Haematologica 2019
Volume 104(12):2410-2417

Clonal hematopoiesis and risk of acute myeloid leukemia

Andrew L. Young,[1,2] R. Spencer Tong,[1,2] Brenda M. Birmann[3*] and Todd E. Druley[1,2*]

[1]Department of Pediatrics, Division of Hematology and Oncology, Washington University School of Medicine, Saint Louis, MO; [2]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, Saint Louis, MO and [3]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

**BMB and TED contributed equally to this work.*

Correspondence:

TODD E. DRULEY
druley_t@wustl.edu

Received: January 18, 2019.

Accepted: April 16, 2019.

Pre-published: April 19, 2019.

doi:10.3324/haematol.2018.215269

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/104/12/2410

©2019 Ferrata Storti Foundation

Material published in Haematologica is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions: https://creativecommons.org/licenses/by-nc/4.0/legalcode. Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions: https://creativecommons.org/licenses/by-nc/4.0/legalcode, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.

ABSTRACT

Nearly all adults harbor acute myeloid leukemia (AML)-related clonal hematopoietic mutations at a variant allele fraction (VAF) of ≥0.0001, yet relatively few develop hematologic malignancies. We conducted a nested analysis in the Nurses' Health Study and Health Professionals Follow-Up Study blood subcohorts, with up to 22 years of follow up to investigate associations of clonal mutations of ≥0.0001 allele frequency with future risk of AML. We identified 35 cases with AML that had pre-diagnosis peripheral blood samples and matched two controls without history of cancer per case by sex, age, and ethnicity. We conducted blinded error-corrected sequencing on all study samples and assessed variant-associated risk using conditional logistic regression. We detected AML-associated mutations in 97% of all participants (598 mutations, 5.8/person). Individuals with mutations ≥0.01 variant allele fraction had a significantly increased AML risk (OR 5.4, 95%CI: 1.8-16.6), as did individuals with higher-frequency clones and those with *DNMT3A* R882H/C mutations. The risk of lower-frequency clones was less clear. In the 11 case-control sets with samples banked ten years apart, clonal mutations rarely expanded over time. Our findings are consistent with published evidence that detection of clonal mutations ≥0.01 VAF identifies individuals at increased risk for AML. Further study of larger populations, mutations co-occurring within the same pre-leukemic clone and other risk factors (lifestyle, epigenetics, etc.), are still needed to fully elucidate the risk conferred by low-frequency clonal hematopoiesis in asymptomatic adults.

Introduction

Clonal hematopoiesis of indeterminate potential (CHIP) has been defined as somatic mutations in the peripheral blood at variant allele fractions (VAF) >0.02 in individuals without evidence of hematologic malignancy.[1] The threshold of 0.02 VAF was arbitrarily derived, reflecting the technical limitations of the standard next generation sequencing (NGS) and not biological risk of leukemic transformation with lower frequency mutations. To date, there have been no systematic screening recommendations for identifying or surveilling CHIP in healthy individuals. Nonetheless, the presence of CHIP has been shown to increase the risk of developing hematologic malignancy (in aggregate) by 0.5-1% per year,[2,3] although the absolute risk of leukemic transformation in individuals with CHIP is very low. Recently, two studies demonstrated an increased risk of developing AML in individuals with CHIP detected using targeted sequencing of peripheral blood samples collected several years prior to diagnosis.[4,5]

Independently, error-corrected sequencing (ECS) has enabled accurate interrogation of the hematologic somatic mutational profile at VAF ≥0.0001[6] and demonstrated that selection of pre-existing clones can lead to therapy-related leukemia.[7] Our ECS-based study of blood samples collected approximately ten years apart

from 20 adult women without AML revealed that nearly all studied individuals harbored somatic mutations frequently observed in myeloid malignancies. The detected hematopoietic clones were often stable over the ten years between blood collections and did not demonstrate positive selection or clonal expansion, regardless of the gene mutated. It is important to note that clonal mutations at a lower frequency than 0.02 VAF are currently not regarded as CHIP, and their clinical significance is even less well understood. The two aforementioned studies of pre-diagnosis CHIP observed associations of increased AML risk for persons with clones ≥0.005 or ≥0.01 VAF over a shorter follow-up period.[4,5] The present investigation examined whether detection of lower-VAF clones or specific mutations are associated with future risk of AML in a nested case-control sample (35 cases, 70 controls) from the Nurses' Health Study (NHS) and Health Professionals Follow-up Study (HPFS) cohorts with up to 22 years of follow up after sample collection.[8,9] We also investigated whether clonal evolution over ten years was associated with long-term future risk of AML in 11 women in the NHS with multiple pre-diagnosis samples.

Methods

Study population

Details of the NHS and HPFS design and data collection and follow-up methods are published elsewhere[8,9] (see also *Online Supplementary Methods*). Biennial questionnaire return rates have been consistently high (>95% in the blood subcohorts described below).

Blood subcohorts

The "blood subcohorts" comprise 32,826 women (NHS) who provided a heparinized whole blood sample from 1989-1990,[10] of whom 18,743 provided a second whole blood sample from 2000-2001,[11] as well as 18,018 men (HPFS) who provided an EDTA whole blood sample from 1993-1995. Participants provided written informed consent. The present study protocol was approved by the Institutional Review Boards of Brigham and Women's Hospital, Harvard TH Chan School of Public Health and Washington University.

