Supplementary Material to:

CLONAL HEMATOPOIESIS IN ELDERLY TWINS: CONCORDANCE, DISCORDANCE AND MORTALITY

Next-Generation Sequencing panel

We designed a TruSeq Custom Amplicon panel (Illumina, San Diego, CA, USA) with an amplicon length of 250 base pairs (bp) covering more than 95% of known mutations identified in publications on clonal hemaptopoiesis[1–6].Library preparation was performed according to the manufacturer's protocol, including pre-PCR addition of random unique molecular identifier (UMI) nucleotide strands. Genes and chromosol coordinates are listed below:

Gene	**Chr**	**hg19 coordinates (start stop)**	
ASXL1	20	(310 25234	310 21081)
ASXL2	2	(25964888	26101101)
DNMT3A	2	(25457124	25536945)
IDH1	2	(209113424	209112994)
IDH2	15	(90631736	90632009)
PPM1D	17	(58740350	58740919)
RAD21	8	(117859733	117866713)
SF3B1	2	(198266563	198267800)
SRSF2	17	(74732237	74733248)
TET2	4	(106155075	106197718)
TP53	17	(7572833	7579961)
JAK2	9	(5073674	5073808)
GNB1	1	Full coding sequence	
GNAS	20	(57415162	57415896)
	20	(57478578	57480535)
	20	(57484212	57485886)
ETV6	12	(11803056	12043986)
CREBBP	16	(3786120	3786810)
	16	(3781300	3781802)
NRAS	1	(115256421	115256599)
	1	(115258671	115258781)
KRAS	12	(25380168	25380346)
	12	(25398208	25398318)
CBL	11	Full coding sequence	
BRCC3	X	(154299797	154348431)
BCOR	X	(39911356	39937188)

Sequence analysis

Each read of the 150bp paired-end raw FASTQ included a 6bp UMI tag. Using the extract program in the UMI-tools software (v0.5.0)[7], the UMI 6mers were removed from the read sequence, combined with the corresponding mate-pair into a 12bp identifier and moved to the read name files of each mate. Corresponding FASTQ files where aligned to the reference hg19 genome with the bwa-mem implementation of the Burrows–Wheeler algorithm[8]. PCR duplicates were removed based on the UMI information using the dedup program in UMI-tools. QC and visual inspection of the alignments and amplicon coverage were assessed with the R package TarSeqQC[9]. Identification of the variants on the filtered BAM files was performed with two variant callers; Freebayes[10] and VarDict[11]. All the variants were annotated using snpEff (v4.2)[12] for gene name, amino acid change, rare variants (ExAC, dbSNP), cancer mutation records (COSMIC) and effect predictor (VEP).

Variant calling

Curation of mutations was performed blinded to clinical data. A manual assessment of all variants was performed, which included visualization in Integrated Genome Viewer (Broad Institute, Cambridge, MA). Exclusion criteria were:

- Coverage below 200x
- Alternative variant with read count below 5
- Variant allele frequency below 2%
- Area with excessive local noise
- Strand biased call explained by sequencing error
- Synonymous changes
- Known sequencing artefact based on >1000 samples analyzed by the same panel
- Known as a common polymorphism in the normal population (>1% frequency in ExAC, TOPMED or 1000G)
- Annotated as benign polymorphism in ClinVar
- Known rare variants and missenses mutations with a variant allele frequency of 40-60%, and not previously described as pathogenic. In MZ twins the presence of rare variants in both twins confirmed a germline origin.

Additional assessment criteria supporting inclusion of variant

- Variants identified in multiple UMI families[13]
- Known COSMIC reference

For each individual gene we used the following filtering:

Gene	Chromosome	Accession	Reported mutations used for variant calling
ASXL1	20	NM_015338	Frameshift/nonsense/splice-site in exon 11-12
ASXL2	2	NM_018263	Frameshift/nonsense/splice-site
DNMT3A	2	NM_022552	Frameshift/nonsense/splice-site; Missense in p.292-350 / p.482-614/ p.634-912
IDH1	2	NM_005896	Missense at R132
IDH2	15	NM_002168	Missense at R140 and R172

PPM1D	17	NM_003620	Frameshift/nonsense in exon 5/6
RAD21	8	NM_006265	Frameshift/nonsense/splice-site
SF3B1	2	NM_012433	Missense in terminal HEAT domains (p.529-1201)
SRSF2	17	NM_003016	Missense/deletion involving p.P95
TET2	4	NM_001127208	TET2 Frameshift/nonsense/splice-site; Missense in conserved domains (p.1104-1481 and p.1843-2002)
TP53	17	NM_001126112	Frameshift/nonsense/splice-site; Missense in DNA-binding domain (p.95-288); Missense at P72 / R337
JAK2	9	NM_004972	JAK2 V617F and Missense/indel in aa range p.536-547
GNB1	1	NM_002074	Missense at K57 / I80
GNAS	20	NM_016592	GNAS Missense at R201
ETV6	12	NM_001987	Frameshift/nonsense/splice-site
CREBBP	16	NM_004380	Frameshift/nonsense/splice-site
NRAS	1	NM_002524	Missense at G12 / G13 / Q61
KRAS	12	NM_033360	Missense at G12 / G13 / Q61 / A146
CBL	11	NM_005188	Missense in Linker/RING finger domains p.345-434
BRCC3	X	NM_024332	Frameshift/nonsense/splice-site
BCOR	X	NM_001123385	Frameshift/nonsense/splice-site

Hematological cancer and cytopenia ICD-10 codes.

All diagnosis from the cohort was identified and we created two groups of diagnoses

Hematological cancers including the following ICD-10 diagnosis.

DD45 - Polycythemia vera
DD46 - Myelodysplastic syndromes
DD47 - Mast cell neoplasms of uncertain behavior
DC81 - Hodgkin lymphoma
DC82 - Follicular lymphoma
DC83 - Non-follicular lymphoma
DC84 - Mature T/NK-cell lymphomas
DC85 - Other specified and unspecified types of non-Hodgkin lymphoma
DC86 - Other specified types of T/NK-cell lymphoma
DC88 - Malignant immunoproliferative diseases and certain other B-cell lymphomas
DC90 - Multiple myeloma and malignant plasma cell neoplasms
DC91 - Lymphoid leukemia
DC92 - Myeloid leukemia
DC93 - Monocytic leukemia
DC94 - Other leukemias of specified cell type
DC95 - Leukemia of unspecified cell type
DC96 - Other and unspecified malignant neoplasms of lymphoid, hematopoietic and related tissue

The group of unspecified cytopenia included the following diagnoses

DD619A – Pancytopenia, unspecified DD693 - Immune thrombocytopenic purpura DD696 - Thrombocytopenia, unspecified DD613 - Idiopathic aplastic anemia DD63 - Anemia in chronic diseases classified elsewhere DD618 - Other specified aplastic anemias and other bone marrow failure syndromes DD649 - Anemia, unspecified DD694 - Other primary thrombocytopenia DD70 – Neutropenia

Supplementary Table 1a and b Cox regression model including the VAF grouped as maximum VAF<10% vs >=10%.

A

Variable	HR	95% CI	P
VAF < 10%	1.16	0.94-1.43	0.115
VAF >=10%	1.120	0.93-1.54	0.205
Age (years)	1.11	1.08-1.13	<0.001
Sex (male)	1.15	0.91-1.45	0.252
Smoking (1st tertile)	1.02	0.81-1.29	0.852
Smoking (2nd tertile)	1.22	0.96-1.55	0.101
Smoking (3rd tertile)	1.70	1.23-2.35	0.001

***VAF = Variant allele frequency**

B

Variable	HR	95% CI	P
VAF < 10%	1.12	0.91-1.38	0.299
VAF >=10%	1.16	0.91-1.49	0.267
Sex (male)	1.14	0.90-1.44	0.267
Smoking (1st tertile)	1.03	0.82-1.29	0.787
Smoking (2nd tertile)	1.26	1.00-1.59	0.053
Smoking (3rd tertile)	1.73	1.26-2.37	0.001

***VAF = Variant allele frequency**

In table **A** time since blood sample is used as the underlaying time scale whereas age is used as the underlying the time scale in **B**, and almost similar results were found.

Supplementary Table 2

Risk of developing hematological cancer, competing risk analysis according to CHIP mutational status

Variable	HR	95% CI	P
Mutation (yes/no)	**1.80**	**0.87-3.75**	**0.115**
Age (years)	**0.94**	**0.86-1.04**	**0.234**
Sex (male)	**0.88**	**0.42-1.82**	**0.722**
Smoking (1st tertile)	**0.88**	**0.32-2.43**	**0.808**
Smoking (2nd tertile)	**0.123**	**0.016-0.95**	**0.045**
Smoking (3rd tertile)	**1.04**	**0.44-2.48**	**0.932**

The risk of developing a hematological cancer (as defined above), was not associated with clonal hematopoiesis after adjusting for age, sex and tobacco consumption. A total of 27 twins developed a hematological cancer during follow up. Of those 19 were classified as a lymphoid neoplasm and eight as a myeloid neoplasm.

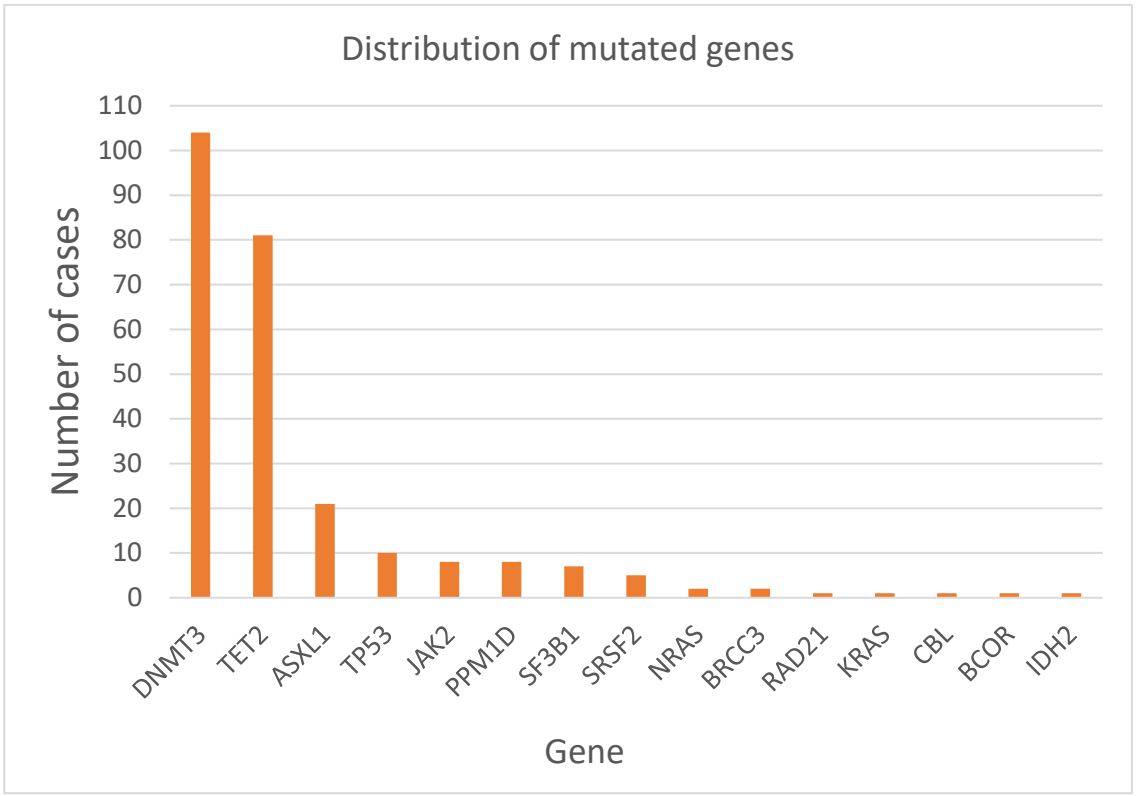
Supplementary Table 3

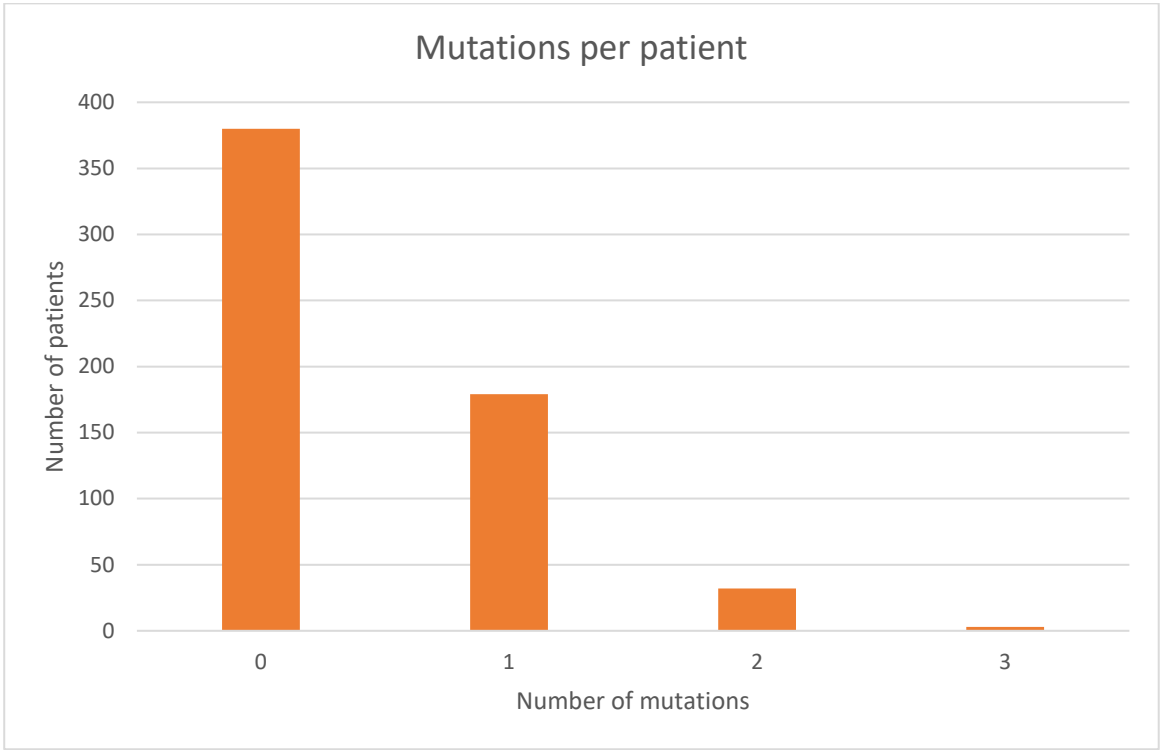
Risk of developing unspecified cytopenia, competing risk analysis according CHIP mutational status

Variable	HR	95% CI	P
Mutation (yes/no)	**1.59**	**0.99-2.54**	**0.052**
Age (years)	**0.96**	**0.91-1.02**	**0.158**
Sex (male)	**1.24**	**0.75-2.04**	**0.403**
Smoking (1st tertile)	**0.92**	**0.48-1.77**	**0.802**
Smoking (2nd tertile)	**0.63**	**0.31-1.30**	**0.211**
Smoking (3rd tertile)	**1.01**	**0.54-1.90**	**0.980**

A total of 74 twins developed an unspecified cytopenia during follow up, and as the only variable clonal hematopoiesis had a borderline significant association with the development of unspecified cytopenia after adjusting for age, sex and tobacco consumption.

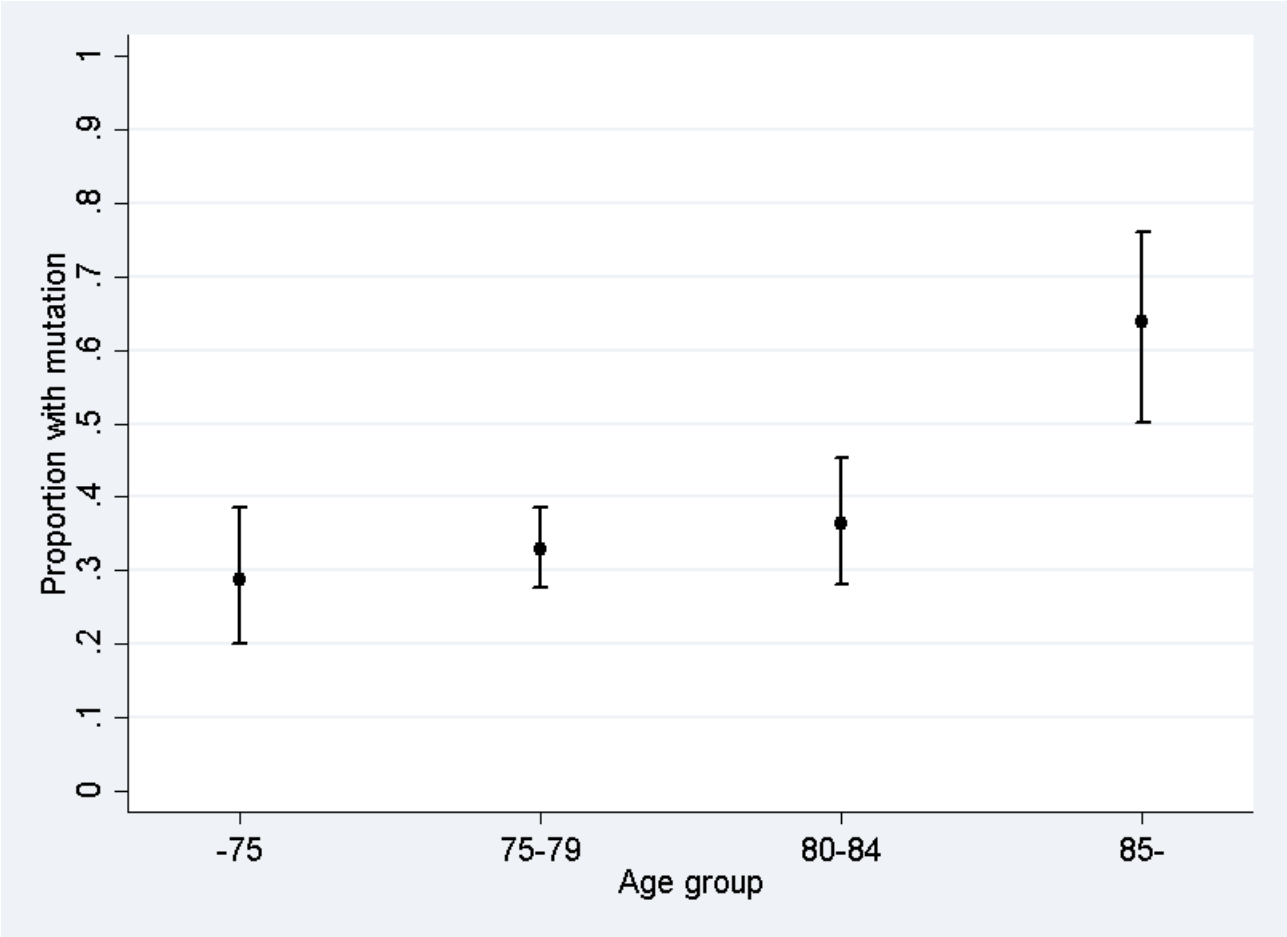
Supplementary Figure 1a and b





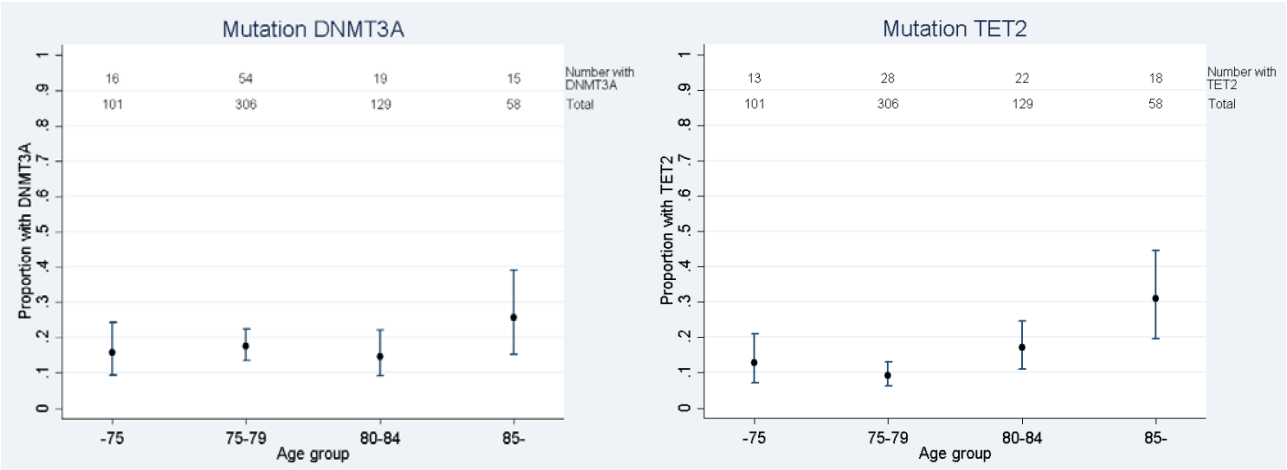
1a cases with a mutation in the affected gene and 1b number of different genes mutated per twin.

Supplementary Figure 2



Proportion of individuals with a mutation according to age group. There were 101 twins between 73-75 years, 306 between 75-79 years, 129 between 80-85 years and 58 were 85 years or above at the time of inclusion in the study in 1997.

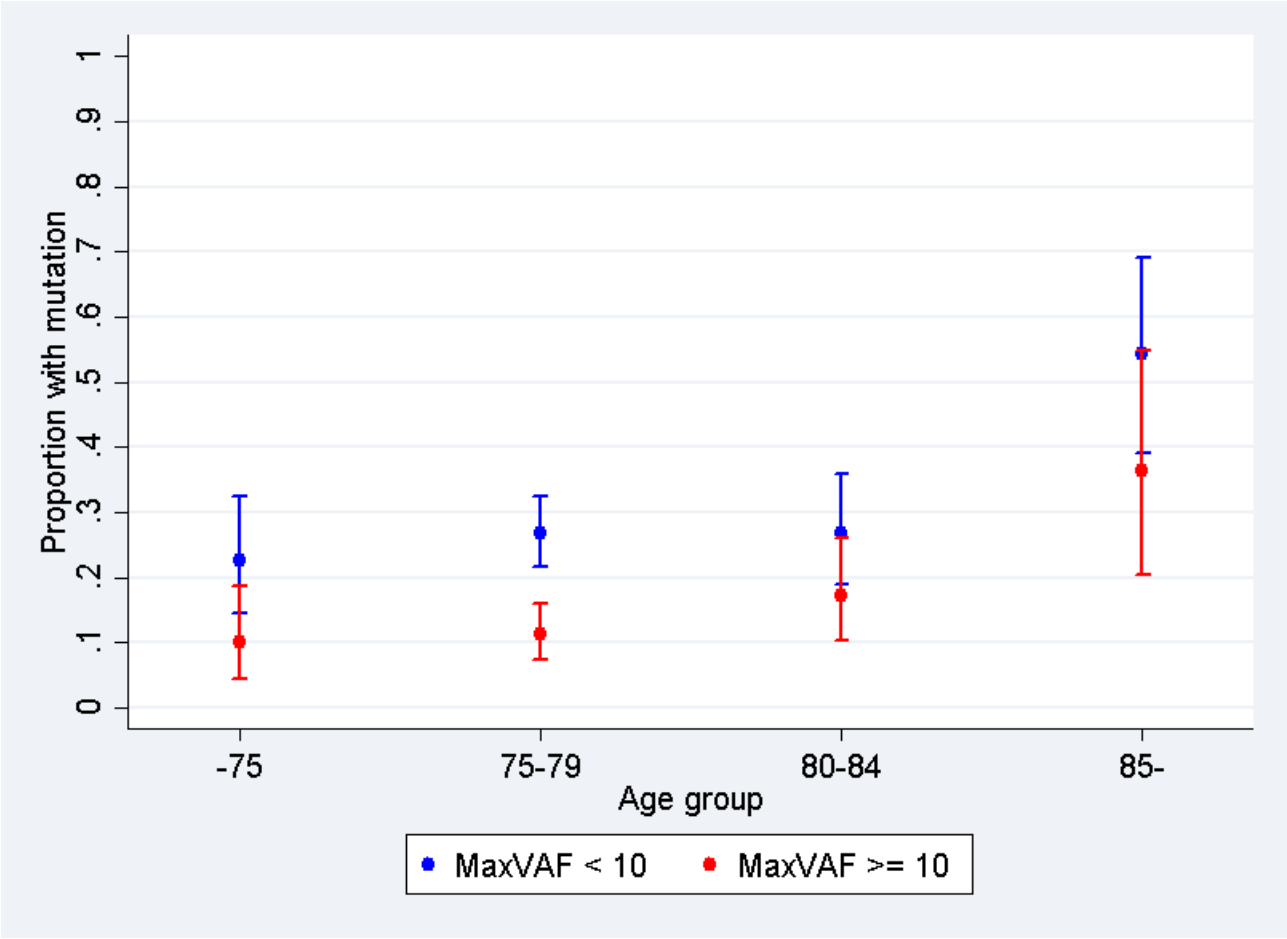
Supplementary Figure 3a and b.



A B

The distribution of mutations in (A) *DNMT3A* and (B) *TET2* according to age group. The total number and number affected in each age group can be seen above the barplot.

Supplementary figure 4



The proportion of samples grouped according to after the allele frequency of the mutations with the highest VAF within each case.

1. 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.

2. 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):2477–87.

3. Gibson CJ, Lindsley RC, Tchekmedyian V, et al. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J Clin Oncol*. 2017;35:.

4. Gillis NK, Ball M, Zhang Q, et al. Clonal haemopoiesis and therapy-related myeloid malignancies in elderly patients: a proof-of-concept, case-control study. *Lancet. Oncol.* 2017;18(1):112–121.

5. Takahashi K, Wang F, Kantarjian H, et al. Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet. Oncol.* 2017;18(1):100–111.

6. Xie M, Lu C, Wang J, et al. a n a ly s i s Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* 2014;20(12):1472–1478.

7. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 2017;27(3):491–499.

8. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.

9. Merino GA, Murua YA, Fresno C, et al. TarSeqQC: Quality control on targeted sequencing experiments in R. *Hum. Mutat.* 2017;38(5):494–502.

10. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012;

11. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 2016;44(11):e108.

12. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin).* 6(2):80–92.

13. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* 2016;7:1–7.