Supplemental Methods

Questionnaires and physical examination

All subjects completed a self-administered questionnaire on medical history, past and current diseases, use of medication, and health behavior at home. Medication use was verified by a certified research assistant, and scored by ATC code. Smoking status was defined as non-smoker, former smoker and current smoker (including the use of cigarettes, cigarillos, cigars and pipe tobacco).[1] Body weight was measured wearing light clothes and without shoes to the nearest 0.1kg. Height was measured to the nearest 0.5 cm. Body mass index was calculated as weight (kg) divided by height squared (m^2). Systolic and diastolic blood pressure were measured every minute for a period of 10 minutes using a DINAMAP Monitor. The average of the last three measurements was reported. At the baseline visit, a resting 12-lead electrocardiogram was made and evaluated as described previously.[2]

Biochemical measurements

Blood samples were collected in the morning after an overnight fast. Hemoglobin, hematocrit, leukocytes and platelets were measured using routine procedures on a XE2100-system (Sysmex, Japan). HbA1c was measured using a turbidimetric inhibition immunoassay on a Cobas Integra 800 CTS analyzer (Roche Diagnostics, the Netherlands). Fasting blood glucose was measured using a hexokinase method. Serum levels of total and high-density lipoprotein cholesterol were measured using an enzymatic colorimetric method, triglycerides using a colorimetric UV method, and low-density lipoprotein (LDL) cholesterol using an enzymatic method, all on a Roche Modular P chemistry analyzer (Roche, Switzerland). Serum creatinine was measured on a Roche Modular P chemistry analyzer (Roche, Switzerland).

Next generation sequencing (NGS)

Gene panel

For NGS we used the Illumina® TruSight® Myeloid Sequencing Panel with a filter on the following 25 genes: *ASXL1* (NM_015338), *BCOR* (NM_001123385), *BCORL1* (NM_021946), *CALR* (NM_004343), *CBL* (NM_005188), *CSF3R* (NM_156039), *DNMT3A* (NM_175629), *EZH2* (NM_004456), *FLT3* (NM_004119), *IDH1* (NM_005896), *IDH2* (NM_002168), *JAK2* (NM_004972), *KIT* (NM_000222), *KRAS* (NM_033360), *NPM1* (NM_002520), *NRAS* (NM_002524), *MPL* (NM_005373), *RUNX1* (NM_001754), *SETBP1* (NM_015559), *SF3B1* (NM_012433) , *SRSF2* (NM_003016), *TET2* (NM_001127208), *TP53* (NM_000546), *U2AF1* (NM_006758), *WT1* (NM_024426).

Sample processing, library preparation, clustering and sequencing

DNA was extracted from whole blood using Autopure LS (Qiagen). DNA quality was analyzed using the 260/280 ratio (Nanodrop 2000) and checked by agarose gel electrophoresis. The genomic library was prepared using 50 ng of DNA template and the Illumina® TruSight® Myeloid Sequencing Panel according to the manufacturer's instructions. Sequencing quality was regarded sufficient in case of a Q30 score of >85%, and if >95% of the sequencing panel was covered at 500x.

Bioinformatic processing

Bioinformatic processing was performed using two different bioinformatic platforms, e.g. NextGENe version 2.3.4.2 (SoftGenetics) combined with Cartagenia Bench Lab NGS (Cartagenia), and Sophia DDM (Sophia Genetics). For both platforms, the human genome build 19 (hg19) was used as a

reference. The minimum variant allele frequency (VAF) was set to 5.0%. Mutations were only reported when present on both platforms.

Variant classification

Variants were classified as mutations based on the level of evidence taking into account the following aspects: type of mutation (frameshift, nonsense, missense or splicing mutation); previous classification; read depth; population frequency, frequency of mutation alone, frequency of mutation in the run, and frequency in our account; in sillico prediction scores (PolyPhen-2, SIFT, MutationTaster, LRT, FATHMM, MutationAssessor, Human Splicing Finder, and Align-GVGD), and information provided by different databases (COSMIC, NCBI, in-house database and Varsome). Depending on the level of evidence, mutations were classified into 5 classes ranging from class 1 (benign) to class 5 (highly pathogenic), according to the Human Genome Variation Society nomenclature.[3,4] Only class 3 (variant of unknown significance), class 4 (likely pathogenic) and class 5 (highly pathogenic) variants were reported and used for subsequent analysis.

R packages for data processing and statistical analyses

The following packages were used for data processing and analysis in R3.5.2.:

- Data reshaping and summarization: *dplyr (*https://github.com/hadley/dplyr*)*, *plyr (*https://github.com/hadley/plyr*)*, *tidyr* and *reshape2 (*https://github.com/hadley/reshape*)*
- Visualisations: *ggplot2 (*https://github.com/tidyverse/ggplot2*)* and *g3viz https://cran.r-project.org/web/packages/g3viz/index.html*.
- Survival analyses and visualization: *survminer* (https://cran.r-project.org/web/packages/survminer/index.html), *survival* (http://cran.r-project.org/web/packages/survival/index.html), *cmprsk* (http://cran.r-project.org/web/packages/cmprsk/index.html) and *riskRegression* (https://cran.r-project.org/web/packages/riskRegression/index.html).

References

1 Slagter SN, van Vliet-Ostaptchouk JV, Vonk JM, et al. Associations between smoking, components of metabolic syndrome and lipoprotein particle size. *BMC Med* 2013; 11: 195.

2 van der Ende MY, Hartman MH, Schurer RA et al. Prevalence of electrocardiographic unrecognized myocardial infarction and its association with mortality. *Int J Cardiol* 2017; 243: 34-39.

3 Den Dunnen JT, Dalgleish R, Maglott DR. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* 2016; 37(6): 564-569.

4 Richards S, Aziz N, Bale S. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17(5): 405-424.

Supplementary Table 1. Baseline characteristics, cardiovascular disease history and all-cause mortality of the study cohort.

