**SRSF2 Mutations in 275 cases with Chronic Myelomonocytic Leukemia (CMML)**

Manja Meggendorfer¹, Andreas Roller¹, Torsten Haferlach¹, Christiane Eder¹, Frank Dicker¹, Vera Grossmann¹, Alexander Kohlmann¹, Tamara Alpermann¹, Kenichi Yoshida², Seishi Ogawa², H. Phillip Koeffler³, Wolfgang Kern¹, Claudia Haferlach¹, Susanne Schnittger¹.

¹MLL Munich Leukemia Laboratory, Munich, Germany
²Cancer Genomics Project, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
³UCLA School of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA
⁴National University Cancer Institute Singapore, National University Hospital, Singapore

°Corresponding author: Susanne Schnittger, PhD, MLL Munich Leukemia Laboratory, Max-Lebsche-Platz 31, 81377 Munich, Germany, Phone +49-89-990-15-300, Fax: +49-89-990-15-309, email: susanne.schnittger@mll.com
URL: www.mll.com

**Running Title:**
*SRSF2* Mutations in CMML
Abstract
We analyzed the mutational hotspot region of SRSF2 (Pro95) in 275 cases with CMML. In addition, ASXL1, CBL, EZH2, JAK2V617F, KRAS, NRAS, RUNX1, and TET2 mutations were investigated in subcohorts. Mutations in SRSF2 (SRSF2mut) were detected in 47% (129/275) of all cases. In detail, 120 cases had a missense mutation at Pro95 leading to a change to Pro95His, Pro95Leu, Pro95Arg, Pro95Ala, or Pro95Thr. In 9 cases, three new in/del mutations were observed: seven cases with a 24 bp deletion, one case with a 3 bp duplication, and one case with a 24 bp duplication. In silico analyses predicted a damaging character for the protein structure of SRSF2 for all mutations. SRSF2mut were correlated with higher age, less pronounced anemia and a normal karyotype. SRSF2mut and EZH2mut were mutually exclusive but SRSF2mut associated with TET2mut. In the total cohort, no impact of SRSF2mut on survival was observed. However in the RUNX1mut subcohort, SRSF2 Pro95His had a favorable impact on overall survival. This comprehensive mutation analysis revealed that 93% of all CMML patients carried at least one somatic mutation in 9 recurrently mutated genes. In conclusion, these data show the importance of SRSF2mut as new diagnostic marker in CMML.
Introduction

Chronic myelomonocytic leukemia (CMML) is a clonal hematopoietic malignancy that can be characterized by features of both, a myelodysplastic syndrome (MDS) and a myeloproliferative neoplasm (MPN). Therefore the WHO classification of 2008 assigned CMML to the mixed category MDS/MPN. A further characteristic feature is the wide heterogeneity of clinical presentations and course, leading to variable prognosis. Beside cytological criteria for diagnosis, the only genetic criterion, until recently, was the absence of the BCR-ABL1 fusion transcript. The number of blasts in the peripheral blood (PB) and bone marrow (BM) is a prognostic factor dividing CMML cases into two morphological categories: CMML-1 with fewer than 5% blasts in PB or 10% in BM, and CMML-2 with 5-19% blasts in PB or 10-19% in BM. Median overall survival is about 20 months in CMML-1 and 15 months in CMML-2, but wide variations exist. In approximately 15-30% of CMML patients, the disease evolves into acute myeloid leukemia (AML). Based on patient characteristics of 213 patients Onida et al. defined a scoring system for CMML, named M. D. Anderson (MDA) Prognostic score, stratifying CMML patients in the four subgroups low, intermediate-1, intermediate-2, and high risk. The level of risk is defined by four scores assigned by the following variables: hemoglobin levels below 12g/dL, lymphocyte count higher than 2.5x10^9/L, peripheral blood IMCs more than 0%, and bone marrow blasts 10% or more.

The vast majority of patients show a normal karyotype in the CMML cells and only 20-40% show clonal cytogenetic abnormalities. Such and coworkers investigated 414 patients with CMML to evaluate the prognostic impact of cytogenetic abnormalities and identified three risk categories. A normal karyotype or loss of the Y-chromosome as a sole abnormality represent the low risk group; trisomy 8, abnormalities of chromosome 7 or a complex karyotype (defined as three or more abnormalities) were related to the high risk group. All other abnormalities were assigned to the intermediate risk category.

In contrast to cytogenetic aberrations, several molecular gene mutations recently have been found to be frequent in CMML (resulting in overall mutation frequencies of >55%); but unfortunately, none of these alterations is specific for CMML. Gene mutations identified in CMML cases affect different cellular targets and processes, like RUNX1 (transcriptional regulation), Isocitrate Dehydrogenases IDH1/2 (metabolism), or KRAS, NRAS, CBL, and JAK2 (tyrosine-signaling pathways). TET2, DNMT3A, ASXL1, UTX, and EZH2 contribute in the broadest sense to epigenetic regulatory mechanisms. All of the cytogenetic changes and molecular mutations have been associated with the pathogenesis of CMML but do not fully explain leukemogenesis.

Thus far, mutations in several of these genes already show prognostic relevance. To date, EZH2 is the best molecularly analyzed gene in CMML and implies an unfavorable prognosis. Mutation of ASXL1 correlates with evolution to AML and a shorter overall
The impact of TET2 mutations remains controversial: in MDS patients it is associated with a favorable outcome, in CMML different studies found favorable to adverse clinical courses for it. Mutations in RUNX1 clearly correlate with a poor outcome in MDS and AML patients. 

