Supplemental files to: Prevalence, predictors and outcomes of clonal hematopoiesis in the oldest old

Isabelle A. van Zeventer, Jonas B. Salzbrunn, Aniek O. de Graaf, Bert A. van der Reijden, H. Marike Boezen, Judith M. Vonk, Pim van der Harst, Jan Jacob Schuringa, Joop H. Jansen and Gerwin Huls

Supplemental Methods

Questionnaires, blood samples and clinical examination

Extensive self-administered questionnaires on medical history, current diseases and health related behavior were filled in by Lifelines participants at time of first screening visit (2007-2014). Current medication use was verified by a certified doctor assistance and coded according to the general Anatomical Therapeutic Chemical Classification System 2008 (ATC) system.[1] The accuracy of self-reported medication use was recently verified in Lifelines upon comparison to the University of Groningen IADB.nl pharmacy prescription database.[2] Education level was classified into four categories: no education, primary, secondary or tertiary education. Smoking status was categorized into non-smokers (not in the last month and never longer than one year), ex-smokers (smoked longer than one year and have not smoked during last month and reported to have stopped smoking) or current smokers (longer than one year and have not reported to have stopped). Self-reported job descriptions were available for 467 individuals and coded into ISCO-88 scores as previously described.[3] Individuals with a Computer Assisted Structured Coding Tool (CASCOT)[4] ≤60 were inspected and where possible manually coded into ISCO-88 scores on basis of the classification table from the International Labour Organization.[5]

For all routine clinical chemistry assays, peripheral blood samples were drawn and directly processed at the University Medical Center Groningen. Total and differential blood counts were measured on a XE2100-system (Sysmex, Japan). Blood glucose was measured according to standard protocols on a Roche Modular P chemistry analyzer (Roche, Switzerland).

At first screening visit, multiple clinical examinations and investigations were performed. Here we made use of Body mass index (BMI), the mini-mental state examination (MMSE) and a 12-lead electrocardiogram. Body weight was measured without shoes to the nearest 0.1 kg. BMI was calculated by dividing weight in kilograms by the squared height in meters (kg/m2). The 11-question Dutch version of the MMSE[6] was performed by a certified research nurse. The clinically validated cut-off score <24 points[7] was defined as cognitive impairment. A 12-lead electrocardiogram was performed at baseline and evaluated for a history of myocardial infarction as reported elsewhere.[8]

Linkage with additional databases

Data on incident diagnosis of hematological malignancy were retrieved by linkage to the nationwide network and registry of histo- and cytopathology in the Netherlands (PALGA). This databank includes reports from all pathology laboratories in the Netherlands and has nationwide coverage since 1991.[9] All Lifelines participants were linked at high certainly level by using pseudonyms based on the first 8 characters of their last name, date of birth, sex, and initials (missing values allowed) and the 4 digits of the postal code. Histopathology reports were collected until October 2019. Causes of death were obtained from the national death statistic registry (Dutch Central Bureau of Statistics), with data available until December 2019. The primary cause of death was classified based on ICD-10 codes for death from hematological malignancies (C81-C96 and D45-D47), death from other malignancies (Cx, except C81-C96), death from cardiovascular diseases (Ix), and death from respiratory disorders (Jx). Calculations for cause of death analyses are based on calculations by the authors using non-public microdata from Statistics Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

Supplemental Table 1 - 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 en 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 en 9

Supplemental Table 2 - Peripheral blood counts and baseline characteristics

Characteristics and peripheral blood counts are given for all individuals with next-generation sequencing data available (n=621). First, individuals without clonal hematopoiesis (CH) are compared with individuals having CH. Second, subgroups for CH were evaluated: individuals carrying a somatic variant in DNMTA and/or TET2 exclusively (isolated DT) and individuals with any other mutational spectrum (others).

	Absence of CH (reference)	Presence of CH	P	Isolated DT	P	Others	P	N
	n=239	n=382		n=264		n=118		
Age (years)	81.0 [80.0;83.0]	82.0 [80.0;85.0]	.01	82.0 [81.0;85.0]	.009	82.0 [80.0;84.0]	.23	621
Male sex	119 (49.8%)	195 (51.0%)	.82	123 (46.6%)	.53	72 (61.0%)	.06	621
Total WBC count (10^9/L)	6.10 [5.10;7.00]	6.10 [5.00;7.10]	.85	6.00 [5.00;7.00]	.55	6.20 [5.05;7.20]	.53	616
Neutrophils (10^9/L)	3.36 [2.69;3.99]	3.37 [2.73;4.17]	.62	3.36 [2.72;4.04]	.90	3.44 [2.75;4.45]	.33	605
Basophils (10^9/L)	0.03 [0.02;0.04]	0.03 [0.02;0.04]	.88	0.03 [0.02;0.04]	.67	0.03 [0.02;0.04]	.68	605
Eosinophils (10^9/L)	0.17 [0.10;0.26]	0.18 [0.11;0.26]	.53	0.18 [0.12;0.25]	.40	0.17 [0.10;0.27]	.99	605
Monocytes (10^9/L)	0.55 [0.45;0.65]	0.55 [0.46;0.67]	.39	0.54 [0.45;0.65]	.88	0.58 [0.48;0.72]	.02	605
Lymphocytes (10^9/L)	1.75 [1.41;2.15]	1.68 [1.32;2.08]	.11	1.67 [1.34;2.10]	.13	1.73 [1.29;2.04]	.25	605
Hemoglobin (g/dL)	14.0 [13.4;14.7]	13.9 [13.1;14.7]	.039	13.9 [13.1;14.7]	.05	13.9 [13.1;14.8]	.14	616
Erythrocytes (10^9/L)	4.61 [4.32;4.85]	4.59 [4.30;4.85]	.91	4.62 [4.31;4.89]	.68	4.57 [4.24;4.81]	.32	616
Hematocrit (L/L)	0.42 [0.40;0.44]	0.42 [0.39;0.44]	.13	0.42 [0.39;0.44]	.23	0.42 [0.39;0.44]	.16	616
Platelets (10^9/L)	219 [192;260]	221 [190;262]	.77	219 [187;256]	.62	230 [194;280]	.11	615
Anemia	16 (6.72%)	42 (11.1%)	.09	24 (9.13%)	.41	18 (15.7%)	.01	616
Neutropenia	9 (3.80%)	23 (6.08%)	.29	14 (5.32%)	.55	9 (7.83%)	.29	615
Thrombocytopenia	6 (2.53%)	19 (5.16%)	.26	14 (5.47%)	.16	5 (4.46%)	.34	605
MCV (fL)	91.5 [89.4;94.1]	91.4 [88.3;94.1]	.22	91.3 [88.0;93.7]	.06	91.9 [88.9;95.4]	.68	616

