Experimental Hematology 2019;80:36–41

BRIEF COMMUNICATION

Inflammatory cytokines promote clonal hematopoiesis with specific mutations in ulcerative colitis patients

Christine R.C. Zhang[a], Darren Nix[b], Martin Gregory[b], Matthew A. Ciorba[b], Elizabeth L. Ostrander[a], Rodney D. Newberry[b], David H. Spencer[a], and Grant A. Challen[a]

[a]*Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO;* [b]*Division of Gastroenterology, Department of Medicine, Washington University School of Medicine, St. Louis, MO*

(Received 22 October 2019; revised 22 November 2019; accepted 25 November 2019)

Epidemiological sequencing studies have revealed that somatic mutations characteristic of myeloid neoplasms can be detected in the blood of asymptomatic individuals decades prior to presentation of any clinical symptoms. This premalignant condition is known as clonal hematopoiesis of indeterminate potential (CHIP). Despite the fact these mutant clones become readily detectable in the blood of elderly individuals (~10% of people over the age of 65), the overall rate of disease progression remains relatively low. Thus, in addition to genetic mutations, there are likely environmental factors that contribute to clonal evolution in people with CHIP. One environmental stress that increases with age is inflammation. Although chronic inflammation is detrimental to the long-term function of normal hematopoietic stem cells, several recent studies in animal models have indicated hematopoietic stem cells with CHIP mutations may be resistant to these deleterious effects. However, direct evidence indicating a correlation between increased inflammation and accelerated CHIP in humans is currently lacking. In this study, we sequenced the peripheral blood cells of a cohort of patients with ulcerative colitis, an autoimmune disease characterized by increased levels of pro-inflammatory cytokines. This analysis revealed that the inflammatory environment of ulcerative colitis promoted CHIP with a distinct mutational spectrum, notably positive selection of clones with DNMT3A and PPM1D mutations. We also show a specific association between elevated levels of serum interferon gamma and DNMT3A mutations. These data add to our understanding of how cell extrinsic factors select for clones with specific mutations to promote clonal hematopoiesis.

Clonal hematopoiesis of indeterminate potential (CHIP) is a pre-malignant condition associated with expansion of blood cell populations driven by age-dependent acquisition of somatic mutations in hematopoietic stem cells (HSCs) [1–7]. Although CHIP mutations are typically present at a low variant allele fraction (VAF), they are cancer drivers and frequently involve genes mutated in myeloid malignancies [1–3,8]. Despite the overall prevalence of CHIP, few of these individuals progress to hematopoietic malignancy [1,2,5,7,8]. In addition to somatic mutations, environmental factors likely contribute to clonal evolution of CHIP. Accumulating evidence suggests inflammation might serve as a selective pressure that contributes to malignant progression of CHIP clones. Individuals with autoimmune diseases [9] or a history of chronic infection [10] have increased lifetime risk of myelodysplastic syndromes (MDS) and acute myeloid

Offprint requests to: Grant A. Challen, Washington University School of Medicine, 660 Euclid Avenue, St. Louis, MO 63110; E-mail: grantchallen@wustl.edu

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.exphem.2019.11.008.

0301-472X/
https://doi.org/10.1016/j.exphem.2019.11.008

leukemia (AML). TET2 is a negative regulator of the inflammatory response in myeloid cells [11], and *TET2*-mutant CHIP is associated with overproduction of promotes mutant HSC expansion [12] and accelerates atherosclerosis [13]. As chronic inflammation compromises HSC function, inflammation could shape hematopoiesis by selecting for HSCs with mutations that increase stress tolerance. However, direct evidence of this is lacking. A recent study suggested that patients with rheumatoid arthritis, a chronic inflammatory autoimmune disease, had no increased incidence of CHIP [14]. But as this patient cohort was relatively small, we sequenced 187 patients with ulcerative colitis (UC), an inflammatory bowel disease characterized by infiltration of T cells in the colon and overproduction of pro-inflammatory cytokines such as tumor necrosis factor α (TNFα) and interferon (IFNγ) to evaluate the role of inflammation in CHIP.

Methods

Sequencing and analysis

Peripheral blood mononuclear cell DNA from ulcerative colitis patients (>50 years old) without prior history of hematologic disease was obtained from the Digestive Diseases Research Core Center (DDRCC) under Washington University Institutional Review Board Protocol 20111078. Samples were sequenced using a panel targeting 40 genes and hotspots recurrently mutated in CHIP, MDS, and AML (Washington University MyeloSeq; Supplementary Table E1, online only, available at www.exphem.org). HaloplexHS (Agilent) amplicon sequencing provides coverage of each target with multiple, distinct amplicons that incorporate unique molecular indexes (UMIs) into sequencing reads for error correction. Libraries were generated using 500 ng DNA, then sequenced on the NovaSeq (Illumina) platform to obtain 16 million reads per sample (10,000 × raw sequence coverage). Sequence data were aligned to build 37 human references with BWA-MEM, creating "read families" from multiple reads with the same UMI sequence, which were used as input for variant identification using VarScan 2, Platypus, and Pindel. Variant filtering was performed on the union of identified variants from these callers using only read families with at least three supporting reads, a VAF >0.5%, at least two supporting HaloplexHS amplicons, more than eight error-corrected reads that support the alternative allele, and at least three supporting read families on each strand. Variants with a population allele frequency >0.1% in the ExAC database or that possessed a VAF >35% were excluded as likely germline variants. The frequency of mutated genes in ulcerative colitis patients was compared with the aggregate incidence across indicated control cohorts.

