The effects of *TET2*-deficiency on neutrophil gene expression and function

By

(Michael) Yi Zhen Luo

A thesis submitted to the Graduate Program in Pathology & Molecular Medicine

in conformity with the requirements for the

Degree of Master of Science

Queen's University

Kingston, Ontario, Canada

August 2020

Abstract

Background: Clonal hematopoiesis of indeterminate potential (CHIP), a potential precursor to myelodysplastic syndrome (MDS), affects over 10% of adults over 65. Loss-of-function ten-eleven-translocation methylcytosine dioxygenase 2 (*TET2*) variants are common in CHIP and associate with epigenetic dysregulation, inflammation, and cardiovascular disease. Although neutrophils are sensitive to epigenetic dysregulation and contribute to many cardiovascular diseases, the effects of *TET2*-deficiency on neutrophils have not been characterized. We hypothesize that *TET2*-deficiency dysregulates gene expression and function in neutrophils.

Methods: With informed consent and ethics approval, serum and peripheral blood (PB) were collected from CHIP subjects and controls. RNA was extracted from whole PB and mRNA expression of two representative neutrophil granule genes, myeloperoxidase (*MPO*) and lactotransferrin (*LTF*), were assessed using quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR). MPO and LTF serum protein titers were quantified with enzyme-linked immunosorbent assays (ELISA). Additionally, the *Tet2*$^{flox/flox}$ *Vav-iCre* mouse model was used to investigate the effects of *Tet2*-deficiency on neutrophils. Motility and phagocytosis were assessed through a *Staphylococcus aureus* phagocytosis assay. RNA sequencing (RNA-seq) was performed to analyze transcriptomic differences in *Tet2*-deficient neutrophils at baseline and after stimulation with *S. aureus*-derived lipoteichoic acid (LTA). DESeq2 and g:Profiler were used for differential expression analysis and functional annotation, respectively.

Results: CHIP subjects had increased serum protein titers of MPO (p=0.006) and LTF (p=0.001) compared to controls, but *MPO* and *LTF* mRNA expression did not differ significantly. *Tet2*-

deficient murine neutrophils exhibited impaired phagocytosis (p=0.034) and motility (p=0.004). Genes related to motility, the interferon response, and neutrophil development were dysregulated in *Tet2*-deficient murine neutrophils. Overall, the results suggest that *Tet2*-deficiency affects neutrophil gene expression and function.

Discussion: This study addresses a gap in the literature by investigating the effects of *TET2*-deficiency on neutrophils. Since high MPO titers associate with cardiovascular disease, elevated MPO serum protein titers in CHIP subjects may exacerbate comorbidities. While *Mpo* was not overexpressed in *Tet2*-deficient murine neutrophils, they exhibited features associated with MDS, such as overexpression of interferon response genes and impaired neutrophil function. We suspect that dysfunctional *Tet2*-deficient neutrophils may also contribute to increased susceptibility to infection, and call for confirmation through human population studies.

Statement of Co-Authorship

Elina Cook conducted a bulk RNA-seq pilot project which found that CHIP subjects show dysregulation of neutrophil granule genes[1]. With her permission, a figure from the RNA-seq study was adapted and presented in the Chapter 1. Elina and I also co-authored a review article on CHIP and inflammation, and with her permission a figure from the review was reproduced in Chapter 2.

Acknowledgements

I would like thank my supervisor Dr. Michael Rauh, for his guidance, kindness, and support. Dr. Rauh has provided mentorship and plenty of learning opportunities throughout my honours thesis project and M.Sc. which have enabled me to broaden and develop my skill set. Thank you for being an amazing supervisor. I also thank our former lab manager Brooke Snetsinger, and current lab manger Dr. Amy McNaughton for their support and advice and also for their hard work ensuring everything in the lab ran smoothly. I thank Elina Cook, who patiently mentored me, and provided a lot of useful advice. I would like to acknowledge past and present lab members, Christina Ferrone, Mack Blydt-Hansen, Anacelia Gomes, and Keara Sutherland for being supportive and contributing to a friendly lab environment.

I would like to acknowledge my M.Sc. Supervisory Committee, Dr. Stephen Archer and Dr. Chandrakant Tayade, for their insightful advice and suggestions on my experimental design and future directions. I acknowledge Dr. Peter Greer, the coordinator of the pathology graduate program for providing opportunities to present my research at departmental seminars.

I would like to acknowledge Jeffrey Mewburn, Dr. Kimberly Dunham-Snary, and Dr. Charles Hindmarch for providing technical expertise in designing and performing experiments. I would also like to thank the Animal Care team for their assistance with maintaining the mouse colony. Lastly, I would like to thank my family for their support and encouragement. I appreciate the help and support that everyone has provided.

Contents

Contributions from others

Elina Cook provided assistance designing and performing the time lapse microscopy (Chapter 5) and RNA-seq (Chapter 6) experiments. Jeffrey Mewburn and Dr. Kimberly Dunham-Snary from the Archer Lab assisted with the *in vitro* phagocytosis assays (Chapter 5). Jeffrey Mewburn performed time lapse microscopy and Dr. Dunham-Snary provided the GFP-labelled *S. aureus*. Dr. Charles Hindmarch and Brooke Snetsinger from the Queen's Cardiopulmonary Unit (QCPU) core facility prepared the libraries and performed bulk RNA-sequencing (Chapter 6). Dr. Charles Hindmarch also processed the raw sequencing data and aligned the reads to generate bam files and gene count tables. Dr. Kathrin Tyryshkin and Justin Wong provided guidance on analyzing the bulk RNA-seq data. Dr. Amy McNaughton, Elina Cook, Christina Ferrone, Brooke Snetsinger and the Queen's Animal Care team managed the animal colony.

