Supplementary Appendix

Supplement to: van Zeventer IA*, de Graaf AO*, Koorenhof-Scheele TN, van der Reijden BA, van der Klauw MM, Dinmohamed AG, Diepstra A, Schuringa JJ, Malcovati L, Huls G* and Jansen JH*.

Monocytosis and its association with clonal hematopoiesis in community-dwelling individuals: data from a prospective population-based cohort study.

Supplementary Methods........................................................................................................................................1

Supplementary Tables.........................................................................................................................................3
Supplementary Table 1......................................................................................................................................3
Supplementary Table 2......................................................................................................................................4

Supplementary Figures....................................................................................................................................5
Supplementary Figure 1...................................................................................................................................5
Supplementary Figure 2...................................................................................................................................6
Supplementary Figure 3...................................................................................................................................7
Supplementary Figure 4...................................................................................................................................8

References..........................................................................................................................................................9
Supplementary Methods

Sample collection and biochemical measurements
Withdrawal of peripheral blood samples was done after an overnight fast. Thereafter, samples were directly processed for routine clinical chemistry assays at the University Medical Center Groningen. Complete and differential blood counts were determined on a XE2100-system (Sysmex, Japan). DNA was isolated from aliquots of peripheral blood specimens and stored in the Lifelines biobank for future use.

Definition of peripheral cytopenia or cytosis
The following cut-offs were used for definition of peripheral cytopenias: anemia, hemoglobin concentration <13.0 g/dL in men and < 12.0 g/dL in women; thrombocytopenia, platelet counts <150 x 10^9/L; and neutropenia, neutrophil counts <1.8 x 10^9/L. For cytosis, the following cut-offs were used: erythrocytosis, hemoglobin concentration >16.5 g/dL or hematocrit ≥48% in women or hemoglobin concentration >18.5 g/dL or hematocrit ≥52% in men; thrombocytosis, platelet count >400 x 10^9/L; leukocytosis, white blood cell count >10 x 10^9/L.

Design of targeted sequencing panel
The molecular inversion probe (MIP) design and protocol are based on previously described methods. Modifications of this protocol are described below. Single molecule tagged molecular inversion probes (smMIPs) were designed for all target regions (Supplementary Table S1) using previously described procedures, in a tiling manner preferentially covering all target nucleotides with two smMIPs targeting both DNA strands independently. MIPs were designed to capture a target of 98 nucleotides and contain a single molecule tag consisting of a stretch of 10N nucleotides inserted between the backbone and ligation probe sequence. In case it was unavoidable to design smMIPs without common single nucleotide polymorphism variants in the respective capture arms, smMIPs were designed recognizing both alleles using degenerate nucleotides.

Library preparation and sequencing
Isolated DNA was sonicated using a Covaris R230 Focused-ultrasonicator (Covaris, Inc. MA, USA) with a standard protocol to obtain 400-bp fragments. Subsequently, library preparation and sequencing were performed as previously described with the following adaptations. A volume of 7 μL containing 100 ng of genomic DNA was used as input per smMIP capture. Subsequently, 20 μL from the exonuclease-treated capture mixture was used for PCR in a total volume of 50 μL. The gDNA:smMIP ratio was set to 800:1 for high-quality blood-derived gDNA. The final pooled smMIP libraries were denaturated and diluted to a concentration of 1.1 nM. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA) according to the manufacturer’s protocol (300 cycles NovaSeq 6000 S1 reagent kit), resulting in 2 × 150 bp paired-end reads.

Data processing and variant calling
Bcl to fastq conversion and demultiplexing of barcoded reads was performed automatically and data was uploaded to a server running commercial analysis software (Sequence Pilot version 5.2.0, build 502 (JSI medical systems, Ettenheim, Germany). Single-molecule–directed assembly of duplicate reads to generate consensus reads after alignment was performed using the same software (Sequence Pilot, JSI medical systems, Ettenheim, Germany). Proper identification of sequence variants greatly relies on the settings used for variant calling, such as the minimal number of mutant unique smMIPs (>5 in our settings) and the minimal variant allele frequency (≥1%).

The following settings were used for variant calling using Sequence Pilot:
(i) Required Coverage/Min abs. cov., 20 combined
(ii) Mutations/Min abs. cov., 5 combined
(iii) (iii) Min % cov., 1% per dir

After variant calling using the commercial software, all variants were manually inspected and curated using a cut-off of 10 reads for minimal absolute coverage and in-house databases for recurrent artifacts, polymorphisms and mutations. Furthermore, Alamut Visual v2.10 (Interactive Biosoftware, 5 Rouen, France) and publicly available databases such as dbSNP, COSMIC, ClinVar, gnomAD, ESP were used to identify rare polymorphisms, which were excluded.
**Isolation and sequencing of white blood cell subfractions**

After blood withdrawal, cells were processed within 24 hours. Full blood was mixed 1:1 with FCS containing 20% DMSO. Samples were frozen in a controlled manner and stored in liquid nitrogen in the Lifelines biobank until further use. Cells were thawed with DNaseI in FSC containing Heparin and MgSO4. After washing twice with FACS buffer (PBS containing 1% FCS and 2mM EDTA), cells were stained in FACS buffer with CD45 - BUV395 (BD biosciences, 563792), CD15 - APCFire750 (Biolegend, 323041), CD14 – fitc (Biolegend, 325604), CD19 – ECD (Beckman Coulter, A07770), CD3 – PE (Beckman Coulter, A07747) and LIVE/DEAD fixable aqua dead cell stain (Invitrogen, L34965). Cells were measured on the cellsorter (BD FACSAria SORP) and viable Monocytes (CD14), Granulocytes (CD15) and T-lymphocytes (CD3) were selected. DNA was isolated from sorted and bulk (sample before sort) and amplified in duplo using the repli-G Midi kit (Qiagen, 150045).

**ICD-O codes used to identify myeloid malignancies**

The Netherlands Cancer Registry (NCR) codes disease morphology according to the International Classification of Diseases for Oncology (ICD-O). The ICD-O first edition was used for case ascertainment until 1992, and from 1993-2000 the second edition (ICD-O-2) was used. The Lifelines study population was included starting in 2006. As a result, the third edition of the ICD-O (ICD-O-3, used in the NCR 2001 to 2011) and the updated ICD-O-3 (used in the NCR from 2012 onwards) were used to identify incident cases of hematological malignancies, including myeloid malignancies. We used ICD-O code 9945 to identify cases of chronic myelomonocytic leukemia.

**R packages for analysis and visualization**

Cox proportional hazards and competing risk regression analyses as well as visualization of survival were performed using R packages 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). The R package compareGroups (https://github.com/isubirana/compareGroups) was used for construction of tables with characteristics. All remaining figures were generated using ggplot2 (https://github.com/tidyverse/ggplot2).
Supplementary Table 1. Overview of genes and regions in the sequencing panel.

| 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                |
| ETVN1  | NM_018638            | ENST00000266517              | 3 (partially)         | codon 243-244                   |
| EZH2   | NM_004456            | ENST00000320356              | 2-20 (all coding exons)| all coding exons                |
| FLT3_835| NM_004119            | ENST00000241453              | 20 (partially)        | codon 835-842                   |
| IDH1   | NM_005896            | ENST00000415913              | 4 (partially)         | codon 132                       |
| IDH2   | NM_002168            | ENST00000330062              | 4 (partially)         | codon 140, 172                  |
| JAK2   | NM_004972            | ENST00000381652              | 12, 14 (partially)    | codon 617 and exon 12           |
| KIT    | NM_000222            | ENST00000288135              | 8 (partially), 17 (partially)| codon 816, 419                 |
| KRAS   | NM_004985            | ENST00000256078              | 2-3 (partially)       | a.o. codon 12, 13, 61           |
| MPL    | NM_005373            | ENST00000372470              | 10 (partially)        | codon 515, 505                  |
| MYD88  | NM_002468.4          | ENST00000417037              | 4-5 (partially)       | codon 265 and 232               |
| NOTCH1 | NM_017617.4          | ENST00000277541              | 34 (partially)        | codon 2514                      |
| NPM1   | NM_002520            | ENST00000517671              | 11 (partially)        | codon 288-290                   |
| NRAS   | NM_002524            | ENST00000369535              | 2-3 (partially)       | a.o. codon 12, 13, 61           |
| RUNX1  | NM_001754            | ENST00000437180              | 2-9 (all coding exons)| all coding exons                |
| SETBP1 | NM_015559            | ENST00000282030              | 4 (partially)         | codon 850-910                   |
| SF3B1  | NM_012433            | ENST00000335508              | 13-16                 | codon 575-790                   |
| SRSF2  | NM_003016            | ENST00000392485              | 1 (partially)         | codon 95, 96                    |
| TET2   | NM_001127208         | ENST00000380013              | 3-11 (all coding exons)| all coding exons                |
| TP53   | NM_000546            | ENST00000269305              | 2-11 (all coding exons)| all coding exons                |
| U2AF1  | NM_006758            | ENST00000291552              | 2, 6 (partially)      | codon 34, 157                   |
| WT1    | NM_024426            | ENST00000332351              | 7, 9                  | exon 7 and 9                    |
Supplementary Table 2. Characteristics and peripheral blood counts of cases with monocytosis (n=167) and controls (n=501), stratified by the presence of clonal hematopoiesis.

| Characteristics                        | Controls without CH | Controls with CH | Monocytosis without CH | Monocytosis with CH |
|----------------------------------------|---------------------|------------------|------------------------|---------------------|
|                                       | N=323               | N=178            | N=82                   | N=85                |
| Age (years)                           | 67.0 [62.0;71.0]    | 69.0 [64.2;73.0] | 0.001 66.0 [62.0;70.0] | 70.0 [66.0;75.0]    | 0.001 |
| Male sex                               | 136 (76.4%)         | 233 (72.1%)      | 0.351 59 (72.0%)       | 64 (75.3%)          | 0.753 |
| Monocyte count (10⁹/L)                 | 0.53 (0.16)         | 0.54 (0.14)      | 0.398 1.10 (0.11)      | 1.14 (0.12)         | 0.023 |
| WBC count (10⁹/L)                      | 6.09 (1.59)         | 6.06 (1.50)      | 0.807 8.58 (1.65)      | 8.61 (1.80)         | 0.896 |
| Neutrophil count (10⁹/L)               | 3.33 (1.21)         | 3.44 (1.23)      | 0.350 4.91 (1.46)      | 4.85 (1.56)         | 0.803 |
| Basophil count (10⁹/L)                 | 0.03 (0.02)         | 0.03 (0.02)      | 0.214 0.04 (0.02)      | 0.04 (0.03)         | 0.168 |
| Eosinophil count (10⁹/L)               | 0.21 (0.14)         | 0.20 (0.15)      | 0.502 0.23 (0.13)      | 0.27 (0.16)         | 0.093 |
| Lymphocyte count (10⁹/L)               | 1.99 (0.59)         | 1.85 (0.59)      | 0.009 2.30 (0.71)      | 2.30 (0.82)         | 0.937 |
| Hemoglobin concentration (g/dL)        | 14.6 (1.25)         | 14.4 (1.25)      | 0.104 14.6 (1.11)      | 14.6 (1.53)         | 0.951 |
| Erythrocyte count (10⁹/L)              | 4.82 (0.38)         | 4.73 (0.41)      | 0.014 4.80 (0.37)      | 4.75 (0.49)         | 0.435 |
| Hematocrit (L/L)                       | 0.44 (0.03)         | 0.43 (0.03)      | 0.299 0.44 (0.03)      | 0.44 (0.04)         | 0.794 |
| Platelet count (10⁹/L)                 | 232 (51.2)          | 224 (52.2)       | 0.097 257 (63.1)       | 251 (66.9)          | 0.541 |
| MCV (fL)                               | 90.4 (4.30)         | 91.5 (3.96)      | 0.003 91.6 (4.58)      | 92.3 (3.71)         | 0.281 |
| Concurrent cytopenia*                  | 39 (12.1%)          | 27 (15.2%)       | 0.40 5 (6.10%)         | 10 (11.8%)          | 0.313 |
| Concurrent cytosis§                    | 7 (2.2%)            | 6 (3.4%)         | 0.558 19 (23.2%)       | 19 (22.4%)          | 1.000 |

MCV, mean corpuscular volume; CH, clonal hematopoiesis; WBC, white blood cell. *A concurrent cytopenia was defined as follows: anemia, hemoglobin concentration <12.0 g/dL in women or <13.0 g/dL in men; thrombocytopenia, platelet count <150 x 10⁹/L; neutropenia, absolute neutrophil count <1.8 x 10⁹/L. §A concurrent cytosis was defined as follows: erythrocytosis, hemoglobin concentration >16.5 g/dL or hematocrit ≥48% in women or hemoglobin concentration >18.5 g/dL or hematocrit ≥52% in men; thrombocytosis, platelet count >400 x 10⁹/L; leukocytosis, white blood cell count >10 x 10⁹/L. 1P-value for comparison of controls with and without CH. 2P-value for the comparison of monocytosis cases with and without CH.
Supplementary Table 2. Characteristics and peripheral blood counts of cases with persistent or corrected monocytosis or loss to follow-up.

|                         | Corrected  | No follow-up | Persistent | P-value | N     |
|-------------------------|------------|--------------|------------|---------|-------|
|                         | N=72       | N=65         | N=30       |         |       |
| Age (years)             | 66.5 [62.8;70.2] | 70.0 [65.0;75.0] | 67.0 [62.0;70.0] | <0.001  | 16644 |
| Male sex                | 54 (75.0%) | 45 (69.2%)   | 24 (80.0%) | <0.001  | 16644 |
| Monocyte count (10^9/L) | 1.10 (0.11) | 1.13 (0.13)  | 1.14 (0.12) | 0.000   | 16644 |
| WBC count (10^9/L)      | 8.42 (1.78) | 8.66 (1.60)  | 8.89 (1.83) | <0.001  | 16643 |
| Neutrophil count (10^9/L)| 4.68 (1.56) | 5.04 (1.46)  | 5.01 (1.47) | <0.001  | 16644 |
| Basophil count (10^9/L) | 0.04 (0.03) | 0.04 (0.03)  | 0.05 (0.03) | <0.001  | 16644 |
| Eosinophil count (10^9/L)| 0.24 (0.13) | 0.25 (0.16)  | 0.29 (0.15) | <0.001  | 16644 |
| Lymphocyte count (10^9/L)| 2.36 (0.79) | 2.19 (0.76)  | 2.41 (0.73) | <0.001  | 16644 |
| Hemoglobin concentration (g/dL) | 14.8 (1.15) | 14.3 (1.59)  | 14.9 (1.06) | <0.001  | 16643 |
| Erythrocyte count (10^12/L) | 4.85 (0.42) | 4.70 (0.48)  | 4.75 (0.33) | <0.001  | 16643 |
| Hematocrit (L/L)        | 0.44 (0.03) | 0.43 (0.04)  | 0.44 (0.03) | <0.001  | 16643 |
| Platelet count (10^9/L) | 250 (61.5)  | 256 (71.3)   | 261 (59.6)  | <0.001  | 16629 |
| MCV (fL)                | 91.7 (3.92) | 91.7 (4.61)  | 93.3 (3.50) | <0.001  | 16643 |
| Concurrent cytopenia*    | 3 (4.17%)   | 9 (13.8%)    | 3 (10.0%)   | 0.107   | 16630 |
| Concurrent cytosis$      | 14 (19.4%)  | 17 (26.2%)   | 7 (23.3%)   | <0.001  | 16629 |
| hsCRP (mg/L)            | 2.90 [2.00;7.80] | 4.15 [1.42;8.00] | 2.45 [1.40;4.05] | <0.001  | 5599 |
| Deceased                | 7 (9.72%)   | 25 (38.5%)   | 4 (13.3%)   | <0.001  | 16644 |
| Number of medications used# | 3.00 [1.00;5.00] | 4.00 [1.00;6.00] | 3.00 [1.25;4.75] | <0.001  | 16644 |

MCV, mean corpuscular volume; CH, clonal hematopoiesis; hsCRP, high sensitive CRP; WBC, white blood cell. *A concurrent cytopenia was defined as follows: anemia, hemoglobin concentration <12.0 g/dL in women or <13.0 g/dL in men; thrombocytopenia, platelet count <150 x 10^9/L; neutropenia, absolute neutrophil count <1.8 x 10^9/L. $A concurrent cytosis was defined as follows: erythrocytosis, hemoglobin concentration >16.5 g/dL or hematocrit ≥48% in women or hemoglobin concentration >18.5 g/dL or hematocrit ≥52% in men; thrombocytosis, platelet count >400 x 10^9/L; leukocytosis, white blood cell count >10 x 10^9/L. #As a proxy for comorbidity.
Supplementary Figure 1. Coverage across all included samples. Graphs below show the number of aligned consensus reads (A) and the number of aligned reads (B) for all genes included in the panel, for the entire sequenced cohort (n=668). Columns and error bars indicate median and interquartile range.
Supplementary Figure 2. Mutational spectra for the entire cohort. The graphs below show (A) the age-related emergence of clonal hematopoiesis, (B) the number of detected somatic variants and (C) the number of affected genes in the entire cohort with next-generation sequencing data (n=668). In addition, (D) the frequency of all somatic variants detected and (E) the distribution in variant allele frequency for all detected somatic variants are shown.
Supplementary Figure 3. Variant allele frequencies for detected variants in the monocytosis and control cohort.
Individual data points correspond to somatic variants detected in the monocytosis (green, n=167) and control (blue, n=501) cohort. The median variant allele frequency for the variants detected in each gene is indicated.
Supplementary Figure 4. Mutational spectrum restricted to "CHIP". Shown below is the mutational spectrum detected in the entire cohort with next-generation sequencing data (n=668) when restricted to n=232 variants detected at ≥2% variant allele frequency (VAF). The graphs below show (A) the frequency of all somatic variants detected and (B) the distribution in variant allele frequency for all detected somatic variants as well as (C) the number of detected somatic variants per individual.
Supplementary Figure 5. Case-control comparison of mutational spectrum restricted to “CHI” \(^9\) \(^\text{(I)}\). Shown below are analyses for the comparison of mutational spectrum between monocytosis cases and controls, restricted to variants detected ≥2% VAF. (A) Prevalence of CH among all individuals with monocytosis (n=167) as compared to 1:3 matched controls (n=501). (B) Prevalence of CH according to age for individuals with monocytosis and controls. (C) Violin plot showing the distribution in the number of mutated genes for individuals with CH in the monocytosis (green) and control (blue) cohort. Gray rectangles indicate the median number. (D) Mutational landscape for the control (blue, top) and monocytosis (green, bottom) cohort. A darker shade indicates multiple mutations in the same gene.
Supplementary Figure 6. Case-control comparison of mutational spectrum restricted to “CHIP” (II). Shown below are analyses for the comparison of mutational spectrum between monocytosis cases and controls, restricted to variants detected ≥2% VAF. (A) Pyramid plot indicating the proportion of individuals with detected gene mutations within the monocytosis (green) and control (blue) cohort. The category of spliceosome mutations includes SF3B1, SRSF2 and U2AF1. The proportion of individuals carrying the gene mutation is given. (B) Bar plot showing the proportion of monocytosis cases (green, top) and controls (blue, bottom) with mutational spectra confined to mutated DNMT3A, TET2 or ASXL1, or multiple mutated genes. The category ‘other’ denotes isolated gene mutations other than DNMT3A, TET2 or ASXL1. The proportion of individuals for each category is given. (C) Highest detected VAF according to mutational spectrum for monocytosis cases (green) and control (blue). Individuals were classified as carrying CH confined to mutated DNMT3A, TET2 or ASXL1 (isolated DTA), CH involving multiple mutated genes and other isolated gene mutations (other). Boxes represent median, first and third quartiles. DTA, DNMT3A, TET2 or ASXL1; VAF, variant allele frequency.
Supplementary Figure 7. Mutational spectra for individuals according to the stability of monocytosis over time.

Subgroups of individuals with corrected monocytosis (n=72), without follow-up (n=65), and with stable monocytosis (n=30) are compared to their respective 1:3 matched controls. The proportion of individuals without clonal hematopoiesis (CH), with CH restricted to mutated DNMT3A, TET2 or ASXL1 (DTA) and with other mutational spectra is shown.
References

1. 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. Clin Epidemiol. 2018;10:981-989.
2. Slagter SN, van Vliet-Ostaptchouk JV, van Beek AP, et al. Health-Related Quality of Life in Relation to Obesity Grade, Type 2 Diabetes, Metabolic Syndrome and Inflammation. PLoS One 2015;10:e0140599.
3. Alberti KGMM, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation. 2009;120:1640-5.
4. O’Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;338:1619-1622.
5. Boyle EA, O’Roak BJ, Martin BK, Kumar A, Shendure J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. Bioinformatics 2014;30:2670-2672.
6. Eijkelenboom A, Kamping EJ, Kastner-van Raaij AW, et al. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. J Mol Diagn 2016;18:851-863.
7. Neveling K, Mensenkamp AR, Derks R, et al. BRCA Testing by Single-Molecule Molecular Inversion Probes. Clin Chem 2017;63:503-512.
8. Fritz AG, Percy C, Jack A, et al. International classification of diseases for oncology 3rd edition. World Health Organization Geneva 2000.
9. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood 2015;126:9-16.