSPCs were edited to model germline CHIP resilience (80 to 100% editing) and acquired *ASXL1* mutations (91 to 95% editing), with most alleles coedited at both loci (fig. S5C). Whereas *ASXL1* editing expanded LT-HSCs in ex vivo cultures by 1.9-fold, a background of ENH-1 or MSI2-KO editing reduced or abolished the net fitness advantage, respectively (Fig. 5D), and reversed an activated HSC signature induced by *ASXL1* mutation (fig. S5, D and E). Consistent with this, ENH-1 editing reduced *ASXL1* mutation-driven clonal expansion as assessed by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

serial colony replating (Fig. 5E and fig. S5F). In vivo, engrafted *ASXL1*-mutant HSPCs displayed expansion of primitive $CD90^+$ HSPCs (Fig. 5F), as well as increased $CD117/KIT^+$ (fig. S5G) and $CD33^+$ cells, consistent with myeloid skewing (Fig. 5G); all of these effects were attenuated under a background of ENH-1 editing. These data show how inherited CHIP resilience that reduces MSI2 attenuates the fitness advantage of acquired *ASXL1* mutations in HSCs.

Conversely, stem cell MSI2 levels are commonly up-regulated in MyMs and drive aggressive disease with poor prognosis (53, 54). Therefore, we generated a mouse model to probe the combined impact of MSI2 overexpression and *Asxl1* mutant CHIP (Fig. 5H and fig. S6A). Whereas MSI2 overexpression or *Asxl1* deletion alone had minimal impact on the frequency of HSC-enriched $Lin^-Sca\text{-}1^+c\text{-}Kit^+$ (LSK) cells, mice transplanted with $Asxl1^{\Delta/\Delta}$+$MSI2^{DOX}$–expressing cells had a 1.5-fold average increase in LSK cell frequency (Fig. 5I). Over 8 months, whereas $Asxl1^{\Delta/\Delta}$ and $MSI2^{DOX}$ mice did not show evidence of hematologic disease, $Asxl1^{\Delta/\Delta}$+$MSI2^{DOX}$ mice developed reduced peripheral leukocyte and red blood cell counts, along with other blood trait alterations (fig. S6B), and blood smears showed MDS features that included hyposegmented Pelger-Huët-like neutrophils and binucleate erythroid precursors (Fig. 5J and fig. S6C). Taken together, our data suggest a model in which MSI2 levels in HSCs functionally interact with CHIP and MyM driver mutations by modifying their clonal advantage both before and after somatic mutation acquisition (Fig. 5K).

Discussion

Genetic variation provides opportunities to uncover natural mechanisms for disease resilience. Notable examples include loss-of-function variants in *CCR5* conferring HIV resistance (55), *PCSK9* protecting against cardiovascular disease (56), and *BCL11A* suppression ameliorating hemoglobin disorders (57, 58), insights that have guided therapeutic development. Although much has been learned about cancer predisposition, inherited cancer resilience is largely unexplored. Because purifying selection and rarity of loss-of-function variants limit discovery (59), we reasoned that large population biobanks may uncover such protective variation. In this study, we identified a protective mechanism against CHIP–an exemplar of somatic mosaicism and precursor to MyMs–through selective genetic down-regulation of the RNA-binding protein MSI2 in HSCs.

Our studies provide critical insight into RNA networks that underlie HSC maintenance and clonal advantage in CHIP. RNA-binding proteins are key posttranscriptional regulators of self-renewal (60), and although previous work established MSI2 as a regulator of HSC numbers in vivo (32, 33, 37), its downstream mechanisms and natural genetic modulation have remained unknown. We mapped an MSI2-regulated network in human HSCs, the systematic perturbation of which is likely to deepen our understanding of mechanisms involved in HSC maintenance and the fitness advantage in CHIP. This may also be of broader relevance in MSI2-driven lymphoid malignancies (42, 61, 62) and solid tumors (63-66).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We show that partial and selective down-regulation of MSI2 in HSCs is adaptive (Fig. 5K), without other apparent adverse implications in variant carriers. Together, our results support two (non-mutually exclusive) models, in which reduced MSI2 levels may (i) reduce the baseline HSC pool (lowering the likelihood of any cell acquiring a CHIP driver mutation) or (ii) attenuate the clonal advantage of drivers once acquired. Our study highlights the potential to target MSI2, through small-molecule inhibition or genome editing at its enhancer, for blood cancer prevention. More broadly, we provide an example of how resilience to cancer can arise through inherited genetic variation, motivating the search for other natural pathways that could be leveraged to prevent or treat malignancy.

Materials and Methods

All Of Us and GWAS meta-analysis

All of Us (AoU) is a longitudinal cohort containing short-read whole genome sequencing data (mean depth of 30x). Informed consent is in place for all AoU participants, and the protocol was reviewed by the Institutional Review Board (IRB) of the AoU Research Program. Variant-level QC metrics included QUAL > 60, ExcessHet < 54.69, GQ > 20, DP > 10, AB > 0.2 for heterozygotes. Variants with population-specific allele frequency (AF) > 1% or population-specific allele count (AC) > 100 were selected. Sites with more than 100 distinct alternate alleles were excluded.

Individual-level CHIP calls in AoU were ascertained, as previously described (68). Only samples with matching reported sex and genetically inferred sex were included, and analyses were performed on individuals of European ancestry with no diagnosis of hematological malignancy (using filtering criteria as previously reported (68)). PLINK (v1.9) was used to perform logistic regression with Firth correction to assess associations with CHIP for the most significant sentinel SNP at 24 loci previously reported (11). An inverse variance-weighted fixed-effect meta-analysis was performed across AoU associations and summary statistics in UK Biobank (UKB) and Geisinger Health Study (GHS) cohorts (11). Additionally, the effect of the 17q22 CHIP-resilience haplotype (tagged by sentinel SNP rs80093687) on prevalence of specific CHIP driver mutations (*DNMT3A, TET2, ASXL1, JAK2*) was evaluated as a meta-analysis (across AoU and UKB). All analyses in AoU were adjusted for age, sex at birth and the first 5 genetic principal components as covariates.

For associations with myeloid malignancy (MyM) risk, effect estimates for the 17q22 CHIP-resilience haplotype were extracted from GWAS summary statistics for myeloproliferative neoplasm (MPN) [3,797 cases and 1,152,977 controls (10)], acute myeloid leukemia (AML) [4,018 cases and 10,488 controls (69)] and myelodysplastic syndromes (MDS) [907 cases and 5,604 controls (70)]. Phenome-wide association study (PheWAS) summary statistics were previously reported (11), and plotted using the PheWAS package in R.

Fine-mapping and variant annotation

Statistical fine-mapping was performed using GWAS summary statistics for CHIP in the UKB European ancestry cohort (11) [GWAS catalogue: GCST90165267], and linkage disequilibrium (LD) matrices for UKB participants of British ancestry (71). Summary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

statistics were converted from hg38 to hg19 using LiftOver. Variants with minor allele frequency (MAF) > 0.001 within 1.5 Mb of the sentinel SNP rs80093687 (7,012 in total) were fine-mapping using SuSIE (72), CARMA (73) and FINEMAP (74), allowing up to 10 causal variants and reporting 95% credible sets for each method. All methods identified the same six variants (rs80093687, rs188761458, rs199691861, rs118121072, rs17834140, rs150497606) in a single 95% credible set.

Fine-mapped variants and those in strong LD with rs80093687 ($R^2 \geq 0.8$, ± 1.5 Mb) were annotated with genomic context, splicing predictions (using SpliceAI (75)), and chromatin accessibility across 13 hematopoietic cell types (76). Normalized mRNA expression for all genes within 1 Mb of rs17834140 was extracted from a single-cell multiome dataset of human HSCs (old-1 and old-2 donors (4)). For ATAC-RNA correlation, hematopoietic cell types were clustered into N=75 'metacells' to address single-cell sparsity, and Pearson's correlation was computed between mean *MSI2* expression and chromatin accessibility at the rs17834140-harboring regulatory element. H3K27ac-HiChIP and ChIP-seq data (LYL1, GATA2, TAL1, LMO2, RUNX1, ERG, FLI1, PU.1) in $CD34^+CD45RA^-$ HSPCs were previously reported (29). Transcription factor motif analysis was conducted using motifBreakR (v2.16) (77), with HOCOMOCO v11 matrices and a statistical threshold of $p<1\times10^{-3}$. Allelic skewing in ATAC-seq data of primary human $CD34^+$ HSPCs was assessed in three heterozygotes for rs17834140 (GEO accessions GSE194122, GSE219015 and GSE156733). Raw FASTQ files were aligned to hg38 using bwa (v0.7.18), BAM files were generated with samtools (v1.20), and bcftools mpileup was used to call 'C' and 'T' alleles at rs17834140.

Primary cell culture

Primary human $CD34^+$ HSPCs from mobilized peripheral blood of healthy donors were obtained from the Fred Hutchinson Cancer Research Center (donor IDs available on GitHub repository, at https://github.com/sankaranlab/chip_resilience). Thawed cells were cultured at $5x10^5$ cells/mL in serum-free StemSpan SFEM II medium (StemCell Technologies) supplemented with 1% L-glutamine (ThermoFisher Scientific, 25-030-081), 1X penicillin/streptomycin (Life Technologies, 15140-122), 1X CC100 (containing the cytokines FLT3L, SCF, IL-3 and IL-6; StemCell Technologies, 02690), 100 ng/mL recombinant thrombopoietin (TPO; PeproTech, 300-18), and 35 nM UM171 (StemCell Technologies).

$CD34^+$ HSPCs from cord blood of healthy newborns (Pasquarello Tissue Bank, Dana-Farber Cancer Institute) were purified using the EasySep Human Cord Blood $CD34^+$ positive selection kit (StemCell Technologies) according to the manufacturer's instructions. For cytokine-free culture (34), cord blood-derived $CD34^+$ HSPCs were cultured at 7×10^4 to 1×10^5 cells/mL in Iscove's Modified Dulbecco's Medium (IMDM; Life Technologies), supplemented with 1% L-glutamine (Thermo Fisher Scientific), 1% penicillin/streptomycin (Life Technologies), 1% insulin-transferrin-selenium-ethanolamine (ITSX; Life Technologies), 1 mg/ml polyvinyl alcohol (PVA; Sigma-Aldrich), 1 μM 740Y-P (MedChemExpress), 0.1 μM butyzamide (MedChemExpress), and 70 nM UM171.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CRISPR interference

For CRISPR interference, dCas9-KRAB (from Addgene plasmid #220838) was subcloned into the backbone of Addgene plasmid #204472, and in vitro transcription was performed to make purified mRNA encoding dCas9-KRAB. On day 2 of culture, HSPCs were washed two times in DPBS, and resuspended in 20 µL of Lonza P3 buffer with supplement. For each nucleofection, HSPCs were mixed with 2 µg dCas9-KRAB mRNA and total 2 µL sgRNA targeting either the *AAVS1* safe-harbor locus or the rs17834140-harboring regulatory element with independent pairs of sgRNAs [*AAVS1*: 1 µL each of sgAAVS1_g1 and sgAAVS1_g2; CRISPRi-ENH-1: 1 µL each of sgENH-1_g1 and sgENH-1_g2; CRISPRi-ENH-2: 1 µL each of sgENH-2_g1 and sgENH-2_g2 sgRNAs] (all sgRNA sequences detailed in Table S7). HSPCs were electroporated in 20 µL Nucleocuvette strips using the Lonza 4D Nucleofector, with the DS-130 program. Immediately after electroporation, 80 µL of prewarmed media were added to the electroporation cuvette, which was placed in an incubator at 37 °C for 5 mins. Cells were then plated at a density of $5x10^5$ cells/mL in adequate complete media. 72 hrs after nucleofection, HSPCs were sorted to enrich for HSCs ($DAPI^{neg}CD34^+CD45RA^-CD90^+$). Subsequently, total RNA was extracted, cDNA was generated, and real-time quantitative PCR (RT-qPCR) was performed to detect *MSI2* expression (as described below).

Lentiviral packaging and quantification

Lentiviral reporter constructs (as previously described (78)) carrying the enhancer element with wild-type (C) and resilience (T) alleles of rs17834140 [hg38: chr17:57,388,104-57,388,632] were cloned upstream of the human *MSI2* promoter driving GFP reporter expression. The *MSI2* promoter element was synthesized as a g-block [hg38: chr17:57,256,155-57,256,741] (IDT Technologies). For lentiviral packaging, HEK-293T cells were cultured at 37 °C in DMEM (Life Technologies) supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were plated into 15 cm^2 plates and grown to ~70% confluency on the day of transfection per lentiviral construct. For each plate, 15 µg of psPAX2 packaging plasmid, 7.5 µg of pMD2.G envelope plasmid and 15 µg of construct were added in the presence of Opti-MEM media (Gibco, 31985-062) and Lipofectamine 3000 Transfection Reagent (Invitrogen, L3000001). Viral supernatants were harvested twice at 48 and 72 hrs post-transfection, and concentrated by ultracentrifugation (24,000 rpm for 2 hrs at 4 °C) in the SW32 rotor of Beckman Coulter ultracentrifuge. After ultracentrifugation, the supernatant was decanted and viral pellets were resuspended in ice cold PBS with 0.1% BSA, and frozen at −80 °C till further use. Lentiviral constructs were titrated using the Lenti-X™ Provirus Quantitation Kit (Takara, 631239) in K562 cells (cultured at 37 °C in IMDM supplemented with 10% FBS and 1% penicillin/streptomycin), as per manufacturer's instructions.

Reporter assay

For lentiviral transduction, concentrated virus was added to $CD34^+$ HSPCs on day 1 of culture at equal titres, in the presence of 8 µg/mL cyclosporin H (Sigma-Aldrich, SML1575). HSPCs were then spinfected at 2,000 rpm for 90 mins at 37 °C. 72 hrs after infection, HSPCs were harvested, washed twice with PBS, and incubated for 30

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mins with different fluorescent-labelled antibodies: 1:40 dilution of anti-human CD34 PerCP-Cyanine5.5 (clone 561, BioLegend, #343612), 1:50 dilution of anti-human CD45RA APC/Cy7 (clone HI100, BioLegend, #304128) and 1:100 dilution of anti-human CD90 PE (clone 5E10, BioLegend, #328110). Immunophenotypic analyses were performed to quantify geometric mean fluorescence intensity (MFI) of GFP^+ transduced cells within $CD34^+CD45RA^-CD90^+$ gated HSC-enriched cells for each lentiviral construct. Analyses were performed on LSRFortessa (BD Biosciences). Data were analyzed using the FlowJo software.

GATA2 ChIP

Chromatin Immunoprecipitation (ChIP) was performed on chromatin from $10–15\times10^6$ primary human $CD34^+$ cells, from mobilized peripheral blood of healthy adults or from cord blood of healthy newborns. On day 5 of serum-free culture, cells were cross-linked with 1% methanol-free formaldehyde (Pierce Life Technologies, 28906) and quenched with glycine. Chromatin was processed according to the iDeal ChIP-seq kit for transcription factors (Diagenode, C01010170) with the following modifications. Lysates were sonicated using an E220 sonicator (Covaris, 500239) in microtube AFA Fiber Pre-Slit Snap-Cap tubes (Covaris, 520045) under these conditions: peak incident power = 165 W, duty factor = 2%, cycles per burst = 200, treatment time = 210 s. Sheared chromatin was immunoprecipitated with 5 μg of either anti-GATA2 antibody (clone EPR2822(2), Abcam, ab109241) or control IgG (clone RIG002B, Diagenode, C15410206), and DNA was eluted in 30 μL nuclease-free water. Validation of GATA2 binding was performed by ChIP-qPCR using primers for: *MSI2* enhancer (forward: TTTCCAAAGACCACTGCGCT, reverse: GAAACACCCCAGGAGCAAGA), or negative control myoglobin exon 2 (Diagenode, C17011006-50).

