Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2018

Genomic Instability in Severe Congenital Neutropenia, a Leukemia Predisposition Syndrome

ADYA SAPRA
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

 Part of the Medicine and Health Sciences Commons

© Adya Sapra

Downloaded from
https://scholarscompass.vcu.edu/etd/5639

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

©Adya Sapra 2018

All Rights Reserved

Genomic Instability in Severe Congenital Neutropenia, a Leukemia Predisposition Syndrome

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science at

Virginia Commonwealth University

By:

Adya Sapra

Bachelor of Technology, Banasthali University, India, 2015

ADVISOR: Dr. Seth Corey

Professor, Department of Pediatrics

Department of Human and Molecular Genetics

Virginia Commonwealth University

Richmond, Virginia

October 2018

DEDICATION

This research is dedicated to my grandfather

'Daddy, thank you for believing in me even when I did not. I hope to make you proud every single day.'

Also, my parents and my sister, Apoorva, who are my pillars of strength.

ACKNOWLEDGEMENT

I extend my sincere gratitude to my advisor, Dr. Seth Corey, for providing me with the opportunity to be his student and work under his guidance. I thank him for his motivation, encouragement and immense knowledge. He is an inspirational scientist and mentor without whose guidance and supervision this research would not have been possible. Besides my advisor, I also thank the other members of my committee: Dr. Larisa Litovchick and Dr. Kristoffer Valerie for their invaluable comments and insights. I'm also grateful to Dr. Gail Christie and Dr. Michael McVoy, from the Molecular Biology and Genetics department for their constant support and guidance.

I would also like to acknowledge the members of Corey lab, past and present, that have helped me along the way and with whom I have built lifelong friendships. I have been very fortunate to have the guidance of Dr. Hrishikesh Mehta, throughout the project. Rishi has mentored me since my first day in lab with more patience and dedication than I deserve. I also thank Dr. Bhavuk Garg and Dr. Usua Oyarbide for all of their support and guidance, both professionally and personally. Our lab has been excellently managed by our lab managers, Sharon, Wilmer and Camden. I'm lucky to have found great friends in my colleagues Ann, Ralph, Abigail, Jessica and Anthony. A big shout out to Borwyn Ann Wang for being my go-to person. Ann has been my colleague, friend, guide, therapist and prankster, all in one. I would also like to acknowledge Ralph, without whom, this thesis would've been completed a few months earlier. But without this distracting, amazing friend, this journey would not have been as memorable.

The completion of this project would be impossible without our collaborators. I thank Dr. Myrna Serrano and the VCU Nucleic Acid Research Facilities for their sequencing services. I'm also grateful to Dr. Roman Jaksik, Systems Engineering Group, Silesian University of Technology, Gliwice, Poland, and Dr. Marek Kimmel and Sara Biesiadny from Department of Statistics and Bioengineering, Rice University for their invaluable contributions to the bioinformatic and statistical analysis of the sequencing results.

A huge thanks to all my friends in Richmond, who have been more than family to me. Thank you Sweta, Nilufa, Hossain, Khushbu and Aneel for making this city home.

Finally, I would like to thank my family for their unconditional love and support all these years. Words are not enough to express my gratitude to my parents, who have supported and nurtured all my dreams. Their confidence and trust in me are my biggest strengths. My grandparents, who have been my inspiration in life. My sister, Apoorva, who's unmatched love and admiration gives me strength and courage in toughest of situations. My best friend Vinny, for sticking with me for the past 2 decades. Whatever I have achieved up till now is dedicated to these people and I hope them to be a part of all the future adventures that life has in store for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2-AP	2-aminopurine
5-BU	5- Bromouracil
AML	Acute Myeloid Leukemia
ATF4	Activating Transcription Factor 6
ATF6	Activating Transcription Factor 6
BiP	Binding Immunoglobulin Protein
cDNA	Complimentary DNA
CHOP	CCAAT/Enhancer-Binding Protein Homologous Protein
CSF3R	Human Granulocyte-Colony Stimulating Factor Receptor
Ct	Threshold Cycle
ddCt	Delta Delta Ct
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
ELANE	Human Neutrophil Elastase Gene
EMS	Ethyl Methanesulfonate
ER	Endoplasmic Reticulum
FACS	Fluorescence activated cell sorting
FBS	Fetal Bovine Serum
GADD153	Growth Arrest and DNA Damage-Inducible Protein 153
G-CSF	Granulocyte-Colony Stimulating Factor
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
HBSS	Hank's Balanced Salt Solution
HSPA5	Heat Shock Protein Family A (HSP70) Member 5

IRE1alpha	Inositol-requiring 1alpha
LD	Lethal Dose
LFC	Log Fold Change
MDS	Myelodysplastic Syndrome
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
NG	Nitrosoguanidine
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
PERK	protein kinase RNA-like endoplasmic reticulum kinase
qPCR	Quantitative-PCR
RNA	Ribonucleic Acid
ROS	Reactive Oxidation Species
RPMI 1640 Medium	Roswell Park Memorial Institute Medium
RUNX1	Runt Related Transcription Factor 1
SCN	Severe Congenital Neutropenia
SCNIR	Severe Chronic Neutropenia International Registry
TBHQ	Tert-butylhydroxyquinone
Tg	Thapsigargin
UV	Ultraviolet
UPR	Unfolded Protein Response
XBP1	X-box binding protein 1

ABSTRACT

Genomic Instability in Severe Congenital Neutropenia, a Leukemia Predisposition Syndrome

By Adya Sapra, B.Tech

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Director: Seth Corey, MD, MPH

Professor, Department of Pediatrics

Department of Human and Molecular Genetics

Virginia Commonwealth University

Severe congenital neutropenia (SCN) is a rare blood disorder characterized by abnormally low levels of circulating neutrophils. Mutations in multiple genes like neutrophil elastase gene (*ELANE*) and granulocyte colony stimulating factor receptor (*CSF3R*) may cause SCN. The treatment of choice for SCN is the administration of granulocyte-colony stimulating factor (G-CSF) which elevates the neutrophil count and hence improves the survival and quality of life. Long term survivorship on G-CSF is however linked to development of MDS (myelodysplastic syndrome)/AML (acute myeloid leukemia). About 70% of MDS/AML patients acquire nonsense mutations affecting the cytoplasmic domain of *CSF3R*. In this project, we hypothesized 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 study the effect of induced oxidative or ER stress on the mutation rate in our hypothesized hotspot of the

exogenous human *CSF3R*, the corresponding region in the endogenous *Csf3r,* and a leukemia-associated gene *Runx1*. Ba/F3 cells transduced with the cDNA for partial C-terminal of *CSF3R* fused in-frame with a Green Fluorescent Protein (GFP) tag was subjected to cellular stress inducing mutagen treatment for a prolonged period of time (30 days). The amplicon based targeted deep sequencing data for days 15 and 30 samples show that although there was increased mutagenesis observed in all genes, there were more mutations in the GFP region as compared to the GC-rich partial *CSF3R* region. Our findings also indicate that there is no correlation between the stress-inducing chemical treatments and mutagenesis in Ba/F3 cells. Thus, we conclude that there are other mechanisms to acquired mutations of *CSF3R* that help drive the evolution of SCN to MDS/AML. To test this hypothesis, further experiments using unique barcoding system are in progress to characterize the clonal competition between different mutant *CSF3R* and *ELANE* expressing cell lines. This study will shed further light on the selection advantage that is provided to cells because cooperativity between mutations in different genes.

.

CHAPTER 1
INTRODUCTION

Severe Congenital Neutropenia: a leukemia predisposition syndrome

Background: Severe Congenital Neutropenia (SCN) is a genetically heterogeneous condition characterized by pathologically low levels of neutrophils. This disorder was first described in 1950 by Rolf Kostmann, who coined the name infantile genetic agranulocytosis to this disease[1]. In his doctoral thesis published in 1956, Kostmann described the clinical and histologic course of SCN in 14 patients, belonging to 9 consanguineous families from Northern Sweden. He also observed the heterogeneity in the genotype and phenotype of SCN[1–3]. In 1970, SCN was first identified as a pre-leukemic syndrome.

As part of the innate immune system, neutrophils provide the first line of defense against microbes by secreting several antimicrobial peptides that can modulate the cytokine and chemokine networks. Thus, neutrophils are critically responsible for providing an early response in both killing antigens and reconstituting tissue integrity[2,4,5]. Clinically significant neutropenia, which can cause an individual immunocompromised is diagnosed when the peripheral absolute neutrophil count (ANC) is lower than 1500/µl. SCN is defined as ANC of less than 500/µl[3]. Patients with SCN are prone to life-threatening, recurrent bacterial infections such as otitis media, bronchitis, pneumonia, osteomyelitis, or cellulitis. Many SCN patients have been also diagnosed with chronic gingivitis and tooth decay. In addition, decreased bone mineral density, leading to ostopenia or osteoporosis and increased propensity to fractionation, are other common clinical problems associated with SCN[3]. Clinical diagnosis of SCN includes examination of bone marrow for accumulation of early myeloid precursors. SCN patients show maturation arrest in the early stages of neutrophil

development in the bone marrow which leads to accumulation of undifferentiated myelocytic progenitors (Figure 1).

Molecular Genetics of SCN

Mutations in multiple genes have been implicated in 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*.

The most frequent mutations in SCN patients involve *ELANE* and are found in approximately 40% of the patients. To date, more than 100 mutations have been identified in *ELANE*-associated SCN[6], which are either transmitted in autosomal dominant pattern or are spontaneously acquired[7,8]. The leading hypotheses state that germline *ELANE* mutations result in misfolding and/or mislocalization 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. 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 of excessive ER stress, these signals pathways get activated in order to maintain homeostasis in the cell. In case of all three UPR mechanisms fail, UPR then promotes apoptosis of the cell. Thus, UPR, activated due to ER stress, causes apoptosis of differentiating neutrophil precursors at the promyelocyte stage, leading to accumulation of promyelocytic progenitors and subsequent deficiency of mature neutrophils[9]. Thus, SCN is associated with both germline and spontaneous *ELANE* mutations. Germline *ELANE* mutations in

SCN are further associated with spontaneous acquired *CSF3R* mutations, which may have a role in the evolution of SCN to AML.

Treatment of SCN

The most important therapeutic goal in patients with SCN is the reconstitution of adequate antibacterial host defense by increasing neutrophil numbers. Marrow cells from SCN patients display a reduced responsiveness to granulocyte-colony-stimulating factor (G-CSF). The first line of treatment of choice for SCN is the administration of recombinant human G-CSF which binds to and activates a specific receptor (*CSF3R)* which further 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].

Although G-CSF successfully induces proliferation, differentiation and survival of myeloid progenitors, however, long term survival with G-CSF is linked to the development of myelodysplastic syndrome/acute myeloid leukemia (MDS/AML). Approximately 30% of SCN patients undergoing G-CSF treatment develop AML or MDS[6,10,11]. Interestingly, about 70% of SCN patients with MDS/AML acquire nonsense mutations in the region of CSF3R that encodes the cytoplasmic domain. Originally, it was believed the main cause of SCN is germline mutations in *CSF3R*[12]. However, subsequent studies[13,14] indicate that the *CSF3R* mutations that have been found in previous SCN patients were somatic or de novo in meiosis, rather than being germline in nature. These non-inherited *CSF3R* mutations are generally characterized by a truncation variant, usually found in the C-terminal of the *CSF3R*, and are associated with a hyperproliferative phenotype in the mutated cells observed in the leukemogenic events seen with SCN[4,11].

Clonal Evolution of SCN to MDS/AML: Literature Review

Previous studies on SCN patients have shown that initial driver mutations involve *CSF3R*, and cell clones harboring *CSF3R* mutations have a growth advantage and acquire additional cooperating mutations that contribute to AML initiation and disease progression[4,5]. As reported by Skakova et al[5] out of 31 SCN patients who developed MDS/AML, 81% patients (23) had acquired *CSF3R* mutations and 65% patients (20) had mutations in *RUNX1*. Furthermore, 17 patients had mutations in both *CSF3R* and *RUNX1*. Another interesting observation was that all the 6 patients who were negative of both *RUNX1* and *CSF3R* mutations developed only MDS, as compared to the group of 17 patients with cooperative *CSF3R* and *RUNX1* mutations, out of which 11 developed AML and 6 developed MDS/AML. Thus, a high frequency of cooperativity between *CSF3R* and *RUNX1* mutations occurs in the transformation of SCN to AML. In another clinical study[5] by the same group, the mutational landscape of one SCN patient under G-CSF therapy who developed AML was done. The sequencing results showed that the *CSF3R* mutations were observed at an initial onset of SCN and *RUNX1* mutations were observed just a few months prior to onset of AML. Furthermore, both *CSF3R* and *RUNX1* mutations diminished immediately after the G-CSF treatment was stopped temporarily and the mutations re-emerged after the treatment was restarted[5]. These studies suggest that G-CSF treatment is directly linked with increased mutagenesis which leads to development of MDS/AML in SCN patients. Also, mutations in different genes work in cooperativity to provide the clonal advantage to cells.