Previously, we have investigated 81 CMML cases and analyzed the mutation frequency of a number of genes that were found to be recurrently mutated in CMML. These comprehensive studies resulted in an overall mutation frequency of 82%, indicating that there is a certain percentage of patients with unknown molecular alterations.

More recently, an additional cellular process was found to be altered in MDS. A whole-exome sequencing approach of 29 MDS specimens and their normal controls detected mutations in several components of the splicing machinery (i.e. spliceosome; like SF3B1 and U2AF1), mostly involved in 3'-splice site recognition. In this context a new candidate gene, SRSF2 (serine/arginine-rich splicing factor 2, also known as SC35, a classical member of the SR-protein family), was identified in close cooperation with our laboratory. Members of the SR-protein family function in constitutive and alternative splicing. They contain a RNA recognition motif (RRM) for binding to RNA and a Arginine/Serine-rich (RS)-domain for interaction with other SR-proteins (Fig. 1). As a component of the spliceosome, SRSF2 binds to exonic splicing enhancers preventing exon skipping and ensuring the correct linear order of exons in spliced mRNA. In our recent study, mutations within the SRSF2 sequence occurred exclusively at position 95 (Pro95), located in a linker sequence between the two functional RRM- and RS-domains. SRSF2 was found to be most frequently mutated in CMML (28%), less frequently in MDS without increased ring sideroblasts (12%), and to some extent in refractory anemia with ring sideroblasts (RARS; 6%) and AML/MDS (7%). It was rarely seen to be mutated in MPN (2%) or de novo AML (1%).

To characterize further the genetic defects of CMML, we analyzed the frequency of SRSF2 mutations, their coincidence with other mutations, and their prognostic relevance in a large cohort of 275 cases.
Patients and Methods

Patient Cohort
In total, 275 cases with CMML were analyzed. All cases were validated on peripheral blood and/or bone marrow smears according to WHO standards and included in all cases May Grünwald Giemsa (MGG) staining, as well as myeloperoxidase (MPO), non-specific esterase (NSE), and iron stains. The cohort comprised 189 males and 86 females with a median age of 72.8 years (range: 21.9 – 93.3 years). 81 patients that have been published previously by our group except for SRSF2 entered the cohort. There is no overlap with the CMML-cohort analyzed in Yoshida et al. Cytogenetic analyses were performed after short-term culture. Karyotypes were analyzed after G-banding and described according to the International System for Human Cytogenetic Nomenclature (ISCN 1995 guidelines). Further parameters are given in Table 1. All patients gave their consent for genetic analyses and the use of laboratory results for research purposes. The study design adhered to the tenets of the Declaration of Helsinki and was approved by our institutional review board before its initiation.

Sequencing Analyses
Isolation of mononuclear cells, DNA and mRNA extraction, as well as random primed cDNA synthesis was performed as described previously. A 187 bp fragment, containing the mutational hotspot region of SRSF2 around Pro95, was amplified with the GC-RICH PCR system (Roche Applied Science, Mannheim, Germany) from either genomic DNA (n=201) or cDNA templates (n=74), using the following primers: SRSF2-for: TTCGCCTTCTCCTCCTTT, SRSF2-rev: TCCGGCGTCCGTAGCCA. The single amplicon was analyzed by Sanger sequencing in all cases using BigDye Term v1.1 cycle sequencing chemistry (Applied Biosystems, Weiterstadt, Germany). Estimation of the mutational load was based on the electropherograms of the forward and reverse reactions. Additionally in 10 cases the mutational load was confirmed by next-generation sequencing, showing the correlation of both methods (Supplemental Figure S2B). Additional mutational data obtained by Sanger sequencing, next-generation deep amplicon sequencing or melting curve analyses were available in subcohorts and are described methodically elsewhere: ASXL1 exon 12 (n=261), CBL (n=274), EZH2 (n=208), IDH1/2 (n=82), JAK2V617F (n=275), KRAS codons 12/13 and 61 (n=266), NRAS codons 12/13 and 61 (n=273), RUNX1 (n=274), and TET2 (n=160). The coding sequence of SF3B1 (n=171) was analyzed by Sanger sequencing. U2AF1 Ser34 and Gln157 (n=265) were analyzed by melting curve analyses.
**In silico analyses**

For protein structure prediction, we used the Robetta prediction server (http://robetta.bakerlab.org). In first iteration, we applied Robetta to predict models for the known RRM-domain (2KN4.pdb) of the SRSF2 wild-type protein. Based on the resulting model, the 3D-full model option was applied to obtain a complete model of SRSF2. Next, we repeated these steps to generate full models for our detected novel mutations. The altered protein sequences were submitted to Robetta and resulting full models were compared to the SRSF2 wild-type model. For each submitted sequence, we selected the best model based on a manual validation process of the RRM domain. Finally, to analyze the differences between the best resulting models we calculated the Cα-Cα distances. For a more detailed report see Supplemental Text S1.

**Statistical analyses**

Statistical analyses were performed using SPSS version 19.0.0 (SPSS by IBM, Ehningen, Germany); the reported *p*-values are two-sided.

Survival curves were calculated for overall survival (OS) according to Kaplan-Meier and compared using the two-sided log rank test. OS was the time from diagnosis to death or last follow-up. Follow-up data was available in 180 cases, which were included in survival analyses. Results were considered significant at *p*<0.05. Adjustment for multiple testing was not done. Dichotomous variables were compared between different groups using the χ²-test and continuous variables by Student’s T-test.
Results

Characterization of 275 CMML patients

According to the WHO classification, the 275 patients were categorized as 193 CMML-1 and 82 CMML-2 cases, respectively. Morphological features of monocytes and monoblasts, and erythroid dysplastic changes are given in Supplemental Text S1. Based on biological parameters 61 patients were categorized to the MDA score, with 10 patients belonging to the low risk group, 15 to the intermediate-1, 25 to the intermediate-2, and 11 to the high risk group.

