Structural aberrations are associated with poor survival in patients with clonal cytopenia of undetermined significance

Stine U. Mikkelsen,[1,2] Setareh Safavi,[3] Konstantinos Dimopoulos,[4] Colm J. O'Rourke,[2] Mette K. Andersen,[3] Mette S. Holm,[5] Claus W. Marcher,[6] Jesper B. Andersen,[2] Jakob W. Hansen[1,2,7#] and Kirsten Grønbæk[1,2,7#]

[1]Department of Hematology, Rigshospitalet, Copenhagen; [2]Biotech Research and Innovation Center (BRIC), Department of Health and Medical Sciences, University of Copenhagen, Copenhagen; [3]Department of Clinical Genetics, Rigshospitalet, Copenhagen; [4]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen; [5]Department of Hematology, Aarhus University Hospital, Aarhus; [6]Department of Hematology, Odense University Hospital, Odense and [7]DanStem, University of Copenhagen, Copenhagen, Denmark

[#]JWH and KG contributed equally as co-senior authors.

Correspondence:

KIRSTEN GROENBAEK - kirsten.groenbaek@regionh.dk

doi:10.3324/haematol.2020.263319

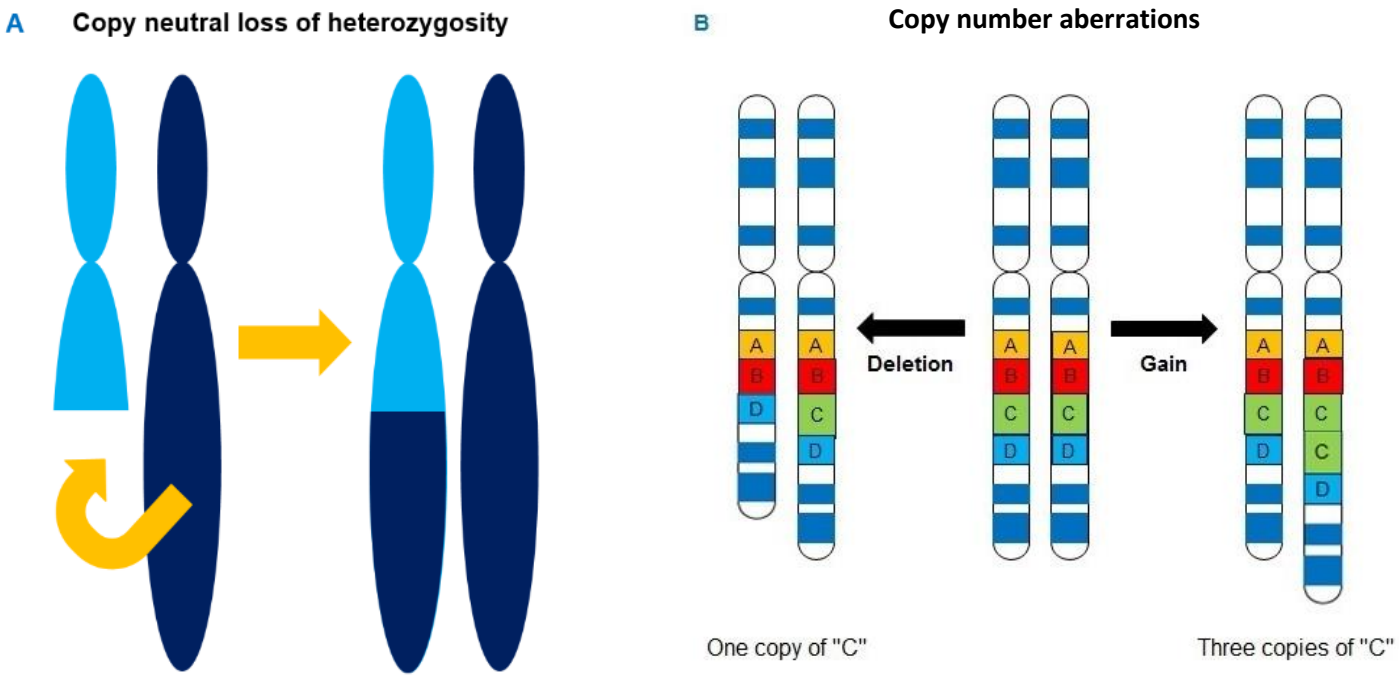
Supplementary Information

Table of Contents

Supplementary Information

Figure S1. Mechanisms of copy neutral loss of heterozygosity and copy number aberrations.

(A) Copy neutral loss of heterozygosity (CNLOH) results from uniparental disomy which describes the phenomenon whereby one homologous chromosome or chromosome region is replaced by an identical copy of the other homologous chromosome. This leads to homozygosity of the genomic region affected while no net change in the copy number occurs. The karyotype, thus, appears normal in the microscope. CNLOH can be non-clonal (constitutional or early embryonic) or clonal (somatically acquired). Potential mechanisms of acquired CNLOH include mitotic recombination or gene conversion, e.g., during DNA double-strand breaks, or trisomy or monosomy rescue in somatic cells. CNLOH can lead to deletion or loss of function of a tumor suppressor gene, or duplication of an activating mutation in an oncogene. Acquired CNLOH is common in cancer. Aside from pinpointing regions harboring possible homozygous mutations, CNLOH by itself appear to portend poor prognosis for patients with myeloid malignancies (Yeung CCS, McElhone S, Chen XY, et al. *Mod. Pathol.* 2018;31(4):569–580). (B) Copy number aberrations describe gains or deletions of whole chromosomes or chromosome regions. Changes in copy number can arise both meiotically (non-clonal) and somatically (clonal) by two general mechanisms; homologous recombination and nonhomologous recombination during repair of DNA breaks and gaps. In DNA repair by homologous recombination, another identical DNA sequence in the sister chromatid or homologous chromosome is used to repair the damaged DNA sequence. Chromosomal structural change can occur by homologous recombination because genomes have tracts of segmental duplications and DNA repair might utilize homologous sequences in different chromosomal positions. Changes in copy number might change the expression levels of genes included in the chromosome region affected. As with CNLOH, this can lead to the development and progression of cancer, e.g., by duplication of an activating mutation in an oncogene, or deletion or loss of function of a tumor suppressor gene in particular if there is a mutation in the allelic gene (Hastings PJ, Lupski JR, Rosenberg SM, Ira G. *Nat. Rev. Genet.* 2009;10(8):551–564).



Adapted from O'Keefe, McDevitt, and Maciejewski. Blood. 2010; 115(14): 2731–2739 and neurowiki2013.wikidot.com.

Supplementary Methods

Patients and samples

Patients referred with persistent cytopenia to the Department of Hematology, Rigshospitalet/University Hospital of Copenhagen, between 2009-2017 and nationwide in Denmark between 2014-2017 were eligible.

Inclusion criteria

- Persistence of cytopenia for >6 months
- Signed informed consent form

Exclusion criteria

- Presence of common causes of cytopenia, e.g., vitamin deficiency, HIV
- Bone marrow morphology diagnostic of myelodysplastic syndrome (MDS) or other disorder
- Abnormal cytogenetics (except from loss of the Y chromosome (LOY) which was accepted)

Cytopenia was defined as hemoglobin <120 g/L (women)/<130 g/L (men), platelet count <150x10^9/L, or neutrophil count <1.8x10^9/L.

Bone marrow morphological investigation, G-band karyotyping, and peripheral blood tests (complete blood count, differential count, serum chemistries) were performed at first visit and in case of follow-up. Bone marrow aspirate and peripheral blood were collected at the same time points for biobanking.

In 11 patients initially included in the study, karyotyping failed or was not performed.

Morphological examination and G-band karyotyping

The procedure for the morphological examination at inclusion and in case of a follow-up investigation was described in detail by Hansen et al.[1] In brief, the three hematopoietic lineages were assessed for the presence and degree of dysplasia, ring sideroblasts, and the percentage of blasts by an experienced hematopathologist. At least 500 nucleated bone marrow cells were counted. The results were reported in accordance with the criteria of the 2008 WHO classification of tumors of hematopoietic and lymphoid tissues.[2]

Cytogenetic analysis by G-band karyotyping at inclusion required examination of at least 20 metaphases if no abnormalities were found. Patients with a normal karyotype in all

examined metaphases or LOY as the only structural abnormality were eligible for inclusion in the study.

Cell separation and DNA extraction

Mononuclear cells and granulocytes were separated from peripheral blood and/or bone marrow aspirates by *Ficoll-Paque* density gradient centrifugation (GE Healthcare, Munich, Germany) and stored at -196°C/-321°F (live mononuclear cells) or -80°C/-112°F (granulocytes).

DNA was extracted from fresh-frozen mononuclear cells or granulocytes from peripheral blood or bone marrow aspirates from first visit (i.e., diagnostic) samples using the DNA Extraction Kit (Qiagen, Manchester, UK), all as per the manufacturer's protocol. DNA was stored at -80°C/-112°F.

In three cases, matched DNA was extracted from sorted T lymphocytes ($CD3^+$ cells) and used as germline DNA reference.

Next generation sequencing

Targeted gene panel next generation sequencing was performed using a custom-designed multiplex Ion Ampliseq panel (Ampliseq designer, Thermo Fischer Scientific, Waltham, MA, USA) including 20 genes recurrently mutated in MDS. Lower limit of allele detection was 2%. Missense variants present in databases of polymorphisms (ExAC, dbSNP) at a population frequency of ≥1% were excluded from further analysis. The table below summarizes genes included in the panel. Methodology and gene sequencing results from 60 patients from our cohort of patients with idiopathic cytopenia of undetermined significance (ICUS) have been previously published.[1]

Genes included in the 20-gene sequencing panel.

List of genes included in the 20-gene sequencing panel	
TET2	*GATA2*
IDH1	*CEBPA*
IDH2	*RUNX1*
DNMT3A	*SF3B1*
ASXL1	*U2AF1*
TP53	*SRSF2*
NRAS	*ZRSR2*
KRAS	*EZH2*
CBL	*SETBP1*
JAK2	*ETV6*

Single nucleotide polymorphism-based array analysis

Single nucleotide polymorphism-based array analysis (SNP-A) was performed on DNA from diagnostic samples using the Infinium CytoSNP-850K v1.1 BeadChip (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. The table below summarizes the specimen sources of DNA used for SNP-A. DNA source (bone marrow vs. peripheral blood) was not significantly associated with SNP-A-detected aberrations (χ^2=0.44; *P*=0.50).

