Mutagenesis, 2020, 35, 381–389
doi:10.1093/mutage/geaa027
Original Manuscript

OXFORD

Original Manuscript

Effect of the unfolded protein response and oxidative stress on mutagenesis in *CSF3R*: a model for evolution of severe congenital neutropenia to myelodysplastic syndrome/acute myeloid leukemia

Adya Sapra[†], Roman Jaksik[1,†], Hrishikesh Mehta, Sara Biesiadny[2], Marek Kimmel[1,2,‡] and Seth J. Corey[‡,*]

Department of Pediatrics, Cancer Biology, and Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, OH, USA, [1]Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland, [2]Department of Statistics, Rice University, Houston, TX, USA

*To whom correspondence should be addressed. Tel: 216-444-2506; Fax: 216-636-3710; Email: coreylab@mac.com
[†]Co-first authors.
[‡]Co-senior authors.

Received 17 December 2019; Editorial decision 11 September 2020; Accepted 21 September 2020.

Abstract

Severe congenital neutropenia (SCN) is a rare blood disorder characterised by abnormally low levels of circulating neutrophils. The most common recurrent mutations that cause SCN involve neutrophil elastase (*ELANE*). The treatment of choice for SCN is the administration of granulocyte-colony stimulating factor (G-CSF), which increases the neutrophil number and improves the survival and quality of life. Long-term survival is however linked to the development of myelodysplastic syndrome/acute myeloid leukemia (MDS/AML). About 70% of MDS/AML patients acquire nonsense mutations affecting the cytoplasmic domain of *CSF3R* (the G-CSF receptor). About 70% of SCN patients with AML harbour additional mutations in *RUNX*1. We hypothesised that this coding region of *CSF3R* constitutes a hotspot vulnerable to mutations resulting from excessive oxidative stress or endoplasmic reticulum (ER) stress. We used the murine Ba/F3 cell line to measure the effect of induced oxidative or ER stress on the mutation rate in our hypothesised hotspot of the exogenous human *CSF3R*, the corresponding region in the endogenous *Csf3r*, and *Runx1*. Ba/F3 cells transduced with the cDNA for partial C-terminal of *CSF3R* fused in-frame with a green fluorescent protein (GFP) tag were subjected to stress-inducing treatment for 30 days (~51 doubling times). The amplicon-based targeted deep sequencing data for days 15 and 30 samples show that although there was increased mutagenesis observed in all the three genes of interest (partial *CSF3R*, *Csf3r* and *Runx1*), there were more mutations in the GFP region compared with the partial *CSF3R* region. Our findings also indicate that there is no correlation between the stress-inducing chemical treatments and mutagenesis in Ba/F3 cells. Our data suggest that oxidative or ER stress induction does not promote genomic instability, affecting partial C-terminal of the transduced *CSF3R*, the endogenous *Csf3R* and the endogenous *Runx1* in Ba/F3 cells that could account for these targets to being mutational hotspots. We conclude that other mechanisms to acquire mutations of *CSF3R* that help drive the evolution of SCN to MDS/AML.

© The Author(s) 2021. Published by Oxford University Press on behalf of the UK Environmental Mutagen Society.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Background

Severe congenital neutropenia (SCN) is a genetically heterogeneous condition characterised by pathologically low levels of neutrophils. Neutrophils provide the first line of defence by killing antigens and reconstituting tissue integrity (1). Clinically significant neutropenia, which can render an individual immunocompromised, is diagnosed when the peripheral absolute neutrophil count (ANC) is <1500/µl. SCN is defined as ANC of <500/µl (2). Clinical diagnosis of SCN includes examination of bone marrow for paucity of early myeloid precursors. SCN patients show maturation arrest in neutrophil development in the bone marrow, which leads to accumulation of promyelocytes. Germline mutations in multiple genes have been found in individuals with SCN. These include the neutrophil elastase *ELANE*, mitochondrial-associated protein *HAX1*, transcription factor *GFI1*, metabolic enzyme G6PC, granulocyte-colony stimulator factor receptor *CSF3R* and Wiskott–Aldrich syndrome protein *WAS* (3). Transmitted in an autosomal dominant pattern or spontaneously acquired, point mutations in *ELANE* have been detected in ~40% of SCN patients. Although the mechanism of induction of apoptosis by *ELANE* mutations is still not fully understood, leading hypotheses state that germline *ELANE* mutations result in misfolding and/or mislocalisation of neutrophil elastase, which triggers the unfolded protein response (UPR). This highly conserved biological process consists of a collection of intracellular signal transduction pathways which are activated in response to endoplasmic reticulum (ER) stress (4–6). UPR is initiated via 3 pathways localised in the ER: IRE1alpha (inositol-requiring 1alpha), PERK (double-stranded RNA-dependent protein kinase [PKR]–like ER kinase) and activating transcription factor 6 (ATF6). In case all of these three UPR pathways fail, and the UPR mechanism is not successful in restoring homeostasis, it will trigger the cell towards apoptosis, which can lead to reduction in the ANC (7–9). Thus, UPR, activated due to ER stress, causes apoptosis of differentiating neutrophil precursors at the promyelocyte stage. This leads to accumulation of promyelocytic progenitors and subsequent deficiency of mature neutrophils.

The standard of treatment for SCN is the administration of recombinant human granulocyte-colony stimulating factor (G-CSF), which binds to and activates a specific receptor (*CSF3R*). CSF3R transduces signals critical for the proliferation and maturation of granulocytic progenitor cells. G-CSF treatment increases the neutrophil count, helps resolve pre-existing infections, diminishes the number of new infections, and significantly improves the survival and quality of life (10). Long-term survival with G-CSF is associated with the development of myelodysplastic syndrome/acute myeloid leukemia (MDS/AML). Approximately 30% of SCN patients undergoing G-CSF treatment develop MDS or AML (9). Of note, ~70% of SCN patients with MDS/AML acquire nonsense mutations in the region of *CSF3R* that encodes the cytoplasmic domain (11–14). These somatic *CSF3R* mutations are characterised by a truncation variant found in the cytoplasmic domain of the CSF3R and are associated with a hyperproliferative phenotype, which might contribute to the leukemogenic transformation (13,15). Cell clones harbouring *CSF3R* mutations have a growth advantage and acquire additional cooperating mutations in other genes like *Runx1* that contribute to transformation to AML (13,16). Clinical studies have also been done to determine the incidence of frequently mutated sites in CSF3R associated with the development of MDS/AML in SCN patients. Nonsense mutations in 17 different nucleotide positions in the *CSF3R* intracellular domain, leading to truncation of one to four tyrosine residues, have been identified (13).