More than 70% of SCN patients who developed MDS/AML had mutations in the cytoplasmic domain of *CSF3R*[11]. In the most comprehensive study to date, Germeshauisen et al analyzed the genomic data of 218 patients with congenital neutropenia to determine the incidence of frequently mutated sites in *CSF3R* associated with the development of MDS/AML. Nonsense mutations in seventeen different nucleotide positions in the *CSF3R* intracellular domain, leading to truncation of one to four tyrosine residues, were identified. Incidentally, all of these mutations were

present in the cytoplasmic domain of the *CSF3R.* Since, this region of the cytoplasmic domain was most highly mutated, we have transduced this specific region in our cell line models for our mutagenesis study. The seventeen most frequently mutated nucleotide positions in *CSF3R* as detected by Germeshausen et al. are shown in Figure 2. The mutations that were detected most frequently in patients with SCN were truncation mutations affecting glutamine at the following amino acid positions: d739, d741, d749 and d754 (d715, d717, d725 and d730 respectively, according to the old nomenclature, which excludes 24 amino acids of the signal peptide region). Other less frequent mutations included truncation mutations in glutamine (725, 730, 743, 759 and 774), tyrosine (727, 752, 767 and 787), arginine (734) and Lysine (785); frameshift mutation in Serine (747); and deletion of 10 bp (2848-2857). These nucleotide changes are represented in Figure 3.

We hypothesize that this region corresponding to the cytoplasmic domain of CSF3R constitutes a mutational hotspot. This vulnerability to mutations may result from excessive oxidative or ER stress. In SCN, myeloid precursors are under stress due to cellular responses like UPR. Under stress conditions, reactive oxidation species (ROS) production increases, and G-CSF treatment can lead to a further increase in ROS production. Thus, the combination of stress pathways and the G-CSFR-mediated signaling transduction pathways creates an environment of enhanced oxidative stress, which may promote genomic instability. This study is focused on determining the mutation rate of cytoplasmic domain of CSF3R under conditions of ROS and UPR stress **(Aim 1)**.

Molecular biology of mutagenesis and mutation hotspots

Mutagenesis is the process of introducing permanent alterations, called mutations, in the sequence of DNA, that might cause genetic instability. These mutations can affect a single base pair to a large segment of a chromosome that includes multiple genes. The mutations in genomic DNA

can be classified into the following groups and subgroups: point mutations (transitions and transversions), deletions/insertions, duplications, inversions, and chromosomal rearrangements. Point mutations are single base-pair substitutions that can be further classified in the following subgroups: silent mutations (single nucleotide substitutions that do not change the amino acid after translation); missense mutations (single nucleotide substitutions that can change the amino acid coded during translation); nonsense mutations (premature stop codon substitution) and frameshift mutations (insertion or deletion of nucleotides causing change in the reading frame of the cDNA). Transition mutations occur when either a purine (A,G) mutates to the other purine or a pyrimidine (C,T) mutates to the other pyrimidine. Transversion mutations are substitution of purines with pyrimidines and vice versa[15]. Complex mutations are rare mutations that consist of combinations of several point mutations. Point mutations are often a direct result of mutagenesis whereas genetic recombination causes other types of mutations like chromosomal rearrangement[15–18].

Mutations can be inherited (germline) or acquired (somatic). As mentioned earlier, SCN is known to be caused due to germline mutations in various genes like *ELANE* and *HAX1*. These mutations can be inherited in an autosomal dominant or recessive manner. Recessive mutations lead to loss of function, whereas dominant mutations often, but not always, result in gain of function[19]. Acquired mutations, on the other hand, occur spontaneously in certain somatic cells, and unlike germline mutations, are not present ubiquitously. These mutations can be induced, i.e. caused due to external factors like ultraviolet radiations or chemical mutagens; or spontaneous, i.e. can occur due to errors during DNA replication, recombination or repair, or due to endogenous DNA damage[20,21].

There are two different classes of mutagenic mechanisms: directly induced base mispairing and misrepair. Spontaneous or induced mispairing can lead to specific base-pair substitutions[21]. Mispairing can occur in various ways. Tautomerization of bases often leads to “mistaken” mispairing. The keto form of each base is normally found in DNA, whereas the enol forms of the

bases are rare. The complementary base pairing of the enols is different from that of the keto forms. For example, as shown in Figure 4, during replication, guanine changes to its enol form, which binds to an incoming thymine instead of its pair cytosine. This mispairing thus leads to an AT mutation instead of GC in the next round of replication[22]. Mispairing also occurs when different bases are ionized simultaneously. 5- Bromouracil (5-BU) and 2-aminopurine (2-AP) are base analogs that can induce mutagenesis. 5-BU is an analog of thymine that can be mistakenly incorporated into DNA as a base. In its normal keto state, 5-BU mimics the pairing behavior of the thymine that it replaces, pairing with adenine. The presence of the bromine atom, however, causes a relatively frequent redistribution of electrons, so 5-BU can spend part of its existence in the rare ionized form. In this state, it pairs with guanine, mimicking the behavior of cytosine and thus inducing mutations in replication[23]. Similarly, 2-AP, an analog of adenine, pairs with thymine, but when protonated, it can mispair with cytosine, thereby inducing mutations during replication. Alkylating agents, such as ethyl methanesulfonate (EMS) and nitrosoguanidine (NG), alter the nucleotide bases, thus causing specific mispairing. For example, EMS-generated ethylation causes alkylation of O-6 position of guanine and O-4 position of thymine, that can lead to direct mispairing of these bases. Another common cause of spontaneous mutation due to misparing is the deamination of cytosine to uracil in the DNA double helix. This leads to the replacement of the wild-type GC base pair with a TA base pair in the subsequent daughter cells[24]. Thus, regions that have higher GC content are more prone to mutagenesis and usually can be considered as mutational hotspots. Another mechanism of inducing mutagenesis is misrepair mutagenesis, which can be inferred either directly, by obtaining appropriate repair defective mutants, or indirectly, by ultraviolet (UV) and ionization radiation[25].

Mutation frequency can vary significantly along nucleotide sequences, often leading to the concentration of maximum mutations in certain positions, which are known as mutational 'hotspots'[26]. The question that arises here is that what makes certain regions in the DNA more prone to mutagenesis? Several studies have been done to answer this question[27–31]. Mutational hotspots

reflect specific mechanisms of clustering of mutations in specific regions. Thus, to study mechanism of mutagenesis and their effects on the functional domains of target proteins, it is of outmost importance to study mutation hotspots.

Serial acquisition and cooperativity of mutations in clonal hematopoiesis

Previous clinical and experimental studies have indicated that different mutations work in cooperativity to provide a clonal advantage to the cells. Beekman et al. in 2012 identified early and late genetic effects associated with leukemic progression in one SCN patient. Germline *ELANE* mutation was detected soon after birth. During the early stages of SCN, four mutations in *CSF3R* (d715 accompanied by *LLGL2* mutation; d717; d7125 and d730) arose. However, as SCN progressed towards AML, only one clone containing *CSF3R* d715 mutations was able to proliferate. This study suggests that additional mutations in *LLGL2* and *ZC3H18* acquired during the progression of SCN to AML provided a selective proliferative advantage to the cells. It can be concluded from this study that remaining three clones were probably unable to survive because of lack of additional cooperating mutations that were present in case of d715 mutations. It is unclear if mutations in specific genes can cooperate with each other to provide the cells with selection advantage. The second part of our study focuses on answering this question by studying the clonal competition between different *CSF3R* and *ELANE*-expressing cell lines. We hypothesize that the mutant *CSF3R* and *ELANE* will provide a selective advantage to the cells **(Aim 2)**. We will investigate the proliferative advantage of different *CSF3R* and *ELANE* mutant expressing cell lines using unique barcoding system which can be detected by Next Generation Sequencing Analysis (NGS).

CHAPTER 2

MATERIALS AND METHODS

Cells and Culture Conditions

Ba/f3 cells were obtained from ATCC and grown and maintained in 1X RPMI-1640 medium (CellGro) which was supplemented by 10% Fetal Bovine Serum (FBS), 1% PenStrep, 1X Glutamax and 2ng/ml mIL3. The cells were passaged every $1^1/_2$ days and the cell density was maintained between $1x10^5$ (lower limit) and $1x10^6$ (upper limit) cells/ml. The cells were maintained and used for not more than 2 months post thaw date.

Transduction of Ba/F3 cells with Partial C terminal of CSF3R construct

Lentiviral transduction was carried out using the spinoculation procedure to transduce the partial C terminal of *CSF3R* fused with a GFP construct (Figure 5) in Ba/f3 Parent cell line. 1-2 million cells were centrifuged and resuspended in 1 ml of the crude supernatant obtained during viral production from 293 FT cells. After resuspension, the cells were plated in a 24 well plate and spinoculation was performed. The cells were centrifuged at 1000 x g four one hour at room temperature. The cells were then incubated at 37°C for another 4 hours. The supernatant was discarded and fresh media (containing mIL3) was added. To confirm confection, cells were checked for fluorescence.

Selection of transduced cells

Transduction was confirmed using FACS, and GFP positive cells were sorted. To select cells, the cells were incubated in the LD100 dose of the antibiotic, Zeocin (1000 µg/mL) for 7 days. The zeocin kill curve was established to calculate the minimum lethal dose of zeocin, i.e. the lowest concentration of zeocin that can cause cell death (LD100). Freshly harvested cells were incubated

in 6 wells of a 24 well plate (around 600,000 cells/well). Serial dilution was used to treat cells in each well with an increasing Zeocin concentrations: 100, 200, 400, 800, and 1000 µg/ml. Negative control cells were treated with only DMSO (solvent). Cell counts were performed on day 2, 4, 6 and 8 of treatment. Respective Zeocin concentrations were maintained in fresh media with every passaging.

Dose response curve for stress inducing chemicals: MTT assay

The dose response was formulated to determine the concentration of the stress inducing drugs, Thapsigargin and TBHQ, that the cells would be exposed to induce ROS and UPR respectively. 1000 cells were plated in each well of a 96 well plate. The cells were treated with increasing concentrations of thapsigargin (0, 0.064, 0.32, 1.6, 5, 8, 40, 200, 1000 nM, respectively). The cells were incubated for 48 hours and then, MTT assay was performed. At the desired time point (48 hours), 10 µL of (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (MTT) reagent was added each well. After 4 hours of incubation at 37°C, 100 µL of MTT detergent was added. The plate was incubated in dark at room temperature overnight. Absorbance was measured at 545 nm. LD10 and LD20 were selected as optimum dose of stress inducing treatment for both drugs. The above steps were repeated with TBHQ treatment.

Flow Cytometry Analysis Measurement of ROS production

200,000 cells were plated in each well of a 6-well plate. The cells were treated with the following concentrations of Thapsigargin respectively: 0 nM (only DMSO) 1nM, 2nM, 5nM and 10nM and were incubated for 48 hours. The cells were then treated with CellRox Deep Red Reagent (Thermo Fisher Scientific) at a final concentration of 5 µM and incubated for 30 minutes at 37°C. The cells were then washed with HBSS buffer and further treated with MitoSox red Reagent (Thermo Fisher

Scientific) for 10 minutes. The cells were washed twice with HBSS and finally suspended in 1ml media before carrying out Flow cytometry analysis.

q-RT-PCR

qPCR was performed to estimate the levels of different UPR markers (ATF4, ATF6, BiP, CHOP, $XBP1_{spliced}$ and $XBP1_{unspliced}$) in Ba/F3 cells treated with different doses UPR inducing chemical, thapsigargin (1 nM, 2 nM and 5 nM). At each desired time point of the treatment, cell pellets were collected and RNA was extracted using the Qiagen RNeasy Plus Mini Kit. Cells were lysed with 350 µl of Qiagen lysis buffer, buffer RLT and homogenized by vortexing. The homogenized lysate was placed in the genomic DNA eliminator spin column and centrifuged at $\geq$ 8000 x g to separate the gDNA and RNA. 1 volume of 70% ethanol was added to precipitate the RNA out of the solution. The solution was then pipetted into a RNeasy Mini spin column and centrifuged. The filter was washed with buffer RW1 and RPE (twice). After drying the membrane, the RNA was eluted using 50 µl of RNase free water. RNA concentrations were measured using NanoDrop One (Thermo Scientific). cDNA was synthesized from the extracted RNA using the iScript™ cDNA Synthesis kit (Bio-Rad). Quantitative PCR was performed on the StepOnePlus™ Real Time PCR System (Applied Biosystems, Thermo-Fisher Scientific). For the relative quantification of ATF4, ATF6, BiP, CHOP and XBP1 expression, Actin was selected as an internal control. Power SYBR Green (Applied Biosystems, Thermo Fisher) was used as a DNA intercalator dye to monitor amplified DNA quantification, and real-time quantitative PCR curves were analyzed by StepOne Software v2.3 (Applied Biosystems) in order 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

Treatment was carried out on Ba/F3 transduced cells. Cell pellets were stored at -80°C on day 0 to serve as control. 1000 cells/well were plated in a 96-well plate using serial dilution. The cells were then treated with the following stress-inducing treatments: thapsigargin 1 nM and 2 nM, TBHQ 1 nM and 2nM, and control (untreated) samples. Each treatment was done in triplicate. The samples were frozen down at regular intervals (every time the cells were split), as shown in Figure 6. For initial screening of mutational landscape, samples from day 0, 15 and 30 (for each treatment) were sent for sequencing. The remaining samples were stored for future follow-up studies.

Next Generation Sequencing: Library Preparation

Mutational analysis of the genes of interest was performed using a sensitive next-generation amplicon deep sequencing assay. The samples were sent to the Nucleic Acid Research Facilities at Virginia Commonwealth University for targeted deep sequencing. For the library preparation, genomic DNA was first extracted from the frozen pellets of all samples using the Qiagen DNeasy Blood and Tissue DNA Extraction kit. Targeted deep sequencing was performed using the tailed amplicon sequencing approach with Illumina MiSeq. The sequencing was carried out in three steps: First round of PCR's to amplify region of interest, second round of PCR's to introduce sequencing adapters, and finally, subsequent sequencing in forward and reverse direction (Figure 7).

The first step was to amplify the region of interest for each of the samples. Our genes of interest were: the exogenous *CSF3R* (transduced construct); endogenous *Csf3r* (Exon 17) and endogenous

Runx1. To amplify the DNA, PCR's were performed using Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB). Since the cDNA of all the 3 genes of interest was too large to amplify the whole region together, so the amplification was carried out in segments for each gene. The primers were designed for overlapping sections in each region of interest.

Each primer contained an overhang adapter sequence next to the locus-specific sequence which helped in pooling the sequences back. The primers were then optimized by performing temperature gradient PCR's for each primer. The primer sequences used for the analysis and the final annealing temperatures for each primer are given in Appendix Table 6.

Once the amplification was carried out and the sequence size was confirmed using PCR, the segments for each gene were pooled together and barcodes were introduced for identification purposes before being sequenced. A pool of 96 samples were sequenced in a single library. The sequencing was carried out in 2 sets to get approximately 100,000 reads/segment.

Next Generation Sequencing Analysis

The bioinformatic analysis on the targeted deep sequencing data was done by Dr. Roman Jaksik, Silesian University, Poland. Broad Institute (GATK v4) guidelines were followed for the analysis. 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.

Statistical Analysis

The statistical analysis of the sequencing data was performed by Sara Biesiadny and Dr. Marek Kimmel, Department of Statistics, Rice University, Texas. The graphs were plotted using R statistical software. The data was analyzed to calculate the trend of mutations that were detected during the treatment period. Comparisons were done between the regions of interest, and between different treatments. The data was divided into different subsets, one for each gene being sequenced. 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.

Preparation of Barcoded Lentiviral Strains

The barcode sequences were obtained from Dr. David Sabatini's lab at Whitehead Institute for Biomedical Research, Cambridge, Massachusetts. Each barcode had a unique 7 base pair sequence integrated in PlkO.1 vector. An example can be seen in Figure 9 which shows the plasmid map of barcode 69, integrated in plko.1 vector. The plasmids were then transformed into One Shot™ STBL3™ chemically competent E.Coli. Approximately 1 ng of barcode DNA was added to 12.5 µl of competent cells and incubated on ice for 30 minutes. The cells were then subjected to heat shock (42°C) for 45 seconds. After incubating the cells on ice for 2 minutes, 250 µl of S.O.C. Medium was added to each vial. The cells were then incubated while shaking at 225 rpm for 1 hour at 37°C. 25-100 µl of each transformation was spread on ampicillin plates and incubated overnight at 37°C. Individual colonies were then selected and the plasmid DNA was purified using QIAprep Spin Miniprep Kit (Qiagen). The barcode sequences were confirmed by sanger sequencing (Genewiz).

Preparation of different mutant CSF3R and ELANE expressing cell lines

Ba/F3 parent cells were divided in 2 sets and transduced with wildtype and G185R *ELANE* respectively. Transduction was performed using polybrene. Approximately 5000 cells (in 100 µl media) were plated in 2 wells of a 96 well plate. The cells were then treated with polybrene at a final concentration of 2 µg/ml. Wildtype and G185R *ELANE* viral particles were then added to each respective set. The cells were centrifuged at 600 x g for 45 minutes to integrate the viral particles into the cells. The cells were then incubated at 37 °C for 1-2 days. Protein expression was confirmed using western blotting technique. Once *ELANE* transduction was confirmed, both wildtype and mutant *ELANE* expressing cell lines were further divided into 2 sets. Each cell line was then transduced with wildtype and d715 *CSF3R.* the cells were selected using fluorescence-activated cell sorting (FACS). Finally, each of the 4 cell lines were then transduced with a unique barcoded lentivirus. Transduced cells were selected by puromycin selection. The cell lines were generated to have specific combinations of *CSF3R* and *ELANE* mutations, as shown in Table 2. Once the cell lines were prepared, single cells were sorted on a FACS instrument (BD LSRFortessa-X20™) into a 96 well plate. Live cells were identified by light scatter characteristics, and doublets were excluded from the analysis. Once the single cell clones were sorted, they were allowed to proliferate, and finally 3 clones from each barcode induced cell line having similar levels of *ELANE* and *CSF3R* expression were chosen for the experiment.

CHAPTER 3

RESULTS

Transduction of Ba/F3 Cells

Lentiviral transduction was carried out using the spinoculation procedure to insert the Fused partial C terminal of CSF3R and mNeon Green region into the Ba/F3 parental cells. Flow cytometry analysis on the transduced Ba/F3 cells showed the presence of the GFP-negative and GFP-positive cells. The cells were then subjected to antibiotic selection for the transduced cells. The chosen antibiotic was zeocin, which belongs to the bleomycin family and is isolated from isolated from Streptomyces verticillus. It is a basic, water-soluble and copper-chelating glycopeptide, and contains a phleomycin D1 frame. Once the Once being activated by reducing Cu^{2+} to Cu^{1+} in cells, Zeocin causes double strand breaks (DSBs) of DNA. Prolonged treatment with high non-lethal doses of zeocin can have a mutagenic effect. Thus, to avoid any additional mutagenesis, the cells were subjected to selection treatment for a short period of time, i.e. 7 days, to avoid any long-term mutagenic effects. For the selection, the LD100 dose (1000μg/ml) of zeocin was used to eliminate all the untransduced cells (Figure 10).

Dose-Response Curves for Stress Inducing Chemicals

TBHQ was used for inducing oxidative stress and thapsigargin for ER stress. The dose-response curve for each chemical was performed and the IC50 values for both the drugs was calculated to be 4.26nM for Thapsigargin and 4.09nM for TBHQ. LD10 (1 nM) and LD20 (2 nM) doses were selected for treatment in order to introduce cellular stress but avoid avid cell death (Figure 11).

Thapsigargin and TBHQ induce increased oxidative stress

Ba/F3 parent cells were treated with increasing doses of TBHQ (1 nM, 2 nM, 5 nM and 10 nM) for 6 hours to induce cellular oxidative stress. ROS production was also determined following treatment with thapsigargin, a non-competitive inhibitor of sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (SERCA), which promotes oxidative stress indirectly by inducing UPR. Thapsigargin leads to decrease in the ER calcium levels, thereby leading to accumulation of excess Calcium in the cytosol. Without calcium, the calcium-dependent ER chaperones lose their protein folding activity, thus leading to accumulation of misfolded proteins in the ER. This triggers the increase in free radicals, leading to oxidative stress conditions.

To estimate the levels of ROS production due to increasing chemical treatment, MitoSOX and CellROX Green oxidative stress probes were used to label the mitochondrial and cytosolic ROS, respectively, followed by detection of fluorescence using flow cytometry. The data shown in Figure 12, confirmed that there is an increase in cellular oxidative stress at our selected doses, 1nM and 2 nM for both the chemicals. Extensive cell death was observed at higher doses, i.e. 5 nM and 10 nM. Furthermore, Thapsigargin was observed to induce comparatively more ROS production than TBHQ. Also, cytosolic ROS was produced at a more consistent rate as compared to mitochondrial ROS, which was somewhat erratic.

Thapsigargin treatment also induces ER stress

Ba/F3 cells were treated with three different doses of thapsigargin: 1 nM, 2 nM and 5 nM for different time intervals (0-48 hours) to study the dose-response and time-dependent effects of thapsigargin on the ER stress levels. qPCR was performed to measure the levels of UPR markers *Atf4*, *Atf6*, *HSPA5 (BiP)* GADD153 (*Chop)*, and *Xbp1* (Figure 13). Although there was no increase

observed in the first 8 hours of treatment, a steep increase in all ER stress marker levels occurred after prolonged treatment (24 hours and 48 hours).

NGS Analysis: New Mutations observed in both treated and untreated samples

No correlation was seen between the treatments with thapsigargin and TBHQ and the total number of new mutations. NGS analysis showed 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 the endogenous genes (*Csf3r* and *Runx1)*. However, in the exogenous transduced gene (*CSF3R)*, there were a greater number of new mutations in the treated samples as compared to the untreated controls on days 15 and 30 (Figure 14). The number of mutations observed in each sample is provided in Appendix Table 3.

GFP region was mutated more than CSF3R

A comparative study was done to study the LFC (Log Fold Change) between the number of mutated positions relative to the region size for *CSF3R* and *GFP*. Surprisingly, for all the samples, the GFP region was more mutated as compared to the hypothesized hotspot, *CSF3R.* The number of mutations calculated were relative to the region size, i.e. 427 bp for *CSF3R* and 617 bp for GFP. The p-values for each treatment was calculated and the samples that had significant p-values, i.e. p-value > 0.05, were represented by triangular markets in Figure 15. The values used to make the plots are given in Appendix Table 4. The table provides the number of mutations in both CSF3R and GFP regions, for each sample and treatment. The proportion of mutations with respect to the region size are also given. These values were used to calculate the LFC proportions.

Mutational Analysis: Three major trends observed

The mutations observed in all the samples could be categorized into 3 disjoint groups as represented in Figure 16: the first represent mutations whose frequency increased by 15 days, followed by diminishing frequency by day 30; the second set represent the mutations for which frequency increased on day 15 and continued to increase on day 30; and the third set consisted of mutations that were dormant until day 15 and emerged only at day 30. The first group of mutations compared to second group, also shows a faster increase in frequency. The values which were used to plot the graphs are provided in Appendix Table 5. The date is divided into different sets according to the gene of interest. For each gene, there are further subsets for each treatment viz.: Untreated, TBHQ 1nM; TBHQ 2 nM; thapsigarging1 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.

Reagent development for future experiments: Clonal Competition between different Ba/F3 mutant cell lines

Ba/F3 cell line was transduced with lentivirus containing *ELANE* wildtype and *ELANE G185R*. Transduction was confirmed using western blotting (Figure 17). Each of these cell lines, were then transduced with *WT and d715 CSF3R,* thus leading to 4 different cell lines. Transduced cells were assessed for GFP expression by FACS (BD LSRFortessa-X20™) to confirm the CSF3R transduction. Selection of transduced cells was performed using FACS where GFP positive cells were sorted and untransduced GFP-negative cells were eliminated (Figure 18). Each of the 4 cell lines were then transduced with lentivirus containing unique barcodes (Table 1). Puromycin selection was then performed to select the transduced cells. To calculate the lethal dose of puromycin for Ba/F3 cells, a dose response curve was formulated as shown in Figure 19. The cells

were then treated with the lethal dose (LD100) of 3 μg/ml for 7-10 days to select the transduced cells.

CHAPTER 4

FIGURES AND TABLES

1.

Figure 1. Maturation block at promyelocyte-myelocyte stage observed in SCN[1]. Granulopoiesis begins when the myeloblaöst differentiates into a neutrophilic promyelocyte that is irreversibly committed to the neutrophilic cell line. Promyelocytes are large cells with purple-staining nonspecific azurophilic granules. Promyelocytes develop into myelocytes that are characterized by the presence of smaller specific or secondary granules. Granule production ceases at the end of the myelocyte stage and the remaining stages are characterized primarily by a reduction in cell size and a change in nuclear shape. When the nucleus becomes flattened and the chromatin further condensed, the cell is called a metamyelocyte. When the nucleus becomes horseshoe-shaped, it is called a mature band cell. The cell is considered a mature neutrophil when the nucleus becomes segmented into lobes. In SCN patients, there is a maturation arrest in the promyelocytic stage, which leads to an accumulation of immature granulocytic precursors in the bone marrow. Neutrophil elastase is expressed at its highest levels in the promyelocytic stage and GCSFR is primarily expressed in neutrophils, although, some traces of CSF3R expression is also observed in neutrophil progenitors[32].

2.

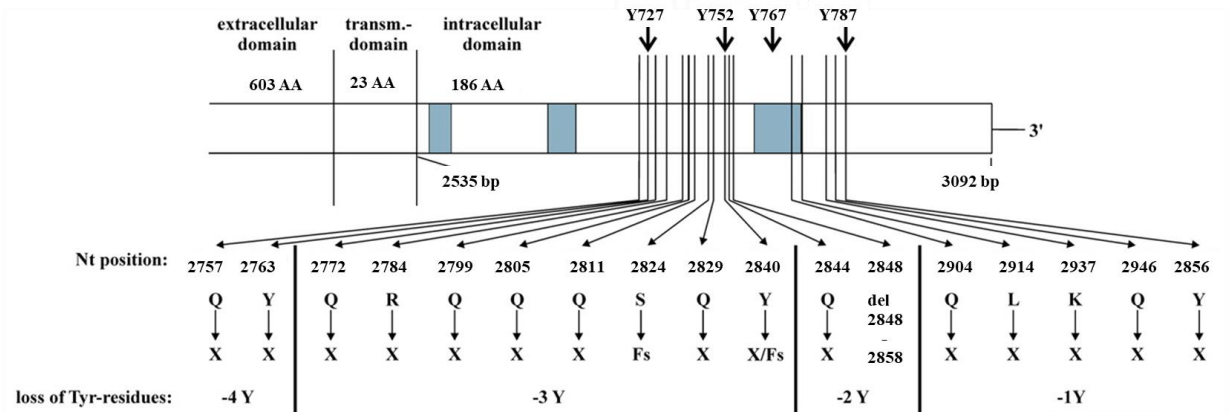


Figure 2. Frequently mutated sites in *CSF3R* in SCN patients that lead to development of MDS/AML (figure from Ref. [11]**).** These mutations were most frequently observed in SCN patients by Germeshausen et al. in 2007 (Blood). Most are nonsense mutations leading to a premature stop codon. Loss of tyrosine residues is observed in these patients as a result of these truncations.

3.

Figure 3. Localization of frequently observed mutations in *CSF3R* Exon 17. The positions of the frequently observed mutations (shown in **Figure 2**) found in SCN patients. The codons highlighted in red undergo truncation mutations. Asterix (*) is placed on the nucleotide that is replaced to get a mutated stop codon. At each site, indicate the mutation C-, T->. Other mutations were also observed. The frameshift mutations were seen in Serine (747 position). Deletion of ten amino acids, highlighted in blue, were also observed in one patient.

4.

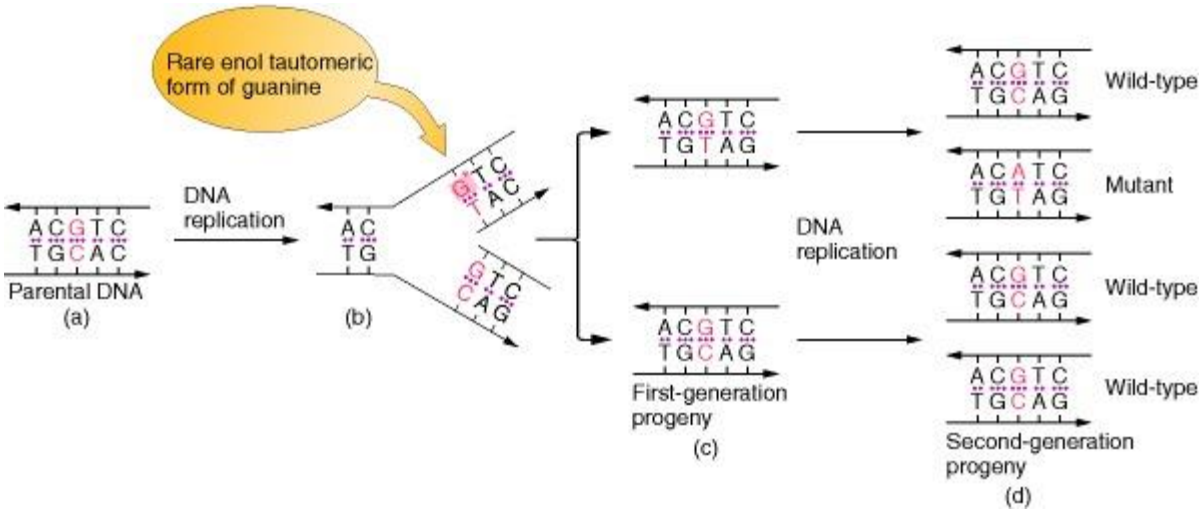


Figure 4. Mutation by tautomeric shifts in the bases of DNA[23]**.** (a) In the example diagrammed, a guanine undergoes a tautomeric shift to its rare enol form (G*) at the time of replication. (b) In its enol form, it pairs with thymine. (c) and (d) In the next replication, the guanine shifts back to its more stable keto form. The thymine incorporated opposite the enol form of guanine, seen in (b), directs the incorporation of adenine in the subsequent replication, shown in (c) and (d). The net result is a G·C → A·T mutation.

5.

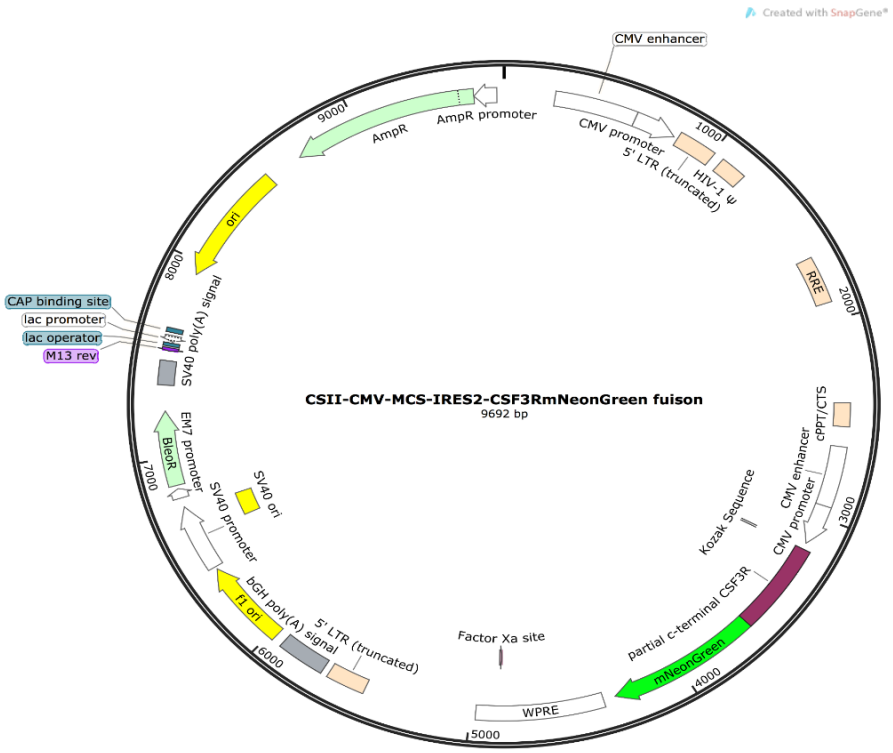


Figure 5. Plasmid map for transduced partial C-terminal of *CSF3R-GFP* construct. The plasmid map shows the transduced construct, which contained the partial C terminal of *CSF3R* (purple) fused with the GFP tag (green). The transduced construct is bleomycin and ampicillin resistant. It also contains the CMV promoter and enhancer region.

6.

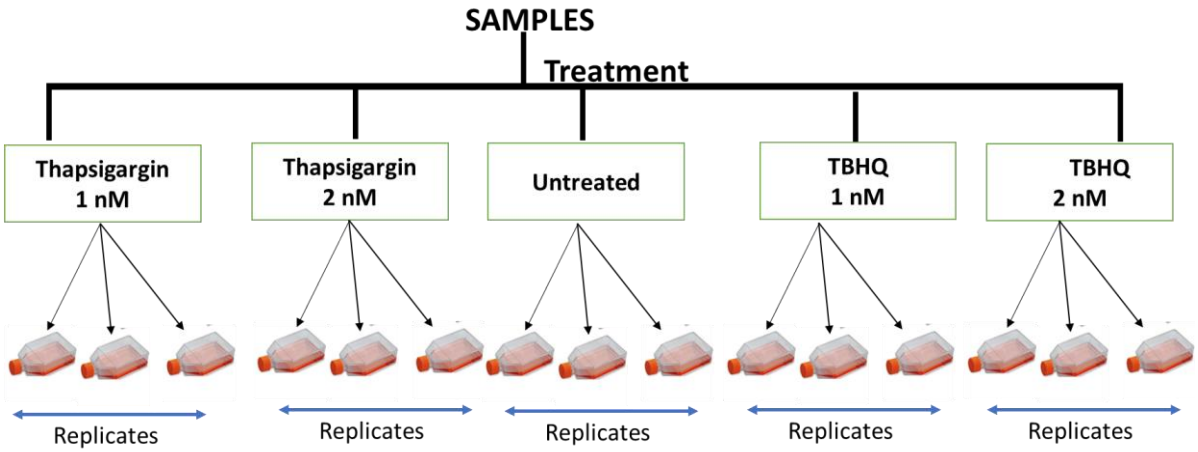


Figure 6. Experimental setup for 30-Day treatment with stress-inducing chemicals. Ba/F3 transduced cells were treated in triplicates for each treatment. Triplicates for untreated control cells were also harvested in identical conditions.

7.

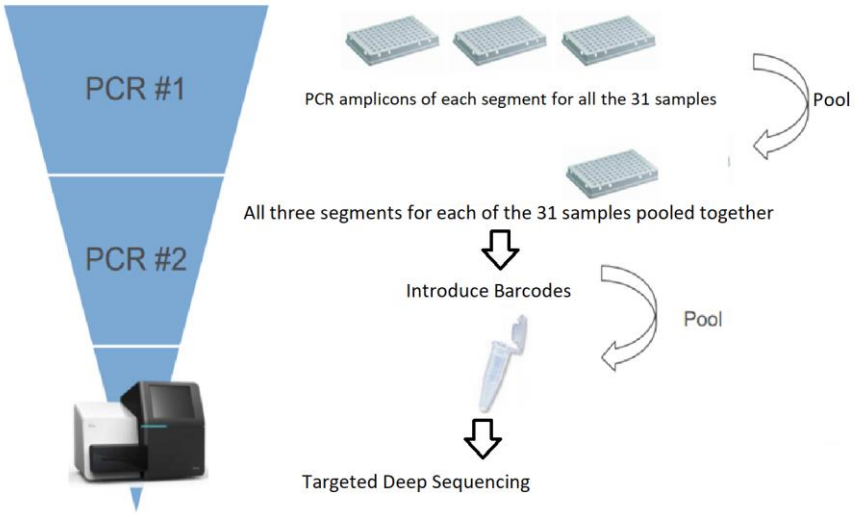


Figure 7. NGS was performed in three steps. The first step was to amplify the genomic DNA of each segment separately for all samples. This was followed by second round of amplification to introduce sequence adapters for segment identification. This was followed by the final sequencing in both forward and reverse directions.

8.

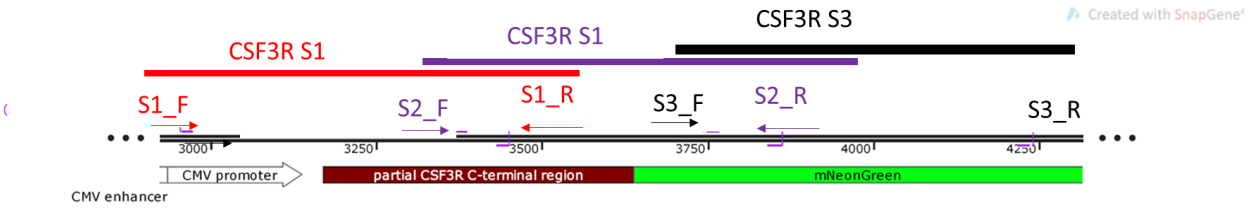


Figure 8. Primer design for the transduce *CSF3R-GFP* construct. The primers were designed to allow an overlap between each segment, thus increasing the efficiency of coverage of complete construct. To validate the primer sequences, PCR was carried out on all samples and the amplicons were confirmed using Sanger sequencing.

9.

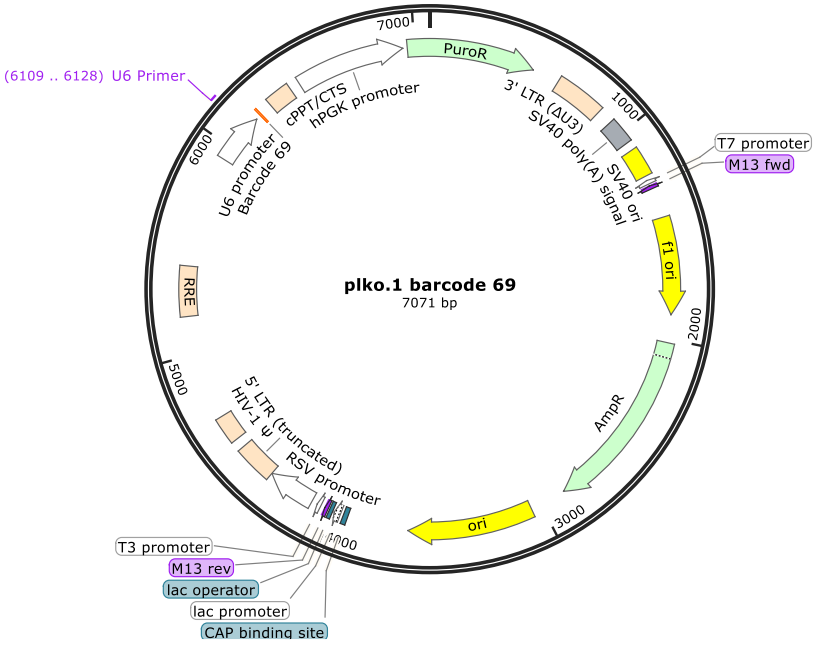


Figure 9. Plasmid map for Barcode 69 in PlkO.1 vector. Barcodes were transduced into Ba/F3 cells. Each barcode consisted of seven unique nucleotides, with common flanking sequence on both ends. The plasmid size for each of the barcodes was approximately 7000 bp in size. Each construct was puromycin and ampicillin resistant.

10.

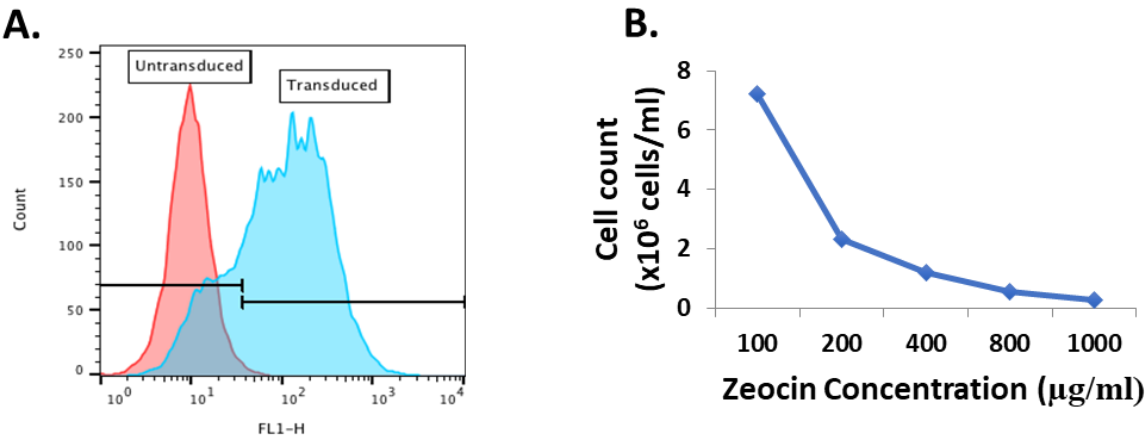


Figure 10. Transduction of Ba/F3 cells was confirmed using flow cytometry and selection was carried out in zeocin. A) Ba/F3 parent cells were transduced with partial C terminal of *CSF3R* fused with GFP positive construct. Flow cytometry analysis confirmed the presence of 2 distinct cell populations: GFP-negative untransduced parent cells (pink) and GFP-positive transduced cells (Blue). B) Zeocin was chosen as the selection antibiotic. The dose-response curve for zeocin showed 1000 µg/ml to be the LD100 dose which caused 100% cell death.