Cytogenetic analyses were carried out in 269/275 cases (in 6 cases, no metaphases were available). As typical in CMML a majority of patients had a normal karyotype (71%; 190/269); yet 29% (79/269) showed an aberrant karyotype. Within the aberrant karyotype group of 79 patients, a loss of the Y-chromosome (n=13), chromosome 7 aberrations (n=9), and a trisomy 8 (n=26) were the most frequent abnormalities (for further parameters, see Table 1). Therefore, 203 cases belong to the low risk category, whereas 27 belong to the intermediate and 39 to the high risk categories, defined by Such et al.4

Characterization and frequency of SRSF2 mutations

To analyze the mutation frequency of SRSF2 in our CMML cohort of 275 patients (Table 1), we investigated the sequence of an amplicon covering the mutation hotspot codon Pro95. Alterations of Pro95 or adjacent sequences were detected in 47% (129/275) of all cases. Mutation frequencies were similar in CMML-1 (47%; 91/193) and CMML-2 (46%; 38/82). In detail, 119 cases had a missense mutation leading to a change of Pro95 to one of the following five residues: p.Pro95His (n=56), p.Pro95Leu (n=38), p.Pro95Arg (n=23), p.Pro95Ala (n=1), p.Pro95Thr (n=1). In all cases, an estimated mutation load of 30-50% in accordance with a heterozygous mutation status was detected. One additional case revealed two different mutations p.[Arg94Pro;Pro95His] in a subset of 50% each. Next-generation sequencing demonstrated a mono-allelic mutation.

Interestingly, beyond the previously described missense mutations leading to alterations of Pro95, three new in/del mutations were observed, affecting the immediate neighboring amino acids (AA) of Pro95. In seven cases a deletion of 24 bp with a start in the codon of Pro95 resulted in deletion of eight AA, ranging from Pro96 to Arg103. All of these seven cases showed an additional missense mutation at Pro107 (p.[Pro96_Arg103del;Pro107His]). Furthermore, one single case showed a 24 bp duplication of the AA Arg86 to Gly93 (p.Arg86_Gly93dup), and another sample had a 3 bp duplication that resulted in an insertion of Arginine between Arg94 and Pro95 (p.Arg94_Pro95insArg). None of these mutations led to a frame shift. Buccal swab controls of two patients, carrying the p.[Pro96_Arg103del;Pro107His] mutation, were SRSF2 wild-type. Furthermore one
patient obtained this mutation during disease course. The NIH dbSNP databases as well as the NHLBI-Exome variant server both report no missense SNPs for the analyzed region (AA 86-107), indicating that these novel mutations are somatic mutations and no germ line polymorphisms. Figure 1 gives a schematic overview of the protein organization (based on information given by UniProtKB Q01130) and the mutation type, localization and frequency. SRSF2 mutated sequences are shown in Supplemental Figure S2A.

**In silico analyses**

In order to estimate the damaging character of these specific missense mutations at Pro95, we used SIFT (http://sift.jcvi.org), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml), and MutationTaster (www.mutationtaster.org) online analysis tools. The straightforward physical and comparative considerations revealed predominantly a damaging character for all missense mutations leading to AA exchanges at position Pro95 (see Supplemental Text S1).

To gain insights into the extent by which other SRSF2 mutations might alter the protein folding and therefore the protein function, we generated and compared structural models of SRSF2wt and SRSF2mut. A crystal structure of the SRSF2 protein is only available for the RRM-domain, and a complete structure for any of the SR-proteins has not yet been achieved. To evaluate any altering character of the three novel mutation types p.[Pro96_Arg103del;Pro107His], p.Arg86_Gly93dup, and p.Arg94_Pro95insArg on the protein structure, we used the Robetta server (http://robetta.bakerlab.org)\textsuperscript{35} to calculate a complete structural model of the wild-type SRSF2 protein and the different mutant SRSF2 proteins (Figure 2A).

The differences between these models were determined by calculating the Cα-Cα distances between the two corresponding AA of SRSF2wt and mutant SRSF2 for AA 88 to 99. This area covers the mutation hotspot Pro95 and represents the linker sequence (AA 92-117); and therefore, reflects the proper folding of the two functional domains (RRM and SR) relative to each other. The three analyzed novel mutations all demonstrated different distances relative to SRSF2wt, summarized in a table in Figure 2B. The 3 bp duplication showed the smallest divergence to the reference model with a distance range of 0.4 – 6.3 Å. The 24 bp deletion and the 24 bp insertion models show greater differences with distances ranging from 0.2 – 20.1 Å and 0.5 – 22.7 Å, respectively. Since only one AA is changed by the missense mutations, the models for the missense mutations show only slight divergences, being very congruent with the wt SRSF2 model. (Cα-Cα distances and a more detailed report about the whole procedure are given in Supplemental Text S1). These data show that all calculated models very well fit the known crystal structure of the RRM-domain up to AA 92, and larger changes appear within the mutated linker sequences.
Taken together, the in silico analyses indicate that the linker sequence, particularly amino acid 95, likely has a relevant impact on protein structure.

