Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work

Marianna Rossi et al. *Clinical relevance of clonal hematopoiesis in the oldest-old population*

Table of Contents

Supplementary File_1 . Study populations and procedures

Study cohorts	Description of study populations and of statistical analyses performed in each study cohort
Study cohort#1 **"Health_&_Anemia" prospective population-based observational study (2003-2017)**	1059 oldest-old individuals (median age 83y, [80-105]) *The following analyses were carried out:* - Prevalence of CHIP - Correlation between CHIP and risk of developing myeloid neoplasms - Study of clonal evolution in 96 subjects with multiple samples available - Relationship between CHIP, coronary heart disease and other chronic inflammatory diseases - Clinical effect of CHIP in subjects affected with unexplained cytopenia
Study cohort#2 **"Monzino_80+" prospective population-based observational study (2002-2017)**	735 oldest-old individuals (median age 90y, [80-104]) *The following analyses were carried out:* - Prevalence of CHIP - Clinical effect of CHIP in subjects affected with unexplained cytopenia
Study cohort#3 **"Health_&_Anemia" prospective population-based observational study (2003-2017)**	727 individuals (median age 77y, [75-79.9]) *The following analyses were carried out:* - Correlation between CHIP and risk of developing myeloid (independent validation cohort)
Study cohort#4 **Retrospective age- and sex-matched cohort of oldest-old patients affected with myeloid neoplasms (myelodysplastic syndromes) from EuroMDS consortium (2000-2018)**	255 patients affected with myeloid neoplasm (myelodysplastic syndromes, median age 84y, [80-94]). An individual matching was performed for each oldest-old individual with Clonal Cytopenia of Unknown Significance (CCUS) from study cohort#1-2 with 5 patients affected with myelodysplastic syndrome with the same year of birth and sex (ratio 1:5) *The following analyses were carried out:* - Comparison of clinical/molecular features and outcome between oldest-old subjects CCUS to those of age- and sex-matched patients with myeloid neoplasms

Study cohort#1 - "Health & Anemia" prospective population-based observational study (2003-2018)

(Haematologica. 2009;94(1):22-28 and Haematologica. 2010;95(11):1849-1856)

"Health & Anemia" is a prospective population-based observational study of all elderly residents in the municipality of Biella, Piedmont, a town in the north-west of Italy, with a population of about 46,000 inhabitants. North African and East European migration is recent and for the most part composed of young people (about 25 years old at the moment of transfer), thus the elderly population of Biella is almost exclusively of Italian origin, though composed not only of local natives but also of immigrant workers mainly from the north-eastern Venetian province of Vicenza (1911-1936) and southern Italian regions (1950-1970) attracted at the time by the flourishing textile industry in Biella. The population is predominantly well off and employment is mainly in the industrial and service sectors. The age structure of the Biella elderly is similar to that of the general Italian elderly population.

Lists of residents were obtained from the registry office of the municipality. The high prevalence of dementia, cognitive impairment, functional disability, and health problems in the oldest old prompted us to separate the investigation of the younger (65-84 years old) from the older subjects. All registered individuals of 65 to 84 years old

residing in Biella at the study enrollment (May, 2003) were eligible for the study (n=10,082). Case ascertainment was made between May 2003 and April 2004. Subsequently the study was extended to all residents aged 85 years or older (n=1,526) at the study enrollment (May, 2007). Case ascertainment was made between May 2007 and July 2008. No exclusion criteria other than age and residence were used.

Elderly in nursing or residential homes were included. In consenting participants, arterial blood pressure (third reading) and heart and respiratory rates were measured and blood samples were taken by trained, registered nurses either at home or in an outpatient clinic at the elderly person's choice and, for institutionalized individuals, in nursing homes. A questionnaire was also administered by the nurses in order to ascertain habits (smoking and alcohol use), present and past diseases, hospital admissions, and interventions.

To face the potential sources of non-participation due to the initial step-limiting request of a blood sample and the poor health condition of many elderly often already recently tested, complete blood count (CBC) results together with age and sex of all elderly residents aged 65 years or older who did not or could not participate, but had a CBC done in the same period and laboratory of the epidemiological study, were obtained in an anonymous way from the central laboratory of Biella Hospital, one of the teams involved in the present research. Biella has only one hospital which is public. Its laboratory is one of the very few in the area, and most residents go there for laboratory investigations.

Venous blood samples were collected from participants in a sitting position by venipuncture. The complete blood count (CBC) was determined using a SISMEX SE-2100 electronic counter (Sysmex Corporation Kobe, Japan) by the central laboratory of Biella Hospital. When a hemoglobin concentration was below WHO reference criteria for anemia, further laboratory investigations were made: creatinine, serum folic acid, vitamin B12, iron, ferritin and transferrin, and transferrin saturation. These laboratory investigations were also assessed in an equal sample of non-anemic individuals matched for age and sex.

Anemia was defined according to the WHO criteria as a hemoglobin concentration lower than 12 g/dL in women and 13 g/dL in men. Mild grade anemia was defined as a hemoglobin concentration between 10.0 and 11.9 g/dL in women and between 10.0 and 12.9 g/dL in men. Iron deficiency anemia was considered present if the elderly had low serum iron (lower than 50 μg/dL in women and 60 μg/dL in men), low ferritin (lower than 15 ng/mL), low transferrin saturation rate (lower than 16%) or increased total iron binding capacity (higher than 450 μg/dL). Anemia of chronic disease was defined as low circulating iron in the presence of increased iron stores (normal or increased ferritin higher than 100 ng/mL, transferrin saturation higher than 25% and lower than 50%) and decreased total iron binding capacity (lower than 250 μg/dL). Thalassemia trait was considered when the following conditions were present: low or very low mean corpuscular volume (MCV) and mean corpuscular hemoglobin (MCH), increased red blood cell (RBC) count, normal or increased circulating iron in the presence of normal or increased iron stores. Since HbA2 levels were not determined, it was not possible to distinguish alpha from beta thalassemia. Anemia associated with folate or vitamin B12 deficiency was defined as concentrations of folate lower than 3.0 ng/mL or vitamin B12 lower than 200 pg/mL and MCV higher than 95 fL. Subjects were classified as having anemia related to chronic renal disease when affected by renal insufficiency. The classification of anemia based on the hematologic findings was supported by the clinical conditions and pharmacological therapies of the elderly. Anemias that could not be classified into any of the previous categories were considered to be of "unexplained origin".

Follow-up of patients was updated to December, 2018. Additional 344,565 laboratory tests were available during this follow-up period (including complete blood count, creatinine, serum folic acid, vitamin B12, iron, ferritin, transferrin, transferrin saturation and protein electrophoresis). Data on hospitalization and mortality (2003-2018) were available for all patients. Information on the development of myeloid neoplasms was based on: i) information provided by the International Classification of Diseases for Oncology ninth edition CM (ICD-9-CM) codes, ii) information provided by local cancer registry (www.cpo.it), and iii) revision of bone marrow biopsy reports provided by Hematology Unit of Biella Hospital. Three investigators (MR, ER and MGDP) reviewed independently all this information and a final diagnosis of myeloid neoplasms was provided by a consensus meeting. The diagnosis of coronary heart disease and chronic inflammatory diseases were based on ICD-9-CM codes. The diagnosis of solid cancer was based on information provided by the ICD-9-CM codes and information provided by local cancer registry.

Study cohort#2- "Monzino 80+" prospective population-based observational study (2002-2018)

(Alzheimers Dement. 2015;11(3):258-270)

The "Monzino 80+" study is a prospective door-to-door population-based survey among all residents 80 years or older in eight neighboring municipalities in the province of Varese, Italy. Some 114,000 people reside (2009) in this predominantly urban area. Migration from other world regions is recent and mostly involves young people, so all but six elderly people included in the study were of Italian origin.

Lists of residents were obtained from the municipal registry offices. All individuals 80 years or older residing in Castellanza, Gorla Maggiore, Gorla Minore, Marnate, Olgiate Olona, and Solbiate Olona and 85 years or older residing in Fagnano Olona on February 12, 2002 were eligible for the study. To increase the number of individuals in the extreme age groups, the survey was subsequently extended to all individuals aged 90 years or older residing in the neighboring municipality of Gallarate on January 1, 2005 and, more recently, to all those aged 100 years or older residing in the remaining municipalities of the province of Varese in 2009. In view of the low number of men aged 95 years or over, the study was further extended to include a random sample of men aged 95 to 99 years, resident in the same municipalities as the centenarians, in the first 9 months of 2010. No exclusion criteria were used other than age and residence.

At first visit, psychologists specifically trained collected information about participants' lifestyle, habits, medical history and health status from both the subject and the proxy informant who was also asked to rate the individual's everyday cognitive ability and functional disability, behavioral disturbances, as well as the stress associated with caregiving and the economic burden associated with the elderly person's health conditions. When the elderly person was not in physical or mental condition to answer the questionnaire, all the information was gathered from the proxy informant. During the same visit, testable participants were also administered a multidomain cognitive test battery. Fasting blood samples were collected by venipuncture from consenting participants in a sitting position. Blood and urine examination included routine investigations as well as laboratory tests aimed at identifying potentially reversible dementias and aggravating factors.

Follow-up of patients was updated to 2018. Laboratory tests at the time of study entry were collected. Where present, information on hospitalization and mortality and information on the development of hematological and solid cancers was collected. The diagnosis coronary heart disease and chronic inflammatory diseases were based on ICD-9-CM codes

Study cohort#3 - "Health & Anemia" prospective population-based observational study (2003-2018) – Validation cohort

Data collection on myeloid neoplasms diagnosis was less accurate in "Monzino_80+" population with respect to "Health_&_Anemia" cohort due to the lack of revision of bone marrow biopsy reports. Therefore, with the aim to validate predictive value of clinical and mutational features on the risk of developing myeloid neoplasms we analyzed a population of 727 subjects aged ≥75<80 years from "Health_&_Anemia" cohort.

Study cohort#4 – Retrospective cohort of age- and sex-matched patients affected with myeloid neoplasms (myelodysplastic syndrome) from EuroMDS consortium (clinicaltrials.gov number:NCT04174547, 2000-2018)

With the aim to compare clinical/molecular features and outcome of oldest-old subjects with Clonal Cytopenia of Unknown Significance (CCUS) from "Health & Anemia" and "Monzino 80+" cohorts to those of age-matched patients affected with myeloid neoplasms, we analyzed an additional population of patients affected with myelodysplastic syndrome from EuroMDS consortium. An individual matching was performed for each oldest-old individual with CCUS with 5 patients affected with myelodysplastic syndrome with the same year of birth and sex (ratio 1:5). Overall, 255 patients with myeloid neoplasm from EuroMDS consortium were included

In the EuroMDS database, data from 2681 patients with myelodysplastic syndromes were retrospectively recorded from 2000 to 2018. The diagnosis was classified according to WHO 2016 criteria. Complete clinical data (including whole blood count at diagnosis, % of bone marrow at diagnosis, cytogenetics at diagnosis, information on transfusion-dependency, information on disease-modifying treatments and outcome) and somatic mutations (targeted gene sequencing, including ASXL1, ATRX, BCOR, BCORL1, BRAF, CALR, CBL, CBLB, CEBPA, CSF3R, DNMT3A, ETV6, EZH2, FBXW7, FLT3, GATA2, GNAS, GNB1, IDH1, IDH2, JAK2, KDM6A (UTX), KIT, KRAS, MPL, NF1, NOTCH1, NPM1, NRAS, PHF6, PIGA, PPM1D, PRPF40B, PTPN11, RAD21, RUNX1 (AML1), SETBP1, SF3B1, SMC1A, SMC3, SRSF2, STAG2, TET2, TP53, U2AF1, WT1, ZRSR2) were available for all patients

Supplementary Table_1. Panel of sequenced genes in the whole study cohort

Gene	Pathway	NCBI ID	Position
ASXL1	chromatin & histones modifier	171023	20q11.1
ASXL2	chromatin & histones modifier	55252	2p23.3
BCOR	chromatin & histones modifier	54880	Xp11.14
BCORL1	chromatin & histones modifier	14616	Xq26.1
EZH2	chromatin & histones modifier	2146	7q35-36
KDM6A	chromatin & histones modifier	7403	Xp11.2
RAD21	cohesin complex	5885	8q24
STAG2	cohesin complex	10735	Xq25
DNMT3A	DNA methylation	1788	2p23
IDH1	DNA methylation	3417	2q33.3
IDH2	DNA methylation	3418	15q26.1
TET2	DNA methylation	54790	4q24
SF3B1	RNA splicing	23451	2q33.1
SRSF2	RNA splicing	6427	17q25.1
U2AF1	RNA splicing	7307	21q22.3
ZRSR2	RNA splicing	8233	Xp22.1
SF3A1	RNA splicing	10291	22q12.2
ATM	signalling	472	11q22.3
CALR	signalling	811	19p13.13
CBL	signalling	867	11q23.3
CSNK1A1	signalling	1452	5q32
ETNK1	signalling	8698	12p12.1
FLT3	signalling	2322	13q12
GNAS	signalling	2778	20q13.3
JAK2	signalling	3717	9p24
KIT	signalling	3815	4q12
KRAS	signalling	3845	12p12.1
MPL	signalling	4352	1p34
MYC	signalling	6359	8q24.21
MYD88	signalling	2674	3p22.2
NF1	signalling	42292	7q11.2
NRAS	signalling	4893	1p13.2
PIGA	signalling	14165	Xp22.2
PPM1D	signalling	11625	17q23.2
PTPN11	signalling	5781	12q24.1
STAT3	signalling	6774	17q21.2
CEBPA	transcription regulation	1050	19q13.1
CREBBP	transcription regulation	1387	16p13.3
CUX1	transcription regulation	1523	7q22.1
ETV6	transcription regulation	2120	12p13.2
GATA1	transcription regulation	2623	Xp11.23
GATA2	transcription regulation	2624	3q21.3
PHF6	transcription regulation	84295	Xq26.2
RUNX1	transcription regulation	861	21q22.3
SETBP1	transcription regulation	45859	18q12.3
TP53	tumor suppressor	7157	17p13.1
WT1	tumor suppressor	7490	1p13

Supplementary File_2. Methods

Mutation Screening

Using peripheral blood DNA we looked for somatic mutations in 47 CHIP-related genes (Gene list available in *Supplementary Table 1*).

Sequencing strategy was performed using a targeted multiplexed amplicon-based approaches (AmpliSeq Custom DNA panels for standard DNA, Illumina, San Diego, CA, USA) starting from genomic DNA; the resulting libraries were sequenced on Illumina platforms (NextSeq500) in paired-end mode.

Median depth of sequencing was 3455x. Variants with a coverage <500x and/or <5 reads carrying the nucleotidic variant were filtered out. Moreover, mutations with low VAF (<0.10) were verified by a re-sequencing on an independent platform (at MML laboratory, Munich). The effectiveness to confirm a variant with whit VAF ≥0.01 and <0.10 by independent re-sequencing was 96.5% (without no significant differences between variants with VAF ≤0.05 vs >0.05), while the technique we used was significantly less performant in and reproducible below the threshold of 0.01 (P=0.02). Basing on these considerations, variants <0.01 VAF were filtered.

Functionally annotated variants were then also excluded based on the information retrieved from public databases (dbSNP, gnomAD) and the expected germ line allele frequency. Single nucleotide polymorphisms (SNP) were annotated according to the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp; Build 137) and gnomAD (http://gnomad.broadinstitute.org; gnomAD r2.0.1) databases.

The remaining variants were considered as possible somatic mutations and their pathogenic value was evaluated in order to differentiate known and putative pathogenic mutations from variants of unclear significance by using a multi-step algorithm:

1) All variants (missense, in-frame insertions/deletions, frameshift, nonsense and splice site) were considered pathogenic if they were previously reported in the publicly accessible Catalogue Of Somatic Mutations In Cancer (COSMIC, version 69) (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic) at least in two hematological sample.

2) Internal-tandem-duplication of FLT3 and in-frame insertions/deletions of CALR (exon 9) genes were included as pathogenic variants.

3) Loss of function mutations (Nonsense, frameshift and splice site) were considered pathogenic.

4) Missense variants and in-frame insertions/deletions not fulfilling these criteria were individually assessed based on the available data from COSMIC (the tissues they were found in, whether any other COSMIC variants were reported affecting the same amino-acid positions or were within 3 amino-acids) and their predicted functional consequences using the Mutation Taster algorithm (http://www.mutationtaster.org).

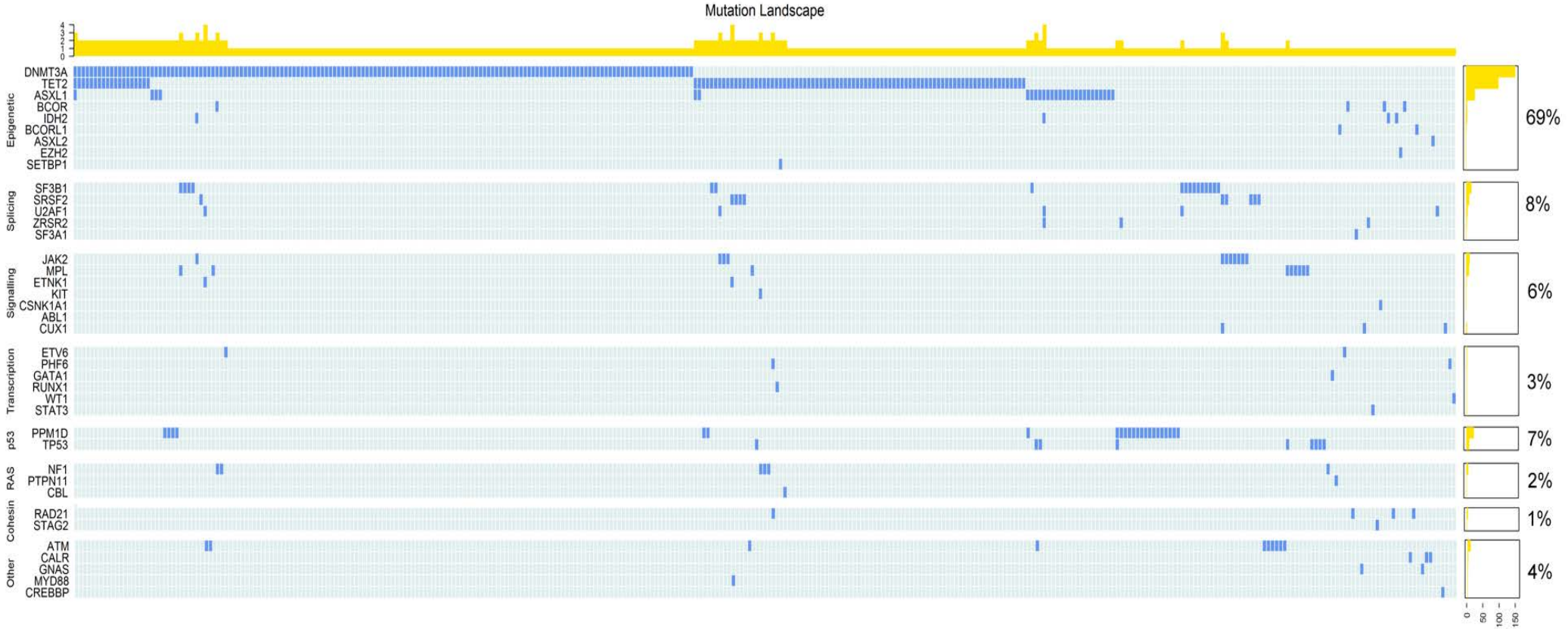
5) Nonsynonymous variants not fulfilling these criteria were then classified on the basis of their functional interpretation using in silico prediction effect by SIFT 1.03 (http://sift.jcvi.org), PolyPhen 2.0 (http://genetics.bwh.harvard.edu/pph2) and MutationTaster 1.0 algorithms (http://www.mutationtaster.org). Variants with less than 2/3 deduced damaging consequences on the amino acid level were discarded.

6) Additionally, TP53 variants were verified using the IARC repository (https://p53.iarc.fr/).

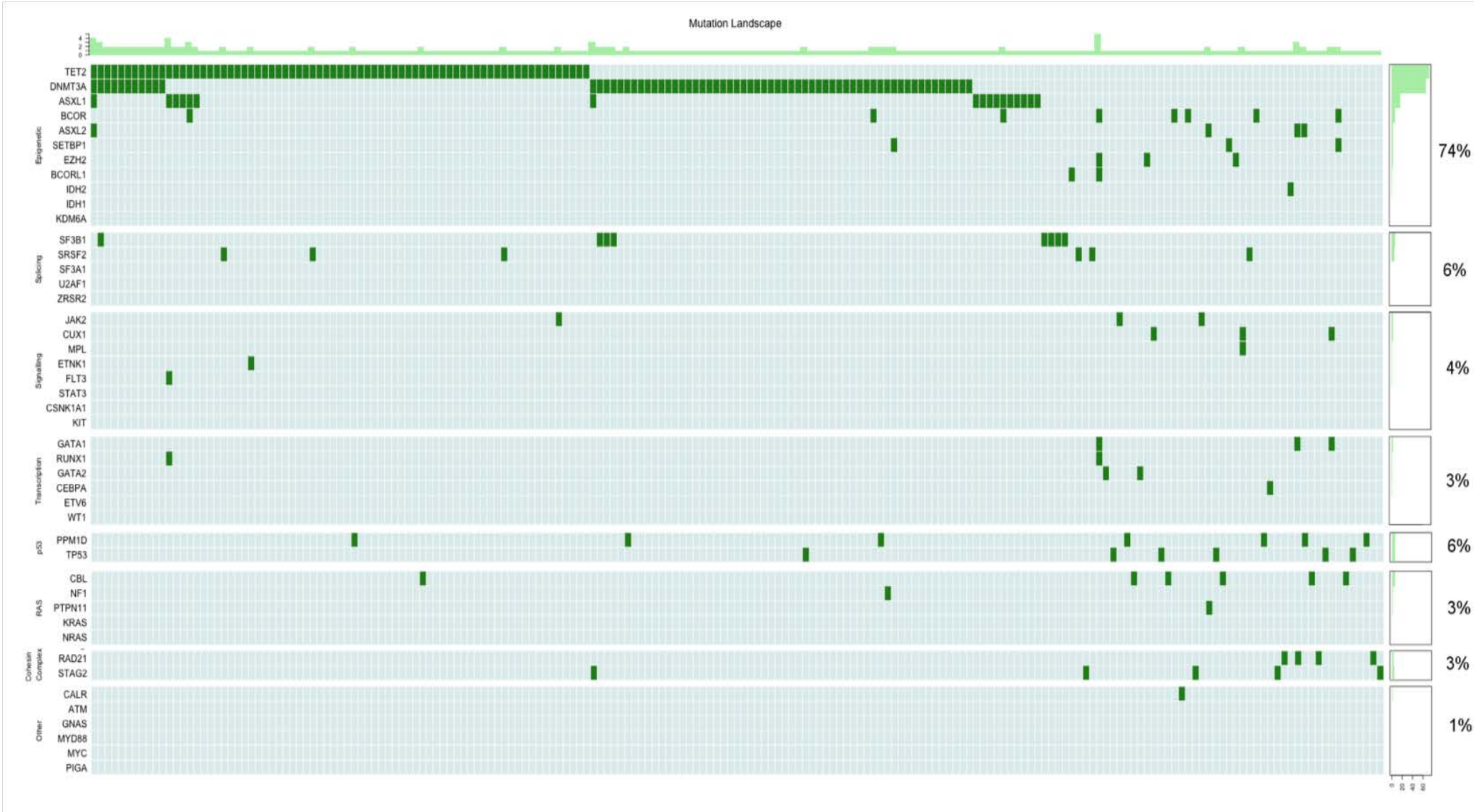
Variants that did not satisfy any of the above criteria were not considered as pathogenic mutations in downstream analyses.

Supplementary Figure_1. Patterns of mutations *(A,B)* and variant allele frequency (VAF) distribution of most frequently mutated genes *(C, D)* identified in the "Health & Anemia" and "Monzino 80+" cohorts. (E) VAF distribution by age groups in oldest old individuals from "Health_&_Anemia" and "Monzino_80+" cohorts. (F) Gene counts in males and females in oldest old individuals from "Health_&_Anemia" and "Monzino_80+" cohorts.

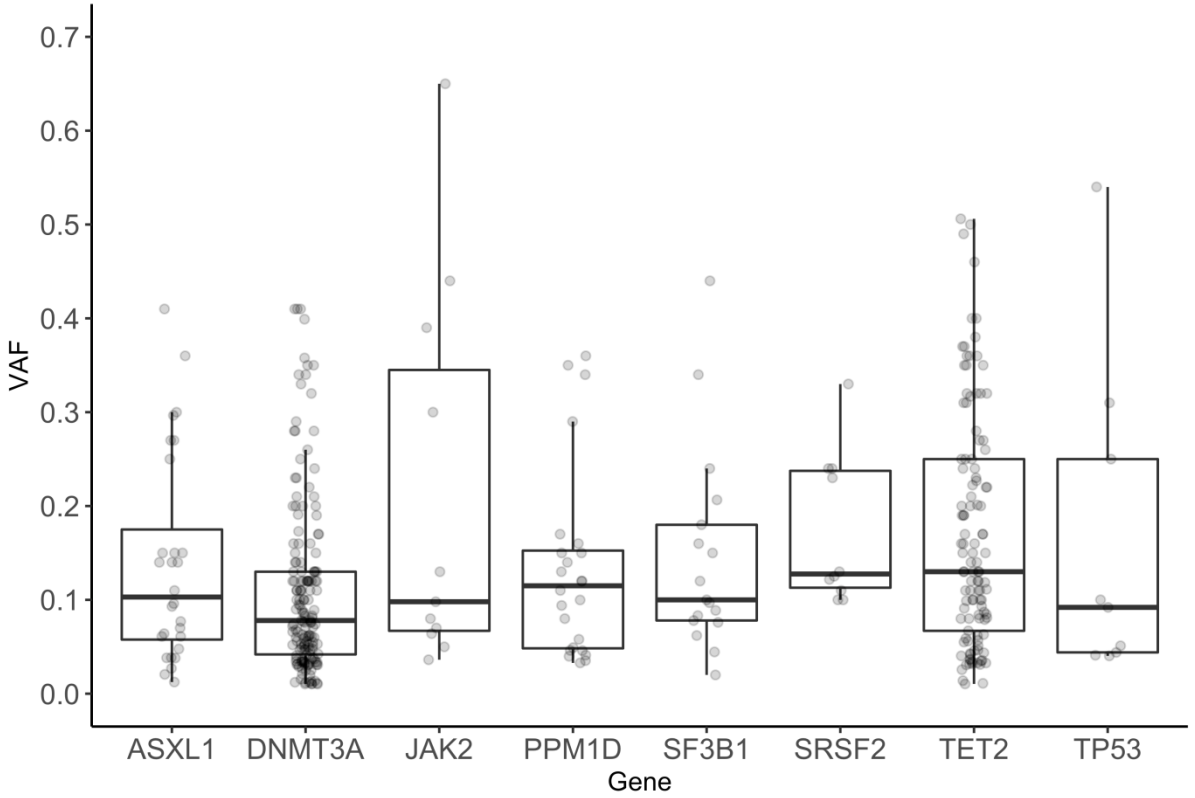
A) *Patterns of mutations identified in the "Health & Anemia" cohort*



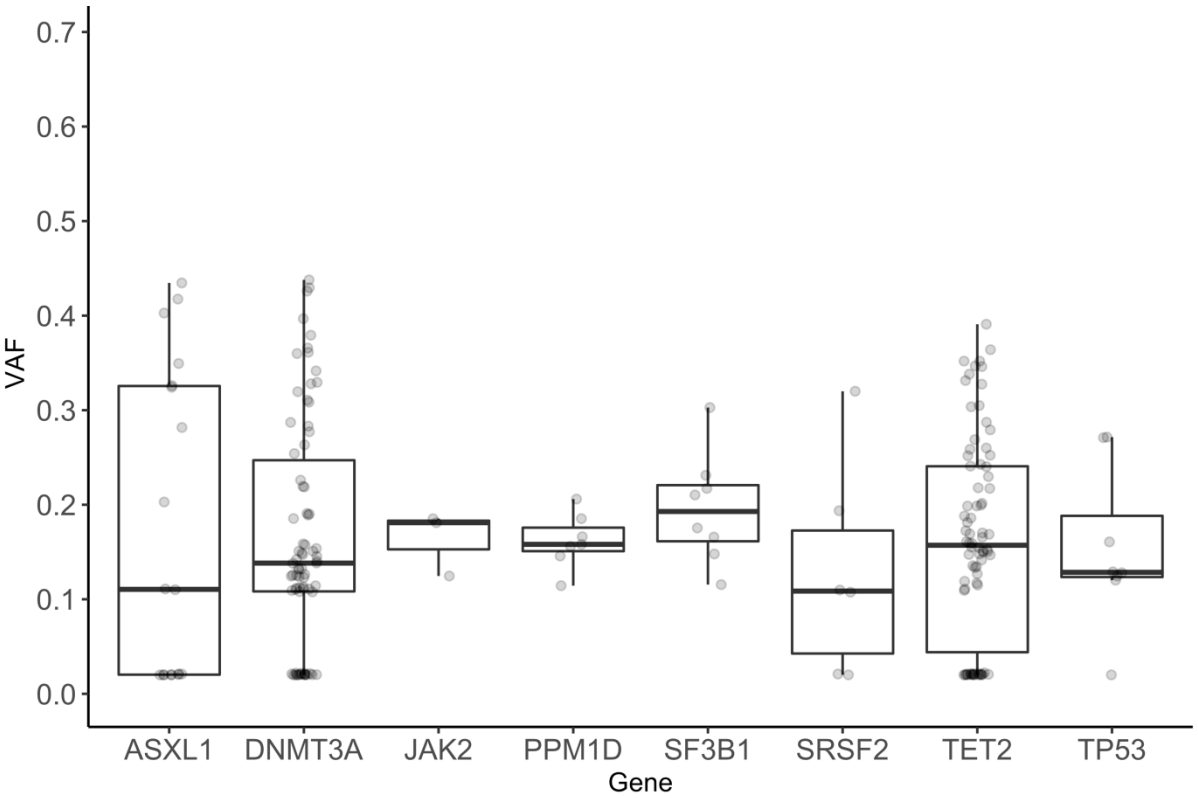
B) *Patterns of mutations identified in the "Monzino 80+" cohort*



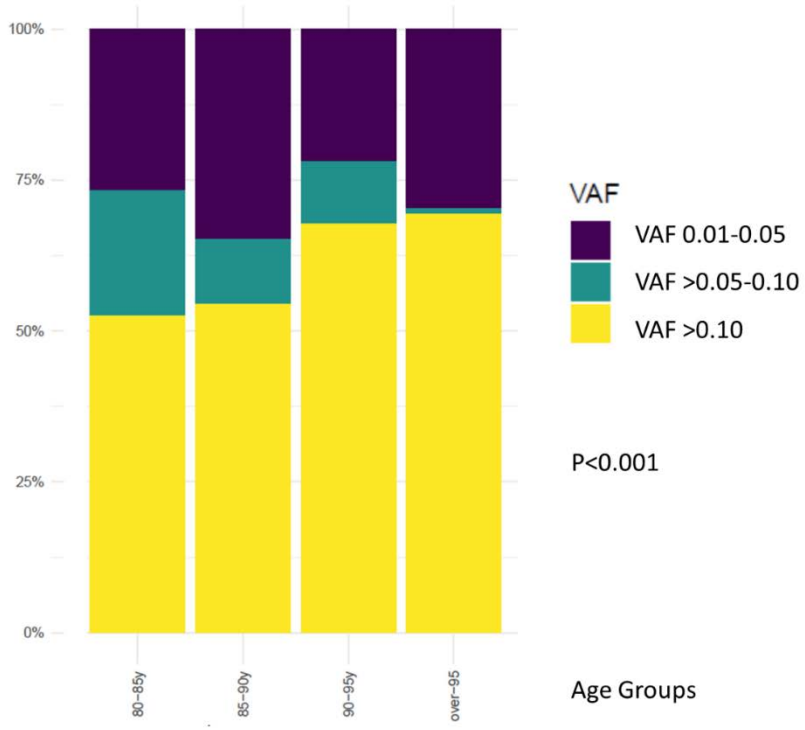
C) *VAF of most frequently reported mutations in the "Health & Anemia" cohort*



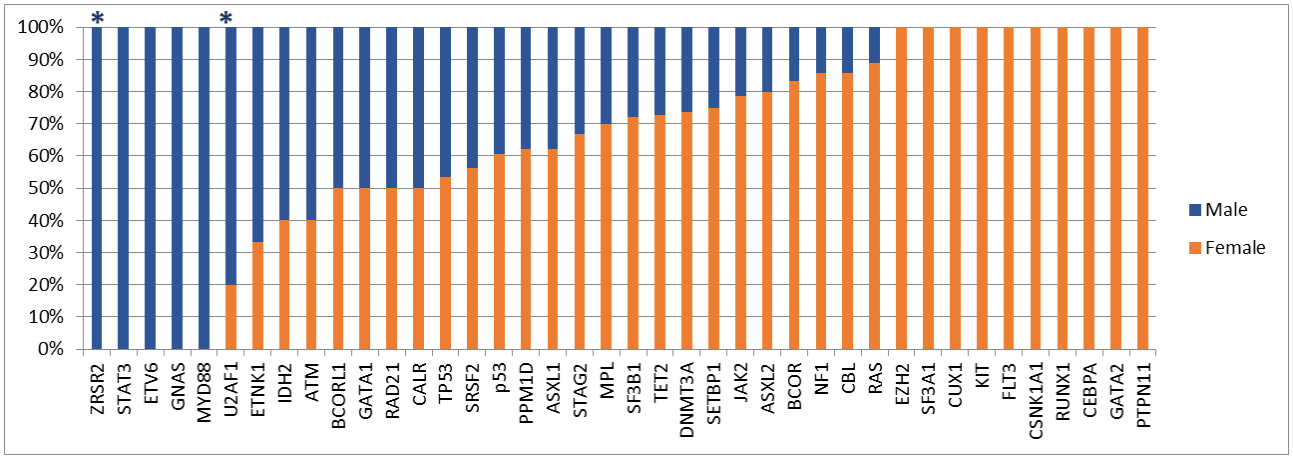
D) *VAF of most frequently reported mutations in the "Monzino 80+" cohort*



E) *VAF distribution by age groups in oldest old individuals from "Health_&_Anemia" and "Monzino_80+" cohorts*



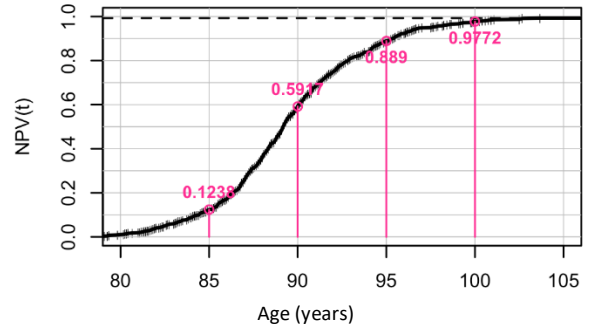
F) *Gene counts in males and females in oldest old individuals from "Health_&_Anemia" and "Monzino_80+" cohorts. Black asterisks denote statistically significant male-biased genes.*



Supplementary Figure_2. (*A,B*) Evaluation of accuracy of specific mutational and hematological features in predicting the risk of developing myeloid neoplasms. These analyses were carried out on oldest-old subjects from "Health_&_Anemia" cohort.

A) *Time-dependent positive predictive value (PPV) and negative predictive value (NPV) of specific categorical variables (mutational features)*

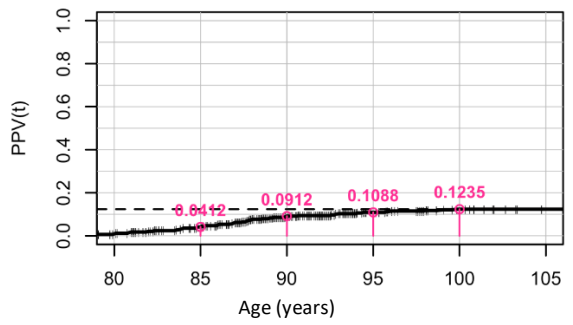
Absence of CHIP (time-dependent NPV)



NPV at selected time-points

85y	0.12[0.09-0-02]
90y	0.59[0.54-0.64]
95y	0.89[0.86-0.92]
100y	0.98[0.96-0.99]

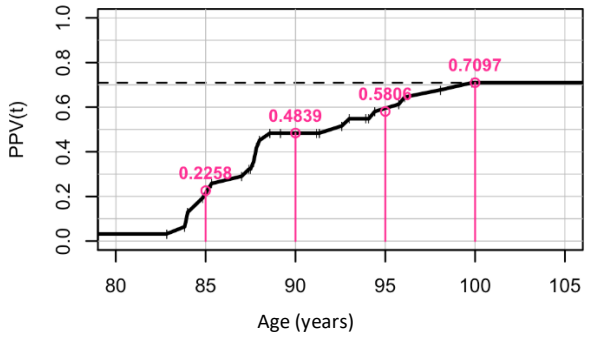
Presence of CHIP (time-dependent PPV)



PPV at selected time-points

85y	0.04[0.02-0.06]
90y	0.09[0.06-0.12]
95y	0.11[0.09-0.16]
100y	0.12[0.09-0.015]

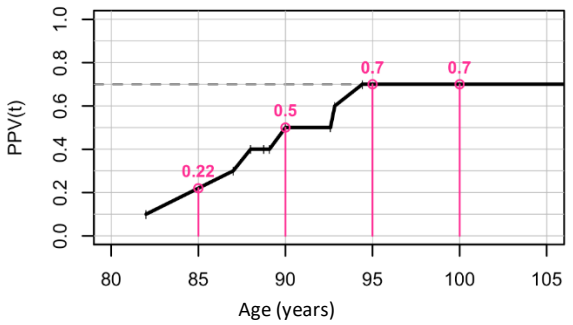
Splicing mutations (time-depedent PPV)



PPV at selected time-points

85y	0.230.09-0.37]
90y	0.48[0.25-0.59]
95y	0.58[0.41-0.63]
100y	0.71[0.44-0.78]

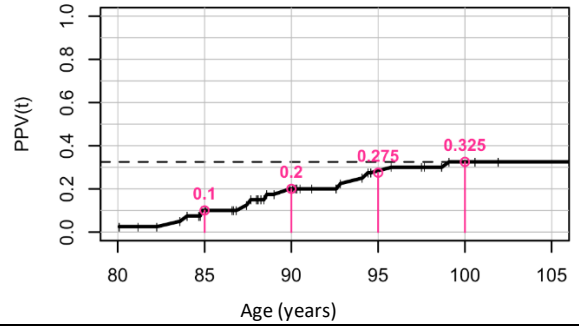
JAK2 mutations (time-dependent PPV)



PPV at selected time-points

85y	0.22[0.04-0.40]
90y	0.50[0.19-0.81]
95y	0.70[0.41-0.98]
100y	0.70[0.41-0.98]

Co-mutational patterns involving TET2, DNMT3A and ASXL1 (time-dependent PPV)



PPV at selected time-points

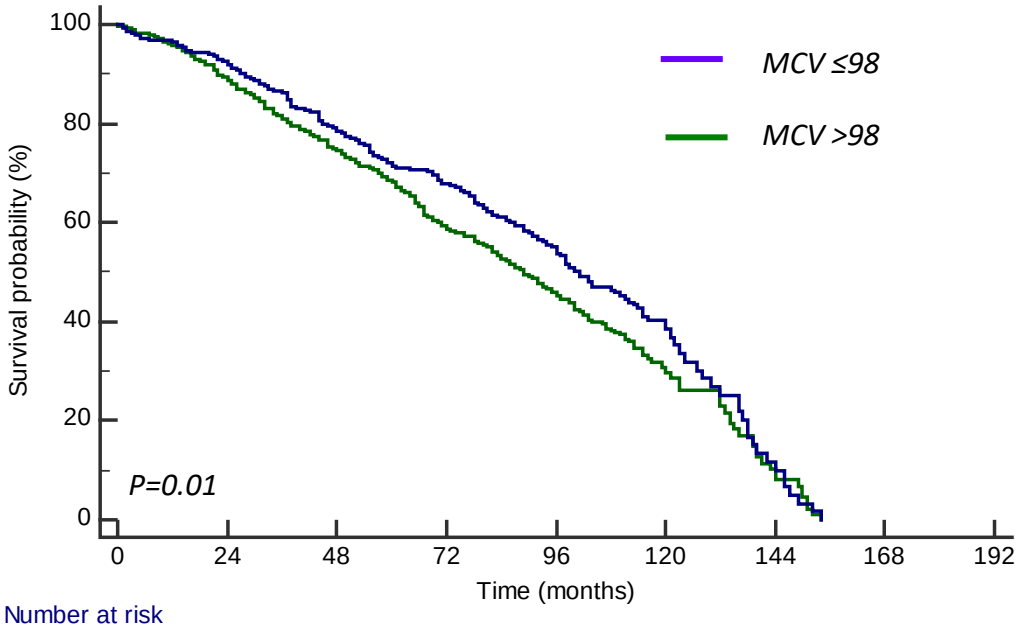
85y	0.10[0.01-0.19]
90y	0.20[0.08-0.32
95y	0.28[0.14-0.41]
100y	0.32[0.18-0.47]

B) *Accuracy of specific hematological and mutational features in predicting the risk of developing myeloid neoplasms (univariable analysis)*

Variable	Predictive value	P
CHIP	PPV 0.11 and NPV 0.89	0.74 and 0.001, respectively
Splicing mutations	PPV 0.58	0.001
Slicing mutations (without co-mutations)	PPV 0.49	0.007
JAK2 mutations	PPV 0.70	<0.001
TET2 mutations	PPV 0.06	0.55
DNMT3A mutations	PPV 0.07	0.78
ASXL1 mutations	PPV 0.04	0.81
Co-mutational patterns involving TET2, DNMT3A and ASXL1	PPV 0.28	0.019
VAF	Cut off 0.096	<0.001
MCV	Cut off >98	0.001
RDW	Cut off >14	0.004

Supplementary Figure_3. Association between increased MCV (>98) and RDW (>14) and probability of overall survival in subjects enrolled in "Health & Anemia" (*A,B*) and "Monzino 80+" *(C,D)* cohorts

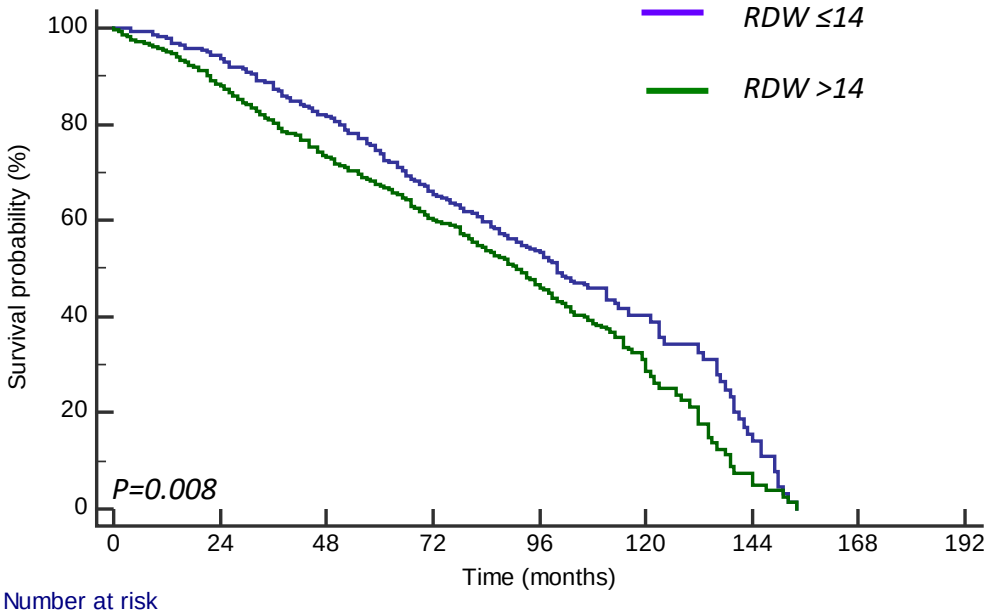
A) *Association between increased MCV (>98) and probability of overall survival in subjects from "Health & Anemia" cohort*



Number at risk

MCV≤98	680	599	503	364	276	26	7	0
MCV >98	361	331	282	240	187	23	6	0

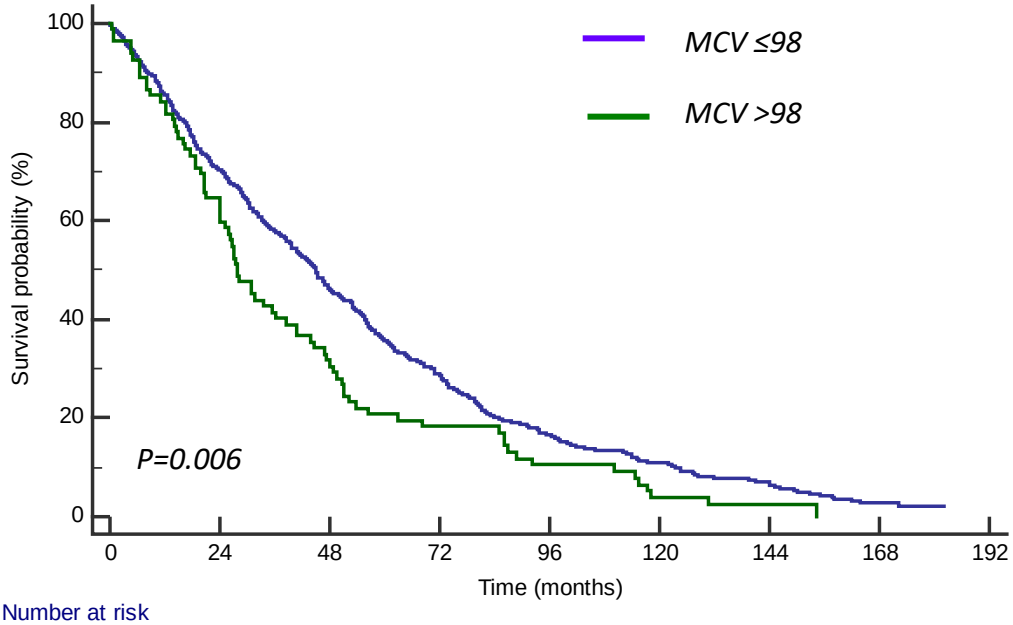
B) *Association between increased RDW (>14) and probability of overall survival in subjects from "Health & Anemia" cohort*



Number at risk

0								
RDW≤14	344	317	277	203	165	26	9	0
1								
RDW >14	697	613	508	401	298	23	4	0

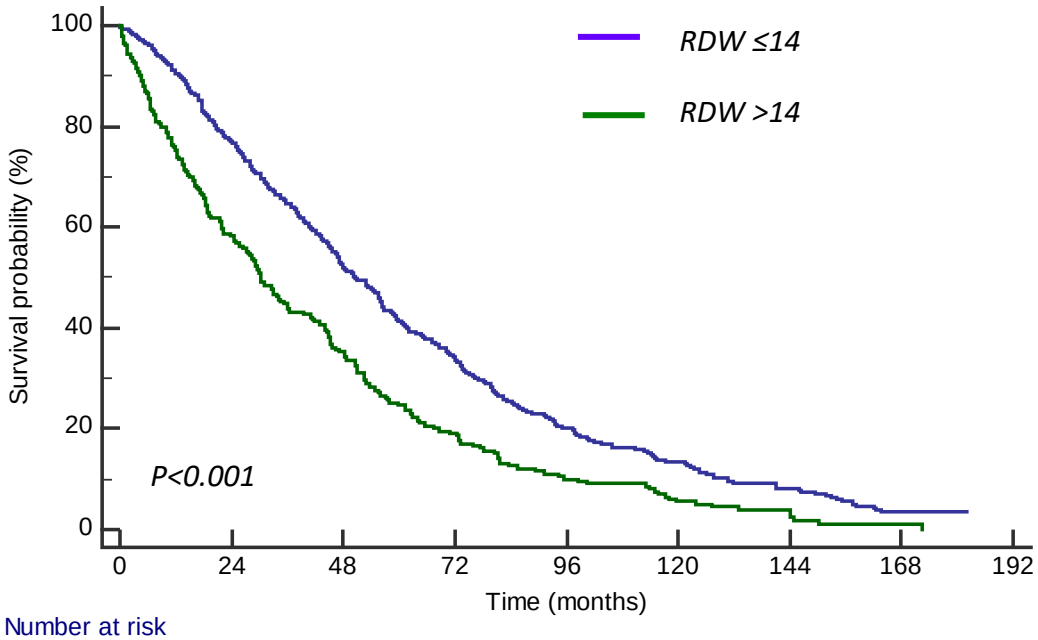
C) *Association between increased MCV (>98) and probability of overall survival in subjects from "Monzino 80+" cohort*



Number at risk

MCV≤98	653	459	299	185	101	66	38	8	0
MCV >98	82	52	25	14	8	3	2	0	0

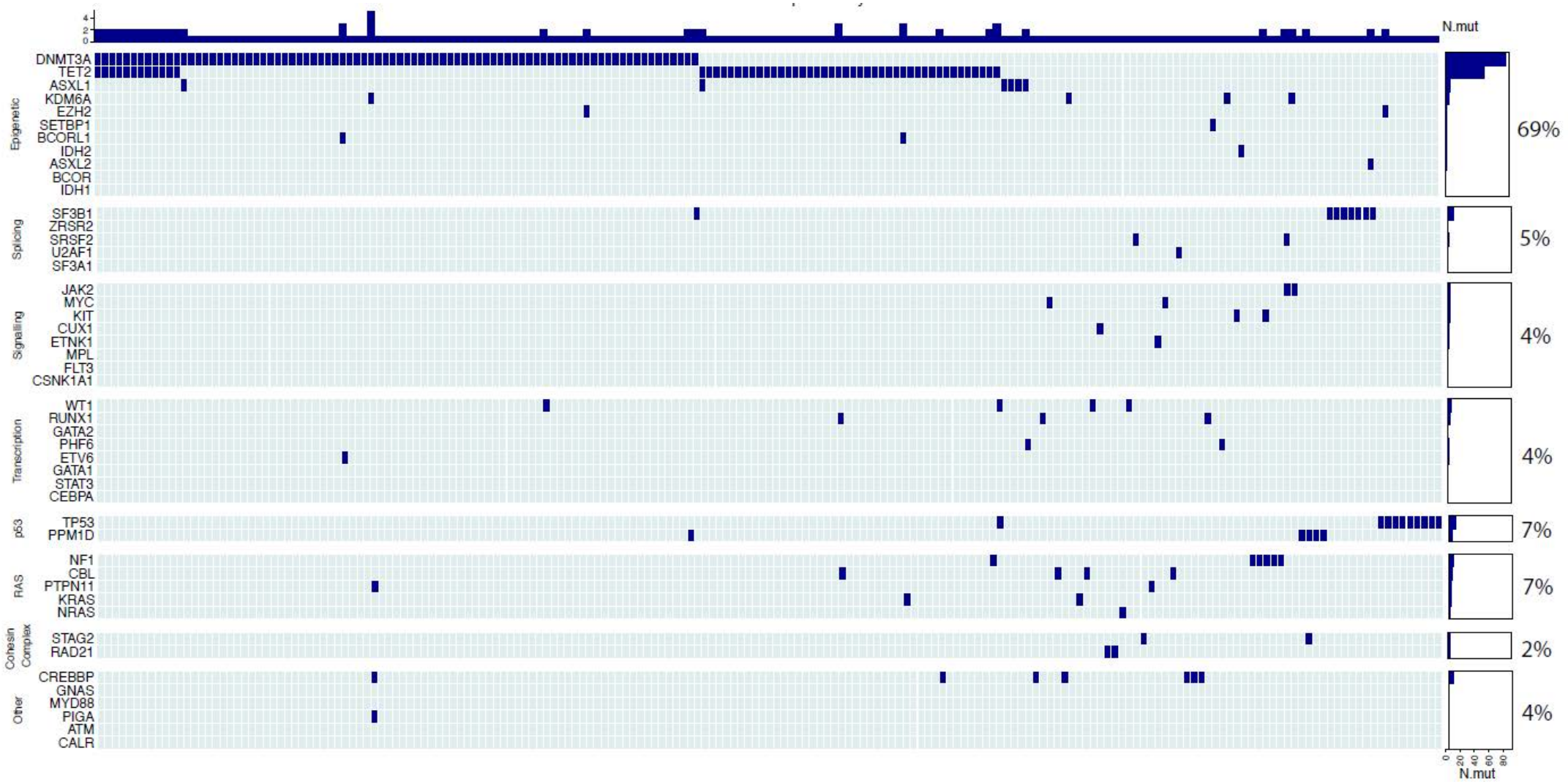
D) *Association between increased RDW (>14) and probability of overall survival in subjects from "Monzino 80+" cohort*



Number at risk

RDW≤14	469	361	241	156	89	58	34	6	0
RDW >14	199	116	70	38	20	11	6	2	0

Supplementary Figure_4. Patterns of mutations identified in the cohort of 727 subjects aged ≥75<80 years from "Health_&_Anemia" cohort.



Supplementary table_4. Clinical features of 255 oldest-old patients with myeloid neoplasms (myelodysplastic syndromes) reported to EUROMDS database (*A*) and pattern of the mutations identified in this patient population (*B*).

A) *Clinical features of 255 oldest-old patients with myelodysplastic syndrome*

Patients	*N=255*
Age	84 (80-94)
Female/Male	105/150 (41%/59%)
WHO diagnosis (2016 criteria) – MDS with del(5q) – MDS-SLD – MDS-RS-SLD – MDS-MLD – MDS-RS-MLD – MDS-EB1 – MDS-EB2	 17 (6.6%) 16 (6.2%) 32 (12.5%) 50 (19.6%) 45 (17.6%) 32 (12.5%) 63 (24.7%)
Leukocytes x10^9/L	4.8 (0.6-60)
Neutrophils x10^9/L	2.5 (0.2-36)
Platelets x10^9/L	136 (3-1068)
Hemoglobin g/dL	9.9 (6.1-15.3)
IPSSR risk group (n=226) – Very low – Low – Intermediate – High – Very high	 34 (15%) 95 (42.1%) 44 (19.4%) 40 (17.7%) 13 (5.8%)
Therapy (n=231) – No treatment – Erythroid stimulating agents/Red blood cell transfusions – Hypomethylating agents – Low-dose chemotherapy – Other (lenalidomide, etc.)	 129 (55.8%) 85 (36.7%) 44 (19%) 30 (12.9%) 15 (6.5%)

MDS= myelodysplastic syndromes; SLD= single lineage dysplasia; MLD=multilineage dysplasia; RS= ring sideroblasts; EB= excess blasts; IPSS-R= Revised International Prognostic Scoring System

B) *Pattern of the mutations identified in 255 oldest-old patients with myelodysplastic syndrome reported to EuroMDS database*