11.

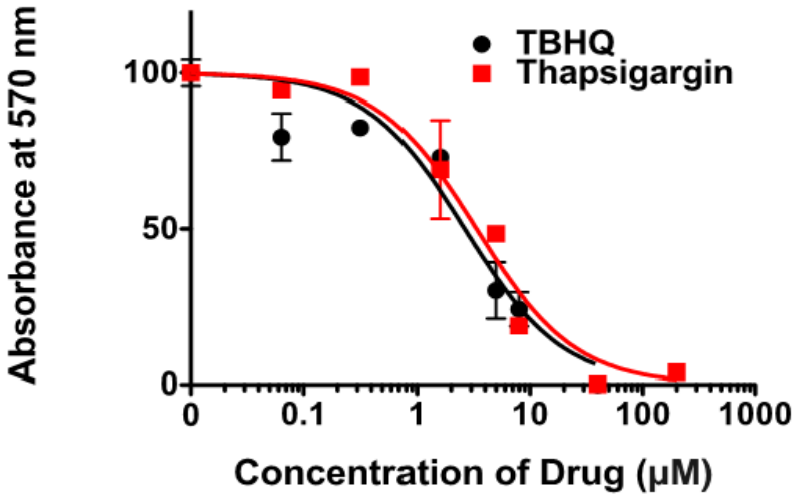


Figure 11. Based on the dose-response curves for both the chemicals, we selected n LD10 (1 nM) and LD20 (2 nM) as the stress inducing dosage. MTT assay was performed to calculate the lethality of each drug. Both Thapsigargin and TBHQ had similar IC50 values (4.26 nM and 4.09 nM respectively). Both chemicals had similar sublethal (LD10 and LD20) doses.

12.

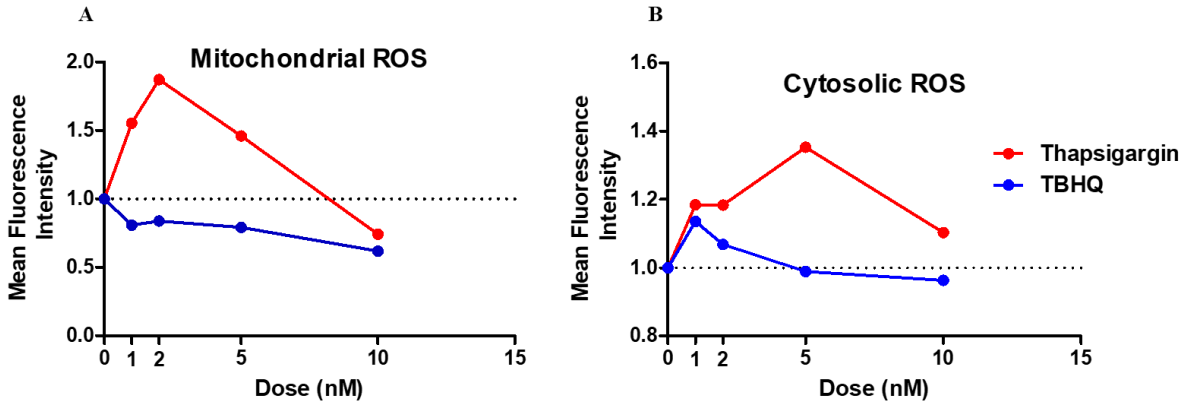
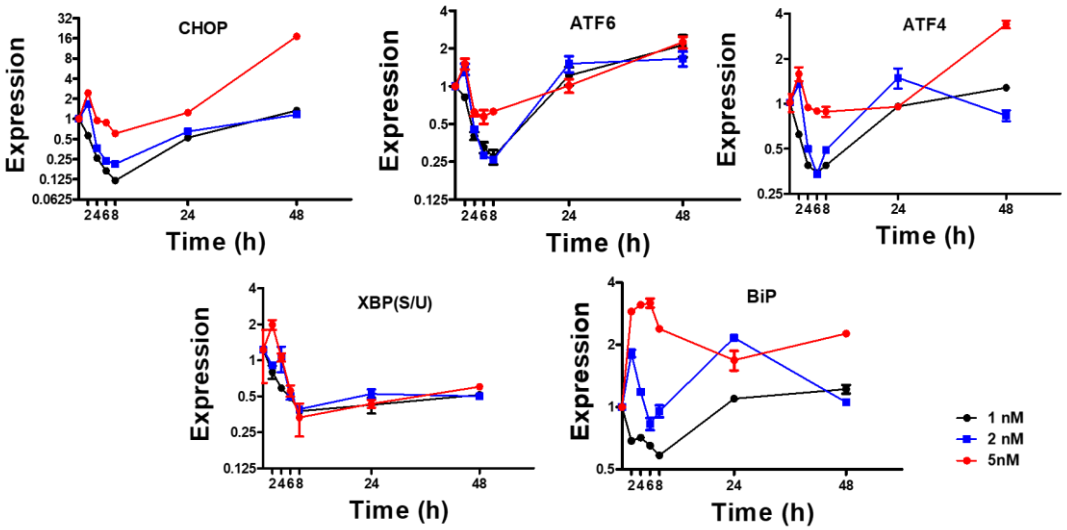


Figure 12. Dose-response for ROS production by TBHQ and thapsigargin. Mitochondrial and cytosolic fluorogenic probes (Mitosox and CellRox respectively) were used to measure the cellular oxidative stress using live cell imaging. Oxidative stress was induced more in the cytosol as compared to the mitochondria. Also, in general, there was an increase in the stress levels at both 1 nM and 2 nM doses. However, higher doses of 5 nM and 10 nM were much more toxic and lead to extensive cell death.

13.



-

Figure 13. Prolonged thapsigargin exposure leads to increased levels of UPR markers. UPR marker expression was measured by real-time PCR. Data is normalized to the expression of actin to get the dCt values. The dCt values were further normalized to the dCt value of day 0 control. All the UPR markers followed the same trend: decrease in the first 8 hours followed by a rapid increase. The experiment was repeated 3 times to confirm the findings.

14.

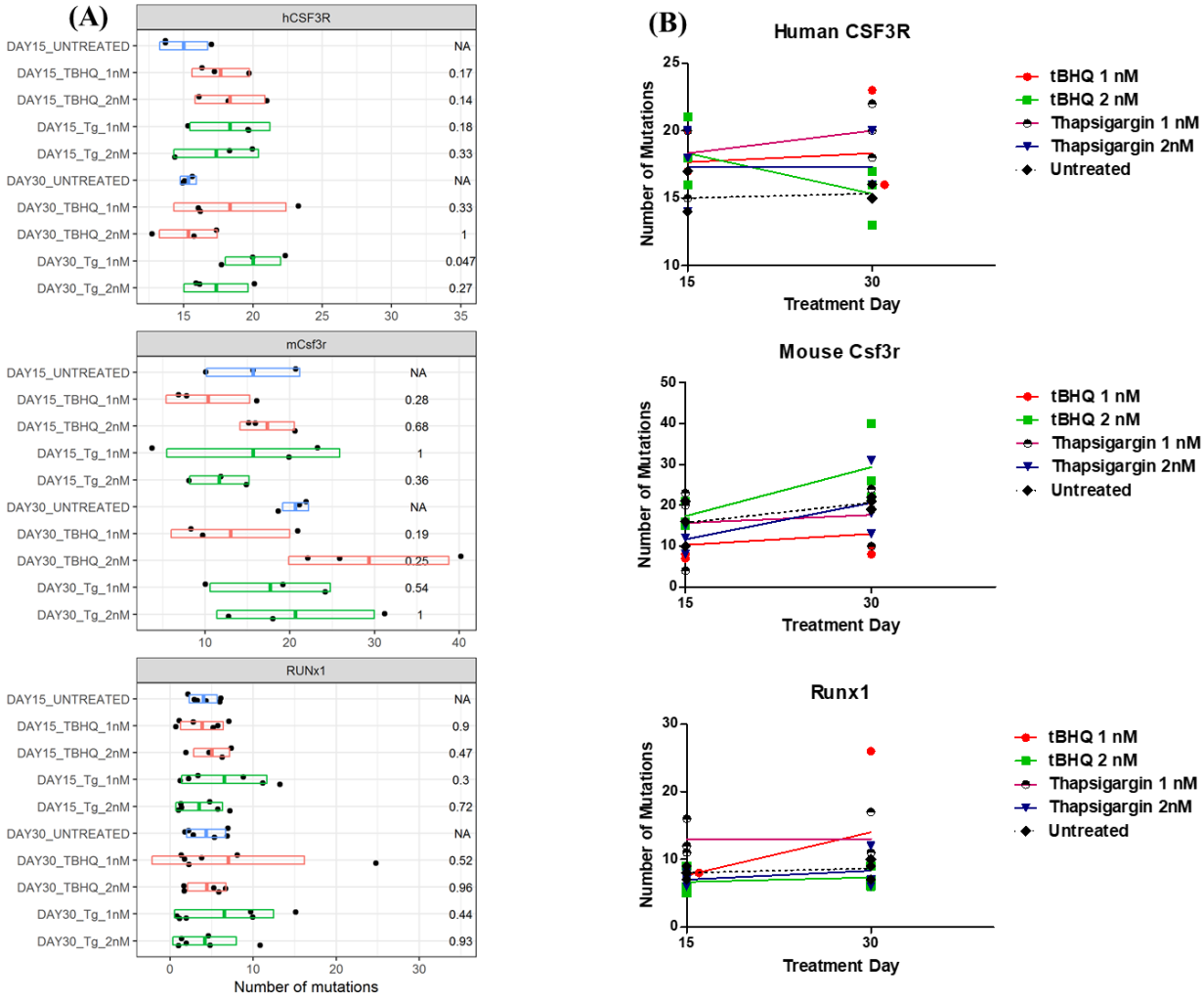


Figure 14. Accumulation of new mutations during the treatment period. Only *CSF3R* shows significant difference in the number of acquired mutations in treated and untreated samples. Although new mutations did arise during the treatment duration, however there was no correlation between the treatment and the total number of new mutations, at least in the endogenous systems. Mutations were observed in both treated and untreated samples for the three genes. In the exogenous CSF3R, the treated samples had more mutations than the untreated sample. However, in the endogenous genes, untreated samples had more mutations than treated samples. (A) The plots show the total number of mutations detected for each sample. Each point corresponds to one replicate, the bars represent mean +/-standard deviation (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 to the untreated samples from the same day. Each color marks different treatment types. (B) These plots are another graphical representation of the same date represented in Figure 6A.

15.

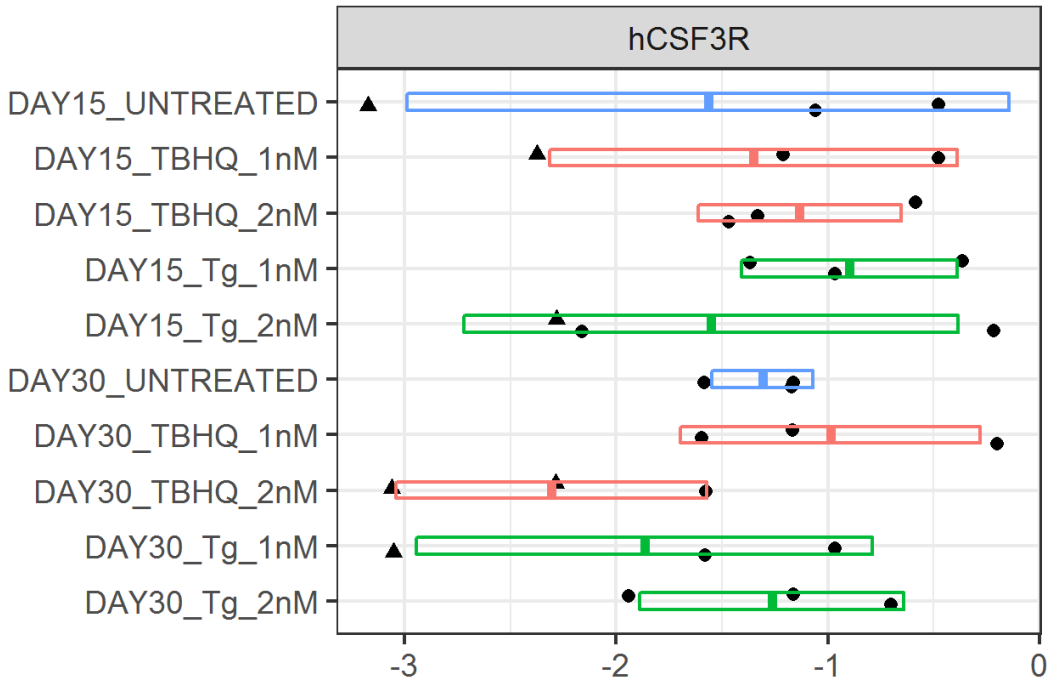


Figure 15. Log Fold Change of mutated sites (GFP vs CSF3R). The plot shows the LFC of the number of mutated positions relative to the region size for GFP vs *CSF3R*. The values below 0 indicate more mutations in the GFP region. The triangle shaped markers correspond to replicates in which the difference between the mutation proportions (standard test for 2 proportions) is significant (unadjusted p value<0.05).

16.

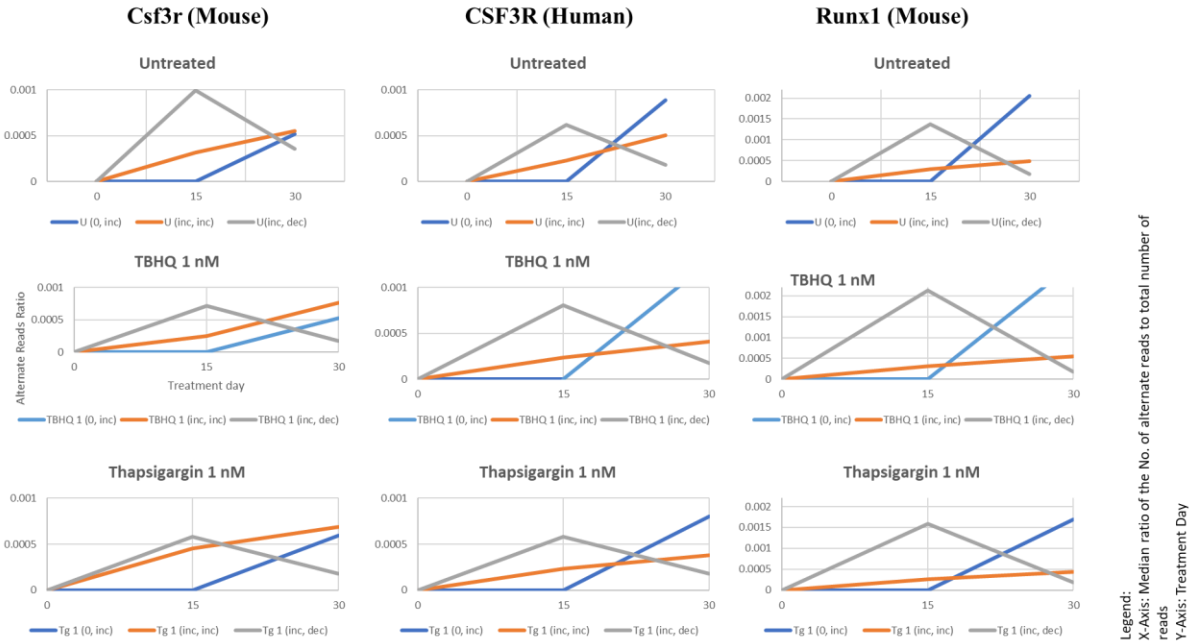


Figure 16. Three major trends were observed in the mutations that were consistent in all samples. The first set, represented by blue line, consisted of mutations that were absent until day 15 and arose only on day 30. The second set (red line) consisted of mutations whose frequency consistently grew during the 30-day period. The third set portrayed by grey line consisted of mutations whose frequency increased on day 15 followed by a decrease on day 30.

17.

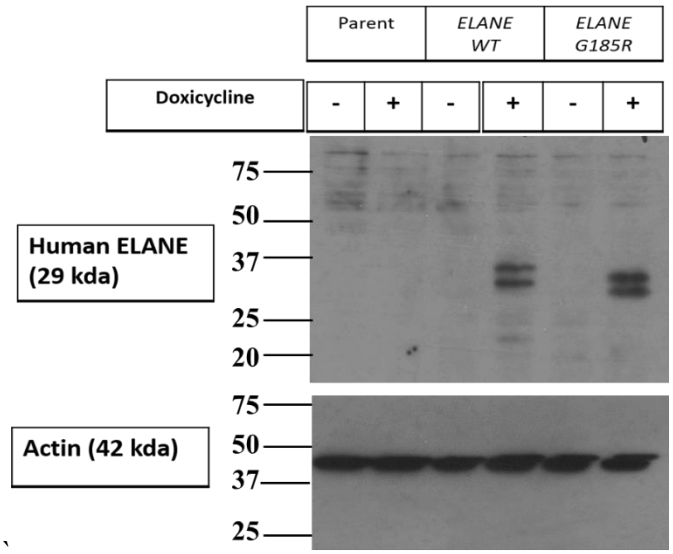


Figure 17. Transduction of *WT* and *G185R ELANE* in Ba/F3 cells was confirmed using Western Blotting. The cells were treated with doxycycline to induce the expression of ELANE. Negative control, i.e. the parent Ba/F3 cells showed no protein expression after doxycycline treatment. Human ELANE expression was observed at the expected size (29 kDa) in both transduced cell lines.

18.

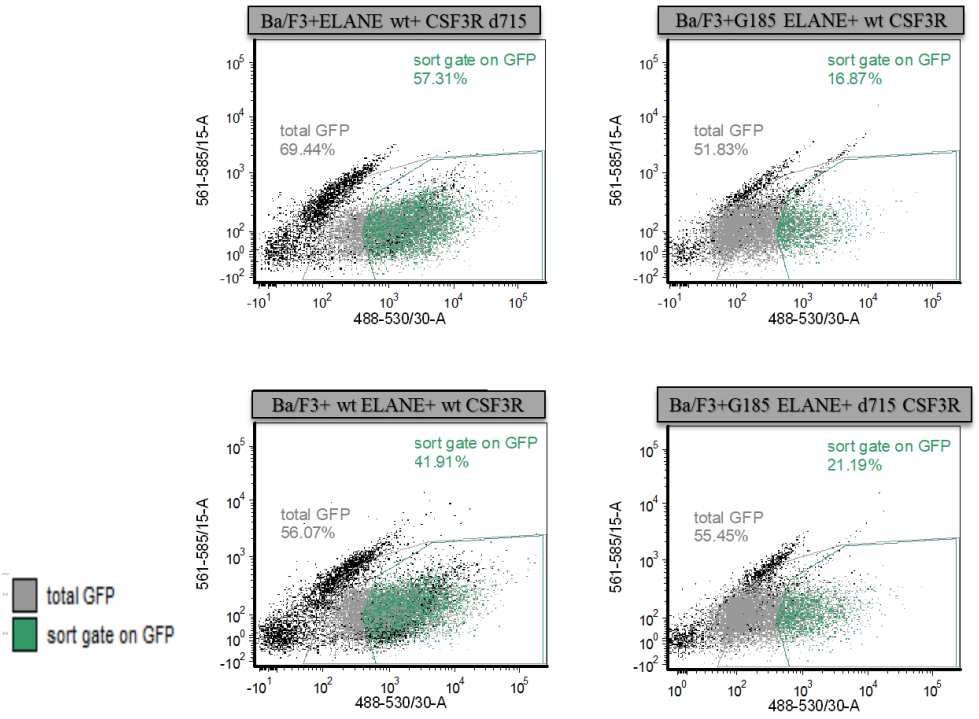


Figure 18. Flow cytometric confirmation of *CSF3R* transduction. The cell line population in black represents GFP negative cells and the population in grey represent GFP positive cells. The GFP positive cells that were selected by flow cytometric cell sorting are represented in green. The black population shows the GFP-negative cells, i.e. the untransduced Ba/F3 parent cells.

19.

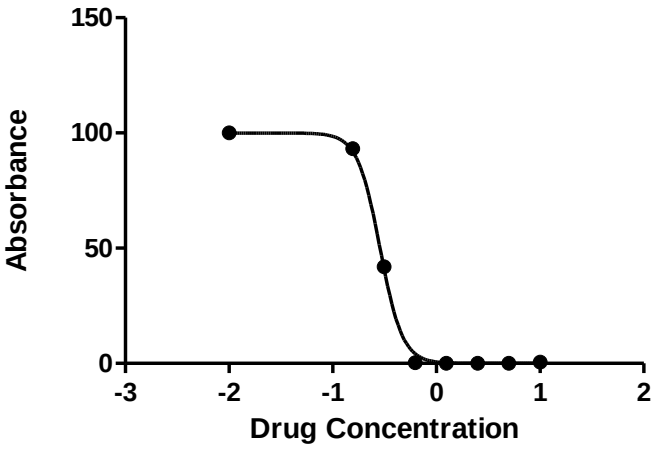


Figure 19. Puromycin dose-response curve for selection of cells transduced with lentivirus containing unique barcode. Lethal dose (LD100) was calculated to be 3 μg/ml. MTT assay was used to calculate the cell viability of cells treated with increasing doses of puromycin. 3 μg/ml dose of puromycin was observed to 100% cell death. The cells were then selected in 3 μg/ml concentration of puromycin for 10 days to eliminate the untransduced cells.

Table 1. The unique barcode sequences transduced in different Ba/F3 cell lines

BARCODE	SEQUENCE
68	CACCTCG
69	AGAAATG
76	GCCGTGC
81	GCCGCCG

Each barcode consists of seven unique nucleotide sequences, which are flanked by a common flanking region on both ends. These unique barcodes were transduced in different cell lines containing a unique combination of different *CSF3R* and *ELANE* mutations (Table 2). After the cell lines are mixed together and subjected to treatment, these unique barcode sequences will be used to identify the most proliferative cell line.

Table 2. Each cell line contains a unique barcode identifier and different combinations of mutations

BARCODE	ELANE	CSF3R
68	wt	d715
69	G185R	wt
76	wt	wt
81	G185R	d715

Ba/F3 cells were first transduced with wildtype and mutant *CSF3R*. The 2 cell lines thus produced, were then further transduced with wildtype and G185R *ELANE*, thus producing 4 cell lines with different combination of mutations, shown in this table. Each combination was then transduced with a unique barcode sequence, which will serve as an identifier.

CHAPTER 5

CONCLUSIONS

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%[33,34]. These clinical data strongly suggest that pharmacological administration of G-CSF for the treatment of SCN is correlated with the development of MDS/AML[5]. According to a study by Severe Chronic Neutropenia International Registry (SCNIR), 21% of SCN patients undergoing G-CSF treatment developed leukemia at one point in their therapy[35]. 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[36–38]. 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 bioinformatic analysis of the next generation sequencing data show that there is no significant increase in the number of mutations due to different cellular stress inducing treatments in Ba/F3 cells. No significant difference was observed between the number of mutations in the samples treated under cellular stress conditions and untreated controls. Surprisingly, the transduced GFP construct, having comparatively low GC content, was more highly mutated as compared to the GC-rich *CSF3R* region. Thus, these findings challenge our hypothesis that the partial C terminal of *CSF3R* is a hotspot for mutations under oxidative and ER stress. Although the published data are limited, this conclusion is supported by a recent clinical study that shows the mutational landscape of the hematopoietic stem cells in SCN patients is similar to the age-matched healthy patients[34].

We observed that the mutations could be classified into three disjoint groups: first group consisted of mutations whose frequency increased by day 15 followed by diminished frequency by

day 30 (increase, decrease). The second group consisted of mutations that continually increased during the 30-day treatment period (increase, increase). The third group of mutations was dormant until day 30 and emerged only on day 30 (dormant, increase). These three trends were consistent for all the samples. It was also observed that the first group of mutations (increase, decrease) are present at a higher frequency as compared to 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 slower rate does not cause cell death, thus giving the cell the time to acquire additional mutations that provide a clonal advantage to them.

This study strongly suggests the possibility of additional factors other than cellular stress induced mutagenesis in the transition from SCN to AML. The question that remains unanswered is why SCN patients with germline *ELANE* mutations are associated with acquired *CSF3R* mutations? This question can probably be answered by the second part of the study, which is focused on the clonal competition between different mutant *CSF3R* expressing cell lines. Our study continues. So far, we have generated cell lines with different combinations of *CSF3R* and *ELANE* mutations. These 4 cell lines will now be mixed together and be subjected to G-CSF treatment. The proliferation rate of each cell line will be measured by estimating the amount of unique barcode sequences using targeted deep sequencing. We expect that *CSF3R* mutations to work in cooperativity of *ELANE* mutations to provide a proliferative advantage to the cells. Once we get these results, we can further analyze the clonal competition between other *CSF3R* mutant expressing cell lines like d717, d725 and d730.

There are several limitations to the study. Most importantly, the cell line model system that we have chosen does not express wild-type or mutant *ELANE*. Although the cell line has its own advantage of rapid doubling time of 12-14 hours, thereby giving us approximately 45-55 generations during the treatment period of 30 days. Thus, for a more objective analysis, we should repeat the experiment

on a myeloid progenitor cell line instead of the lymphoid Ba/F3 cells. To confirm ROS production, ROS inducing mutagens like hydrogen peroxide will be used as a positive control. Because this study is limited to the analysis of mutational landscape in three genes, it is possible that stress induction is causing genomic instability in other genes. The initial cell genotoxicity will be measured by performing the HPRT-Forward-Mutation-Assay to study the nucleotide oxidation. The future experiments will be designed for broader sequencing analysis on these samples, either by performing whole-exome sequencing or targeted deep sequencing for a larger number of genes. We also aim to sequence at intermediate time intervals to study the trends that have been observed at day 15 and day 30. We particularly want to focus on day 10 samples, as until then, no selection was carried due to passaging and thus these sample would contain all the mutations that occurred since day 0. Current work seeks to measure the clonal competition between different mutant *CSF3R* and *ELANE* expressing cell lines. Since the cell lines have been prepared and single clones have been isolated, the next step in the study is to treat the cells with G-CSF treatment followed by NGS analysis to quantify which cell line proliferative advantage over others.

To summarize, our data suggests that *in vivo* cellular stress induction does not promote the genomic instability in Ba/F3 cells. We observed no increase in the mutagenesis in cells that were treated with cellular stress inducing chemicals as compared to untreated samples. Surprisingly, we observed that the GC-rich region in *CSF3R* had fewer genomic alterations as compared to the GFP construct with a considerably low GC content. This study also revealed that slowly acquiring mutations does not affect the cell viability, thus providing the time for persistence of a clone that provides proliferative advantage to the cells. Thus, our study is consistent with the conclusion of the study by Xia et al.[34], that both intrinsic and non-intrinsic cellular changes could be responsible for the development of MDS/AML in SCN patients. However, since *CSF3R* mutations are specific only to SCN associated leukemias, it is evident that these acquired mutations are associated with germline *ELANE* mutations in SCN. This project 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.

REFERENCES

1. Touw, I. P. Game of clones: the genomic evolution of severe congenital neutropenia. - PubMed - NCBI. 1–7 (2015).
2. Klein, C. Genetic Defects in Severe Congenital Neutropenia: Emerging Insights into Life and Death of Human Neutrophil Granulocytes. *Annu. Rev. Immunol.* **29,** 399–413 (2011).
3. Klein, C. Congenital neutropenia. *Hematology* 344–350 (2009). doi:10.1182/asheducation-2009.1.344
4. Beekman, R. *et al.* Sequential gain of mutations in severe congenital neutropenia progressing to acute myeloid leukemia. *Blood* **119,** 5071–5077 (2012).
5. Skokova, J. et al. Cooperativity of RUNX1 and CSF3R mutations in severe congenital neutropenia: a unique pathway in myeloid leukemogenesis. **123,** 2229–2238 (2014).
6. Germeshausen, M. *et al.* The Spectrum of ELANE Mutations and their Implications in Severe Congenital and Cyclic Neutropenia. *Hum. Mutat.* **34,** 905–914 (2013).
7. Triot, A. *et al.* Inherited biallelic CSF3R mutations in severe congenital neutropenia. *Blood* **123,** 3811–3818 (2015).
8. Boxer, L. No Title Role of neutrophils in genetic disorders of phagocyte function leading to IBD. *J Pediatr Gastroenterol Nutr* **46,** Suppl 1:E17 (2008).
9. Walter, P. & Ron, D. The Unfolded Protein Response : *Science (80-.).* **334,** 1081–1086 (2012).
10. Bonilla, M. A., Gillio, A. P. & Kernan, N. et al. Effects of recombinant human granulocyte colony-stimulating factor on neutropenia in patients with congenital agranulocytosis. *N. Engl. J. Med.* **317,** 185–191 (1987).
11. Germeshausen, M., Ballmaier, M., Welte, K. & Dc, W. Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis : results of a long-term survey Incidence of CSF3R mutations in severe congenital neutropenia and relevance for leukemogenesis : results of a long-term survey. **109,** 93–99 (2012).
12. Dong, F. *et al.* Identification of a nonsense mutation in the granulocyte-colony-stimulating factor receptor in severe congenital neutropenia. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 4480–4 (1994).
13. Tidow, N. *et al.* 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 (1997).
14. Dong, F. et al. Receptor in Patients With Acute Myeloid Leukemia Preceded By Severe Congenital Neutropenia. **333,** (2013).
15. Rogozin, I. B. & Pavlov, Y. I. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* **544,** 65–85 (2003).
16. Ripley, L. S. & Glickman, B. W. Unique Self-complementarity of Palindromic Sequences Provides DNA Structural Intermediates for Mutation. *Cold Spring Harb Symp Quant Biol* (1980).

17. Salganik, R. I. & Dianov, G. L. Cluster of point mutations predetermined by a quasipalindromic nucleotide sequence in plasmid pBR322 DNA. **261,** 28–30 (1990).

18. Gordenin, D. A. & Resnick, M. A. Congenital agranulocytosis: prolonged survival and terminal acute leukemia. (1998).

19. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology. 4th edition.* (2000).

20. Maki, H. O RIGINS OF S PONTANEOUS M UTATIONS : Specificity and Directionality of. *Annu. Rev. Genet* 279–303 (2002). doi:10.1146/annurev.genet.36.042602.094806

21. Drake, J. W. The biochemistry of mutagenesis. *Annu. Rev. Biochem* (1976).

22. Griffiths AJF, Gelbart WM, Miller JH, et al. Modern Genetic Analysis. New York: W. H. Freeman; in (1999).

23. Gardner, E. J. & Snustad, D. P. *Principles of Genetics, 5th ed, by John Wiley & Sons, New York.* (1984).

24. Pfeifer, G. P. DNA Methylation and Mutation Mutations at Methylated CpG. *eLS* 1–5 (2017). doi:10.1002/9780470015902.a0006159.pub2

25. Baltz, R. H. & Drake, J. W. The biochemistry of mut agenesis. *Annu. Rev. Biochem* **45,** 11–37 (1976).

26. Benzer, B. Y. S. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. U.S.A.* **47,** 403–415 (1961).

27. Markiewicz, P., Kleina, L. G., Miller, J., Kisters-woike, B. & Mu, B. Genetic Studies of the Lac Repressor XV †: 4000 Single Amino Acid Substitutions and Analysis of the Resulting Phenotypes on the Basis of the Protein Structure. *J. Mol. Biol.* **261,** 509–523 (1996).

28. Walker, D. R. *et al.* Evolutionary conservation and somatic mutation hotspot maps of p53 : correlation with p53 protein structural and functional features. *Oncogene* **19,** 211–218 (1999).

29. Demarini, D. M., Shelton, M. L., Abu-shakra, A., Szakmary, A. & Levine, J. G. Spectra of Spontaneous Frameshift Mutations at the hisD3052 Allele of Salmonella typhimurium in Four DNA Repair Backgrounds. *Genetics* 17–36 (1998).

30. Rogozin, I. B., Kondrashov, F. A. & Galina, V. Use of Mutation Spectra Analysis Software. *Hum. Mutat.* **102,** 83–102 (2001).

31. Zavolan, M. & Kepler, T. B. Statistical inference of sequence-dependent mutation rates. *Curr. Opin. Genet. Dev* **11,** 612–615 (2001).

32. Mehta, H., Futami, M., Glaubach, T. & Lee, D. W. et al. Alternatively spliced, truncated GCSF receptor promotes leukemogenic properties and sensitivity to JAK inhibition HM. *Leukemia.* **28,** 1041–1051 (2014).

33. Rosenberg, P. S. *et al.* Stable long-term risk of leukaemia in patients with severe congenital neutropenia maintained on G-CSF therapy. *Br. J. Haematol.* **150,** no-no (2010).

34. Xia, J. *et al.* Somatic mutations and clonal hematopoiesis in congenital neutropenia. *Blood* **131,** 408–416 (2018).

35. Mehta, H. M., Malandra, M. & Corey, S. J. HHS Public Access. **195,** 1341–1349 (2016).

36. Wong, W. Y. *et al.* Terminal acute myelogenous leukemia in a patient with congenital agranulocytosis. *Am. J. Hematol.* **43,** 133–138 (1993).

37. Gilman, P., Jackson, D. & Guild, H. Congenital agranulocytosis: prolonged survival and terminal acute leukemia. *Blood* **36,** 85 (1970).

38. Donadieu, J. *et al.* Analysis of risk factors for myelodysplasias, leukemias and death from infection among patients with congenital neutropenia. **90,** 45–53 (2005).

APPENDIX

Table 3. Number of mutations for each sample, for all 3 genes

Group	Sample	Day	Treatment	Replicate	No. of Mutations
***CSF3R* (Human)**	DAY15_TBHQ_1nM_A	15	TBHQ_1nM	A	17
	DAY15_TBHQ_1nM_B	15	TBHQ_1nM	B	16
	DAY15_TBHQ_1nM_C	15	TBHQ_1nM	C	20
	DAY15_TBHQ_2nM_A	15	TBHQ_2nM	A	16
	DAY15_TBHQ_2nM_B	15	TBHQ_2nM	B	18
	DAY15_TBHQ_2nM_C	15	TBHQ_2nM	C	21
	DAY15_Tg_1nM_A	15	Tg_1nM	A	15
	DAY15_Tg_1nM_B	15	Tg_1nM	B	20
	DAY15_Tg_1nM_C	15	Tg_1nM	C	20
	DAY15_Tg_2nM_A	15	Tg_2nM	A	20
	DAY15_Tg_2nM_B	15	Tg_2nM	B	18
	DAY15_Tg_2nM_C	15	Tg_2nM	C	14
	DAY15_UNTREATED_A	15	UNTREATED	A	14
	DAY15_UNTREATED_B	15	UNTREATED	B	14
	DAY15_UNTREATED_C	15	UNTREATED	C	17
	DAY30_TBHQ_1nM_A	30	TBHQ_1nM	A	16
	DAY30_TBHQ_1nM_B	30	TBHQ_1nM	B	23
	DAY30_TBHQ_1nM_C	30	TBHQ_1nM	C	16
	DAY30_TBHQ_2nM_A	30	TBHQ_2nM	A	13
	DAY30_TBHQ_2nM_B	30	TBHQ_2nM	B	16
	DAY30_TBHQ_2nM_C	30	TBHQ_2nM	C	17
	DAY30_Tg_1nM_A	30	Tg_1nM	A	20
	DAY30_Tg_1nM_B	30	Tg_1nM	B	22
	DAY30_Tg_1nM_C	30	Tg_1nM	C	18
	DAY30_Tg_2nM_A	30	Tg_2nM	A	16
	DAY30_Tg_2nM_B	30	Tg_2nM	B	20
	DAY30_Tg_2nM_C	30	Tg_2nM	C	16
	DAY30_UNTREATED_A	30	UNTREATED	A	15
	DAY30_UNTREATED_B	30	UNTREATED	B	15
	DAY30_UNTREATED_C	30	UNTREATED	C	16
***Csf3r* (Mouse)**	DAY15_TBHQ_1nM_A	15	TBHQ_1nM	A	8
	DAY15_TBHQ_1nM_B2	15	TBHQ_1nM	B2	7
	DAY15_TBHQ_1nM_C	15	TBHQ_1nM	C	16
	DAY15_TBHQ_2nM_A	15	TBHQ_2nM	A	15
	DAY15_TBHQ_2nM_B	15	TBHQ_2nM	B	21
	DAY15_TBHQ_2nM_C	15	TBHQ_2nM	C	16
	DAY15_Tg_1nM_A	15	Tg_1nM	A	20
	DAY15_Tg_1nM_B2	15	Tg_1nM	B2	4

	DAY15_Tg_1nM_C	15	Tg_1nM	C	23
	DAY15_Tg_2nM_A2	15	Tg_2nM	A2	8
	DAY15_Tg_2nM_B	15	Tg_2nM	B	15
	DAY15_Tg_2nM_C2	15	Tg_2nM	C2	12
	DAY15_UNTREATED_A	15	UNTREATED	A	16
	DAY15_UNTREATED_B	15	UNTREATED	B	10
	DAY15_UNTREATED_C	15	UNTREATED	C	21
	DAY30_TBHQ_1nM_A	30	TBHQ_1nM	A	8
	DAY30_TBHQ_1nM_B	30	TBHQ_1nM	B	10
	DAY30_TBHQ_1nM_C	30	TBHQ_1nM	C	21
	DAY30_TBHQ_2nM_A	30	TBHQ_2nM	A	40
	DAY30_TBHQ_2nM_B	30	TBHQ_2nM	B	26
	DAY30_TBHQ_2nM_C	30	TBHQ_2nM	C	22
	DAY30_Tg_1nM_A	30	Tg_1nM	A	10
	DAY30_Tg_1nM_B	30	Tg_1nM	B	19
	DAY30_Tg_1nM_C	30	Tg_1nM	C	24
	DAY30_Tg_2nM_A	30	Tg_2nM	A	31
	DAY30_Tg_2nM_B	30	Tg_2nM	B	18
	DAY30_Tg_2nM_C	30	Tg_2nM	C	13
	DAY30_UNTREATED_A	30	UNTREATED	A	19
	DAY30_UNTREATED_B	30	UNTREATED	B	21
	DAY30_UNTREATED_C	30	UNTREATED	C	22
Runx1 (Mouse)	DAY15_TBHQ_1nM_A	15	TBHQ_1nM	A	3
	DAY15_TBHQ_1nM_A	15	TBHQ_1nM	A	5
	DAY15_TBHQ_1nM_B	15	TBHQ_1nM	B	1
	DAY15_TBHQ_1nM_B	15	TBHQ_1nM	B	7
	DAY15_TBHQ_1nM_C	15	TBHQ_1nM	C	1
	DAY15_TBHQ_1nM_C	15	TBHQ_1nM	C	6
	DAY15_TBHQ_2nM_A	15	TBHQ_2nM	A	5
	DAY15_TBHQ_2nM_B	15	TBHQ_2nM	B	6
	DAY15_TBHQ_2nM_C	15	TBHQ_2nM	C	2
	DAY15_TBHQ_2nM_C	15	TBHQ_2nM	C	7
	DAY15_Tg_1nM_A	15	Tg_1nM	A	2
	DAY15_Tg_1nM_A	15	Tg_1nM	A	9
	DAY15_Tg_1nM_B	15	Tg_1nM	B	3
	DAY15_Tg_1nM_B	15	Tg_1nM	B	13
	DAY15_Tg_1nM_C	15	Tg_1nM	C	1
	DAY15_Tg_1nM_C	15	Tg_1nM	C	11
	DAY15_Tg_2nM_A	15	Tg_2nM	A	1
	DAY15_Tg_2nM_A	15	Tg_2nM	A	7
	DAY15_Tg_2nM_B	15	Tg_2nM	B	1
	DAY15_Tg_2nM_B	15	Tg_2nM	B	6
	DAY15_Tg_2nM_C	15	Tg_2nM	C	1
	DAY15_Tg_2nM_C	15	Tg_2nM	C	5
	DAY15_UNTREATED_A	15	UNTREATED	A	3

	DAY15_UNTREATED_A	15	UNTREATED	A	6
	DAY15_UNTREATED_B	15	UNTREATED	B	3
	DAY15_UNTREATED_B	15	UNTREATED	B	4
	DAY15_UNTREATED_C	15	UNTREATED	C	2
	DAY15_UNTREATED_C	15	UNTREATED	C	6
	DAY30_TBHQ_1nM_A	30	TBHQ_1nM	A	2
	DAY30_TBHQ_1nM_A	30	TBHQ_1nM	A	4
	DAY30_TBHQ_1nM_B	30	TBHQ_1nM	B	2
	DAY30_TBHQ_1nM_B	30	TBHQ_1nM	B	8
	DAY30_TBHQ_1nM_C	30	TBHQ_1nM	C	1
	DAY30_TBHQ_1nM_C	30	TBHQ_1nM	C	25
	DAY30_TBHQ_2nM_A	30	TBHQ_2nM	A	2
	DAY30_TBHQ_2nM_A	30	TBHQ_2nM	A	7
	DAY30_TBHQ_2nM_B	30	TBHQ_2nM	B	6
	DAY30_TBHQ_2nM_C	30	TBHQ_2nM	C	2
	DAY30_TBHQ_2nM_C	30	TBHQ_2nM	C	5
	DAY30_Tg_1nM_A	30	Tg_1nM	A	1
	DAY30_Tg_1nM_A	30	Tg_1nM	A	10
	DAY30_Tg_1nM_B	30	Tg_1nM	B	2
	DAY30_Tg_1nM_B	30	Tg_1nM	B	15
	DAY30_Tg_1nM_C	30	Tg_1nM	C	1
	DAY30_Tg_1nM_C	30	Tg_1nM	C	10
	DAY30_Tg_2nM_A	30	Tg_2nM	A	1
	DAY30_Tg_2nM_A	30	Tg_2nM	A	11
	DAY30_Tg_2nM_B	30	Tg_2nM	B	1
	DAY30_Tg_2nM_B	30	Tg_2nM	B	5
	DAY30_Tg_2nM_C	30	Tg_2nM	C	2
	DAY30_Tg_2nM_C	30	Tg_2nM	C	5
	DAY30_UNTREATED_A	30	UNTREATED	A	2
	DAY30_UNTREATED_A	30	UNTREATED	A	7
	DAY30_UNTREATED_B	30	UNTREATED	B	2
	DAY30_UNTREATED_B	30	UNTREATED	B	5
	DAY30_UNTREATED_C	30	UNTREATED	C	3
	DAY30_UNTREATED_C	30	UNTREATED	C	7

Table 4. Log Fold Change (LFC) between number of mutated positions in *CSF3R* vs *GFP*

Column 1 gives all the samples being tested. nMut_CSF3R column contains the number of mutations in the CSF3R region for each sample. Similarly, nMut_GFP contains the number of mutations in the GFP region, for each sample. Prop_CSF3R and Prop_GFP columns contain the number of mutations with respect to the size (NO. of mutations/No of base pairs) of CSF3R and GFP regions, respectively. The unadjusted p-values are provided in the pval column, followed by the adjusted P values in padj column.

Sample	Day	nMut_CSF3R	nMut_GFP	Prop_CSF3R	Prop_GFP	PropLFC	pval	padj
DAY15_TBHQ_1nM_A	15	2	15	0.004673	0.024272	-2.376895	0.026682	0.190595
DAY15_TBHQ_1nM_B	15	3	10	0.007009	0.016181	-1.206970	0.301753	0.411481
DAY15_TBHQ_1nM_C	15	6	12	0.014019	0.019417	-0.470004	0.675673	0.750748
DAY15_TBHQ_2nM_A	15	4	16	0.009346	0.025890	-1.470004	0.090750	0.263796
DAY15_TBHQ_2nM_B	15	6	13	0.014019	0.021036	-0.585481	0.548445	0.658134
DAY15_TBHQ_2nM_C	15	3	11	0.007009	0.017799	-1.344473	0.222648	0.351549
DAY15_Tg_1nM_A	15	4	15	0.009346	0.024272	-1.376895	0.123105	0.263796
DAY15_Tg_1nM_B	15	7	13	0.016355	0.021036	-0.363089	0.753610	0.807440
DAY15_Tg_1nM_C	15	5	14	0.011682	0.022654	-0.955431	0.284178	0.405968
DAY15_Tg_2nM_A	15	2	13	0.004673	0.021036	-2.170444	0.054341	0.232890
DAY15_Tg_2nM_B	15	2	14	0.004673	0.022654	-2.277359	0.038119	0.190595
DAY15_Tg_2nM_C	15	6	10	0.014019	0.016181	-0.206970	0.980850	0.980850
DAY15_UNTREATED_A	15	1	13	0.002336	0.021036	-3.170444	0.020668	0.190595
DAY15_UNTREATED_B	15	3	9	0.007009	0.014563	-1.054966	0.405011	0.528276
DAY15_UNTREATED_C	15	6	12	0.014019	0.019417	-0.470004	0.675673	0.750748
DAY30_TBHQ_1nM_A	30	4	13	0.009346	0.021036	-1.170444	0.221909	0.351549

DAY30_TBHQ_1nM_B	30	3	13	0.007009	0.021036	-1.585481	0.118481	0.263796
DAY30_TBHQ_1nM_C	30	6	10	0.014019	0.016181	-0.206970	0.980850	0.980850
DAY30_TBHQ_2nM_A	30	3	13	0.007009	0.021036	-1.585481	0.118481	0.263796
DAY30_TBHQ_2nM_B	30	1	12	0.002336	0.019417	-3.054966	0.030164	0.190595
DAY30_TBHQ_2nM_C	30	2	14	0.004673	0.022654	-2.277359	0.038119	0.190595
DAY30_Tg_1nM_A	30	1	12	0.002336	0.019417	-3.054966	0.030164	0.190595
DAY30_Tg_1nM_B	30	3	13	0.007009	0.021036	-1.585481	0.118481	0.263796
DAY30_Tg_1nM_C	30	5	14	0.011682	0.022654	-0.955431	0.284178	0.405968
DAY30_Tg_2nM_A	30	2	11	0.004673	0.017799	-1.929436	0.109531	0.263796
DAY30_Tg_2nM_B	30	6	14	0.014019	0.022654	-0.692396	0.439479	0.549349
DAY30_Tg_2nM_C	30	4	13	0.009346	0.021036	-1.170444	0.221909	0.351549
DAY30_UNTREATED_A	30	4	13	0.009346	0.021036	-1.170444	0.221909	0.351549
DAY30_UNTREATED_B	30	4	13	0.009346	0.021036	-1.170444	0.221909	0.351549
DAY30_UNTREATED_C	30	3	13	0.007009	0.021036	-1.585481	0.118481	0.263796

Table 5. Median of ratio of altered reads over sequencing coverage

For each gene, the date is divided into further subsets, according to treatment type. 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. These median values were then plotted in Figure 16.

A. **Human *CSf3R***

Human *CSF3R*			
Untreated			
Day	Untreated (0, inc)	Untreated (inc, inc)	Untreated (inc, dec)
0	0	0	0
15	0	0.000231289	0.000615185
30	0.000887725	0.000500624	0.00017998
TBHQ 1 nM			
	TBHQ 1 (0, inc)	TBHQ 1 (inc, inc)	TBHQ 1 (inc, dec)
0	0	0	0
15	0	0.000233828	0.000807103
30	0.001277289	0.000412307	0.00017998
TBHQ 2 nM			
	TBHQ 2 (0, inc)	TBHQ 2 (inc, inc)	TBHQ 2 (inc, dec)
0	0	0	0
15	0	0.000259646	0.000669902
30	0.000687286	0.000444045	0.00017998
Thapsigargin 1 nM			
	Tg 1 (0, inc)	Tg 1 (inc, inc)	Tg 1 (inc, dec)
0	0	0	0
15	0	0.000234742	0.000577848
30	0.00079911	0.000382576	0.00017998
Thapsigargin 2 nM			

	Tg 2 (0, inc)	Tg 2 (inc, inc)	Tg 2 (inc, dec)
0	0	0	0
15	0	0.000260518	0.000521002
30	0.001185711	0.000419425	0.00017998

B. Mouse Csf3r

Mouse *Csf3r*			
Untreated			
Day	Untreated (0, inc)	Untreated (inc, inc)	Untreated(inc, dec)
0	0	0	0
15	0	0.000313599	0.000995043
30	0.000518643	0.000551422	0.000358031
TBHQ 1 nM			
	TBHQ 1 (0, inc)	TBHQ 1 (inc, inc)	TBHQ 1 (inc, dec)
0	0	0	0
15	0	0.000253049	0.000719266
30	0.000526455	0.000762185	0.000173521
TBHQ 2 nM			
	TBHQ 2 (0, inc)	TBHQ 2 (inc, inc)	TBHQ 2 (inc, dec)
0	0	0	0
15	0	0.000276113	0.000673067
30	0.00048037	0.000593368	0.000261271
Thapsigargin 1 nM			
	Tg 1 (0, inc)	Tg 1 (inc, inc)	Tg 1 (inc, dec)
0	0	0	0
15	0	0.000451671	0.000582751
30	0.000593884	0.000685648	0.00017998
Thapsigargin 2 nM			
	Tg 2 (0, inc)	Tg 2 (inc, inc)	Tg 2 (inc, dec)
0	0	0	0
15	0	0.000261467	0.001691814
30	0.000626228	0.000400772	0.00017998

C. Runx1

Runx1			
Day	U (0, inc)	U (inc, inc)	U(inc, dec)
0	0	0	0
15	0	2.22E-04	6.37E-04
30	9.78E-04	3.81E-04	8.48E-05
Day	TBHQ 1 (0, inc)	TBHQ 1 (inc, inc)	TBHQ 1 (inc, dec)
0	0	0	0
15	0	2.36E-04	9.36E-04
30	1.06E-03	4.68E-04	4.78E-05
Day	TBHQ 2 (0, inc)	TBHQ 2 (inc, inc)	TBHQ 2 (inc, dec)
0	0	0	0
15	0	0.000212619	0.000961766
30	0.000544782	0.000348164	0
Day	Tg 1 (0, inc)	Tg 1 (inc, inc)	Tg 1 (inc, dec)
0	0	0	0
15	0	2.22E-04	8.82E-04
30	1.04E-03	4.22E-04	4.29E-05
Day	Tg 2 (0, inc)	Tg 2 (inc, inc)	Tg 2 (inc, dec)
0	0	0	0
15	0	2.41E-04	6.99E-04
30	7.96E-04	4.11E-04	6.23E-05

Table 6. 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.

Genotype	Segment	Forward Primer (Sense)	Reverse Primer (Anti-sense)	Annealing Temperature
CSF3R (Human)	1	ACATCAATGGGCGTGGATAG	TGGAGTCACAGCGGAGATA	66ºC
	2	CCAGCGATCAGGTCCTTTATG	CAGGTACTGATGGAAGCCATAC	64ºC
	3	CGGCAATCCAAATGATGGTTAT	CGGCTGGTTCTTCAGATAGTT	62ºC
Csf3r (Mouse)	1	CCAGGCAAGTATGTCCCTTT	GGGTCTTGAATGGAACCAGAT	64ºC
	2	CTCTATGGTCAGGTGCTTGAG	TTGGGCTGTAAGCAGTATCTATC	64ºC
	3	CTTGAAGACCCTGCCCTATTC	GGCCTCGAACTCAGAAATCTAC	72ºC
	4	TGGTCACCATTCGTCTTTCATA	GAGTCCACTTCTCAAGATCCAC	71.3ºC
Runx1 (Mouse)	1	GGCAACTAACTGCTGGAACT	GCTCGTGCTGGCATCTAC	63ºC
	2	CTCCGGTAGTAATAAAGGCTTCTG	GACCTCCACCTGCTCCT	66.4ºC
	3	GACGGACGGCTCTTCTATAAAC	ACGTGATAATATCCAGTTGTTAGGT	62.4ºC
	4	TAGGTGTACCAGGAGAACAAG	GAGAGGTGAGAGAGTTGACC	61.9ºC
	5	GGGTGAAGGGAGAGGGA	CCCTTAGAGCACCCAGAAAG	72ºC
	6	ATACTTGGGCTCTAGTGGT	CCTGTGTCTTTCTGTGTCTC	61.5ºC
	7	CACACACTCATGCCTTCTCTT	CATGTTGGTGGGCGAGTT	72ºC
	8	CTGCACCAACGCATCCA	GCAATCGATAAGGTGCGGAAA	65ºC

VITA

Adya Sapra was born on August 7, 1993 in New Delhi, India. She graduated from Modern School Vasant Vihar, New Delhi, in 2011. She received her Bachelor of technology (B.Tech) degree from Banasthali University, Rajasthan India in 2015.