**Correlation of SRSF2 with karyotype**

As shown before, the majority of patients had a normal karyotype (71%; 190/269), whereas 29% (79/269) showed an aberrant karyotype. Within the aberrant karyotype group, the most frequent abnormalities were a loss of the Y-chromosome, aberrations of chromosome 7, and trisomy 8. Therefore, we correlated SRSF2mut with both a normal or aberrant karyotype and with subgroups exhibiting the respective chromosomal changes. These analyses revealed a normal karyotype in 81% of the SRSF2mut cases. Stated differently, in the group with a normal karyotype, 53% (101/190) had a SRSF2 mutation, whereas only 30% (24/79) were SRSF2mut in the aberrant karyotype group (p=0.001) (Table 1 and Fig. 3). Therefore, SRSF2mut correlated significantly with the low risk group (composed of normal karyotype and loss of the Y-chromosome) compared to the intermediate risk group (103/203, 51% vs. 8/27, 30%, p=0.043). No correlation of SRSF2mut was noted for the subcohorts with either loss of the Y-chromosome, chromosome 7 aberrations, or trisomy 8.

**Correlation of SRSF2 with biologic parameters**

Mutations in SRSF2 correlated with higher age (73.6 years vs. 71.5 years in the SRSF2wt cases, p=0.011) and higher hemoglobin levels (11.3 vs. 10.2 g/dl in the SRSF2wt cases, p=0.006), whereas white blood cell and platelet counts were not different. No correlations were observed between cases with SRSF2mut and the CMML categories 1 and 2 or gender (Table 1). There was also no significant correlation of SRSF2mut with other morphological features (Supplemental Text S1), any MDA risk category, or proliferative CMML (WBC counts >13,000/µl) and dysplastic CMML (WBC counts <13,000/µl).

**Coincidence of SRSF2 with other mutations**

We further investigated our CMML cohort regarding mutations in genes that have been recently described to be relevant in CMML. ASXL1, CBL, EZH2, KRAS, NRAS, IDH1/2, JAK2V617F, RUNX1, SF3B1, TET2, and U2AF1 were analyzed in large fractions of the 275 cases (Table 1). Comparison of the mutation frequencies of these genes showed that SRSF2 is the second most frequently mutated gene in this cohort (47%; 129/275) after TET2 (61%; 97/160), followed by ASXL1 (44%; 115/261), RUNX1 (22%; 61/274), CBL (19%; 51/274), NRAS (16%; 43/273), KRAS (11%; 28/266), EZH2 (10%; 20/208), and JAK2 (7%; 18/275). The mutation frequencies and associations are shown in Table 1 and Fig. 3, respectively. Mutations in IDH1/2, U2AF1, and SF3B1 occurred in ≤5% of patients and are therefore depicted in Supplemental Figure S3.
Analyses of coincidences showed that SRSF2 mutations were nearly mutually exclusive of EZH2 mutations. Of the 20 cases with an EZH2 mutation only one had a SRSF2 mutation. In counter-distinction, in the 208 cases with wt EZH2, SRSF2 was mutated in 106 samples (56%; p<0.001). In contrast, a high coincidence of SRSF2 mutations occurred with TET2 mutations as 62% (60/97) of the samples with TET2mut had a SRSF2 mutation; while in the TET2wt group, only 35% (22/63) also carried a mutation in SRSF2 (p=0.001). With respect to associations with all the other genes, no specific associations were observed (Fig. 3B). In a further analysis the coincidences of SRSF2mut with any other gene mutation were analyzed separately for CMML-1 and CMML-2 cases. Both groups reflect the same associations as observed in the total cohort (Supplemental Figure S4).

**Comprehensive analysis of gene mutations**

In a subset of 148 cases of the cohort, the mutational status data of nine genes were available (SRSF2, ASXL1, CBL, EZH2, JAK2, KRAS, NRAS, RUNX1, TET2). Overall, 93% (137/148) of the samples had at least one mutation in any of these genes, whereas only 7% (11/148) showed no molecular mutation. Eight of these eleven patients without mutation had a normal karyotype, three patients carried an aberrant karyotype. This consequently leads to a combined detection rate of alterations in 140/148 (95%) of patients with CMML having cytogenetic and/or molecular genetic aberrations. 12% (18/148) showed mutations in one gene; but in none of these 18 cases did a sole mutation of either SRSF2 or RUNX1 occur. Most of the cases had simultaneous mutations in two (33%; 49/148) or three (28%; 42/148) genes. In cases with mutations involving two genes, one of the two mutated genes was SRSF2 in 49% (24/49) of the samples. In these cases the mutational load of SRSF2mut was equal or beneath the mutational load of the second mutated gene. Four mutations occurred in 22/148 cases (15%). In only five patients mutations in five genes were observed (5/148; 3%), one patient carried mutations in seven genes (1/148; 1%).

**Impact of SRSF2 mutation on clinical outcome**

Follow-up data was available in 180 cases (median follow up of 12 months; median overall survival (OS): 29.6 months). This cohort comprised 117 CMML-1 (65%) and 63 CMML-2 (35%) cases, and 93 patients had mutations in SRSF2 (52%). Calculation of the OS for prognostic relevance of ASXL1, EZH2, TET2, and RUNX1 mutations in the total CMML cohort revealed an adverse impact of ASXL1mut compared to ASXL1 wild-type (wt) (median OS: 17.3 months vs. not reached (n.r.), p=0.001) and a slightly adverse impact of EZH2mut relative to EZH2wt (median OS: 18.3 vs. 29.3 months, p=0.073). TET2 and RUNX1 mutations showed no impact on OS (Supplemental Figure S5).
Finally, the influence of SRSF2 mutation on survival was analyzed. In the total cohort, no impact of SRSF2 mutations on OS was observed (Fig. 4A). Due to the high coincidence of SRSF2 mutations with TET2 mutations and the prognostic relevance of RUNX1 and ASXL1 alterations in MDS and CMML, respectively, we additionally analyzed these specific subcohorts, resulting in no statistically significant differences. Further, the three most frequently appearing missense mutations (Pro95His, Pro95Leu, and Pro95Arg) were analyzed separately. The OS curve of Pro95His mutated cases showed a slightly better course compared to the wt SRSF2 cases, whereas the OS of Pro95Leu and Pro95Arg was slightly shorter than of the wt (see Supplemental Figure S6). Based on these finding, we calculated the prognostic relevance of Pro95His separately in the above mentioned subcohorts: TET2mut, RUNX1mut, and ASXL1mut. Pro95His tends to have a favorable impact on OS in the RUNX1mut group compared to other SRSF2mut or SRSF2wt cases (median OS: n.r. vs. 18.3 months, p=0.066) (Fig. 4B). SRSF2mut had no influence on OS within any of the cytogenetic risk categories or MDA risk groups.
Discussion

A number of molecular targets have been identified that are frequently mutated in MDS or MDS/MPN. Thereby, some cellular pathways became apparent that are affected by mutations of several genes, including tyrosine kinase signaling and epigenetic regulation.\textsuperscript{5,6,8,20} Very recently, components of the splicing machinery were found to be frequently mutated in MDS, including mutations in \textit{U2AF1}, \textit{ZRSR2}, \textit{SF3B1}, and \textit{SRSF2}.\textsuperscript{7} All of these factors are involved in 3'-splice site recognition of pre-mRNA, inducing abnormal RNA splicing.

In the present study, we analyzed 275 CMML patients for mutations in \textit{SRSF2} and found a high frequency of mutations (47%). This frequency is even higher than the 28% that is described in the primary publication of Yoshida \textit{et al.}\textsuperscript{7} This difference in frequencies may be caused by ethnic differences of the two cohorts, more stringent patient selection and diagnostic procedures (using in all cases non-specific esterase for calculation of monocytes) or partially by methodological differences. For example the next-generation short read sequencing platform which was used in the previous study may have missed the in/del mutations. \textit{SRSF2}, therefore, belongs to the most frequently mutated genes in CMML together with \textit{TET2} and \textit{ASXL1}, with incidences of 61% and 44%, respectively, which is comparable to the frequencies of 44-50\%\textsuperscript{6,22} and 49\%\textsuperscript{20} in previous studies.

Of note, all other results of our mutational screening were in line with recently published data. \textit{RUNX1} was mutated in 22\% of the cases, which is in the range of findings reported by Kohlmann \textit{et al.},\textsuperscript{6} and Gelsi-Boyer \textit{et al.},\textsuperscript{37} with frequencies of 9\%, and 30\% respectively. Likewise, \textit{CBL} was mutated in 19\% of the cases; and therefore in the range of 13\% and 22\%, reported by Grand \textit{et al.}\textsuperscript{32} and Kohlmann \textit{et al.},\textsuperscript{6} respectively. In this paper with our enlarged cohort we also confirmed the mutation frequency of \textit{RAS} gene mutations of 30\%, also observed by Kohlmann \textit{et al.}\textsuperscript{6} (16\% for \textit{NRAS} and 11\% for \textit{KRAS}). Grossmann \textit{et al.}\textsuperscript{5} found a mutation frequency of 11\% for \textit{EZH2}, which was confirmed with 10\%. Levine \textit{et al.}\textsuperscript{38} noted a mutation frequency of 8\% for \textit{JAK2}, which is in line with the 7\% mutated cases observed in this study. \textit{IDH1/2} showed a mutation frequency of 5\%,\textsuperscript{5} being in line with 4\% presented in Jankowska \textit{et al.}.\textsuperscript{8}

The cytogenetic risk stratification suggested by Such \textit{et al.}\textsuperscript{4} could not be confirmed in this cohort by Kaplan-Meier analysis, which may be due to small case numbers for the intermediate (n=18) and low risk (n=27) categories. The median OS for the low and intermediate risk groups were not reached and the median OS for the high risk group was 21.1 months, but there was no statistically significant difference between the three cohorts. This is also true for the MDA risk stratification, where the case numbers were even smaller (low, n=7; intermediate-1, n=8; intermediate-2, n=17; high, n=8). The median OS for the low
and intermediate-1 risk groups were not reached and was 11.6 months for the intermediate-2 and 17.3 months for the high risk groups.

For functional insights of the SRSF2 mutations various computational analyses were performed. Since a crystal structure of SR-proteins is not available, bioinformatic tools were used to predict the character of the missense mutations and to generate SRSF2 structural models based on the AA sequence. All missense mutations of Pro95 in this study were predicted to be damaging. Recently, Daubner et al. analyzed the RNA binding mode of SRSF2 and indicated that Pro95 forms extensive contact with RNA. Also the three newly described mutations with deletions and insertions are suggestive of being even more deleterious. Comparison of the calculated models indicated that the mutations affected the linker sequence. Therefore, the topography of the two domains (RRM and RS) might have changed as a result of an altered number or structure of the AA. Considering the fact that no frame-shift or nonsense mutations occurred, the protein probably retains both structural integrity and any other modified function.