Specimen sources of DNA for SNP-A in 153 ICUS patients.

Specimen	**Bone Marrow**		**Peripheral Blood**
	Mononuclear Cells	**Granulocytes**	**Granulocytes**
No. of samples, (%)	25 (16)	10 (7)	118 (77)

Illumina intensity files (.idat files) were analyzed visually with the GenomeStudio software version 2011.1 (Illumina, San Diego, CA, USA) based on the human genome build hg19/GRCh37.

Abnormalities were called using modified inclusion criteria applied in a recent systematic review of copy number aberrations (CNA) and copy neutral loss of heterozygosity (CNLOH) in myeloid malignancies[3] and guidelines for reporting of acquired CNA and CNLOH in the clinical setting[4].

Criteria for reporting of CNA and CNLOH detected with SNP-A in our study were:

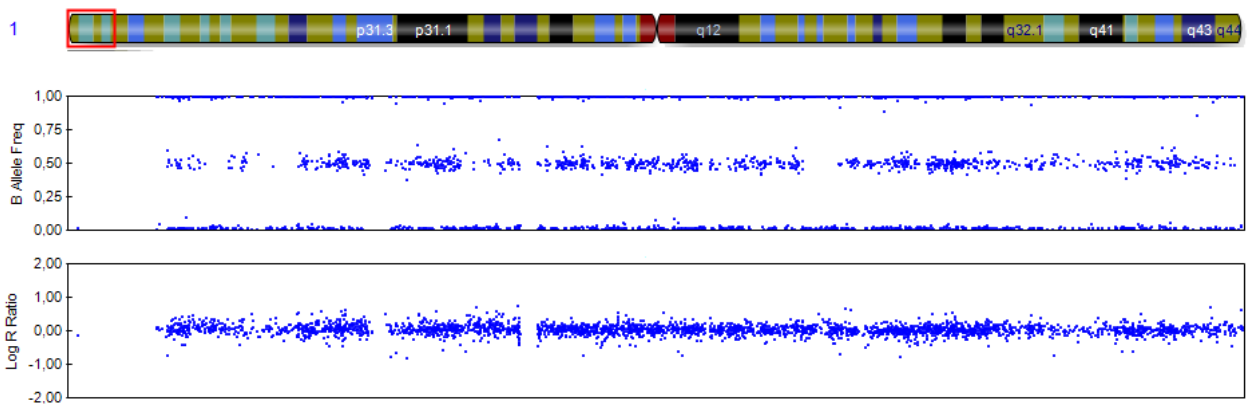
a. CNA ≥100 Kb in size (including at least 30 probes for deletions, and 90 probes for gains) and/or CNA encompassing genes related to myeloid malignancies.

b. CNLOH ≥10 Mb extending to the telomeres and/or in mosaic state.

c. Focal CNA in gene poor regions or in T cell receptor or immunoglobulin genes were not reported.

d. In the case of recurrent focal CNA (in ≥2 subjects) involving regions or genes not previously associated with myeloid malignancies, SNP-A was performed on matched DNA from sorted T lymphocytes from the same patient. CNA were considered constitutional/germline and not reported if the identical CNA was present in matched T lymphocytes. If matched T lymphocyte DNA was not available, the CNA was searched in the Database of Genomic Variants[5]. CNA with ≥50% overlap with previously reported genomic variants were considered germline and excluded.

Interpretation of single nucleotide polymorphism-based array analysis data



Illustrative example of data output from single nucleotide polymorphism-based array analysis of section of chromosome 1 (indicated by red square) with no aberrations apparent as it appears in the GenomeStudio software version 2011.1 (Illumina, San Diego, CA, USA).

Each blue dot/marker on the histogram represents a single nucleotide polymorphism (SNP) position for which the B allele frequency (BAF; upper) and the logR ratio (lower) is plotted. The genotype is determined from the BAF plot, and the copy number is determined from the logR ratio plot. The BAF of a single SNP is generated based on the ratio of signal intensity of B allele to signal intensity of A+B alleles, with values of 0, 0.5, and 1 corresponding to genotypes AA, AB, and BB, respectively. As humans are diploid, we expect at each locus to see either a homozygous ratio (0 or 1) or a heterozygous ratio (0.5). The logR ratio of a single SNP is generated as log2(observed R/expected R), where R is the normalized intensity for each SNP. The signal intensity for each SNP reflects the copy number of the SNP. With an observed signal intensity of a given SNP being equal to the expected signal intensity, the logR ratio would be 0.

Copy number deletion would be visible as loss of heterozygosity (loss of the “band” of data points at 0.5) in the BAF plot and a relative decrease in signal intensity in the logR ratio plot.

Copy number gain would be depicted in the BAF plot by the middle “band” of data points splitting into two new populations representing the genotypes ABB and AAB (in case of trisomy) and a concurrent increase in signal intensity in the logR ratio plot.

Copy neutral loss of heterozygosity would be apparent in the BAF plot as loss of heterozygosity with no deflection in the logR ratio.

Structural aberrations in mosaic state would give rise to more than three clusters of data points in the BAF plot reflecting a greater number of genotypes, and a deviation or no deviation in the logR ratio plot in accordance with the type of aberration.

Statistics

Mann-Whitney U-test or unpaired t-test was used to compare continuous variables and Chi-squared test or Fisher's exact test were used to compare categorical variables. Overall survival (OS) was measured from first bone marrow investigation (=inclusion) to death from any cause. Progression-free survival (PFS) was measured from first bone marrow investigation to progression to a myeloid cancer or death from any cause. Survival curves were estimated with the Kaplan-Meier method and compared with the log-rank test. The Cox proportional hazards regression model was used to estimate hazard ratios (HRs) and associated 95% confidence intervals (CI) for univariable and multivariable analyses.
All *P* values were 2-sided, and *P* values <0.05 denoted statistical significance. Statistical analyses and graphics were performed using R (version 3.5.2).[6]

Supplementary Results

Figure S2. Workflow and distribution of patients.

A total of 171 ICUS patients were included. Of these, 18 patients were excluded from analysis due to insufficient amount of DNA (n=12) or insufficient SNP-A data quality (n=5). Mosaic whole chromosomal trisomy affecting chromosomes 8 and 1q was identified in one patient lacking G-band karyotyping. This patient was also excluded from further analysis, as these extensive gains probably would have been visible by G-band karyotyping and excluded the patient from the study cohort.

In the remaining 10 patients with no cytogenetic analysis available, no CNA or CNLOH were detected, thus, these patients remained part of the analysis population.

Abbreviations: ICUS, idiopathic cytopenia of unknown significance; pts, patients; MDS, myelodysplastic syndromes; 850K SNP-A, 850,000 markers single nucleotide polymorphism-based array analysis; CCUS, clonal cytopenia of undetermined significance

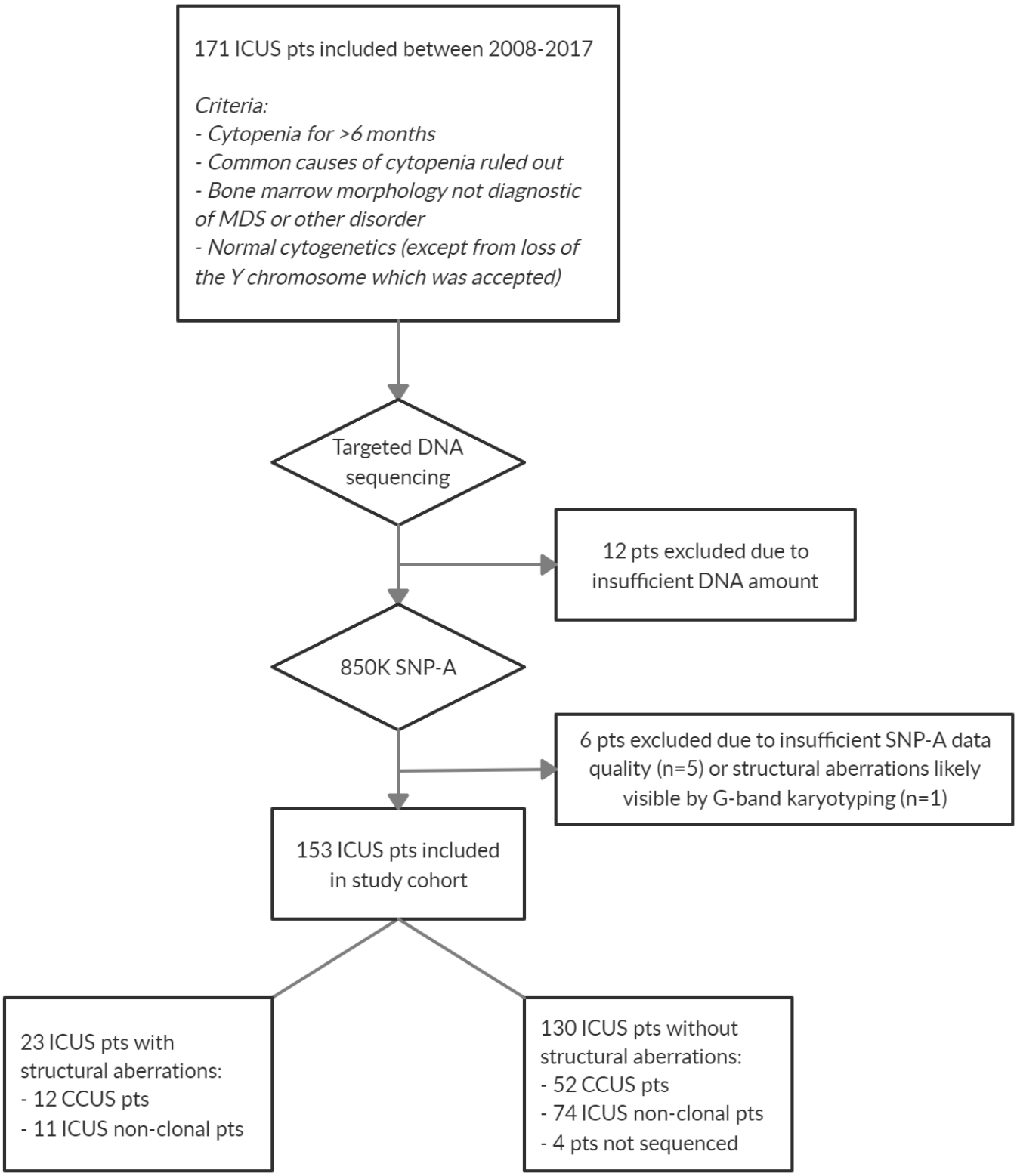

171 ICUS pts included between 2008-2017
Criteria:
- Cytopenia for >6 months
- Common causes of cytopenia ruled out
- Bone marrow morphology not diagnostic of MDS or other disorder
- Normal cytogenetics (except from loss of the Y chromosome which was accepted)
Targeted DNA sequencing
12 pts excluded due to insufficient DNA amount
850K SNP-A
6 pts excluded due to insufficient SNP-A data quality (n=5) or structural aberrations likely visible by G-band karyotyping (n=1)
153 ICUS pts included in study cohort
23 ICUS pts with structural aberrations:
- 12 CCUS pts
- 11 ICUS non-clonal pts
130 ICUS pts without structural aberrations:
- 52 CCUS pts
- 74 ICUS non-clonal pts
- 4 pts not sequenced

Table S1. Baseline characteristics of the 153 ICUS patients.