In this study, we hypothesised that the region corresponding to the cytoplasmic domain of CSF3R constitutes a mutational hotspot. This vulnerability to mutations may result from excessive oxidative or ER stress. Previous studies have shown that enhanced reactive oxidation species (ROS) production can be one of the mechanisms that cause the evolution of MDS/AML in SCN patients with a truncated G-CSF receptor (17). In SCN, myeloid precursors may be under stress due to cellular responses like UPR associated with mutant *ELANE*. Under stress conditions, ROS production increases, and G-CSF treatment can lead to even greater increase in ROS production. Thus, the combination of stress pathways and the G-CSFR-mediated signalling transduction pathways may create an environment of enhanced oxidative stress that promotes genomic instability. In this study, we aim to measure the mutation rates of partial C-terminal of *CSF3R* under oxidative and ER stress conditions and compare them with the endogenous genes (*Csf3r* and *Runx1*) and the tagged mNeonGreen region. The GC content percentage of the CSF3R region is significantly higher (64%) than the mNeonGreen region (54%). Previous studies have shown that CpG dinucleotides mutate at a high rate because cytosine is vulnerable to undergo deamination and generate thymine (18). We expect *CSF3R* to have higher variant allele frequency (VAF) when compared with the mNeonGreen region.

Methods

Cell culture

Murine pro-B cell line Ba/f3 cells (ATCC) were maintained and grown in 1× RPMI-1640 medium (CellGro) supplemented by 10% fetal bovine serum, 1% PenStrep, 1× Glutamax and 2 ng/ml mIL3. Ba/f3 cells were transduced with lentivirus to introduce the partial C-terminal of *CSF3R* (c.2628–3092, NM_000760.3) fused with mNeonGreen (GenBank EU482389.1) (25). Positively transduced cells were sorted for green fluorescence by fluorescence-activated cell sorting using a Becton Dickinson FACSAria II.

Flow cytometry analysis measurement of ROS production

Untransduced Ba/F3 cells were treated with the increasing concentrations of thapsigargin (0, 1, 2, 5 and 10 nM) for 48 h. The cells were then treated with CellRox Deep Red Reagent (ThermoFisher) at a final concentration of 5 µM and incubated for 30 min at 37°C. The cells were washed with Hanks balanced salt solution (HBSS) and incubated with MitoSox Red Reagent (ThermoFisher) for 10 min. The cells were washed twice with HBSS and finally suspended in 1 ml HBSS before carrying out flow cytometry on a Becton Dickinson LSRII.

qRT–PCR analysis to measure UPR marker levels

qPCR was performed to quantitate the levels of different UPR markers (ATF4, ATF6, BiP, CHOP, XBP1spliced and XBP1unspliced) in untransduced Ba/F3 cells treated with increasing doses of UPR-inducing chemical, thapsigargin (1, 2 and 5 nM) for 0, 24 and 48 h. At each desired time point of the treatment, cell pellets were collected, and RNA was extracted using the Qiagen RNeasy Plus Mini Kit. RNA concentrations were measured using NanoDrop One (Thermo Scientific). cDNA was synthesised from the extracted RNA using the iScriptTM cDNA Synthesis kit (Bio-Rad). Quantitative PCR was performed on the StepOnePlus Real Time PCR System (Applied Biosystems, ThermoFisher Scientific). Power SYBR Green (Applied Biosystems, ThermoFisher) was used

as a DNA intercalator dye to monitor amplified DNA quantification, and real-time quantitative PCR curves were analysed by StepOne Software v2.3 (Applied Biosystems) to obtain threshold cycle (Ct) values for each sample. Comparative Ct method was used to calculate the mRNA expression levels for each marker. The housekeeping gene, Actin expression Ct values were comparable among all experiments. The ddCt values were calculated for all samples with respect to housekeeping gene expression values and then day zero Ct values.

Stress-inducing chemical treatment

Two stress-inducing chemicals, thapsigargin (26) and tert-butylhydroquinone (TBHQ) (27,28), were selected to induce ROS and UPR, respectively. Treatment was carried out on Ba/F3 transduced cells. Cell pellets were stored at −80°C on day 0 to serve as control. One thousand cells/well were plated in a 96-well plate using serial dilution. After formulating a dose–response curve, lethal dose (LD) 10 and LD20 were selected as optimum doses of stress-inducing treatment for both the drugs. Transduced Ba/f3 cells were treated continuously for 30 days with the following stress-inducing treatments: 1 and 2 nM thapsigargin, 1 and 2nM TBHQ, and control (untreated) samples. The samples from days 0, 15 and 30 were then sent for next-generation sequencing (NGS) analysis.

NGS: library preparation

The samples were submitted to the Genomics Core (Nucleic Acid Research Facilities) at Virginia Commonwealth University for amplicon deep sequencing using NGS technology. Targeted deep sequencing was performed using the tailed amplicon sequencing approach (a.k.a. two-step PCR approach). Genomic DNA was extracted from frozen cell pellets using the Qiagen DNeasy Blood and Tissue DNA Extraction kit. The PCR approach was carried out in two steps: the first round of PCR uses primers containing a locus-specific sequence and a common 5′ tail and the second round of PCR is performed to introduce the indexes (barcodes) as well as the Illumina-specific sequencing adapters. The first step was performed to amplify the region of interest for each of the samples. Our genes of interest were as follows: the exogenous *CSF3R* (transduced construct); endogenous *Csf3r* (Exon 17); and endogenous *Runx1*. Since the cDNA of all these genes was too large to amplify the whole region together, the amplification was carried out in segments for each