SRSF2 belongs to the SR-protein family and is therefore a splicing factor involved in alternative splicing (reviewed in ). Alternative splicing is an essential process by which eukaryotes generate high protein diversity from a single gene through the selective joining of different exons. Over 60% of human genes have been estimated to be alternatively spliced, indicating that regulation of alternative splicing is an important event. Mutations in both the nucleotide sequence of splicing regulatory elements and the components of the cellular splicing machinery can result in aberrant splicing. In addition, aberrant splicing has been found to be associated with various diseases, including cancer. Many cancer-related genes are regulated by alternative splicing and changes in the splicing pattern appear to be unique to the malignant state. Daubner et al. report, that mutations of affecting the RNA recognition motif (RRM) also affect the function, showing a decreased splicing activity of the protein. More recently, Makishima et al. showed that SRSF2mut leads to defective splicing of the RUNX1 gene. Moreover, miss-expression of SR-proteins changes the alternative splicing pattern, and is associated with the development of cancer. Increased expression of SR-proteins correlates with cancer progression, as was shown for SRSF2 in ovarian cancers. On the other hand, depletion of SRSF2 in the thymus of a mouse model changed the alternative splicing of CD45, causing a defect in T-cell maturation. Lareau et al. reported that SRSF2 directs the splicing of its own transcripts and autoregulates its own expression by coupling alternative splicing with RNA decay. Recent reports indicate further functions of SRSF2 in transcription, promoting RNA Pol II elongation, genome stability, and cell-cycle progression (reviewed in ). Taken together, mutations in SRSF2, although occurring in a region without any obvious functional domain, may cause changes in protein
function or expression levels, both possibly contributing to a change of alternative splicing patterns, leading to developmental defects and the onset of cancer.

SRSF2 mutations frequently overlapped with other mutations in our cohort of 275 CMML patients. Only mutations of EZH2 did not overlap, pointing to their mutual exclusiveness. One may speculate that this occurs because either no advantageous cooperating effect results from both proteins being altered, or concomitant mutations of both proteins is lethal for the cell. Overall, in 18 cases only one mutation was detected and this was never SRSF2. Thus, SRSF2 never occurs as a sole mutation, either indicating that SRSF2 mutations are not early events in the pathogenesis of CMML, or that a sole mutation in SRSF2 results in no clinical manifestation. This is further supported by the fact, that the mutational load of SRSF2mут was always equal or below the mutational load of the second mutated gene in cases with only two mutations. On the other hand, SRSF2 is frequently mutated in cases with either two or three mutations. SRSF2 mutations may result in a dysfunction of the protein affecting transcriptional elongation and therefore genome stability. Depletion of SRSF2 has been reported to trigger overwhelming double strand breaks (reviewed in49).

Mutations in SRSF2 were highly associated with TET2 mutations, a protein converting 5-methyl-cytosin to 5-hydroxymethyl-cytosin. Depletion of TET2 in bone marrow progenitor cells promotes an expansion of monocyte/macrophage cells,50 indicating that loss of function can promote clonal expansion of mutant cells. Addressing the white blood cell (WBC) count in cases with TET2wt + SRSF2wt (n=32) revealed a mean of 16,036 cells per µl, whereas TET2mut + SRSF2wt cases (n=33) showed 31,112 WBC/µl (p=0.044). SRSF2mут seems to antagonize this leukocytosis, mostly by monocytosis, since the mean WBC count was 16,864/µl in cases with TET2mut + SRSF2mut (n=57; TET2mut + SRSF2mut vs. TET2wt + SRSF2wt, p=0.047). As mentioned above, SRSF2 depletion has been reported to cause genome instability by triggering double strand breaks, which induced the S phase checkpoint and ended in cell cycle arrest or apoptosis (reviewed in49). Furthermore, SRSF2 mutation is correlated with higher hemoglobin levels. Thus, patients with SRSF2mut show a less pronounced leuko/monocytosis in the presence of a concomitant TET2 mutation, and have a less pronounced anemia, both indicating a better state of health.

The median OS of our cohort is 29.6 months, indicating that the outcome of our cohort is somehow better than in other data sets published (e.g.3). This may be due to the fact that 60% of the patients sent for diagnosis to our institution are referred from outpatient units and hematologists´ practices at first suspect of CMML and thus are diagnosed very early. Many of our patients were not treated upfront but followed a watch-and-wait-strategy. This may in part explain the differences in the survival curves in comparison to other studies published from centers to which the patients were referred in order to receive treatment.
including also enrolment into clinical trials. The mutational status of \textit{SRSF2} did not affect OS, although the median OS was not reached in \textit{SRSF2} wt cases in contrast to 29.6 months in those with a mutated \textit{SRSF2} (\(p=0.858\)). In \textit{RUNX1} mutated cases the addition of a \textit{SRSF2} mutation prolonged the OS. Analyzing the most frequently occurring \textit{SRSF2} missense mutations (Pro95His, Pro95Leu, and Pro95Arg) separately revealed that CMML with a Pro95His showed a better outcome compared to the other two frequent mutations as well as the wild-type \textit{SRSF2}, in the \textit{RUNX1}, \textit{TET2}, and \textit{ASXL1} mutated groups. This goes in line with the idea that a \textit{SRSF2} mutation, especially Pro95His, affects protein function; and this may result in a favorable impact in cases with concomitant (adverse) mutation, possibly due to inhibition of cell cycle progression.

In summary, \textit{SRSF2} mutations are very common in CMML and seem to have a deleterious impact on protein structure and function. This may on the one hand result in promoting further gene mutations and therefore disease progression. On the other hand, it could have a favorable impact on the OS of patients with an additional (adverse) mutation. \textit{SRSF2} mut further correlated with a normal karyotype and confined the cytogenetic categories low and intermediate. \textit{SRSF2}, therefore, represents a novel molecular marker that is helpful for diagnosis of CMML or suspected CMML and for further genetic characterization of this disease. A possible positive prognostic effect in cases with other, partially adverse mutations (\textit{RUNX1}, \textit{TET2}, and \textit{ASXL1}), that was suggested based on our results has to be validated in further independent studies. Of note, based on this data, overall 93% of CMML patients in the present cohort carried at least one mutated gene. However, cases are still found without any detectable genetic defect, warranting further efforts to identify new genetic aberrations that are essential to better understand the molecular pathology of this disease.
**Authorship**