*From targeted next generation sequencing using a 20-gene panel with a lower limit of detection at 2%

Abbreviations: WBC, white blood cell; MCV, mean corpuscular volume; LDH, lactate dehydrogenase

Parameter	n=153
Years of age, median (range)	69 (17-94)
Sex, male/female, n (%)	97/56 (63%/37%)
Karyotype by G-band karyotyping, n patients (%)	
• Normal (male: 46,XY; female: 46,XX)	132 (86%)
• Normal except from loss of the Y chromosome	11 (11% of males)
• *Not analyzed*	*10 (7%)*
Patients with ≥1 somatic mutation*, n (%)	64 (42%)
Gene mutated* in ≥5 patients, n patients (%)	
• *TET2*	32 (21%)
• *DNMT3A*	19 (12%)
• *SRSF2*	13 (8%)
• *SF3B1*	7 (5%)
• *ZRSR2*	6 (4%)
• *ASXL1*	6 (4%)
• *Not analyzed*	*4 (3%)*
Hematologic parameters, median (range)	
• Hemoglobin, g/L	117.6 (54.8-170.8)
• WBC count, x 10^9/L	4.1 (1.6-10.6)
• Neutrophil count, x 10^9/L	2.1 (0.3-8.2)
• Platelet count, x 10^9/L	126 (9-503)
MCV, median (range), fL	91 (66-116)
Ferritin, median (range), µg/L	185 (5-2980)
LDH, median (range), U/L	193 (9-391)
Number of cytopenias, n (%)	
• One	80 (52%)
Anemia only	39 (26%)
Thrombocytopenia only	22 (14%)
Neutropenia only	19 (12%)
• Two	62 (41%)
• Pancytopenia	11 (7%)

Table S2. List of all mutations detected in 64 out of a total of 153 ICUS patients by targeted 20-gene panel next generation sequencing (Thermo Fischer Scientific, Waltham, MA, USA) with a lower limit of allele detection at 2%. Abbreviations: VAF, variant allele frequency

Patient ID	Transcript	Gene	DNA Sequence Change	Amino Acid Change	Variant Type	VAF, %
14	NM_001127208.2	*TET2*	c.5014_5015insT	p.Pro1673fs	Frameshift	42
14	NM_012433.2	*SF3B1*	c.1998G>T	p.Lys666Asn	Missense	28
40	NM_005089.3	*ZRSR2*	c.598_600delCTT	p.Leu200del	Inframe deletion	84
40	NM_001127208.2	*TET2*	c.1904_1904delA	p.Val636fs	Frameshift	45
40	NM_001127208.2	*TET2*	c.5618T>C	p.Ile1873Thr	Missense	41
41	NM_022552.4	*DNMT3A*	c.938G>A	p.Trp313Ter	Nonsense	14
41	NM_001242532.1	*SRSF2*	c.340G>A	p.Gly114Ser	Missense	12
42	NM_022552.4	*DNMT3A*	c.2645G>A	p.Arg882His	Missense	18
42	NM_006758.2	*U2AF1*	c.101C>T	p.Ser34Phe	Missense	6
1	NM_005188.3	*CBL*	c.1178T>C	p.Ile393Thr	Missense	47
1	NM_001127208.2	*TET2*	c.4120_4121insAAAC	p.Cys1374fs	Frameshift	34
1	NM_001127208.2	*TET2*	c.2474_2474delC	p.Ser825fs	Frameshift	34
100	NM_022552.4	*DNMT3A*	c.2723A>G	p.Tyr908Cys	Missense	4
43	NM_001127208.2	*TET2*	c.2068C>T	p.Gln690Ter	Nonsense	25
43	NM_001127208.2	*TET2*	c.4145A>C	p.His1382Pro	Missense	20
43	NM_004456.4	*EZH2*	c.668T>A	p.Ile223Asn	Missense	5
44	NM_001127208.2	*TET2*	c.4246C>A	p.His1416Asn	Missense	33
44	NM_006758.2	*U2AF1*	c.470A>C	p.Gln157Pro	Missense	23
44	NM_012433.2	*SF3B1*	c.2098A>G	p.Lys700Glu	Missense	5
44	NM_001127208.2	*TET2*	c.521C>A	p.Pro174His	Missense	5
45	NM_005896.3	*IDH1*	c.395G>A	p.Arg132His	Missense	40
45	NM_001242532.1	*SRSF2*	c.284_307delCCCCGGA CTCACACCACAGCCGCC	p.Pro95_Arg102del	Inframe deletion	27
9	NM_022552.4	*DNMT3A*	c.1906G>T	p.Val636Leu	Missense	3
46	NM_001242532.1	*SRSF2*	c.284C>G	p.Pro95Arg	Missense	50
46	NM_001987.4	*ETV6*	c.632G>A	p.Arg211His	Missense	49
46	NM_001127208.2	*TET2*	c.4082G>A	p.Gly1361Asp	Missense	47
46	NM_002168.2	*IDH2*	c.419G>A	p.Arg140Gln	Missense	45
47	NM_001127208.2	*TET2*	c.5035T>C	p.Tyr1679His	Missense	48
47	NM_022552.4	*DNMT3A*	c.1924G>C	p.Gly642Arg	Missense	4
48	NM_004364.4	*CEBPA*	c.265_267delGAG	p.Glu89del	Inframe deletion	50
48	NM_015338.5	*ASXL1*	c.2030_2034delCCAGG	p.Pro677fs	Frameshift	26
48	NM_004456.4	*EZH2*	c.2186T>C	p.Phe729Ser	Missense	27
48	NM_001127208.2	*TET2*	c.1648C>T	p.Arg550Ter	Nonsense	27
48	NM_001127208.2	*TET2*	c.5421_5422insATCA	p.Arg1808fs	Frameshift	26

Patient ID	Transcript	Gene	DNA Sequence Change	Amino Acid Change	Variant Type	VAF, %
48	NM_005188.3	*CBL*	c.1211G>A	p.Cys404Tyr	Missense	3
49	NM_001127208.2	*TET2*	c.2890_2891insA	p.Thr965fs	Frameshift	3
50	NM_001127208.2	*TET2*	c.3017_3018insT	p.Thr1007fs	Frameshift	45
50	NM_003016.4	*SRSF2*	c.283_291delCCCCCGGAC	p.Pro95_Asp97del	Inframe Insertion	42
51	NM_005089.3	*ZRSR2*	c.1344_1345insGAGCCG	p.Arg448_Arg449insGluPro	Inframe Insertion	45
52	NM_022552.4	*DNMT3A*	c.2093G>A	p.Trp698Ter	Nonsense	14
22	NM_001127208.2	*TET2*	c.3176C>G	p.Ser1059Ter	Nonsense	92
22	NM_001242532.1	*SRSF2*	c.284C>A	p.Pro95His	Missense	46
54	NM_015338.5	*ASXL1*	c.3187C>T	p.Gln1063Ter	Nonsense	38
55	NM_001754.4	*RUNX1*	c.167T>C	p.Leu56Ser	Missense	46
56	NM_012433.2	*SF3B1*	c.1996A>C	p.Lys666Gln	Missense	35
57	NM_004364.4	*CEBPA*	c.364G>C,	p.Gly122Arg	Missense	49
58	NM_022552.4	*DNMT3A*	c.1648G>A,	p.Gly550Arg	Missense	9
59	NM_000546.5	*TP53*	c.743G>A	p.Arg248Gln	Missense	6
59	NM_006758.2	*U2AF1*	c.470A>C	p.Gln157Pro	Missense	5
7	NM_000546	*TP53*	c.79C>T	p.Pro27Ser	Missense	7
7	NM_000546	*TP53*	c.797_798insG	p.Arg267fs	Frameshift	3
61	NM_001987.4	*ETV6*	c.791G>A	p.Arg264His	Missense	50
62	NM_022552.4	*DNMT3A*	c.2711C>T	p.Pro904Leu	Missense	5
63	NM_022552.4	*DNMT3A*	c.2096G>A	p.Gly699Asp	Missense	3
64	NM_003016.4	*SRSF2*	c.284C>T	p.Pro95Leu	Missense	8
65	NM_003016.4	*SRSF2*	c.284C>A	p.Pro95His	Missense	49
65	NM_001127208.3	*TET2*	c.1127dup	p.Met376Ielfs	Frameshift	44
65	NM_001127208.3	*TET2*	c.5770A>G	p.Lys1924Glu	Missense	22
65	NM_001127208.3	*TET2*	c.3951del	p.Glu1318Argfs*45	Frameshift	18
66	NM_022552.4	*DNMT3A*	c.1127C>A	p.Ala376Asp	missense	16
66	NM_001127208.3	*TET2*	c.3532del	p.Glu1178Lysfs*48	Frameshift	12
32	NM_001127208.3	*TET2*	c.3781C>T	p.Arg1261Cys	Missense	43
32	NM_001127208.3	*TET2*	c.3577T>C	p.Cys1193Arg	Missense	27
67	NM_001127208.3	*TET2*	c.1648C>T	p.Arg550Ter	Nonsense	4
67	NM_005188.3	*CBL*	c.1150T>A	p.Cys384Ser	Missense	3
18	NM_001127208.3	*TET2*	c.3764_3765insA	p.Tyr1255Ter	Nonsense	38
18	NM_001754.4	*RUNX1*	c.862delC	p.Leu288fs	Frameshift	18
68	NM_001127208.2	*TET2*	c.2479G>A	p.Ala827Thr	Missense	49
68	NM_022552.4	*DNMT3A*	c.2644C>T	p.Arg882Cys	Missense	46
69	NM_015338.5	*ASXL1*	c.1748G>A	p.Trp583Ter	Nonsense	22
69	NM_001127208.2	*TET2*	c.4150G>A	p.Asp1384Asn	Missense	19
70	NM_022552.4	*DNMT3A*	c.2193_2195del	p.Phe732del	Inframe deletion	10
71	NM_003016.4	*SRSF2*	c.284C>A	p.Pro95His	Missense	24
71	NM_001127208.2	*TET2*	c.2266delA	p.Ile756fs	Frameshift	24
71	NM_001127208.2	*TET2*	c.4317delA	p.Lys1439fs	Frameshift	16