Case and control selection

The present study utilized a nested case-control design for which the case definition included all blood subcohort participants with confirmed diagnoses of AML (ICD-8=205.0) occurring after blood draw. We matched two controls per case on cohort (sex), race, birthdate (±1 year), and blood draw details (date ±1 year, time ±4 hours, fasting status). For NHS cases with a second collection sample, we matched controls with a second sample using the same criteria. These protocols selected 35 cases (16 NHS, 19 HPFS) and 70 controls (32 NHS, 38 HPFS), including 11 matched sets (NHS) with two samples (n=137 total samples after excluding four with insufficient volume).

Clonal hematopoiesis of indeterminate potential determination and validation

Sequencing libraries were prepared as previously described[6] using the Illumina TruSight Myeloid Sequencing Panel for targeted capture from 54 leukemia-associated genes (*Online Supplementary Methods* and *Online Supplementary Table S1*). Libraries were sequenced on the Illumina HiSeq 3000 platform per manufacturer specifications; with technical replicate libraries sequenced on different machine runs. ECS analysis of raw sequencing results was performed as previously described,[6] except that, to improve rare SNV identification at potential "hot spot" loci, we re-called variants from the binomial error model after removing the variants already identified until a subsequent iteration revealed no additional new variants. We reported single nucleotide variants (SNV) and insertions and deletions (indels) identified in both technical replicates for a given sample. To validate our ECS-based variant calls, we performed droplet digital polymerase chain reaction (ddPCR) for 61 variants.

Statistical analysis

We combined NHS and HPFS data to maximize statistical power. We analyzed mutations detected in both technical replicates for ≥4 participants and selected VAF thresholds (≥0.001, ≥0.005, ≥0.01, ≥0.02), in the first collection samples and in samples from either collection. We used conditional logistic regression, conditioning on matched sets, to calculate odds ratios (OR) and 95% confidence intervals (CI) for the relative risk of AML for a given detected variant or VAF threshold. Sparse data precluded evaluation of confounding by other AML risk factors[3,12] or effect modification. Exploratory and sensitivity analyses are detailed in the *Online Supplementary Methods*. We utilized SAS version 9.3 for statistical analyses and the ggplot2 and ppcor packages of R version 3.3.3[13] for graphical descriptive analyses. Hypothesis tests assumed a two-tailed α-error of 0.05.

Results

Study samples

Due to the matched design, the cases and controls had similar distributions of sex, age at blood collection, and interval from blood draw to case diagnosis or control index date (Table 1). The median age of sample collection was 61 years for the first collection and 70 years for the second collection. The median age of AML diagnosis was 76 years (range: 53-87 years). More than 90% of cases and controls had one year or more of follow up after blood draw, and >88% of each group had follow-up intervals of five or more years. All the women with repeat blood samples had at least one year of follow up after the second blood collection (Table 1). All the participants selected into the study sample had self-reported their race/ethnicity as White.

Error-corrected sequencing results

During ECS library preparation, we generated an average of 60 million raw sequenced reads, yielding 3.9 million ECS reads, per library, which translated into approximately 8,000x ECS read coverage of the target space. We identified 563 single nucleotide variants and 35 insertion/deletion (indel) variants by ECS; this corresponded to detection of AML-associated mutations in 97% of all participants (598 mutations, 5.8/person), with an average of 7.4 (range: 1-14) per case and an average of 5.0 (range: 0-15) per control (*Online Supplementary Table S2*). As expected, due to the targeted enrichment sequencing scheme, these mutations predominantly occurred in exonic regions (*Online Supplementary Figure S1A*). Most detected mutations were predicted to change the underlying amino acid sequence in cases and controls (*Online Supplementary Figure S1B*). Of the 252 clonal mutations detected in the cases, we identified 144 non-synonymous SNV (57%), 40 stop gain variants (16%), 22 intronic variants (9%), 18 indels

Table 1. Selected characteristics of study participants by case-control status.

	Cases (N=34)	Controls (N=69)
Sex (cohort), N (%)		
Female (NHS)	15 (44.1)	31 (44.9)
Male (HPFS)	19 (55.9)	38 (55.1)
Age (years) at blood draw, median (range)		
Collection 1	61 (48-70)	61 (48-71)
Collection 2[a]	70 (62-78)	70 (62-79)
Age (years) at case diagnosis date, median (range)[b]	76 (53-87)	75 (53-87)
Years, blood collection to case diagnosis		
Collection 1,		
Mean (±SD)	14.2 (±6.1)	14.1 (±6.0)
N (%) by interval[c]		
<1	2 (5.9)	6 (8.7)
1 to <5	2 (5.9)	2 (2.9)
5 to <10	2 (5.9)	4 (5.8)
10 to <15	9 (26.5)	18 (26.1)
15 to <20	15 (44.1)	31 (44.9)
≥20	4 (11.8)	8 (11.6)
Collection 2,[a]		
Mean (±SD)	6.4 (±2.6)[a]	6.3 (±2.8)[a]
N (%) by interval[c]		
<1	0 (0.0)	0 (0.0)
1 to <5	2 (16.7)	6 (26.1)
5 to <10	8 (66.7)	13 (56.5)
≥10	2 (16.7)	4 (17.4)

N: number; AML: acute myeloid leukemia; HPFS: Health Professionals Follow-up Study; NHS: Nurses' Health Study; SD: standard deviation. [a]Second blood samples were available for 11 cases and 21 controls in the NHS. [b]In controls, defined as date of AML diagnosis for the matched case. [c]Case percentages do not sum to 100 due to rounding.