List of Figures

List of Abbreviations

$\Delta\Delta C_T$ = Delta-delta cycle threshold

5-hmC = 5-hydroxymethylcytosine

5-mC = 5-methylcytosine

ABCA13 = ATP binding cassette subfamily A member 13

ACTB = Beta actin

ADAR = Adenosine deaminase RNA specific

AML = Acute myeloid leukemia

APC = Allophycocyanin

ARCH = Age-related clonal hematopoiesis

BPI = Bactericidal permeability increasing protein

C/EBPα = CCAAT/enhancer binding protein alpha

C/EBPβ = CCAAT/enhancer binding protein beta

C/EBPε = CCAAT/enhancer binding protein epsilon

CASS4 = Cas scaffold protein family member 4

CD11b = Integrin subunit alpha M

CDC42EP2 = CDC42 effector protein 2

cDNA = Complementary DNA

CH = Clonal hematopoiesis

CHIP = Clonal hematopoiesis of indeterminate potential

CMML = Chronic myelomonocytic leukemia

CRISP3 = Cysteine rich secretory protein 3

CXCR = C-X-C motif receptor

CXCL = C-X-C motif ligand

DEFA4 = Defensin alpha 4

DNMT3A = DNA methyltransferase 3A

DOCK9 = Dedicator of cytokinesis 9

EDTA = Ethylenediaminetetraacetic acid

ELISA = Enzyme-linked immunosorbent assay

FBS = Fetal bovine serum

FER = Fps/Fes related tyrosine kinase

FITC = Fluorescein isothiocyanate

FLT3 = Fms Related Receptor Tyrosine Kinase 3

fMLP = N-formyl-methionyl-leucyl-phenylalanine

GAPDH = Glyceraldehyde 3-phosphate dehydrogenase

G-CSF = Granulocyte colony stimulating factor

GFI-1 = Growth factor independent 1 transcriptional repressor

GFP = Green fluorescent protein

GMP = Granulocyte-monocyte progenitor

GZMM = Granzyme M

H & E = Hematoxylin and eosin

H3K9 = Histone 3 Lysine 9

HOXA = Homeobox A

HSC = Hematopoietic stem cell

IFITM1 = Interferon induced transmembrane protein 1

IGHG3 = Immunoglobulin heavy constant gamma 3

IGTP = Interferon gamma-induced GTPase

IGV = Integrative genomics viewer

IL = Interleukin

JAK2 = Janus Kinase 2

KIT = CD117 KIT receptor tyrosine kinase

lncRNA = Long non-coding RNA

LTA = Lipoteichoic acid

LTF = Lactotransferrin

Ly6G = Lymphocyte antigen 6 complex locus G6D

MDS = Myelodysplastic syndrome

MeDIP = Methylated DNA immunoprecipitation

MMP2 = Matrix metalloproteinase 2

MPO = Myeloperoxidase

MΦ = Macrophages and monocytes

NETosis = Neutrophil extracellular trap release

NF-κB = Nuclear factor kappa B

NLRP3 = NOD-like receptor protein 3

NRAS = Neuroblastoma RAS viral oncogene homolog

OAS3 = 2'-5'-oligoadenylate synthetase 3

OLFM4 = Olfactomedin 4

Oligo-dT = Oligo-deoxythymidine

Padj = Adjusted p value

PAH = Pulmonary arterial hypertension

PB = Peripheral blood

PBS = Phosphate buffered saline

PI = Propidium iodide

PKN2 = Protein Kinase N2

QCPU = Queen's Cardiopulmonary Unit

qRT-PCR = Quantitative reverse-transcriptase polymerase chain reaction

RASGEF1A = Ras guanine exchange factor domain-containing family member 1A

RNA-seq = RNA sequencing

SCAMP = Secretory carrier membrane protein

SELP = P-selectin

SOX = SRY-Box transcription factor

STAR = Spliced Transcripts Alignment to a Reference

SVA = Seminal vesicle antigen

TESC = Tescalcin

TET2 = Ten-eleven-translocation methyl cytosine dioxygenase 2

TET2-CHIP = *TET2* mutant CHIP

Tet2-KO = *Tet2*-deficient

Tet2-FF = *Tet2*-sufficient

THBS1 = Thrombospondin 1

TLR = Toll-like receptor

TNFα = Tumour necrosis factor alpha

VAF = Variant allele frequency

Chapter 1: Introduction

CHIP is a hematological disorder characterized by the disproportionate expansion of a mutant hematopoietic clone present at a variant allele frequency (VAF) ≥ 2% in the peripheral blood[2,3]. CHIP arises when a hematopoietic stem cell (HSC) acquires a leukemia-associated driver mutation, one of the most common being *TET2*[4,5]. TET2 plays a role in epigenetic regulation by oxidizing methylcytosine to mark it for removal[6]. Loss-of-function *TET2* variants are frequently found in MDS and acute myeloid leukemia (AML) and are predicted to be early hits in leukemogenesis[7-11]. HSCs carrying loss-of-function *TET2* variants show myeloid-skewed differentiation, increased self-renewal, and resistance to apoptosis which enables them to outcompete their wildtype counterparts[12-14]. Individuals who carry CHIP mutations have 10 times greater risk of developing leukemia[15,16] and recent studies have revealed unexpected associations between *TET2*-mutant CHIP (*TET2*-CHIP) and atherosclerotic cardiovascular disease[17,18]. Moreover, CHIP is prevalent among elderly individuals, affecting 10-20% of adults over 65[15,16,19]. To provide context, Statistics Canada estimated that there were 6.6 million Canadians over 65 in 2019 (Table 17-10-0005-01, https://doi.org/10.25318/1710000501-eng) which means that between 660,000 and 1,320,000 Canadians would be predicted to carry CHIP mutations. Given its widespread prevalence and association with numerous diseases, we investigated the pathogenesis of *TET2*-CHIP.