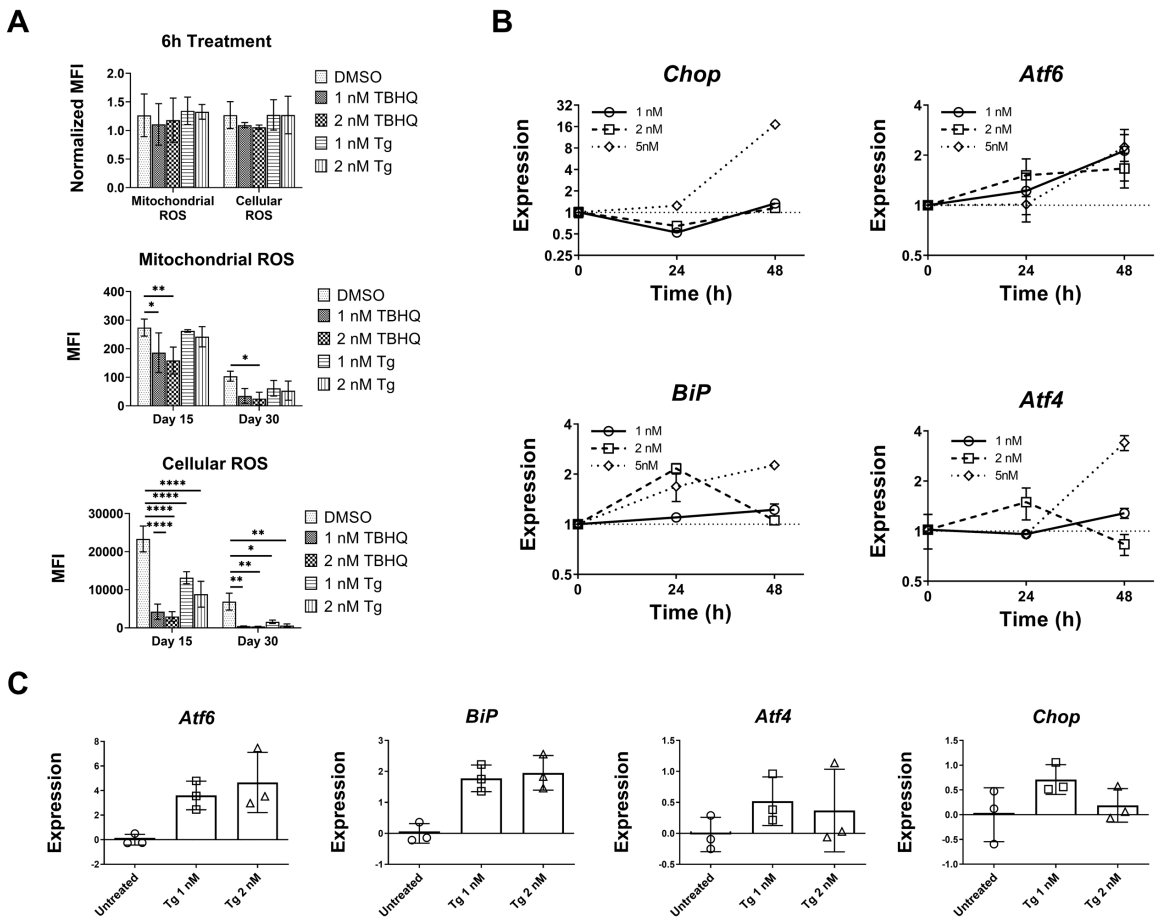

Figure 1. Measurement of oxidative and ER stress produced by thapsigargin and TBHQ treatment. (**A**) Mitochondrial and cytosolic fluorogenic probes (Mitosox and CellRox, respectively) were used to measure the cellular oxidative stress using flow cytometry in cells treated with either DMSO, 1 and 2 nM TBHQ, or 1 and 2 nM Thapsigargin for 6 h, 15 days and 30 days. Statistical significances were computed using two-way ANOVA followed by Dunnett's multiple comparison test comparing all treatments with DMSO (*$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$). (**B**) UPR marker expression was measured by real-time PCR. Data are normalised to the expression of actin to get the dCt values. The dCt values were further normalised to the dCt value of day 0 control. Most of the UPR markers were observed to have elevated levels after 48 h of treatment. (**C**) ER stress was also measured after treatment with thapsigargin for 28 days. Thapsigargin treatment led to an increase in UPR stress levels.

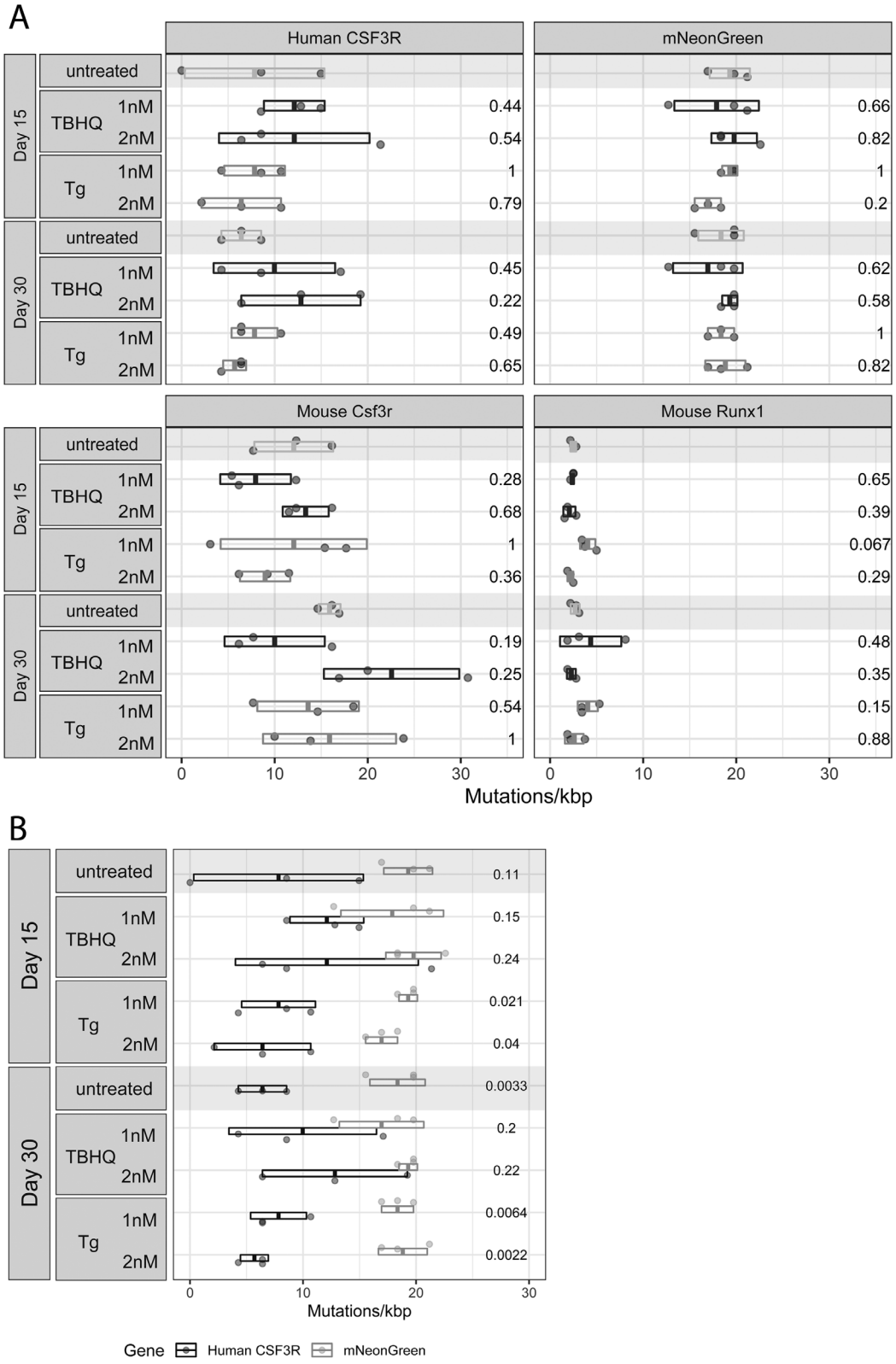

Figure 2. Accumulation of new mutations during the treatment period. (**A**) Although new mutations did arise during the treatment duration, however, there was no correlation between the treatment and the total number of new mutations in both the endogenous and the exogenous systems. The plots show the total number of mutations detected for each sample. Each point corresponds to one replicate; the bars represent mean ± SD. The values on the right *Y* axis are the *P*-values from a *t*-test between the mean values of each treated sample (either thapsigargin or TBHQ) compared with the untreated samples from the same day. Each colour marks different treatment types. (**B**) The plot shows the comparison between the average number of mutations per kbp between CSF3R and mNeonGreen regions using *t*-test. Boxplots represent mean ± SD. The figure shows that there were more mutations in the mNeonGreen region whem compared with *CSF3R* consistently for all treatments and both days.

gene. The primers were designed for overlapping sections in each region of interest. PCR was performed using Phusion High-Fidelity PCR Master Mix with HF Buffer from New England BioLabs (2×). PCR thermocycler profile used was 98°C 30 s, 34 cycles of 98°C for 10 s, optimum annealing temperature for 30 s and 72°C for 30 s, followed by a final extension at 72°C for 10 min. The primer sequencing and the annealing temperatures for each segment are provided in Supplementary Table 1. Samples were purified using Agencourt® AMPure XP beads using DNA:beads ratio of 1.2 to remove primer and unincorporated deoxyribonucleotide triphosphate (dNTPs). Once the amplification was carried out and the sequence size was confirmed using PCR, the segments for each gene were pooled together and the second round of PCR was performed. The second PCR reactions were performed using 1 µl of a 1/10 diluted first-round PCR product as template and Applied Biosystems 9700 PCR System with 1 U Thermo Scientific Phusion Hot Start II High-Fidelity PCR Master Mix, and 0.2 µM of each primer. PCR thermocycler profile used was 95°C for 2 min, 10 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by a final extension at 72°C for 5 min. Samples were submitted to a second PCR clean-up using Agencourt® AMPure XP beads using DNA:beads ratio of 1.12 to remove primer and unincorporated dNTPs. Libraries were quantified using fluorometric quantification and pooled equimolarly. Libraries were sequenced using Illumina MiSeq v3 600 cycle kit.

NGS analysis

Broad Institute (GATK v4) guidelines were followed for the analysis of the targeted deep sequencing data. The Illumina Nextera adapter sequences introduced during PCR amplification were first removed using Trim Galore, which is based on Cutadapt too. The filtered sequences were then aligned against the respective reference genomes, using BWA-MEM algorithm. The alignment reads that were bound by primers were then trimmed using the BAMClipper tool. Finally, the variants were discovered using Mutect2 tool. Day 0 samples were used as reference, to eliminate the mutations that were already present before the treatment started.

Results

Oxidative and ER stress produced by thapsigargin and TBHQ treatment

Cellular oxidative stress was induced in Ba/F3 cells using increasing doses of thapsigargin and TBHQ (1 and 2 nM). Thapsigargin is a known mediator of ER stress, whereas TBHQ has been identified as a mediator of mitochondrial ROS and can promote DNA damage (19). Flow cytometry was then used to measure the amount of ROS that were induced. Flow cytometry analysis at 6 h of treatment shows that ROS across treatments is statistically not different (Figure 1A). However, days 15 and days 30 measurement of ROS demonstrates a statistically significant reduction in ROS for cells treated with thapsigargin or TBHQ in comparison to DMSO. These data suggest reduced ROS levels present with TBHQ and thapsigargin treatment. ER stress was also measured following treatment of cells with thapsigargin (1, 2 and 5 nM doses) at different time points (0, 24 and 48 h). As shown in Figure 1B, qPCR demonstrated increased expression in the UPR markers: *Atf4*, *Atf6*, *HSPA5 (BiP)* and GADD153 *(Chop)*. We next determined if the effect of thapsigargin treatment on the UPR stress levels persisted for prolonged period of time, i.e.

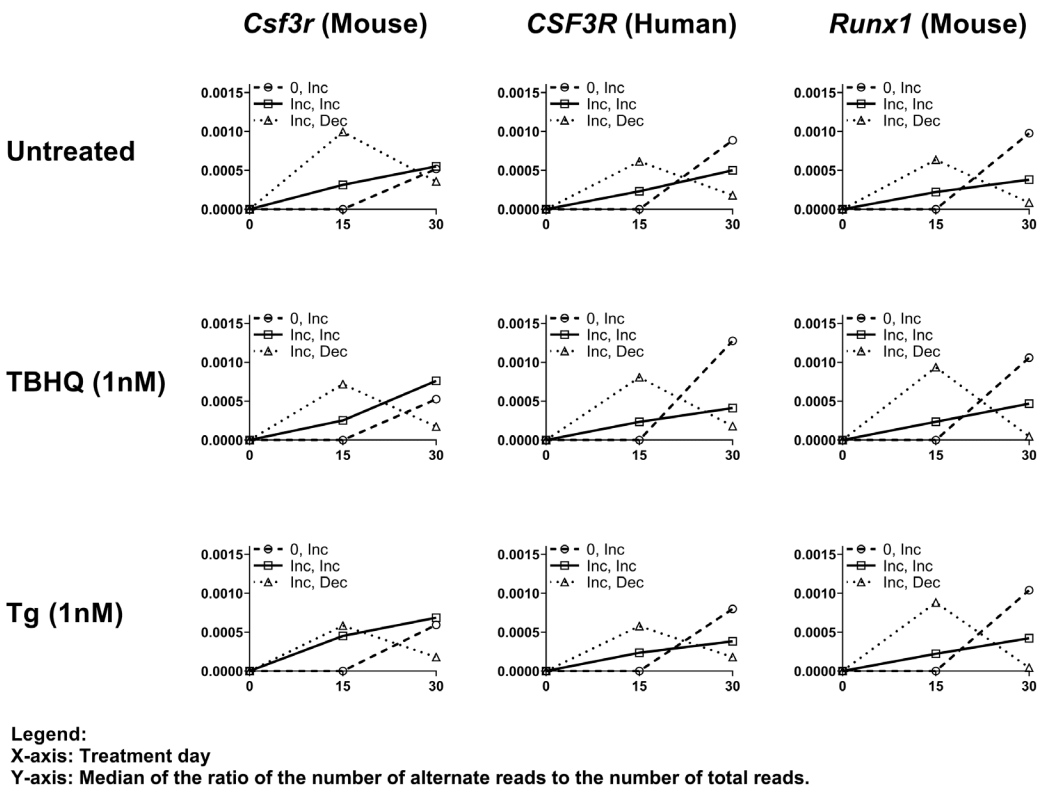

Figure 3. Three major trends were observed in the mutations that were consistent in all samples. The first set (blue line) consisted of mutations that were absent until day 15 and arose only on day 30. The second set (red line) consisted of mutations whose frequency consistently grew during the 30-day period. The third set (grey line) consisted of mutations whose frequency increased on day 15 followed by a decrease on day 3. Tg, thapsigargin.

Table 1. Types of mutations present

Mutations	*CSF3R*	GFP	*Csf3r*	*Runx1*	
A→G	6.06%	7.91%	4.60%	20.07%	Transition
G→A	2.02%	8.68%	10.91%	16.79%	
T→C	Absent	Absent	6.01%	2.50%	
C→T	Absent	Absent	9.35%	16.79%	
A→T	Absent	4.08%	1.34%	Absent	Transversion
T→A	Absent	Absent	Absent	0.36%	
A→C	1.01%	Absent	2%	Absent	
C→A	80.81%	58.16%	40.09%	0.04%	
T→G	4.04%	15.81%	19.82%	1.46%	
G→T	5.05%	4.59%	4.90%	28.1	
G→C	1.01%	0.77%	0.45%	10.20%	
C→G	Absent	Absent	0.45%	Absent	

Distribution of transition and transversion mutations that were observed in all three genes.

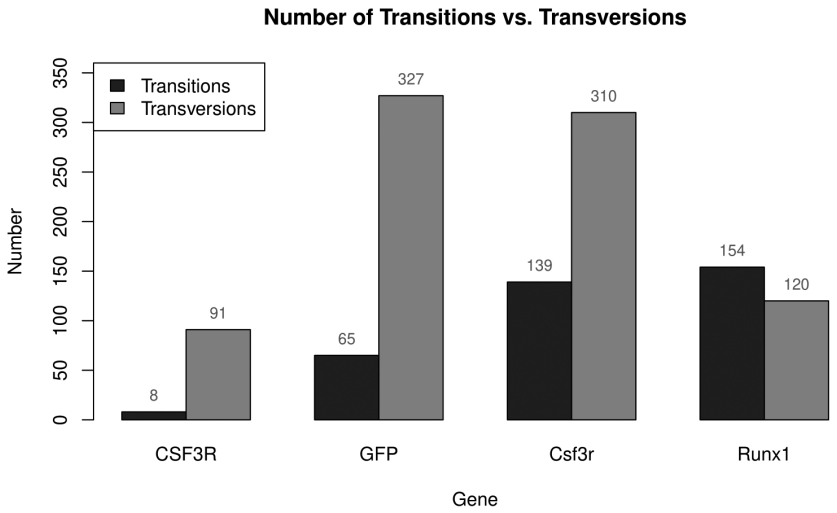

Figure 4. Comparison between total number of transition and transversion mutations. The total number of transition mutations was much higher than the total number of transversion mutations in all three genes and the GFP region.

28 days. As seen in Figure 1C, there was an increase in the following UPR markers *Atf4*, *Atf6*, *BiP* and *Chop*. The cells were treated with 1 and 2 nM doses of thapsigargin, and the UPR marker levels were compared with the untreated cells.

Acquired mutations in both treated and untreated samples

No correlation was seen between treatment with thapsigargin or TBHQ and the total number of new mutations. The *P*-values from a *t*-test between the mean values of each sample treated with thapsigargin and TBHQ compared with the untreated samples were plotted for each gene (Figure 2A). We observed that although many new mutations were present in both day 15 and 30 samples, the number of mutations was statistically not different between treated and untreated samples for all three regions of interest: (partial transduced *CSF3R*, *Csf3r* and *Runx1*) and the green fluorescent protein (GFP) region. Moreover, GFP region having comparatively lower GC content was more highly mutated when compared with our proposed hotspot, *CSF3R* (having higher GC content percentage). Figure 2B shows a comparison between the mutation rates observed in *CSF3R* and the GFP region.

Three major patterns of mutation acquisition

The mutations observed in all the samples could be categorised into three patterns as represented in Figure 3. The first pattern represents mutations whose frequency increased by 15 days, followed by diminishing frequency by day 30; the second pattern represents mutations for which frequency increased on day 15 and continued to increase on day 30; and the third pattern consisted of mutations that were dormant until day 14 and emerged only at day 30. The first group of mutations compared with second group, also shows a faster increase in frequency. The data are divided into different sets according to the gene of interest. For each gene, there are further subsets for each treatment viz.: untreated, TBHQ 1 nM; TBHQ 2 nM; thapsigargin 1 nM; and thapsigargin 2 nM. For each subset, a histogram of the log (base 10) of the ratio of altered reads over coverage was developed. Once separated according to days, the median of each subgroup was calculated and plotted. These three trends were consistent for all the samples. It was also observed that the first group of mutations (increase, decrease) is presented at a higher frequency when compared with the second group (increase, increase). This indicates that the fast increase in mutations could be affecting the cell growth and viability, thus providing a clonal disadvantage to these cells. On the other hand, the mutations that are acquired at a

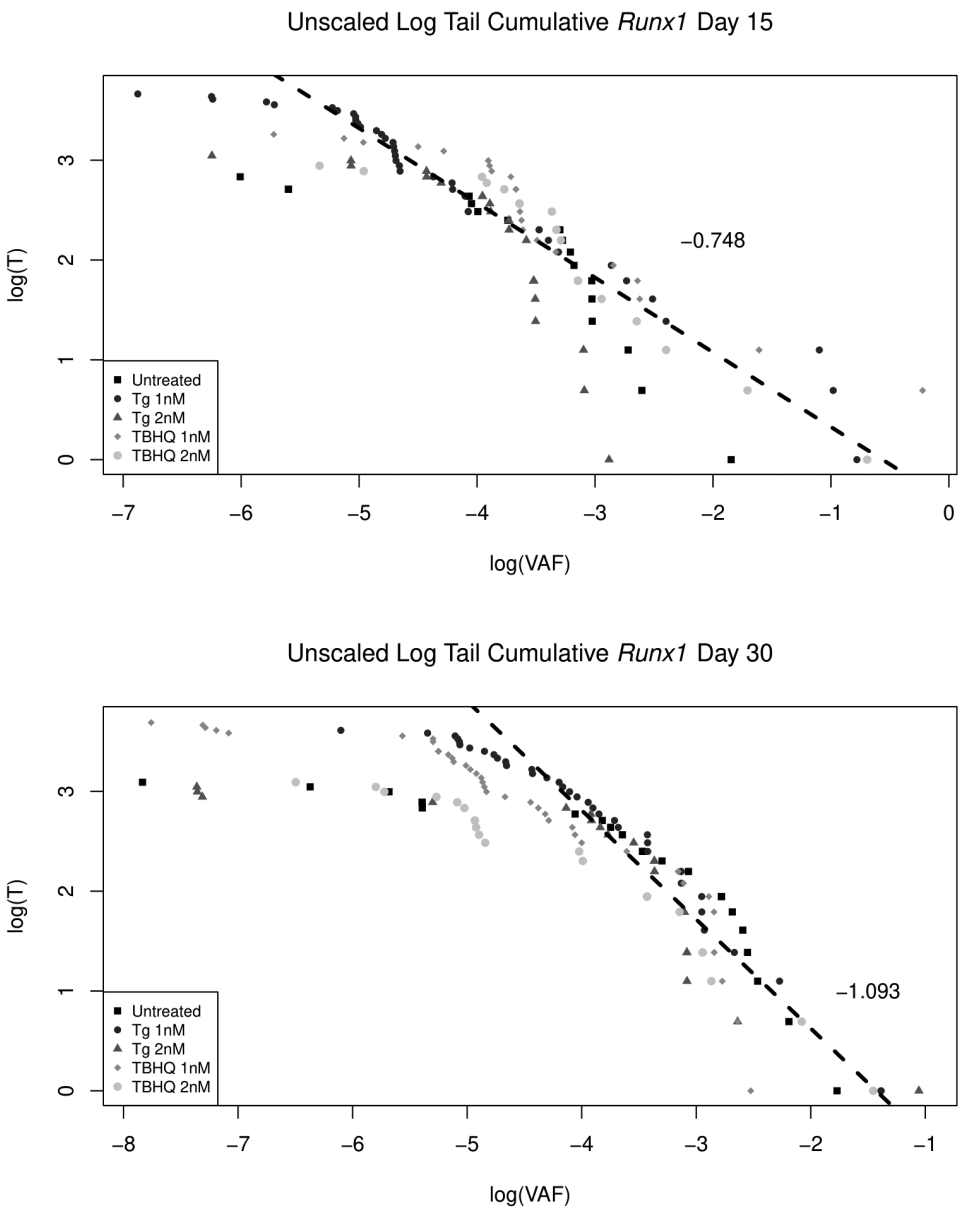

Figure 5. Log–log plots of the SFS cumulative tails for mutations in the *Runx1* gene for $t = 15$ days (top) and $t = 30$ days (bottom), as a function of VAF. As explained in the text, under exponential growth of cell and neutrality of mutations, in log–log coordinates, the SFS tails should have slopes approximated by −1. As depicted, it is the case for the *Runx1* gene, but not true of any other genes we tested (Supplementary Figure S1), which have markedly different slopes, possibly indicating selection effects.

slower rate do not cause cell death, thus giving the cell the time to acquire additional mutations that provide a clonal advantage to them.

Number of transition mutations is consistently higher than transversion mutations

Further statistical analysis was done to study the different types of mutations that were occurring most frequently. Table 1 shows the different types of nucleotide substitutions that were seen in all the three regions of interest. Surprisingly, the mutations replacing cytosine to thymine, and adenine to guanine, which were found to have the highest frequency in clinical studies (leading to substitution of glutamine to stop codon), were absent (or present in very few numbers) in *CSF3R*. Furthermore, the total number of transition mutations was much higher than the transversion mutations for all the three genes (Figure 4).

Signatures of selection in mutational data

Although there do not seem to exist systematic differences in mutation rates with respect to the dose of thapsigargin or TBHQ, or gene (except for those related to gene length), there seems to exist two types of signatures of selection. First, as seen in all panels in Figure 3 (grey lines) mutations at some sites increase in frequency over the initial 15 days, but decrease over day 16–day 30 period. The trend seems identical in all experiments (untreated, both mutagens and all doses), and it concerns similar proportions of sites in each case. It is consistent with mutagen-independent mutation rate, distributed over the genome, with some mutations being deleterious. As a consequence, cells with these mutations might gradually start dividing slower, or dying, after the deleterious effects take over. The second effect is gene specific. It concerns the site-frequency spectra (SFS), which are the frequency graphs of mutations with given VAF.