(7%), 12 synonymous SNV (5%), 11 splice variants (4%), 4 UTR variants (2%), and one stop loss variant (<1%). Of the 346 clonal mutations detected in the matched controls, we identified 208 non-synonymous SNV (60%), 47 intronic variants (14%), 34 synonymous SNV (10%), 30 stop gain variants (9%), 13 indels (4%), 12 splice variants (3%), and 2 UTR variants (<1%). As expected, C to T (G to A) substitutions were by far the most common in both cases and controls (*Online Supplementary Figure S1C*).

Droplet digital polymerase chain reaction validation

Spearman correlation coefficients reflect a high correspondence between ECS and ddPCR variant calls at both blood collections (collection 1, r=0.97, *P*<0.0001; collection 2, r=0.95, *P*<0.0001) (*Online Supplementary Table S3* and *Online Supplementary Figure S2*).

Gene-specific mutations

As expected, the most frequently observed mutations occurred in the epigenetic regulators *DNMT3A* and *TET2*, although we observed mutations in most of the genes targeted by the assay (Figure 1 and *Online Supplementary Figure S3*). In cases, we observed 58 *DNMT3A* and 56 *TET2* clonal variants, comprising 23% and 22% of the 252 clonal variants detected in cases, respectively. In controls, we observed 128 *DNMT3A* and 57 *TET2* clonal mutations, comprising 37% and 16% of the 346 clonal mutations detected in controls, respectively. Most mutations occurred in exonic regions and were predominantly non-synonymous and nonsense mutations (*Online Supplementary Figure S3*). The observed exonic variants in *DNMT3A* occurred predominantly in the functional domains (*Online Supplementary Figure 4A*). The observed exonic variants in *TET2* occurred across the entire amino acid sequence (*Online Supplementary Figure 4B*). No single mutation in *TET2* was observed in more than two individuals.

Association of individual variants and clonal hematopoiesis with acute myeloid leukemia risk

For the mutations that occurred in at least four individuals, associations with AML were similar in magnitude whether we classified mutation status according to the first or according to either blood collection (Table 2). Thus, for brevity, we focus herein on the findings based on either blood collection. Detecting the *DNMT3A* R822H variant at either time point was associated with a 14-fold increased risk of AML (OR: 14.0, 95%CI: 1.7-113.8; *P*=0.01). Participants with either a *DNMT3A* R882H or R882C variant ("R882H/C") at either collection had a more than 7-fold increased risk of AML relative to individuals without either variant (OR: 7.3, 95%CI: 1.5-34.7; *P*=0.01). For the *DNMT3A* W860R and *ASXL1* E1183K variants, the sparse counts prevented estimation of 95%CI by the conditional logistic regression models (implying a 95%CI range from zero to infinity). The *JAK2* V617F vari-

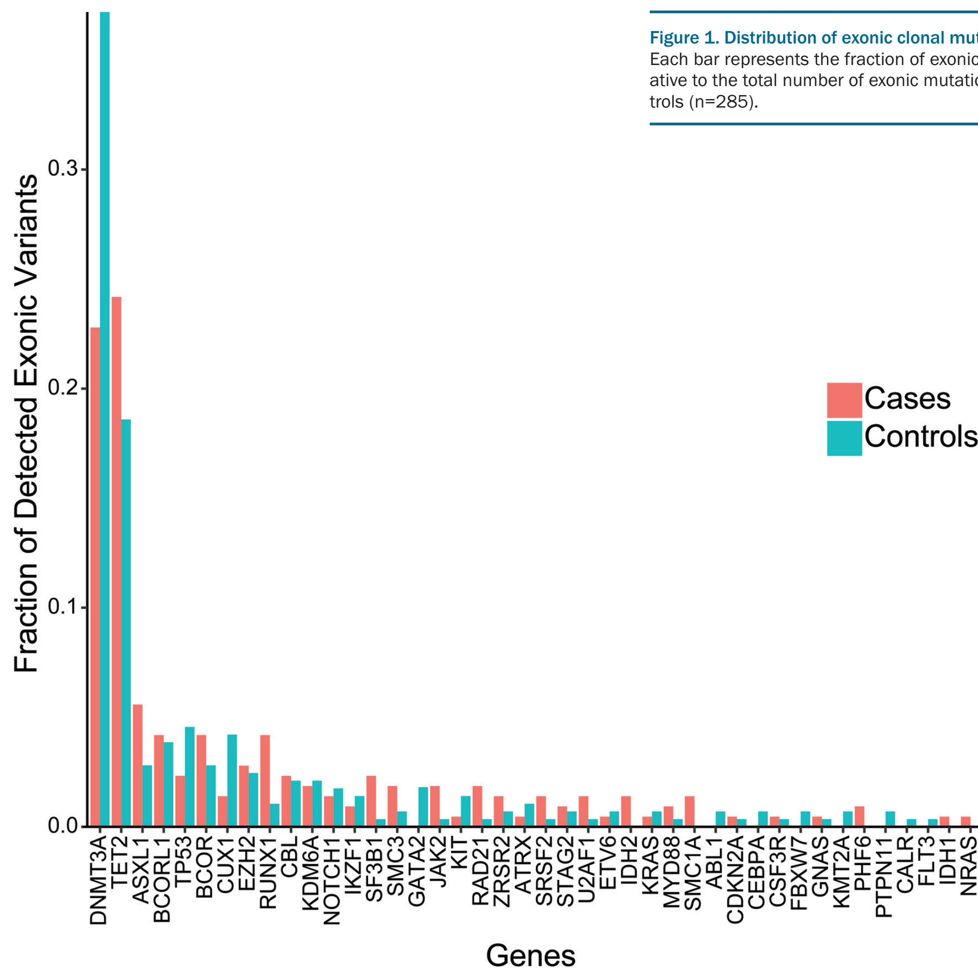

Figure 1. Distribution of exonic clonal mutations by gene in cases and controls. Each bar represents the fraction of exonic mutations detected in each gene relative to the total number of exonic mutations detected in cases (n=215) or controls (n=285).

ant, which we observed only in men in the present study sample, had a non-significant positive association with future development of AML (Table 2).

Individuals with clonal mutations detected at ≥0.01 (OR: 5.4, 95%CI: 1.8-16.6; *P*=0.003) or ≥0.02 VAF (OR: 5.6, 95%CI: 1.8-17.2; *P*=0.003) had a significantly increased risk of AML compared to those without a mutation detected at or above those thresholds (Table 2). The association with AML risk for VAF lower than 0.01 was unclear; for example, individuals with mutations at a VAF of ≥0.005 at either blood collection had a 2.5-fold increase in AML risk that was not statistically significant (OR: 2.5, 95%CI: 1.0-6.3; *P*=0.05). Further, nearly every case and most controls had at least one clonal mutation at ≥0.001 VAF. Of interest, the *ASXL1* E1183K variant noted above, which we observed in five women (4 cases, 1 control) and for which an association with AML risk could not be well quantified due to sparse counts, occurred at VAF between 0.001 and 0.002.

Sensitivity analyses that omitted records for the cases and controls with less than one year of follow up after blood collection did not materially change the main findings. One omitted case and two omitted controls were positive for *DNMT3A* R882H/C, whereas the remaining omitted case and four omitted controls were negative for that variant. Even after omitting these participants, detecting the *DNMT3A* R882H/C variant at one or both blood collections remained a statistically significant risk factor for AML (OR: 14.0, 95%CI: 1.7-113.8; *P*=0.01), as did detecting any mutation with a VAF ≥0.01 (OR: 5.1, 95%CI: 1.6-15.9; *P*=0.005) or ≥0.02 (OR: 5.3, 95%CI: 1.7-16.4; *P*=0.004).

In the exploratory analyses restricted to AML cases, we did not observe marked differences in time to AML diagnosis by *DNMT3A* R882H/C mutation status (detected *vs.* not detected) or by detection of any mutation at VAF of ≥0.005, ≥0.01 or ≥0.02 at either blood draw (*Online Supplementary Figure S5A-D*).

Clonal stability

We examined clonal evolution of mutations over time in 11 matched sets of women (NHS) with samples banked approximately ten years apart (Figure 2A). The VAF of mutations detected in these cases at blood collection one (median: 0.0021; range: 0.0003-0.0782) and blood collection two (median: 0.0037; range: 0.0006-0.2992) was very similar to controls at collection one (median: 0.0017; range: 0.0003-0.0731) and collection two (median: 0.0023; range: 0.0003-0.2689). In the cases with two blood collections, 31 clonal mutations occurred only at the first blood draw, 37 occurred only at the second blood draw, and 22 occurred at both time points (see Figure 2A; yellow data points connected with a line). Of the latter 22 clonal mutations, in the approximately ten years between the first and second blood draw, five (23%) increased by >0.01 VAF, none decreased by >0.01 VAF, and 17 (77%) were

unchanged. In controls with two blood collections, 27 clonal mutations occurred only at the first blood draw, 58 only at the second blood draw, and 29 at both time points (Figure 2A; blue and red data points connected by lines). Of the latter 29 clonal mutations, in the approximately ten years between the first and second blood draw, five (17%) increased by >0.01 VAF, none decreased by >0.01 VAF, and 24 (83%) were unchanged. In the 22 matched sets of men (HPFS) with only one banked sample, we again observed a similar VAF for clonal mutations detected in cases [median (range) VAF: 0.0020 (0.0002-0.3280)] and controls [median (range) VAF: 0.0014 (0.0002-0.3513)] (Figure 2B).

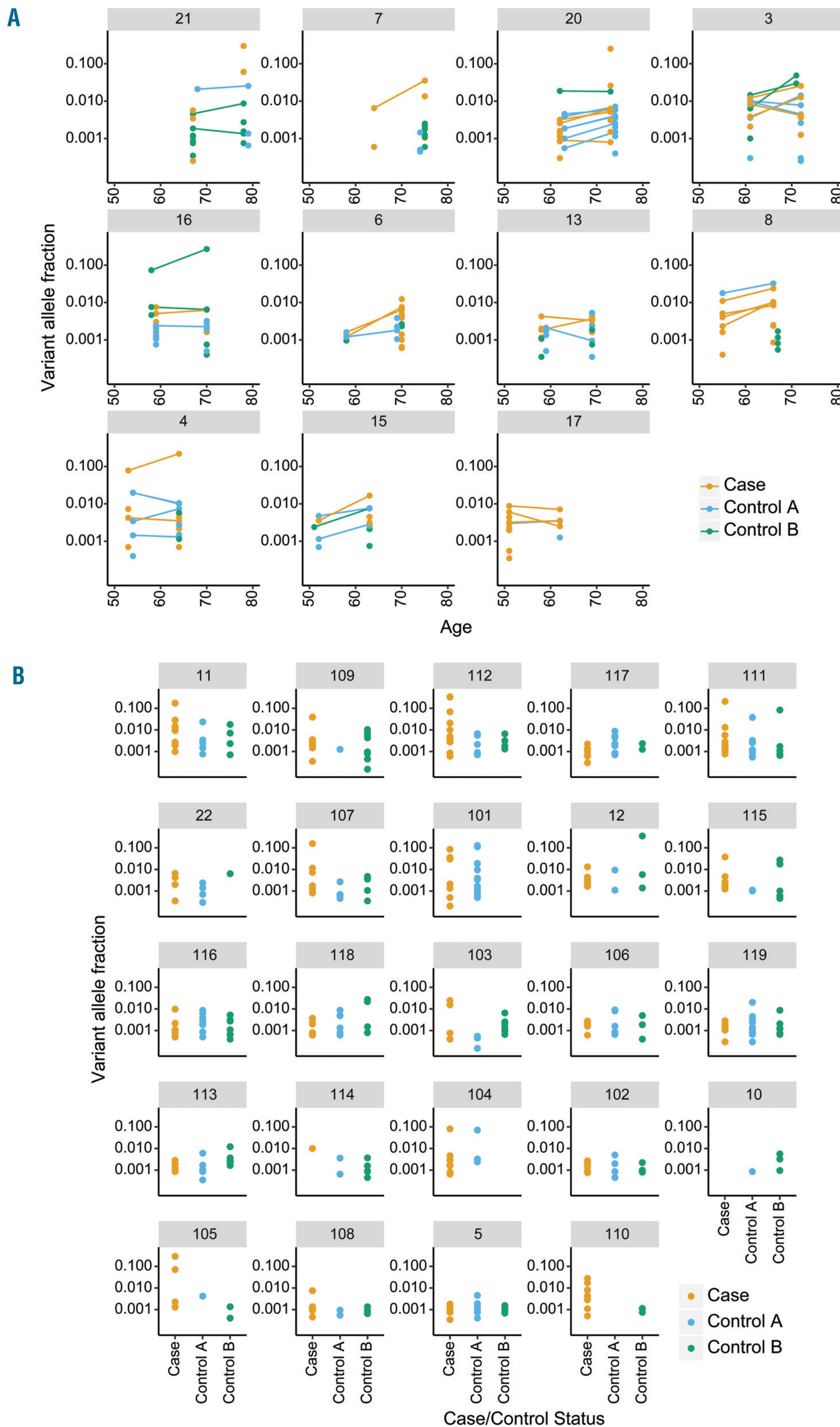

Figure 2. Clonal mutations detected in study participants. (A) Clonal dynamics for women with two blood samples collected approximately ten years apart. Trios of one case and two controls are shown in each panel. The ages at first and second collection are on the x-axis while the variant allele fraction is on the y-axis. Dots connected by a line represent the same mutation seen in both blood collections. Individual dots represent mutations only seen in a single blood collection. (B) Clonal mutations for individuals with only a single blood collection. Trios of one matched case and two controls are shown in each panel with female trios numbered <100 and male trios numbered >100. The y-axis portrays variant allele fraction and each dot represents a mutation seen in the respective individual.

In exploratory case-only analyses (see *Online Supplementary Methods*), the VAF for the most abundant clone observed at the first blood draw (i.e. the largest VAF observed at collection one) did not correlate to the time to diagnosis of AML (partial Spearman r = -0.11, *P*=0.55, adjusted for age and sex) (Figure 3A). In the NHS cases with a second collection blood sample, the maximum VAF at the second time point and time to AML diagnosis was not correlated (partial Spearman r=0.34, *P*=0.33, adjusted for age) (Figure 3B). The largest change in VAF between collections with time to AML diagnosis was also not correlated (partial Spearman r=0.30, *P*=0.39, adjusted for age) (Figure 3C).

Discussion

In this study, we investigated associations of clonal hematopoiesis with long-term risk of AML, leveraging ECS-determined clonal variants and up to 22 years of follow up after blood draw in 34 matched case-control sets from the NHS and HPFS. Surprisingly, we found no clear differences in clonal mutation abundance, location or VAF between cases and controls. As expected, *DNMT3A* and *TET2* were the genes with the most frequently detected clonal mutations in both cases and controls,[2,6,14] and overall, cases and controls showed abundant mutation across the rest of the coding sequence. Few individual variants occurred frequently enough for separate analysis of AML risk, but among those occurring in at least four participants, *DNMT3A* R882H/C had a strong association with AML risk. We also observed statistically significant associations with AML risk for individuals with any variant with a VAF ≥0.01. Contrary to expectation, in the 11 matched sets with two banked blood samples, we did not observe a signature of clonal evolution over time that distinguished cases from controls or predicted latency to AML diagnosis in the cases.

Two recent studies reported findings for clonal hematopoiesis and future risk of AML.[4,5] Briefly, both studies observed an increased risk of AML for increasing numbers of clonal mutations, higher VAF and detection or number of mutations in known driver genes. Of note, Abelson *et al*.[5] observed an increased AML risk for individuals with clones of VAF ≥0.005 detected by ECS, and Desai *et al*.[4] reported an increased risk of AML for women with clones of VAF ≥0.01 detected by targeted deep sequencing. We detected an increase in AML risk for persons with clonal mutations at ≥0.01 VAF and those with *DNMT3A* R882 mutations, and our observed effect estimates had a similar magnitude and precision as those reported by the previous studies. Other prior studies reported that these mutations raise AML risk by 0.5-1%/year.[2,3] For mutations detected at VAF <0.01, our findings were less clear due to limited statistical precision. Additionally, in the subset of women with repeat blood

Table 2. Future risk of acute myeloid leukemia associated with individual variants and selected variant allele frequencies detected in pre-diagnosis blood samplesa in a pooled sample from the NHS and HPFS cohorts.

Gene or VAF criterion[b]	Polymorphism	Major/minor allele	Total testing positive[a] Cases (N=34)[c]	Total testing positive[a] Controls (N=69)[c]	OR (95% CI)[d]	*P*
First blood collection only						
DNMT3A	R882H or R882C	C/T or G/A	7	3	6.3 (1.3, 30.7)	0.02
	R882H only	C/T	6	1	12.0 (1.4, 99.7)	0.02
	W860R	A/G	0	4	0.0 (NC)	
ASXL1	E1183K	G/A	4	1	2.0×10^7(NC)	
JAK2	V617F	G/T	4[e]	1[e]	8.0 (0.9, 71.6)	0.06
Any VAF ≥0. 001			33	57	2.3×10^7 (NC)	
Any VAF ≥0.005			22	31	2.4 (1.0, 6.1)	0.06
Any VAF ≥0.01			14	17	2.5 (0.9, 7.1)	0.07
Any VAF ≥0.02			11	10	3.2 (1.1, 9.7)	0.04
First or second blood collection[f]						
DNMT3A	R882H or R882C	C/T or G/A	8	3	7.3 (1.5, 34.7)	0.01
	R882H only	C/T	7	1	14.0 (1.7, 113.8)	0.01
	W860R	A/G	0	5	0.0 (NC)	
Any VAF ≥0.001			34	63	1.7×10^7 (NC)	
Any VAF ≥0.005			26	39	2.5 (1.0, 6.3)	0.05
Any VAF ≥0.01			20	19	5.4 (1.8, 16.6)	0.003
Any VAF ≥0.02			17	13	5.6 (1.8, 17.2)	0.003

AML: acute myeloid leukemia; CI: confidence interval; HPFS: Health Professionals Follow-up Study; NHS: Nurses' Health Study; OR: odds ratio; NC: not calculable due to zero or sparse cell counts; VAF: variant allele fraction. [a]A participant was considered positive for a given mutation if both technical repeats for the same collection time tested positive; otherwise the participant was classified as negative for that mutation and collection time. [b]Polymorphism-specific analyses were limited to individual polymorphisms detected in at least four individuals in a given blood collection; VAF cut-off point variables were defined according to all mutations detected in a given person in both technical repeats for the given blood collection. [c]The pooled N for cases includes 15 women in the NHS and 19 men in the HPFS; the pooled N for the controls includes 31 women in the NHS and 38 men in the HPFS. A second blood sample was available for 11 cases and 21 controls from the NHS. [d]The OR, 95%CI and *P*-values were calculated using conditional logistic regression, conditioning on the matched sets [matched on cohort (e.g. sex), age, and date of blood draw]. [e]*JAK2* V617F was detected only in men. [f]Polymorphism-specific results were tabulated only for the polymorphisms with additional positive case or control samples in NHS blood collection 2.

samples, we did not observe clonal expansion over ten years and found no evidence among the AML cases that the most abundant clone at either an early or late time point, or the largest difference in VAF between time points for any clone, correlated with time to AML onset. Similarly, neither of the previous studies observed differences in clonal expansion in individuals with serial samples who did or did not eventually develop AML.[4,5] However, Desai *et al*.[4] observed striking differences in time to AML diagnosis for individuals with any baseline mutation and noted that the degree of diminished latency varied by mutation and clonal complexity. With our smaller sample size, we lacked resolution to perform as detailed an analysis of mutational complexity of clonal hematopoiesis, or of temporal changes, as the prior larger studies. Nonetheless, our findings extended, by several years, the pre-diagnosis period during which detection of clonal hematopoiesis could be informative for identifying individuals at an increased risk for AML.

Notably, we observed relatively similar VAF of clonal hematopoietic mutations in cases and controls, whereas the Abelson *et al*. and Desai *et al*. studies[4,5] reported more striking differences in the overall VAF and mutational complexity of CHIP in cases and controls. The explanation for these discrepancies is not immediately clear, although differences in methodology for control matching or differing average lengths of follow up across the three studies may have contributed. In the present study, we did not have sufficient sample size to compare mutational profiles of cases *versus* controls within more proximal and more distal follow-up periods, but it is plausible that contrasts in clonal hematopoiesis profiles between individuals who do and do not progress may deepen as diagnosis of malignancy approaches.

Our observation of an increased risk of AML in individuals with variants at the *DNMT3A* R882 locus is unsurprising, given the prevalence of *DNMT3A* R882 hotspot mutations in AML,[15] but also highlights that different mutations in the same gene do not convey the same risk and should not be viewed as equivalent *a priori*. Of interest, one of the few other individual variants that occurred relatively frequently in the present study sample, *DNMT3A* W860R, occurred more commonly in controls than in cases. This raises the question as to whether the aggregation of variable mutations across any single gene is appropriate to evaluate true AML risk. Larger studies with sufficient statistical power to examine individual mutations at varying VAF (and, perhaps, combinations of individual variants) may prove informative for further refining the interpretation of clonal hematopoiesis for stratifying risk of AML.

The strengths of this research include studying two large, well-characterized population-based cohorts with many years of follow up after blood collection. We matched cases and controls carefully on potential confounding variables including age, ethnicity, sex, and date(s) of blood collection and utilized conditional logistic regression for efficient control of confounding by those variables in the analysis. Further, for a subset of women in the NHS, we explored and compared temporal changes in clonal hematopoiesis over an approximately 10-year interval in those who did or did not subsequently develop AML. We conducted ECS assays and ddPCR validation in a blinded manner and observed strong reproducibility of variant calls across orthogonal platforms, affirming the

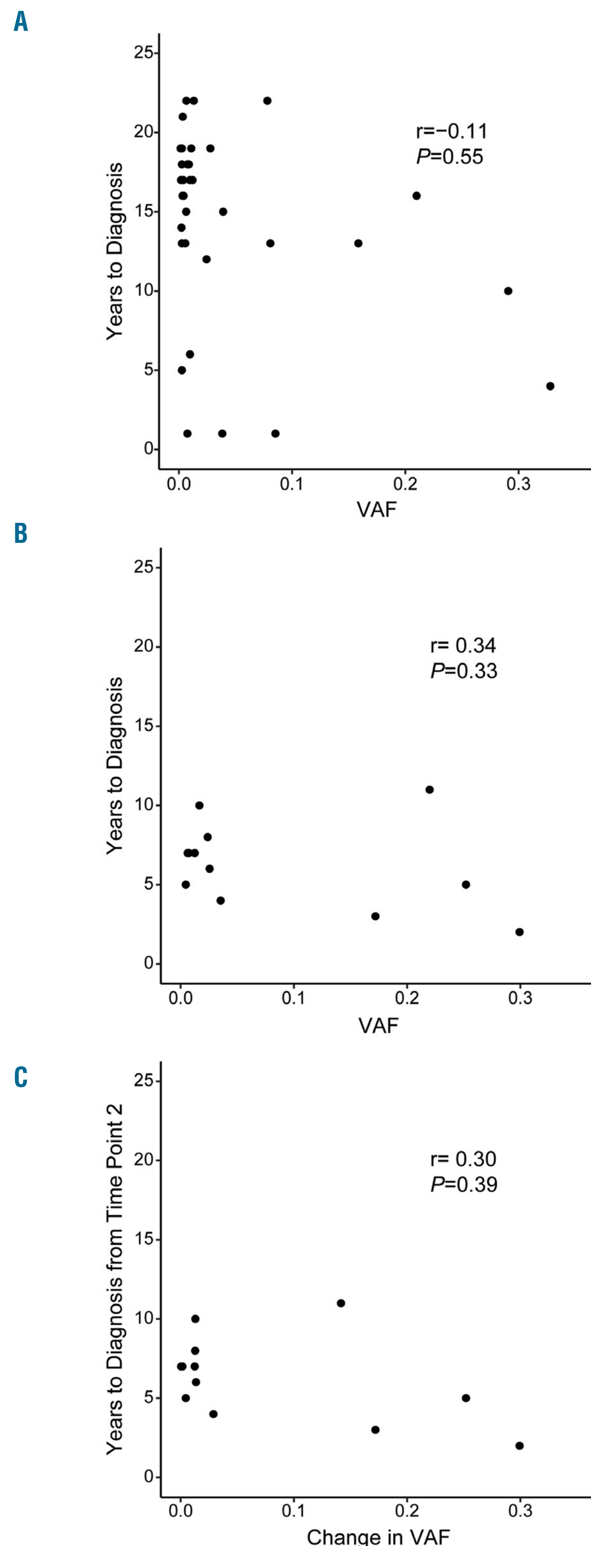

Figure 3. Comparison between the largest variant allele fraction (VAF) or largest change in VAF per individual with acute myeloid leukemia (AML) (x-axis) and the time to AML diagnosis (y-axis). (A) The VAF for the most common mutation observed in a first or only blood collection sample is plotted relative to the time to AML diagnosis from the date of first or only blood collection. (B) The VAF for the most common mutation observed in a second blood collection sample is plotted relative to the time to AML diagnosis from the date of second blood collection. (C) The time from the second blood draw to AML diagnosis in participants with two blood samples, based on the largest observed increase in VAF between the first and second blood collections (regardless of the corresponding variant).

credibility of variants detected at very low VAF.

The most notable limitations of the study relate to sample size and statistical precision, as previously noted. We had insufficient statistical power for concurrent interrogation of multiple mutations and for more than exploratory analysis of clone size and time to AML diagnosis. Likewise, we could not control for potential confounding variables (other than matching factors), such as body mass index or history of cigarette smoking,[10,12] or stratify by those variables or by follow-up time. Additionally, myeloproliferative neoplasms and myelodysplastic syndrome were not routinely reported in the NHS and HPFS, so we were unable to identify which participants had clonal hematopoiesis attributable to one of these pre-malignant disorders. Further, we did not have access to AML diagnostic samples for the cases in this study, making it impossible to determine which, if any, clonal hematopoietic mutations detected prior to diagnosis appeared in the founding AML clone. The study was limited technically by the sequencing panel, which targeted 54 genes recurrently mutated in AML. Future studies should expand the panel to target the entire exome or at least include additional genes that have been observed in CHIP, such as *PPM1D*,[1] to more fully characterize the spectrum of mutations in clonal hematopoiesis. However, any increase in panel size must be balanced with the cost of sequencing, which is higher for ECS compared to conventional NGS. Lastly, ECS, while precise, cannot co-localize mutations within the same cell. Future single-cell sequencing studies would provide further insights into the evolution of preleukemic clones and potentially improve screening for risk of developing AML.

In summary, we demonstrated that detection of AML-associated variants at VAF as low as 0.01 is associated with long-term risk of AML in concordance with other recent reports. Additionally, our study has extended by several years the period of follow up over which this increased risk applies and provided evidence that even individual variants in known driver genes may be associated with AML risk, suggesting that not all clonal somatic variants have equivalent associations with AML. The collective data from this and previous reports underscore that, while clonal hematopoiesis is associated with a markedly increased long-term risk of AML, the vast majority of individuals with detectable clonal hematopoiesis will not develop AML. Likewise, further detailed investigation is needed to incorporate detection of clonal hematopoiesis into AML-risk assessment for healthy individuals. Such studies will require considerably larger populations, ideally with serial samples and sufficient sample size to analyze multiple features of clonal hematopoiesis (including individual variants, gene-level mutational profiles and temporal evolution of variant clones) as well as additional genetic, epigenetic and environmental factors that may influence the stepwise progression of healthy cells to leukemic clones. Future work will also need to incorporate single cell sequencing technology to identify, which rare clonal mutations occur in the same cells and tease out the sequence of mutation acquisition driving the transformation from clonal hematopoiesis to AML.

Funding

This project was supported in part by the National Institutes of Health (R01 CA211711, UM1 CA186107, P01 CA87969, R01 CA49449, UM1 CA167552 and R01 CA149445), the Children's Discovery Institute of Washington University and St Louis Children's Hospital (MC-II-2015-461), and Hyundai Hope on Wheels (2015Q3-3).

References

1. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015; 126(1):9-16.
2. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371(26):2488-2498.
3. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014; 371(26):2477-2487.
4. 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.
5. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature. 2018; 559(7714):400-404.
6. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun. 2016; 712484.
7. Wong TN, Ramsingh G, Young AL, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature. 2015;518(7540):552-555.
8. Hu FB, Willett WC. Diet and coronary heart disease: findings from the Nurses' Health Study and Health Professionals' Follow-up Study. J Nutr Health Aging. 2001;5(3):132-138.
9. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. Nat Rev Cancer. 2005;5(5):388-396.
10. Hankinson SE, Willett WC, Manson JE, et al. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal Women. JNCI J Natl Cancer Inst. 1995;87(17):1297-1302.
11. Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. Breast Cancer Res Treat. 2013;137(3):883-892.
12. Poynter JN, Richardson M, Blair CK, et al. Obesity over the life course and risk of acute myeloid leukemia and myelodysplastic syndromes. Cancer Epidemiol. 2016; 40:134-140.
13. R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/.
14. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20(12):1472-1478.
15. Cancer Genome Research Atlas Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N Engl J Med. 2013;368(22):2059-2074.