The current understanding of *TET2*-CHIP is primarily based on findings from the *Tet2*-knockout murine model[5]. *Tet2*-deficient macrophages/monocytes (MΦ) are hyper-inflammatory and exacerbate atherosclerotic cardiovascular disease and heart failure in murine models[18,20,21]. Although many leukocyte populations can carry *TET2* mutations, research to date has focused on the role of MΦs in disease development, while other leukocytes remain poorly characterized[14,22].

This study focused on neutrophils because they bear the second highest mutation burden in *TET2*-CHIP (following MΦs)[22], are the most abundant leukocyte in adults[23], and are involved in the pathogenesis of many cardiovascular and inflammatory diseases[24-26]. Furthermore, MDS patients are susceptible to infection and their neutrophils exhibit impaired immune function[27,28]. As CHIP can be a precursor to MDS[2,7], we sought to investigate the origins of this neutrophil dysfunction. We suspected that *TET2*-deficient neutrophils may contribute to comorbidities associated with CHIP and MDS.

Given that *TET2* is recurrently mutated in MDS[10], we used the *Tet2$^{flox/flox}$ Vav-iCre* conditional mouse knockout model[5] to investigate the neutrophil dysfunction phenotype associated with MDS. Bacterial infections, particularly those involving Gram-positive bacteria, are common in MDS[29,30], so we performed *in vitro* phagocytosis assays using *Staphylococcus aureus* to evaluate the phagocytic function and motility of *Tet2*-deficient neutrophils. Moreover, we performed bulk RNA-seq on isolated bone marrow neutrophils to explore transcriptomic changes associated with *Tet2*-deficiency. We compared gene expression between *Tet2*-deficient neutrophils and controls under baseline conditions and after stimulation with *S. aureus* derived lipoteichoic acid (LTA).

We also quantified expression of neutrophil genes in PB samples from CHIP subjects because CHIP associates with epigenetic dysregulation[31] which is predicted to disrupt neutrophil development and function[32]. A bulk RNA-seq study analyzing whole PB from CHIP subjects and controls previously found that mRNA expression of neutrophil granule genes were significantly upregulated in CHIP subjects[1] (Figure 1). We aimed to validate those findings using qRT-PCR and serum ELISA. We focused on key targets such as MPO which can contribute to disease when inappropriately expressed[33,34].

In summary, this study aimed to address a gap in the literature by performing transcriptional and functional analyses in CHIP subjects and knockout mouse models to characterize how *TET2*-deficiency impacts neutrophil gene expression and function.

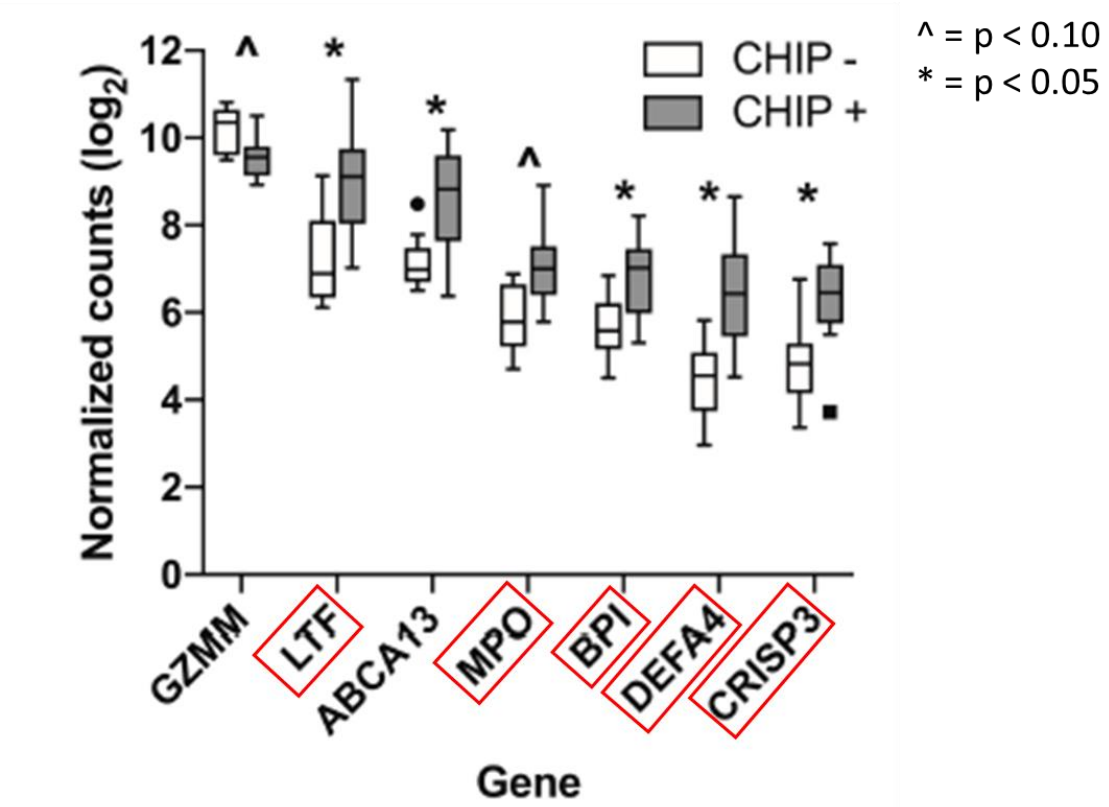


Figure 1. Upregulation of neutrophil granule genes in CHIP subjects
Bulk RNAseq was performed on PB samples collected in PAXGene RNA tubes from 30 CHIP subjects and controls (*TET2* = 12, *DNMT3A* = 8, Other CHIP = 1, No CHIP = 9). Neutrophil granule genes are outlined in red. *MPO* and *LTF*, neutrophil granule genes with known biological significance were among the differentially expressed genes. ^ indicates $p < 0.10$, while * indicates $p < 0.05$. Figure adapted from Cook et al. 2018[1], with permission from the authors.

Chapter 2: Literature Review

2.1 The evolving understanding of CHIP

Healthy hematopoiesis is polyclonal, whereby thousands of HSCs produce a trillion mature blood cells each day[35]. In disorders of clonal expansion such as CHIP, a mutated HSC outcompetes wildtype HSCs and dominates hematopoiesis, passing its mutations on to billions of progeny, collectively referred to as its clone[2,3]. Clonal expansion occurs when an HSC acquires somatic mutations granting growth or survival advantages[2,4,5,12,36].

CHIP is a type of clonal expansion defined by the presence of a mutant clone in the peripheral blood at a VAF of 2% or greater[2,3]. In diseases of clonal expansion, clone size often correlates with disease severity[2]. The 2% cut-off of CHIP was assigned because it associated with comorbidities and it also was the limit at which standard genome sequencing approaches could reliably detect variants[3]. While CHIP is detectable in 10-20% of individuals over 65, more sensitive deep sequencing approaches have detected smaller clones in up to 95% of individuals over 50, however these small clones may not have any or as much biological significance[19,37]. There are some closely related terms such as age-related clonal hematopoiesis (ARCH) and clonal hematopoiesis (CH) that include clones with VAF < 2%, but this report focuses on "CHIP" (VAF ≥ 2%).

The main risk factor for CHIP is age[15,16]. This could be due to the accumulation of mutations over time in HSCs due to replication errors and genotoxic environmental exposures. CHIP may not emerge until old age because HSCs acquire mutations slowly; they divide infrequently and have robust DNA repair mechanisms that correct most mutations[38]. For example, C:T base pair transitions are associated with age-related mutagenesis[39] and this is thought to be due to the accumulation of spontaneous 5-methylcytosine (5-mC) deamination

mutations over the individual's lifespan[40,41]. Additionally, aging exerts a selective pressure for HSCs that retain their self-renewal ability[42]. *TET2* and DNA methyltransferase 3A (*DNMT3A*) are commonly mutated in CHIP and these variants often have potent effects on HSC biology such as increasing self-renewal potential and reducing apoptosis[5,43,44].

There is evidence that CHIP can be a precursor to MDS or AML. Multiple mutations are needed to trigger malignant transformation, and CHIP may be the first hit in this process[4,7]. Genes like *TET2* and *DNMT3A*, which are commonly mutated in CHIP, are also recurrently mutated in malignant conditions such as MDS and AML[9,10,45]. One study showed that HSCs carried many of the same mutations as the leukemic cells, thereby suggesting that the malignant clone arose from a mutated HSC[46]. Interestingly, these pre-leukemic HSCs were still capable of contributing to hematopoiesis, thus they may reside in patients for a long time without causing symptoms[46]. This may explain the sudden presentation of AML as the disease process could remain unnoticed until the HSC acquires enough mutations to trigger malignant transformation[47]. In one striking example where subjects were sequenced then followed, somatic CHIP variants were detected in healthy subjects nearly a decade before they were diagnosed with AML[48]. These findings corroborate the population-based studies which found that CHIP was associated with 10 times greater risk of developing hematological malignancy[15,16].

Further supporting the model of CHIP as a pre-leukemic state, *TET2* mutations have been reported to synergize with other oncogenic mutations[31,49,50]. *TET2* mutations co-occurring with CD117 KIT receptor tyrosine kinase (*KIT*) mutations were associated with mastocytosis in humans and mice[49]. Experiments in mouse models that express the AML1-ETO fusion oncoprotein showed that *Tet2* inactivation had a synergistic effect and accelerated the development of leukemia[31]. Interestingly, the order in which mutations are acquired also

influences the disease phenotype[50]. One study analyzed patients that had both *TET2* and Janus Kinase 2 (*JAK2*) mutations and they found that patients who acquired the *JAK2* mutation first had an excess of erythrocytes[50]. In contrast, patients who acquired the *TET2* mutation first exhibited an abundance of platelets and were less responsive to targeted anti-JAK2 therapies[50]. This finding underscores the importance of studying CHIP mutations with a comprehensive scope that considers the overall mutation landscape and interactions with other variants. Taken together, there is strong evidence suggesting that CHIP mutations are early hits that prime HSCs for leukemogenesis.

Up until 2014, *TET2*- and *DNMT3A*-associated clonal expansion had primarily been considered in the context of hematological malignancies, however a large sequencing study made the surprising finding that this disorder was also associated with cardiovascular disease[15]. They found that CHIP was associated with 2 times greater risk of coronary heart disease and 2.6 times greater risk of ischemic stroke[15]. Although the risks of malignancy and cardiovascular disease are increased, not all carriers go on to develop these diseases, hence the authors coined the term "CHIP" to note the "indeterminate potential"[15-17]. The discovery drew attention towards investigating comorbidities associated with CHIP and the mechanisms through which they arise. Subsequent studies found that CHIP is associated with rampant and dysregulated inflammation which potentially contributes to disease development[17,18,20]. Recently, there has been much excitement around CHIP and many groups have explored associations between CHIP and inflammatory and cardiovascular diseases. Retrospective cohort studies with thousands of participants show that pulmonary disease[51], deep vein thrombosis[52], atherosclerotic cardiovascular disease[15,17] and related complications such as myocardial infarction[15,17] and stroke[15,17] are associated with CHIP. Smaller exploratory studies have reported that CHIP

associates with diabetes[53], gastroesophageal reflux disease[53], and thyroid dysfunction[53], and is a disease-modifying factor in vasculitis[54], and valvulopathy[55]. In summary, CHIP has emerged as a risk factor for many comorbid diseases and it would be beneficial to investigate the pathogenesis to gain a better understanding of its "indeterminate potential".

2.2 Epigenetic dysregulation in CHIP

DNMT3A and *TET2*, the most commonly mutated genes in CHIP, code for enzymes that respectively, add or remove 5-mC marks associated with transcriptional suppression[6,31,56,57]. DNMT3A is a *de novo* methyltransferase that establishes methylation marks on DNA[56,58]. On the other hand, TET2 iteratively oxidizes 5-mC to produce 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine and finally 5-carboxycytosine which can then be removed through base excision repair and replaced with unmethylated cytosine[59] (Figure 2). The removal of 5-mC is predicted to enable expression[57]. Together, DNMT3A and TET2 maintain the DNA methylation code[60].

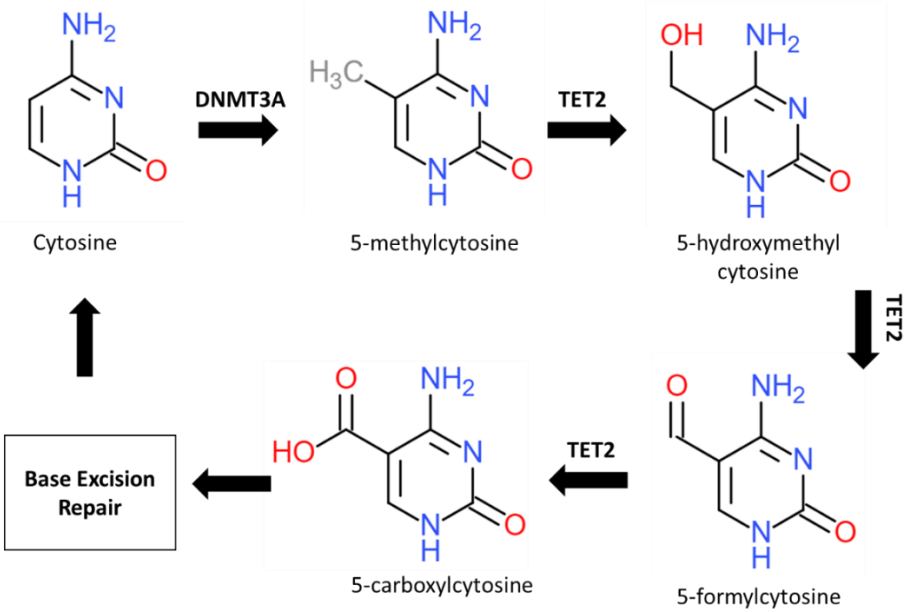


Figure 2. Epigenetic modification of cytosine by DNMT3A and TET2
DNMT3A methylates cytosine creating 5-mC. TET2 oxidizes 5-mC first to 5-hydroxymethyl cytosine, then to 5-formylcytosine and finally to 5-carboxylcytosine. 5-carboxylcytosine is removed through base excision repair and replaced with an unmethylated cytosine. Together DNMT3A and TET2 regulate cytosine methylation.

It follows that loss-of-function *DNMT3A* and *TET2* variants associate with aberrant methylation patterns and dysregulated transcription[6,45,61]. Although DNMT3A and TET2 have opposing enzymatic functions, mutations in either gene have similar effects; HSCs carrying either mutation exhibit a differentiation block and increased self-renewal which provides them a significant fitness advantage in the bone marrow niche[5,43,62]. These functional changes are predicted to contribute to leukemogenesis[4,7].

Currently there is a gap in the understanding of how these mutations contribute to leukemia development: inactivation of *DNMT3A* and *TET2* are associated with altered methylation patterns and the phenotypes associated with these mutations are known, but it remains unclear how these epigenetic changes lead to dysfunction. Loss-of-function *TET2* variants are associated with reduction of 5-hmC, particularly at enhancer sites[31,63], but the transcriptional effects are not fully understood. Further complicating the mechanism, the TET2 enzyme is known to recruit other epigenetic regulators which also influence gene expression[64,65]. On the other hand, loss-of-function *DNMT3A* variants associate with increased methylation in some regions and decreased methylation in others[62]. This finding seems counter-intuitive, but it is possible that there is a complex mechanism regulating methylation and some feedback loops may have become dysregulated with the loss of *DNMT3A*. Furthermore, *DNMT3A* is predicted to have a role in defining the boundaries of "methylation canyons" which are conserved genomic regions with low methylation[66]. Loss of *Dnmt3a* was associated with expansion of the canyons and inappropriate gene expression in murine HSCs[66]. These studies show that loss of *TET2* or *DNMT3A* can have widespread effects on the DNA methylation landscape, thus the pathogenesis of *TET2*- or *DNMT3A*-mutant CHIP could be quite complex.

In addition, there is extensive crosstalk between pathways involving DNA methylation and histone methylation; therefore the disruption of DNA methylation patterns may perturb histone methylation[67,68]. For example, there is a strong correlation between DNA methylation and the methylation of histone 3 at lysine 9 (H3K9), both of which are associated with repression of transcription[69]. Studies in mammalian embryonic stem cells and plant models have revealed some mechanisms underlying this association. DNA methylation may promote histone methylation in nearby regions as certain H3K9 methyltransferase enzymes recognize methylated DNA and are recruited to those regions[70]. On the other hand, DNA methylation may also follow histone methylation because some histone methyltransferase complexes can recruit DNA methyltransferases like DNMT3A and DNMT3B[71]. These findings demonstrate that DNA and histone methylation patterns are mechanistically interconnected, which suggests that aberrant DNA methylation patterns associated with CHIP could disrupt histone methylation patterns, further dysregulating transcription. Further studies exploring associations between epigenetic and transcriptional changes would be an important step towards understanding the mechanism of how CHIP mutations translate to changes in function.

2.3 The role of TET2-deficiency in disease

TET2-inactivating mutations are predicted to contribute to the development of leukemia by promoting clonal expansion. Evidence from mouse models shows that *Tet2*-deficient HSCs have increased self-renewal and proliferative capacity and skewing of differentiation toward the granulocyte-monocyte lineage, thereby leading to a disease phenotype similar to chronic myelomonocytic leukemia (CMML) in humans[5,72]. These mice show splenomegaly and an abundance of neutrophils and monocytes[5]. Another study suggested that *TET2*-deficiency could be a primary driver of this CMML-like phenotype by using RNA-interference to reversibly

downregulate *TET2* expression; they showed that the aberrant self-renewal and proliferation phenotype in *TET2*-deficient HSCs was corrected after *TET2* expression was allowed to return to normal levels[13]. Together these findings highlight the importance of *TET2* as a regulator of HSC function and show how loss of *TET2* could promote disease development.

The pro-inflammatory effects of *TET2*-CHIP are well characterized and while *TET2*-CHIP has primarily been thought of as a driver of inflammation, there is evidence that the relationship may be bi-directional (Figure 3)[12,44,73]. *Tet2*-deficient HSCs have a selective advantage in environments with high levels of inflammatory mediators like tumour necrosis factor alpha (TNFα), because they can withstand the inflammatory stress while wildtype HSCs undergo apoptosis[12,44]. It was predicted that as the clones expand, the intensity of the inflammation will increase, further amplifying the selective advantage for the mutant HSC[73]. In effect, the CHIP clone is modifying the bone marrow environment to facilitate its own expansion.

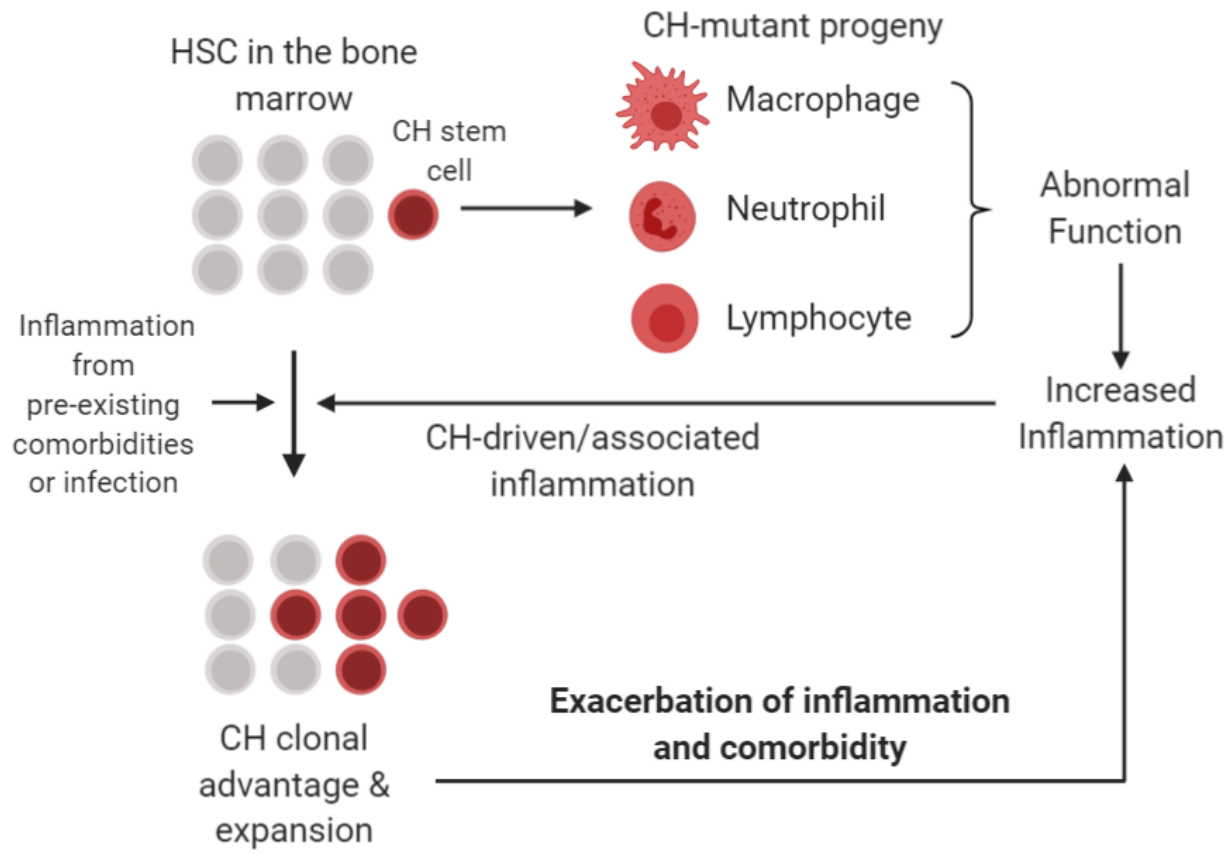


Figure 3. Synergistic relationship between inflammation and clonal expansion
Inflammation provides a selective advantage to CHIP-mutant HSCs allowing them to expand in the bone marrow. The CHIP-mutant HSCs then produce hyper-inflammatory leukocytes which further exacerbate the systemic inflammation. Thus, inflammation and clonal expansion work in a synergistic cycle. Reproduced from Cook et al. 2020[73] with permission from co-authors.

Mechanistic studies analyzing how *Tet2*-deficiency drives inflammation have primarily been performed in *Tet2*-deficient mouse models and have identified MΦs as major contributors to the dysregulated inflammation[18,20,74,75]. *Tet2* controls expression of pro-inflammatory cytokines in macrophages[74] and plays a role in the resolution of inflammation by recruiting histone deacetylases to the *Il6* promoter[75]. Notably, *Tet2*-deficient peritoneal macrophages exhibit constitutive release of inflammatory mediators without a triggering stimulus[74]. Other studies have also shown that NOD-like receptor protein 3 (NLRP3) inflammasome activity is upregulated in *Tet2*-deficient MΦs[18,20]. Together, these studies demonstrate the importance of

TET2 as a regulator of inflammation, and when *TET2* is inactivated, the resulting inflammation may contribute to comorbidities.

Some studies have also provided evidence of causality between *Tet2* inactivation, inflammation, and cardiovascular disease. Fuster and colleagues performed bone marrow transplants with wildtype and *Tet2*-deficient HSCs on atherosclerosis-prone mice and found that mice receiving *Tet2*-deficient HSCs had higher concentrations of inflammatory cytokines and developed more severe disease[18]. Moreover, they showed that administration of NLRP3 inhibitors reduced disease severity, thereby suggesting that inflammation is driving disease progression[18]. Corroborating these mouse model findings, another study found that *TET2*-CHIP was associated with elevated serum concentrations of inflammatory cytokines and increased risk of atherosclerotic disease[17]. Taken together, these results convincingly support a model where *TET2*-CHIP leads to increased inflammation which then contributes to atherosclerotic disease.

In addition to the well characterized association between atherosclerosis and *TET2*-CHIP, there is also mechanistic evidence supporting associations between *TET2* inactivation, pulmonary disease, and heart failure. Potus *et al*. reported that loss-of-function *TET2* variants were enriched in pulmonary arterial hypertension (PAH) patients and that PAH developed spontaneously in *Tet2*-deficient mice[51]. They suggested that elevated interleukin-(IL) 1β levels are central to the pathology because treatment with antibodies targeting IL-1β reduced disease severity[51]. Another group reported that mice receiving *Tet2*-deficient HSC transplants showed increased expression of *Il1b*, pathological cardiac remodelling, and reduced cardiac function; these disease symptoms were also mitigated with NLRP3 inhibitor treatment[20]. These independent studies of atherosclerosis, PAH, and heart failure all reached the same conclusion

that *Tet2*-deficiency-associated inflammation, particularly involving the NLRP3-IL-1β pathway, drives comorbid diseases[18,20,51].

The finding that inflammation promotes comorbid diseases associated with CHIP has implications for potential interventions. There are plenty of anti-inflammatory drugs including antibodies targeting cytokines like IL-1β, but they can have severe side effects such as increased susceptibility to infection[76]. Interestingly, a follow up analysis to a major clinical trial showed that individuals with *TET2*-CHIP benefited more from the cardioprotective effects of an anti-IL-1β antibody compared to matched non-CHIP controls[77], which raises the possibility that the benefits of these anti-inflammatory therapies may outweigh the risks in high risk populations like *TET2*-CHIP carriers.

While many studies have found that *TET2*-CHIP is associated with cardiovascular disease, a recent case study of a family with germline *TET2* mutations did not find an increased incidence of atherosclerotic disease nor elevated levels of inflammatory cytokines and chemokines[78]. This finding was quite surprising because the current literature suggests that the cardiovascular disease association would be stronger due to the increased mutational burden[2,17,18]. A possible explanation is that these individuals might have developed compensatory mechanisms to offset the effects of the *TET2* mutation. These conflicting findings point to the need to study CHIP more in humans because much of the mechanistic knowledge comes from knockout mouse models.

Another limitation is that CHIP research has focused on MΦs in disease development while neglecting other leukocytes. Although neutrophils are the most numerous leukocyte in the peripheral blood and bear the second highest mutation burden in *TET2*-CHIP (following MΦs)[22], the effects on *TET2*-deficiency on this population have not been characterized. As neutrophil

dysfunction plays a role in cardiovascular and inflammatory diseases, we are interested in investigating whether *TET2*-deficient neutrophils also have a role in CHIP-associated comorbidities[24-26].

2.4 Neutrophil development

Neutrophil development is a multi-stage process in which specific genes need to be expressed at the appropriate stages[32,79]. HSCs produce common myeloid progenitor cells which can differentiate into myeloblasts, the earliest cells in the granulocyte lineage[23,80]. Myeloblasts progress through the following differentiation stages: promyelocyte, myelocyte, metamyelocyte, banded cells, and segmented cells (Figure 4). Segmented cells develop into polymorphonuclear leukocytes which are released into the peripheral blood[23].

Stage	Myeloblast	Promyelocyte	Myelocyte	Metamyelocyte	Banded cell	Segmented cell	Polymorphonuclear leukocyte
Key transcription factors	C/EBPα	C/EBPα	C/EBPα C/EBPε	C/EBPα C/EBPε	C/EBPα	C/EBPβ C/EBPδ	C/EBPβ C/EBPδ
Granules being produced	Azurophilic Granules (MPO)	Azurophilic Granules (MPO)	Specific Granules (LTF)	Specific Granules (LTF)	Gelatinase Granules (MMP2)	Secretory Vesicles (SCAMP)	Secretory Vesicles (SCAMP)
Location	Bone marrow	Bone marrow	Bone marrow	Bone marrow	Bone marrow	Bone marrow	Peripheral blood

Figure 4. Stages of neutrophil development
Myeloblasts are the earliest cells committed to the granulocyte lineage. Myeloblasts progress through the following developmental stages: promyelocyte, myelocyte, metamyelocyte, band cell, segmented cell and finally they are released into the peripheral blood as polymorphonuclear leukocytes. The transcription factor C/EBPα is expressed during the early stages of development, C/EBPε during intermediate stages, and C/EBPβ, C/EBPδ, in the final stages of development. Neutrophil granules are produced during specific developmental stages with azurophilic granules being the earliest, followed by specific granules, gelatinase granules and secretory vesicles. MPO and LTF are key proteins contained in azurophilic and specific granules, respectively.