Continuous data are presented as median [IQR] and categorical variables are presented as number (%). The following classification criteria for peripheral cytopenia have been used: anemia, <13g/dL in males and <12g/dL in females; neutropenia, <1.8x10^9/L and thrombocytopenia, <150x10^9/L. P-values are given for the comparison between each subgroup and the cohort of individuals without clonal hematopoiesis (CH). Isolated DT, clonal hematopoiesis confined to DNMT3A and/or TET2 variants; others, clonal hematopoiesis involving other mutational spectra; WBC, white blood cell; MCV, mean corpuscular volume; N, number of evaluable individuals.

Supplemental Table 3 - Regression analyses for the association between clonal hematopoiesis and COPD status

Univariable and multivariable logistic regression analysis for the association between clonal hematopoiesis (CH) and history of chronic obstructive pulmonary disease (COPD).

	Univariable		Multivariable 1		Multivariable 2		
	OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P	N
Presence of CH							621
No (ref)							239
Yes	2.18 [1.06-4.95]	.046	2.31 [1.11-5.26]	.033	2.26 [1.09-5.18]	.038	382
Age			0.90 [0.78-1.01]	.105	0.90 [0.78-1.02]	.122	621
Sex							621
Female (ref)							307
Male			1.23 [0.64-2.39]	.532	1.31 [0.60-2.84]	.490	314
Smoking status							610
Never smoker (ref)							250
Ex-smoker					1.10 [0.48-2.51]	.823	329
Current smoker					1.52 [0.32-5.32]	.545	31

CH, clonal hematopoiesis; OR, odds ratio; CI, confidence interval; N, number of evaluable individuals.

Supplemental Table 4 - Self-reported history of cancer and incident hematological malignancies

(A) Prevalent and self-reported diagnoses of cancer within the Lifelines cohort ≥80 years. The mutational spectrum of individuals with a prevalent (B) as well as incident (C) diagnosis of hematological malignancy is shown. Data on hematological malignancies are retrieved by linkage to the Dutch Nationwide Network and Registry of Histo- and Cytopathology (PALGA, Supplemental Methods).

A

Self-reported history of cancer	
	N
Gastrointestinal	19
Breast	18
Prostate	21
Urological	15
Others	15

B

ID	Mutational spectrum (mutated gene and VAF)	Prevalent diagnosis of hematological neoplasm	
1	Absence of CH	Lymphoid	B-cell lymphoma
2	DNMT3A# 17%	Lymphoid	B-cell lymphoma
3	CBL 3.2%	Lymphoid	B-cell lymphoma

# indicates multiple mutations in the same gene, for which the highest variant allele frequency (VAF) is given. CH, clonal hematopoiesis

C

ID	Mutational spectrum (mutated gene and VAF)	Incident diagnosis of hematological neoplasm	
1	MPL 42%	Myeloid	Myeloproliferative neoplasm
2	TP53 26%	Lymphoid	B-cell lymphoma
3	JAK2 1%, SF3B1 1.4%	Myeloid	Myeloproliferative neoplasm
4	ASXL1 43%, CALR 6%, MPL 37%	Lymphoid	B-cell lymphoma
		Myeloid	Myeloproliferative neoplasm
5	TET2 1.2%	Lymphoid	B-cell lymphoma
6	DNMT3A# 1.5%	Lymphoid	Plasma cell dyscrasia

# indicates multiple mutations in the same gene, for which the highest variant allele frequency (VAF) is given.

Supplemental Table 5 - Regression analyses for the association between clonal hematopoiesis and cause-specific death

Subdistribution hazard ratios from competing risk regression as well as cause-specific hazard ratios from Cox proportional hazard regression are shown. Cause of death was coded according to the International Classification of Diseases, tenth revision (ICD-10). Classification was performed as follows: death from cardiovascular diseases (ICD-10 codes Ix); death from hematological malignancies (ICD-10 codes C81-C96 and D45-D47); death from solid cancers (ICD-10 codes Cx, except for C81-C96); death from respiratory disorders (ICD-10 codes Jx). Results are based on calculations by the authors using non-public microdata from Statistics Netherlands. Under certain conditions, these microdata are accessible for statistical and scientific research. For further information: microdata@cbs.nl.