Allele-specific ChIP at rs17834140 was performed as above using cord blood $CD34^+$ cells from two healthy newborn donors confirmed to be heterozygous for rs17834140 (C/T) by Sanger sequencing. Following ChIP, a 96 bp amplicon spanning rs17834140 was obtained by PCR (primers as above) and next-generation sequencing (Premium PCR Sequencing, Plasmidsaurus; Oxford Nanopore Technology) was performed. Raw FASTQ files were aligned to hg38 using bwa (v0.7.18), BAM files were generated with samtools (v1.20), and bcftools mpileup was used to quantify skewing of ‘C’ and ‘T’ alleles at rs17834140.

CRISPR/Cas9 RNP nucleofection and editing analysis

Electroporation was performed on either day 2 (for $CD34^+$ from mobilized peripheral blood of healthy donors) or day 3 (for cord blood $CD34^+$ from healthy newborns) of culture, using the Lonza 4D Nucleofector with 20 μL Nucleocuvette strips. The Cas9 ribonucleoprotein (RNP) complexes were prepared by combining 2.1 μL of Lonza P3 primary cell nucleofection reagent (Lonza, V4XP-3032), 1.7 μL of 62μM Alt-R S.p. HiFi Cas9 Nuclease V3 (IDT, 1081061) and total 1.0 μL of 100 μM sgRNA in IDTE pH 7.5 (IDT) [*AAVS1*: 0.5 μL each of sgAAVS1_g1 and sgAAVS1_g2; ENH-1 microdeletion: 0.5 μL each of sgENH-1_g1 and sgENH-1_g2; ENH-2 microdeletion: 0.5 μL each of sgENH-2_g1 and sgENH-2_g2; MSI2-KO: 1 μL of sgMSI2-KO]. HSPCs were washed two times in DPBS, then resuspended in 20 μL of Lonza P3 buffer with supplement (in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

presence of nucleofection enhancer). Cells were electroporated using the DZ-100 program in a 4D-Nucleofector X Unit (20 μL cuvettes). Immediately after electroporation, 80 μL of prewarmed media were added to the electroporation cuvette, which was placed in an incubator at 37 °C for 5 mins. Cells were then plated at a density of $5x10^5$ cells/mL in adequate complete media.

For analysis of DNA editing, $CD34^+CD45RA^-CD90^+$ HSC-enriched cells were harvested for genomic DNA extraction at least 72 hrs post-nucleofection. PCR fragments flanking the editing site (at least 250 bp upstream and downstream) were amplified (with primer sequences detailed in Table S7) and editing frequencies and outcomes were detected by either Sanger sequencing (and analyzed with the ICE analysis tool from Synthego) or next-generation sequencing (NGS). For NGS, Premium PCR Sequencing was performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation, and analysis was performed using CRISPResso2 (79).

Real-time quantitative PCR (RT-qPCR) for MSI2 Expression

Total RNA was obtained on day 5 of HSPC culture (72 hrs following CRISPR/Cas9 editing or dCas9-KRAB targeting) using the Total RNA Purification Micro Kit (Norgen, 35300) as per manufacturer's instructions, including DNase I digestion (Norgen, 25710). 100 to 500 ng of total RNA was used for reverse transcription using the iScript cDNA synthesis (BioRad, # 1708890) or PrimeScript™ RT Master Mix (Takara, RR036A). The cDNA product was used for real-time PCR analysis using iQ SYBR green supermix (BioRad, #1708882), with primers specific to *MSI2* or *GAPDH* (sequences detailed in Table S7). Three technical replicates were performed for each sample, and the mean value was selected for further analysis. The relative expression of each target gene was first normalized to *GAPDH* housekeeping gene expression and then represented as fold changes (2^{-ddCt}) relative to the indicated control conditions.

Intracellular flow cytometry

MSI2 intracellular protein abundance was quantified on day 5 of culture (72 hrs after editing). HSPCs were harvested, washed once in DPBS, and incubated for 30 mins with different fluorescent-labelled antibodies: 1:40 dilution of anti-human CD34 PerCP-Cyanine5.5 (clone 561, BioLegend, #343612), 1:50 dilution of anti-human CD45RA APC/Cy7 (clone HI100, BioLegend, #304128) and 1:100 dilution of anti-human CD90 PE (clone 5E10, BioLegend, #328110). Cells were washed, fixed with 4% paraformaldehyde for 15 mins at RT, washed, permeabilized with 0.2% Tween 20 for 15 mins at RT and washed again. Next, cells were then stained with a 1:100 dilution of rabbit anti-MSI2 monoclonal antibody (clone EP1305Y, Abcam, ab76148) for 30 mins at RT, washed, then incubated with a 1:2,000 dilution of goat anti-rabbit AlexaFluor-488 antibody (Invitrogen, A-11034) for 30 mins at RT and washed. Immunophenotypic analyses were performed to quantify the geometric mean fluorescence intensity (MFI) of GFP within $CD34^+CD45RA^-CD90^+$ gated HSC-enriched cells. Analyses were performed on LSRFortessa (BD Biosciences). Data were analyzed using the FlowJo software.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Flow cytometry and cell sorting

Immunophenotypic analyses for long-term (LT)-HSCs were performed on adult $CD34^+$ HSPCs (on day 8 of serum-free culture, 6 days after editing) or cord blood $CD34^+$ HSPCs (on day 14 of cytokine-free culture, 11 days after editing). Cells were harvested, washed and incubated for 30 mins with fluorescent-labelled antibodies: 1:40 dilution of anti-human CD34 PerCP-Cyanine5.5 (clone 561, BioLegend, #343612), 1:50 dilution of anti-human CD45RA AlexaFluor-488 (clone HI100, BioLegend, #304114), 1:100 dilution of anti-human CD90 PE/Cy7 (clone 5E10, BD Biosciences, # 561558), 1:100 dilution of anti-human CD201 (EPCR) PE (clone RCR-401, BioLegend, #351904), 1:33 dilution of anti-human CD49c (ITGA3) APC (clone ASC-1, BioLegend, #343808). Adult-derived HSPCs in serum-free culture were additionally stained with 1:20 dilution of anti-human CD133 Super Bright 436 (clone 7, BioLegend, #372808). All analyses were performed on LSRFortessa (BD Biosciences). Data were analyzed using the FlowJo software. Total LT-HSC numbers were calculated as a product of the frequency of LT-HSCs by flow cytometry and total cell number in culture.

Fluorescence-activated cell sorting (FACS) was conducted to enrich for HSCs, for DNA editing analysis, RT-PCR of *MSI2* expression, and for single-cell and bulk RNA-sequencing experiments. FACSymphony S6 Cell Sorter (BD Biosciences) was used to sort with a 100 μm nozzle for viable cells (negative for DAPI) with surface expression of $CD34^+CD45RA^-CD90^+$ markers.

Xenotransplantation and animal models

All animal procedures were performed under a protocol approved by the Boston Children's Hospital Institutional Animal Care and Use Committee (IACUC). $CD34^+$ HSPCs purified from human cord blood of healthy donors were edited on day 2 of culture (as indicated below), and cultured for 48 hr. On day 4 of culture, flow cytometry was used to assess % $CD34^+$ in culture for each condition, and 200,000 $CD34^+$ cells per mouse were injected via tail vein into Kit-mutant and immunodeficient NOD.Cg-Kit^{W41J}Tyr+ $Prkdc^{scid}Il2rg^{tm1Wjl}$/ThomJ (NBSGW) mice (JAX#026622). To prevent infections, the mice were provided with autoclaved sulfatrim antibiotic water, which was changed weekly.

For experiments assessing impact of CHIP-resilience on HSC function, HSPCs were edited at AAVS1, ENH-1 or MSI2-KO. Animals were euthanized at 16 weeks post-transplantation, and their bone marrows (BMs) and spleens were collected for analysis. BM cells were obtained by flushing the bilateral femurs and tibias, while spleens were carefully minced. Cell suspensions were counted prior to further analyses. For immunophenotypic analyses, cells were stained with the following antibodies: 1:200 dilution of anti-human CD45 APC (clone HI30, BioLegend, #304037), 1:400 dilution of anti-mouse CD45 FITC (clone 30-F11, BioLegend, #103018), 1:400 dilution of anti-human CD3 BV786 (clone SK7, Fisher Scientific, BDB563800), 1:100 dilution of anti-human CD19 PerCP-Cyanine5.5 (clone HIB19, BD Biosciences, #561295), 1:100 dilution of anti-human CD34 APC-Cyanine7 (clone 561, BioLegend, #343614, 1:100 dilution of anti-human CD33 PE (clone HIM3-4, Invitrogen, #12-0339-42), and 1:100 dilution of anti-human CD15 AlexaFluor700 (clone HI98, BioLegend, #301920).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For experiments assessing how CHIP-resilience modifies the effect of *ASXL1* mutation, HSPCs were editing (as detailed under 'Modeling human *ASXL1* mutant CHIP' section below) at: i) sgAAVS1_g1+sgAAVS1_g2; ii) sgAAVS1_g1+sgASXL1-ex12; iii) ENH-1+AAVS1_g2; iv) ENH-1+sgASXL1-ex12). Animals were euthanized at 9 weeks post-transplantation, and their BMs were collected for analysis. BM cells were obtained by flushing the bilateral femurs and tibias. Cell suspensions were counted prior to further analyses. For immunophenotypic analyses of lineages, cells were stained with the following antibodies: 1:200 dilution of anti-human CD45 APC (clone HI30, BioLegend, #304037), 1:400 dilution of anti-mouse CD45 FITC (clone 30-F11, BioLegend, #103018), 1:400 dilution of anti-human CD3 BV786 (clone SK7, Fisher Scientific, BDB563800), 1:100 dilution of anti-human CD19 PE-Cyanine5 (clone HIB19, BD Biosciences, #555414), 1:100 dilution of anti-human CD117 PE (clone 104D2, BioLegend, #313204, 1:100 dilution of anti-human CD33 Super Bright 436 (clone HIM3-4, Invitrogen, #62-0339-42), and 1:100 dilution of anti-human CD11b PE/Cy7 (clone ICRF44, BioLegend, #301322). For immunophenotypic analyses of primitive HSPCs, cells were stained with the following antibodies: 1:200 dilution of anti-human CD45 APC (clone HI30, BioLegend, #304037), 1:40 dilution of anti-human CD34 BV421 (clone 561, BioLegend, #343610), 1:50 dilution of anti-human CD45RA AlexaFluor-488 (clone HI100, BioLegend, #304114), and 1:100 dilution of anti-human CD90 PE (clone 5E10, BioLegend, #328110). Analyses were performed on LSRFortessa (BD Biosciences). Data were analyzed using the FlowJo software.

Colony-forming unit cell assay

Three days post-nucleofection, edited $CD34^+$ HSPCs were plated at a density of 500 cells/mL in methylcellulose medium (H4034, Stem Cell Technologies) according to manufacturer's instructions. Cultures were incubated for 10 days at 37 °C with 5% CO_2 prior to colony quantification. For serial replating, cells from primary colonies were harvested and replated at 10,000 cells/mL (equalized between conditions). Secondary colonies were counted after 8-10 days.

For CFU assays from long-term xenotransplant recipients, human $CD34^+$ cells were enriched from mouse bone marrow 16 weeks post-transplant using the Human CD34 MicroBead Kit (Miltenyi Biotec, 130-046-702). From each mouse, 5,000 $CD34^+$ HSPCs were plated in 1 mL methylcellulose medium (H4434, StemCell Technologies) according to the manufacturer's instructions, and colonies were counted 14 days post-plating.

MSI2-HyperTRIBE

$CD34^+$ HSPCs from cord blood were purified using the CD34 MicroBead Kit (Miltenyi Biotec, 130-046-703) according to the manufacturer's instructions. Purified $CD34^+$ cells were cultured for 2 days in IMDM supplemented with 20 % BIT 9500 (StemCell Technologies, #09500), 1X penicillin/streptomycin (Corning, 30-003-CI), and cytokines including 100 ng/mL Human SCF (Peprotech, 300-07-10UG), 10 ng/mL hFLT3L (Peprotech, 300-19-10UG), 100 ng/mL hTPO (Peprotech, 300-18-10UG), and 20 ng/mL hIL-6 (Peprotech, 200-06-20UG). Cells were then transduced with high-titer, concentrated retroviral suspensions encoding either MIG (empty vector control) or MSI2-ADAR

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

constructs in the presence of 8 μg/mL polybrene (Millipore, TR1003G), and spinfected at 700 g for 1.5 hrs. A second round of transduction was performed the following day. GFP-positive cells were sorted using a BD FACSAria cell sorter 24 hrs after second transduction. Three independent biological replicates were performed. RNA was extracted from bulk sorted cells using the SMARTer RNA extraction method. cDNA was subjected to automated paired-end library construction for sequencing on an Illumina HiSeq 2000 platform (PE100 read length). Sequencing was performed at a depth of 30–40 million reads per sample.

MSI2-ADAR editing site calling

Sequencing reads from three MSI2-ADAR and three control (empty vector) samples were aligned to the human T2T reference genome chm13v2.0 (80) using minimap2 (v.2.26) (81). All possible single-nucleotide mismatches were piled up with cellsnp-lite (v1.2.3) (82) in mode 2b in each sample individually, requiring a minimum coverage of 10 and a minor allele frequency of at least 0.01. Mismatches were annotated against the GENCODE v35 GTF to determine their overlap with gene regions. Those annotated as common or rare single-nucleotide polymorphisms (SNPs) in human dbSNP (v155) were excluded from further analysis. Because editing sites in MSI2-ADAR samples may not appear in controls, where they remain unedited and do not appear as mismatches, bcftools (v1.20) (83) was used to extract allele counts for any mismatch present in at least two MSI2-ADAR samples, across all six samples. We retained only A>G mismatches on the forward strand or T>C mismatches on the reverse strand in control samples to ensure that A or T was the major allele in the control.

Allele counts of retained mismatches were modelled using a negative binomial generalized linear model (GLM) with a log link (via MASS::glm.nb). Specifically, for each mismatch i:

$$k_i \sim \text{NegBinom}(\mu_i, \theta), \log(\mu_i) = \beta_0 + \beta_1 \text{group}_i + \log(n_i)$$

where k_i is the minor allele count, μ_i is the expected count of k_i, θ is the dispersion parameter estimated by the model, n_i is coverage (with $\log(n_i)$ serving as an offset), and group_i indicates whether the sample is MSI2-ADAR or control. P-values of the MSI2-ADAR group coefficient β_1 were adjusted with the Benjamini-Hochberg method. A smaller p-value reflects a larger difference in editing frequency between MSI2-ADAR and control samples. To estimate editing frequency in each group, we regressed the observed edited counts (k_i) on the total coverage (n_i) without an intercept: $k_i \sim 0 + n_i$. Ultimately, we kept mismatches that met all of the following criteria: 1) the difference in editing frequencies between MSI2-ADAR and control samples was ≥0.1; 2) the editing frequency in control samples was <0.02; 3) FDR < 0.05 in the GLM test.

The $CD34^+$ MSI2-HyperTRIBE score was calculated by taking the sum of total differential editing frequencies per gene (for instance, if a target mRNA was edited at two separate nucleotides in 50% and 70% of reads, then the total score was 1.2). For orthogonal validation, targets and edit sites were presented for MSI2-CLIP-seq in hematopoietic cells reported previously (33). To identify motifs surrounding MSI2-ADAR editing sites, we used

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a similar strategy as previously reported (38). Briefly, sequences extending ±100 bp of each editing site in the 3′ UTR were extracted, and overlapping regions were merged. As a background set, 201-bp segments were randomly selected from 3′ UTRs in the reference genome that did not overlap with the target sequence pool. The HOMER2 software (84) was used to search for enriched RNA motifs of length 6, 7, or 8.

Single-cell RNA sequencing and analysis

Droplet-based digital 3'-end single cell RNA sequencing (scRNA-seq) was performed on a Chromium Single-Cell Controller (10X Genomics) using the Chromium Next GEM Single Cell 3' Reagent Kit v3.1 according to the manufacturer's instructions. $CD34^+$ HSPCs derived from mobilized peripheral blood of 2 distinct healthy donors were edited at AAVS1, with ENH-1 microdeletion or MSI2-KO, on day 2 of culture. 72 hrs after editing, cells were sorted with a 100 μm nozzle for viable cells (DAPI negative) and $CD34^+CD45RA^-CD90^+$ cell-surface markers. Approximately $1.6x10^4$ viable cells from each sample were utilized for subsequent scRNA-seq processing, with an estimated recovery of 8,000-12,000 cells per condition. Briefly, single cells were partitioned in Gel Beads in Emulsion (GEMs) and lysed, followed by RNA barcoding, reverse transcription, and PCR amplification (11 cycles). scRNA-Seq libraries were prepared according to the manufacturer's instructions, checked, and quantified on BioAnalyzer instrument. Sequencing was performed on a Nova Seq S2 (Illumina).

The raw scRNA-seq FASTQ files were processed with the CellRanger (v8.0.1) pipeline to map in the reference genome (GRCh38). We excluded cells with unique molecular identifier (UMI) counts less than 2,000 or greater than 25,000, or mitochondrial UMI fraction higher than 20%, and removed potential doublets by a threshold of doublet score > 0.25 using ScrubletR, which resulted in a total of 57,412 cells for AAVS1 (donor 1 = 10,059 and donor 2 = 12,399), ENH-1 (donor 1 = 7,281 and donor 2 = 8,345), and MSI2-KO (donor 1 = 11,330 and donor 2 = 7,998) edited cells. Symphony R package was used to project the cells on a human BM reference (https://github.com/andygxzeng/BoneMarrowMap) (85), to annotate molecularly defined HSCs. A standard Seurat framework (v4.4.0) was used to conduct normalization, principal component analysis (PCA), and dimensionality reduction. The feature-barcode matrix was normalized by the total read count and log-transformed, and the top 3,000 variable features were selected by the vst method in the FindVariableFeatures function. The normalized expression was scaled by Seurat's ScaleData function, and ribosomal (RPS/RPL) and mitochondrial (MT-) genes were filtered. PCA was performed using the RunPCA function (npc = 30). The sample-dependent technical variation was corrected by using Harmony (86). Uniform Manifold Approximation and Projection (UMAP) was conducted to reduce dimensions to embed the cells into two-dimensional space. Seurat's FindMarkers function using "wilcox" method was applied within molecularly defined HSCs to identify differentially expressed genes between ENH-1 vs. AAVS1-edited cells, or MSI2-KO vs. AAVS1-edited cells, with a significance threshold of Benjamini & Hochberg (BH)-adjusted $P < 0.05$, an absolute change in relative expression of ≥ 5%, and minimum percent of expressed cells ≥ 10%. Gene set enrichment analysis was performed using the fGSEA package (https://github.com/ctlab/fgsea/) using the 2024 Hallmark gene sets. Figures were generated using R (v4.4).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Mapping and validating CHIP-resilience network

The MSI2-DOWN CHIP-resilience network (total 208 genes) was defined by mRNA targets with $CD34^+$ MSI2-HyperTRIBE score ≥0.2, as well as significantly downregulated following enhancer perturbation (relative expression ≤0.95 and adjusted $P < 0.05$ in ENH-1 vs. AAVS1 HSCs). Inter-individual variation and correlation between the z-normalized expression of *MSI2* and its network was analysed from a reference single cell dataset of circulating HSCs from healthy donors (39). Ribo-seq data was generated in $CD34^+CD45RA^-CD90^+$ human HSCs, as previously reported (36); raw counts of ribosome protected fragments (RPFs) were normalized to transcript length, and the mean log-normalized RPKM (reads per kilobase per million mapped reads) across three replicates was ranked, followed by GSEA using the fGSEA package.

To validate clinical relevance of the MSI2-regulated network, TARGET-seq data (targeted high-sensitivity single cell mutational analysis with parallel RNA-seq) for *TET2* mutant vs. non-mutant HSC/MPPs was previously reported (47). The transcriptomic profile of *ASXL1* mutant vs. non-mutant $CD34^+$ cells was generated through re-analysis of a previously published RNA-seq dataset (48). The gene set of functional $Dnmt3a^{R882H}$ vulnerabilities was previously described (49). Survival analysis in the TARGET-AML dataset was performed using Survival Genie 2.0 (87), within diagnostic BM samples from *N*=1,713 patients (50); overall survival from time of diagnosis was stratified by median normalized expression of *MSI2*, or the downstream MSI2-DOWN CHIP-resilience network (*N*=208 genes), with all other default settings. Expression of *MSI2* or its network was z-normalized from RNA-seq of diagnostic BM samples from the adult BEAT-AML cohort (51), and stratified as either ‘mature’ or ‘primitive’ as previously reported through single cell deconvolution (88).

Longitudinal clonal growth rate analysis

Targeted, error-corrected sequencing was performed on two serial blood samples from 3,000 people in the Vanderbilt BioVU biobank, with custom-designed probes for 22 CHIP-associated genes, as previously described (89, 90). Vanderbilt University Medical Center’s Institutional Review Board oversees BioVU and approved this project (IRB #201783). UMIs were used for error correction, excluding mutations detected from a single UMI. The mean coverage depth was 1725x after de-duplication. CHIP mutations were called for variants with 100x total read depth, 3 variant allele reads, and variant allele fraction (VAF) ≥ 2% in at least one blood draw. The median time between blood draws was 5.7 years (range: 0.7-13). 657 CHIP mutations (VAF ≥ 2%) were identified at the first time point, of which at the second time point 513 mutations had a VAF ≥ 2% (persistent CHIP) and 144 had a VAF < 2% (transient CHIP). For those with persistent CHIP, growth rate r was modelled with a compound interest formula $r = \frac{VAF_2{}^{1/t}}{VAF_1} - 1$ where t is the duration in years between blood draws, as previously done (89, 90). In those with CHIP mutations at baseline, rs17834140 genotypes were extracted using GATK’s HaplotypeCaller (91). For persistent CHIP mutations, we modeled growth rate $r \sim \text{age} + \text{age}^2 + \text{VAF} + \text{sex} + \text{rs17834140-T} + \text{driver gene}$ (encoded as a factor), and extracted the coefficient of rs17834140-T. For all CHIP mutations at baseline, we performed logistic regression of transient_CHIP (0/1) ~ age + age^2 + VAF

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

+ sex + rs17834140-T + driver gene (encoded as a factor). We extracted the odds ratio of transient CHIP by exponentiating the coefficient of rs17834140-T. We repeated the models for the DNMT3A, TET2, ASXL1, and other driver genes separately, and removed driver gene as a covariate for these models.

Modeling human ASXL1 mutant CHIP and bulk RNA-seq

To model germline CHIP-resilience and acquired *ASXL1* mutations, CRISPR/Cas9 was used to edit $CD34^+$ cells (from mobilized peripheral blood) on day 1 of culture. HSPCs were edited (as described in previous section) with total 2.0 μL of 100 μM sgRNA, with 1.0 μL to model germline effects [AAVS1: 1.0 μL of sgAAVS1_g1; ENH-1 microdeletion: 0.5 μL each of sgENH-1_g1 and sgENH-1_g2; MSI2-KO: 1 μL of sgMSI2-KO], and 1.0 μL to model somatic mutation [control AAVS1: sgAAVS1_g2; ASXL1-exon 12: sgASXL1-ex12]. Immunophenotyping of LT-HSCs was performed 7 days after editing.

For bulk RNA-sequencing of *ASXL1* edited human HSPCs, FACS was performed on day 8 of culture (7 days after editing) for viable cells (negative for DAPI) and $CD34^+CD45RA^-CD90^+$ markers to enrich for HSCs. Total RNA was obtained from sorted cells using the Total RNA Purification Micro Kit (Norgen, 35300) as per manufacturer's instructions. Ultra-low input RNA-seq was conducted with ~30M reads per sample, across 2 independent experiments in edited HSPCs (AAVS1-sg1+AAVS1-sg2, AAVS1-sg1+ASXL1, ENH-1+ASXL1 and MSI2-KO+ASXL1). FASTQ files were processed for quality control using FastQC (v0.12.1) to assess sequence quality. Salmon (v1.5.2) was used to align reads to the reference transcriptome (GRCh38, Ensembl v104), to quantify gene expression levels. Transcript abundance estimates were imported into R for differential expression analysis using DESeq2 (v1.34.0). Gene set enrichment analysis was performed using the GO Biological Process 2021 database, and the activated HSC signature previously reported (43).

Asxl1-mutant mouse model

We crossed Mx1-Cre^- $Asxl1^{f/f}$ or Mx1-Cre^+ $Asxl1^{f/f}$ mice with Col-tet on-*MSI2*/ROSA-rTTA mice, to generate Mx1-Cre^- $Asxl1^{f/f}$ $MSI2^{wt/ht}$ (Control), Mx1-Cre^- $Asxl1^{f/f}$ $MSI2^{ht/ht}$ ($MSI2^{DOX}$), Mx1-Cre^+ $Asxl1^{f/f}$ $MSI2^{wt/ht}$ ($Asxl1^{f/f}$) and Mx1-Cre^+ $Asxl1^{f/f}$ $MSI2^{ht/ht}$ ($Asxl1^{f/f}$+$MSI2^{DOX}$) mice. 1 million whole BM cells from donor plus 200K CD45.1 helper BM cells were transplanted into lethally irradiated B6-CD45.1 recipient mice. Depletion of *Asxl1* was initiated one-month post-transplantation by administering three intraperitoneal injections of pIpC HMW (InVivogen, vac-pic) at a dose of 10 mg/kg. MSI2 overexpression was induced by administration of 2 mg/mL Doxycycline hyclate (Fisher Scientific, AC446060050) supplemented with 10 mg/mL sucrose in the drinking water, starting from one month after pIpC treatment. Peripheral blood or BM cells were collected at designated time points and analyzed by flow cytometry.

Quantification and statistical analysis

In all experiments, data were presented as mean ± standard error of mean. When comparing two samples, a two-tailed Student's t-test was used to test statistical significance. When one of the two samples was a default value (as in fold change comparison), the one sample t- and Wilcoxon test was applied. When comparing three or more samples, Levene's test was first

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

used to test equality of variance. If the variance across all samples were tested insignificantly differed, one-way or two-way ANOVA with Dunnett's test (for multiple comparisons where no reference group is defined) or Tukey's test (for multiple comparisons where reference group is defined) as post-hoc analysis was used. If the variance across samples was tested to be significantly different, the Kruskal-Wallis test was used instead of ANOVA, with the Dunn test as the post-hoc multiple comparison test. Log-rank test was used for survival analysis. All statistical tests were performed in Graphpad software or R when statistical tests were not available through Graphpad.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank members of the Sankaran and Kharas labs for valuable comments and suggestions.

Funding:

This work was supported by the Howard Hughes Medical Institute (V.G.S.), the Mather's Foundation (V.G.S.), the Alex's Lemonade Stand Foundation (V.G.S.), the Edward P. Evans Foundation (V.G.S.), Blood Cancer United (V.G.S. and M.G.K.), the Bill and Melinda Gates Foundation (V.G.S.), and National Institutes of Health (NIH) grants R01 DK103794, R01 CA265726, R01 CA292941, R33 CA278393, and R01 HL146500 (V.G.S.). X.X. was supported by a NYSTEM Postdoctoral Fellowship award. A-L.N. was supported by the EMBO Postdoctoral Fellowship (ALTF 209-2024). A.G.B. was supported by NIH grants DP5 OD029586, R01 AG088657, R01 AG083736; a Burroughs Wellcome Fund Career Award; a Pew-Stewart Scholar for Cancer Research award; and a Hevolution/AFAR New Investigator Award in Aging Biology and Geroscience Research. M.G.K was supported by Blood Cancer United, the American Society of Hematology, Break Through Cancer, and NIH grants R01 DK101989, R01 CA193842, R01 HL135564, R01 CA274249, R01 CA186702, R01 CA283578, R01 CA225231, and P30 CA008748. V.G.S. is an Investigator of the Howard Hughes Medical Institute.

References and notes

1. Martincorena I, Somatic mutation and clonal expansions in human tissues. Genome medicine 11, 35 (2019). [PubMed: 31138277]
2. Jaiswal S, Ebert BL, Clonal hematopoiesis in human aging and disease. Science 366, eaan4673 (2019). [PubMed: 31672865]
3. Mitchell E, Spencer Chapman M, Williams N, Dawson KJ, Mende N, Calderbank EF, Jung H, Mitchell T, Coorens THH, Spencer DH, Machado H, Lee-Six H, Davies M, Hayler D, Fabre MA, Mahbubani K, Abascal F, Cagan A, Vassiliou GS, Baxter J, Martincorena I, Stratton MR, Kent DG, Chatterjee K, Parsy KS, Green AR, Nangalia J, Laurenti E, Campbell PJ, Clonal dynamics of haematopoiesis across the human lifespan. Nature 606, 343–350 (2022). [PubMed: 35650442]
4. Weng C, Yu F, Yang D, Poeschla M, Liggett LA, Jones MG, Qiu X, Wahlster L, Caulier A, Hussmann JA, Schnell A, Yost KE, Koblan LW, Martin-Rufino JD, Min J, Hammond A, Ssozi D, Bueno R, Mallidi H, Kreso A, Escabi J, Rideout WM, Jacks T, Hormoz S, van Galen P, Weissman JS, Sankaran VG, Deciphering cell states and genealogies of human haematopoiesis. Nature 627, 389–398 (2024). [PubMed: 38253266]
5. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, McCarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL, Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. New England Journal of Medicine 371, 2488–2498 (2014). [PubMed: 25426837]
6. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, McConkey M, Gupta N, Gabriel S, Ardissino D, Baber U, Mehran R, Fuster V, Danesh J, Frossard P, Saleheen D, Melander O,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sukhova GK, Neuberg D, Libby P, Kathiresan S, Ebert BL, Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. New England Journal of Medicine 377, 111–121 (2017). [PubMed: 28636844]

7. Ferrari B, Peyvandi F, Association of clonal hematopoiesis with chronic obstructive pulmonary disease. Blood 139, 357–368 (2022). [PubMed: 34855941]
8. Vlasschaert C, Robinson-Cohen C, Chen J, Akwo E, Parker AC, Silver SA, Bhatraju PK, Poisner H, Cao S, Jiang M, Wang Y, Niu A, Siew E, Van Amburg JC, Kramer HJ, Kottgen A, Franceschini N, Psaty BM, Tracy RP, Alonso A, Arking DE, Coresh J, Ballantyne CM, Boerwinkle E, Grams M, Zhang MZ, Kestenbaum B, Lanktree MB, Rauh MJ, Harris RC, Bick AG, Clonal hematopoiesis of indeterminate potential is associated with acute kidney injury. Nat Med 30, 810–817 (2024). [PubMed: 38454125]
9. Díez-Díez M, Ramos-Neble BL, de la Barrera J, Silla-Castro JC, Quintas A, Vázquez E, Rey-Martín MA, Izzi B, Sánchez-García L, García-Lunar I, Mendieta G, Mass V, Gómez-López N, Espadas C, González G, Quesada AJ, García-Álvarez A, Fernández-Ortiz A, Lara-Pezzi E, Dopazo A, Sánchez-Cabo F, Ibáñez B, Andrés V, Fuster V, Fuster JJ, Unidirectional association of clonal hematopoiesis with atherosclerosis development. Nat Med 30, 2857–2866 (2024). [PubMed: 39215150]
10. Bao EL, Nandakumar SK, Liao X, Bick AG, Karjalainen J, Tabaka M, Gan OI, Havulinna AS, Kiiskinen TTJ, Lareau CA, de Lapuente Portilla AL, Li B, Emdin C, Codd V, Nelson CP, Walker CJ, Churchhouse C, de la Chapelle A, Klein DE, Nilsson B, Wilson PWF, Cho K, Pyarajan S, Gaziano JM, Samani NJ, Palotie A, Daly MJ, Jacob H, Matakidou A, Runz H, John S, Plenge R, McCarthy M, Hunkapiller J, Ehm M, Waterworth D, Fox C, Malarstig A, Klinger K, Call K, Mäkelä T, Kaprio J, Virolainen P, Pulkki K, Kilpi T, Perola M, Partanen J, Pitkäranta A, Kaarteenaho R, Vainio S, Savinainen K, Kosma VM, Kujala U, Tuovila O, Hendolin M, Pakkanen R, Waring J, Riley-Gillis B, Matakidou A, Runz H, Liu J, Biswas S, Hunkapiller J, Waterworth D, Ehm M, Diogo D, Fox C, Malarstig A, Marshall C, Hu X, Call K, Klinger K, Gossel M, Ripatti S, Schleutker J, Perola M, Arvas M, Carpén O, Hinttala R, Kettunen J, Laaksonen R, Mannermaa A, Tuovila O, Hendolin M, Pakkanen R, Soininen H, Julkunen V, Remes A, Kälviäinen R, Hiltunen M, Peltola J, Tienari P, Rinne J, Ziemann A, Waring J, Esmaeeli S, Smaoui N, Lehtonen A, Eaton S, Runz H, Lahdenperä S, van Adelsberg J, Biswas S, Michon J, Kerchner G, Hunkapiller J, Bowers N, Teng E, Eicher J, Mehta V, Gormley P, Linden K, Whelan C, Xu F, Pulford D, Färkkilä M, Pikkarainen S, Jussila A, Blomster T, Kiviniemi M, Voutilainen M, Georgantas B, Heap G, Waring J, Smaoui N, Rahimov F, Lehtonen A, Usiskin K, Maranville J, Lu T, Bowers N, Oh D, Michon J, Mehta V, Kalpala K, Miller M, Hu X, McCarthy L, Eklund K, Palomäki A, Isomäki P, Pirilä L, Kaipiainen-Seppänen O, Huhtakangas J, Georgantas B, Waring J, Rahimov F, Lertratanakul A, Smaoui N, Lehtonen A, Close D, Hochfeld M, Bowers N, Michon J, Diogo D, Mehta V, Kalpala K, Bing N, Hu X, Gordillo JE, Mars N, Laitinen T, Pelkonen M, Kauppi P, Kankaanranta H, Harju T, Smaoui N, Close D, Greenberg S, Chen H, Bowers N, Michon J, Mehta V, Betts J, Ghosh S, Salomaa V, Niiranen T, Juonala M, Metsärinne K, Kähönen M, Junttila J, Laakso M, Pihlajamäki J, Sinisalo J, Taskinen MR, Tuomi T, Laukkanen J, Challis B, Peterson A, Hunkapiller J, Bowers N, Michon J, Diogo D, Chu A, Mehta V, Parkkinen J, Miller M, Muslin A, Waterworth D, Joensuu H, Meretoja T, Carpén O, Aaltonen L, Auranen A, Karihtala P, Kauppila S, Auvinen P, Elenius K, Popovic R, Waring J, Riley-Gillis B, Lehtonen A, Matakidou A, Schutzman J, Hunkapiller J, Bowers N, Michon J, Mehta V, Loboda A, Chhibber A, Lehtonen H, McDonough S, Crohns M, Kulkarni D, Kaarniranta K, Turunen J, Ollila T, Seitsonen S, Uusitalo H, Aaltonen V, Uusitalo-Järvinen H, Luodonpää M, Hautala N, Runz H, Strauss E, Bowers N, Chen H, Michon J, Podgornaia A, Mehta V, Diogo D, Hoffman J, Tasanen K, Huilaja L, Hannula-Jouppi K, Salmi T, Peltonen S, Koulu L, Harvima I, Kalpala K, Wu Y, Choy D, Michon J, Smaoui N, Rahimov F, Lehtonen A, Waterworth D, Davis JW, Riley-Gillis B, Quarless D, Petrovski S, Liu J, Chen CY, Bronson P, Yang R, Maranville J, Biswas S, Chang D, Hunkapiller J, Bhangale T, Bowers N, Diogo D, Holzinger E, Gormley P, Wang X, Chen X, Hedman Å, Auro K, Wang C, Xu E, Auge F, Chatelain C, Kurki M, Ripatti S, Karjalainen J, Jalanko A, Palin K, Palta P, della Briotta Parolo P, Zhou W, Lemmelä S, Rivas M, Harju J, Lehisto A, Ganna A, Llorens V, Karlsson A, Kristiansson K, Arvas M, Hyvärinen K, Ritari J, Wahlfors T, Koskinen M, Carpén O, Kettunen J, Pylkäs K, Kalaoja M, Karjalainen M, Mantere T, Kangasniemi E, Heikkinen S, Mannermaa A, Laakkonen E, Kononen J, Kallio L, Soini S, Partanen J, Pitkänen K, Vainio S, Savinainen K, Kosma VM, Kuopio T, Jalanko A, Kajanne R, Lyhs U, Kurki M, Karjalainen J, della Briotta Parola P, Rüeger S, Lehistö A, Zhou W, Kanai

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

M, Laivuori H, Lemmelä S, Kiiskinen T, Kaunisto M, Harju J, Kilpeläinen E, Sipilä TP, Brein G, Dada OA, Awaisa G, Shcherban A, Sipilä T, Donner K, Loukola A, Laiho P, Sistonen T, Kaiharju E, Laukkanen M, Järvensivu E, Lähteenmäki S, Männikkö L, Wong R, Mattsson H, Kristiansson K, Lemmelä S, Hiekkalinna T, Jiménez MG, Palta P, Pärn K, Nunez-Fontarnau J, Laitinen T, Siirtola H, Tabuenca JG, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Huber KE, Kleinman A, Litterman NK, McCreight JC, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Tung JY, Vacic V, Wilson CH, Regev A, Neale BM, Dick JE, Natarajan P, O'Donnell CJ, Milyavsky M, Kathiresan S, Sankaran VG, Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. Nature 586, 769–775 (2020). [PubMed: 33057200]

11. Kessler MD, Damask A, O'Keeffe S, Banerjee N, Li D, Watanabe K, Marketta A, Van Meter M, Semrau S, Horowitz J, Tang J, Kosmicki JA, Rajagopal VM, Zou Y, Houvras Y, Ghosh A, Gillies C, Mbatchou J, White RR, Verweij N, Bovijn J, Parikshak NN, LeBlanc MG, Jones M, Glass DJ, Lotta LA, Cantor MN, Atwal GS, Locke AE, Ferreira MAR, Deering R, Paulding C, Shuldiner AR, Thurston G, Ferrando AA, Salerno W, Reid JG, Overton JD, Marchini J, Kang HM, Baras A, Abecasis GR, Jorgenson E, Common and rare variant associations with clonal haematopoiesis phenotypes. Nature 612, 301–309 (2022). [PubMed: 36450978]
12. Thompson DJ, Genovese G, Halvardson J, Ulirsch JC, Wright DJ, Terao C, Davidsson OB, Day FR, Sulem P, Jiang Y, Danielsson M, Davies H, Dennis J, Dunlop MG, Easton DF, Fisher VA, Zink F, Houlston RS, Ingelsson M, Kar S, Kerrison ND, Kinnersley B, Kristjansson RP, Law PJ, Li R, Loveday C, Mattisson J, McCarroll SA, Murakami Y, Murray A, Olszewski P, Rychlicka-Buniowska E, Scott RA, Thorsteinsdottir U, Tomlinson I, Moghadam BT, Turnbull C, Wareham NJ, Gudbjartsson DF, Kamatani Y, Hoffmann ER, Jackson SP, Stefansson K, Auton A, Ong KK, Machiela MJ, Loh PR, Dumanski JP, Chanock SJ, Forsberg LA, Perry JRB, Genetic predisposition to mosaic Y chromosome loss in blood. Nature 575, 652–657 (2019). [PubMed: 31748747]
13. Brown DW, Cato LD, Zhao Y, Nandakumar SK, Bao EL, Gardner EJ, Hubbard AK, DePaulis A, Rehling T, Song L, Yu K, Chanock SJ, Perry JRB, Sankaran VG, Machiela MJ, Shared and distinct genetic etiologies for different types of clonal hematopoiesis. Nat Commun 14, 5536 (2023). [PubMed: 37684235]
14. Zhao J, Cato LD, Arora UP, Bao EL, Bryant SC, Williams N, Jia Y, Goldman SR, Nangalia J, Erb MA, Vos SM, Armstrong SA, Sankaran VG, Inherited blood cancer predisposition through altered transcription elongation. Cell 187, 642–658 (2024). [PubMed: 38218188]
15. Bick AG, Weinstock JS, Nandakumar SK, Fulco CP, Bao EL, Zekavat SM, Szeto MD, Liao X, Leventhal MJ, Nasser J, Chang K, Laurie C, Burugula BB, Gibson CJ, Lin AE, Taub MA, Aguet F, Ardlie K, Mitchell BD, Barnes KC, Moscati A, Fornage M, Redline S, Psaty BM, Silverman EK, Weiss ST, Palmer ND, Vasan RS, Burchard EG, Kardia SLR, He J, Kaplan RC, Smith NL, Arnett DK, Schwartz DA, Correa A, de Andrade M, Guo X, Konkle BA, Custer B, Peralta JM, Gui H, Meyers DA, McGarvey ST, Der Chen IY, Shoemaker MB, Peyser PA, Broome JG, Gogarten SM, Wang FF, Wong Q, Montasser ME, Daya M, Kenny EE, North KE, Launer LJ, Cade BE, Bis JC, Cho MH, Lasky-Su J, Bowden DW, Cupples LA, Mak ACY, Becker LC, Smith JA, Kelly TN, Aslibekyan S, Heckbert SR, Tiwari HK, Yang IV, Heit JA, Lubitz SA, Johnsen JM, Curran JE, Wenzel SE, Weeks DE, Rao DC, Darbar D, Moon JY, Tracy RP, Buth EJ, Rafaels N, Loos RJF, Durda P, Liu Y, Hou L, Lee J, Kachroo P, Freedman BI, Levy D, Bielak LF, Hixson JE, Floyd JS, Whitsel EA, Ellinor PT, Irvin MR, Fingerlin TE, Raffield LM, Armasu SM, Wheeler MM, Sabino EC, Blangero J, Williams LK, Levy BD, Sheu WHH, Roden DM, Boerwinkle E, Manson JAE, Mathias RA, Desai P, Taylor KD, Johnson AD, Abe N, Albert C, Almasy L, Alonso A, Ament S, Anderson P, Anugu P, Applebaum-Bowden D, Arking D, Ashley-Koch A, Aslibekyan S, Assimes T, Avramopoulos D, Barnard J, Barr RG, Barron-Casella E, Barwick L, Beaty T, Beck G, Becker D, Beer R, Beitelshees A, Benjamin E, Benos P, Bezerra M, Bielak L, Bowler R, Brody J, Broeckel U, Bunting K, Bustamante C, Cardwell J, Carey V, Carty C, Casaburi R, Casella J, Castaldi P, Chaffin M, Chang C, Chang YC, Chasman D, Chavan S, Chen BJ, Chen WM, Choi SH, Chuang LM, Chung M, Chung RH, Clish C, Comhair S, Cornell E, Crandall C, Crapo J, Curtis J, Damcott C, Das S, David S, Davis C, DeBaun M, Deka R, DeMeo D, Devine S, Duan Q, Duggirala R, Dutcher S, Eaton C, Ekunwe L, El Boueiz A, Emery L, Erzurum S,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Farber C, Flickinger M, Franceschini N, Frazar C, Fu M, Fullerton SM, Fulton L, Gabriel S, Gan W, Gao S, Gao Y, Gass M, Gelb B, Geng X. (Priscilla), Geraci M, Germer S, Gerszten R, Ghosh A, Gibbs R, Gignoux C, Gladwin M, Glahn D, Gong DW, Goring H, Graw S, Grine D, Gu CC, Guan Y, Gupta N, Haessler J, Hall M, Harris D, Hawley NL, Heavner B, Hernandez R, Herrington D, Hersh C, Hidalgo B, Hobbs B, Hokanson J, Hong E, Hoth K, Hsiung C. (Agnes), Hung YJ, Huston H, Hwu CM, Jackson R, Jain D, Jaquish C, Jhun MA, Johnson C, Johnston R, Jones K, Kang HM, Kelly S, Kessler M, Khan A, Kim W, Kinney G, Kramer H, Lange C, LeBoff M, Lee SS, Lee WJ, LeFaive J, Levine D, Lewis J, Li X, Li Y, Lin H, Lin H, Lin KH, Lin X, Liu S, Liu Y, Lunetta K, Luo J, Mahaney M, Make B, Manichaikul A, Margolin L, Martin L, Mathai S, May S, McArdle P, McDonald ML, McFarland S, McGoldrick D, McHugh C, Mei H, Mestroni L, Mikulla J, Min N, Minear M, Minster RL, Moll M, Montgomery C, Musani S, Mwasongwe S, Mychaleckyj JC, Nadkarni G, Naik R, Naseri T, Nekhai S, Nelson SC, Neltner B, Nickerson D, O'Connell J, O'Connor T, Ochs-Balcom H, Paik D, Pankow J, Papanicolaou G, Parsa A, Perez M, Perry J, Peters U, Peyser P, Phillips LS, Pollin T, Post W, Becker JP, Boorgula MP, Preuss M, Qasba P, Qiao D, Qin Z, Rasmussen-Torvik L, Ratan A, Reed R, Regan E, Sefuiva Reupena M, Rice K, Roselli C, Ruczinski I, Russell P, Ruuska S, Ryan K, Saleheen D, Salimi S, Salzberg S, Sandow K, Scheller C, Schmidt E, Schwander K, Sciurba F, Seidman C, Seidman J, Sheehan V, Sherman SL, Shetty A, Shetty A, Silver B, Smith J, Smith T, Smoller S, Snively B, Snyder M, Sofer T, Sotoodehnia N, Stilp AM, Storm G, Streeten E, Sung YJ, Sylvia J, Szpiro A, Sztalryd C, Taliun D, Tang H, Taylor M, Taylor S, Telen M, Thornton TA, Threlkeld M, Tinker L, Tirschwell D, Tishkoff S, Tiwari H, Tong C, Tsai M, Vaidya D, Van Den Berg D, VandeHaar P, Vrieze S, Walker T, Wallace R, Walts A, Wang H, Watson K, Weir B, Weng LC, Wessel J, Willer C, Williams K, Wilson C, Wu J, Xu H, Yanek L, Yang R, Zaghloul N, Zhang Y, Zhao SX, Zhao W, Zhi D, Zhou X, Zhu X, Zody M, Zoellner S, Auer PL, Kooperberg C, Laurie CC, Blackwell TW, Smith AV, Zhao H, Lange E, Lange L, Rich SS, Rotter JI, Wilson JG, Scheet P, Kitzman JO, Lander ES, Engreitz JM, Ebert BL, Reiner AP, Jaiswal S, Abecasis G, Sankaran VG, Kathiresan S, Natarajan P, Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature 586, 763–768 (2020). [PubMed: 33057201]

16. Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, Iyer V, Barcena C, Vijayabaskar MS, Fabre MA, Carter P, Petrovski S, Burgess S, Vassiliou GS, Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat Genet 54, 1155–1166 (2022). [PubMed: 35835912]

17. Weinstock JS, Gopakumar J, Burugula BB, Uddin MM, Jahn N, Belk JA, Bouzid H, Daniel B, Miao Z, Ly N, Mack TM, Luna SE, Prothro KP, Mitchell SR, Laurie CA, Broome JG, Taylor KD, Guo X, Sinner MF, von Falkenhausen AS, Kääb S, Shuldiner AR, O'Connell JR, Lewis JP, Boerwinkle E, Barnes KC, Chami N, Kenny EE, Loos RJF, Fornage M, Hou L, Lloyd-Jones DM, Redline S, Cade BE, Psaty BM, Bis JC, Brody JA, Silverman EK, Yun JH, Qiao D, Palmer ND, Freedman BI, Bowden DW, Cho MH, DeMeo DL, Vasan RS, Yanek LR, Becker LC, Kardia SLR, Peyser PA, He J, Rienstra M, Van der Harst P, Kaplan R, Heckbert SR, Smith NL, Wiggins KL, Arnett DK, Irvin MR, Tiwari H, Cutler MJ, Knight S, Muhlestein JB, Correa A, Raffield LM, Gao Y, de Andrade M, Rotter JI, Rich SS, Tracy RP, Konkle BA, Johnsen JM, Wheeler MM, Smith JG, Melander O, Nilsson PM, Custer BS, Duggirala R, Curran JE, Blangero J, McGarvey S, Williams LK, Xiao S, Yang M, Gu CC, Chen YDI, Lee WJ, Marcus GM, Kane JP, Pullinger CR, Shoemaker MB, Darbar D, Roden DM, Albert C, Kooperberg C, Zhou Y, Manson JAE, Desai P, Johnson AD, Mathias RA, Abe N, Abecasis GR, Aguet F, Albert C, Almasy L, Alonso A, Ament S, Anderson P, Anugu P, Applebaum-Bowden D, Ardlie K, Arking D, Arnett DK, Ashley-Koch A, Aslibekyan S, Assimes T, Auer P, Avramopoulos D, Ayas N, Balasubramanian A, Barnard J, Barnes KC, Barr RG, Barron-Casella E, Barwick L, Beaty T, Beck G, Becker D, Becker LC, Beer R, Beitelshees A, Benjamin E, Benos T, Bezerra M, Bielak L, Bis J, Blackwell TW, Blangero J, Boerwinkle E, Bowden DW, Bowler R, Brody JA, Broeckel U, Broome JG, Brown D, Bunting K, Burchard E, Bustamante C, Buth E, Cade BE, Cardwell J, Carey V, Carrier J, Carty C, Casaburi R, Casas Romero JP, Casella J, Castaldi P, Chaffin M, Chang C, Chang YC, Chasman D, Chavan S, Chen BJ, Chen WM, Chen YDI, Cho MH, Choi SH, Chuang LM, Chung M, Chung RH, Clish C, Comhair S, Conomos M, Cornell E, Correa A, Crandall C, Crapo J, Cupples LA, Curran JE, Curtis J, Custer BS, Damcott C, Darbar D, David S, Davis C, Daya M, de Andrade M, de las Fuentes L, de Vries P, DeBaun M, Deka R, DeMeo DL, Devine S, Dinh H, Doddapaneni H, Duan Q, Dugan-Perez S, Duggirala R, Durda JP, Dutcher SK, Eaton C, Ekunwe L, El Boueiz A, Ellinor

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

P, Emery L, Erzurum S, Farber C, Farek J, Fingerlin T, Flickinger M, Fornage M, Franceschini N, Frazar C, Fu M, Fullerton SM, Fulton L, Gabriel S, Gan W, Gao S, Gao Y, Gass M, Geiger H, Gelb B, Geraci M, Germer S, Gerszten R, Ghosh A, Gibbs R, Gignoux C, Gladwin M, Glahn D, Gogarten S, Gong DW, Goring H, Graw S, Gray KJ, Grine D, Gross C, Gu CC, Guan Y, Guo X, Gupta N, Haas DM, Haessler J, Hall M, Han Y, Hanly P, Harris D, Hawley NL, He J, Heavner B, Heckbert SR, Hernandez R, Herrington D, Hersh C, Hidalgo B, Hixson J, Hobbs B, Hokanson J, Hong E, Hoth K, Hsiung CA, Hu J, Hung YJ, Huston H, Hwu CM, Irvin MR, Jackson R, Jain D, Jaquish C, Johnsen JM, Johnson AD, Johnson C, Johnston R, Jones K, Kang HM, Kaplan R, Kardia SLR, Kelly S, Kenny EE, Kessler M, Khan A, Khan Z, Kim W, Kimoff J, Kinney G, Konkle BA, Kooperberg C, Kramer H, Lange C, Lange E, Lange L, Laurie C, Laurie CA, LeBoff M, Lee J, Lee S, Lee WJ, LeFaive J, Levine D, Levy D, Lewis JP, Li X, Li Y, Lin H, Lin H, Lin X, Liu S, Liu Y, Liu Y, Lubitz S, Lunetta K, Luo J, Magalang U, Mahaney M, Make B, Manichaikul A, Manning A, Manson JAE, Martin L, Marton M, Mathai S, Mathias RA, May S, McArdle P, McDonald ML, McFarland S, McGarvey S, McGoldrick D, McHugh C, McNeil B, Mei H, Meigs J, Menon V, Mestroni L, Metcalf G, Meyers DA, Mignot E, Mikulla J, Min N, Minear M, Minster RL, Mitchell BD, Moll M, Momin Z, Montasser ME, Montgomery C, Muzny D, Mychaleckyj JC, Nadkarni G, Naik R, Naseri T, Natarajan P, Nekhai S, Nelson SC, Neltner B, Nessner C, Nickerson D, Nkechinyere O, North K, O'Connell J, O'Connor T, Ochs-Balcom H, Okwuonu G, Pack A, Paik DT, Palmer ND, Pankow J, Papanicolaou G, Parker C, Peloso G, Peralta JM, Perez M, Perry J, Peters U, Peyser PA, Phillips LS, Pleiness J, Pollin T, Post W, Becker JP, Boorgula MP, Preuss M, Psaty BM, Qasba P, Qiao D, Qin Z, Rafaels N, Raffield LM, Rajendran M, Ramachandran VS, Rao DC, Rasmussen-Torvik L, Ratan A, Redline S, Reed R, Reeves C, Regan E, Reiner AP, Sefuiva Reupena M, Rice K, Rich SS, Robillard R, Robine N, Roden DM, Roselli C, Rotter JI, Ruczinski I, Runnels A, Russell P, Ruuska S, Ryan K, Sabino EC, Saleheen D, Salimi S, Salvi S, Salzberg S, Sandow K, Sankaran VG, Santibanez J, Schwander K, Schwartz D, Sciurba F, Seidman C, Seidman J, Sériès F, Sheehan V, Sherman SL, Shetty A, Shetty A, Sheu WHH, Shoemaker MB, Silver B, Silverman EK, Skomro R, Smith AV, Smith J, Smith J, Smith NL, Smith T, Smoller S, Snively B, Snyder M, Sofer T, Sotoodehnia N, Stilp AM, Storm G, Streeten E, Su JL, Sung YJ, Sylvia J, Szpiro A, Taliun D, Tang H, Taub M, Taylor KD, Taylor M, Taylor S, Telen M, Thornton TA, Threlkeld M, Tinker L, Tirschwell D, Tishkoff S, Tiwari H, Tong C, Tracy RP, Tsai M, Vaidya D, Van Den Berg D, VandeHaar P, Vrieze S, Walker T, Wallace R, Walts A, Wang FF, Wang H, Wang J, Watson K, Watt J, Weeks DE, Weinstock JS, Weir B, Weiss ST, Weng LC, Wessel J, Willer C, Williams K, Williams LK, Wilson C, Wilson J, Winterkorn L, Wong Q, Wu J, Xu H, Yanek LR, Yang I, Yu K, Zekavat SM, Zhang Y, Zhao SX, Zhao W, Zhu X, Zody M, Zoellner S, Blackwell TW, Abecasis GR, Smith AV, Kang HM, Satpathy AT, Natarajan P, Kitzman JO, Whitsel EA, Reiner AP, Bick AG, Jaiswal S, Aberrant activation of TCL1A promotes stem cell expansion in clonal haematopoiesis. Nature 616, 755–763 (2023). [PubMed: 37046083]

18. Young AL, Challen GA, Birmann BM, Druley TE, Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun 7, 12484 (2016). [PubMed: 27546487]
19. Sousos N, Ní Leathlobhair M, Simoglou Karali C, Louka E, Bienz N, Royston D, Clark SA, Hamblin A, Howard K, Mathews V, George B, Roy A, Psaila B, Wedge DC, Mead AJ, In utero origin of myelofibrosis presenting in adult monozygotic twins. Nat Med 28, 1207–1211 (2022). [PubMed: 35637336]
20. Williams N, Lee J, Mitchell E, Moore L, Baxter EJ, Hewinson J, Dawson KJ, Menzies A, Godfrey AL, Green AR, Campbell PJ, Nangalia J, Life histories of myeloproliferative neoplasms inferred from phylogenies. Nature 602, 162–168 (2022). [PubMed: 35058638]
21. Fabre MA, de Almeida JG, Fiorillo E, Mitchell E, Damaskou A, Rak J, Orrù V, Marongiu M, Chapman MS, Vijayabaskar MS, Baxter J, Hardy C, Abascal F, Williams N, Nangalia J, Martincorena I, Campbell PJ, McKinney EF, Cucca F, Gerstung M, Vassiliou GS, The longitudinal dynamics and natural history of clonal haematopoiesis. Nature 606, 335–342 (2022). [PubMed: 35650444]
22. Uddin MM, Saadatagah S, Niroula A, Yu B, Hornsby WE, Ganesh S, Lannery K, Schuermans A, Honigberg MC, Bick AG, Libby P, Ebert BL, Ballantyne CM, Natarajan P, Long-term longitudinal analysis of 4,187 participants reveals insights into determinants of clonal hematopoiesis. Nature Communications 15, 7858 (2024).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

23. Robertson NA, Latorre-Crespo E, Terradas-Terradas M, Lemos-Portela J, Purcell AC, Livesey BJ, Hillary RF, Murphy L, Fawkes A, MacGillivray L, Copland M, Marioni RE, Marsh JA, Harris SE, Cox SR, Deary IJ, Schumacher LJ, Kirschner K, Chandra T, Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. Nat Med 28, 1439–1446 (2022). [PubMed: 35788175]
24. Hormaechea-Agulla D, Matatall KA, Le DT, Kain B, Long X, Kus P, Jaksik R, Challen GA, Kimmel M, King KY, Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling. Cell Stem Cell 28, 1428–1442 (2021). [PubMed: 33743191]
25. Takahashi K, Nakada D, Goodell M, Distinct landscape and clinical implications of therapy-related clonal hematopoiesis. J Clin Invest. 134, 19 (2024).
26. Mesbah Uddin M, Yu Z, Weinstock JS, Nakao T, Urbut SM, Koyama S, Zekavat SM, Paruchuri K, Silver AJ, Mack TM, Wong MY, Haidermota SM, Doroodgar Jorshery S, Raddatz MA, Hornsby WE, Jinye Zhang M, Sankaran VG, Griffin K, Gibson CJ, Kresge HA, Ellinor PT, V. Affairs, Germline genomic and phenomic landscape of clonal hematopoiesis in 323,112 individuals. [Preprint] (2022). 10.1101/2022.07.29.22278015.
27. Menendez-Gonzalez JB, Vukovic M, Abdelfattah A, Saleh L, Almotiri A, anne Thomas L, Agirre-Lizaso A, Azevedo A, Menezes AC, Tornillo G, Edkins S, Kong K, Giles P, Anjos-Afonso F, Tonks A, Boyd AS, Kranc KR, Rodrigues NP, Gata2 as a Crucial Regulator of Stem Cells in Adult Hematopoiesis and Acute Myeloid Leukemia. Stem Cell Reports 13, 291–306 (2019). [PubMed: 31378673]
28. Souroullas GP, Salmon JM, Sablitzky F, Curtis DJ, Goodell MA, Adult Hematopoietic Stem and Progenitor Cells Require Either Lyl1 or Scl for Survival. Cell Stem Cell 4, 180–186 (2009). [PubMed: 19200805]
29. Subramanian S, Thoms JAI, Huang Y, Cornejo-P ´ P, Koch FC, Jacquelin S, Shen S, Song E, Joshi S, Brownlee C, Woll PS, Chacon-Fajardo D, Beck D, Curtis DJ, Yehson K, Antonenas V, O'brien T, Trickett A, Powell JA, Lewis ID, Pitson SM, Gandhi MK, Lane SW, Vafaee F, Wong ES, Göttgens B, Alinejad-Rokny H, Wong JWH, Pimanda JE, Genome-wide transcription factor–binding maps reveal cell-specific changes in the regulatory architecture of human HSPCs. Blood 142, 1448–1462 (2023). [PubMed: 37595278]
30. De Andrés-Aguayo L, Varas F, Kallin EM, Infante JF, Wurst W, Floss T, Graf T, Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood 118, 554–564 (2011). [PubMed: 21613258]
31. Hope KJ, Cellot S, Ting SB, MacRae T, Mayotte N, Iscove NN, Sauvageau G, An RNAi screen identifies msi2 and prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell 7, 101–113 (2010). [PubMed: 20621054]
32. Kharas MG, Lengner CJ, Al-Shahrour F, Bullinger L, Ball B, Zaidi S, Morgan K, Tam W, Paktinat M, Okabe R, Gozo M, Einhorn W, Lane SW, Scholl C, Fröhling S, Fleming M, Ebert BL, Gilliland DG, Jaenisch R, Daley GQ, Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nat Med 16, 903–908 (2010). [PubMed: 20616797]
33. Rentas S, Holzapfel NT, Belew MS, Pratt GA, Voisin V, Wilhelm BT, Bader GD, Yeo GW, Hope KJ, Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. Nature 532, 508–511 (2016). [PubMed: 27121842]
34. Sakurai M, Ishitsuka K, Ito R, Wilkinson AC, Kimura T, Mizutani E, Nishikii H, Sudo K, Becker HJ, Takemoto H, Sano T, Kataoka K, Takahashi S, Nakamura Y, Kent DG, Iwama A, Chiba S, Okamoto S, Nakauchi H, Yamazaki S, Chemically defined cytokine-free expansion of human haematopoietic stem cells. Nature 615, 127–133 (2023). [PubMed: 36813966]
35. Voit RA, Liao X, Caulier A, Antoszewski M, Cohen B, Armant M, Lu HY, Fleming TJ, Kamal E, Wahlster L, Roche AM, Everett JK, Petrichenko A, Huang M-M, Clarke W, Myers KC, Forester C, Perez-Atayde A, Bushman FD, Pellin D, Shimamura A, Williams DA, Sankaran VG, Regulated GATA1 expression as a universal gene therapy for Diamond-Blackfan anemia. Cell Stem Cell 32, 38–52 (2025). [PubMed: 39532107]
36. Zhao J, Jia Y, Mahmut D, Deik AA, Jeanfavre S, Clish CB, Sankaran VG, Human hematopoietic stem cell vulnerability to ferroptosis. Cell 186, 732–747 (2023). [PubMed: 36803603]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

37. Park SM, Deering RP, Lu Y, Tivnan P, Lianoglou S, Al-Shahrour F, Ebert BL, Hacohen N, Leslie C, Daley GQ, Lengner CJ, Kharas MG, Musashi-2 controls cell fate, lineage bias, and TGF-β signaling in HSCs. Journal of Experimental Medicine 211, 71–87 (2014). [PubMed: 24395885]
38. Nguyen DTT, Lu Y, Chu KL, Yang X, Park SM, Choo ZN, Chin CR, Prieto C, Schurer A, Barin E, Savino AM, Gourkanti S, Patel P, Vu LP, Leslie CS, Kharas MG, HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. Nat Commun 11, 2026 (2020). [PubMed: 32332729]
39. Furer N, Rappoport N, Milman O, Tavor S, Lifshitz A, Bercovich A, Ben-Kiki O, Danin A, Kedmi M, Shipony Z, Lipson D, Meiri E, Yanai G, Shapira S, Arber N, Berdichevsky S, Tyner J, Joshi S, Landau D, Ganesan S, Dusaj N, Chamely P, Kaushansky N, Chapal-Ilani N, Shamir R, Tanay A, Shlush L, A reference model of circulating hematopoietic stem cells across the lifespan with applications to diagnostics. Nat Med 31, 2442–2451 (2025). [PubMed: 40579545]
40. Kwon HY, Bajaj J, Ito T, Blevins A, Konuma T, Weeks J, Lytle NK, Koechlein CS, Rizzieri D, Chuah C, Oehler VG, Sasik R, Hardiman G, Reya T, Tetraspanin 3 Is Required for the Development and Propagation of Acute Myelogenous Leukemia. Cell Stem Cell 17, 152–164 (2015). [PubMed: 26212080]
41. Vu LP, Prieto C, Amin EM, Chhangawala S, Krivtsov A, Calvo-Vidal MN, Chou T, Chow A, Minuesa G, Park SM, Barlowe TS, Taggart J, Tivnan P, Deering RP, Chu LP, Kwon JA, Meydan C, Perales-Paton J, Arshi A, Gönen M, Famulare C, Patel M, Paietta E, Tallman MS, Lu Y, Glass J, Garret-Bakelman FE, Melnick A, Levine R, Al-Shahrour F, Järås M, Hacohen N, Hwang A, Garippa R, Lengner CJ, Armstrong SA, Cerchietti L, Cowley GS, Root D, Doench J, Leslie C, Ebert BL, Kharas MG, Functional screen of MSI2 interactors identifies an essential role for SYNCRIP in myeloid leukemia stem cells. Nat Genet 49, 866–875 (2017). [PubMed: 28436985]
42. Sureda-Gómez M, Balsas P, Rodríguez ML, Nadeu F, De Bolòs A, Eguileor Á, Kulis M, Castellano G, López C, Giné E, Demajo S, Jares P, Martín-Subero JI, Beà S, Campo E, Amador V, Tumorigenic role of Musashi-2 in aggressive mantle cell lymphoma. Leukemia 37, 408–421 (2023). [PubMed: 36509891]
43. García-Prat L, Kaufmann KB, Schneiter F, Voisin V, Murison A, Chen J, Chan-Seng-Yue M, Gan OI, McLeod JL, Smith SA, Shoong MC, Parris D, Pan K, Zeng AGX, Krivdova G, Gupta K, Takayanagi SI, Wagenblast E, Wang W, Lupien M, Schroeder T, Xie SZ, Dick JE, TFEB-mediated endolysosomal activity controls human hematopoietic stem cell fate. Cell Stem Cell 28, 1838–1850 (2021). [PubMed: 34343492]
44. Wang S, He X, Wu Q, Jiang L, Chen L, Yu Y, Zhang P, Huang X, Wang J, Ju Z, Min J, Wang F, Transferrin receptor 1-mediated iron uptake plays an essential role in hematopoiesis. Haematologica 105, 2071–2082 (2020). [PubMed: 31601687]
45. Carrascoso-Rubio C, Zittersteijn HA, Pintado-Berninches L, Fernández-Varas B, Lozano ML, Manguan-Garcia C, Sastre L, Bueren JA, Perona R, Guenechea G, Generation of dyskeratosis congenita-like hematopoietic stem cells through the stable inhibition of DKC1. Stem Cell Res Ther 12, 1–8 (2021). [PubMed: 33397467]
46. Guo P, Liu Y, Geng F, Daman AW, Liu X, Zhong L, Ravishankar A, Lis R, Barcia Durán JG, Itkin T, Tang F, Zhang T, Xiang J, Shido K, sen Ding B, Wen D, Josefowicz SZ, Rafii S, Histone variant H3.3 maintains adult haematopoietic stem cell homeostasis by enforcing chromatin adaptability. Nat Cell Biol 24, 99–111 (2022). [PubMed: 34961794]
47. Jakobsen NA, Turkalj S, Zeng AGX, Stoilova B, Metzner M, Rahmig S, Nagree MS, Shah S, Moore R, Usukhbayar B, Angulo Salazar M, Gafencu GA, Kennedy A, Newman S, Kendrick BJL, Taylor AH, Afinowi-Luitz R, Gundle R, Watkins B, Wheway K, Beazley D, Murison A, Aguilar-Navarro AG, Flores-Figueroa E, Dakin SG, Carr AJ, Nerlov C, Dick JE, Xie SZ, Vyas P, Selective advantage of mutant stem cells in human clonal hematopoiesis is associated with attenuated response to inflammation and aging. Cell Stem Cell 31, 1127–1144 (2024). [PubMed: 38917807]
48. Binder M, Carr RM, Lasho TL, Finke CM, Mangaonkar AA, Pin CL, Berger KR, Mazzone A, Potluri S, Ordog T, Robertson KD, Marks DL, Fernandez-Zapico ME, Gaspar-Maia A, Patnaik MM, Oncogenic gene expression and epigenetic remodeling of cis-regulatory elements in ASXL1-mutant chronic myelomonocytic leukemia. Nat Commun 13, 1434 (2022). [PubMed: 35301312]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

49. Gozdecka M, Dudek M, Wen S, Gu M, Stopforth RJ, Rak J, Damaskou A, Grice GL, McLoughlin MA, Bond L, Wilson R, Giotopoulos G, Shanmugiah VM, Bakar RB, Yankova E, Cooper JL, Narayan N, Horton SJ, Asby R, Pask DC, Mupo A, Duddy G, Marando L, Georgomanolis T, Carter P, Ramesh AP, Dunn WG, Barcena C, Gallipoli P, Yusa K, Petrovski S, Wright P, Quiros PM, Frezza C, Nathan JA, Kaser A, Kar S, Tzelepis K, Mitchell J, Fabre MA, Huntly BJP, Vassiliou GS, Mitochondrial metabolism sustains DNMT3A-R882-mutant clonal haematopoiesis. Nature 642, 431–441 (2025). [PubMed: 40239706]
50. Bolouri H, Farrar JE, Triche T, Ries RE, Lim EL, Alonzo TA, Ma Y, Moore R, Mungall AJ, Marra MA, Zhang J, Ma X, Liu Y, Liu Y, Auvil JMG, Davidsen TM, Gesuwan P, Hermida LC, Salhia B, Capone S, Ramsingh G, Zwaan CM, Noort S, Piccolo SR, Kolb EA, Gamis AS, Smith MA, Gerhard DS, Meshinchi S, The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med 24, 103–112 (2018). [PubMed: 29227476]
51. Bottomly D, Long N, Schultz AR, Kurtz SE, Tognon CE, Johnson K, Abel M, Agarwal A, Avaylon S, Benton E, Blucher A, Borate U, Braun TP, Brown J, Bryant J, Burke R, Carlos A, Chang BH, Cho HJ, Christy S, Coblentz C, Cohen AM, d'Almeida A, Cook R, Danilov A, Dao KHT, Degnin M, Dibb J, Eide CA, English I, Hagler S, Harrelson H, Henson R, Ho H, Joshi SK, Junio B, Kaempf A, Kosaka Y, Laderas T, Lawhead M, Lee H, Leonard JT, Lin C, Lind EF, Liu SQ, Lo P, Loriaux MM, Luty S, Maxson JE, Macey T, Martinez J, Minnier J, Monteblanco A, Mori M, Morrow Q, Nelson D, Ramsdill J, Rofelty A, Rogers A, Romine KA, Ryabinin P, Saultz JN, Sampson DA, Savage SL, Schuff R, Searles R, Smith RL, Spurgeon SE, Sweeney T, Swords RT, Thapa A, Thiel-Klare K, Traer E, Wagner J, Wilmot B, Wolf J, Wu G, Yates A, Zhang H, Cogle CR, Collins RH, Deininger MW, Hourigan CS, Jordan CT, Lin TL, Martinez ME, Pallapati RR, Pollyea DA, Pomicter AD, Watts JM, Weir SJ, Druker BJ, McWeeney SK, Tyner JW, Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. Cancer Cell 40, 850–864.e9 (2022). [PubMed: 35868306]
52. Tothova Z, Krill-Burger JM, Popova KD, Landers CC, Sievers QL, Yudovich D, Belizaire R, Aster JC, Morgan EA, Tsherniak A, Ebert BL, Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. Cell Stem Cell 21, 547–555 (2017). [PubMed: 28985529]
53. Byers RJ, Currie T, Tholouli E, Rodig SJ, Kutok JL, MSI2 protein expression predicts unfavorable outcome in acute myeloid leukemia. Blood 118, 2857–2867 (2011). [PubMed: 21753187]
54. Ito T, Kwon HY, Zimdahl B, Congdon KL, Blum J, Lento WE, Zhao C, Lagoo A, Gerrard G, Foroni L, Goldman J, Goh H, Kim SH, Kim DW, Chuah C, Oehler VG, Radich JP, Jordan CT, Reya T, Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature 466, 765–768 (2010). [PubMed: 20639863]
55. Hütter G, Nowak D, Mossner M, Ganepola S, Müßig A, Allers K, Schneider T, Hofmann J, Kücherer C, Blau O, Blau IW, Hofmann WK, Thiel E, Long-Term Control of HIV by CCR5 Delta32/ Delta32 Stem-Cell Transplantation. New England Journal of Medicine 360, 692–698 (2009). [PubMed: 19213682]
56. Sabatine MS, PCSK9 inhibitors: clinical evidence and implementation. Nature reviews cardiology 16, 155–165 (2019). [PubMed: 30420622]
57. Lu HY, Orkin SH, Sankaran VG, Fetal Hemoglobin Regulation in Beta-Thalassemia. Hematol Oncol Clin North Am 37, 301–312 (2023). [PubMed: 36907604]
58. Sankaran VG, Menne T, Xu J, Akie T, Lettre G, Van Handel B, Mikkola H, Hirschhorn J, Cantor A, Orkin S, Human Fetal Hemoglobin Expression is Regulated by the Developmental Stage-Specific Repressor BCL11A. Science (1979) 322, 1839–1842 (2008).
59. Minikel EV, Karczewski KJ, Martin HC, Cummings BB, Whiffin N, Rhodes D, Alföldi J, Trembath RC, van Heel DA, Daly MJ, Alföldi J, Armean IM, Banks E, Bergelson L, Cibulskis K, Collins RL, Connolly KM, Covarrubias M, Cummings BB, Donnelly S, Farjoun Y, Ferriera S, Francioli L, Gabriel S, Gauthier LD, Gentry J, Gupta N, Jeandet T, Kaplan D, Karczewski KJ, Laricchia KM, Llanwarne C, Minikel EV, Munshi R, Neale BM, Novod S, O'Donnell-Luria AH, Petrillo N, Poterba T, Roazen D, Ruano-Rubio V, Saltzman A, Samocha KE, Schleicher M, Seed C, Solomonson M, Soto J, Tiao G, Tibbetts K, Tolonen C, Vittal C, Wade G, Wang A, Wang Q, Ware JS, Watts NA, Weisburd B, Whiffin N, Aguilar Salinas CA, Ahmad T, Albert CM, Ardissino D, Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonnycastle LL, Bottinger EP,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Bowden DW, Bown MJ, Chambers JC, Chasman D, Cho J, Chan JC, Cohen B, Correa A, Dabelea D, Darbar D, Duggirala R, Dupuis J, Ellinor PT, Elosua R, Erdmann J, Esko T, Färkkilä M, Florez J, Franke A, Getz G, Glaser B, Glatt SJ, Goldstein D, Gonzalez C, Groop L, Haiman C, Hanis C, Harms M, Hiltunen M, Holi MM, Hultman CM, Kallela M, Kaprio J, Kathiresan S, Kim BJ, Kim YJ, Kirov G, Kooner J, Koskinen S, Krumholz HM, Kugathasan S, Kwak SH, Laakso M, Lehtimäki T, Loos RJF, Lubitz SA, Ma RCW, MacArthur DG, Marrugat J, Mattila KM, McCarroll S, McCarthy MI, McGovern D, McPherson R, Meigs JB, Melander O, Metspalu A, Nilsson PM, O'Donovan MC, Ongur D, Orozco L, Owen MJ, Palmer CNA, Palotie A, Park KS, Pato C, Pulver AE, Rahman N, Remes AM, Rioux JD, Ripatti S, Roden DM, Saleheen D, Salomaa V, Samani NJ, Scharf J, Schunkert H, Shoemaker MB, Sklar P, Soininen H, Sokol H, Spector T, Sullivan PF, Suvisaari J, Tai ES, Teo YY, Tiinamaija T, Tsuang M, Dan Turner T, Tusie-Luna T, Vartiainen E, Watkins H, Weersma RK, Wessman M, Wilson JG, Xavier RJ, Schreiber SL, Evaluating drug targets through human loss-of-function genetic variation. Nature 581, 459–464 (2020). [PubMed: 32461653]

60. Joshi P, Keyvani Chahi A, Liu L, Moreira S, Vujovic A, Hope KJ, RNA binding protein-directed control of leukemic stem cell evolution and function. HemaSphere 8, e116 (2024). [PubMed: 39175825]
61. Palacios F, Yan XJ, Ferrer G, Chen SS, Vergani S, Yang X, Gardner J, Barrientos JC, Rock P, Burack R, Kolitz JE, Allen SL, Kharas MG, Abdel-Wahab O, Rai KR, Chiorazzi N, Musashi 2 influences chronic lymphocytic leukemia cell survival and growth making it a potential therapeutic target. Leukemia 35, 1037–1052 (2021). [PubMed: 33504942]
62. Zhang J, Duan Y, Wu P, Chang Y, Wang Y, Hu T, Liu C, Chen X, Zong S, Chen X, Wu Y, Jin L, Lan Y, Liu X, Cheng X, Ding F, Li T, Chen X, Guo Y, Chen Y, Yang W, Zhang L, Zou Y, Cheng T, Zhu X, Zhang Y, Clonal evolution dissection reveals that a high MSI2 level promotes chemoresistance in T-cell acute lymphoblastic leukemia. Blood 143, 320–335 (2024). [PubMed: 37801708]
63. Rajbhandari N, Hamilton M, Quintero CM, Ferguson LP, Fox R, Schürch CM, Wang J, Nakamura M, Lytle NK, McDermott M, Diaz E, Pettit H, Kritzik M, Han H, Cridebring D, Wen KW, Tsai S, Goggins MG, Lowy AM, Wechsler-Reya RJ, Von Hoff DD, Newman AM, Reya T, Single-cell mapping identifies MSI+ cells as a common origin for diverse subtypes of pancreatic cancer. Cancer Cell 41, 1989–2005 (2023). [PubMed: 37802055]
64. Li M, Li AQ, Zhou SL, Lv H, Wei P, Yang WT, RNA-binding protein MSI2 isoforms expression and regulation in progression of triple-negative breast cancer. Journal of Experimental and Clinical Cancer Research 39, 92 (2020). [PubMed: 32448269]
65. Ferrari A, Fiocca R, Bonora E, Domizio C, Fonzi E, Angeli D, Domenico Raulli G, Mattioli S, Martinelli G, Molinari C, Detection of a Novel MSI2-C17orf64 Transcript in a Patient with Aggressive Adenocarcinoma of the Gastroesophageal Junction: A Case Report. Genes (Basel) 14, 918 (2023). [PubMed: 37107676]
66. Wang S, Li N, Yousefi M, Nakauka-Ddamba A, Li F, Parada K, Rao S, Minuesa G, Katz Y, Gregory BD, Kharas MG, Yu Z, Lengner CJ, Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. Nat Commun 6, e6517 (2015).
67. Agarwal G, Antoszewski M, Xie X, Pershad Y, Arora UP, Poon C-L, Lee AJ, Guo C-J, Ye T, Norford LB, Neehus A-L, Volpe L. d., Wahlster L, Ranasinghe D, Ho T-C, Barlowe TS, Chow A, Schurer A, Taggart J, Durham BH, Abdel-Wahab O, McGraw KL, Allan JM, Soldatov R, Bick AG, Kharas MG, Sankaran VG, sankaranlab/chip_resilience, Version v1, Zenodo (2025). 10.5281/zenodo.17443511
68. Vlasschaert C, Mack T, Brett Heimlich J, Niroula A, Mesbah Uddin M, Weinstock J, Sharber B, Silver AJ, Xu Y, Savona M, Gibson C, Lanktree MB, Rauh MJ, Ebert BL, Natarajan P, Jaiswal S, Bick AG, A practical approach to curate clonal hematopoiesis of indeterminate potential in human genetic data sets. Blood 141, 2214–2223 (2023). [PubMed: 36652671]
69. Lin WY, Fordham SE, Hungate E, Sunter NJ, Elstob C, Xu Y, Park C, Quante A, Strauch K, Gieger C, Skol A, Rahman T, Sucheston-Campbell L, Wang J, Hahn T, Clay-Gilmour AI, Jones GL, Marr HJ, Jackson GH, Menne T, Collin M, Ivey A, Hills RK, Burnett AK, Russell NH, Fitzgibbon J, Larson RA, Le Beau MM, Stock W, Heidenreich O, Alharbi A, Allsup DJ, Houlston RS, Norden J, Dickinson AM, Douglas E, Lendrem C, Daly AK, Palm L, Piechocki K, Jeffries S, Bornhäuser M, Röllig C, Altmann H, Ruhnke L, Kunadt D, Wagenführ L, Cordell HJ, Darlay R, Andersen

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

MK, Fontana MC, Martinelli G, Marconi G, Sanz MA, Cervera J, Gómez-Seguí I, Cluzeau T, Moreilhon C, Raynaud S, Sill H, Voso MT, Lo-Coco F, Dombret H, Cheok M, Preudhomme C, Gale RE, Linch D, Gaal-Wesinger J, Masszi A, Nowak D, Hofmann WK, Gilkes A, Porkka K, Milosevic Feenstra JD, Kralovics R, Grimwade D, Meggendorfer M, Haferlach T, Krizsán S, Bödör C, Stölzel F, Onel K, Allan JM, Genome-wide association study identifies susceptibility loci for acute myeloid leukemia. Nat Commun 12, e6233 (2021).

70. McGraw KL, Cheng CH, Chen YA, Hou HA, Nilsson B, Genovese G, Cluzeau T, Pellagatti A, Przychodzen BP, Mallo M, Arenillas L, Mohamedali A, Adès L, Sallman DA, Padron E, Sokol L, Moreilhon C, Raynaud S, Tien HF, Boultwood J, Ebert BL, Sole F, Fenaux P, Mufti GJ, Maciejewski JP, Kanetsky PA, List AF, Non-del(5q) myelodysplastic syndromes–associated loci detected by SNP-array genome-wide association meta-analysis. Blood Adv 3, 3579–3589 (2019). [PubMed: 31738830]
71. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Márquez-Luna C, O'Connor L, Pirinen M, Finucane HK, Price AL, Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nat Genet 52, 1355–1363 (2020). [PubMed: 33199916]
72. Wang G, Sarkar A, Carbonetto P, Stephens M, A simple new approach to variable selection in regression, with application to genetic fine mapping. Journal of the Royal Statistical Society Series B: Statistical Methodology 82, 1273–1300 (2020). [PubMed: 37220626]
73. Yang Z, Wang C, Liu L, Khan A, Lee A, Vardarajan B, Mayeux R, Kiryluk K, Ionita-Laza I, CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses. Nat Genet 55, 1057–1065 (2023). [PubMed: 37169873]
74. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M, FINEMAP: Efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32, 1493–1501 (2016). [PubMed: 26773131]
75. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, Chow ED, Kanterakis E, Gao H, Kia A, Batzoglou S, Sanders SJ, Farh KKH, Predicting Splicing from Primary Sequence with Deep Learning. Cell 176, 535–548 (2019). [PubMed: 30661751]
76. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, Majeti R, Chang HY, Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet 48, 1193–1203 (2016). [PubMed: 27526324]
77. Coetzee SG, Coetzee GA, Hazelett DJ, MotifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics 31, 3847–3849 (2015). [PubMed: 26272984]
78. de Smith AJ, Wahlster L, Jeon S, Kachuri L, Black S, Langie J, Cato LD, Nakatsuka N, Chan TF, Xia G, Mazumder S, Yang W, Gazal S, Eng C, Hu D, Burchard EG, Ziv E, Metayer C, Mancuso N, Yang JJ, Ma X, Wiemels JL, Yu F, Chiang CWK, Sankaran VG, A noncoding regulatory variant in IKZF1 increases acute lymphoblastic leukemia risk in Hispanic/Latino children. Cell Genomics 4, e100526 (2024).
79. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L, CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nature biotechnology 37, 224–226 (2019).
80. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, Hook PW, Koren S, Rautiainen M, Alexandrov IA, Allen J, Asri M, Bzikadze AV, Chen NC, Chin CS, Diekhans M, Flicek P, Formenti G, Fungtammasan A, Garcia Giron C, Garrison E, Gershman A, Gerton JL, Grady PGS, Guarracino A, Haggerty L, Halabian R, Hansen NF, Harris R, Hartley GA, Harvey WT, Haukness M, Heinz J, Hourlier T, Hubley RM, Hunt SE, Hwang S, Jain M, Kesharwani RK, Lewis AP, Li H, Logsdon GA, Lucas JK, Makalowski W, Markovic C, Martin FJ, Mc Cartney AM, McCoy RC, McDaniel J, McNulty BM, Medvedev P, Mikheenko A, Munson KM, Murphy TD, Olsen HE, Olson ND, Paulin LF, Porubsky D, Potapova T, Ryabov F, Salzberg SL, Sauria MEG, Sedlazeck FJ, Shafin K, Shepelev VA, Shumate A, Storer JM, Surapaneni L, Taravella Oill AM, Thibaud-Nissen F, Timp W, Tomaszkiewicz M, Vollger MR, Walenz BP, Watwood AC, Weissensteiner MH, Wenger AM, Wilson MA, Zarate S, Zhu Y, Zook JM, Eichler EE, O'Neill

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RJ, Schatz MC, Miga KH, Makova KD, Phillippy AM, The complete sequence of a human Y chromosome. Nature 621, 344–354 (2023). [PubMed: 37612512]

81. Li H, Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100 (2018). [PubMed: 29750242]
82. Huang X, Huang Y, Cellsnp-lite: An efficient tool for genotyping single cells. Bioinformatics 37, 4569–4571 (2021). [PubMed: 33963851]
83. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Twelve years of SAMtools and BCFtools. Gigascience 10, giab008 (2021). [PubMed: 33590861]
84. Duttke SH, Guzman C, Chang M, Delos Santos NP, McDonald BR, Xie J, Carlin AF, Heinz S, Benner C, Position-dependent function of human sequence-specific transcription factors. Nature 631, 891–898 (2024). [PubMed: 39020164]
85. Zeng AGX, Iacobucci I, Shah S, Mitchell A, Wong G, Bansal S, Chen D, Gao Q, Kim H, Kennedy JA, Arruda A, Minden MD, Haferlach T, Mullighan CG, Dick JE, Single-cell Transcriptional Atlas of Human Hematopoiesis Reveals Genetic and Hierarchy-Based Determinants of Aberrant AML Differentiation. Blood Cancer Discovery 6, 307–324 (2025). [PubMed: 40294241]
86. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S, Fast, sensitive and accurate integration of single-cell data with Harmony. Nature methods 16, 1289–96 (2019). [PubMed: 31740819]
87. Dwivedi B, Mumme H, Satpathy S, Bhasin SS, Bhasin M, Survival Genie, a web platform for survival analysis across pediatric and adult cancers. Sci Rep 12, e3069 (2022).
88. Zeng AGX, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, Chan-Seng-Yue M, Voisin V, van Galen P, Tierens A, Cheok M, Preudhomme C, Dombret H, Daver N, Futreal PA, Minden MD, Kennedy JA, Wang JCY, Dick JE, A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nat Med 28, 1212–1223 (2022). [PubMed: 35618837]
89. Mack T, Pershad Y, Vlasschaert C, Bejan CA, Heimlich JB, Li Y, Mickels NA, Van Amburg JC, Ulloa J, Silver AJ, Luo LY, Jones A, Ferrell PB, Kishtagari A, Xu Y, Savona MR, Bick AG, Germline genetics, disease, and exposure to medication influence longitudinal dynamics of clonal hematopoiesis. Haematologica 110, 1010–1018 (2025). [PubMed: 39540221]
90. Mack T, Vlasschaert C, von Beck K, Silver AJ, Heimlich JB, Poisner H, Condon HR, Ulloa J, Sochacki AL, Spaulding TP, Kishtagari A, Bejan CA, Xu Y, Savona MR, Jones A, Bick AG, Cost-Effective and Scalable Clonal Hematopoiesis Assay Provides Insight into Clonal Dynamics. Journal of Molecular Diagnostics 26, 563–573 (2024).
91. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303 (2010). [PubMed: 20644199]
92. Marcus JH, Novembre J, Visualizing the geography of genetic variants. Bioinformatics 33, 594–595 (2017). [PubMed: 27742697]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

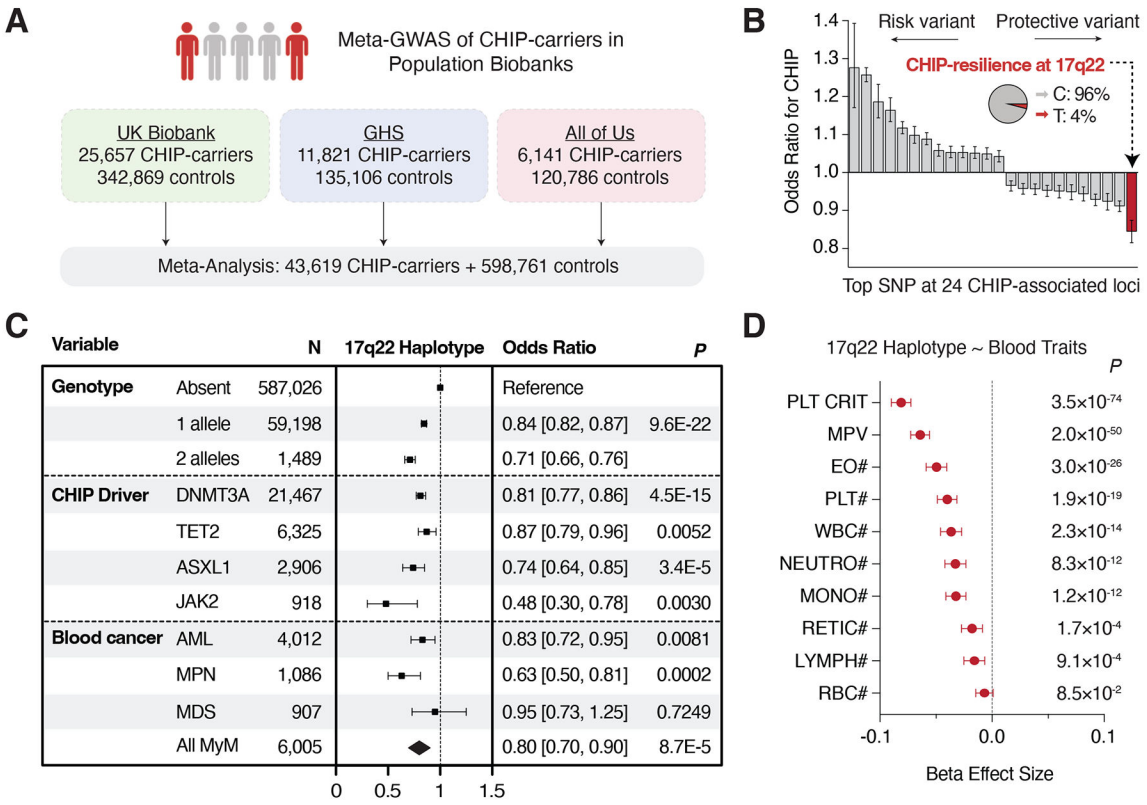


Fig. 1. Inherited resilience to CHIP and MyM at the 17q22 locus.

(**A**) Approach for meta-analysis of CHIP across UK Biobank, Geisinger Health Study (GHS), and All of Us cohorts. (**B**) Waterfall plot showing meta-analysis odds ratio effect sizes (95% CI) for the most significant variant at 24 CHIP-associated loci. (**C**) Protective effect of 17q22 haplotype for CHIP by number of alleles (top) and per-allele effects by driver mutation (middle) or for MyMs (bottom). MPN = myeloproliferative neoplasm. (**D**) Per-allele effect sizes (beta coefficients) for associations between CHIP-protective 17q22 haplotype and blood cell traits. #, count; PLT CRIT, plateletcrit; MPV, mean platelet volume; EO, eosinophil; PLT, platelet; WBC, white blood cell; NEUTRO, neutrophil, MONO, monocyte; RETIC, reticulocyte; LYMPH, lymphocyte; RBC, red blood cell. Error bars represent mean ± 95% CI.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

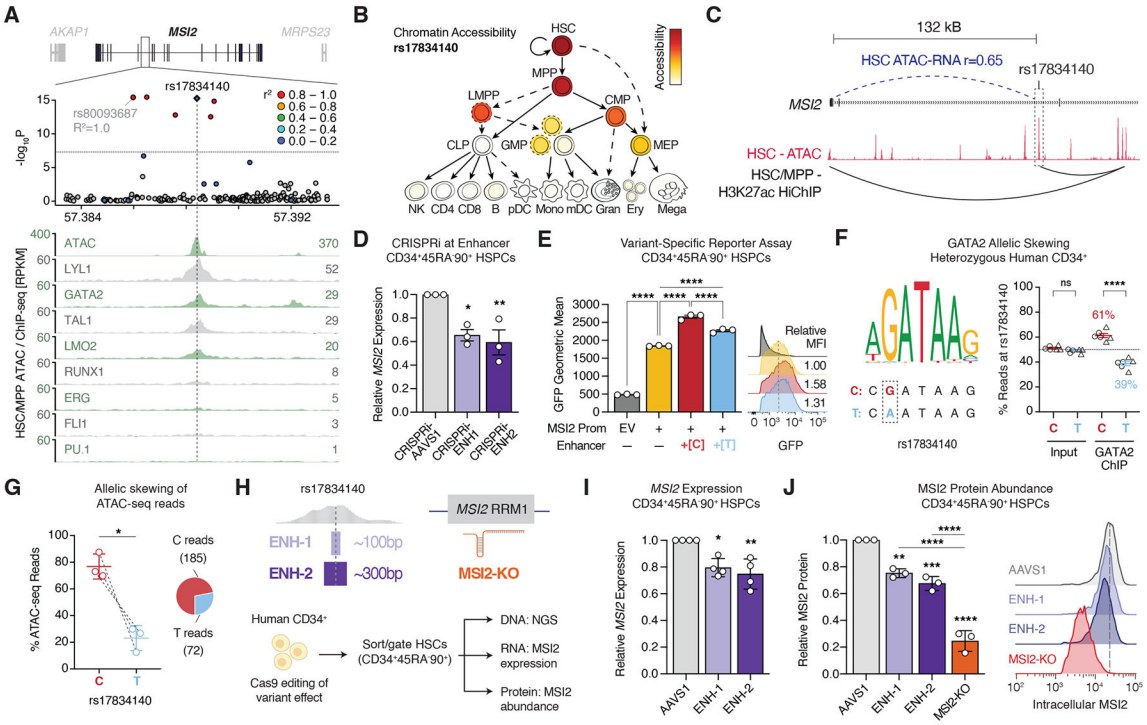


Fig. 2. Regulatory variant rs17834140-T protects from CHIP through HSC-selective down-regulation of *MSI2*.

(**A**) LocusZoom plot identifying rs17834140 at the 17q22 CHIP-resilience haplotype as a putative causal regulatory variant, with tracks for chromatin accessibility (ATAC) and transcription factor binding (ChIP-seq) in primary human HSPCs. Normalized reads per kilobase of transcript per million (RPKM) counts at rs17834140 are shown on the right side. (**B**) Normalized chromatin accessibility of the rs17834140-harboring regulatory element in hematopoietic cells, indicating HSC-selective regulatory activity. CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor; LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; MPP, multipotent progenitor. (**C**) H3K27ac-HiChIP track showing physical contact between rs17834140 and the *MSI2* promoter in human HSPCs (29). (**D**) Reduced *MSI2* mRNA expression in $CD34^{+}CD45RA^{-}CD90^{+}$ HSC-enriched cells, 3 days after CRISPR interference targeting the rs17834140-harboring regulatory element. $n = 3$. (**E**) Reporter assay showing enhancer activity of rs17834140 wild-type (C) or CHIP-resilience (T) alleles in $CD34^{+}CD45RA^{-}CD90^{+}$ HSC-enriched cells, with relative geometric mean fluorescence intensity (MFI) relative to *MSI2*-promoter alone on the right. N=3. EV, empty vector. (**F**) ChIP showing reduced GATA2 binding at rs17834140 in primary $CD34^{+}$ HSPCs, across $n = 2$ heterozygous donors (marked as triangles or circles). **(G)** Allele-specific ATAC-seq reads at rs17834140 in HSPCs showing reduced chromatin accessibility with the resilience (T) allele. $n = 3$. (**H**) Editing strategy modeling rs17834140-T effect with ~100-bp (ENH-1) or ~300-bp (ENH-2) microdeletions, or MSI2-KO. NGS, next-generation sequencing. (**I**) Reduced *MSI2* mRNA expression in $CD34^{+}CD45RA^{-}CD90^{+}$ HSC-enriched cells, 3 days after enhancer microdeletions. $n = 4$. (**J**) Flow cytometry showing reduced intracellular MSI2 protein abundance in $CD34^{+}CD45RA^{-}CD90^{+}$ gated HSC-enriched cells,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3 days after editing. $n = 3$. Error bars represent mean ± SEM. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns is not significant.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

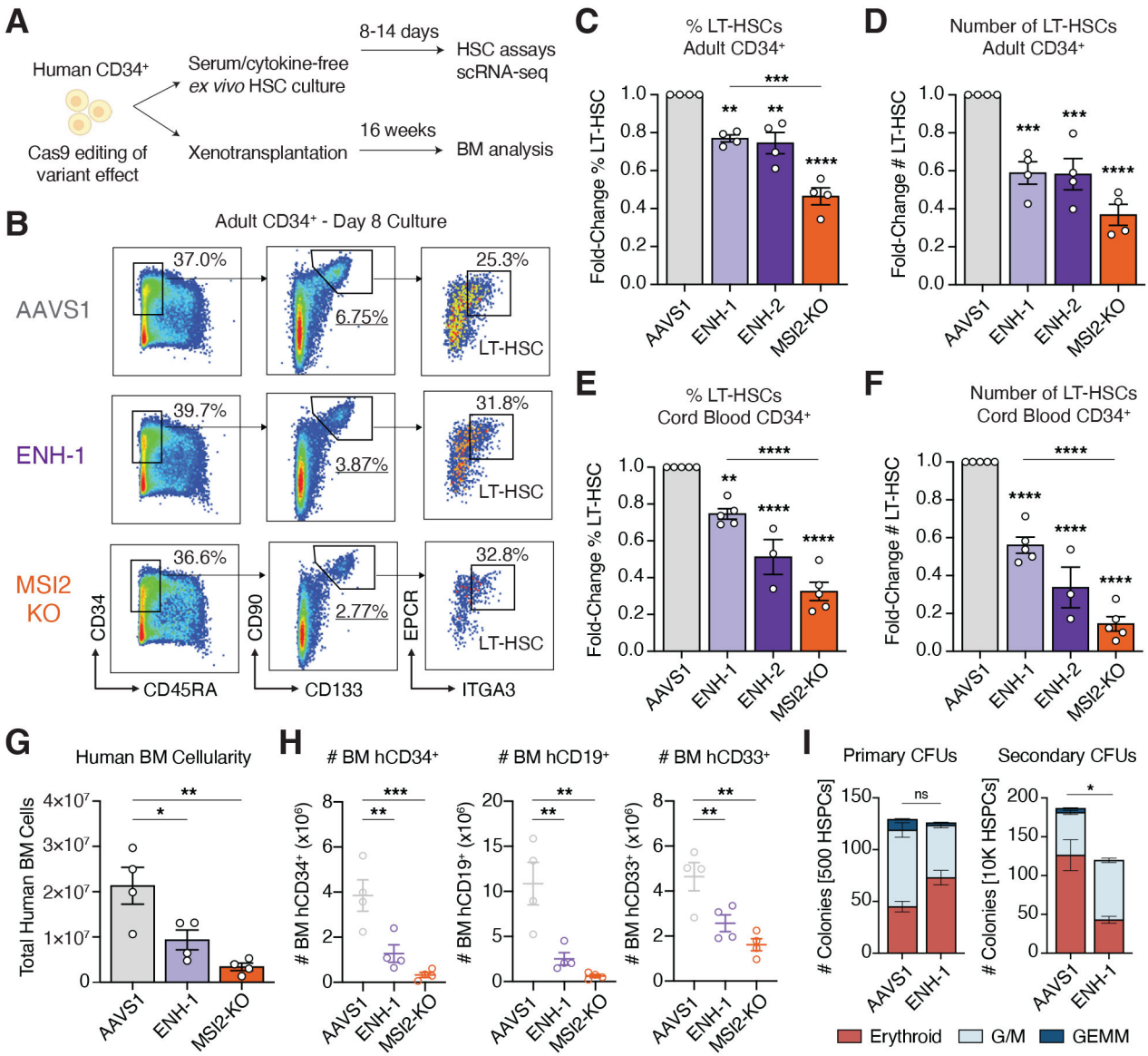


Fig. 3. Genetic variation–driven loss of MSI2 enhancer reduces human HSC fitness.

(**A**) Outline of HSC assays used to assess CHIP-resilience variant effect. (**B**) Representative flow cytometry gating for phenotypic LT-HSC ($CD34^+CD45RA^-CD90^+CD133^+EPCR^+ITGA3^+$) immunophenotyping, 6 days after editing adult $CD34^+$ HSPCs. (**C**) Proportion of LT-HSCs on day 8 of serum-free adult $CD34^+$ cell culture. $n = 4$. (**D**) Total numbers of LT-HSCs on day 8 of serum-free adult $CD34^+$ cell culture. $n = 4$. (**E**) Proportion of LT-HSCs on day 14 of cytokine-free cord blood $CD34^+$ cell culture. $n = 3$ to 5. (**F**) Total numbers of LT-HSCs on day 14 of cytokine-free cord blood $CD34^+$ cell culture. $n = 3$ to 5. (**G**) Total human cells achieving long-term engraftment in bone marrow BM of NBSGW mice, 16 weeks after transplantation with AAVS1-, ENH-1– or MSI2-KO–edited cord blood HSPCs. $n = 12$ mice. (**H**) Total human $CD45^+$ engrafted cells in BM expressing lineage markers ($CD34^+$ HSPCs, $CD19^+$ B cells, $CD33^+$ myeloid cells). $n = 12$ mice. **(I)** Colony-forming units (CFUs) formed in primary

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

plating (from 500 HSPCs) and secondary replating (from 10,000 HSPCs), showing loss of self-renewal capacity in ENH-1–edited cells. G/M, granulocyte-macrophage; GEMM, granulocyte-erythrocyte-monocyte-megakaryocyte. $N = 3$. Error bars represent mean ± SEM. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns is not significant.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

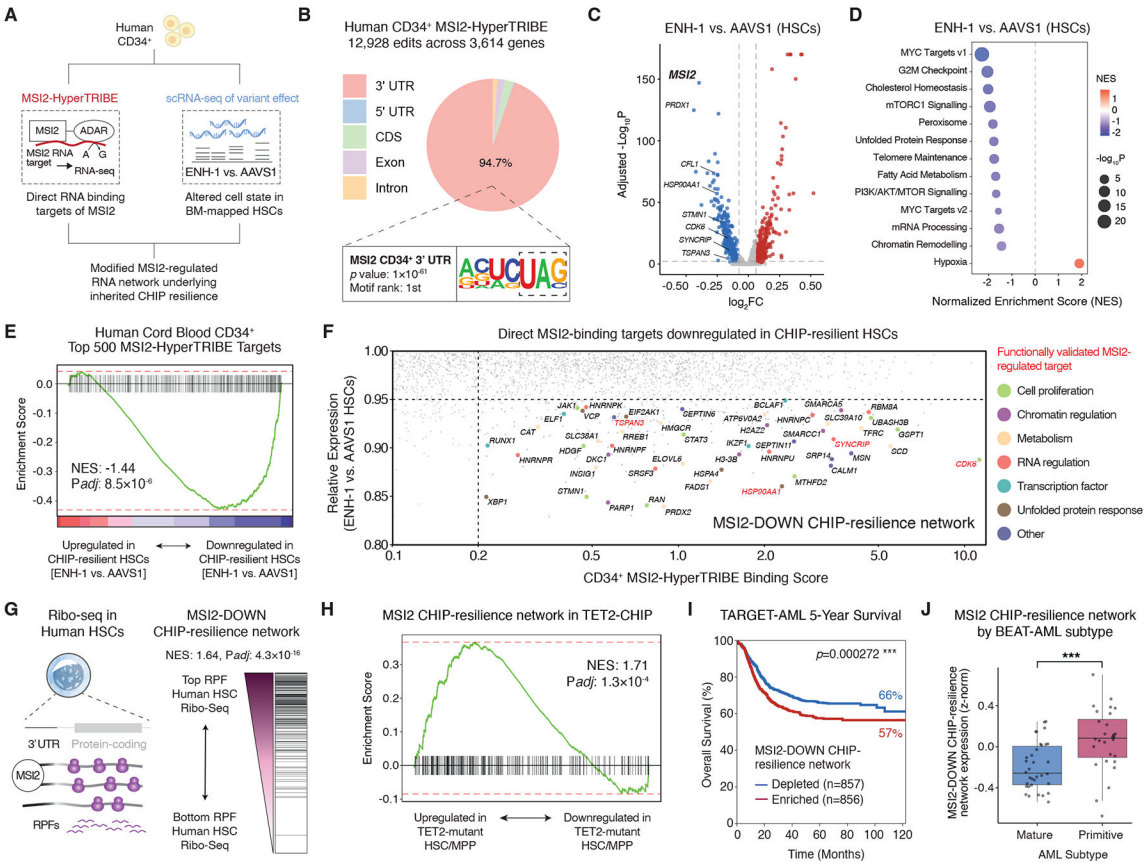


Fig. 4. Modified RNA regulation underlying CHIP-resilience in human HSCs.

(**A**) Overview of approach to determine MSI2-regulated RNA network in human HSCs. (**B**) MSI2-HyperTRIBE edit sites are enriched in the 3’ UTR of target mRNAs with the UAG MSI2-recognition motif. (**C**) Volcano plot of differentially expressed genes (DEGs) between ENH-1– versus AAVS1-edited HSCs. (**D**) Gene set enrichment analysis of ENH-1– versus AAVS1-edited HSCs. (**E**) CHIP-resilient HSCs (ENH-1 versus AAVS1) show negative enrichment for CD34^{+} MSI2-HyperTRIBE mRNA targets. (**F**) Identification of an RNA network that is MSI2-bound and functionally down-regulated in CHIP-resilient HSCs. **(G)** Ribosome sequencing (Ribo-seq) profiling of human HSCs showing that CHIP-resilience network mRNAs are enriched for ribosome protected fragments (RPFs). (**H**) *TET2* mutant HSC and MPPs show positive enrichment for the MSI2-DOWN CHIP-resilience network, relative to nonmutant cells from the same donors. (**I**) Overall survival of pediatric patients with AML in the TARGET-AML cohort, stratified by median expression of the MSI2-DOWN CHIP-resilience network. **(J)** Expression of the MSI2-DOWN CHIP-resilience network in BEAT-AML diagnostic BM samples, stratified by predominantly mature versus primitive cellular profiles. The center line represents the median, box limits are upper and lower quartiles, and whiskers indicate 1.5× the interquartile range. ***$P < 0.001$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

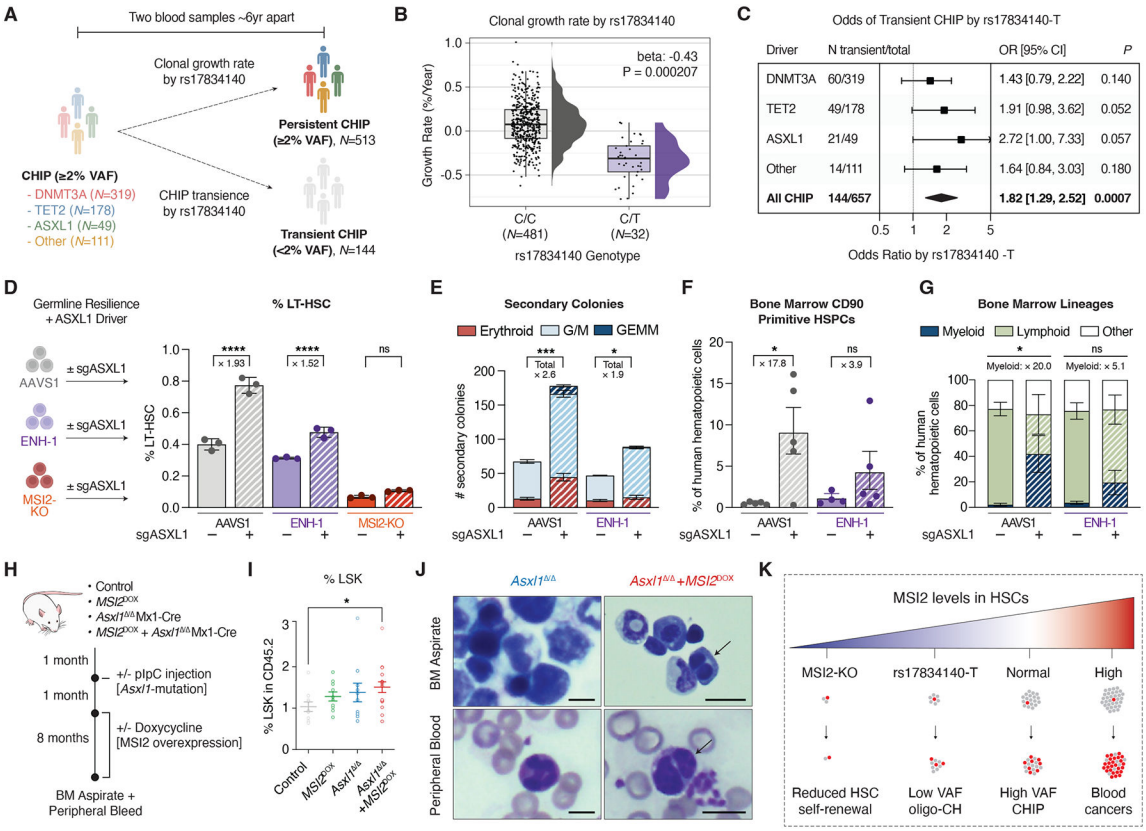


Fig. 5. Stem cell MSI2 levels modify clonal dominance of *ASXL1*-mutant HSCs.

(A) Approach to assess associations between rs17834140-T and longitudinal CHIP progression in a cohort with serial blood samples ~6 years apart. (**B**) Growth rate (% VAF/year) of CHIP-mutations stratified by C/C (n = 481) or C/T (n = 32) genotype at rs17834140. The center line represents the median, box limits are upper and lower quartiles, and whiskers indicate 1.5× the interquartile range. (**C**) Odds of CHIP transience (≥2% VAF at baseline, <2% VAF at follow-up) by rs17834140-T for driver mutations. (**D**) Dual modeling of germline CHIP-resilience (ENH-1 microdeletion, MSI2-KO, AAVS1 control) with *ASXL1* mutation in primary adult $CD34^+$ HSPCs, quantifying LT-HSCs 7 days afer editing. N = 3. (**E**) Colony-forming units formed after secondary replating, showing that ENH-1 editing attenuates expansion induced through *ASXL1* mutation. G/M, granulocyte-macrophage; GEMM, granulocyte-erythroid-monocyte-megakaryocyte. N = 3. **(F)** Proportions of engrafted human cells expressing CD90 in BM of NBSGW mice after xenotransplantation. n = 19 mice. **(G)** Proportions of human engrafted cells expressing myeloid (CD33) or lymphoid (CD19 or CD3) lineage markers in BM of NBSGW mice after xenotransplantation. n = 19 mice. **(H)** Workflow of murine model used to evaluate MSI2 and Asxl1 cooperativity. (**I**) Quantification of LSK cells, 1 month after doxycycline treatment in murine model. n = 47 mice. (**J**) Representative smears from BM and peripheral blood; arrows indicate MDS features (hyposegmented Pelger-Huët-like neutrophils and binucleate erythroid precursors) in $Asxl1^{\Delta/\Delta}$+$MSI2^{DOX}$ mice. Scale bars are 20 μm. (**K**) Model of how MSI2 levels in HSCs regulate stem cell pool size (gray) and clonal advantage of somatic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

CHIP driver mutations (red). Error bars represent mean ± SEM. *$P < 0.05$; ***$P < 0.001$; ****$P < 0.0001$; ns is not significant.