**Contribution:**
M.M. investigated the molecular mutations of *SRSF2* and *ASXL1*, analyzed the data, and wrote the manuscript; A.R. made the bioinformatic analyses; T.H. was responsible for cytomorphologic analysis and was involved in the collection of clinical data; C.E., F.D., V.G., and A.K. contributed to molecular analyses of the *ASXL1*, *CBL*, *EZH2*, *JAK2V617F*, *KRAS*, *NRAS*, *RUNX1*, and *TET2* mutations; T.A. collected and documented clinical data and compiled statistical analyses; K.Y., S.O., and H.P.K. originally detected *SRSF2* gene mutations and shared unpublished data; W.K. was responsible for immunophenotyping and was involved in statistical analyses; C.H. was responsible for chromosome banding analysis; S.S. was the principle investigator of the study and wrote the manuscript. All authors read and contributed to the final version of the manuscript.

**Conflict of interest:**
References


Table 1

<table>
<thead>
<tr>
<th>Parameter (available cases n)</th>
<th>Patient numbers (% or ranges)</th>
<th>total cohort (n=275)</th>
<th>SRSF2mut (n=129, 47%)</th>
<th>SRSF2wt (n=146, 53%)</th>
<th>p=</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male/female (ratio)</td>
<td></td>
<td>189/86 (2.2)</td>
<td>91/38 (2.4)</td>
<td>98/48 (2.0)</td>
<td>n.s.</td>
</tr>
<tr>
<td>CMML-1</td>
<td></td>
<td>193 (70%)</td>
<td>91 (47%)</td>
<td>102 (53%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>CMML-2</td>
<td></td>
<td>82 (30%)</td>
<td>38 (46%)</td>
<td>44 (54%)</td>
<td></td>
</tr>
<tr>
<td>median age (years)</td>
<td></td>
<td>72.8 (21.9-93.3)</td>
<td>73.6 (49.9-89.5)</td>
<td>71.5 (21.9-93.3)</td>
<td>0.011</td>
</tr>
<tr>
<td>median WBC (x10^3/µl)</td>
<td>(n=247)</td>
<td>15.3 (0.9-160.0)</td>
<td>17.4 (2.2-113.2)</td>
<td>12.9 (0.9-160.0)</td>
<td>n.s.</td>
</tr>
<tr>
<td>median platelets (x10^3/µl)</td>
<td>(n=227)</td>
<td>90.0 (3.0-1,385)</td>
<td>80.5 (3.0-1,119)</td>
<td>105.0 (5.0-1,385)</td>
<td>n.s.</td>
</tr>
<tr>
<td>median Hb (g/dl)</td>
<td>(n=226)</td>
<td>11.0 (4.0 – 18.2)</td>
<td>11.3 (6.0-15.5)</td>
<td>10.2 (4.0-18.2)</td>
<td>0.006</td>
</tr>
<tr>
<td><strong>Cytogenetics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normal karyotype</td>
<td></td>
<td>190 (71%)</td>
<td>101 (53%)</td>
<td>89 (47%)</td>
<td>0.001</td>
</tr>
<tr>
<td>aberrant karyotype</td>
<td></td>
<td>79 (29%)</td>
<td>24 (30%)</td>
<td>55 (70%)</td>
<td></td>
</tr>
<tr>
<td>trisomy 8</td>
<td></td>
<td>26 (33%)</td>
<td>9 (35%)</td>
<td>17 (65%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>-Y</td>
<td></td>
<td>13 (17%)</td>
<td>2 (15%)</td>
<td>11 (85%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>chromosome 7 aberration</td>
<td></td>
<td>9 (11%)</td>
<td>4 (44%)</td>
<td>5 (56%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>complex</td>
<td></td>
<td>4 (5%)</td>
<td>1 (25%)</td>
<td>3 (75%)</td>
<td></td>
</tr>
<tr>
<td>all other</td>
<td></td>
<td>27 (34%)</td>
<td>8 (30%)</td>
<td>19 (70%)</td>
<td></td>
</tr>
<tr>
<td><strong>Molecular mutations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASXL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=261)</td>
<td>115 (44%)</td>
<td>56 (49%)</td>
<td>59 (51%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>146 (56%)</td>
<td>68 (47%)</td>
<td>78 (53%)</td>
<td></td>
</tr>
<tr>
<td>CBL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=274)</td>
<td>51 (19%)</td>
<td>27 (53%)</td>
<td>24 (47%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>223 (81%)</td>
<td>101 (45%)</td>
<td>122 (55%)</td>
<td></td>
</tr>
<tr>
<td>EZH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=208)</td>
<td>20 (10%)</td>
<td>1 (5%)</td>
<td>19 (95%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>188 (90%)</td>
<td>106 (56%)</td>
<td>82 (44%)</td>
<td></td>
</tr>
<tr>
<td>JAK2V617F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=275)</td>
<td>18 (7%)</td>
<td>9 (50%)</td>
<td>9 (50%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>257 (93%)</td>
<td>120 (47%)</td>
<td>137 (53%)</td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=266)</td>
<td>28 (11%)</td>
<td>10 (36%)</td>
<td>18 (64%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>238 (89%)</td>
<td>117 (49%)</td>
<td>121 (51%)</td>
<td></td>
</tr>
<tr>
<td>NRAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=273)</td>
<td>43 (16%)</td>
<td>17 (40%)</td>
<td>26 (60%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>230 (84%)</td>
<td>111 (48%)</td>
<td>119 (52%)</td>
<td></td>
</tr>
<tr>
<td>RUNX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=274)</td>
<td>61 (22%)</td>
<td>34 (56%)</td>
<td>27 (44%)</td>
<td>n.s.</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>213 (78%)</td>
<td>94 (44%)</td>
<td>119 (56%)</td>
<td></td>
</tr>
<tr>
<td>TET2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mutated</td>
<td>(n=160)</td>
<td>97 (61%)</td>
<td>60 (62%)</td>
<td>37 (38%)</td>
<td>0.001</td>
</tr>
<tr>
<td>wild-type</td>
<td></td>
<td>63 (39%)</td>
<td>22 (35%)</td>
<td>41 (65%)</td>
<td></td>
</tr>
</tbody>
</table>
Table and Figure Legends:

Table 1
Clinical characteristics, cytogenetics, and molecular mutations of 275 CMML patients. Parameters describing the three groups: total cohort, SRSF2mut, and SRSF2wt are listed. Analyzed case numbers n and percentages or ranges are provided in parenthesis. p-values are given for significant differences. n.s. = not significant.

Figure 1
Schematic overview of SRSF2 protein organization, mutation type, and mutation frequency. SRSF2 consists of an RNA Recognition Motive (RRM, AA 14-92, light grey box), a linker region (white box) and an Arginin/Serine-rich domain (RS, AA 117-221, dark grey box). The four different mutation types and their mutation localization are indicated in red. The mutation frequency is listed and also illustrated as black triangles above each mutation type. Each triangle is representing one mutated case. From top to bottom Pro95 missense mutations, p.Pro96_Arg103del;Pro107His, p.Arg86_Gly93dup, and p.Arg94_Pro95insArg alterations are depicted.

Figure 2
In silico analyses of the structural models of the SRSF2 in/del mutations. The structural changes of the in/del mutations were calculated with the Robetta server algorithm (http://robetta.bakerlab.org).35 The calculated model for the complete SRSF2wt protein (white structure) is depicted in (A), the mutational hotspot Pro95 is marked in red. In addition, the enlargement shows the structure of AA 61-129 of the calculated models: p.[Pro96_Arg103del;Pro107His]; p.Arg86_Gly93dup; and p.Arg94_Pro95insArg. The Cα-Cα distance measurement of the corresponding AA of SRSF2wt to SRSF2mut is illustrated exemplarily for Pro95. All Å-values of the Cα-Cα distance measurements for AA 88-99 are given in a table in (B).

Figure 3
Alignment of gene mutations, karyotype information, and CMML category for 275 patients. (A) Each column represents one of the 275 analyzed samples. Analyses of nine investigated genes, the karyotype, and CMML category-1 or -2 are depicted by colored bars. Red bar: mutated gene; dark grey bar: non-mutated gene; white bar: no data available; light grey bar: normal karyotype; black bar: aberrant karyotype; grey bar: CMML-1; anthracite bar: CMML-2. (B) Concomitant events of SRSF2 with other mutations are also shown as a bar chart. The grey part represents SRSF2wt, the red one SRSF2mut within the analyzed
subcohorts. SRSF2mut frequencies and significances (p-values) are denoted; numbers of mutated/analyzed cases of the subcohorts are given in parenthesis below the bars.

Figure 4
Overall survival by Kaplan-Meier analyses of CMML patients according to SRSF2 mutations. (A) Overall survival of patients with SRSF2mut did not significantly differ from patients with SRSF2wt. (B) Overall survival of SRSF2 Pro95His positive patients compared to patients with all other SRSF2 mutations and SRSF2wt (= all other) within the RUNX1 mutated subcohort showed a favorable trend. Overall survival is indicated in months and was compared using the two-sided log rank test. p-values are denoted in each graph, respectively.
Fig. 1

- p.Pro95His (n=57)
- p.Pro95Leu (n=38)
- p.Pro95Arg (n=23)
- p.Pro95Glu (n=1)
- p.Pro95Trp (n=1)

(n=120)

- p.Pro96_Arg103del;Pro107His (n=7)

- p.Arg86_Gly93dup (n=1)

- p.Arg94_Pro95insArg (n=1)
Fig. 2

A

B

<table>
<thead>
<tr>
<th>Calculated mutation model</th>
<th>Cα - Cα distance of reference to mutation model for AA 88 to 99 in Å</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>RRM</td>
</tr>
<tr>
<td>p.Pro66_Arg103del:Pro107His</td>
<td>0.2</td>
</tr>
<tr>
<td>p.Arg86_Gly93dup</td>
<td>0.7</td>
</tr>
<tr>
<td>p.Arg94_Pro95insArg</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Fig. 3

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutations</th>
<th>Normal Karyotype</th>
<th>CMML-1</th>
<th>CMML-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRSF2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TET2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EZH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUNX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASXL1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAK2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Karyotype

CMML-1/2

B

Bar chart showing the distribution of mutations in different genes:

- **TET2mut** (97/160): 62%
- **EZH2mut** (20/208): 5%
- **RUNX1mut** (61/274): 57%
- **ASXL1mut** (115/281): 49%
- **CBLmut** (51/274): 53%
- **KRASmut** (28/266): 36%
- **NRASmut** (43/273): 40%
- **JAK2mut** (18/275): 50%

**p = 0.001**

**p = 0.001**
Fig. 4

A  total cohort

\[
p = 0.858
\]

Overall Survival

SRSF2wt (n=87; median OS n.r.)

SRSF2mut (n=93; median OS=29.6)

Months

B  RUNX1 mutated

\[
p = 0.066
\]

Overall Survival

SRSF2 Pro95His (n=10; median OS n.r.)

all other (n=35; median OS=18.3)

Months
SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML)