Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia

Renée Beekman¹, Marijke G. Valkhof¹, Mathijs A. Sanders¹, Paulette M.H. van Strien¹, Jurgen R. Haanstra¹, Lianne Broeders¹, Wendy M. Geertsma-Kleinekoort¹, Anjo J.P. Veerman², Peter J.M. Valk¹, Roel G. Verhaak³, Bob Löwenberg¹, Ivo P. Touw¹

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

²Department of Pediatric Hematology-Oncology, Vrije Universiteit Medical Center, Amsterdarm, The Netherlands;

³Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston TX, USA.

Running title: Sequential gain of mutations in leukemogenesis.

Correspondence:
Ivo P. Touw, PhD
Department of Hematology, Erasmus University Medical Center
Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
Phone: +31-10-7043837
Fax: +31-10-7044745
Email: i.touw@erasmusmc.nl

Abstract word count: 199
Text word count: 3572
Number of tables: 1
Number of figures: 4
Number of references: 43
Abstract

Severe congenital neutropenia (SCN) is a bone marrow failure syndrome with a high risk to progress towards acute myeloid leukemia (AML). The underlying genetic changes involved in SCN evolution to AML are largely unknown. We obtained serial hematopoietic samples of an SCN patient who developed AML 17 years after initiation of granulocyte-colony stimulating factor (G-CSF) treatment. Next-generation sequencing was done to identify mutations during disease progression. In the AML phase, we found 12 acquired non-synonymous mutations. Three of these, in CSF3R, LLGL2 and ZC3H18, co-occurred in a subpopulation of progenitor cells already in the early SCN phase. This population expanded in time, whereas clones solely harboring CSF3R mutations disappeared from the bone marrow. The other 9 mutations were only apparent in the AML and affected known AML-associated genes (RUNX1 and ASXL1) and chromatin remodelers (SUZ12 and EP300). In addition, a novel CSF3R mutation was found conferring autonomous proliferation to myeloid progenitors. We conclude that progression from SCN towards AML is a multistep process with distinct mutations arising early during the SCN phase and others later in AML development. Sequential gain of two CSF3R mutations implicates abnormal G-CSF signaling as a driver of leukemic transformation in this case of SCN.
Introduction

Severe congenital neutropenia (SCN) is a bone marrow failure syndrome characterized by strongly reduced neutrophil counts and recurrent, potentially life threatening, opportunistic bacterial infections. Treatment with granulocyte-colony stimulating factor (G-CSF) elevates peripheral neutrophil counts and reduces the risk of infections. Leukemic progression of SCN is a major concern, with an estimated overall cumulative incidence of approximately 20% after 15 years of G-CSF treatment.

Constitutional mutations in the gene encoding neutrophil elastase (ELANE) are common defects in SCN. In addition, the acquisition of nonsense mutations in the gene encoding the granulocyte-colony stimulating factor receptor (CSF3R) is a unique feature in SCN patients. These mutations lead to expression of truncated CSF3R proteins, also known as delta forms. In cell line models, truncated CSF3R are hampered in transducing signals required for proper neutrophil differentiation. Additionally, they confer increased proliferative responses to G-CSF treatment but do not cause leukemia in mice. CSF3R delta mutations can be detected in approximately 30% of SCN patients. In some cases, distinct clones with different CSF3R delta mutations are present for many years. After evolution of SCN towards AML, CSF3R delta mutations are found in approximately 80% of the cases. Until now, all reported SCN/AML cases harboring a CSF3R delta mutation in the SCN phase also carry this mutation in the leukemic phase. These observations suggest that leukemic progression in SCN follows a unique pattern, with CSF3R delta mutations as an early event, followed by additional genetic and epigenetic events that are essential for full leukemic transformation. Chromosomal aberrations, e.g., loss of chromosome 7 and gain of chromosome 21, are apparent in AML arising from SCN and other bone marrow failure syndromes like Fanconi anemia and Shwachman-Diamond syndrome. However, mutations that are quite commonly seen in de novo AML have not been reported in AML arising from SCN. Thus, the additional molecular events involved in leukemic progression of SCN remain largely unknown.

To identify the sequential genetic events in leukemic progression of SCN towards AML, we collected serial hematopoietic samples of an SCN patient who developed AML after 17 years of G-CSF therapy. Using whole exome sequencing, we found 12 somatic non-synonymous mutations in the leukemic blasts of this patient. Three of these mutations, the known CSF3R mutation and mutations in LLGL2 and ZC3H18, were already present at low frequencies in the early SCN phase,
15 years before AML was diagnosed. Myeloid colony analysis showed that these 3 “early” mutations co-existed in the same hematopoietic progenitors in a small subpopulation of bone marrow cells. Six years later, in the “intermediate” SCN phase, still 9 years before the AML became overt, we observed an expansion of the clone harboring all 3 mutations. The other 9 of 12 mutations were only apparent in the AML. The latter “late” appearing mutations comprise a second, novel, CSF3R mutation in addition to a series of new and known AML-associated mutations. The novel CSF3R mutation is located on the already mutated CSF3R-d715 allele and causes growth factor independence of myeloid progenitors.
Materials and Methods

Case report

A 27-year old male SCN patient was diagnosed with AML 17 years after the start of G-CSF treatment (10μg/kg/day), on which he reached normal neutrophil counts. The patient had a constitutional heterozygous ELANE mutation, G174R. At the age of 12, 2 years after G-CSF treatment was initiated, a CSF3R truncation or delta mutation (CSF3R-d715) was discovered in the bone marrow. At the time of AML diagnosis, the peripheral blood contained 24% blasts and dysplasia was observed in the bone marrow. G-CSF treatment was stopped at this point. Six weeks later, a bone marrow analysis revealed 17% blasts. Immunophenotypically, these blasts were of myeloid nature, i.e., positive for CD34, CD117, CD13, CD133, CD33, MPO and CD90. Because no HLA-identical donor was available, the patient received a matched unrelated donor (MUD) allogeneic bone marrow transplant. Induction therapy was given according to the induction therapy scheme HOVON42A of the Hemato-Oncology Foundation for Adults in the Netherlands. At initiation of induction therapy, the bone marrow contained 15.7% blasts, with 10-50% dysplasia in all lineages. Routine cytogenetic and molecular diagnostics revealed a trisomy 21 (47, XY, +21 /46, XY), with no additional abnormalities (AML-ETO, CBF/MYH11, FLT3/ITD, FLT3/TKD, mutations in NPM1, NRAS, KRAS, c-KIT, JAK2 and CEBPA). After the second induction cycle trisomy 21 was undetectable in a marrow cytogenetic analysis. The MUD transplant was administered after myeloablative conditioning with chemotherapy and total body irradiation. Two months after the transplant 28% blast were detected in the bone marrow, indicating a recurrence of the AML and the patient died 3.5 months after the transplant. Figure 1 gives a schematic overview of the disease history.

Patient cell samples

Ficoll-gradient separated bone marrow cells from the SCN phases and CD34+ leukemic blasts from the peripheral blood in the leukemic phase were used. Control DNA was isolated from bone marrow-derived fibroblasts. All cell samples were obtained and frozen according to established procedures for viable cell cryopreservation as previously described. The study was performed under the permission of the Institutional Review Board of the Erasmus MC, registration number MEC-2008-
387 for biobanking and MEC-2012-030 for the genetic analysis of leukemic progression in SCN patients.

**Nucleotide sequencing**

**Whole Exome sequencing (WES)**

Sequencing libraries were prepared according to the SureSelect Target Enrichment system for Illumina, protocol version 2.2.1, Nov. 2010. In short, 3 μg genomic DNA was sheared to fragments of approximately 170 base pairs using the Covaris S-series Single Tube Sample Preparation System, Model S2 (Covaris, Woburn, MA, USA). Fragment sizes were checked on the Bioanalyzer (Agilent, Santa Clara, CA). Adapter ligated libraries were prepared according to the manufacturer’s protocol using the Paired-End Genomic DNA Sample Prep Kit PE-102-1001 (Illumina, San Diego, CA); 5 cycles of amplification were used. Five hundred ng of prepped library was taken for hybridization with the SureSelect Human All Exon Kit (G3362A, Agilent). A sample concentration of 5.5 picomolar was loaded for sequencing on the Hiseq2000 (Illumina) using 101-bp paired-end reads.

Sequencing reads were processed with the Casava pipeline (version 1.7, Illumina). For alignment the Hg18/NCBI36 assembly (March 2006) was used. Detection of single nucleotide variants, deletions and insertions was performed with otherwise default settings, while snpCovCutoff and indelsCovCutoff were switched off. Variations detected in the AML sample in 2 independent sequence runs were further analyzed after removal of germ-line variations (present in the fibroblasts) and single nucleotide polymorphisms (SNPs, dbSNP)\textsuperscript{17}. Next, non-synonymous variants were determined. Integrative Genome Browser was used for sequence read visualization\textsuperscript{18}.

**Sanger sequencing**

WES results were validated by Sanger sequencing, performed according to the manufacturer’s protocol (Applied Biosystems, Foster City, CA, USA) using primers indicated in Table S1. Before amplicon generation, genomic DNA or cDNA was first amplified using a Whole Genome Amplification kit (WGA2, Sigma-Aldrich, Zwijndrecht, The Netherlands). DNA was purified with a PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol and diluted to 50 ng/μl. Hundred nanograms of amplified DNA was used for amplicon generation; cycling conditions were 30” at 95°C, 30” at the indicated annealing temperature (Table S1) and 45’ at 72°C for 35
cycles. In some instances, the unamplified material was used directly for Sanger sequencing (Table S1).

**Amplicon-based deep sequencing**

Amplicons were generated and purified according to the Amplicon Library Preparation Method Manual (version May 2010, Roche, Basel, Switzerland). Primers and annealing temperatures are indicated in Table S2; 35 cycles were used for amplification. DNA enriched beads, carrying the amplification products, were generated according to the emPCR Amplification Method Manual – Lib-A (version May 2010, Roche); a beads to amplicon ratio of 1:2 was used. Amplicons were analyzed with the GS junior (Roche). Sequence reads were analyzed using the GS Amplicon Variant Analyzer (Roche). For the SCN samples, coverage of at least 1600 was achieved to identify mutations present in minor clones within the bone marrow. For the AML sample coverage of 80 was considered sufficient to validate mutations.

**Human myeloid colony assay**

Bone marrow was thawed at 37°C, washed twice with IMDM (Gibco Invitrogen, San Diego, CA) with 10% FCS (PAA laboratories, Pasching, Austria). Per 4 ml of culture medium, 2.9 ml methocult (H4230, Stem Cell Technologies, Vancouver, Canada), 980 μl IMDM and human GM-CSF (Immunex, Seattle, WA), human G-CSF (Neupogen, Amgen, Thousand Oaks, CA) and human IL-3 (R&D Systems, Minneapolis, USA) in final concentrations of respectively 2 ng/ml, 200 ng/ml and 25 ng/ml were used. Cells were plated at a density of 0.8 x 10^5 /ml. After 2 weeks genomic DNA of single colonies was isolated, followed by amplification using the Whole Genome Amplification kit and Sanger sequencing of CSF3R-d715, LLGL2 and ZC3H18, as described above. Results were validated in an independent round of whole genome amplification for i) colonies harboring a mutation, ii) colonies with unclear sequences and iii) a number of randomly chosen non-mutated colonies to rule out amplification artifacts. All colonies harboring mutations in CSF3R, LLGL2 or ZC3H18 were also analyzed for the presence of the remaining 9 mutations found in the AML sample.

**Murine colony assays**
Four different CSF3R expression constructs (WT, d715, T595I, d715-T595I) were generated and retrovirally transduced into bone marrow cells of Csf3r deficient FVB/N mice. Colony assays of these transduced progenitors were performed as previously described.

Further details of these procedures are given in the Materials and Methods section of the Supplementary Appendix.
Results

Whole exome sequencing reveals acquired mutations in SCN/AML

Whole exome sequencing was done on genomic DNA from the CD34+ leukemic blast fraction and the fibroblast control sample. Acquired non-synonymous mutations were detected by identification of single nucleotide variants and small insertions and deletions, followed by subtraction of variants present in the control fibroblasts and known single nucleotide polymorphisms. Twelve non-synonymous acquired mutations were identified and validated by Sanger sequencing (Table 1, Figure S1). Except for the mutation in \textit{FBXO18}, all mutations occurred in evolutionary conserved amino acids (Figure S2). With the exception of \textit{LAMBI}, all mutant transcripts were detectably expressed in the leukemic blasts (Table 1, Figure S3). Mutations in \textit{ASXL1} and \textit{RUNX1} are known in myeloid malignancies. Deletions in \textit{EP300}, distinct from the 7-bp deletion found in this patient, have been reported in lymphomas. The ATT insertion in \textit{SUZ12} duplicates an isoleucine at amino acid position 597, located in the conserved VEFS-box. Mutations in this region, which is involved in the interaction between SUZ12 and the histone methyltransferase EZH2 in the polycomb repressor complex 2 (PRC2), have recently also been identified in myelodysplastic/myeloproliferative neoplasms (MDS/MPN) with 17q abnormalities. As expected, the previously identified \textit{CSF3R} delta mutation (\textit{CSF3R-d715}) was present in the leukemic blasts, but remarkably a new \textit{CSF3R} mutation, T595I, was now also present. Furthermore, the \textit{CSF3R}-T595I mutation was located on the same allele as the delta mutation, as determined by Sanger sequencing of single amplicons generated from cDNA. Using exome sequencing data from 199 AML cases reported by The Cancer Genome Atlas (TCGA), a similar single \textit{CSF3R}-T595I mutation was detected. Additionally, mutations in \textit{ASXL1} (n=5), \textit{CCDC155} (n=1), \textit{LLGL2} (n=1), \textit{MGA} (n=1), \textit{RUNX1} (n=17), \textit{SUZ12} (n=2) and \textit{ZC3H18} (n=2) were found in the TCGA data set (Table S3; R.G.V. and The Cancer Genome Atlas disease working group, unpublished data).

Amplicon-based sequencing reveals an early pre-leukemic clone that expands over time

Amplicon-based deep sequencing was applied to analyze the presence of all 12 somatic mutations in bone marrow samples obtained at 15 and 9 years before AML was diagnosed (Figure 1). Not only the known \textit{CSF3R-d715} mutation, but also mutations in \textit{LLGL2} and \textit{ZC3H18} were already present in these earlier disease phases (Figure 2A, Table S4). We investigated the clonal hierarchy of
these mutations in single myeloid colonies cultured from the earliest bone marrow sample (15 years before AML). In the individual colonies (n=88), the mutation status of CSF3R-d715, LLGL2 and ZC3H18 was determined. Fifteen colonies (17%) harbored both the CSF3R-d715 and the LLGL2 mutation, whereas none of the colonies exhibited exclusively either the LLGL2 or the CSF3R-d715 mutation (Figure 2B, Table S5). Two of the CSF3R-d715 and LLGL2 mutated colonies also carried the ZC3H18 mutation (Figure 2B, Table S5), indicating that this mutation had emerged later in time. None of the other 9 mutations found in the AML cells were apparent in these colonies (Table S5).

A previous report has shown that multiple CSF3R delta mutations can be present in distinct progenitors in the bone marrow of an individual SCN patient. In line with this, we found myeloid colonies with CSF3R-d717 (n=2) and CSF3R-d725 (n=1) (Figure 2B, Table S5). Each of these mutations and yet an additional delta mutation (CSF3R-d730) were detected in the SCN phase at low frequencies by amplicon-based deep sequencing (Figure 2C, Table S6). None of these variant CSF3R mutant clones harbored LLGL2 or ZC3H18 mutations, nor were they seen as dominant clones in the AML (Figure 2, Table S5-S6). No viably frozen cells were available from the bone marrow sample obtained 9 years before AML development and colony analysis could not be performed at this stage. However, by amplicon-based deep sequencing we observed a parallel increase of the CSF3R-d715, LLGL2 and ZC3H18 mutations from 15 to 9 years before AML development (Figure 2A). Together with the finding that these mutations are present in the same myeloid progenitor cells (Figure 2B), this observation is consistent with a selective outgrowth of clones carrying these 3 mutations.