Neutral mutations in exponentially growing populations are characterised by the cumulative tail distribution of the SFS approximated by 1/VAF (20). Therefore, in log–log coordinates, the SFS tails should have slopes approximated by −1. This is true of *Runx1* (Figure 5), but not true for other genes, which have markedly different values, possibly indicating selection effects (21) (Supplemental Figure S1).

Discussion

Several clinical studies have demonstrated that leukemia transformation is significantly higher in SCN patients. The age-adjusted incidence of MDS/AML in SCN patients being treated with G-CSF is reported as 22% (22). These clinical data strongly suggest that pharmacological administration of G-CSF for the treatment of SCN is correlated with the development of MDS/AML (16). According to a study by Severe Chronic Neutropenia International Registry (SCNIR), 21% of SCN patients undergoing G-CSF treatment developed leukemia at some point during their therapy. The precise contribution of G-CSF is however uncertain because there are reports of MDS/AML arising in patients with SCN prior to the availability of G-CSF (23). The exact mechanisms of clonal evolution causing the transition from SCN to MDS/AML remain unknown. This study was done to estimate the number, timeline and proliferative advantage of mutations in lymphoid progenitors under cellular stress conditions. The bioinformatics analysis of the NGS data shows that there is no significant increase in the number of mutations due to different cellular stress-inducing treatments in Ba/F3 cells. No correlation was seen between treatment with thapsigargin or TBHQ and the total number of new mutations in partially transduced CSF3R, Csf3r and Runx1. Moreover, the GFP region, which contains lower number of CPG islands, was significantly more mutated than the proposed hotspot, *CSF3R*. Although the published data are limited, this conclusion is supported by a recent clinical study by Xia *et al*. that shows the patients with SCN did not display increased age-matched rates of mutagenesis and that both intrinsic and non-intrinsic cellular changes could contribute to the transformation of SCN to MDS/AML (22). This study strongly suggests that in vivo cellular stress induction does not promote the genomic instability in Ba/F3 cells. Since CSF3R mutations are specific only to SCN-associated MDS/AML and not to MDS/AML associated with less-severe neutropenia as seen in Shwachman–Diamond syndrome, there may be other factors involved, including mutant ELANE variants.

As mentioned in Results, there is a notable difference in the slopes of log-log plots of the SFS tails between the *Runx1* (slope equals approximately −1; Figure 5) and the other genes (slopes equal approximately −2; Supplementary Figure S1). This confirms apparent selective neutrality in *Runx1* and hints at a departure from neutrality in the other genes. However, this signature (slope −2) might be consistent with population size reduction, such as the gradual loss of exogenous constructs (21).

The premise of cell division is the accurate duplication of the whole genetic apparatus, and error in this duplication is one of the main causes of the somatic mutations. In our study, we had chosen Ba/F3 as our cell line model system because it has a short doubling time of 12–14 h. Thus, the cells were able to replicate more number of times (~50–60 replications) in the 30-day treatment period, thereby increasing the chances of replication errors and the number of mutations. Additionally, in the case of Ba/F3 cells, other factors that could have potentially caused these new mutations can be further investigated. As shown by Watanabe *et al*. (24), in Ba/F3 cells, mutated receptors are highly susceptible to acquiring additional mutations, which can be of unknown functional significance. One hypothesis is that since these weakly transforming mutations are occurring along with the acquired mutations, which might be providing a clonal advantage to them. Our study lays the groundwork for further understanding the possible mechanisms behind the association of germline ELANE and spontaneous CSF3R mutations that may be responsible for the evolution of MDS/AML from SCN.

Conclusions

Our studies show that *in vivo* oxidative or ER stress induction does not promote genomic instability affecting both transduced partial *CSF3R* and the endogenous systems (*Csf3R* and *Runx1*) that might account for it being a mutational hotspot. Since *CSF3R* mutations are specific to SCN-associated MDS/AML, other factors including stochastic processes are likely to be involved. This study provides a model to study the mechanisms behind the association of *ELANE* and *CSF3R* mutations responsible for the evolution of MDS/AML from SCN.

Supplementary data

Supplementary data are available at *Mutagenesis* Online.

Supplemental Figure S1. Log-log plots of the site frequency spectra (SFS) cumulative tails for mutations in the Csf3r (A), CSF3R (B) and GFP (C) sequences for t=15 days (left) and t=30 days (right), as a function of variant allele frequencies (VAF). As explained in the text, under exponential growth of cell and neutrality of mutations, in log-log coordinates the SFS tails should have slopes approximated by -1. This is not true of these genes (see the slope coefficients marked in the graph fields). Compare Figure 5, in which the Runx 1 sequence is shown to have slope coefficient -1. Departure from the -1 value may indicate selection effects (21)

Supplementary Table 1. Primers for *CSF3R*, *Csf3R* and *Runx1* used for library preparation of NGS. The primer sequences were ordered from IDT. Temperature gradient PCR's were performed to select the optimum annealing temperature of each primer

Funding

M.K. and S.J.C. were supported by NIH R01HL128173. K.D. and S.J.C. were supported by Alex's Lemonade Stand Foundation Innovation Award. M.K. was supported by the grant 2012/04/A/ST7/00353 and R.J. by the grant 2016/23/D/ST7/03665 from the National Science Center (Poland). S.B. was supported by the training grant T32 CA096520-14A1, R01 HL136333 and R01 HL134880.

Conflict of interest statement: None declared.

References

1. Klein, C. (2011) Genetic defects in severe congenital neutropenia: emerging insights into life and death of human neutrophil granulocytes. *Annu. Rev. Immunol.*, 29, 399–413.
2. Klein, C. (2009). Congenital neutropenia. *Hematol. Am. Soc. Hematol. Educ. Program* 344–350. doi:10.1182/asheducation-2009.1.344. PMID: 20008220.
3. Xia, J., Bolyard, A. A., Rodger, E., Stein, S., Aprikyan, A. A., Dale, D. C. and Link, D. C. (2009) Prevalence of mutations in ELANE, GFI1, HAX1, SBDS, WAS and G6PC3 in patients with severe congenital neutropenia. *Br. J. Haematol.*, 147, 535–542.
4. Köllner, I., Sodeik, B., Schreek, S., *et al*. (2006) Mutations in neutrophil elastase causing congenital neutropenia lead to cytoplasmic protein accumulation and induction of the unfolded protein response. *Blood*, 108, 493–500.

5. Nanua, S., Murakami, M., Xia, J., Grenda, D. S., Woloszynek, J., Strand, M. and Link, D. C. (2011) Activation of the unfolded protein response is associated with impaired granulopoiesis in transgenic mice expressing mutant *Elane*. *Blood*, 117, 3539–3547.
6. Tidwell, T., Wechsler, J., Nayak, R. C., *et al*. (2014) Neutropenia-associated ELANE mutations disrupting translation initiation produce novel neutrophil elastase isoforms. *Blood*, 123, 562–569.
7. Makaryan, V., Kelley, M. L., Fletcher, B., Bolyard, A. A., Aprikyan, A. A. and Dale, D. C. (2017) Elastase inhibitors as potential therapies for ELANE-associated neutropenia. *J. Leukoc. Biol.*, 102, 1143–1151.
8. Massullo, P., Druhan, L. J., Bunnell, B. A., Hunter, M. G., Robinson, J. M., Marsh, C. B. and Avalos, B. R. (2005) Aberrant subcellular targeting of the G185R neutrophil elastase mutant associated with severe congenital neutropenia induces premature apoptosis of differentiating promyelocytes. *Blood*, 105, 3397–3404.
9. Qiu, Y., Zhang, Y., Hu, N. and Dong, F. (2017) A truncated granulocyte colony-stimulating factor receptor (G-CSFR) inhibits apoptosis induced by neutrophil elastase G185R mutant: implication for understanding csf3r gene mutations in severe congenital neutropenia. *J. Biol. Chem.*, 292, 3496–3505.
10. Bonilla, M. A., Gillio, A. P., Ruggeiro, M., *et al*. (1989) Effects of recombinant human granulocyte colony-stimulating factor on neutropenia in patients with congenital agranulocytosis. *N. Engl. J. Med.*, 320, 1574–1580.
11. Cassinat, B., Bellanné-Chantelot, C., Notz-Carrère, A., *et al*. (2004) Screening for G-CSF receptor mutations in patients with secondary myeloid or lymphoid transformation of severe congenital neutropenia. A report from the French neutropenia register. *Leukemia*, 18, 1553–1555.
12. Dong, F., Brynes, R. K., Tidow, N., Welte, K., Löwenberg, B. and Touw, I. P. (1995) Mutations in the gene for the granulocyte colony-stimulating-factor receptor in patients with acute myeloid leukemia preceded by severe congenital neutropenia. *N. Engl. J. Med.*, 333, 487–493.
13. Germeshausen, M., Ballmaier, M. and Welte, K. (2007) Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis: results of a long-term survey. *Blood*, 109, 93–99.
14. Tidow, N., Pilz, C., Teichmann, B., Müller-Brechlin, A., Germeshausen, M., Kasper, B., Rauprich, P., Sykora, K. W. and Welte, K. (1997) Clinical relevance of point mutations in the cytoplasmic domain of the granulocyte colony-stimulating factor receptor gene in patients with severe congenital neutropenia. *Blood*, 89, 2369–2375.
15. Beekman, R., Valkhof, M. G., Sanders, M. A., *et al*. (2012) Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. *Blood*, 119, 5071–5077.
16. Skokowa, J., Steinemann, D., Katsman-Kuipers, J. E., *et al*. (2014) Cooperativity of RUNX1 and CSF3R mutations in severe congenital neutropenia: a unique pathway in myeloid leukemogenesis. *Blood*, 123, 2229–2237.
17. Zhu, Q. S., Xia, L., Mills, G. B., Lowell, C. A., Touw, I. P. and Corey, S. J. (2006) G-CSF induced reactive oxygen species involves Lyn-PI3-kinase-Akt and contributes to myeloid cell growth. *Blood*, 107, 1847–1856.
18. Fryxell, K. J. and Moon, W. J. (2005) CpG mutation rates in the human genome are highly dependent on local GC content. *Mol. Biol. Evol.*, 22, 650–658.
19. Eskandani, M., Hamishehkar, H. and Ezzati Nazhad Dolatabadi, J. (2014). 'Cytotoxicity and DNA damage properties of tert-butylhydroquinone (TBHQ) food additive'. *Food Chem.* 153: 315–320.
20. Durrett, R. (2013) Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann. Appl. Probab.*, 23, 230–250.
21. Spence, J. P., Kamm, J. A. and Song, Y. S. (2016) The site frequency spectrum for general coalescents. *Genetics*, 202, 1549–1561.
22. Xia, J., Miller, C. A., Baty, J., *et al*. (2018) Somatic mutations and clonal hematopoiesis in congenital neutropenia. *Blood*, 131, 408–416.
23. Donadieu, J., Leblanc, T., Bader Meunier, B., *et al*.; French Severe Chronic Neutropenia Study Group. (2005) Analysis of risk factors for myelodysplasias, leukemias and death from infection among patients with congenital neutropenia. Experience of the French Severe Chronic Neutropenia Study Group. *Haematologica*, 90, 45–53.
24. Watanabe-Smith, K., Godil, J., Agarwal, A., Tognon, C. and Druker, B. (2017) Analysis of acquired mutations in transgenes arising in Ba/F3 transformation assays: findings and recommendations. *Oncotarget*, 8, 12596–12606.
25. Shaner, N. C., Lambert, G. G., Chammas, A., *et al*. (2013) A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods*, 10, 407–409.
26. Sehgal, P., Szalai, P., Olesen, C., Praetorius, H. A., Nissen, P., Christensen, S. B., Engedal, N. and Møller, J. V. (2017) Inhibition of the sarco/endoplasmic reticulum (ER) Ca^{2+}-ATPase by thapsigargin analogs induces cell death via ER Ca^{2+} depletion and the unfolded protein response. *J. Biol. Chem.*, 292, 19656–19673.
27. Koh, K., Cha, Y., Kim, S. and Kim, J. (2009) tBHQ inhibits LPS-induced microglial activation via Nrf2-mediated suppression of p38 phosphorylation. *Biochem. Biophys. Res. Commun.*, 380, 449–453.
28. Shih, A. Y., Li, P. and Murphy, T. H. (2005) A small-molecule-inducible Nrf2-mediated antioxidant response provides effective prophylaxis against cerebral ischemia in vivo. *J. Neurosci.*, 25, 10321–10335.