Serum cytokine analysis

Serum TNFα and IFNγ levels were examined using a cytokine-specific ProQuantum Immunoassay kit (Invitrogen, USA). Patients groups were matched for age and sex.

Statistics

Comparisons of CHIP prevalence in this study with that in published data sets were performed using a χ^2 test for each age range and gene independently, followed by Bonferroni correction for multiple hypothesis testing. One-way analysis of variance (ANOVA) with Bonferroni multiple test correction was used for analysis of serum cytokine levels. Generalized linear models for multivariate analysis were established using R (version 3.5.3).

Table 1. Demographic information of UC patient cohort

Parameter	CH− *n* = 145	CH+ *n* = 42	Patients without information available No. (%)
Males (% of patients)	72.7	27.3	0 (0)
Females (% of patients)	83.0	17.0	0 (0)
Age (y)	63.1 ± 0.68 (50–84)	66.6 ± 1.18 (55–81)	0 (0)
Time from diagnosis to blood draw (y)	11.9 ± 1.1 (0.087–48.4)	12.8 ± 1.67 (0.70–39.1)	47 (25.1)
Colectomy (% of patients)	55.9	52.0	69 (36.9)
Flare at time of blood collection (% of patients)	64.7	61.3	59 (31.6)
Blood counts			
WBC (× 1000/mL)	9.6 ± 0.47 (3.7–27.3)	9.1 ± 1.1 (3.7–34.5)	60 (32.1)
Neutrophils (× 1000/mL)	7.7 ± 0.49 (1.4–25.3)	7.3 ± 1.1 (2.2–32.1)	66 (35.3)
Lymphocytes (× 1000/mL)	1.2 ± 0.070 (0.10–3.4)	1.3 ± 0.12 (0.30–3.0)	66 (35.3)
Hemoglobin B (g/dL)	11.6 ± 0.25 (6.6–16.0)	11.6 ± 0.38 (8.2–15.8)	61 (32.6)
Hematocrit (%)	34.5 ± 0.69 (19.0–47.1)	36.8 ± 2.2 (23.0–44.6)	61 (32.6)
MCV (fL)	90.6 ± 0.78 (70.8–107.7)	89.8 ± 2.7 (33.4–113.4)	62 (33.2)
Platelets (× 1000/mL)	301.1 ± 12.9 (97.0–706.0)	308.1 ± 24.2 (98.0–580.0)	61 (32.6)
Treatments			
Current anti-TNF (% of patients)	27.9	23.1	75 (40.1)
Previous anti-TNF (% of patients)	24.4	31.2	77 (41.2)
Current immuno-modulator (% of patients)	27.9	23.3	76 (40.6)
Previous immuno-modulator (% of patients)	29.4	23.7	77 (41.2)
Current steroids (% of patients)	30.3	19.6	75 (40.1)
Previous steroids (% of patients)	29.0	23.8	76 (40.6)

Values are expressed as mean ± SEM (range) unless otherwise indicated.

Results and discussion

We sequenced 40 CHIP-associated genes in 187 UC patients (Table 1) to identify genetic variants in their blood compartment. As germline DNA was unavailable, variants with VAFs >35% that were identified in the ExAC database were excluded as likely germline variants. To compare these data with the incidence of CHIP in the general population, we used the following definition of CHIP (CHIP-classical or CHIP-c) as outlined by a prior study with similar limitations [15]. CHIP-c is defined here as follows:

- VAF >2%
- Variant reported in COSMIC in "hematopoietic and lymphoid" category OR
  - Damaging variants (frameshift and stop_gain mutations) in *DNMT3A* gene within exons 7–23
  - Any damaging variants in the genes *ASXL1*, *TET2*, *PPM1D*, and *TP53*
  - CALR exon 9 indels.

Because of the sequencing depth and error correction, the limit of detection of 0.5% VAF facilitated a more complete view of clonal hematopoiesis in these patients. Such variants were classified as CHIP-high sensitivity (CHIP-hs) mutations by satisfying the following criteria; VAF >0.5% and variant reported in COSMIC in "hematopoietic and lymphoid" category >2 times.

CHIP-c was identified in 12.8% of patients in this cohort, with 22.5% of patients being positive for CHIP-hs. Twenty-six CHIP-c variants were identified in 24 patients (Figure 1A), with an additional 31 CHIP-hs variants harbored in 18 patients (Figure 1B; Supplementary Table E2, online only, available at www.exphem.org). C>T transitions were the most prevalent single-nucleotide variants, consistent with an age-dependent mutational signature (Figure 1A, B). The most recurrently mutated gene in UC patients with CHIP-c (Figure 1C) and CHIP-hs (Figure 1D) was *DNMT3A*, followed by *PPM1D*. As a comparable control cohort was not available for this study, to gain a sense of the prevalence of CHIP in UC patients compared with the general population, we compared our data with epidemiological sequencing cohorts using the quantitative definition of CHIP outlined above across all data sets. The frequency of CHIP-c in UC patients trended higher after the sixth decade, although sample size restrictions in older individuals in the UC cohort limit this comparison (Figure 1E). There was an age-dependent increase in CHIP-hs in UC patients without sex bias (Figure 1F). The frequency of *DNMT3A*$^{+}$ and *PPM1D*$^{+}$ CHIP was higher in UC patients (Figure 1G) compared with the aggregate incidence across the other studies [1–7]. *TET2* is typically the second most mutated gene (after *DNMT3A*) in CHIP sequencing studies of healthy individuals [1,2,7,8], but was less frequent in UC patients (Figure 1C, D). This was intriguing because certain inflammatory cytokines (e.g., IL-6) enhance the competitive fitness of *Tet2*-mutant HSCs [12]. It is possible the inflammatory milieu of UC is not conducive for the clonal expansion of *TET2*-mutant clones. The relative incidence of *TET2*$^{+}$ CHIP was lower in UC because certain mutations were positively selected, notably *PPM1D* (Figure 1G). *PPM1D*-mutant clones undergo strong positive selections in response to stress such as chemotherapy. The stress of chronic inflammation in UC patients may provide another environment that promotes growth of *PPM1D*-mutant clones.

The incidence of CHIP-hs in UC patients was not related to colectomy, a disease flare at time of blood collection (diarrhea, blood in stool, abdominal pain), current therapy, or treatment history (Figure 2A). CHIP-hs+ patients in this cohort were associated with increased age (Figure 2B), but not time since UC diagnosis (Figure 2C). No difference in blood counts was noted between CHIP-hs + and CHIP− individuals (Figure 2D). The VAF of most variants was <5%; however, there were several cases where clones dominated the blood, particularly for *ASXL1* mutations and *DNMT3A*^{R882H} (Figure 2E). The distribution of CHIP-hs+ mutations in UC patients was not clustered with age or sex for any given gene (Figure 2F). The mutational spectrum of *DNMT3A* in UC patients reflected the typical pattern for this gene in CHIP, with mutations clustered in functional domains without enrichment for the R882 hotspot (Figure 2G). *PP1MD* mutations were concentrated in the C-terminus (Figure 2G), mirroring those in therapy-related myeloid neoplasms [16,17]. We sought to determine if inflammatory signals selected for specific clones in UC patients. UC patients were stratified into three groups matched for age and sex: CHIP− ($n = 21$), CHIP-hs + without *DNMT3A* mutations (CH+*DNMT3A*−, $n = 17$), and CHIP-hs+ with *DNMT3A* mutations (CH+*DNMT3A*+, $n = 19$). We focused on TNFα and IFNγ because these cytokines have described roles in the symptoms of UC, and treatment can involve targeting these molecules. Serum TNFα levels did not differ among the three patient groups (Figure 2H). However, CH+*DNMT3A*+ patients were associated with significantly higher levels of serum IFNγ (Figure 2H). Average clone size did not contribute to the observed variances (Figure 2H), although *DNMT3A* VAF was a significant contributor to increased IFNγ levels in multivariate analysis using a generalized linear model ($p = 0.026$). These data imply that increased IFNγ may select for clones with *DNMT3A* mutations. Focused mechanistic studies will be required to show this relationship is causative and not correlative. Cumulatively, these data indicate that UC patients may harbor slightly higher levels of CHIP than the general population and that the

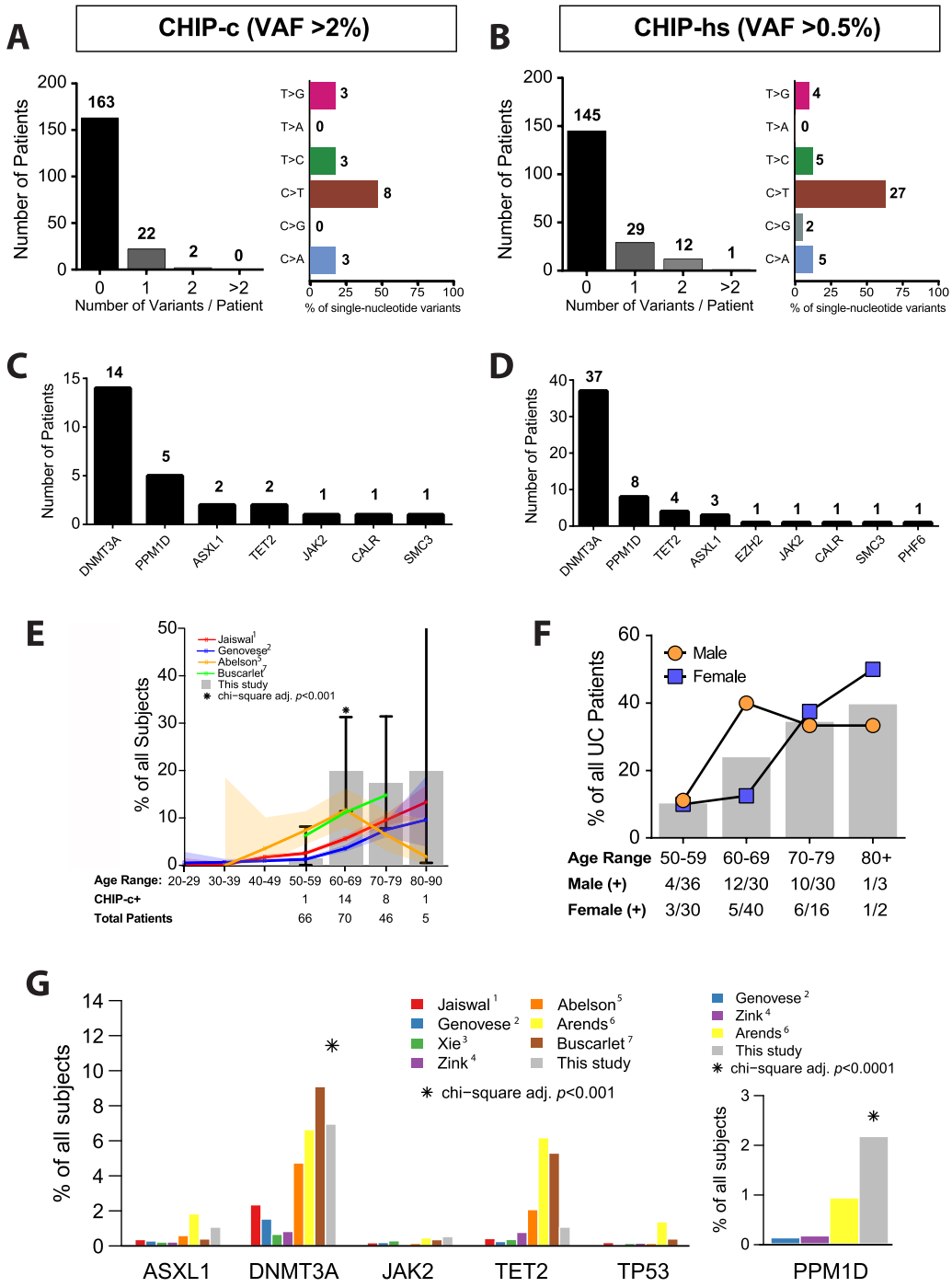

Figure 1. Clonal hematopoiesis in UC patients. **(A)** Number of CHIP-c variants identified in UC patients (*left*), and categories of those single-nucleotide variants (*right*). **(B)** Number of CHIP-hs variants identified in UC patients (*left*), and categories of those single-nucleotide variants (*right*). **(C)** Number of CHIP-c mutations identified per gene. **(D)** Number of CHIP-hs mutations identified per gene. **(E)** CHIP incidence by age in referenced studies. Error bars represent the 95% binomial confidence interval of the point estimate in UC patients. The conference interval bands for referenced studies represent the 95% confidence interval around the point estimate. **(F)** CHIP-hs incidence by sex in UC patients over age brackets. Gray bars denote the CHIP-hs incidence in given age bracket. **(G)** Incidence of CHIP-c mutations by gene across indicated studies. Significance is indicated for incidence of CHIP-c mutations in given gene in UC patients compared with aggregate incidence across other reference studies.

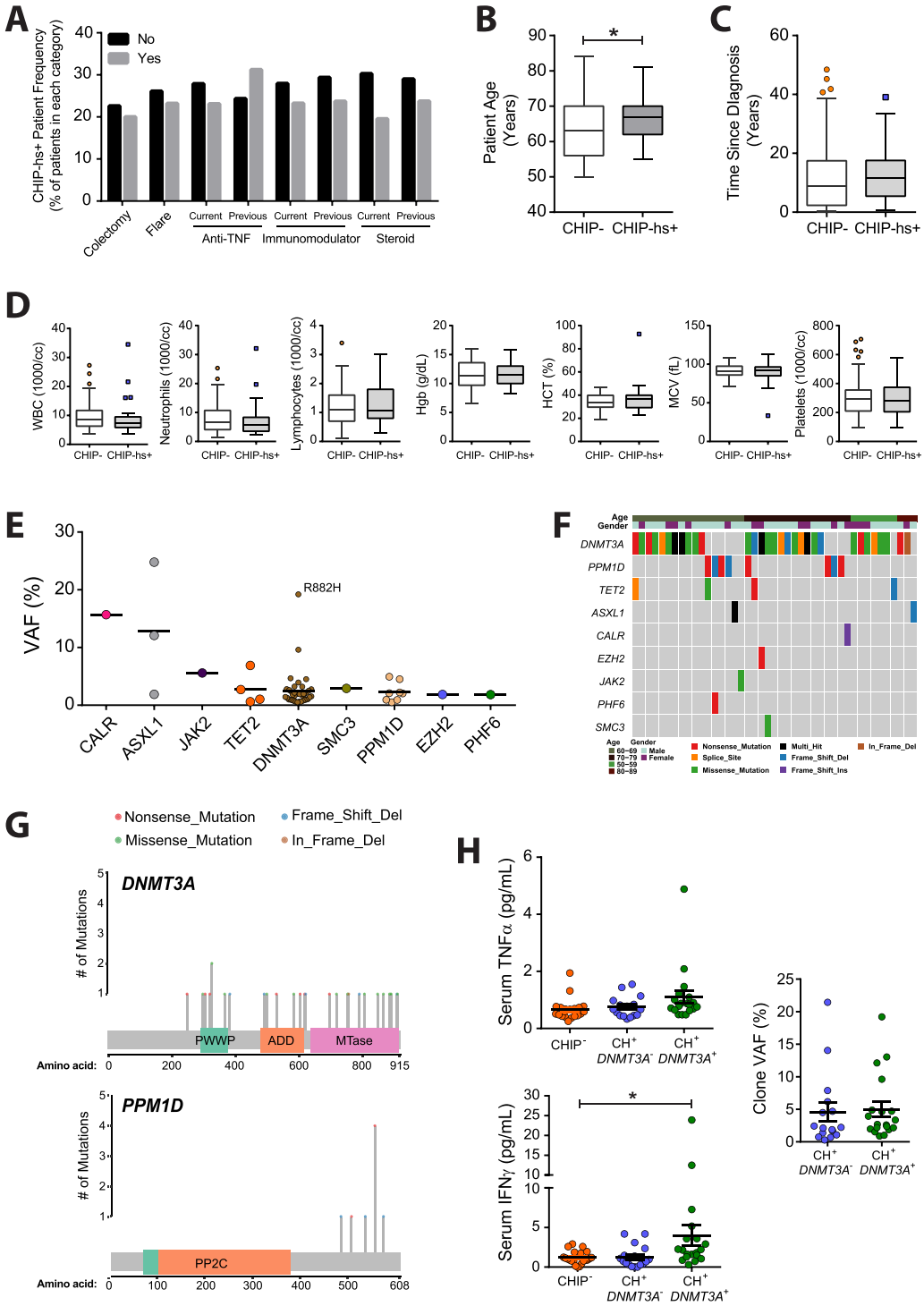

Figure 2. CHIP dynamics in UC patients. **(A)** Correlation of clonal hematopoiesis in UC patients with clinical variables and treatment history. **(B)** Tukey plot revealing ages of CH− and CH+ UC patients. *$p < 0.05$, two-tailed t test. **(C)** Tukey plot revealing time since diagnosis of CH− and CH+ UC patients. **(D)** Tukey plots of blood count parameters of CH− and CH+ UC patients. **(E)** VAFs of CH variants in UC patients. Each dot represents a single mutation; the line represents the mean VAF. **(F)** Mutational profile of CH variants in UC patients. **(G)** Mutational spectrum of somatic variants identified in *DNMT3A* and *PPM1D* in UC patients. **(H)** TNFα and IFNγ levels in serum of UC samples, and clone size (by VAF) in patients surveyed in this experiment. Each dot indicates the level of a given cytokine in the serum of one patient. Graphs represent means ± SEM. *$p < 0.05$, one-way ANOVA with Bonferroni multiple test correction.

inflammatory environment of UC potentially selects for the growth of HSC clones with specific mutations.

Acknowledgments

The Washington University Digestive Diseases Research Cores Center (DDRCC) BioBank Core is supported by National Institutes of Health (NIH) Grant P30 DK052574. This work was supported by the Washington University McDonnell Genome Institute and the NIH (Grant R01 DK102428 to GAC). ELO was supported by NIH Grant F31DK114951. Philanthropic support comes from the Lawrence C. Pakula IBD Research Innovations Fund (MAC). GAC is a scholar of the Leukemia and Lymphoma Society.

Conflict of interest disclosure

The authors declare no competing interests.

Author contributions

GAC was responsible for the project conceptualization and experimental design. Experiments were performed by CRCZ, ELO, and GAC. Critical reagents were provided by DN and RDN; CRCZ, MG, MAC, ELO, DHS, and GAC performed the data analysis. CRCZ prepared the original draft of the manuscript and GAC reviewed and edited it. GAC was responsible for project administration and funding acquisition.

References

1. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–2498.
2. Genovese G, Kahler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371:2477–2487.
3. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20:1472–1478.
4. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130:742–752.
5. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature. 2018;559:400–404.
6. Arends CM, Galan-Sousa J, Hoyer K, et al. Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia. 2018;32:1908–1919.
7. Buscarlet M, Provost S, Zada YF, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood. 2017;130:753–762.
8. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun. 2016;7:12484.
9. Anderson LA, Pfeiffer RM, Landgren O, Gadalla S, Berndt SI, Engels EA. Risks of myeloid malignancies in patients with autoimmune conditions. Br J Cancer. 2009;100:822–828.
10. Kristinsson SY, Bjorkholm M, Hultcrantz M, Derolf AR, Landgren O, Goldin LR. Chronic immune stimulation might act as a trigger for the development of acute myeloid leukemia or myelodysplastic syndromes. J Clin Oncol. 2011;29:2897–2903.
11. Cull AH, Snetsinger B, Buckstein R, Wells RA, Rauh MJ. Tet2 restrains inflammatory gene expression in macrophages. Exp Hematol. 2017;55:56–70, e13.
12. Cai Z, Kotzin JJ, Ramdas B, et al. Inhibition of inflammatory signaling in Tet2 mutant preleukemic cells mitigates stress-induced abnormalities and clonal hematopoiesis. Cell Stem Cell. 2018;23:833–849, e835.
13. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. N Engl J Med. 2017;377:111–121.
14. Savola P, Lundgren S, Keranen MAI, et al. Clonal hematopoiesis in patients with rheumatoid arthritis. Blood Cancer J. 2018;8:69.
15. Coombs CC, Zehir A, Devlin SM, et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. Cell Stem Cell. 2017;21:374–382, e374.
16. Kahn JD, Miller PG, Silver AJ, et al. PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. Blood. 2018;132:1095–1105.
17. Hsu JI, Dayaram T, Tovy A, et al. PPM1D mutations drive clonal hematopoiesis in response to cytotoxic chemotherapy. Cell Stem Cell. 2018;23:700–713, e706.

Supplementary Table E1. Amplicons covered by MyeloSeq (Washington University, St. Louis, MO) custom gene panel

Gene	Category	Target(s)	Frequency in AML*	Frequency in MDS**
BRAF	Activated signaling	V600E	0.0%	0.6%
FLT3	Activated signaling	TKD and ITD	28.0%	1.3%
JAK2	Activated signaling	V617, exon 12	0.5%	3.3%
KIT	Activated signaling	exons 2, 8-13, 17	4.0%	1.4%
KRAS	Activated signaling	G12, G13, Q61	4.0%	2.1%
MPL	Activated signaling	exon 10	0.0%	2.3%
NF1	Activated signaling	whole gene	1.5%	4.8%
NRAS	Activated signaling	G12, G13, Q61	8.0%	4.1%
PTPN11	Activated signaling	exons 3, 13, 14	4.0%	1.2%
ASXL1	Chromatin modifiers	whole gene	2.5%	14.4%
EZH2	Chromatin modifiers	whole gene	1.0%	7.2%
SUZ12	Chromatin modifiers	whole gene	1.4%	0.0%
CSF3R	Chromatin modifiers	whole gene	0.5%	
RAD21	Cohesin	whole gene	2.0%	1.2%
SMC1A	Cohesin	whole gene	0.5%	1.5%
SMC3	Cohesin	whole gene	1.0%	1.5%
STAG2	Cohesin	whole gene	3.0%	6.5%
DNMT3A	DNA methylation	whole gene	22.0%	11.6%
IDH1	DNA methylation	R132	10.0%	2.1%
IDH2	DNA methylation	R140, R172	10.0%	3.7%
TET2	DNA methylation	whole gene	9.0%	23.9%
CALR	Other genes	exon 9	1.0%	0.0%
CBL	Other genes	exons 8, 9	1.0%	4.2%
NPM1	Other genes	exon 11	27.0%	2.5%
PIGA	Other genes	whole gene	0.0%	0.6%
PPM1D	Other genes	exon 6	0.0%	0.1%
CUX1	Other genes	whole gene		
SF3B1	Spliceosome	whole gene	0.5%	21.5%
SRSF2	Spliceosome	exon 1	0.5%	11.0%
U2AF1	Spliceosome	exons 2, 6	4.0%	8.2%
ZRSR2	Spliceosome	whole gene	0.0%	5.0%
BCOR	Transcriptional regulator	whole gene	1.0%	5.2%
BCORL1	Transcriptional regulator	whole gene	0.5%	4.2%
CEBPA	Transcriptional regulator	whole gene	6.0%	0.7%
ETV6	Transcriptional regulator	whole gene	1.0%	2.5%
GATA2	Transcriptional regulator	whole gene	0.0%	1.6%
RUNX1	Transcriptional regulator	whole gene	12.0%	8.4%
PHF6	Tumor suppressors	whole gene	2.0%	2.6%
TP53	Tumor suppressors	whole gene	8.0%	10.1%
WT1	Tumor suppressors	whole gene	6.0%	1.2%

*Frequency obtained from TCGA.

**Frequency averaged from Haferlach et al. (Leukemia, 2014), Papaemmanuil et al. (Blood, 2013), and Walter et al. (Leukemia, 2013).

Supplementary Table E2. Annotation of variants identified from MyeloSeq analysis of 187 patients with UC