**Sequential gain of a second CSF3R mutation results in G-CSF independence**

A new CSF3R mutation, acquired at the CSF3R-d715 mutant allele, was found exclusively in the AML blasts and changed a polar threonine residue at amino acid position 595 into a nonpolar isoleucine. This residue is located in a highly conserved threonine-rich region in the extracellular domain of the G-CSF receptor (Figure S2). Introduction of human CSF3R mutant receptors, carrying this new T595I mutation (Figure 3A), in Csf3r deficient primary mouse bone marrow progenitors resulted in the autonomous outgrowth of myeloid colony-forming cells (Figure 3, Table S7). Thus, in the AML phase of disease evolution two different co-existing mutations, i.e., the T595I single amino acid substitution and the CSF3R-d715 mutation, had accumulated in the gene of the G-CSF receptor.
Because expression of the new CSF3R mutant without the delta mutation conferred G-CSF independence as did the mutant receptor carrying both the delta and the extracellular mutation, this gain of function can entirely be attributed to the T595I mutation. However, the T595I/d715 colonies were bigger than the T595I colonies (Figure S4), suggestive of a higher proliferation capacity by the addition of the CSF3R-d715 mutant.
Discussion

This study has elucidated non-synonymous mutations acquired in an SCN patient who progressed to AML. The availability of sequential hematopoietic samples from the childhood SCN phase to overt AML, spanning a period of 17 years, provided the unique opportunity to identify the early and late genetic defects associated with leukemic progression (Figure 4). The *CSF3R*-d715 mutation and a mutation in *LLGL2*, encoding the human homologue of the Drosophila lethal giant larvae (*Lgl*) gene, were the first 2 acquired mutations in the early SCN phase. Loss of *Lgl* in Drosophila leads to inadequate distribution of the cell polarity protein Numb, resulting in inappropriate cell fate determinations and tumor formation in epithelial tissues and the brain. In man, the NUMB protein has been implicated in controlling the balance between symmetric versus asymmetric hematopoietic stem cell divisions. Interestingly, deregulation of NUMB expression has been associated with blast transformation of chronic myeloid leukemia. How the *LLGL2* mutation found in this study affects hematopoietic stem cell divisions is still unknown; the fact that *CSF3R*-d715 and *LLGL2* mutations were uniformly present in the same myeloid cell clones could suggest that they cooperate. Hierarchically, the next genetic abnormality occurring in the early SCN phase in the *CSF3R*-d715 and *LLGL2* mutated clone was a mutation in *ZC3H18*. *ZC3H18* is a putative mRNA binding protein with a still unknown function, but in trypanosomes it is shown to be essential for differentiation.

Additionally, we found small subpopulations harboring distinct *CSF3R* delta mutations in the bone marrow at the early SCN stage. Of interest is that all these clones disappeared during the disease course, except the *CSF3R*-d715 clone which evolved towards AML. The different *CSF3R* delta mutations cause expression of distinct G-CSF truncated receptors that all have similar consequences for signaling; they result in a sustained activation of signal transducer and activator of transcription 5 (STAT5). STAT5 is a transcription factor, implicated in abnormal signaling responses of leukemic cells with mutated forms of the FLT3 receptor (FLT3-ITD) in AML and the BCR-ABL fusion protein in CML. Furthermore, why one of these *CSF3R* delta mutant clones survived in vivo and progressed towards a fully transformed AML clone while the other *CSF3R* delta variants extinguished during disease development currently remains unexplained. However, it is conceivable that the additional mutations in *LLGL2* and *ZC3H18*, exclusively present in the *CSF3R*-d715 clone, conferred a
competitive growth advantage of this particular subclone representative of essential early steps in leukemic progression that cooperate with the aberrant signaling from the truncated G-CSF receptor.

Besides early genetic events, we found 9 mutations that occurred later in the process of leukemic transformation. Of particular interest is the novel CSF3R mutation (T595I), which appeared exclusively in the AML stage and imposed growth factor independence on an already functionally defective G-CSF receptor. A different mutation in the CSF3R transmembrane domain, CSF3R-T617N, with a similar downstream effect was previously found as a constitutive mutation in a family with hereditary chronic neutrophilia and as an acquired mutation in 2 AML patients. This mutation is suggested to cause ligand independent homodimerization and induces growth factor independent proliferation and differentiation. The major difference between the T617N and the T595I mutation in our patient is that the latter one is located on the already affected CSF3R-d715 allele, which has been shown to cause increased proliferation and impaired differentiation in cell line and animal models, which could explain the increase in colony size between the T595I mutant and the T595I/d715 mutant. The acquisition of autonomous growth abilities by myeloid progenitor cells that already express a hyper-responsive G-CSF receptor mutant strongly suggests that perturbed G-CSF signaling was of vital importance for malignant transformation in this case of SCN. To our knowledge, this is the first example of a gain of 2 different mutations in the same receptor in the process of malignant transformation. An important but still open question is whether the administration of G-CSF to this patient had contributed to the acquisition of this additional mutation. Possibly, the continuous proliferative pressure imposed by G-CSF on clones carrying mutations in CSF3R-d715 and LLGL2 and later also in ZC3H18 may have provided the context for the selection of a clone harboring this self-activating CSF3R mutation, pushing it to become an autonomously proliferating and dominant leukemic clone.

Abnormalities appearing in the AML phase included mutations in ASXL1, SUZ12, and EP300, genes encoding proteins involved in chromatin modification. Mutations in ASXL1 have been reported previously in AML and are associated with an unfavorable prognosis. SUZ12 is a member of the PRC2 complex that also contains EZH2, the histone methyl transferase responsible for the di- and tri-methylation of lysine 27 in the tail of histone 3 (H3K27), imposing a chromatin mark that represses gene expression. Mutations affecting EZH2 and less frequently SUZ12 have been detected in MDS/MPN patients. In contrast, mutations in EP300 and the highly related CREBBP, encoding
histone acetyl transferases that act as transcriptional co-activators, have not yet been reported in myeloid malignancies but are the most frequent structural abnormalities in follicular lymphoma and diffuse large B cell lymphoma\textsuperscript{23, 24}. Mutations in \textit{CCDC155}, encoding coiled-coil domain containing protein 155 with unknown function, \textit{FBXO18}, encoding a DNA helicase involved in DNA repair and genomic integrity, \textit{LAMB1}, encoding an extracellular matrix protein, and \textit{MGA}, encoding a Max gene associated antagonist of Myc oncoproteins, all represent novel mutations with currently unknown functional significance.

Recurrence is an important criterion to discriminate driver from passenger mutations in the process of malignant transformation. Interestingly, mutations in \textit{CCDC155}, \textit{LLGL2}, \textit{MGA} and \textit{ZC3H18} were recently also reported by the TCGA consortium in a panel of AML patients (n=199), albeit at low frequencies. Because frequencies of specific mutations have been shown to vary with the natural history of AML, e.g. \textit{de novo} versus secondary to MDS/MPN or different bone marrow failure syndromes\textsuperscript{14, 42}, it will be of interest to establish how often the newly identified genes are affected in distinct subtypes of secondary AML. Specifically, it will be important to determine whether \textit{LLGL2}, \textit{ZC3H18} or functionally related genes are more generally affected in bone marrow failure syndromes prone to progress to AML and to establish how these mutations contribute to malignant transformation in conjunction with cooperative gene defects.

\textbf{Acknowledgements}

This research was supported by the Center of Translational Molecular Medicine (CTMM), The Dutch Cancer Society “KWF kankerbestrijding” and the E-RARE project ELA2-CN.
Author contributions

R. Beekman: Designed research, performed research, collected data, analyzed and interpreted data and wrote the manuscript.
M.G. Valkhof: Performed research and collected data.
M.A. Sanders: Contributed analytical tools, analyzed and interpreted data.
P.M.H. van Strien: Performed research and collected data.
J.R. Haanstra: Performed research.
C.C.A.M. Broeders: Sample collection.
W.M. Geertsma-Kleinekoort: Sample collection.
A.J.P. Veerman: Contributed samples and patient information.
P.J.M. Valk: Sample collection, analyzed and interpreted data.
R.G. Verhaak: Analyzed and provided TCGA consortium sequencing data.
B. Löwenberg: Analyzed and interpreted data and wrote the manuscript.
I.P. Touw: Supervised the project, sample collection, designed research, analyzed and interpreted data and wrote the manuscript.

Disclosure of conflicts of interest

The authors declare to have no conflicts of interest.
References


**Table 1. Somatic non-synonymous mutations in SCN/AML.** All 12 somatic non-synonymous mutations identified in the AML phase are listed. For each mutation, Ensembl and Refseq reference transcripts, the position of the mutation on genomic DNA, cDNA and protein level, the mutation type and the effect on the protein are indicated. The last column denotes whether the mutated transcript is expressed in the leukemic blasts. See also Figure S1-S3. $\textit{Amino acid numbers based on earlier publications}^4, 6$