Patient ID	Transcript	Gene	DNA Sequence Change	Amino Acid Change	Variant Type	VAF, %
72	NM_005089.3	*ZRSR2*	c.1343_1344insGAGCCG	p.Gly438_Ser439insSerArg	Inframe Insertion	18
34	NM_022552.4	*DNMT3A*	c.1903C>G	p.Arg635Gly	Missense	45
34	NM_001127208.2	*TET2*	c.1496delC	p.Pro499fs	Frameshift	42
34	NM_001127208.2	*TET2*	c.5690T>A	p.Ile1897Asn	Missense	27
73	NM_001127208.2	*TET2*	c.4121G>A	p.Cys1374Tyr	Missense	39
73	NM_001127208.2	*TET2*	c.2141C>G	p.Ser714Ter	Nonsense	36
73	NM_003016.4	*SRSF2*	c.284C>A	p.Pro95His	Missense	29
74	NM_022552.4	*DNMT3A*	c.2047T>G	p.Tyr683Asp	Missense	10
74	NM_022552.4	*DNMT3A*	c.1976G>A	p.Arg659His	Missense	4
36	NM_003016.4	*SRSF2*	c.284C>G	p.Pro95Arg	Missense	27
36	NM_001127208.2	*TET2*	c.4925_4926insGACAACTG	p.Cys1642fs	Frameshift	26
36	NM_001127208.2	*TET2*	c.4354C>T	p.Arg1452Ter	Nonsense	3
75	NM_001127208.2	*TET2*	c.4528C>G	p.Gln1510Glu	Missense	49
75	NM_001127208.2	*TET2*	c.2083_2084delAT	p.Met695fs	Frameshift	16
76	NM_001127208.2	*TET2*	c.5734C>T	p.His1912Tyr	Missense	48
76	NM_001127208.2	*TET2*	c.5606G>A	p.Gly1869Glu	Missense	46
77	NM_022552.4	*DNMT3A*	c.1969G>A	p.Val657Met	missense	15
3	NM_005089.3	*ZRSR2*	c.708_709delCC	p.Leu237fs	Frameshift	90
3	NM_015559.2	*SETBP1*	c.2608G>A	p.Gly870Ser	Missense	44
3	NM_005188.3	*CBL*	c.1099C>A	p.Gln367Lys	Missense	35
3	NM_015338.5	*ASXL1*	c.1934dup	p.Gly646fs	Frameshift	*Sanger seq*
78	NM_001127208.2	*TET2*	c.4767_4768insG	p.Ser1591fs	Frameshift	48
78	NM_001127208.2	*TET2*	c.4011_4011delT	p.Tyr1337fs	Frameshift	46
78	NM_004456.4	*EZH2*	c.1633C>T	p.Gln545Ter	Nonsense	41
78	NM_005089.3	*ZRSR2*	c.196_197delAG	p.Arg68fs	Frameshift	10
78	NM_015338.5	*ASXL1*	c.1934dup	p.Gly646fs	Frameshift	*Sanger seq*
79	NM_001754.4	*RUNX1*	c.292_292delC	p.Leu98fs	Frameshift	33
79	NM_001127208.2	*TET2*	c.3149_3150insTT	p.Gln1051fs	Frameshift	32
79	NM_001127208.2	*TET2*	c.911_911delC	p.Ser305fs	Frameshift	30
79	NM_001242532.1	*SRSF2*	c.284C>T	p.Pro95Leu	Missense	5
80	NM_001127208.2	*TET2*	c.3781C>T	p.Arg1261Cys	Missense	16
80	NM_015338.5	*ASXL1*	c.3403C>T	p.Gln1135Ter	Nonsense	4
81	NM_005896.3	*IDH1*	c.395G>A	p.Arg132His	Missense	20
11	NM_001127208.2	*TET2*	c.4546C>T	p.Arg1516Ter	Nonsense	50
11	NM_001242532.1	*SRSF2*	c.284C>G	p.Pro95Arg	Missense	49
11	NM_001127208.2	*TET2*	c.2352T>A	p.Cys784Ter	Nonsense	49
11	NM_001754.4	*RUNX1*	c.274A>G	p.Thr92Ala	Missense	37
11	NM_001754.4	*RUNX1*	c.496C>T	p.Arg166Ter	Nonsense	11
82	NM_012433.2	*SF3B1*	c.1986C>A	p.His662Gln	Missense	26
83	NM_001127208.2	*TET2*	c.822_822delC	p.Asn275fs	Frameshift	33
84	NM_005896.3	*IDH1*	c.394C>T	p.Arg132Cys	Missense	8

Patient ID	Transcript	Gene	DNA Sequence Change	Amino Acid Change	Variant Type	VAF, %
85	NM_001127208.2	*TET2*	c.3646C>T	p.Arg1216Ter	Nonsense	27
85	NM_001127208.2	*TET2*	c.2746C>T	p.Gln916Ter	Nonsense	5
86	NM_012433.3	*SF3B1*	c.2098A>G	p.Lys700Glu	Missense	15
87	NM_001127208.2	*TET2*	c.1978delG	p.Val660fs	Frameshift	46
87	NM_003016.4	*SRSF2*	c.284C>T	p.Pro95Leu	Missense	41
87	NM_002168.3	*IDH2*	c.419G>A	p.Arg140Gln	Missense	17
88	NM_022552.4	*DNMT3A*	c.2719G>A	p.Glu907Lys	Missense	7
89	NM_022552.4	*DNMT3A*	c.1863_1876delGGTTTACCCACCTG	p.Lys621fs	Frameshift	13
90	NM_012433.3	*SF3B1*	c.2098A>G	p.Lys700Glu	Missense	16
90	NM_022552.4	*DNMT3A*	c.2389A>G	p.Asn797Asp	Missense	14
91	NM_005089.3	*ZRSR2*	c.1354C>T	p.Arg452Cys	Missense	100
91	NM_012433.3	*SF3B1*	c.1998G>T	p.Lys666Asn	Missense	32
91	NM_005089.3	*ZRSR2*	c.591T>G	Ser197Arg	Missense	13
23	NM_001127208.2	*TET2*	c.1637A>G	Lys546Arg	Missense	38
29	NM_022552.4	*DNMT3A*	c.1752C>A	p.Tyr584Ter	Nonsense	25

Table S3. Copy number aberrations and copy neutral loss of heterozygosity in ICUS patients.

All chromosomal aberrations, including loss of the Y chromosome, with base pair location of first and last abnormal single nucleotide polymorphism (SNP) found with 850K SNP-based array analysis in our study cohort of 153 individuals with ICUS. Grey cells mark individuals with at least one somatic mutation present in genes recurrently affected in MDS (i.e., CCUS patients). The mutated gene(s) in the CCUS patients is listed in the far-right column.

[1]From targeted next generation sequencing using a 20-gene panel with a lower limit of detection at 2%

Abbreviations: SNP, single nucleotide polymorphism; CNLOH, copy neutral loss of heterozygosity; qter, q arm terminal; pter, p arm terminal

Patient ID	Chromosome and band	Type	First abnormal SNP	Last abnormal SNP	Comment	Gene(s) mutated[1]
1	Y	Deletion	2655180	28817458	Mosaic	*TET2*, *CBL*
3	9p21.1	Deletion	28194546	28342179		*ASXL1*, *CBL*, *ZRSR2*, *SETBP1*
3	Y	Deletion	2655180	28817458		[Same as above]
4	Y	Deletion	2655180	28817458		
6	2p23.3	Deletion	25448050	26053166	Mosaic	
7	16q23.2	Deletion	80785448	81008860		*TP53*
8	Y	Deletion	2655180	28817458	Mosaic	
9	15q11.2	Deletion	22750303	23272733		*DNMT3A*
10	9p24.3	Deletion	535866	722946		
16	Y	Deletion	2655180	28817458	Mosaic	
18	Y	Deletion	2655180	28817458		*TET2*, *RUNX1*
23	12p12.2	Deletion	20942465	21074038		*TET2*
25	7q31.1	Deletion	111012622	111285540		
29	7p14.3	Deletion	33772241	34467317		*DNMT3A*
31	Y	Deletion	2655180	28817458		
32	3p14.2	Deletion	60192208	60379253		*TET2*
32	Y	Deletion	2655180	28817458	Mosaic	[Same as above]
33	6q27	Deletion	168699291	168879958		
34	7q22.1	Deletion	99971313	102025660	Mosaic	*TET2*, *DNMT3A*
35	Y	Deletion	2655180	28817458	Mosaic	
36	4q24-q25	Deletion	105217098	108084614	Mosaic	*TET2*, *SRSF2*
37	Y	Deletion	2655180	28817458	Mosaic	
1	11q12.3-qter	CNLOH	62778397	135000000	Mosaic	*TET2*, *CBL*
1	20q13.2-qter	CNLOH	51361284	63000000		[Same as above]
11	15 (entire q arm)	CNLOH	20071673	102461000	Mosaic	*TET2*, *SRSF2*, *RUNX1*
13	19pter-p13.11	CNLOH	260912	19887292	Mosaic	
14	1pter-p31.1	CNLOH	752721	84111016	Mosaic	*TET2*, *SF3B1*
22	4q21-qter	CNLOH	84518315	190963766		*TET2*, *SRSF2*
34	4 (entire q arm)	CNLOH	53859826	190963766	Mosaic	*TET2*, *DNMT3A*
39	1q21.1-qter	CNLOH	142632577	249219000	Mosaic	
5	12q14.1-q14.2	Gain	62825830	64764152		
8	3p14.2	Gain	59709050	61023335		
19	22q11.21	Gain	18886915	21458082		
21	4q35.1-q35.2	Gain	186504058	187331659		
37	11q25	Gain	131276764	131915260		

Figure S3. Copy number aberrations and copy neutral loss of heterozygosity in ICUS patients. Genome-wide overview of all genetic aberrations.