	General Lifelines population (n=142 172)	Individuals with erythrocytosis (wide) (n=4 995)	*P*
Male/female (%/%)	39.8/60.2	93.8/6.2	<.001
Age (years)	44.8 ± 13.1	45.1 ± 13.1	.18
BMI (kg/m^2)	26.0 ± 4.3	27.2 ± 3.9	<.001
Systolic BP (mmHg)	125 ± 15	134 ± 15	<.001
Diastolic BP (mmHg)	74 ± 9	79 ± 10	<.001
Creatinine (µmol/L)	73.1 ± 13.6	83.1 ± 11.8	<.001
Number of medications used	1 (0 – 2)	0 (0 – 1)	<.001
Use of antihypertensive drugs (%)	12.5	13.8	.010
Use of drugs for obstructive airway diseases (%)	7.0	7.3	.28
Use of androgens (%)	0.1	0.6	<.001
Current smokers (%)	20.4	29.7	<.001
Diabetes (%)	3.3	4.1	.006
Metabolic syndrome (%)	16.8	31.0	<.001
Family history of CVD (%)	8.8	12.6	<.001
Hemoglobin (g/dL)			
Male	15.0 ± 0.8	16.9 ± 0.5	<.001
Female	13.4 ± 0.9	16.2 ± 0.6	<.001
Hematocrit (%)			
Male	44.4 ± 2.3	49.0 ± 1.6	<.001
Female	40.5 ± 2.5	48.2 ± 2.1	<.001
MCV (fL)	89.9 ± 4.2	90.2 ± 3.9	<.001
Leukocytes (10^9/L)	6.1 ± 1.8	6.7 ± 1.9	<.001
Leukocytosis (%)	2.9	5.9	<.001
Platelets (10^9/L)	251 ± 56	227 ± 55	<.001
Thrombocytosis (%)	1.2	0.5	<.001
Fasting glucose (mmol/L)	5.0 ± 0.8	5.2 ± 1.1	<.001
HbA1c (%)	5.5 ± 0.4	5.5 ± 0.5	<.001
Total cholesterol (mmol/L)	5.1 ± 1.0	5.3 ± 1.1	<.001
LDL Cholesterol (mmol/L)	3.2 ± 0.9	3.5 ± 0.9	<.001
HDL Cholesterol (mmol/L)			
Male	1.3 ± 0.3	1.2 ± 0.3	<.001
Female	1.6 ± 0.4	1.5 ± 0.4	<.001
Triglycerides (mmol/L)	1.2 ± 0.8	1.7 ± 1.3	<.001
Thrombosis (%)	1.7	1.4	.13
Stroke (%)	0.6	0.5	.75
Myocardial infarction (%)	1.2	2.1	<.001
Composite cardiovascular events (%)	3.3	3.9	.016
Antithrombotic agent use (%)			
Vitamin K antagonists	0.9	1.7	<.001
Platelet aggregation inhibitors	3.2	4.5	<.001
Direct factor Xa inhibitors	0.0	0.0	.57
All-cause mortality (%)	1.9	2.7	<.001

Data are given as mean ± SD, median (IQR) when not normally distributed, or percentage. Leukocytosis was defined as leukocytes ≥$10x10^9$/L. Thrombocytosis was defined as a platelet count ≥$400x10^9$/L. Abbreviations: BMI, body mass index; BP, blood pressure; CVD, cardiovascular disease; HDL, high-density lipoprotein; LDL, low-density lipoprotein; HbA1c, glycated hemoglobin; MCV, mean corpuscular volume.

Supplementary Table 2. Univariable and multivariable regression models for the association between erythrocytosis (strict criteria) and cardiovascular mortality.

		Unadjusted HR (95% CI)	Adjusted HR (95% CI)[1]	Adjusted HR (95% CI)[2]
Subdistribution hazard (Fine and Gray)	**No erythrocytosis**	1	1	1
	Erythrocytosis	6.3 (3.1-12.6)*	3.5 (1.7-6.9)*	2.2 (1.0-4.6)*
Cause-specific hazard (Cox regression)	**No erythrocytosis**	1	1	1
	Erythrocytosis	6.4 (3.2-13.0)*	3.6 (1.8-7.3)*	2.3 (1.1-4.8)*

Cause of death analysis were performed in a subset of evaluable Lifelines participants for which linkage to Statistics Netherlands succeeded (n=147 090): n=408 with and n=146 682 without erythrocytosis. 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. Data are shown as hazard ratio (HR) and 95% confidence interval (95% CI). The absence of erythrocytosis was used as reference. [1]Adjusted for age and sex (n=147 090). [2]Adjusted for age, sex, smoking status, diastolic and systolic blood pressure, medical history of diabetes, number of medications, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, low-density lipoprotein cholesterol, family history of cardiovascular disease and body mass index (n=144 463). *P<.05.

Supplementary Table 3. Baseline characteristics, cardiovascular disease and all-cause mortality of individuals with erythrocytosis with concurrent leuko- and/or thrombocytosis and matched individuals with isolated erythrocytosis.

	Individuals with erythrocytosis with concurrent leuko- and/or thrombocytosis (n=47)	**Matched individuals with isolated erythrocytosis (n=94)**	***P***
Male/female (%/%)	27.7/72.3	27.7/72.3	1.00
Age (years)	50.7 ± 11.5	50.9 ± 10.7	.89
BMI (kg/m^2)	28.6 ± 5.9	27.5 ± 5.5	.25
Systolic BP (mmHg)	136 ± 18	132 ± 16	.18
Diastolic BP (mmHg)	79 ± 12	77 ± 9	.17
Number of medications used	2 (1 – 4)	1 (0 – 3)	.40
Use of antihypertensive drugs (%)	25.5	29.8	.60
Use of drugs for obstructive airway diseases (%)	14.9	11.7	.59
Use of androgens (%)	0.0	0.0	NA
Current smokers (%)	63.6	63.4	.93
LDL Cholesterol (mmol/L)	3.5 ± 1.2	3.6 ± 1.0	.75
Diabetes (%)	12.8	6.4	.20
Family history of CVD (%)	13.0	19.1	.37
Hemoglobin (g/dL)			
Male	17.9 ± 1.2	17.9 ± 0.7	.95
Female	16.5 ± 0.9	16.2 ± 0.5	.04
Hematocrit (%)			
Male	53.6 ± 4.3	53.0 ± 0.8	.46
Female	49.4 ± 2.5	48.8 ± 1.1	.10
MCV (fL)	93.6 ± 5.2	92.8 ± 3.8	.34
Leukocytes (10^9/L)	11.2 ± 2.4	7.2 ± 1.5	<.001
Leukocytosis (%)	87.2	0.0	<.001
Platelets (10^9/L)	330 ± 121	249 ± 52	<.001
Thrombocytosis (%)	19.1	0.0	<.001
Composite cardiovascular events (%)	12.8	6.4	.20
All-cause mortality (%)	8.5	4.3	.30

Data are given as mean ± SD, median (IQR) when not normally distributed, or percentage. Leukocytosis was defined as leukocytes ≥10x10^9/L. Thrombocytosis was defined as platelet count ≥400x10^9/L. Abbreviations: BMI, body mass index; BP, blood pressure; CVD, cardiovascular disease; LDL, low-density lipoprotein; MCV, mean corpuscular volume.

Supplementary Table 4. Results of multivariable regression analysis showing the associations between erythrocytosis (wide criteria) and cardiovascular disease and mortality.

		Unadjusted OR/HR (95% CI)	Adjusted OR/HR (95% CI)
Erythrocytosis[1]	**Thrombosis**	0.8 (0.7 – 1.1)	1.0 (0.8 – 1.3)
	Stroke	0.9 (0.6 – 1.3)	0.8 (0.5 – 1.2)
	Myocardial infarction	1.7 (1.4 – 2.1)*	1.3 (1.0 – 1.6)*
	Composite cardiovascular events	1.2 (1.0 – 1.4)*	1.1 (0.9 – 1.3)
	Anti-thrombotic agents	1.5 (1.3 – 1.7)*	1.3 (1.2 – 1.6)*
	All-cause mortality	1.5 (1.3 – 1.8)*	1.2 (1.0 – 1.4)
	Cardiovascular mortality[3]	1.6 (1.1 – 2.3)*	1.1 (0.7 – 1.7)
Erythrocytosis with concurrent leuko- and/or thrombocytosis[2]	**Composite cardiovascular events**	1.3 (0.7 – 2.2)	1.0 (0.5 – 1.9)
	All-cause mortality	1.4 (0.7 – 2.5)	0.9 (0.5 – 1.9)

Adjusted regression models include sex, age, BMI, systolic and diastolic blood pressure, smoking status, medical history of diabetes, number of medications used, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, family history of cardiovascular disease and low-density lipoprotein cholesterol as covariates. Data are shown as odds ratio/hazard ratio (OR/HR) and 95% confidence interval (95% CI). [1]The absence of erythrocytosis was used as reference. [2]Isolated erythrocytosis was used as reference. [3]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.*P <.05.

Supplementary Table 5. Univariable and multivariable regression models for the association between erythrocytosis (wide criteria) and cardiovascular mortality.

		Unadjusted HR (95% CI)	Adjusted HR (95% CI)[1]	Adjusted HR (95% CI)[2]
Subdistribution hazard (Fine and Gray)	**No erythrocytosis**	1	1	1
	Erythrocytosis	1.6 (1.1-2.3)*	1.3 (0.9-1.9)	1.1 (0.7-1.7)
Cause-specific hazard (Cox regression)	**No erythrocytosis**	1	1	1
	Erythrocytosis	1.6 (1.1-2.3)*	1.3 (0.9-1.9)	1.1 (0.7-1.7)

Cause of death analysis were performed in a subset of evaluable Lifelines participants for which linkage to Statistics Netherlands succeeded (n=147 090): n=4 994 with and n=142 096 without erythrocytosis. 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. Data are shown as hazard ratio (HR) and 95% confidence interval (95% CI). The absence of erythrocytosis was used as reference. [1]Adjusted for age and sex (n=147 090). [2]Adjusted for age, sex, smoking status, diastolic and systolic blood pressure, medical history of diabetes, number of medications, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, low-density lipoprotein cholesterol, family history of cardiovascular disease and body mass index (n=144 463). *P<.05.

Supplementary Table 6. Results of multivariable regression analysis showing the associations between erythrocytosis (strict criteria) and cardiovascular disease and all-cause mortality in individuals not using diuretics and without metabolic syndrome (sensitivity analyses).

		Individuals not using diuretics[2]	Individuals not using diuretics and without metabolic syndrome[3]
		Adjusted OR/HR (95% CI)	**Adjusted OR/HR (95% CI)**
Erythrocytosis[1]	**Thrombosis**	1.5 (0.9 – 2.6)	2.4 (1.3 – 4.4)*
	Stroke	0.6 (0.2 – 2.2)	2.5 (0.7 – 8.8)
	Myocardial infarction	2.0 (1.1 – 3.7)*	2.7 (1.0 – 6.9)*
	Composite cardiovascular events	1.7 (1.1 – 2.6)*	2.7 (1.7 – 4.4)*
	Anti-thrombotic agents	2.0 (1.2 – 3.2)*	3.1 (1.7 – 5.5)*
	All-cause mortality	1.8 (1.2 – 2.8)*	1.3 (0.7 – 2.7)

Adjusted regression models include sex, age, BMI, systolic and diastolic blood pressure, smoking status, medical history of diabetes, number of medications used, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, family history of cardiovascular disease and low-density lipoprotein cholesterol as covariates. Data are shown as odds ratio/hazard ratio (OR/HR) and 95% confidence interval (95% CI). [1]The absence of erythrocytosis was used as reference. [2] 378 individuals with erythrocytosis. [3] 253 individuals with erythrocytosis.

Supplementary Table 7. Baseline characteristics of individuals with a *JAK2* V617F mutation VAF <5% compared to individuals with a *JAK2* V617F mutation VAF≥5%.

	Individuals with *JAK2* V617F <5% (n=7)	**Individuals with *JAK2* V617F ≥5% (n=7)**	***P***
Male/female (%/%)	0.0/100.0	42.9/57.1	.051
Age (years)	51.6 ± 11.9	61.3 ± 10.4	.13
BMI (kg/m^2)	26.3 ± 2.8	25.2 ± 2.4	.44
Number of medications used	1 (1 – 3)	3 (0 – 4)	.90
Current smokers (%)	71.4	14.3	.053
Hemoglobin (g/dL)			
Male	NA	17.8 ± 0.7	NA
Female	16.1 ± 0.5	16.2 ± 0.4	.75
Hematocrit (%)			
Male	NA	57.5 ± 9.0	NA
Female	48.5 ± 1.1	49.6 ± 1.4	.19
MCV (fL)	93.2 ± 4.7	89.3 ± 4.7	.10
Leukocytes (10^9/L)	6.5 ± 1.4	9.8 ± 3.0	.024
Leukocytosis (%)	0.0	57.1	.018
Platelets (10^9/L)	255 ± 37	524 ± 180	.002
Thrombocytosis (%)	0.0	71.4	.005

Data are given as mean ± SD, median (IQR) when not normally distributed, or percentage. Leukocytosis was defined as leukocytes ≥10x10^9/L. Thrombocytosis was defined as a platelet count ≥400x10^9/L. Abbreviations: BMI, body mass index; MCV, mean corpuscular volume.

Supplementary Table 8. Baseline characteristics, cardiovascular disease and all-cause mortality of individuals with erythrocytosis with and without clonal hematopoiesis.

	Individuals with clonal hematopoiesis (n=51)	Individuals without clonal hematopoiesis (n=82)	*P*
Male/female (%/%)	25.5/74.5	30.5/69.5	.54
Age (years)	54.4 ± 10.5	48.4 ± 10.3	.001
BMI (kg/m^2)	27.5 ± 5.4	27.8 ± 5.6	.77
Systolic BP (mmHg)	133 ± 18	133 ± 16	.97
Diastolic BP (mmHg)	76 ± 11	78 ± 9	.15
Number of medications used	2 (0 – 4)	1 (1 – 3)	.40
Use of antihypertensive drugs (%)	33.3	25.6	.34
Use of drugs for obstructive airway diseases (%)	7.8	15.9	.18
Use of androgens (%)	0.0	0.0	NA
Current smokers (%)	58.8	66.7	.54
LDL Cholesterol (mmol/L)	3.3 ± 0.9	3.8 ± 1.1	.004
Diabetes (%)	11.8	6.1	.25
Family history of CVD (%)	19.6	17.3	.74
Hemoglobin (g/dL)			
Male	18.2 ± 0.6	17.8 ± 0.6	.21
Female	16.2 ± 0.4	16.3 ± 0.8	.30
Hematocrit (%)			
Male	54.0 ± 4.3	52.8 ± 0.7	.17
Female	49.0 ± 1.0	49.1 ± 2.1	.75
MCV (fL)	92.4 ± 4.3	93.5 ± 4.3	.17
Leukocytes (10^9/L)	8.3 ± 2.7	8.6 ± 2.7	.54
Leukocytosis (%)	29.4	29.2	.94
Platelets (10^9/L)	286 ± 126	270 ± 62	.31
Thrombocytosis (%)	9.8	4.9	.51
Glucose (mmol/L)	5.2 ± 0.9	5.1 ± 0.7	.68
HbA1c (%)	5.7 ± 0.6	5.6 ± 0.4	.38
Total cholesterol (mmol/L)	5.1 ± 1.0	5.7 ± 1.3	.011
LDL Cholesterol (mmol/L)	3.3 ± 0.9	3.8 ± 1.1	.004
HDL Cholesterol (mmol/L)			
Male	1.2 ± 0.2	1.1 ± 0.3	.10
Female	1.5 ± 0.4	1.4 ± 0.4	.06
Triglycerides (mmol/L)	1.3 ± 0.6	1.7 ± 1.2	.017
Cardiovascular event (%)	15.7	4.9	.034
All-cause mortality (%)	7.8	4.9	.48

Data are given as mean ± SD, median (IQR) when not normally distributed, or percentage. Leukocytosis was defined as leukocytes ≥10x10^9/L. Thrombocytosis was defined as a platelet count ≥400x10^9/L. Abbreviations: BMI, body mass index; BP, blood pressure; CVD, cardiovascular disease; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MCV, mean corpuscular volume.

Supplementary Table 9. Incident hematological malignancies in the cohort of individuals with erythrocytosis.

ID	Definition of erythrocytosis	Leukocytosis	Thrombocytosis	NGS	Gene	VAF (%)	pNomen	cNomen	Diagnosis		Time to diagnosis (months)
1	Strict	-	X	X	*JAK2*	43.2	p.V617F	c.1849G>T	Myeloid	MPN	4
					TP53	9.5	p.D313G	c.938A>G			
2	Strict	-	-	-					Myeloid	MPN	4
3	Strict	X	-	X	*JAK2*	77.5	p.V617F	c.1849G>T	Myeloid	MPN	7
4	Strict	X	X	X	*JAK2*	39.2	p.V617F	c.1849G>T	Myeloid	MPN	13
5	Strict	-	X	X	*JAK2*	26.6	p.V617F	c.1849G>T	Myeloid	MPN	18
					TP53	5.7	p.Q5*	c.13C>T			
6	Strict	-	-	-					Myeloid	MPN	59
7	Wide	-	-	-					Myeloid	MPN	39
8	Wide	X	-	-					Myeloid	MPN	44
9	Wide	-	-	-					Myeloid	MPN	54
10	Wide	X	-	-					Myeloid	MPN	56
11	Wide	-	-	-					Myeloid	MPN	80
12	Strict	-	-	-					Lymphoid	B-cell	54
13	Strict	-	-	X	*BCOR*	15.4	p.C1525R	c.4573T>C	Lymphoid	B-cell	80
14	Strict	-	-	-					Lymphoid	B-cell	89
15	Wide	-	-	-					Lymphoid	B-cell	11
16	Wide	-	-	-					Lymphoid	B-cell	32
17	Wide	-	-	-					Lymphoid	B-cell	37
18	Wide	-	-	-					Lymphoid	B-cell	43
19	Wide	-	-	-					Lymphoid	B-cell	49
20	Wide	-	-	-					Lymphoid	B-cell	52
21	Wide	-	-	-					Lymphoid	B-cell	70
22	Wide	X	-	-					Lymphoid	B-cell	79
23	Wide	-	-	-					Lymphoid	B-cell	80
24	Wide	-	-	-					Lymphoid	B-cell	82

25	Wide	-	-	-	Lymphoid	Cutaneous B-cell	52
26	Wide	-	-	-	Lymphoid	Hodgkin	27

Abbreviations: MPN, myeloproliferative neoplasms; NGS, next generation sequencing; VAF, variant allele frequency.

Supplementary Table 10. Studies evaluating the prevalence of JAK2 V617F mutation in the general population (sample size ≥100) and in individuals with abnormal hematological parameters.

General population								
Author, year	n	Selected population	Age (years)	% male	Mutation analysis method	N mutated *JAK2* V617F (%)	SNV VAF threshold (%)	Ref
Levine, 2005	269	General population from the International HapMap Consortium	NA	NA	MALDI-TOF MS	0 (0.0)	NA	1
Xu, 2007	3 935	Chinese Hospital Population	Mean 49	NA	PCR	37 (0.9)	0.25	2
Rapado, 2008	149	Healthy individuals	NA	NA	(qRT)-PCR	3 (2.0)	0.01	3
Martinaud, 2010	198	Blood donors and hospitalized patients for non-hematological reasons	Median 37	69.0	(qRT)-PCR	5 (2.5)	0.035	4
Nielsen, 2011	10 507	Copenhagen City Heart Study	Median 59	44.0	(qRT)-PCR	18 (0.2)	0.1	5
Weinberg, 2012	142	Smoking and nonsmoking patients admitted to the clinical ward	Median 61	63.4	(qRT)-PCR	38 (26.8)	1.0	6
Nielsen, 2013	49 488	Copenhagen General Population Study	Median 56	45.0	(qRT)-PCR	68 (0.1)	0.8	7
Genovese, 2014	12 380	Swedish cohort with individuals with schizophrenia (n=4 970), bipolar disorder (n=1 165) and controls (n=6 245)	Mean 55	53.3	WES	24 (0.19)	10.0	8
Jaiswal, 2014	17 182	22 population-based cohorts, mainly type 2 diabetes studies	Median 58	49.1	WES	31 (0.18)	3.5	9
McKerrell, 2015	4 219	Blood donors aged 17–70 and unselected individuals aged 60–98 years	Median 50-59	NA	Targeted sequencing	25 (0.59)	0.8	10
Van den Akker, 2016	864	Older individuals from the population-based Rotterdam Study (≥ 80 years, n=646) and Leiden Longevity study (≥ 89 years, n=218)	Mean 87	34.8	WES/WGS	4 (0.46)	NA	11
Acuna-Hidalgo, 2017	2 007	Population-based Nijmegen Biomedical Study	Mean 45	50.0	Targeted sequencing (error-corrected)	7 (0.35)	0.08	12
Buscarlet, 2017	2 530	Females without any known hematological disorder including family-based participants	Mean 69	0.0	Targeted sequencing	9 (0.36)	2.0	13
Zink, 2017	11 262	Icelanders participating in various disease	NA	NA	WGS	7 (0.06)	10.0	14

		projects at deCODE genetics						
Abelson, 2018*	676	Age- and sex- matched controls of pre-AML cases selected from EPIC study	Median 60	44.7	Targeted sequencing (error-corrected)	3 (0.44)	0.5	15
Desai, 2018*	181	Age-matched controls of AML cases selected from Women's Health Initiative cohort	Median 66	0.0	Targeted sequencing	1 (0.55)	1.0	16
Cordua, 2019	19 958	Danish General Suburban Population Study	Mean 56	45.6	Pooled multiplex droplet digital PCR	613 (3.1)	0.009	17
Cook, 2019	359	Adults ≥65 years from family practice clinics	Mean 80	34.3	Targeted sequencing	2 (0.56)	2.0	18
Cohorts selected for abnormal hematological parameters (hemoglobin or hematocrit)								
Kralovics, 2005	11	Secondary erythrocytosis	NA	NA	PCR	0 (0.0%)	NA	19
James, 2005	35	Secondary erythrocytosis	NA	NA	PCR	0 (0.0%)	NA	20
Tefferi, 2005	19	Secondary erythrocytosis with associated comorbidity and bone marrow inconsistent with primary myeloid disorder	Median 57	84.2	PCR	0 (0.0%)	NA	21
McClure, 2006	22	Individuals with high oxygen affinity Hb variant	NA	NA	PCR	0 (0.0%)	0.1	22
Percy, 2006	63	Individuals with a raised red cell mass and Hct without identifiable secondary cause of erythrocytosis	Mean 35	NA	PCR	1 (1.6%)	1-2	23
Bianchi, 2007	83	Repeat blood donors with Hct >47% (males) or >42% (females)	NA	66.3	PCR	4 (4.8%)	0.01	24
Strobbe, 2007	1	Outpatient clinic patients with elevated Hb level on more than two occasion, not meeting WHO criteria for PV.	82	100	(qRT)-PCR	1 (100.0%)	0.8	25
Tagariello, 2009	103	Blood donors with Hct >50% (males) or >46% (females)	Mean 44	81.6	Amplification refractory mutation	1 (1.0%)	1-2	26

					system			
Magnussen, 2013	46	Blood donors with Hb above 18.5 g/dl (male) and 16.5 g/dL (female) on two succeeding donations	Mean 52	28.2	PCR	1 (2.2%)	NA	27
Kamaruzzaman, 2018	45	Blood donors with erythrocytosis (Hb >17/16 g/dL), leukocytosis (>12x10^9/L) and thrombocytosis (>450x10^9/L)	NA	NA	PCR	0 (0.0%)	NA	28

Abbreviations: AML, acute myeloid leukemia; Hb, hemoglobin; Hct, hematocrit; Indels, insertions and deletions; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; NA, not available; NGS, next generation sequencing; (qRT)-PCR, (quantitative real-time) polymerase chain reaction; PV, polycythemia vera; SNV, single-nucleotide variant; WES, whole-exome sequencing; WGS, whole-genome sequencing. *Only the control cases from these studies are included in this table.

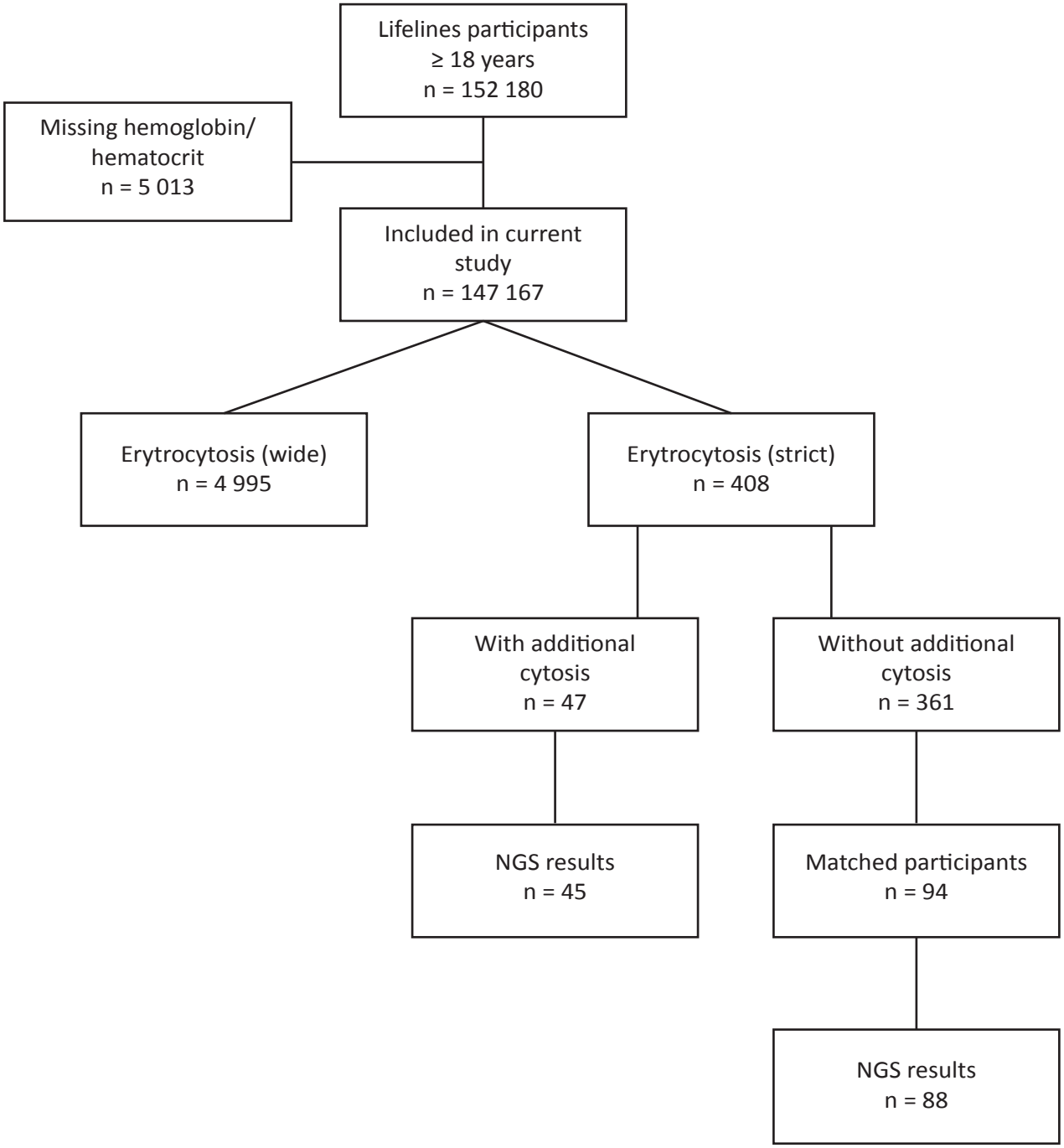
References

1 Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell.* 2005;7(4):387-397.
2 Xu X, Zhang Q, Luo J, et al. JAK2V617F: prevalence in a large Chinese hospital population. *Blood.* 2007; 109(1):339-342.
3 Rapado I, Albizua E, Ayala R, et al. Validity test study of JAK2 V617F and allele burden quantification in the diagnosis of myeloproliferative diseases. *Ann Hematol.* 2008;87(9):741-749.
4 Martinaud C, Brisou P, Mozziconacci MJ. Is the JAK2(V617F) mutation detectable in healthy volunteers? *Am J Hematol.* 2010;85(4):287-288.
5 Nielsen C, Birgens HS, Nordestgaard BG, Bojesen SE. Diagnostic value of JAK2 V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. *Br J Haematol.* 2013;160(1):70-79.
6 Weinberg I, Borohovitz A, Krichevsky S, et al. Janus Kinase V617F mutation in cigarette smokers. *Am J Hematol.* 2012;87(1):5-8.
7 Nielsen C, Birgens HS, Nordestgaard BG, Kjaer L, Bojesen SE. The JAK2 V617F somatic mutation, mortality and cancer risk in the general population. *Haematologica.* 2011;96(3):450-453.
8 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.
9 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.
10 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.
11 van den Akker E, 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.
12 Acuna-Hidalgo R, Sengul H, Steehouwer M, et al. Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. *Am J Hum Genet.* 2017;101(1):50-64.

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 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.
15 Abelson S, Collord G, Ng SWK, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature.* 2018 Jul;559(7714):400-404.
16 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.
17 Cordua S, Kjaer L, Skov V, et al. Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood.* 2019;134(5):469-479.
18 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.
19 Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005;352(17):1779-1790.
20 James C, Ugo V, Le Couédic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature.* 2005;434(7037):1144-1148.
21 Tefferi A, Sirhan S, Lasho TL, et al. Concomitant neutrophil JAK2 mutation screening and PRV-1 expression analysis in myeloproliferative disorders and secondary polycythaemia. *Br J Haematol.* 2005;131(2):166-171.
22 McClure RF, Hoyer JD, Mai M. The JAK2 V617F mutation is absent in patients with erythrocytosis due to high oxygen affinity hemoglobin variants. *Hemoglobin.* 2006;30(4):487-489.
23 Percy MJ, Jones FGC, Green AR, Reilly JT, McMullin MF. The incidence of the JAK2 V617F mutation in patients with idiopathic erythrocytosis. *Haematologica.* 2006;91:413–414.
24 Bianchi P, Fermo E, Mozzi F, et al. Presence of JAK2 V617F Mutation in peripheral blood of blood donors with upper-limit hematocrit and platelet values. *Blood.* 2007;110(11):2533.
25 Strobbe L, Lestrade P, Hermans MH, Fijnheer R. Detection of the JAK2V617F mutation in patients with slightly elevated platelets or hemoglobin without a secondary cause. *Ann Hematol.* 2007;86(11):801-803.
26 Tagariello G, Di Gaetano R, Sartori R, et al. The JAK2(V617F) tyrosine kinase mutation in blood donors with upper-limit haematocrit levels. *Blood Transfus.* 2009;7(2):111-116.
27 Magnussen K, Hasselbalch HC, Ullum H, Bjerrum OW. Characterization of blood donors with high haemoglobin concentration. *Vox Sang.* 2013;104(2):110-114.
28 Kamaruzzaman SF, Noor NH, Yusoff SM, et al. Detection of JAK2 V617F mutation among donors with erythrocytosis. *J. Clin. Heal. Sci.* 2018;3(2):19-25.

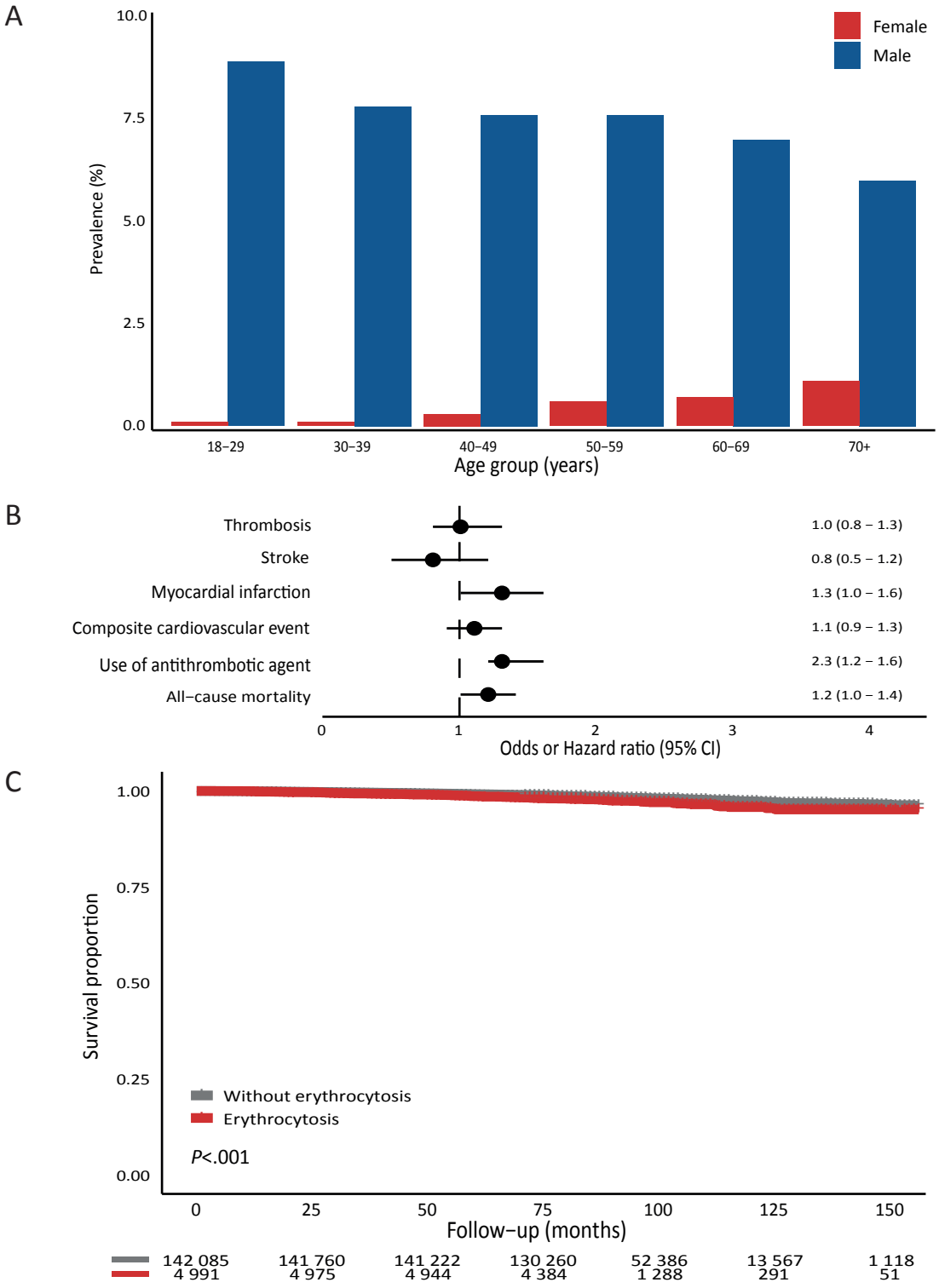
Supplementary Figure 1. A flowchart of the study.

Individuals having erythrocytosis (strict criteria) with concurrent leuko- or thrombocytosis were 1:2 matched with individuals having isolated erythrocytosis for age, sex, body mass index, smoking status and number of medications used. Available DNA samples were used for next-generation sequencing. Erythrocytosis strict: hemoglobin concentration >18.5 g/dL or hematocrit ≥52% in males and hemoglobin concentration >16.5 g/dL or hematocrit ≥48% in females. Erythrocytosis (wide): hemoglobin concentration >16.5 g/dL or hematocrit >49% in males and hemoglobin concentration >16.0 g/dL or hematocrit >48% in females.



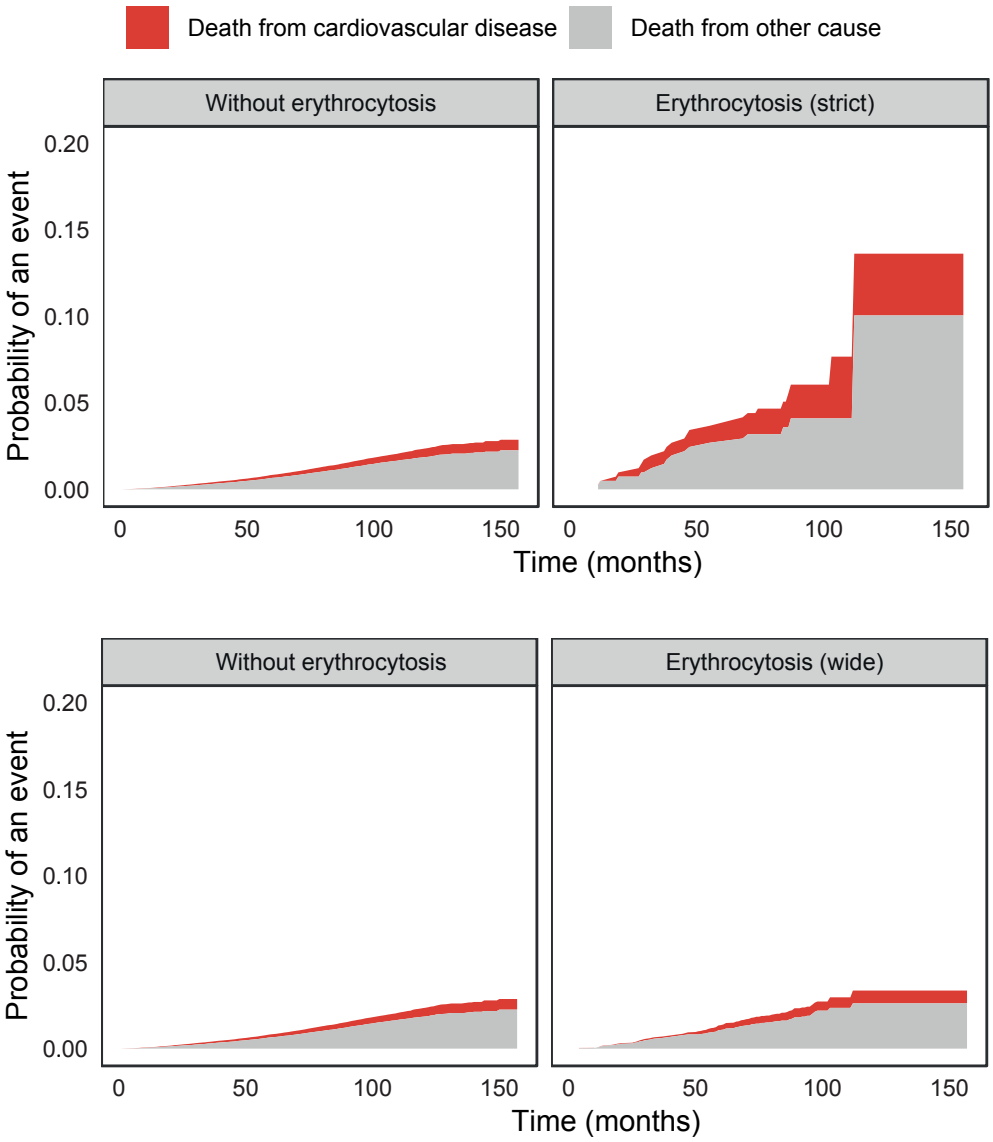
Supplementary Figure 2. Prevalence of erythrocytosis (wide criteria) as function of sex and age and the association of erythrocytosis with cardiovascular disease and survival.

(A) Prevalence of erythrocytosis according to sex and age categories. (B) Forest plot for the risk of cardiovascular disease and all-cause mortality. Logistic regression analyses and cox proportional hazards regression included age, sex, body mass index, systolic and diastolic blood pressure, smoking status, medical history of diabetes, number of medications used, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, a family history of cardiovascular disease and low-density lipoprotein cholesterol as covariates. Absence of erythrocytosis was used as a reference. Circles indicate the odds/hazard ratio, with horizontal lines corresponding to 95% confidence intervals. (C) Kaplan-Meier curve for overall survival, stratified according to the presence of erythrocytosis.



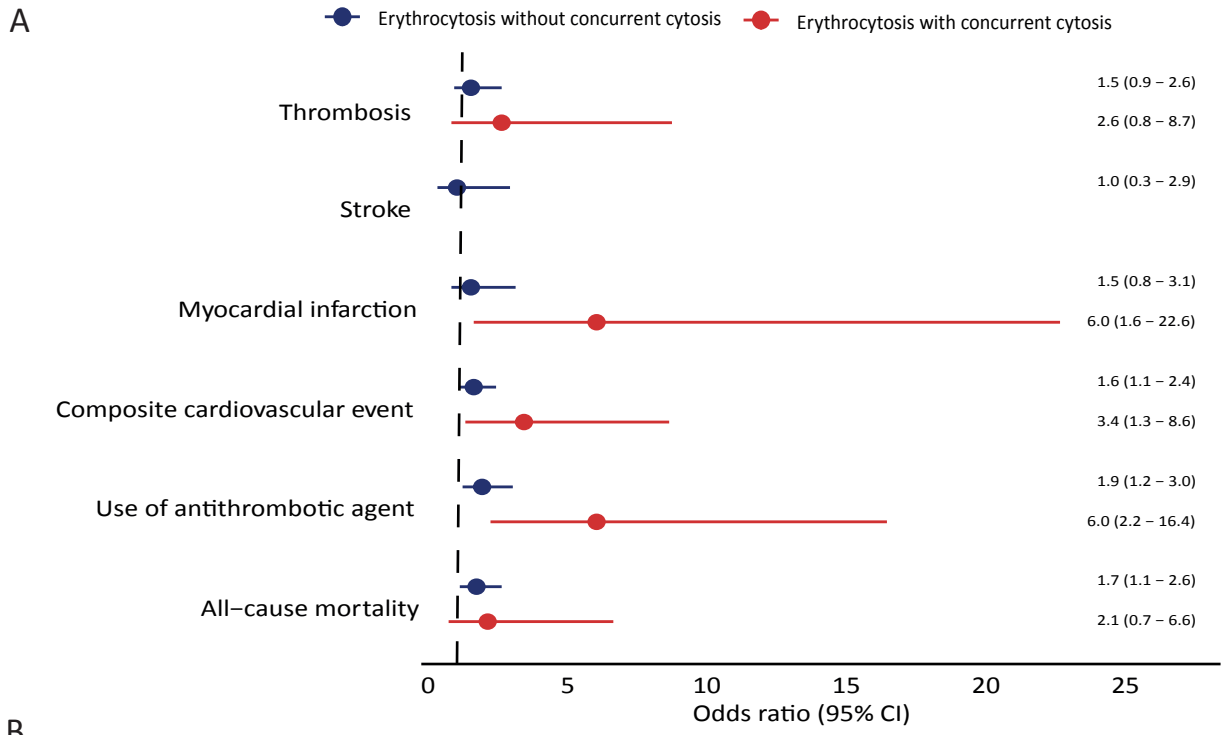
Supplementary Figure 3. Cumulative incidence graphs for mortality from cardiovascular disease, according to the presence of erythrocytosis.

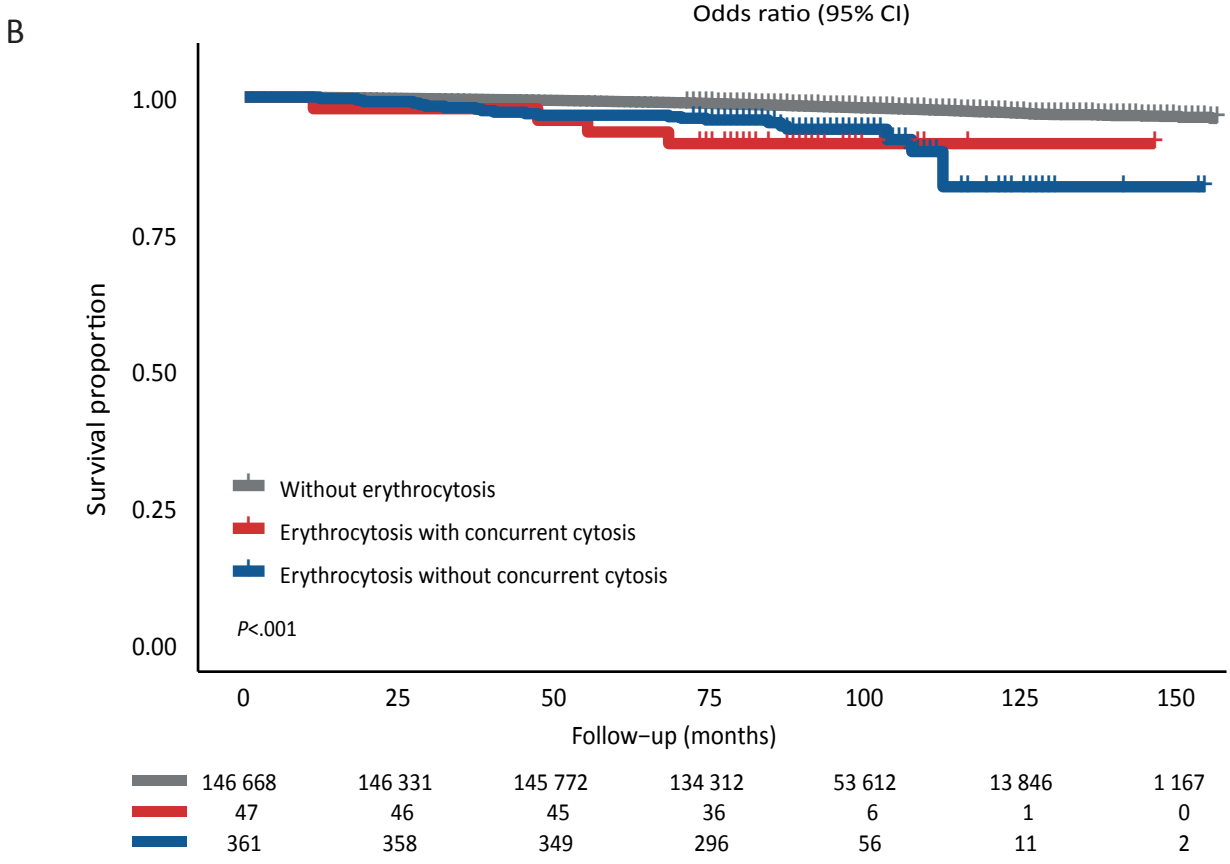
Data on reported primary cause of death were obtained by linkage to the national registry of death statistics. Results for this analysis are based on calculations by the authors using non-public microdata from Statistics Netherlands. Colors indicate death from cardiovascular disease (red) or other causes (grey).



Supplementary Figure 4. The association of erythrocytosis with and without concurrent cytosis with cardiovascular disease and survival.

(A) Forest plot for the risk of cardiovascular disease and all-cause mortality. Logistic regression analyses and cox proportional hazards regression included age, sex, body mass index, systolic and diastolic blood pressure, smoking status, medical history of diabetes, number of medications used, androgen drug use, drugs for obstructive airway disease, antihypertensive drug use, a family history of cardiovascular disease and low-density lipoprotein cholesterol as covariates. Absence of erythrocytosis was used as a reference. Circles indicate the odds/hazard ratio, with horizontal lines corresponding to 95% confidence intervals. (B) Kaplan-Meier curve for overall survival, stratified according to the presence of erythrocytosis with or without concurrent cytosis.





Supplementary Figure 5. Lollipop plots showing the distribution of mutations in *BCOR* and *BCORL1*.

Truncating mutations include nonsense, frameshift and splice site mutations.