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Ensembl Reference Transcript</th>
<th>RefSeq Reference Transcript</th>
<th>Genomic DNA Change (NCBI36/hg18)</th>
<th>cDNA Change</th>
<th>Mutation Type</th>
<th>Amino Acid Change</th>
<th>Protein Change</th>
<th>Expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASXL1</td>
<td>ENST00000375 687</td>
<td>NM_015338.5</td>
<td>g.chr20:30465948dupA</td>
<td>c.1772dupA</td>
<td>Indel</td>
<td>Frameshift and premature stop</td>
<td>p.Y591*</td>
<td>yes</td>
</tr>
<tr>
<td>CCDC155</td>
<td>ENST00000447 857</td>
<td>NM_144688.4</td>
<td>g.chr19:54601976C&gt;T</td>
<td>c.820C&gt;T</td>
<td>Missense</td>
<td>Arg&gt;Trp</td>
<td>p.R274W</td>
<td>yes</td>
</tr>
<tr>
<td>CSF3R-T</td>
<td>ENST00000373 106</td>
<td>NM_000760.3</td>
<td>g.chr1:3670621G&gt;A</td>
<td>c.1853C&gt;T</td>
<td>Missense</td>
<td>Thr&gt;ile</td>
<td>p.T595I</td>
<td>yes</td>
</tr>
<tr>
<td>CSF3R-d</td>
<td>ENST00000373 106</td>
<td>NM_000760.3</td>
<td>g.chr1:36704841G&gt;A</td>
<td>c.2215C&gt;T</td>
<td>Nonsense</td>
<td>Gln&gt;*</td>
<td>p.Q716*</td>
<td>yes</td>
</tr>
<tr>
<td>EP300</td>
<td>ENST00000263 253</td>
<td>NM_001429.3</td>
<td>g.chr22:39902447_3902453delTGAGAGAC</td>
<td>c.5030_503 6delTGAGAGAC</td>
<td>Indel</td>
<td>Frameshift and premature stop</td>
<td>p.V1677Dfs30</td>
<td>yes</td>
</tr>
<tr>
<td>FBXO18</td>
<td>ENST00000379 999</td>
<td>NM_032807.3</td>
<td>g.chr10:6003435C&gt;G</td>
<td>c.2372C&gt;G</td>
<td>Missense</td>
<td>Ala&gt;Gly</td>
<td>p.A791G</td>
<td>yes</td>
</tr>
<tr>
<td>LAMB1</td>
<td>ENST00000222 399</td>
<td>NM_002291.2</td>
<td>g.chr7:107387385delG</td>
<td>c.2445delC</td>
<td>Indel</td>
<td>Frameshift and premature stop</td>
<td>p.P195Pfs5</td>
<td>no</td>
</tr>
<tr>
<td>LLGL2</td>
<td>ENST00000167 462</td>
<td>NM_004524.2</td>
<td>g.chr17:71070826G&gt;C</td>
<td>c.665G&gt;C</td>
<td>Missense</td>
<td>Arg&gt;Pro</td>
<td>p.R222P</td>
<td>yes</td>
</tr>
<tr>
<td>MGA</td>
<td>ENST00000219 905</td>
<td>NM_001164273.1</td>
<td>g.chr15:39787311C&gt;T</td>
<td>c.2282G&gt;T</td>
<td>Missense</td>
<td>Pro&gt;Leu</td>
<td>p.P761L</td>
<td>yes</td>
</tr>
<tr>
<td>RUNX1</td>
<td>ENST00000300 305</td>
<td>NM_001754.4</td>
<td>g.chr21:35153662C&gt;T</td>
<td>c.592G&gt;A</td>
<td>Missense</td>
<td>Asp&gt;Asn</td>
<td>p.D198N</td>
<td>yes</td>
</tr>
<tr>
<td>SUZ12</td>
<td>ENST00000322 652</td>
<td>NM_015355.2</td>
<td>g.chr17:27344689_2734691dupATT</td>
<td>c.1789_179 1dupATT</td>
<td>Indel</td>
<td>Insertion Ile</td>
<td>p.597dup</td>
<td>yes</td>
</tr>
<tr>
<td>ZC3H18</td>
<td>ENST00000301 011</td>
<td>NM_144604.3</td>
<td>g.chr16:87192175delC</td>
<td>c.777delC</td>
<td>Indel</td>
<td>Frameshift and premature stop</td>
<td>p.P259Pfs15</td>
<td>yes</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Chronological overview of the clinical course of the SCN/AML patient. Distinct events in the disease course are indicated above the timeline, i.e., the diagnosis of SCN, the initiation of G-CSF therapy, the discovery of the CSF3R-d715 mutation and the diagnosis of AML.

