Supplemental Materials

Supplemental Methods

Supplemental Table 1. Patient characteristics, clinical parameters and variants calls for all study participants.

Supplemental Figure 1. Timeline and number of sample collections for all study patients.

Supplemental Figure 2. Blood count indices for patients with CH in comparison to patients without CH.

Supplemental Figure 3. Age, comorbidities, and medical history of patients with CH in comparison to patients without CH.

Supplemental References

Supplemental Methods

Study Population and Sample Collection

Patients with planned hip replacement surgeries at the University of Chicago Medical Center were enrolled on an Institutional Review Board-approved study from April/2018 to March/2021. Peripheral blood (PB) samples were collected at the time of the pre-operative anesthesia visit, which generally occurred about three weeks prior to surgery, or on the day of surgery. PB samples were also collected from a subset of patient at their post-operative follow up visit and at subsequent visits to our center for other purposes. Hematopoietic cells from bone marrow (BM) samples were collected from femoral heads that were bisected during surgery and rocked in RPMI supplemented with 10% Penicillin/Streptomycin for 2-4 hours. A few femoral heads were incubated in RPMI supplemented with 10% Penicillin/Streptomycin overnight at 4 degrees Celsius due to logistical aspects. The resultant cell suspension was passed through a 100µm cell strainer and centrifuged at 1200rpm for 5 minutes. Cell pellets were suspended in 5mL of cold PBS and passed through a 70µm cell strainer and spun again. RBC lysis was performed on the resultant pellet using an RBC lysis buffer consisting of 155 mM NH_4CL, 12 mM $NaHCO_3$ and 0.1 mM EDTA. Pellets were resuspended in 25mL of 1x RBC lysis buffer and incubated for 10 minutes at room temperature. Similarly, RBC lysis was performed on peripheral blood samples through resuspension in RBC lysis buffer. Resuspended cells were then centrifuged at 1200rpm for 5 minutes. Supernatant was aspirated and cells were resuspended in 5mL of PBS. Cells were counted using trypan blue dye (10uL in 90uL trypan blue). Resuspended cells were centrifuged at 1200rpm for 5 minutes. Resultant pellets were resuspended in freezing media (10% v/v DMSO in FBS) and viably frozen. Cells were thawed rapidly at 37°C and centrifuged at 1200rpm for 5 minutes. Freezing media was removed, and cells were washed with PBS and resuspended in lysis buffer (1mg/ml proteinase K in buffer consisting of 50 mM Tris, pH 8.0, 100 mM EDTA, 100 mM NaCl and 1% SDS). Phenol chloroform DNA extraction was performed on all BM and PB samples collected in separate batches. Resultant DNA was quantified using fluorimetric methods.

Next Generation Sequencing and Variant Calling

DNA libraries containing unique molecular identifiers allowing for error correction were prepared with the Kapa HyperPrep Kit (Kapa Biosystems) followed by target enrichment (Twist Biosciences) with a customized assay designed specifically to detect clonal hematopoiesis (CH) and other disease associated risk alleles [1]. The final enriched material was sequenced to >2,000X mean overall depth of coverage. Gencore v0.13. and fastp v0.20.1[2,3] were used to trim adaptor sequences and remove duplicates. After alignment with BWA MEM v0.7.17 [4], VarDict-Java v1.7.0 [5] was used to call variants in the reads. Bcftools v1.10.2[6], SnpEff v4.3 [7], and Vcflib v1.0.1 [8] were used to filter out reads for mapping quality, depth, and strand bias. Variant-Effect-Predictor (VEP) v100.3 was used to annotate variants [9]. Filtration of artifacts and potential SNPs was performed as described [1,10]. Variants showing recurrence or detected in NA12878 reference DNA were excluded as suggestive of artifacts unless reported ≥ 10 in Cosmic database v90 or reported once in The Cancer Genome Atlas (TCGA) AML study [11]. SNPs were stringently filtered out when reported as > 0.25% in population allele frequency databases (dbSNP v153 or ExAC adjusted population allele frequency). Variants with Variant Allele Fraction (VAF) > 30% were excluded as potential germline alleles or exhibiting artifactual VAF under-representation as a result of capture bias, unless reported ≥ 10 times in COSMIC. Finally, variants with VAF ≥ 1% and VAF < 30% were reported. After these filtering steps, 3 instances of DNMT3A R882H mutation were detected above 35% VAF in one patient (patient 55), and the mutation was detected at similar VAF in the PB and BM samples. For tracking of variants in longitudinal or paired PB and BM samples, variants with VAF < 1% were included provided they had been detected in at least one of the paired/longitudinal samples passing all filtering criteria and above the VAF threshold. Manual revision of read piles using IGV software v 2.10.2 (Integrative Genomics Viewer) was performed in those cases showing discrepancies between compartments after force calling to verify absence of variants.

Genes Analyzed: *ANGPTL4, ANKRD26, APC, ASXL1, ATM, BAP1, BARD1, BMPR1A, BRAF, BRCA1, BRCA2, BRIP1, CBL, CDH1, CDK4, CDKN2A, CEBPA, CHEK2, DDX41, DNMT3A, DSC2, DSG2, EPCAM, ETV6, FBN1, FLT3, GATA1, GATA2, GLA, GREM1, IDH1, IDH2, JAK2, KCNH2, KCNQ1, KDM1A, KRAS, MITF, MLH1, MPL, MSH2, MSH6, MUTYH, MYL2, MYL3, NBN, NPC1L1, NPM1, NRAS, PALB2, PCSK9, PKP2, PMS2, POLD1, POLE, PPM1D, PRKAG2, PTEN, RAD51C, RAD51D, RUNX1, SCN5A, SF3B1, SMAD3, SMAD4, SRP72, SRSF2, STK11, TET2, TGFBR1, TGFBR2, TMEM43, TP53, TPM1, U2AF1, ZRSR2*. Depth of coverage statistics were determined using Picard tools v2.6.055.

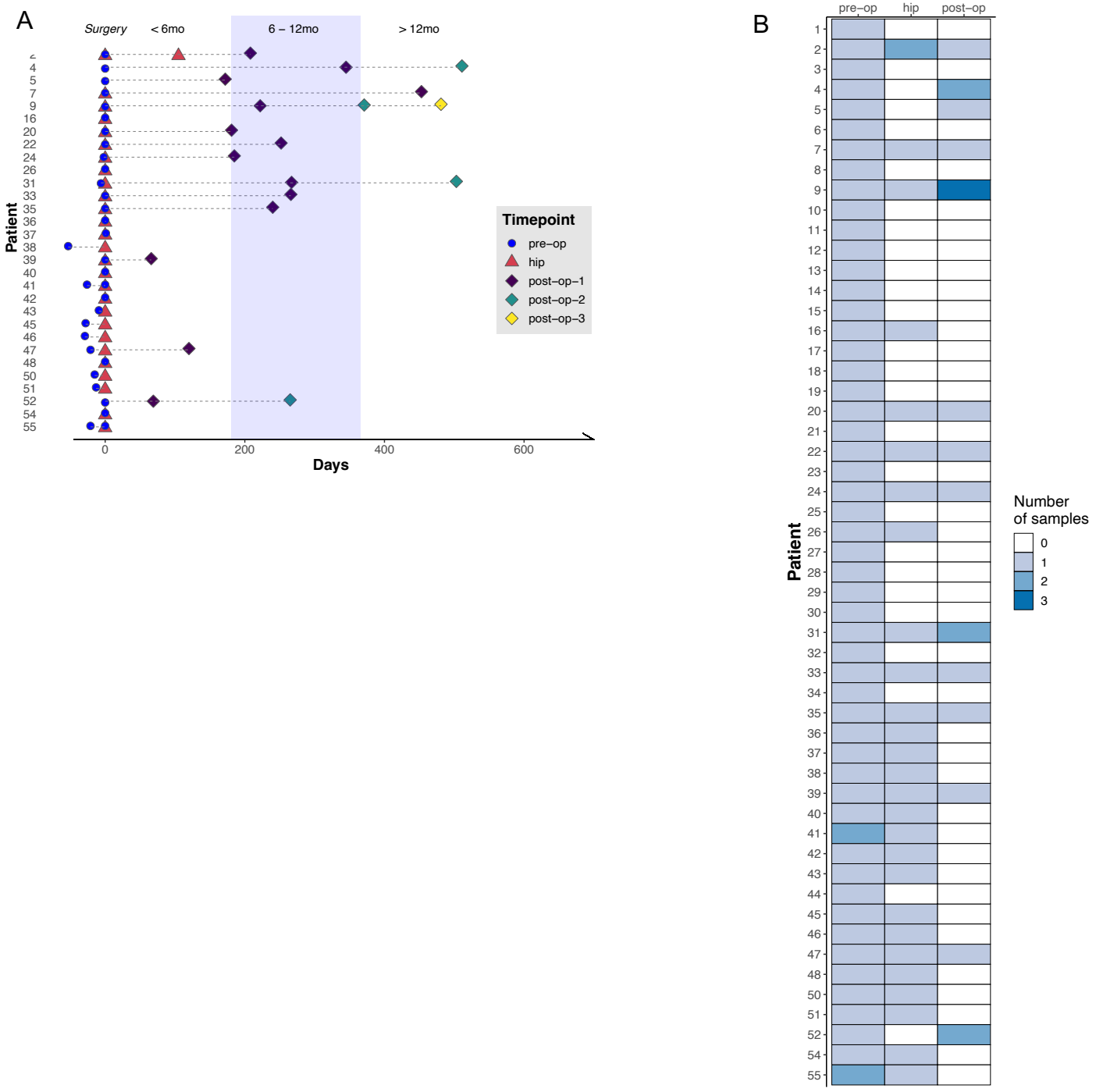
Statistical Methods

T-tests were used for continuous variables, and the Fisher's exact test for categorical variables. The Pearson correlation test was used to compare the VAFs between the PB and BM and the VAFs between pre-operative and post-operative samples in patients with CH.

All statistical analyses were performed using the *R* software v3.6.3 (R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/). Plots were produced using *ggplot2* v3.3.5 (https://github.com/tidyverse/ggplot2) and *ComplexHeatmap* v2.2.0[12]. Data summarization and reshaping were performed using *plyr* v1.8.6 (https://github.com/hadley/plyr), *dplyr* v1.0.7 (https://github.com/hadley/dplyr) and *reshape2* v1.4.4 (https://github.com/hadley/reshape). All p-values were two-sided with statistical significance set *a priori* at the 0.05 level. Ninety-five percent confidence intervals (95 % C.I.) were calculated.

Supplemental Table 1. Patient characteristics, clinical parameters and variants calls for all study participants.

P: patient identifier. WBC: white blood cell count (K/ul), Hgb: hemoglobin (g/dL), MCV: mean corpuscular volume (fL), RDW: red cell distribution width (%), PLT: platelets (K/ul). Patient 49 was excluded due to quality control (QC) failure of all samples. Patient 53 was omitted due to use of duplicate deidentifiers for the same patient.

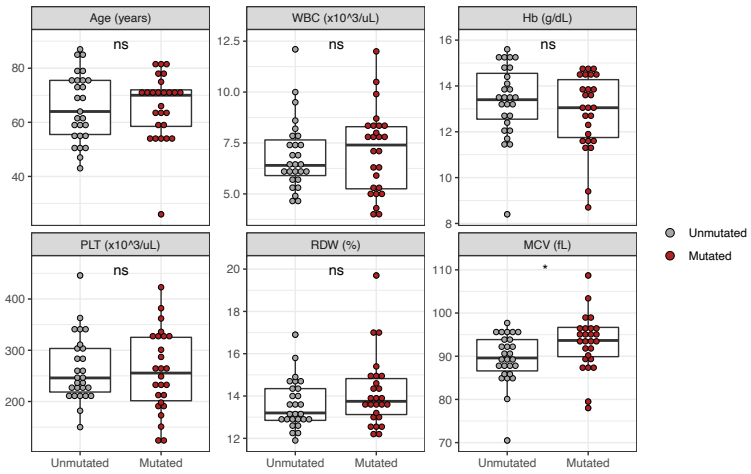


Supplemental Figure 1. Timeline and number of sample collections for all study patients. (A) Each line on the vertical axis represents a study participant. The horizontal axis represents the timeline of the study in days. Day 0 is the day of hip replacement surgery. The time of sample collections and the type of sample collected are displayed on the graph for each study participant. **(B)** Diagram showing the type of samples collected for each study participant. Each row represents a unique study participant. Patient identifier numbers are on the left side of the panel.

A

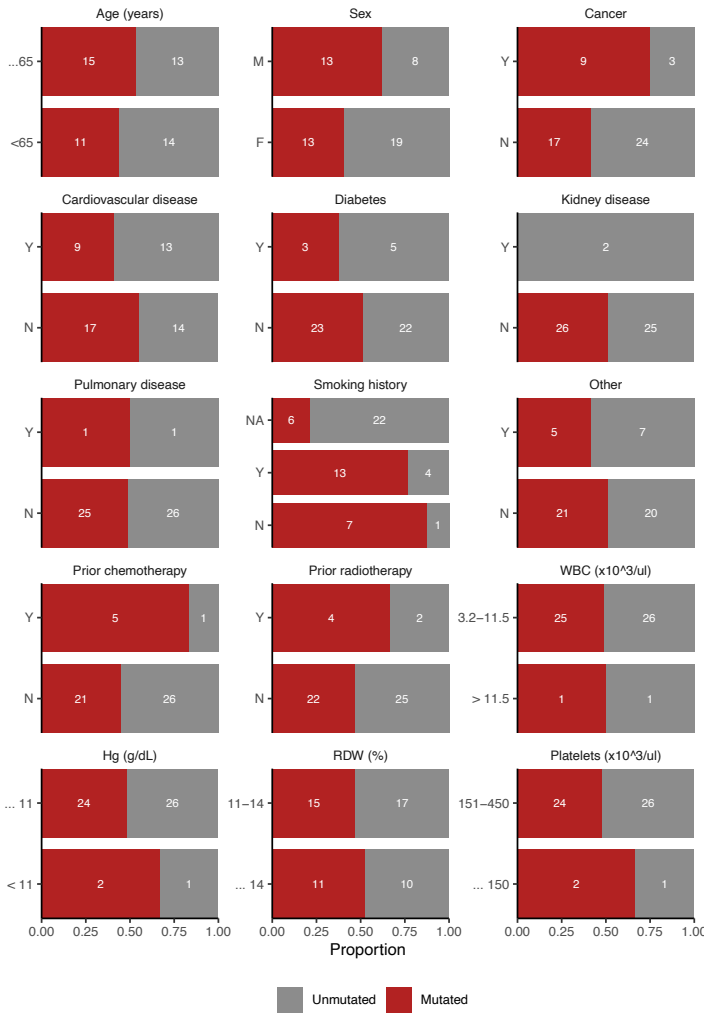
Parameter	CH- vaf 1% n = 27	CH+ vaf 1% n = 26
Age (years)	65.59 (43-87)	65.88 (26-82)
WBC (x10^3/uL)	6.89 (4.6-12.1)	7.09 (3.9-12)
Hb (g/dL)	13.36 (8.4-15.6)	12.88 (8.7-14.8)
PLT (x10^3/uL)	262.85 (150-446)	258.42 (124-423)
RDW (%)	13.63 (11.9-16.9)	14.16 (12.1-19.7)
MCV (fL)	89.56 (70.5-97.7)	93.16 (78-108.7)

B



Supplemental Figure 2. Blood count indices for patients with CH in comparison to patients without CH. (A-B) Age and blood parameters in patients with CH in comparison to patients without CH. MCV is significantly more elevated in patients with CH. WBC: white blood cell count (K/ul), Hgb: hemoglobin (g/dL), MCV: mean corpuscular volume (fL), RDW: red cell distribution width (%), PLT: platelets (K/ul).

Parameter		CH+ vaf 1% n = 26	CH- vaf 1% n = 27	OR (95%CI)	p
Age (years)	<65 (n=25)	11 (44%)	14 (56%)	1.46 (0.44-4.98)	0.5860
	≥65 (n=28)	15 (53.6%)	13 (46.4%)		
Cancer	N (n=41)	17 (41.5%)	24 (58.5%)	4.12 (0.86-27.18)	0.0537
	Y (n=12)	9 (75%)	3 (25%)		
Prior chemotherapy	N (n=47)	21 (44.7%)	26 (55.3%)	6 (0.6-303.3)	0.1000
	Y (n=6)	5 (83.3%)	1 (16.7%)		
Cardiovascular disease	N (n=31)	17 (54.8%)	14 (45.2%)	0.58 (0.16-1.97)	0.4060
	Y (n=22)	9 (40.9%)	13 (59.1%)		
Diabetes	N (n=45)	23 (51.1%)	22 (48.9%)	0.58 (0.08-3.4)	0.7040
	Y (n=8)	3 (37.5%)	5 (62.5%)		
Kidney disease	N (n=51)	26 (51%)	25 (49%)	0 (0-5.5)	0.4910
	Y (n=2)	NA	2 (100%)		
Other	N (n=41)	21 (51.2%)	20 (48.8%)	0.69 (0.15-2.99)	0.7450
	Y (n=12)	5 (41.7%)	7 (58.3%)		
Prior chemo and/or radiotherapy	N (n=45)	20 (44.4%)	25 (55.6%)	3.66 (0.57-40.93)	0.1420
	Y (n=8)	6 (75%)	2 (25%)		
Pulmonary disease	N (n=51)	25 (49%)	26 (51%)	1.04 (0.01-84.68)	1.0000
	Y (n=2)	1 (50%)	1 (50%)		
Prior radiotherapy	N (n=47)	22 (46.8%)	25 (53.2%)	2.24 (0.29-27.01)	0.4200
	Y (n=6)	4 (66.7%)	2 (33.3%)		
Sex	F (n=32)	13 (40.6%)	19 (59.4%)	0.43 (0.12-1.49)	0.1660
	M (n=21)	13 (61.9%)	8 (38.1%)		
Smoking history	N (n=8)	7 (87.5%)	1 (12.5%)	0.48 (0.01-6.21)	1.0000
	NA (n=28)	6 (21.4%)	22 (78.6%)		
	Y (n=17)	13 (76.5%)	4 (23.5%)		



Supplemental Figure 3. Age, comorbidities, and medical history of patients with CH in comparison to patients without CH. Summary of age, comorbidities and history of prior cancer, chemotherapy, radiation, and smoking for all study participants by CH status. Patients with CH appeared to be more likely to have a history of cancer than those without CH (p=0.053). The two groups are otherwise similar in terms of age and comorbidities.

References:

1. Singh A, Mencia-Trinchant N, Griffiths EA, et al. Mutant PPM1D- and TP53-Driven Hematopoiesis Populates the Hematopoietic Compartment in Response to Peptide Receptor Radionuclide Therapy. *JCO Precis Oncol*. 2022;6:e2100309.
2. Chen S, Zhou Y, Chen Y, et al. Gencore: an efficient tool to generate consensus reads for error suppressing and duplicate removing of NGS data. *BMC Bioinformatics*. 2019;20(Suppl 23):606.
3. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890.
4. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*. 2013.
5. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108.
6. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
7. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.
8. Vcflib GE. a simple C++ library for parsing and manipulating VCF files (MIT). Vol. 2021; 2016.
9. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
10. Desai P, Mencia-Trinchant N, Savenkov O, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med*. 2018;24(7):1015-1023.
11. Forbes SA, Beare D, Boutselakis H, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777-D783.
12. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849.