Genome-wide overview of all genetic aberrations, including loss of the Y chromosome (LOY), found with 850K SNP-based array analysis in our study cohort of 153 individuals with ICUS. Each line represents a different individual. Green lines are deletions, red lines represent regions of acquired copy neutral loss of heterozygosity and blue lines are copy number gains. Image created with the online software GREVE (Cazier JB, Holmes CC, Broxholme J. *Bioinformatics*. 2012;28(22):2981–2982).

A total of 25 structural aberrations (excluding LOY) were identified by SNP-A; 12 deletions, 8 CNLOH, and 5 gains, in 23/153 patients (15%) with no aberrations identified in chromosomes 5, 8, 10, 13, 14, 17, 18, 21, and X. LOY was detected in 10 patients by SNP-A of whom 8 were overlapping with the cases identified by G-band karyotyping. Two cases of LOY were identified by SNP-A only, and three cases were identified by cytogenetic analysis only; all five were in mosaic state. Previous studies have also reported a variable number of aberrations, mainly low-level, that were "missed" by SNP-A compared to metaphase karyotyping (Kanagal-Shamanna R, Hodge JC, Tucker T, et al. *Cancer Genet.* 2018;228–229:197–217).

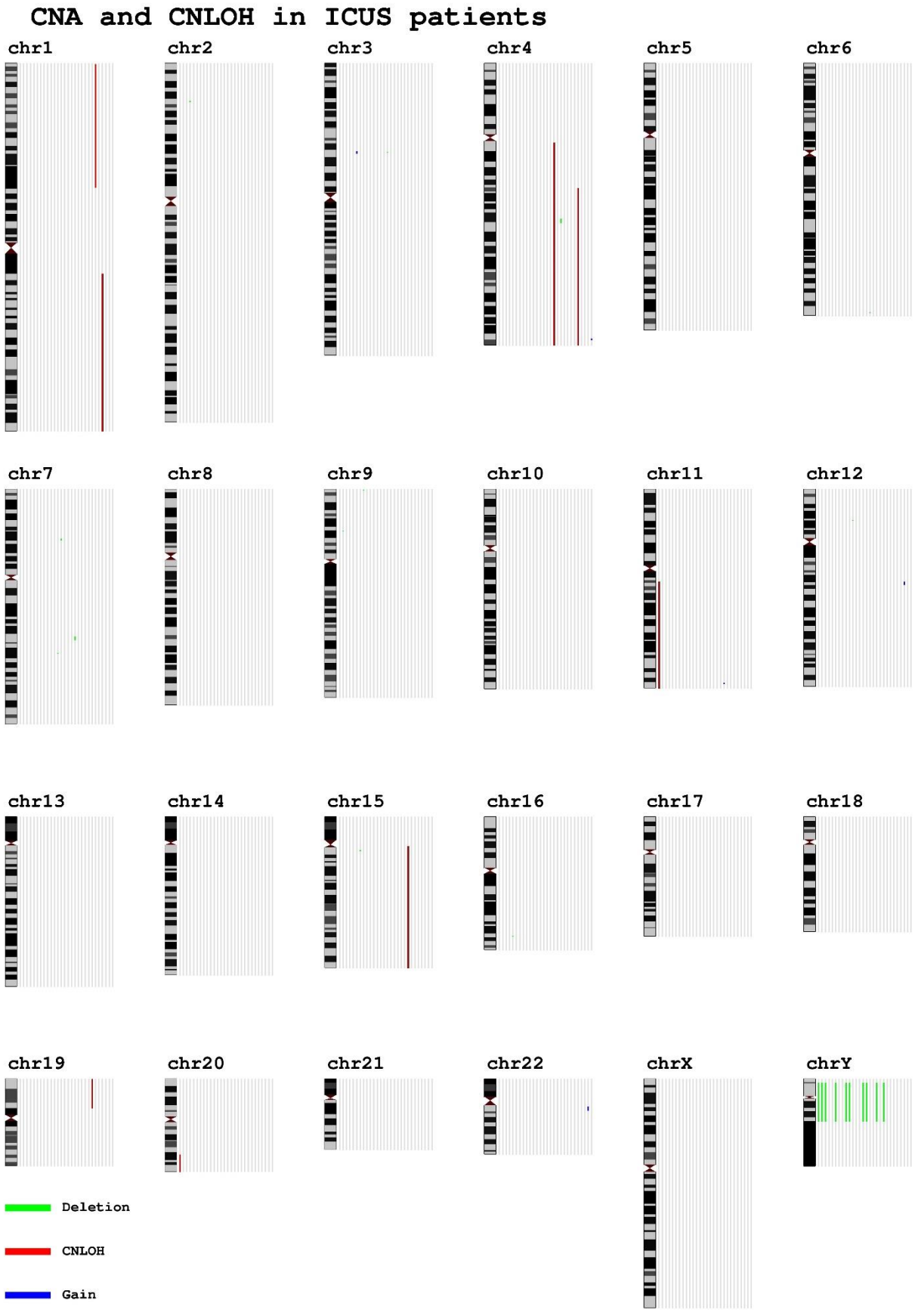

CNA and CNLOH in ICUS patients
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
Deletion
CNLOH
Gain

Figure S4. Proportion of ICUS patients with a marker of clonal hematopoiesis identified by targeted sequencing alone and in combination with SNP-A.

Proportion of total ICUS patients (n=153) with a marker of clonal hematopoiesis identified by targeted 20-gene panel next generation sequencing alone (n=64) and in combination with SNP-A (n=75; loss of the Y chromosome ignored). Frequency and types of structural aberrations in ICUS patients detected by SNP-A.

Of note, all patients had a normal karyotype by G-band cytogenetics (or loss of the Y chromosome was the only abnormality), or G-band karyotyping had failed/was not attempted (n=10). The inserted pie in the diagram on the right shows the distribution of abnormalities attributed to copy number aberrations (CNA) and copy neutral loss of heterozygosity (CNLOH), respectively. The hatched insert in the diagram on the right shows the proportion of ICUS patients (n=11) with abnormalities detected only by SNP-A (no somatic mutations identified).

Abbreviations: SNP-A, single nucleotide polymorphism-based array analysis; CNA, copy number aberration; CNLOH, copy neutral loss of heterozygosity

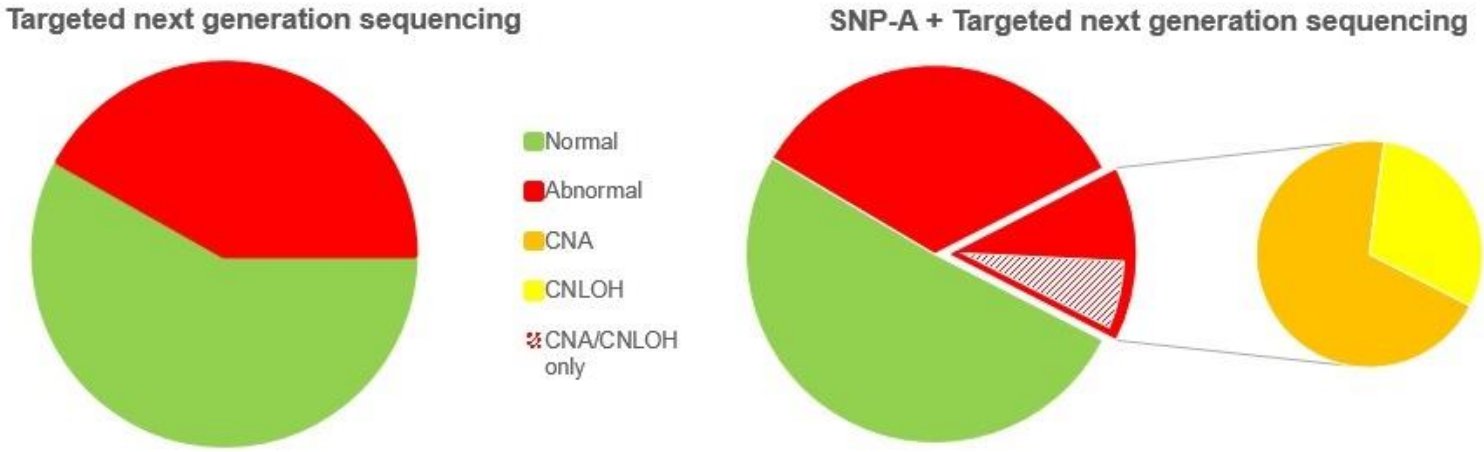


Figure S5. Structural aberrations of chromosome 4q24 involving the *TET2* locus for three different patients.

Chromosome 4 is shown in top. Each pair of histograms represents a single case. The histogram sections marked in salmon show the single nucleotide polymorphisms/chromosomal regions involved in the structural aberrations. The *TET2* locus is marked with a stippled vertical line.

The upper panel shows partial CNLOH of chromosome 4q. In the middle panel, mosaic CNLOH of chromosome 4q is visible, and the lower panel shows a focal mosaic deletion involving the *TET2* locus. The insert shows a magnified portion of chromosome 4q24 containing the *TET2* deletion. All three patients also had a *TET2* mutation (with variant allele frequencies at 92%, 42%, and 26%, respectively) and one other mutation, in line with the findings in previous studies (Jankowska AM, Szpurka H, Tiu R V., et al. *Blood*. 2009;113(25):6403–6410; Mohamedali AM, Smith AE, Gaken J, et al. *J. Clin. Oncol.* 2009;27(24):4002–4006).