Figure 2. Acquisition of mutations in the evolution of SCN towards AML. (A) The 12 somatic non-synonymous mutations identified in the leukemic blasts were analysed in the SCN phase, using amplicon-based deep sequencing. Per mutation, the percentage of mutated amplicons is shown. Based on their frequencies in the AML population, all mutations are considered to be heterozygous, implying that the number of cells carrying the mutations is estimated to be twice the number of mutated amplicons. (B) Single myeloid colonies grown from the bone marrow sample obtained 15 years before leukemia development were analysed for the presence of mutations in CSF3R, LLGL2 and ZC3H18. See also Table S5. (C) The presence of different CSF3R mutations in the bone marrow obtained 15 and 9 years before leukemia development as well as in the leukemic phase was investigated by amplicon-based deep sequencing. Per mutation, the percentage of mutated amplicons is shown. T595i: CSF3R mutation T595i, d715-d730: CSF3R delta mutations at amino acid position 715 to 730.

Figure 3. Functional analysis of CSF3R mutants in myeloid progenitor cell assays. In vitro colony growth of Csf3r deficient murine hematopoietic progenitor cells expressing different CSF3R mutants. (A) Graphical representation of the different CSF3R constructs. Wild type (wt), T595i (containing the extracellular mutation at amino acid position 595), d715 (containing the intracellular mutation, Q716X, causing the introduction of a stop codon at amino acid position 716) and T595i/d715, containing both mutations as found in the SCN/AML patient. Ig: Immunoglobulin like domain; CRH: cytokine receptor homology domain; FNIII: fibronectin type III repeats; TM: transmembrane domain; cyto: cytoplasmic domain. Nomenclature has been adopted from Layton et al. (B) Colonies were grown in the presence of puromycin, either without growth factor (no GF) or with 100ng/ml human G-CSF. The induced colony growth is dependent on the transduction efficiency and the type of CSF3R construct. The transduction efficiency can be deduced from the number of GM-CSF-induced colonies under puromycin selection as the CSF3R constructs confer puromycin resistance, but do not affect GM-CSF-induced colony growth. Hence, by dividing the number of colonies by the number of GM-CSF induced colonies the transduction efficiency was corrected for.

Figure 4. Schematic representation of the clonal evolution of SCN towards AML. The sequential genetic events, starting with the presence of a germ line mutation in the gene encoding neutrophil elastase (ELANE) are incidated. A sequential gain of CSF3R delta mutations and an LLGL2 mutation is observed in the early SCN phase. Only the clone harboring the CSF3R-d715 and the LLGL2 mutation gained an additional mutation in ZC3H18, followed by its expansion in the intermediate SCN phase. Gain of 9 additional mutations and trisomy 21 in the mutated population preceded complete transformation towards AML. CSF3R-d715-d730: CSF3R delta mutations at amino acid position 715 to 730.
Figure 1

![Diagram showing the timeline of diagnosis, G-CSF treatment, CSF3R-d715 mutation, and AML phases, along with the transition from SCN phase to AML phase.]

Figure 2

![Graph showing the percentage of mutated amplifications for various genes, including ASML1, CCDC15, CSF3R-d715, CSF3R, CSF3R-wt, EP300, FBXO18, LAMB1, LLGL2, MGA, RUNX1, SUZ12, and ZC3H18, and the distribution of myeloid colonies and CSF3R mutations.]

- Early SCN (15 years before AML)
- Intermediate SCN (9 years before AML)
- AML
Figure 3

A

B

No GF

G-CSF (100 ng/ml)
Figure 4

Birth

Early SCN

Germ line mutation

ELANE

Gain of first mutations

CSF3R-d715

LLGL2

CSF3R-d717

CSF3R-d725

CSF3R-d730

Gain of additional mutation

ZC3H18

Intermediate SCN

Expansion of mutated clone

Gain of late genetic alterations

ASXL1

CCDC155

CSF3R-T595I

EP300

FBXO18

LAMB1

MGA

RUNX1

SUZ12

Trisomy 21

AML
Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia