Supplemental Digital Content (SDC)

Supplement to: Sex differences in the spectrum of clonal hematopoiesis. Priscilla Kamphuis, Isabelle A. van Zeventer, Aniek O. de Graaf, Jonas B. Salzbrunn, Maaike G.J.M. van Bergen, Avinash G. Dinmohamed, Bert A. van der Reijden, Jan Jacob Schuringa, Joop H. Jansen, Gerwin Huls

Supplementary Materials and Methods

This study was performed within the Lifelines Cohort study, which is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. The Lifelines cohort study was shown to be representative for the general population living in the Northern part of the Netherlands.[1] The study was performed in compliance with the Declaration of Helsinki and approved by the medical ethical committee of the University Medical Center Groningen. Within this Lifelines population-based cohort we used targeted error-corrected next-generation sequencing to acquire data on clonal hematopoiesis (CH) from 4715 individuals of ≥60 years, as described earlier (Suppl. Figure S2).[2] A custom panel of single-molecule-tagged molecular inversion probes covering target regions in 27 myeloid and lymphoid malignancy associated genes was used (Suppl. Table S2). The mean number of aligned consensus reads was 8827, with a coverage >500x for 97.8% of all targeted regions (Suppl. Figure S3). Variants were called with a variant allele frequency (VAF) of at least 1% and with at least 10 consensus reads and curated for artefacts and polymorphisms. Given the interest in CH in the context of peripheral blood cell abnormalities,[3-6] the cohort was enriched for individuals with blood count abnormalities (Suppl. Figure S1).

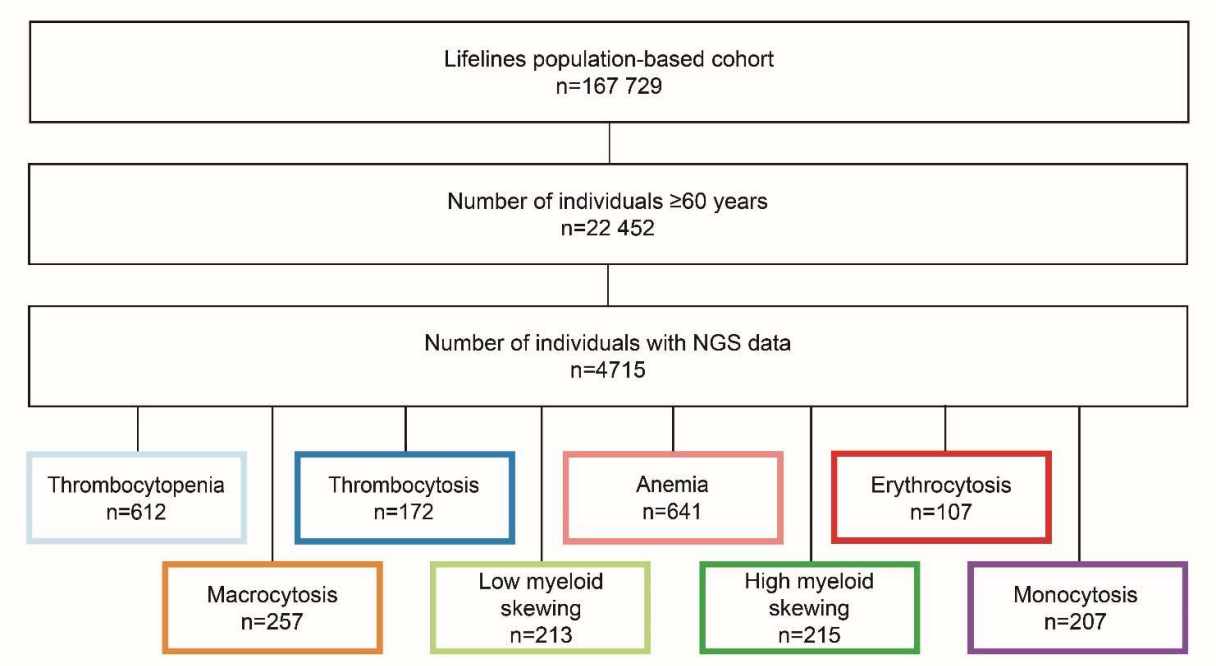
All statistical analyses were performed with R version 4.0.2. For data reshaping and summarization, we used the R packages *dplyr v1.0.7* (https://github.com/tidyverse/dplyr/), *tidyr v1.1.4* (https://github.com/tidyverse/tidyr), *stringr v1.4.0* (https://github.com/tidyverse/stringr) and *data.table v1.14.2* (https://github.com/Rdatatable/data.table). Visualizations were made using *ggplot2 v3.3.5* (https://github.com/tidyverse/ggplot2) and *RColorBrewer v1.1-2* (https://github.com/cran/RColorBrewer). The circos plots to show co-mutational patterns were made using the R package *circlize v0.4.14* (https://github.com/jokergoo/circlize). Survival analyses and visualizations were performed with R packages *survival v3.2-13* (https://github.com/therneau/survival) and *survminer v0.4.9* (https://github.com/kassambara/survminer).

By linkage with the Netherlands Cancer Registry (IKNL) we retrieved the incidence of hematological malignancies until December 2019. IKNL has a nationwide coverage of at least 95% and receives notifications of all newly diagnosed malignancies, confirmed by histology and/or cytology, in the Netherlands since 1989. Participants were linked using pseudonyms based on the first 8 characters of their last name, date of birth, sex and postal code at time of diagnosis. We used the ICD-O morphology codes 9590-9999 to identify hematological malignancies. Time to malignancy was calculated from inclusion in the Lifelines study until registered diagnosis of hematological malignancy. Individuals with a recorded prevalent malignancy were excluded from the analysis. Linkage with IKNL succeed for 4666/4715 individuals (99%) of our cohort. For males, we obtained linkage for 992 individuals with CH and 1445 individuals without CH. Linkage was successful for 937 females with CH and 1292 females without CH.

Supplementary Figure 1 - Cohort overview and selection of cases for next-generation sequencing

(A) Overview of selected individuals and blood count abnormalities in the next-generation sequencing (NGS) cohort. Remaining individuals were selected as population-based matched controls without a blood count abnormality. (B) Number and percentage of males and females among included individuals with peripheral blood count abnormalities.

A



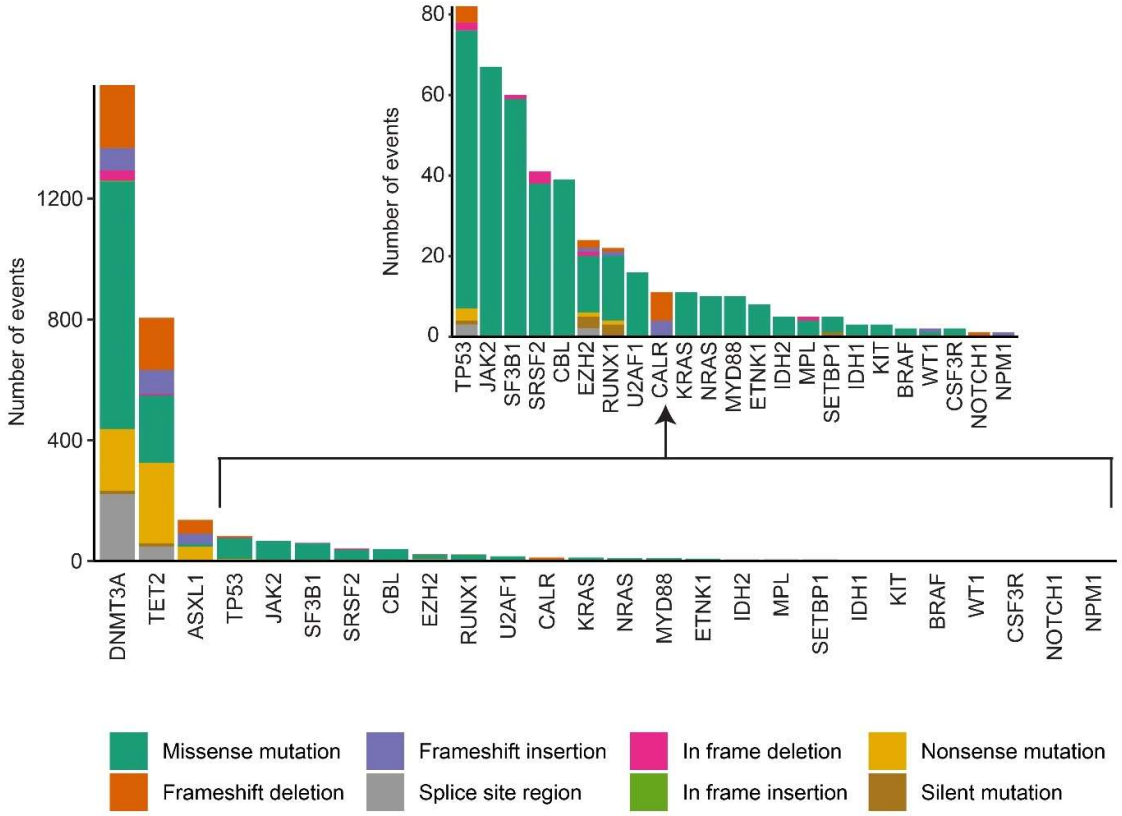
B

Blood count abnormality	Males (%)	Females (%)	p-value
Monocytosis[1]	153 (73.9%)	54 (26.1%)	<0.001
Thrombocytopenia[2]	481 (78.6%)	131 (21.4%)	<0.001
Thrombocytosis[3]	43 (25%)	129 (75%)	<0.001
High myeloid skewing[4]	135 (62.8%)	80 (37.2%)	0.002
Low myeloid skewing[5]	44 (20.7%)	169 (79.3%)	<0.001
Macrocytosis[6]	131 (51%)	126 (49%)	0.700
Erythrocytosis[7]	37 (34.6%)	70 (65.4%)	<0.001
Anemia[8]	285 (44.5%)	356 (55.5%)	<0.001

[1]Monocytosis was defined as peripheral monocyte count ≥1 x10^9/L.[3] [2]Thrombocytopenia was defined as peripheral platelet count <150 x10^9/L. [3]Thrombocytosis was defined as peripheral platelet count >400 x10^9/L. [4]High myeloid skewing was defined as the highest percentile (>99%) of myeloid cell percentage in individuals ≥60 years. [5]Low myeloid skewing was defined as the lowest percentile (<1%) of myeloid cell percentage in individuals ≥60 years. [6]Macrocytosis was defined as MCV >100 fL. [7]Erythrocytosis was defined in females as hemoglobin >16.5 g/dL or hematocrit ≥48% and in males as hemoglobin >18.5 g/dL or hematocrit ≥52%.[4] [8]Anemia was defined in accordance to WHO definitions: hemoglobin levels in females <12.0 g/dL or <13.0 g/dL in males.[5]

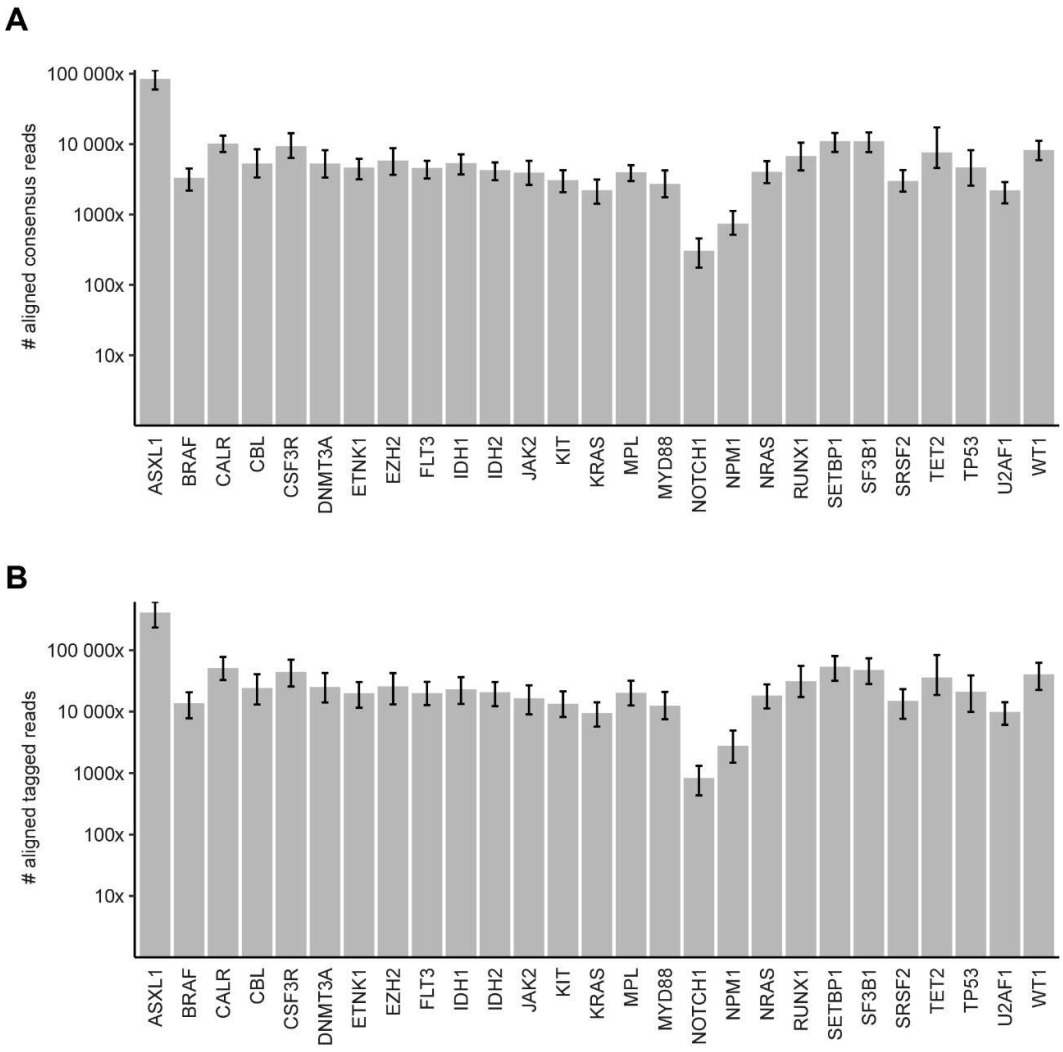
Supplementary Figure 2 – Overview of mutated genes

Bar plot to show the number and type of mutations per gene in the complete NGS cohort (n=4715). NGS, next-generation sequencing.



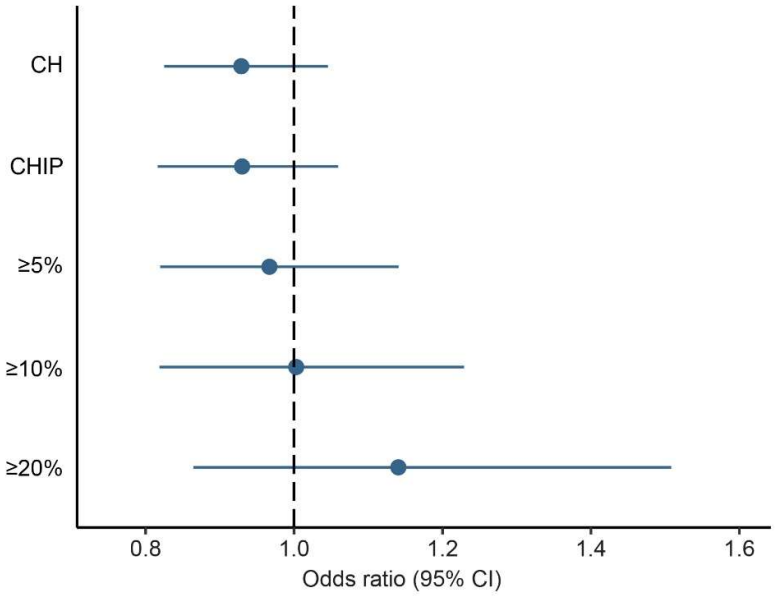
Supplementary Figure 3 - Coverage

Number of aligned consensus reads (A) and the raw number of aligned reads (B) for all genes included in the sequencing panel for the NGS cohort (n=4715). Columns and error bars indicate median and interquartile range respectively. NGS, next-generation sequencing.



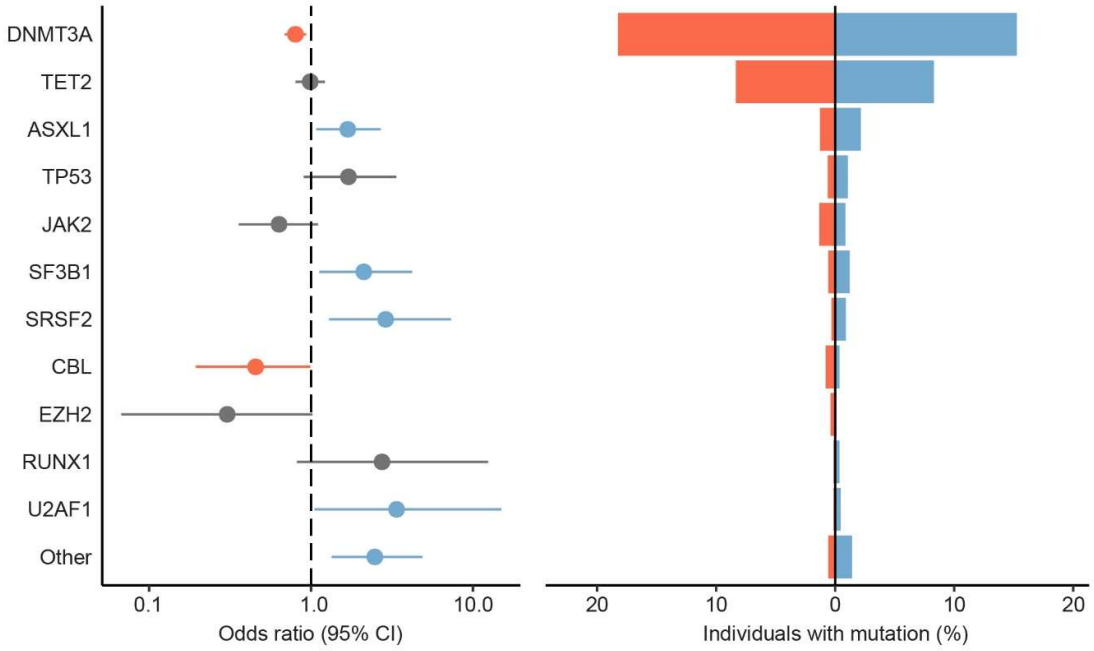
Supplementary Figure 4 – Odds ratios for the association between male sex and CH with different VAF cut-offs

Forest plot indicating the OR and 95% CIs (horizontal lines) for the association between male sex and CH, CHIP (VAF cut-off ≥2%), CH with VAF cut-off ≥5%, CH with VAF cut-off ≥10% and CH with VAF cut-off ≥20% respectively, as derived from multivariable logistic regression with correction for age. OR, odds ratio; CI, confidence interval; CH, clonal hematopoiesis; CHIP, Clonal hematopoiesis of indeterminate potential; VAF, variant allele frequency.



Supplementary Figure 5 – Odds ratios and pyramid plot for the association between CHIP and sex (VAF ≥2%)

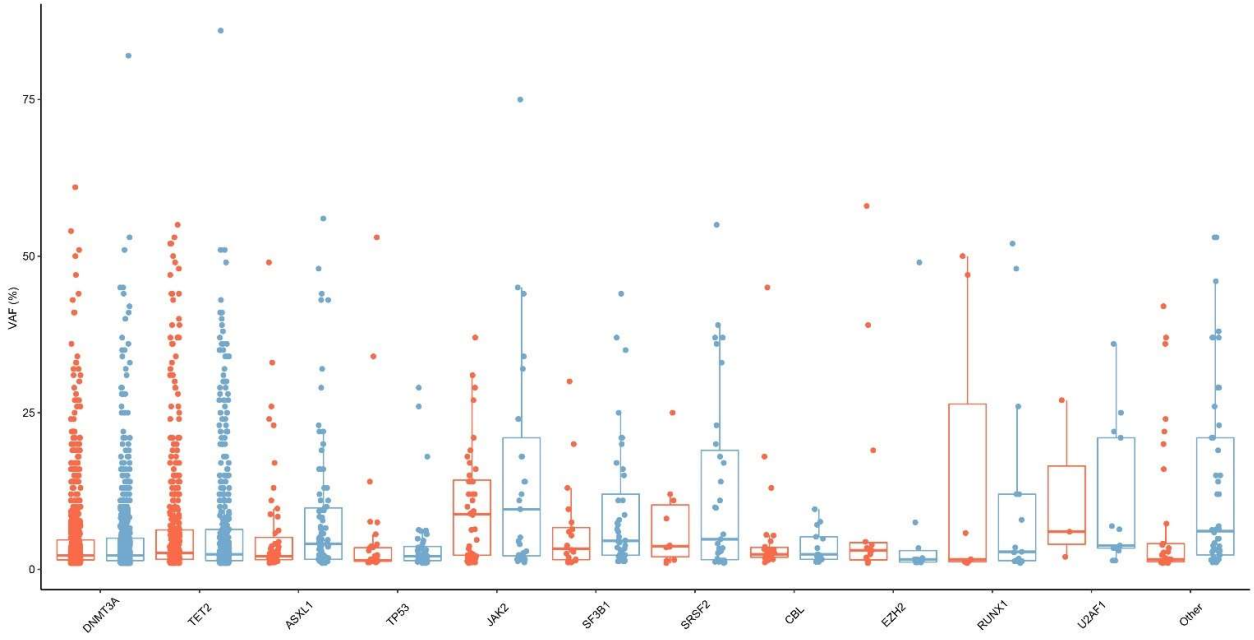
Forest plot indicating the odds ratio (OR) and 95% confidence intervals (CI) for the association between recurrently mutated genes and male sex. A blue line indicates a significantly higher frequency in males, whereas orange-colored lines indicate a significantly higher frequency in females. Pyramid plot indicating the proportion of individuals with a mutation in the most commonly mutated genes for males (blue) and females (orange). Mutations with a VAF of at least 2% are included in the analyses. VAF, variant allele frequency.



Supplementary Figure 6 – VAF distribution per gene

(A) VAFs for all detected somatic mutations in males (blue) and females (orange). Genes with <10 mutations in either males or females in this cohort were grouped in the category *other*. (B) Median VAFs detected in males and females for the genes that show sex-specific differences in prevalence. P-values for the difference in VAF distribution between males and females are given. VAF, variant allele frequency.

A

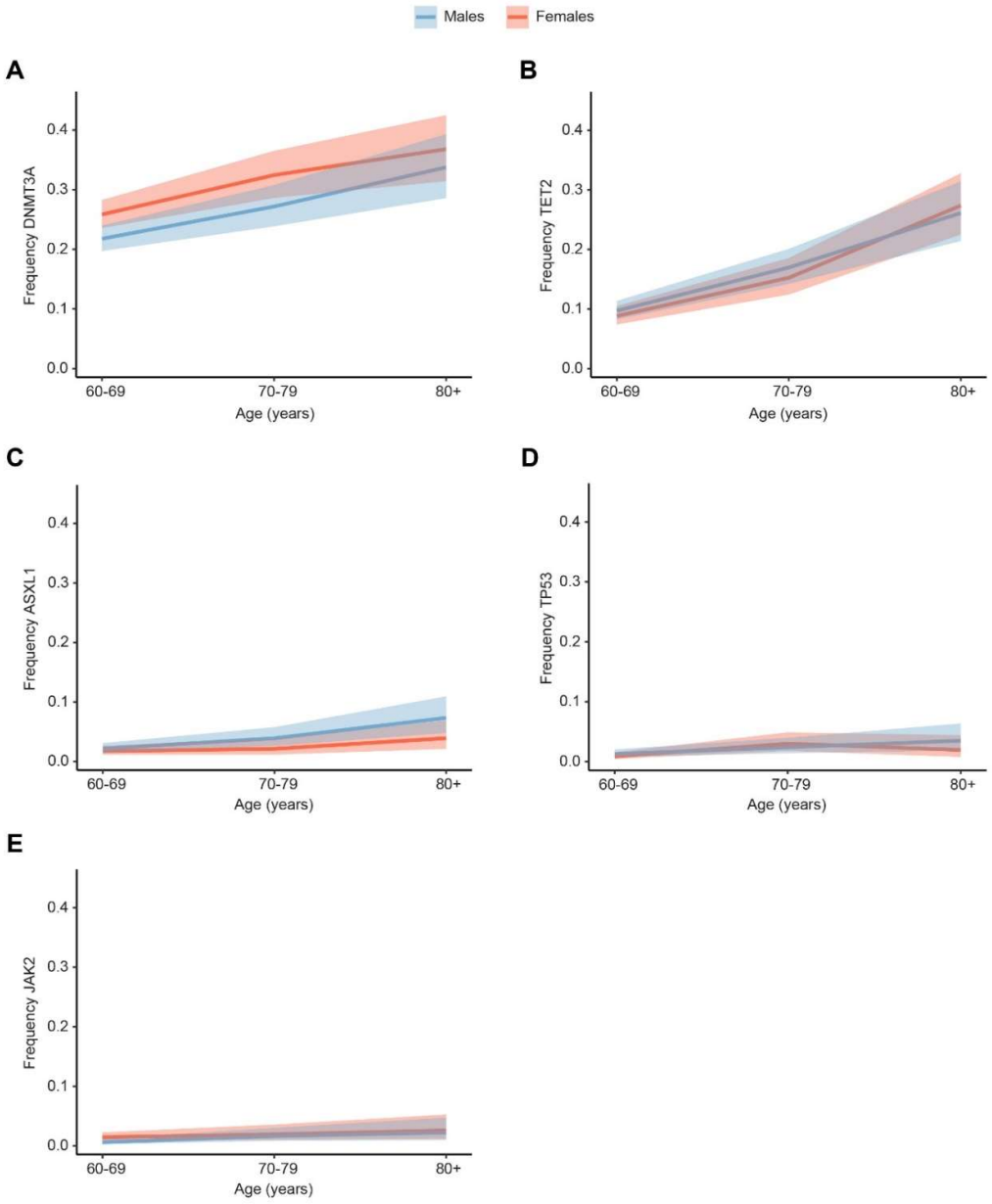


B

Gene	Median VAF males	Median VAF females	p-value
DNMT3A	2.55	2.70	0.723
CBL	2.40	2.30	0.882
ASXL1	4.30	2.50	0.118
SF3B1	4.60	3.30	0.196
SRSF2	4.80	3.70	0.627
U2AF1	3.80	6.00	0.893

Supplementary Figure 7 – Age-related increase in the prevalence of most frequently mutated genes, stratified by sex

(A) Age-related increase in the prevalence of *DNMT3A* mutations in males (blue, n=610 mutated individuals) and females (orange, n=653 mutated individuals). (B) Age-related increase in the prevalence of *TET2* mutations in males (detected in n=339) and females (detected in n=292). (C) Age-related increase in the prevalence of *ASXL1* mutations in males (n=81 mutated individuals) and females (n=48 mutated individuals). (D) Age-related increase in the prevalence of *TP53* mutations in males (n=46 mutated individuals) and females (n=35 mutated individuals). (E) Age-related increase in the prevalence of *JAK2* mutations in males (detected in n=27) and females (detected in n=39). Shaded areas represent 95% confidence intervals.



Supplementary Table 1 – Sex-associated mutations

The following tables summarizes a selection of literature describing sex differences for mutations detected in (A) MDS, (B) Unclassifiable MDS/MPN syndrome, (C) MDS/MPN overlap syndrome, (D) MPN and (E) AML.

A

MDS	
Male	**Female**
U2AF1[7]	DNMT3A[7]
ZRSR2[7]	TP53[7]

B

Unclassifiable MDS/MPN syndrome	
Male	**Female**
ASXL1[8]	DNMT3A[8]
	IDH2[8]

C

MDS/MPN overlap syndrome	
Male	**Female**
EZH2[9]	

D

MPN	
Male	**Female**
IDH2[10]	SF3B1[10]
ASXL1[10,11]	
EZH2[10]	
SRSF2[10]	
U2AF1[10]	

E

AML	
Male	**Female**
ZRSR2[7]	KDM6A[7]
PHF6[7]	FLT3_ITD[7]
U2AF1[7]	NPM1[7]
SRSF2[7]	DNMT3A[7]
STAG2[7]	TP53[7]
ASXL1[7]	
BCOR[7]	
RUNX1[7]	
TET2[7]	

Supplementary Table 2 – Overview of sequenced genes and regions

Gene	Reference transcript	ENSEMBL reference transcript	Exon	Targeted codons/region
ASXL1	NM_015338	ENST00000375687	13 (partially)	exon 13
BRAF	NM_004333.4	ENST00000288602	15 (partially)	codon 600
CALR	NM_004343	ENST00000316448	9	exon 9
CBL	NM_005188	ENST00000264033	8-9	exon 8 and 9
CSF3R	NM_156039	ENST00000373103	14, 17	codon 618, 615 and exon 17
DNMT3A	NM_175629	ENST00000264709	2-23 (all coding exons)	all coding exons
ETNK1	NM_018638	ENST00000266517	3 (partially)	codon 243-244
EZH2	NM_004456	ENST00000320356	2-20 (all coding exons)	all coding exons
FLT3_835	NM_004119	ENST00000241453	20 (partially)	codon 835-842
IDH1	NM_005896	ENST00000415913	4 (partially)	codon 132
IDH2	NM_002168	ENST00000330062	4 (partially)	codon 140, 172
JAK2	NM_004972	ENST00000381652	12, 14 (partially)	codon 617 and exon 12
KIT	NM_000222	ENST00000288135	8 (partially), 17 (partially)	codon 816, 419
KRAS	NM_004985	ENST00000256078	2-3 (partially)	a.o. codon 12, 13, 61
MPL	NM_005373	ENST00000372470	10 (partially)	codon 515, 505
MYD88	NM_002468.4	ENST00000417037	4-5 (partially)	codon 265 and 232
NOTCH1	NM_017617.4	ENST00000277541	34 (partially)	codon 2514
NPM1	NM_002520	ENST00000517671	11 (partially)	codon 288-290
NRAS	NM_002524	ENST00000369535	2-3 (partially)	a.o. codon 12, 13, 61
RUNX1	NM_001754	ENST00000437180	2-9 (all coding exons)	all coding exons
SETBP1	NM_015559	ENST00000282030	4 (partially)	codon 850-910
SF3B1	NM_012433	ENST00000335508	13-16	codon 575-790
SRSF2	NM_003016	ENST00000392485	1 (partially)	codon 95, 96
TET2	NM_001127208	ENST00000380013	3-11 (all coding exons)	all coding exons
TP53	NM_000546	ENST00000269305	2-11 (all coding exons)	all coding exons
U2AF1	NM_006758	ENST00000291552	2, 6 (partially)	codon 34, 157
WT1	NM_024426	ENST00000332351	7, 9	exon 7 and 9

Supplementary Table 3 – Associations between CH/CHIP and sex corrected for age and blood count abnormalities

(A) Odds ratios (OR) and 95% confidence intervals (CI) for the association between male sex and CH (VAF ≥1%) or CHIP (VAF ≥2%), derived from logistic regression corrected for age and the presence of selected blood count abnormalities. (B) Odds ratios (OR) and 95% confidence intervals (CI) for the association between male sex and most commonly mutated genes in CH and CHIP derived from logistic regression with age and the presence of selected blood count abnormalities as covariables. Blood count abnormalities included monocytosis, thrombocytopenia, thrombocytosis, low myeloid skewing, high myeloid skewing, high MCV/macrocytosis, erythrocytosis and anemia (Suppl. Figure S1). CH, clonal hematopoiesis; CHIP, clonal hematopoiesis of indeterminate potential; VAF, variant allele frequency.

A

Variable	OR CH (95% CI)	OR CHIP (95% CI)
Clonal hematopoiesis	0.95 (0.84 - 1.08)	0.94 (0.82 – 1.08)

B

Gene	OR CH (95% CI)	OR CHIP (95% CI)
DNMT3A	0.81 (0.71 – 0.93)	0.79 (0.68 – 0.93)
TET2	1.05 (0.88 – 1.26)	1.00 (0.80 – 1.24)
ASXL1	1.57 (1.07 – 2.32)	1.70 (1.06 – 2.78)
TP53	1.07 (0.68 – 1.71)	1.57 (0.81 – 3.18)
JAK2	1.03 (0.59 – 1.79)	1.09 (0.59 – 2.02)
SF3B1	1.58 (0.90 – 2.85)	1.86 (0.96 – 3.82)
SRSF2	2.62 (1.28 – 5.82)	2.78 (1.21 – 7.21)
CBL	0.45 (0.22 – 0.89)	0.44 (0.19 – 0.98)
EZH2	0.67 (0.28 – 1.55)	0.39 (0.09 – 1.34)
RUNX1	1.47 (0.59 – 3.96)	2.53 (0.73- 11.58)
U2AF1	3.86 (1.12 – 18.38)	2.69 (0.78 – 12.47)
Other	1.34 (0.81 – 2.25)	3.17 (1.59 – 6.67)

Supplementary Table 4 – Associations between CH/CHIP and sex corrected for age, smoking and blood count abnormalities

(A) Odds ratios (OR) and 95% confidence intervals (CI) for the association between male sex and CH (VAF ≥1%) or CHIP (VAF ≥2%), derived from logistic regression corrected for age, smoking and the presence of selected blood count abnormalities. (B) Odds ratios (OR) and 95% confidence intervals (CI) for the association between male sex and most commonly mutated genes in CH and CHIP derived from logistic regression with age, smoking and selected blood count abnormalities as covariables. Blood count abnormalities included monocytosis, thrombocytopenia, thrombocytosis, low myeloid skewing, high myeloid skewing, high MCV/macrocytosis, erythrocytosis and anemia (Suppl. Figure S1). The smoking status of participants was categorized as current smoker, former smoker or never smoker.[12] CH, clonal hematopoiesis; CHIP, clonal hematopoiesis of indeterminate potential; VAF, variant allele frequency.

A

Variable	OR CH (95% CI)	OR CHIP (95% CI)
Clonal hematopoiesis	0.93 (0.81 - 1.06)	0.92 (0.80 - 1.06)

B

Gene	OR CH (95% CI)	OR CHIP (95% CI)
DNMT3A	0.77 (0.67 - 0.89)	0.77 (0.65 – 0.91)
TET2	1.11 (0.92 – 1.34)	1.02 (0.81 – 1.29)
ASXL1	1.43 (0.96 – 2.16)	1.45 (0.88 – 2.44)
TP53	0.92 (0.57 – 1.52)	1.21 (0.60 – 2.55)
JAK2	1.08 (0.60 – 1.95)	1.18 (0.60 – 2.28)
SF3B1	1.38 (0.76 – 2.57)	1.62 (0.81 – 3.42)
SRSF2	2.30 (1.08 – 5.30)	2.56 (1.05 – 6.93)
CBL	0.41 (0.19 – 0.86)	0.41 (0.16 – 0.97)
EZH2	0.71 (0.28 – 1.74)	0.46 (0.10 – 1.75)
RUNX1	1.71 (0.65 – 4.86)	2.63 (0.71 – 12.60)
U2AF1	3.79 (1.00 – 19.52)	2.84 (0.77 – 13.90)
Other	1.45 (0.85 – 2.50)	3.18 (1.55 – 6.91)

References

1. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. *PLoS One*. 2015;10(9):e0137203. doi:10.1371/journal.pone.0137203
2. van Zeventer IA, Salzbrunn JB, de Graaf AO, et al. Prevalence, predictors, and outcomes of clonal hematopoiesis in individuals aged ≥80 years. *Blood Adv*. Apr 27 2021;5(8):2115-2122. doi:10.1182/bloodadvances.2020004062
3. van Zeventer IA, de Graaf AO, Koorenhof-Scheele TN, et al. Monocytosis and its association with clonal hematopoiesis in community-dwelling individuals. *Blood Adv*. May 13 2022;doi:10.1182/bloodadvances.2021006755
4. Wouters H, Mulder R, van Zeventer IA, et al. Erythrocytosis in the general population: clinical characteristics and association with clonal hematopoiesis. *Blood Adv*. Dec 22 2020;4(24):6353-6363. doi:10.1182/bloodadvances.2020003323
5. van Zeventer IA, de Graaf AO, Wouters H, et al. Mutational spectrum and dynamics of clonal hematopoiesis in anemia of older individuals. *Blood*. Apr 2 2020;135(14):1161-1170. doi:10.1182/blood.2019004362
6. van Zeventer IA, de Graaf AO, van der Klauw MM, et al. Peripheral blood cytopenias in the aging general population and risk of incident hematological disease and mortality. *Blood Adv*. Sep 14 2021;5(17):3266-3278. doi:10.1182/bloodadvances.2021004355
7. De-Morgan A, Meggendorfer M, Haferlach C, Shlush L. Male predominance in AML is associated with specific preleukemic mutations. *Leukemia*. Mar 2021;35(3):867-870. doi:10.1038/s41375-020-0935-5
8. Bose P, Nazha A, Komrokji RS, et al. Mutational landscape of myelodysplastic/myeloproliferative neoplasm-unclassifiable. *Blood*. Nov 8 2018;132(19):2100-2103. doi:10.1182/blood-2018-05-848473
9. Karantanos T, Gondek LP, Varadhan R, et al. Gender-related differences in the outcomes and genomic landscape of patients with myelodysplastic syndrome/myeloproliferative neoplasm overlap syndromes. *Br J Haematol*. Jun 2021;193(6):1142-1150. doi:10.1111/bjh.17534
10. Karantanos T, Chaturvedi S, Braunstein EM, et al. Sex determines the presentation and outcomes in MPN and is related to sex-specific differences in the mutational burden. *Blood Adv*. Jun 23 2020;4(12):2567-2576. doi:10.1182/bloodadvances.2019001407
11. Stein BL, Williams DM, O'Keefe C, et al. Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. *Haematologica*. Oct 2011;96(10):1462-9. doi:10.3324/haematol.2011.045591
12. Slagter SN, van Vliet-Ostaptchouk JV, Vonk JM, et al. Associations between smoking, components of metabolic syndrome and lipoprotein particle size. *BMC Med*. Sep 3 2013;11:195. doi:10.1186/1741-7015-11-195