SAMPLE ID	Gender	Age	CHROM	POS	REF	ALT	VAF	SYMBOL	Consequence	IMPACT	BIOTYPE	CDS_ position	Protein_ position	Amino_ acids	Codons	Existing_variation
1012-1-112706	Female	57	2	25463541	G	C	1.53%	DNMT3A	missense_variant	MODERATE	protein_coding	2141	714	S/C	tCc/tGc	rs367909007& COSM442677& COSM5580894& COSM5580895& COSM87011
1079-2-022108	Male	66	4	106190906	T	G	6.89%	TET2	splice_donor_ variant&NMD_ transcript_variant	HIGH	nonsense_ mediated_decay					
			2	25467497	G	A	4.65%	DNMT3A	stop_gained	HIGH	protein_coding	1579	527	Q/*	Cag/Tag	COSM5095526& COSM5095527
109-101609-1	Female	70	19	13054627	A	ATTGTC	15.68%	CALR	frameshift_variant	HIGH	protein_coding	1154-1155	385	K/NCX	aag/aaTTGTCg	rs765476509
1103-1-071509	Male	68	17	58740749	C	T	0.50%	PPM1D	stop_gained	HIGH	protein_coding	1654	552	R/*	Cga/Tga	rs779070661& COSM982226
1163-1-062813	Male	51	4	106155939	TAACTCTG	T	0.62%	TET2	frameshift_variant& NMD_transcript_ variant	HIGH	nonsense_ mediated_decay	841-847	281-283	NSE/X	AACTCTGag/ag	
1382-1-070705	Male	80	2	25467073	C	T	0.99%	DNMT3A	stop_gained	HIGH	protein_coding	1802	601	W/*	tGg/tAg	rs941325374
1422-1-042408	Female	67	2	25457252	T	C	3.20%	DNMT3A	missense_variant	MODERATE	protein_coding	2635	879	N/D	Aac/Gac	COSM1583135& COSM5878743
1637-1-100407	Male	78	2	25470498	G	A	1.94%	DNMT3A	missense_variant	MODERATE	protein_coding	976	326	R/C	Cgc/Tgc	rs747448117& COSM4169721& COSM4383600
1645-2-110607	Male	75	2	25468121	C	T	1.72%	DNMT3A	splice_donor_variant	HIGH	protein_coding					rs766110518& COSM4775128& COSM4775129
177-2-092806	Male	68	2	25471016	G	A	0.97%	DNMT3A	stop_gained	HIGH	protein_coding	745	249	Q/*	Cag/Tag	rs759747476
1785-2-050806	Male	62	9	5073770	G	T	5.60%	JAK2	missense_variant	MODERATE	protein_coding	1849	617	V/F	Gtc/Ttc	rs77375493& CM123094& COSM12600& COSM29117
1792-1-051611	Male	73	17	58740603	C	A	4.94%	PPM1D	stop_gained	HIGH	protein_coding	1508	503	S/*	tCa/tAa	rs375618423
1811-1-062807	Female	70	2	25470560	C	T	3.23%	DNMT3A	stop_gained	HIGH	protein_coding	914	305	W/*	tGg/tAg	rs765341003& COSM1169636
			7	148523591	G	A	1.87%	EZH2	stop_gained	HIGH	protein_coding	862	288	R/*	Cga/Tga	COSM1000721& COSM4384289
			2	25470581	C	T	1.60%	DNMT3A	missense_variant	MODERATE	protein_coding	893	298	G/E	gGg/gAg	COSM5878868& COSM5878869
			2	25466766	C	T	1.06%	DNMT3A	splice_donor_variant	HIGH	protein_coding					
1846-1-121707	Male	64	2	25457242	C	T	19.20%	DNMT3A	missense_variant	MODERATE	protein_coding	2645	882	R/H	cGc/cAc	rs147001633& COSM1583129& COSM3356083& COSM442676& COSM52944& COSM99740
2028-1-101513	Male	70	2	25466849	GT	G	2.27%	DNMT3A	frameshift_variant& splice_region_ variant	HIGH	protein_coding	1853	618	D/X	gAc/gc	
2207-2-082307	Male	68	2	25469918	A	C	2.17%	DNMT3A	splice_donor_variant	HIGH	protein_coding					COSM4766077
			2	25469919	C	T	1.81%	DNMT3A	splice_donor_variant	HIGH	protein_coding					rs747220514& COSM5878851& COSM5878852
2225-3-083106	Female	67	2	25463241	A	C	2.80%	DNMT3A	missense_variant	MODERATE	protein_coding	2252	751	F/C	tTc/tGc	rs765813304
2716-1-032008	Male	71	2	25458595	A	G	1.38%	DNMT3A	missense_variant	MODERATE	protein_coding	2578	860	W/R	Tgg/Cgg	rs373014701& COSM231568& COSM4383524
2862-1-092310	Male	74	2	25458648	T	C	4.50%	DNMT3A	missense_variant	MODERATE	protein_coding	2525	842	Q/R	cAg/cGg	rs771174392
			17	58740749	C	T	0.95%	PPM1D	stop_gained	HIGH	protein_coding	1654	552	R/*	Cga/Tga	rs779070661& COSM982226
3096-4-091406	Female	71	17	58740689	AC	A	4.50%	PPM1D	frameshift_variant	HIGH	protein_coding	1595	532	T/X	aCa/aa	
3179-1-032107	Female	70	2	25469646	C	T	0.61%	DNMT3A	splice_acceptor_ variant	HIGH	protein_coding					

(*continued on next page*)

Supplementary Table E2 (*Continued*)

SAMPLE ID	Gender	Age	CHROM	POS	REF	ALT	VAF	SYMBOL	Consequence	IMPACT	BIOTYPE	CDS_ position	Protein_ position	Amino_ acids	Codons	Existing_variation
3205-1-092208	Male	68	20	31022786	GCCATGCC-AGGCCTT	G	24.79%	ASXL1	frameshift_variant	HIGH	protein_coding	2257-2270	753-757	PCQAL/X	CCATGCCAG-GCCTTg/g	
			20	31024021	C	CT	1.89%	ASXL1	frameshift_variant	HIGH	protein_coding	3491-3492	1164	S/SX	tct/tcTt	
3240-2-102606	Female	55	2	25470516	G	A	2.61%	DNMT3A	stop_gained	HIGH	protein_coding	958	320	R/*	Cga/Tga	rs778270132& COSM1318922& COSM133721
3352-1-100107	Male	62	4	106164764	G	A	2.70%	TET2	missense_variant& NMD_transcript_ variant	MODERATE	nonsense_ mediated_decay	3541	1181	V/I	Gtt/Att	COSM87124
			17	58740749	C	T	2.04%	PPM1D	stop_gained	HIGH	protein_coding	1654	552	R/*	Cga/Tga	rs779070661& COSM982226
3602-1-012606	Female	53	2	25469946	G	T	1.86%	DNMT3A	missense_variant	MODERATE	protein_coding	1096	366	R/S	Cgc/Agc	COSM5944976
3876-1-102212	Male	70	10	112361545	T	C	2.90%	SMC3	missense_variant	MODERATE	protein_coding	2795	932	L/P	cTa/cCa	rs758156711& COSM5487671
			2	25457176	G	A	0.97%	DNMT3A	missense_variant	MODERATE	protein_coding	2711	904	P/L	cCg/cTg	rs149095705& COSM1741211& COSM87007
3880-2-031308	Male	75	17	58740749	C	T	0.92%	PPM1D	stop_gained	HIGH	protein_coding	1654	552	R/*	Cga/Tga	rs779070661& COSM982226
3925-1-072506	Female	84	2	25463235	CAGA	C	1.28%	DNMT3A	inframe_deletion	MODERATE	protein_coding	2255-2257	752-753	FW/W	tTCTgg/tgg	rs749132507& COSM133723& COSM5351731
3943-2-072706	Female	70	4	106157827	C	T	1.05%	TET2	stop_gained&NMD_ transcript_variant	HIGH	nonsense_ mediated_decay	2728	910	Q/*	Caa/Taa	COSM4383868& COSM4383869
			2	25468898	TC	T	0.87%	DNMT3A	frameshift_variant	HIGH	protein_coding	1464	488	R/X	cgG/cg	
401-2-100807	Female	69	2	25469624	TC	T	2.77%	DNMT3A	frameshift_variant	HIGH	protein_coding	1143	381	G/X	ggG/gg	rs745939351
			2	25468186	C	T	0.88%	DNMT3A	missense_variant	MODERATE	protein_coding	1490	497	C/Y	tGt/tAt	rs779323387& COSM1318925& COSM1318926
404-012207-2	Male	81	20	31021176	CAG	C	12.09%	ASXL1	frameshift_variant	HIGH	protein_coding	1161-1162	387-388	SV/SX	tcAGtg/tctg	
4564-1-032014	Female	74	2	25467032	G	A	2.16%	DNMT3A	stop_gained	HIGH	protein_coding	1843	615	Q/*	Cag/Tag	rs768548187& COSM1318929& COSM1318930
			2	25459804	C	T	0.75%	DNMT3A	splice_donor_variant	HIGH	protein_coding				rs762213449& COSM4766078	
482-1-092210	Male	64	2	25457181	GA	G	2.51%	DNMT3A	frameshift_variant	HIGH	protein_coding	2705	902	F/X	tTc/tc	
			2	25463248	G	A	1.91%	DNMT3A	missense_variant	MODERATE	protein_coding	2245	749	R/C	Cgc/Tgc	rs754613602& COSM219133& COSM3580252
4907-4527	Male	55	2	25463602	G	C	0.74%	DNMT3A	splice_region_ variant&intron_ variant	LOW	protein_coding					
528-1-061413	Female	60	2	25457236	G	T	3.94%	DNMT3A	missense_variant	MODERATE	protein_coding	2651	884	A/E	gCg/gAg	COSM231547
559-2-061306	Male	71	2	25470498	G	A	0.61%	DNMT3A	missense_variant	MODERATE	protein_coding	976	326	R/C	Cgc/Tgc	rs747448117& COSM4169721& COSM4383600
588-1-101906	Female	60	17	58740532	TA	T	2.21%	PPM1D	frameshift_variant	HIGH	protein_coding	1438	480	K/X	Aaa/aa	rs749097315
677-1-011714	Male	69	17	58740800	ACCTCACAG	A	2.10%	PPM1D	frameshift_variant	HIGH	protein_coding	1706-1713	569-571	TSQ/X	aCCTCACAG/a	
			X	133551267	C	A	1.84%	PHF6	stop_gained	HIGH	protein_coding	903	301	Y/*	taC/taA	COSM144570& COSM3558102& COSM3558103& COSM5945628
706-1-080306	Male	66	2	25463247	C	T	9.62%	DNMT3A	missense_variant	MODERATE	protein_coding	2246	749	R/H	cGc/cAc	rs34843713& COSM221577
747-1-013112	Male	56	2	25467134	A	C	1.97%	DNMT3A	missense_variant	MODERATE	protein_coding	1741	581	W/G	Tgg/Ggg	COSM1583082& COSM231554& COSM231558
858-2-070507	Male	62	2	25464439	G	A	1.44%	DNMT3A	stop_gained	HIGH	protein_coding	2074	692	Q/*	Cag/Tag	
916-1-052406	Male	70	2	25462054	CT	C	0.61%	DNMT3A	frameshift_variant	HIGH	protein_coding	2352	784	E/X	gaA/ga	
			2	25462006	T	C	0.51%	DNMT3A	missense_variant	MODERATE	protein_coding	2401	801	M/V	Atg/Gtg	rs753567076& COSM5944905