Abbreviations: CNLOH, copy neutral loss of heterozygosity; BAF, B allele frequency

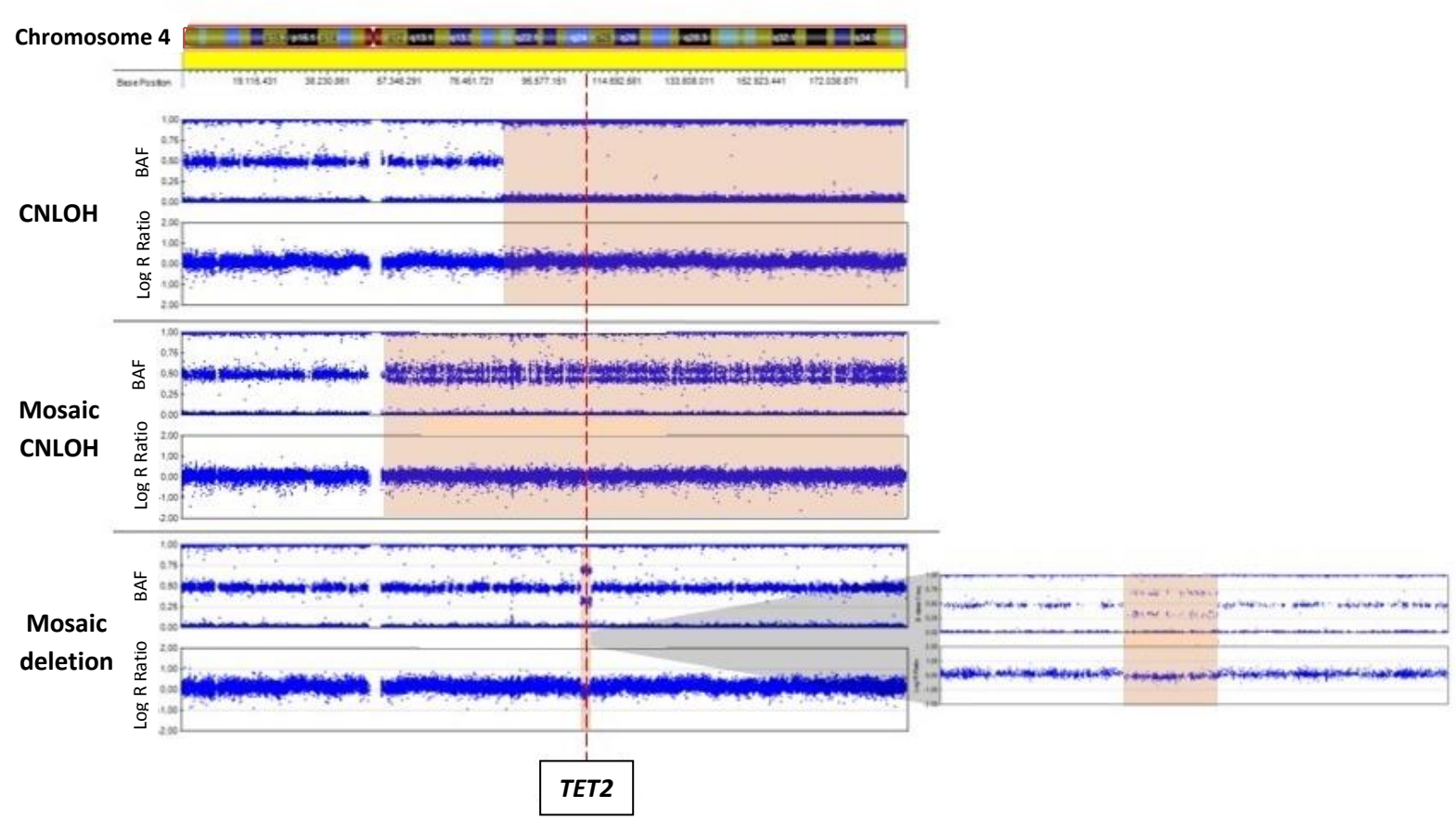


Table S4. Clinical, hematological, and molecular features of ICUS patients grouped according to the presence of copy number aberrations (CNA) or copy neutral loss of heterozygosity (CNLOH).

[1]Data available for 92 ICUS patients without CNA/CNLOH and 19 ICUS patients with CNA/CNLOH
[2]From G-band karyotyping. Data available for 75 male ICUS patients without CNA/CNLOH and all 17 male ICUS patients with CNA/CNLOH
[3]From targeted next generation sequencing using a panel of 20 MDS-associated genes with a lower limit of detection at 2%. Data available for 126 ICUS patients without CNA/CNLOH and all 23 ICUS patients with CNA/CNLOH
[4]Adverse mutation definition adopted from Bejar *et al.* Current Opinion in Hematology. 2017 encompassing mutation in *ASXL1, NRAS, SRSF2, U2AF1, TP53, RUNX1, EZH2, IDH2*, or *GATA2*

Variable	ICUS without CNA/CNLOH (n=130)	ICUS with CNA/CNLOH (n=23)	*P* value
Age, median (range), years	68 (17-89)	72 (25-94)	0.19
Sex, male/female, n	80/50	17/6	0.37
Hemoglobin, mean (range), g/L	118.1 (54.8-171.0)	113.1 (72.5-168.0)	0.32
White blood cell count, median (range), $x10^9$/L	4.1 (1.6-10.6)	3.9 (1.7-10.1)	0.98
Absolute neutrophil count, median (range), $x10^9$/L	2.1 (0.1-8.2)	2.1 (0.6-7.8)	0.74
Platelet count, median (range), $x10^9$/L	127 (9-503)	110 (73-277)	0.38
Mean corpuscular volume, median (range), fL	90 (66-114)	97 (70-116)	0.014
Ferritin, median (range), µg/L	173 (5-2980)	260 (16-2590)	0.043
Lactate dehydrogenase, median (range), U/L	192 (60-497)	201 (121-386)	1.00
Smoking ever[1], n (%)	60 (65)	12 (63)	1.00
Loss of the Y chromosome[2], n (% of males)	6 (8)	5 (29)	0.028
Unmutated[3], n (%) Mutated[3], n (%)	74 (59) 52 (41)	11 (48) 12 (52)	0.46
No. of patients with adverse mutation(s)[4], (%)	21 (17)	4 (17)	1.00
No. of patients with ≥2 mutations[3], (%)	25 (20)	7 (30)	0.27

Figure S6. Forest plot of hazard ratios and 95% confidence intervals for progression or death in multivariable analysis in ICUS patients (n=109 with complete data).

Severe anemia defined as hemoglobin <100 g/L.

Mutation identified by targeted next generation sequencing using a 20-gene panel with a lower limit of detection at 2%.

Abbreviation: HR, hazard ratio; CI, confidence interval; CNA, copy number aberration; CNLOH, copy neutral loss of heterozygosity

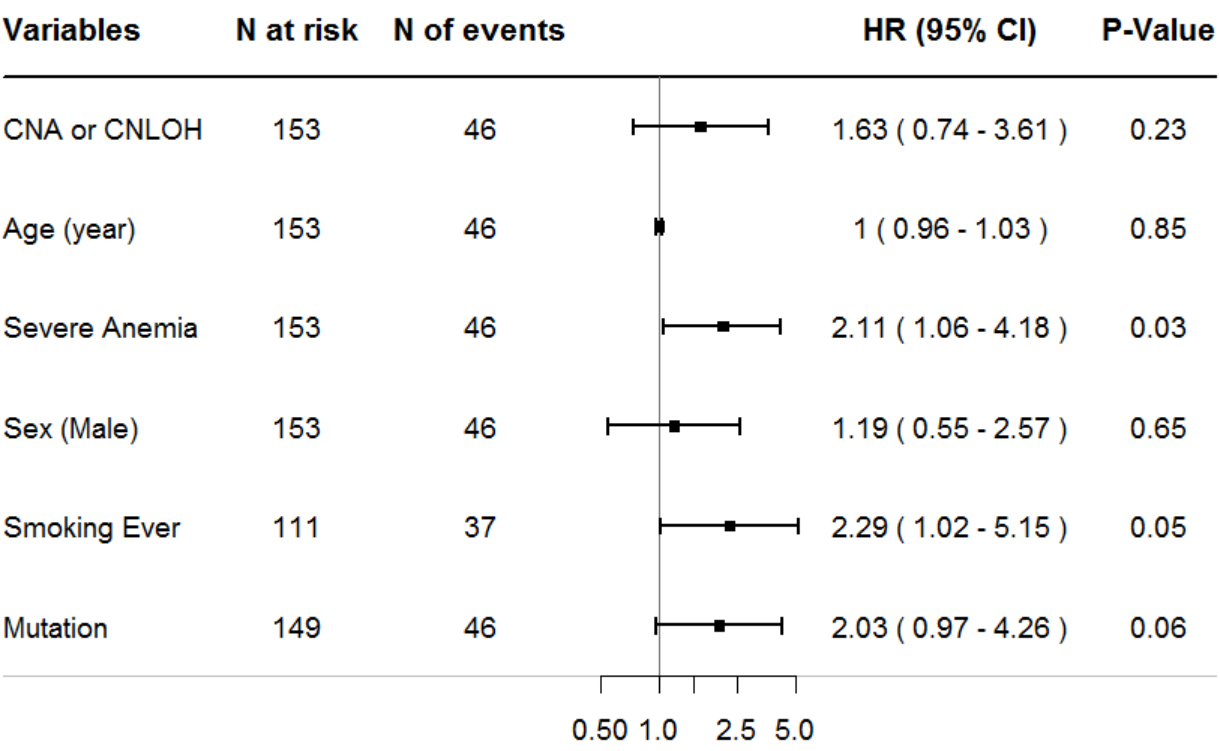


Table S5. Isolated impact of MDS-associated CNA and CNLOH on survival.

MDS-associated CNA and CNLOH (listed in Table 1A and B) were not as strongly associated with an adverse outcome as the presence of any CNA/CNLOH. However, number of patients (n=9) with MDS-associated CNA or CNLOH and number of deaths (n=5) among these were small. (A) Univariable and (B) multivariable analysis, including (C) forest plot, of the impact of MDS-associated CNA and CNLOH on Overall Survival and Progression-Free Survival in ICUS patients.

Loss of the Y chromosome was not included in analysis.

*From two-sided log-rank test

†CCUS defined as ICUS (criteria in Supplementary Methods) in the presence of one or more somatic mutations in genes associated with myeloid malignancies detected through targeted next generation sequencing using a 20-gene panel with a lower limit of detection at 2%

‡Four patients lacked mutational profiling and were not included in subgroup analysis

Abbreviations: MDS, myelodysplastic syndromes; CNAs/CNLOH, copy number aberrations or copy neutral loss of heterozygosity; ICUS, idiopathic cytopenia of unknown significance; CCUS, clonal cytopenia of undetermined significance; CI, confidence interval

A

Impact of MDS-associated CNAs/CNLOH versus absence of MDS-associated CNAs/CNLOH on survival in ICUS patients and subgroups of CCUS and ICUS non-clonal patients in univariable analysis

Sample	No. of Patients	Overall Survival		Progression-Free Survival	
		Hazard Ratio (95% CI)	*P**	Hazard Ratio (95% CI)	*P**
All patients	153	2.74 (1.04-7.21)	0.03	1.63 (0.63-4.17)	0.3
CCUS†	64	2.18 (0.62-7.68)	0.2	1.09 (0.33-3.61)	0.9
ICUS non-clonal‡	85	2.46 (0.53-11.45)	0.2	-	

B

Impact of MDS-associated CNAs/CNLOH versus absence of MDS-associated CNAs/CNLOH on survival in ICUS patients and in the subgroup of CCUS patients in multivariable analysis

Sample	No. of Patients	Overall Survival		Progression-Free Survival	
		Adjusted Hazard Ratio (95% CI)	*P**	Adjusted Hazard Ratio (95% CI)	*P**
All patients	109	1.43 (0.39-5.24)	0.58	0.86 (0.24-3.06)	0.82
CCUS†	51	1.84 (0.18-18.79)	0.61	0.46 (0.05-3.83)	0.47

C

Forest plot of hazard ratios (HR) and 95% CI for all-cause mortality in multivariable analysis in ICUS patients (n=109 with complete data).

Severe anemia defined as hemoglobin <100 g/L; Mutation identified by targeted next generation sequencing using a 20-gene panel with a lower limit of detection at 2%

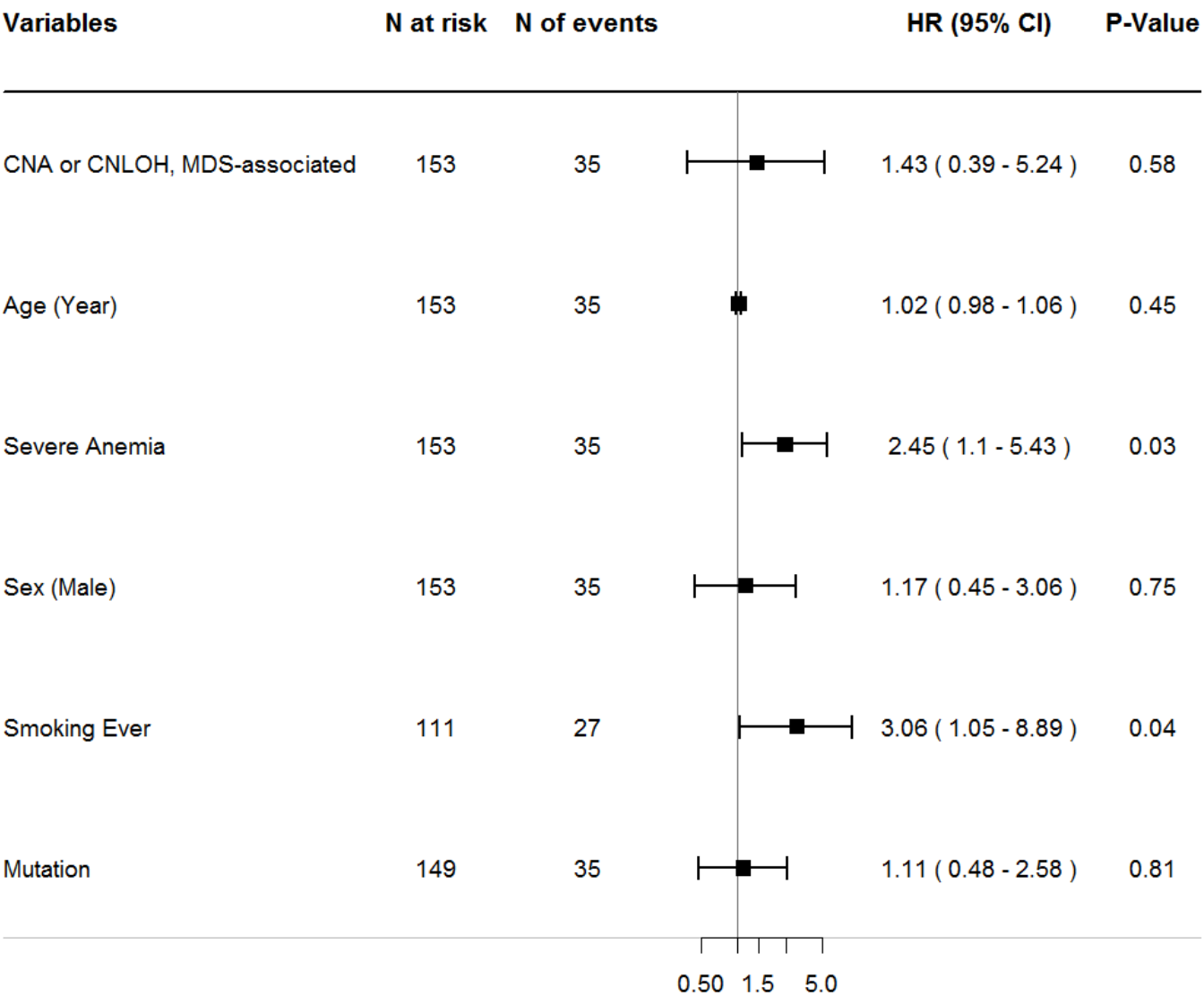


Figure S7. Progression-free survival in CCUS patients with and without CNA or CNLOH.

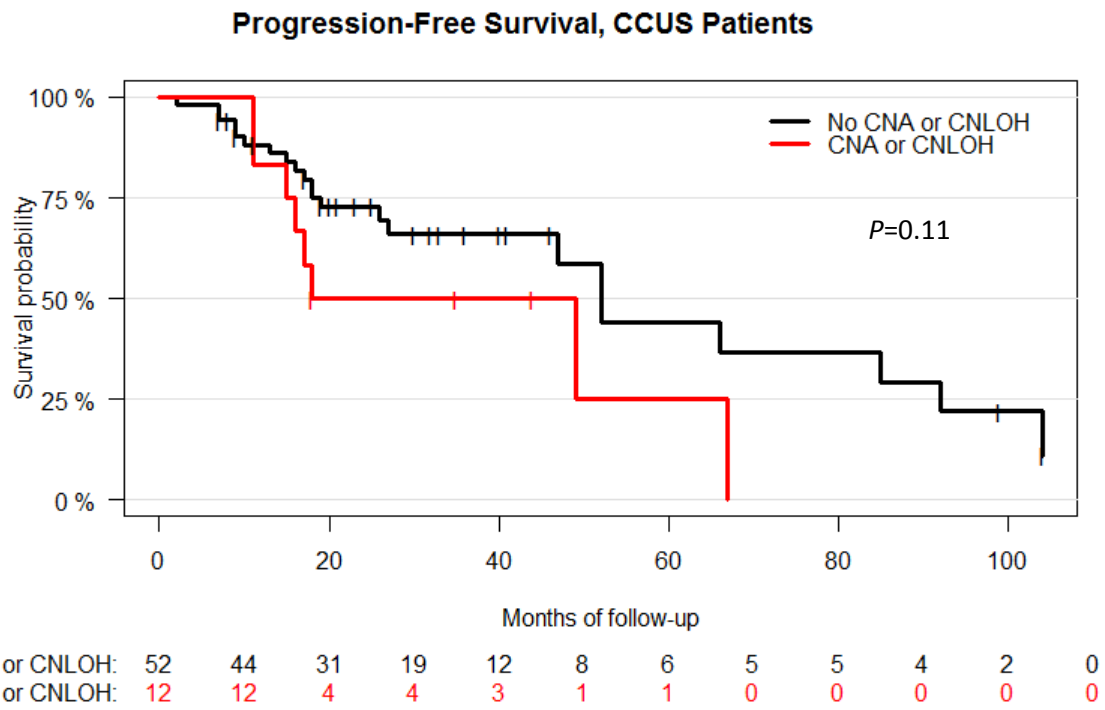
(A) Kaplan-Meier curves for progression-free survival for the group of CCUS patients with CNA or CNLOH (excluding loss of the Y chromosome; red curve) and the group of CCUS patients with no CNA or CNLOH (black curve). (B) Forest plot of hazard ratios including 95% confidence intervals for progression or death in multivariable analysis in CCUS patients (n=51 with complete data).

P-values from two-sided log-rank tests.

Severe anemia defined as hemoglobin <100 g/L.

Abbreviations: CCUS, clonal cytopenia of undetermined significance; CNA, copy number aberrations; CNLOH, copy neutral loss of heterozygosity; HR, hazard ratio; CI, confidence intervals

A



B

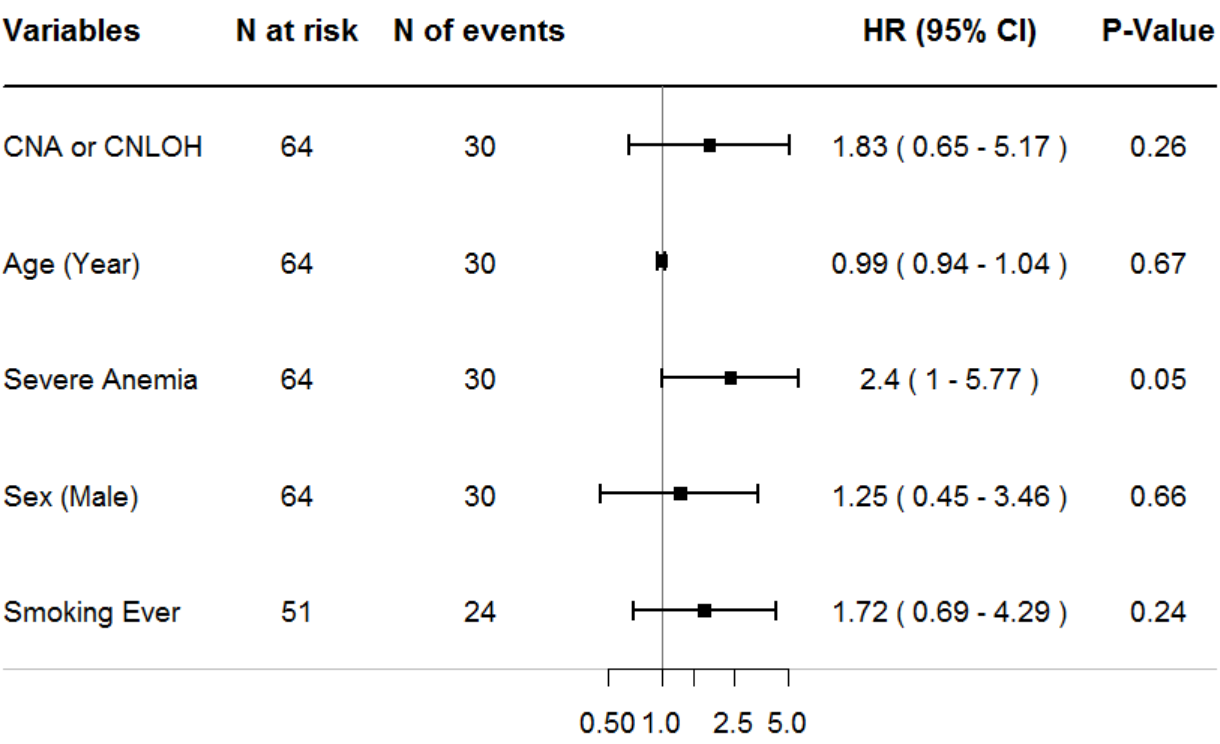

Variables
N at risk
N of events
HR (95% CI)
P-Value
CNA or CNLOH
64
30
1.83 (0.65 - 5.17)
0.26
Age (Year)
64
30
0.99 (0.94 - 1.04)
0.67
Severe Anemia
64
30
2.4 (1 - 5.77)
0.05
Sex (Male)
64
30
1.25 (0.45 - 3.46)
0.66
Smoking Ever
51
24
1.72 (0.69 - 4.29)
0.24
0.50
1.0
2.5
5.0

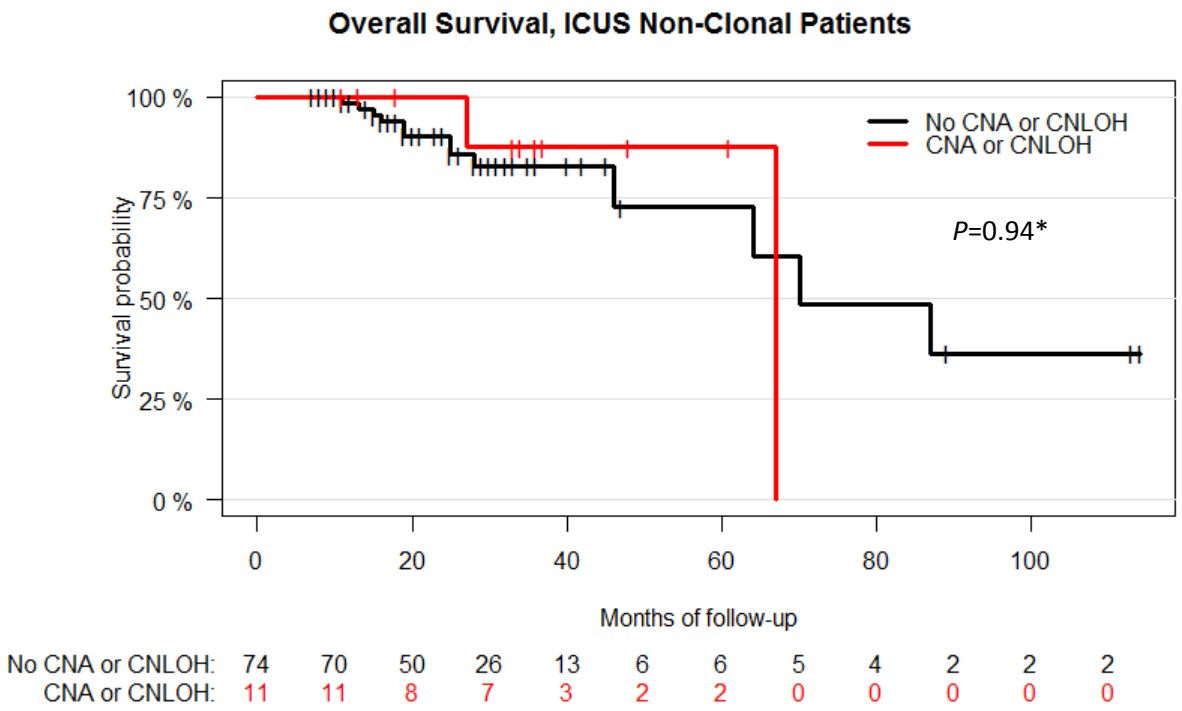
Figure S8. Overall survival in ICUS non-clonal patients with and without CNA or CNLOH.

(A) Kaplan-Meier curves for overall survival for the group of ICUS non-clonal patients with CNA or CNLOH (excluding loss of the Y chromosome; red curve) and the group of ICUS non-clonal patients with no CNA or CNLOH (black curve). (B) Univariable and multivariable analysis of the impact of CNA or CNLOH on overall survival in patients with ICUS non-clonal.

*From two-sided log-rank test
†Adjusted for age, sex and severe anemia (hemoglobin <100 g/L). Smoking ever was not included as a covariate in this multivariable analysis in patients with ICUS non-clonal as numbers were too few.

Abbreviations: CNA, copy number aberration; CNLOH, copy neutral loss of heterozygosity; ICUS, idiopathic cytopenia of unknown significance; CI, confidence interval

A



B

Impact of CNAs/CNLOH versus absence of CNAs/CNLOH on overall survival in patients with ICUS non-clonal in univariable and multivariable analysis

Sample	No. of Patients	Overall Survival			
		Hazard Ratio (95% CI)	*P**	Adjusted Hazard Ratio† (95% CI)	*P**
ICUS non-clonal	85	0.94 (0.21-4.28)	0.9	0.66 (0.14-3.10)	0.6

Table S6. Clinical, hematological, and molecular features of CCUS patients grouped according to the presence of CNA and CNLOH detected by SNP-A.

[1]From Chi-squared test/Fisher's exact test (categorical variables) or Mann-Whitney U-test/unpaired t-test (continuous variables)
[2]Defined as mean corpuscular volume >100 fL. Data available for 49 CCUS patients without CNA/CNLOH and all 12 CCUS patients with CNA/CNLOH
[3]Data available for 41 CCUS patients without CNA/CNLOH and 10 CCUS patients with CNA/CNLOH
[4]From G-band karyotyping. Data available for 27 male CCUS patients without CNA/CNLOH and all 10 male CCUS patients with CNA/CNLOH
[5]From targeted next generation sequencing using a panel of 20 MDS-associated genes with a lower limit of detection at 2%
[6]Adverse mutation definition adopted from Bejar *et al.* Current Opinion in Hematology. 2017 encompassing mutations in *ASXL1, NRAS, SRSF2, U2AF1, TP53, RUNX1, EZH2, IDH2,* or *GATA2*

Variable	CCUS without CNA/CNLOH (n=52)	CCUS with CNA/CNLOH (n=12)	*P* value[1]
Age, median (range), years	71 (39-89)	76 (53-94)	0.12
Sex, male/female, n	29/23	10/2	0.11
Hemoglobin, mean (range), g/L	111.7 (70.9-153.0)	112.8 (72.5-139.0)	0.87
Absolute neutrophil count, median (range), $x10^9$/L	1.9 (0.5-8.0)	1.8 (0.6-4.0)	0.26
Platelet count, median (range), $x10^9$/L	126 (18-427)	106 (73-207)	0.25
Macrocytosis[2], n (%)	8 (16)	6 (50)	0.022
Smoking ever[3], n (%)	22 (54)	6 (60)	1.00
Loss of the Y chromosome[4], n (% of males)	2 (7)	4 (40)	0.035
Variant allele frequency[5], median (range), %	24 (3-100)	36 (3-92)	0.039
No. of patients with adverse mutation(s)[6], (%)	21 (40)	4 (33)	0.75
No. of patients with ≥2 genes mutated[5], (%)	25 (48)	7 (58)	0.75

Supplementary Information, References

1. Hansen JW, Westman MK, Sjö LD, et al. Mutations in idiopathic cytopenia of undetermined significance assist diagnostics and correlate to dysplastic changes. *Am. J. Hematol.* 2016;91(12):1234–1238.

2. Swerdlow SH, Campo E, Harris NL, et al. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon, France: IARC Press; 2008.

3. Kanagal-Shamanna R, Hodge JC, Tucker T, et al. Assessing copy number aberrations and copy neutral loss of heterozygosity across the genome as best practice: An evidence based review of clinical utility from the cancer genomics consortium (CGC) working group for myelodysplastic syndrome, myelodysplastic/myeloproliferative and myeloproliferative neoplasms. *Cancer Genet.* 2018;228–229:197–217.

4. Schoumans J, Suela J, Hastings R, et al. Guidelines for genomic array analysis in acquired haematological neoplastic disorders. *Genes Chromosom. Cancer*. 2016;55(5):480–491.

5. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986-D992.

6. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.