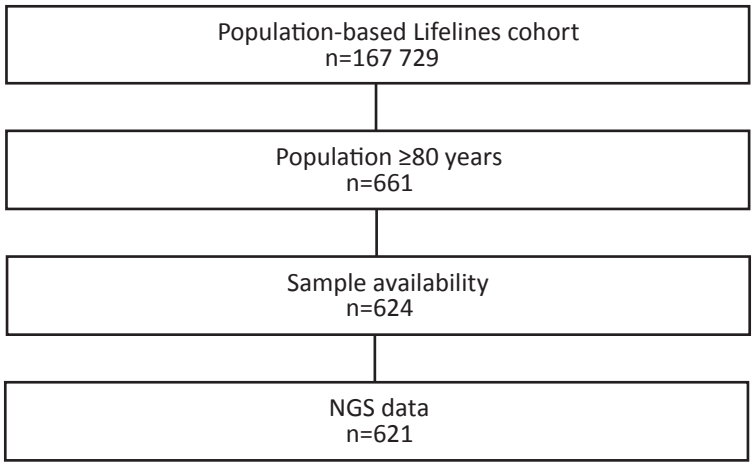
	Univariable HR (95% CI)	P	Multivariable* HR (95% CI)	P
Cardiovascular diseases				
Cause-specific hazard (Cox regression)	0.90 (0.58-1.38)	.615	0.78 (0.51-1.20)	.263
Subdistribution hazard (Fine and Gray regression)	0.87 (0.57-1.33)	.530	0.79 (0.52-1.22)	.290
Solid malignancies				
Cause-specific hazard (Cox regression)	1.21 (0.69-2.11)	.503	1.12 (0.64-1.96)	.696
Subdistribution hazard (Fine and Gray regression)	1.19 (0.69-2.08)	.530	1.13 (0.65-1.99)	.660
Respiratory disorders				
Cause-specific hazard (Cox regression)	1.49 (0.65-3.40)	.344	1.33 (0.58-3.05)	.504
Subdistribution hazard (Fine and Gray regression)	1.48 (0.65-3.38)	.350	1.39 (0.61-3.18)	.430

HR, hazard ratio; CI, confidence interval. *Age and sex were used as covariables in multivariable regression models. The absence of clonal hematopoiesis was used as the reference.

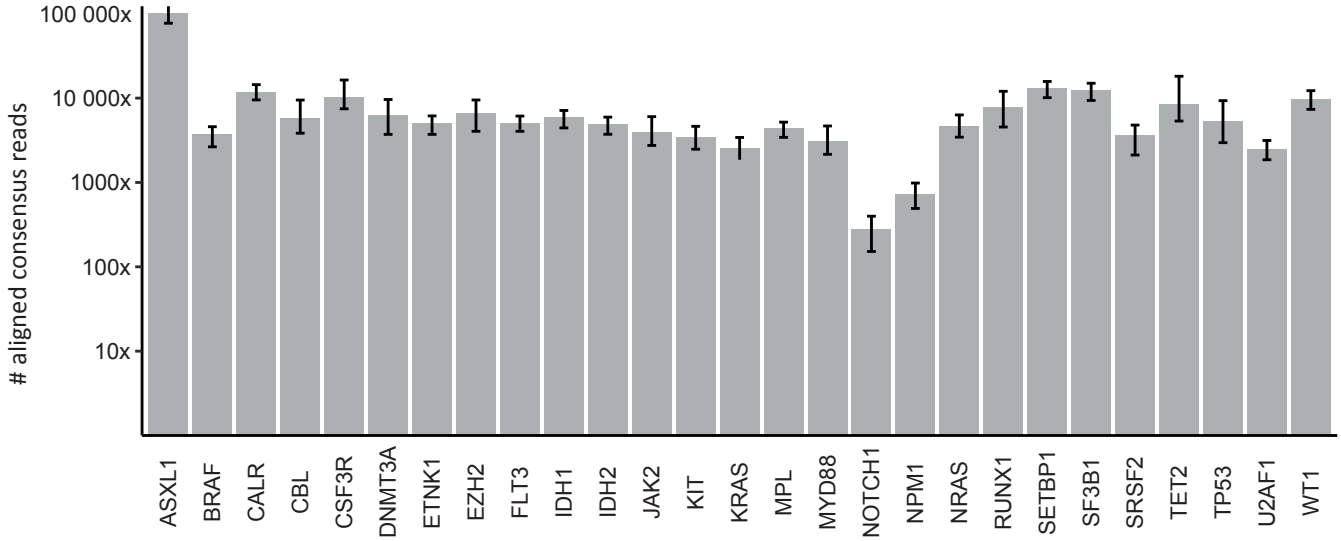
Supplemental Figure 1 - Methods

(A) Flowchart for the overview of all individuals included in the study and sequencing cohort. (B-C) Number of aligned consensus reads and the raw number of aligned reads for all genes included in the sequencing panel. Columns and error bars indicate median and interquartile range respectively.

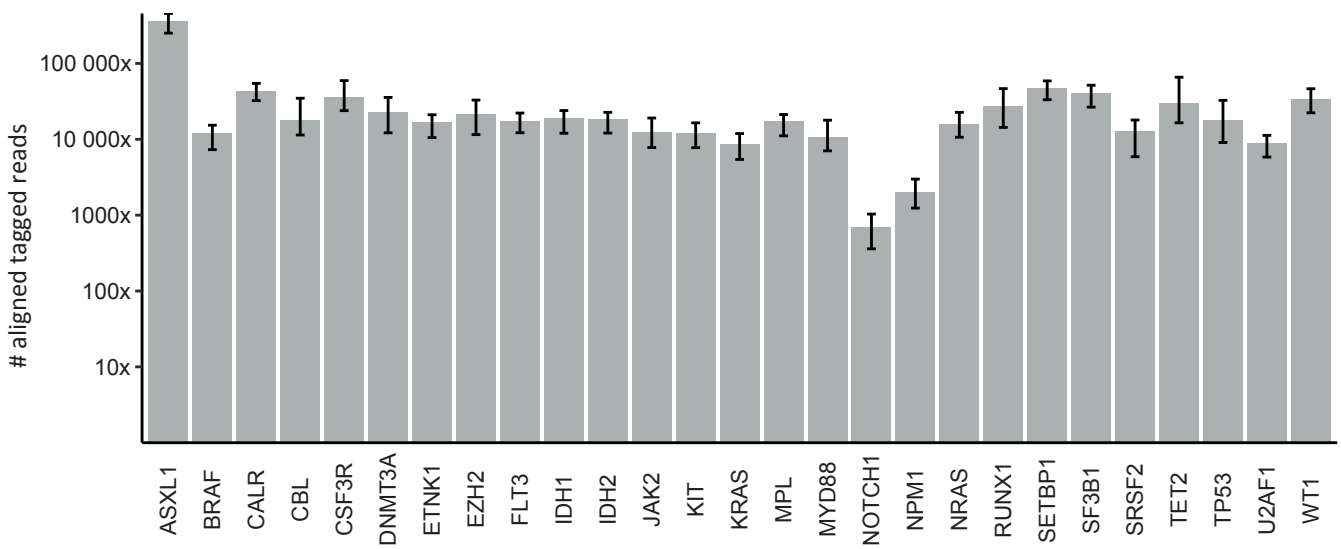
A



B

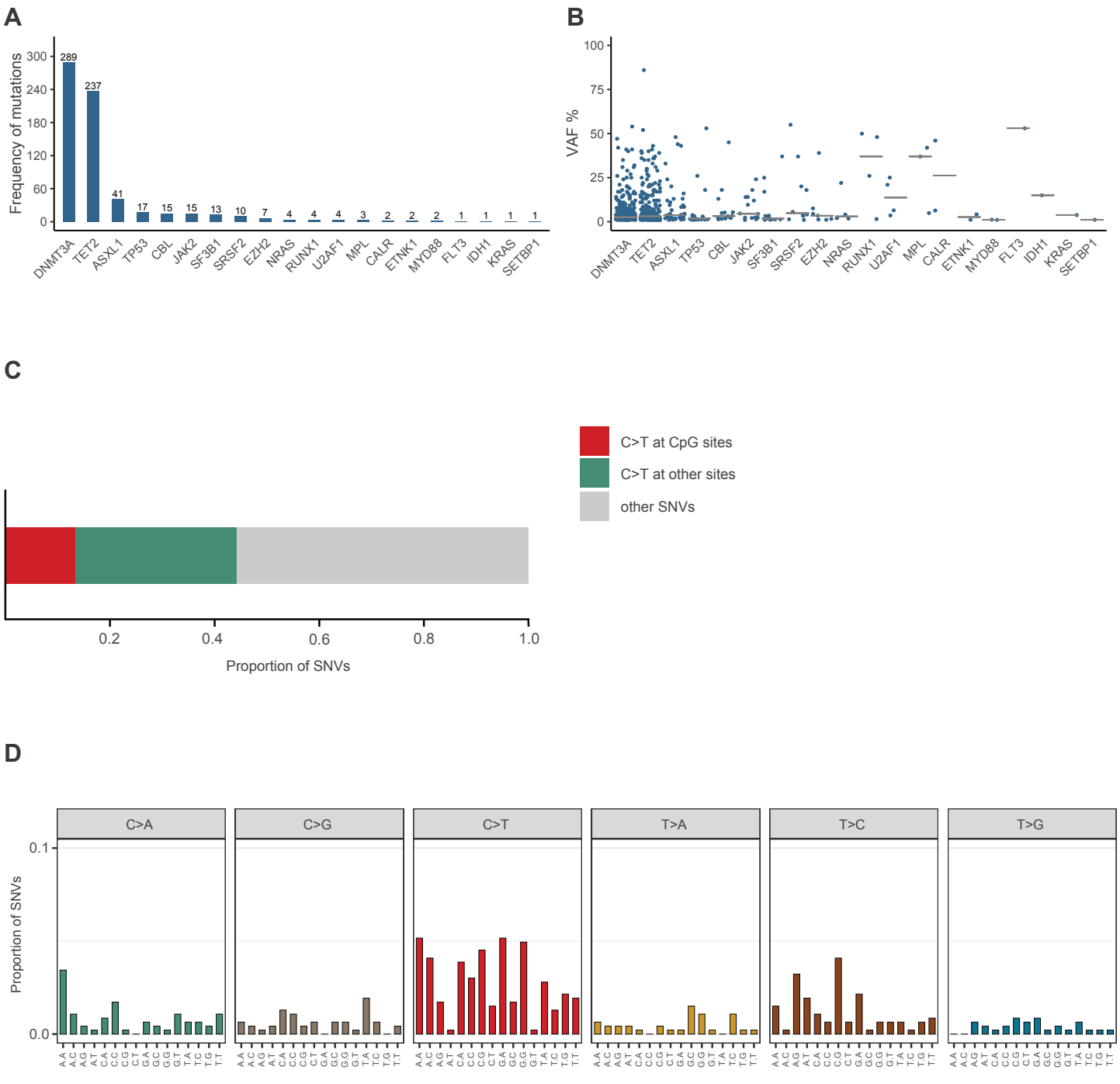


C



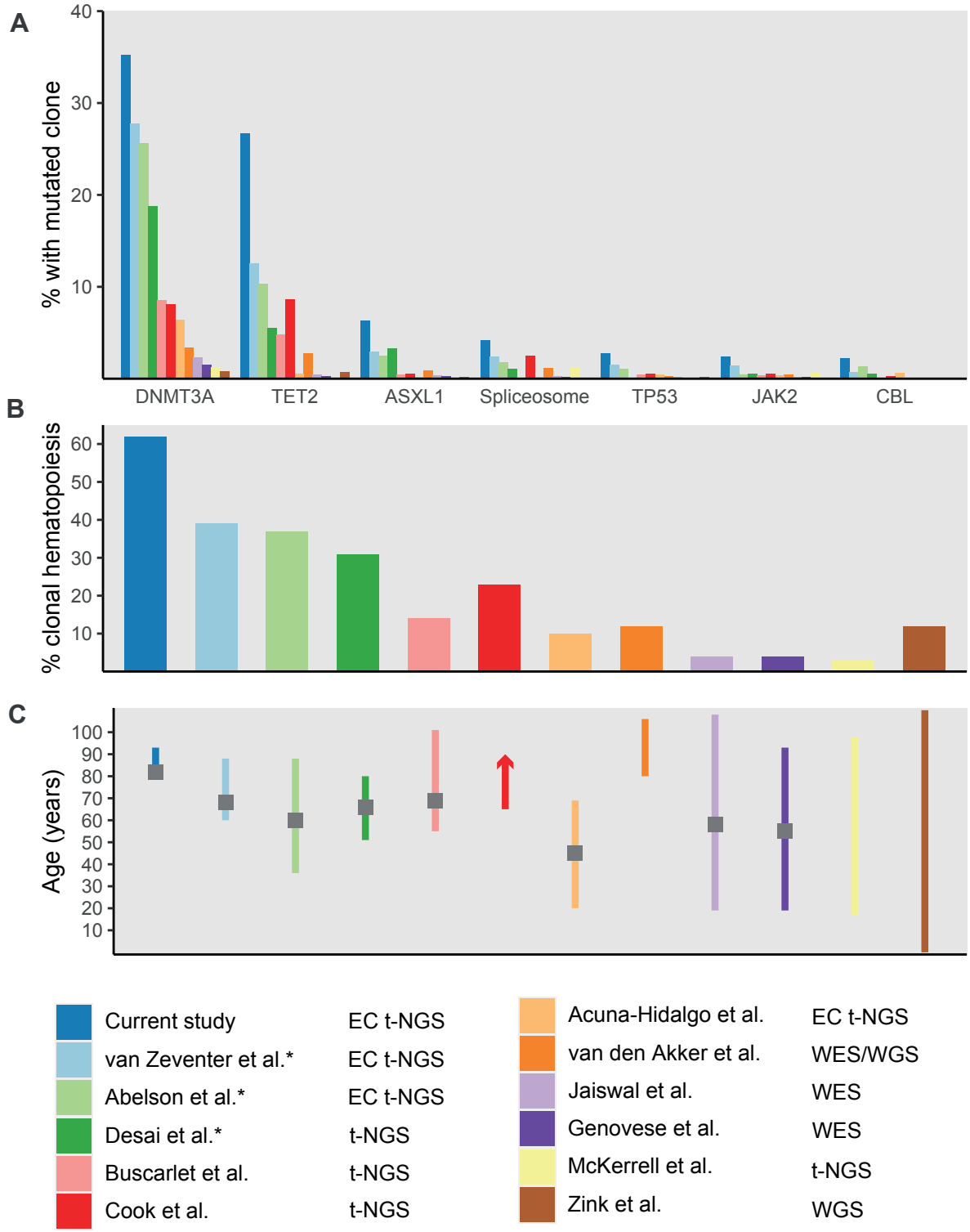
Supplemental Figure 2 - Overview of all variants detected

(A) Number of somatic variants detected in each gene in the entire cohort of sequenced individuals (n=621). (B) Variant allele frequencies of detected variants per gene in the entire cohort of sequenced individuals (n=621). The horizontal line represents the median. (C) Relative contribution of C>T and other base substitutions to all detected single nucleotide variants (SNVs) (n=465). (D) Frequencies of all 96 possible substitution patterns for SNVs in their trinucleotide context (ie. the bases immediately 5' and 3' to the mutated base). The mutational signatures were determined using the R package MutationalPatterns.



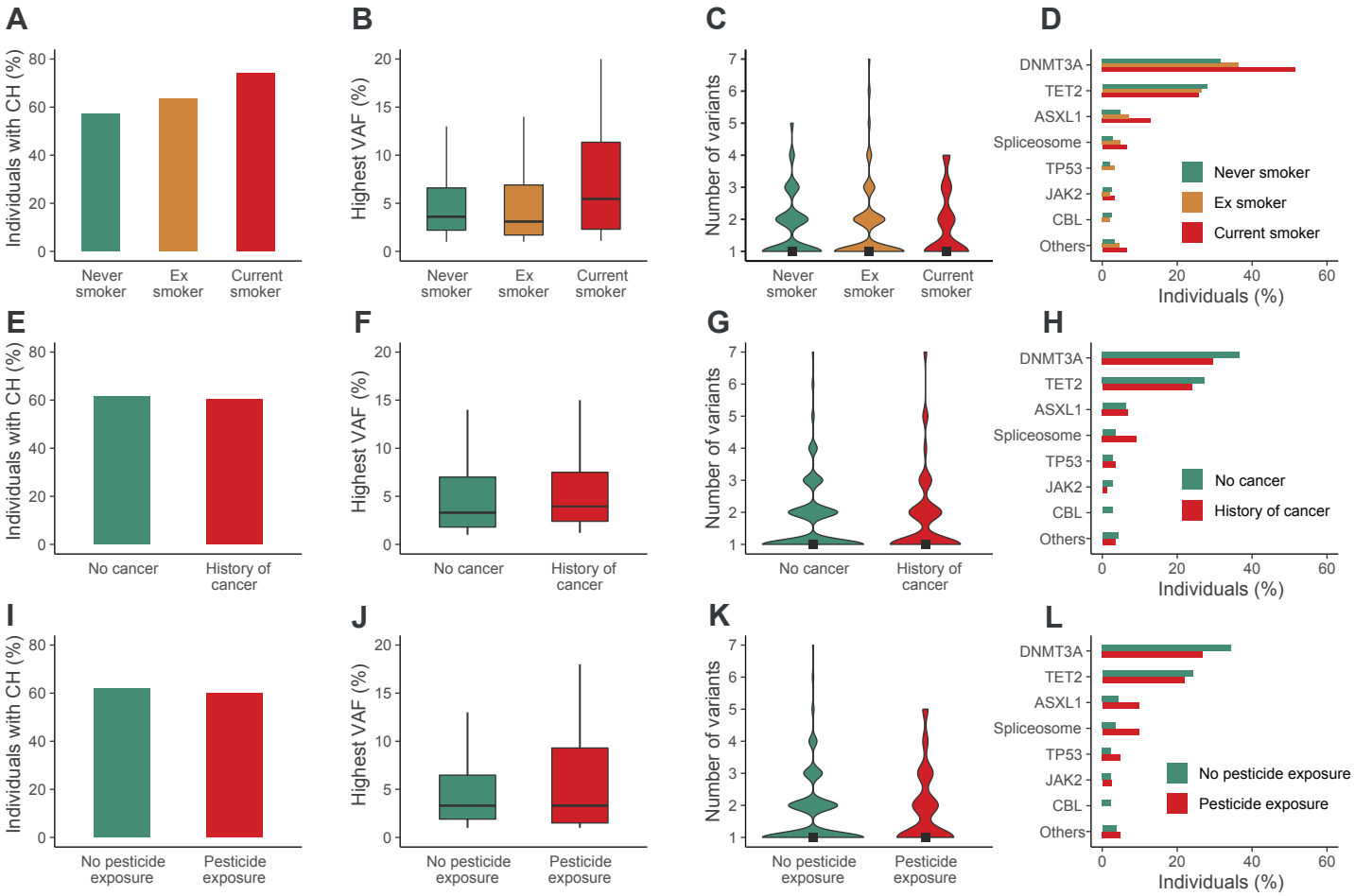
Supplemental Figure 3 - Comparison of mutational frequencies with previous published cohorts

Apart from age of included participants, studies differ considerably in sequencing sensitivity and variant calling criteria. This hampers a direct comparison with the current study cohort, with the exception of the study by van Zeventer et al. that used identical sequencing methodology. We aimed to summarize the mutational frequencies and characteristics of previous studies[10-20], in comparison to the current cohort. The proportion of individuals with a mutation in recurrently mutated genes (A) and presence of clonal hematopoiesis (B) is shown. In C, the age distribution of participants is show. Minimum, maximum and mean or median age are displayed, when reported in the original article. Sequencing methodologies are displayed in the table below. *From these studies, only control cases were included in this figure. The category of spliceosome variants includes SF3B1, SRSF2, and U2AF1. t-NGS, targeted next-generation sequencing; WES, whole exome sequencing; WGS, whole genome sequencing, EC, error-corrected.



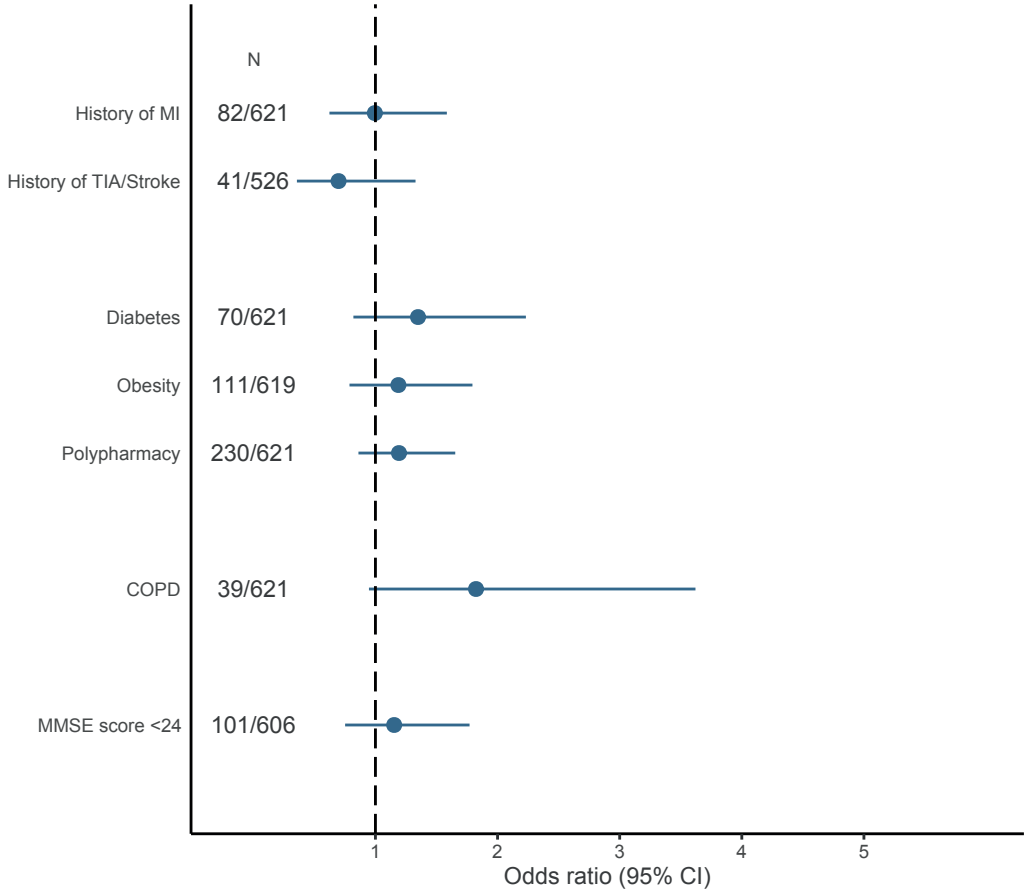
Supplemental Figure 4 - Mutational spectrum according to different estimated DNA damaging toxicities

Smoking status classified into never (n=250), ex (n=329) and current smokers (n=31) for A-D. For E-H, individuals were categorized based on self-reported history of cancer (n=88/615). I-L show job related pesticide exposure classified into low-risk (n=441) and medium to high-risk of exposure (n=35). (A,E,I) The proportion of individuals with clonal hematopoiesis (CH). (B,F,J) Highest observed VAF per individual, for individuals with detectable CH. Boxplots show the median and 25th to 75th percentile, with whiskers extending to the highest and lowest 5th percentile. (C,G,K) Violin plots showing the number of somatic variants per individual, with rectangles indicating the median. (D,H,L) The proportion of individuals carrying recurrent (>10x) gene mutations. The category of spliceosome variants includes SF3B1, SRSF2, and U2AF1. VAF, variant allele frequency.



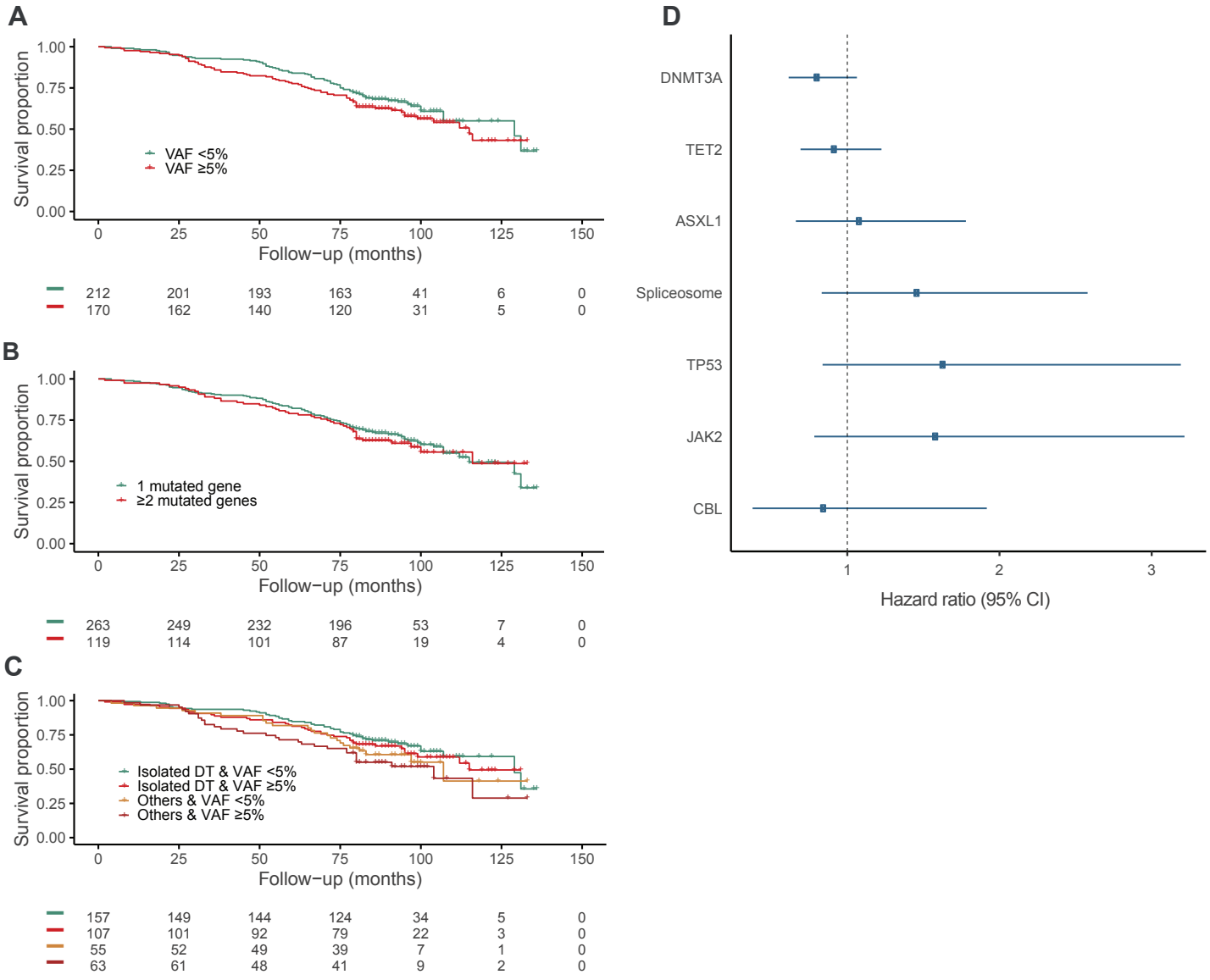
Supplemental Figure 5 - Comorbid profile for VAF ≥2%

Forest plots show odds ratios for the association between clonal hematopoiesis (CH) at a VAF ≥2% and prevalent age-related comorbidities, derived from univariable logistic regression. Circles indicate the odds ratio, with horizontal lines corresponding to the 95% confidence interval (CI). Polypharmacy was defined as ≥5 medications. Obesity was defined as BMI ≥30 kg/m2. MI, myocardial infarction; TIA, transient ischemic attack; COPD, chronic obstructive pulmonary disease; MMSE, mini mental state examination; N, number of individuals with respective medical history and total number of evaluable individuals; VAF, variant allele frequency.



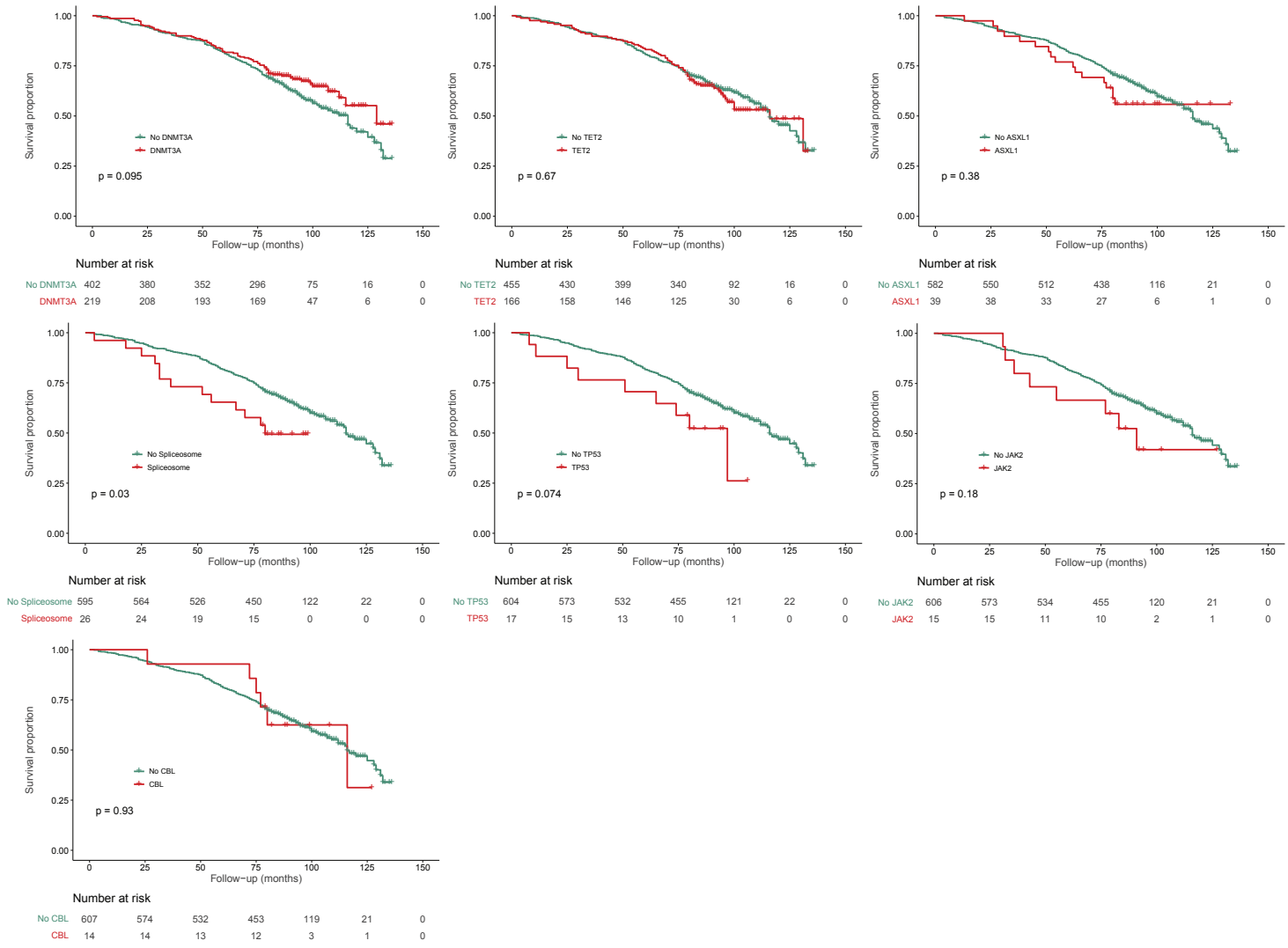
Supplemental Figure 6 - Survival analysis according to clonal hematopoiesis spectrum (I)

(A-C) Kaplan-Meier curves for overall survival (OS) stratified into subgroups for individuals with clonal hematopoiesis (CH). (A) Individuals with a highest variant allele frequency (VAF) of <5% versus ≥5%. (B) Individuals carrying 1 mutated gene versus individuals carrying ≥2 mutated genes. (C) Individuals carrying somatic variants in DNMTA and/or TET2 exclusively (isolated DT) versus other mutational spectra (others), according to their highest VAF (<5% or ≥5%). (D) Forest plot displaying hazard ratios from multivariable Cox proportional hazards regression (with age and sex as covariates), according to the presence of recurrent (>10x) gene mutations.



Supplemental Figure 7 - Survival analysis according to clonal hematopoiesis spectrum (II)

Kaplan-Meier curves for visualization of overall survival (OS) for all gene mutations detected ≥10x. Individuals were classified according to presence or absence of a mutation in the respective gene. The category of spliceosome variants includes SF3B1, SRSF2, and U2AF1.



References

1. World Health Organization. The Anatomical Therapeutic Chemical (ATC) classification 2008. http://www.whocc.no/atc_ddd_index. Accessed 28 September 2020.
2. Sediq R, van der Schans J, Dotinga A, et al. Concordance assessment of self-reported medication use in the Netherlands three-generation Lifelines Cohort study with the pharmacy database iaDB.nl: The PharmLines initiative. *Clinical Epidemiology*. 2018;10:981-989.
3. Faruque MO, Vonk JM, Bültmann U, Boezen HM. Airborne occupational exposures and inflammatory biomarkers in the Lifelines cohort study. *Occup Environ Med*. Epub ahead of print: 07 August 2020.
4. Jones R and Elias P. CASCOT: computer-assisted structured coding tool. *Coventry, UK: Institute for Employment Research, University of Warwick*. 2004.
5. International Labour Organization. The revised international standard classification of occupations (ISCO-88). Geneva: International Labour Organization; 1990.
6. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198.
7. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA*. 1993;269(18):2386-2391.
8. van der Ende MY, Hartman M, Schurer RAJ, et al. Prevalence of electrocardiographic unrecognized myocardial infarction and its association with mortality. *International journal of cardiology*. 2017;243:34-39.
9. Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol*. 2007;29(1):19-24.
10. van Zeventer IA, de Graaf AO, Wouters HJ, et al. Mutational spectrum and dynamics of clonal hematopoiesis in anemia of older individuals. *Blood*. 2020;135(14):1161-1170.
11. Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature*. 2018;559(7714):400-404.
12. 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.
13. 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(6):753-762.
14. Cook EK, Izukawa T, Young S, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood Adv*. 2019;3(16):2482-2486.
15. Acuna-Hidalgo R, Sengul H, Steehouwer M, et al. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *American Journal of Human Genetics*. 2017;101(1):50-64.
16. van den Akker EB, Pitts SJ, Deelen J, et al. Uncompromised 10-year survival of oldest old carrying somatic mutations in DNMT3A and TET2. *Blood*. 2016;127(11):1512-1515.
17. 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.
18. 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.
19. McKerrell T, Park N, Moreno T, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep*. 2015;10(8):1239-1245.
20. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*. 2017;130(6):742-752.