Two transcription factors, CCAAT/enhancer binding protein alpha (C/EBPα) and epsilon (C/EBPε) have important roles in directing neutrophil development[81]. *CEBPα* expression begins early in neutrophil development and is highly expressed in all developmental stages until the segmented cell and PMN stage where *CEBPα* expression ceases[32,81,82]. C/EBPα is essential for driving granulocytic differentiation as *Cebpα*-deficient mice are unable to produce neutrophils; instead they generate immature myeloid cells that are arrested at an early stage of differentiation[81,82]. While C/EBPα is responsible for initiating differentiation, *C/EBPε* is expressed only during the intermediate myelocyte and metamyelocyte stages and is essential for neutrophil maturation because it directs differentiation towards neutrophils rather than other granulocytes[83-85]. *C/EBPε* also shuts down the proliferative program which is characteristic of immature progenitors while upregulating expression of granule proteins such as LTF and collagenase[83,84]. After the metamyelocyte stage, *CEBPε* expression ceases and other transcription factors are upregulated to allow for the final stages of differentiation[23,80]. Taken together, these studies suggest that tightly regulated control of gene expression is critical in neutrophil development.

Epigenetic regulation plays an important role in orchestrating the neutrophil developmental program[23,32,80]. Grassi and colleagues analyzed neutrophil development using a multimodal approach involving transcriptomic profiling, epigenetic analysis, and functional assays[32]. They found that histone acetylation patterns dynamically change during development to allow expression of genes at appropriate times[32]. While they predict that dynamic histone acetylation is the main regulatory mechanism, DNA methylation was found to have a modest role in regulation of expression, mostly in a repressive capacity[32]. Other studies have shown the

importance of DNA methylation in regulating neutrophil function[86,87]. Based on these studies, we hypothesize that *TET2*-CHIP-driven epigenetic dysregulation may disturb neutrophil function.

Following the report by Grassi *et al*. in 2018, a subsequent study explored correlations between the transcriptome and proteome during neutrophil development[79]. As neutrophils mature, metabolic functions including transcription, translation, and mitochondrial activity decrease while immune effector functions such as chemotaxis, generation of reactive oxygen species, and granule release are developed[79]. For example, neutrophils show increasing surface expression of chemotactic receptors including C-X-C motif receptor (CXCR) 1 and CXCR2 during later stages of development[79]. An important implication of these findings is that mRNA expression may not correlate with protein abundance within mature neutrophils. For example, MPO is one of the most abundant proteins in neutrophils, but mature neutrophils do not express *MPO* mRNA; rather MPO was synthesized at an early developmental stage[79]. Neutrophil granule proteins like MPO are potent and have important anti-microbial functions, but to limit collateral damage to the host, expression of these granule proteins is regulated through a principle called "targeting by timing", discussed in detail below[80,88].

2.5 Neutrophil granule biology

Neutrophils can release granules containing cytotoxic and microbicidal proteins in response to pathogens and other danger signals. These granules are essential for neutrophil effector function and individuals with deficiencies in granule proteins are vulnerable to recurrent infections[88-90]. There are four types of granules and they are synthesized during different stages of development: azurophilic granules are synthesized first, followed by specific granules, then gelatinase granules, and finally secretory vesicles[80,88,91]. The "targeting by timing" principle states that the expression of each set of granule proteins is separated temporally and only one

type of granule is synthesized at a time[32,79,80,88]. Transcription factors and epigenetic mechanisms regulate the timing of expression and coordinate this process[32,92]. The expression of genes related to azurophilic granules is regulated primarily by C/EBPα and growth factor independent 1 transcriptional repressor (GFI-1)[93,94]. The former cooperates with granulocyte colony stimulating factor (G-CSF) to upregulate expression of azurophilic granule proteins, while the latter downregulates expression[93,94]. C/EBPε is responsible for activating the expression of specific and gelatinase granules, while secretory vesicles are regulated by transcription factors expressed at the terminal stages of neutrophil differentiation[79,83]. These studies show the complexity of the system that regulates granule synthesis and we suspect that the epigenetic dysregulation associated with CHIP could disturb this process.

Each granule type serves different functions and are mobilized under different conditions[88,91]. Azurophilic granules contain potent bactericidal and oxidative agents including MPO, defensins, and serine proteases[79,88]. The majority of azurophilic granules fuse with the phagosome to destroy the microbes, but some are also exocytosed[88]. Some of the proteins in azurophilic granules, especially MPO and neutrophil elastase, have the potential to cause severe tissue damage, so these granules are only released in response to strong stimuli[33,88,91]. Specific granules have the next highest threshold for mobilization, followed by gelatinase granules, and lastly secretory vesicles[91]. Specific granules and gelatinase granules contain bactericidal enzymes including LTF and cytochrome b558, as well as metalloproteinases and collagenases which facilitate movement through the extracellular matrix[79,88]. Readily released secretory vesicles contain many transmembrane receptors for chemokines and inflammatory mediators that prime the neutrophil for activation and chemotaxis[79,88]. Currently, the mechanism controlling the different thresholds for exocytosis is not fully understood, but it is believed to be based on the

density of membrane fusion proteins, actin, and microtubule attachments on the granule membrane[80]. This system controlling granule exocytosis enables neutrophils to respond proportionately to the severity of the threat, thereby limiting collateral damage to the host while effectively fighting off the infection.

2.6 The essential role of neutrophils in innate immunity

Neutrophils form the backbone of the innate immune system and are essential for effective immunity against bacterial and fungal pathogens[23,95]. Individuals with neutropenia or neutrophil dysfunction are at an increased risk of developing serious and potentially life-threatening infections[23,96]. In addition to releasing granules, neutrophils employ numerous other mechanisms to fight against infection including phagocytosis and the release of neutrophil extracellular traps (NETs)[95]. However, the first step in the response to severe infections is increasing neutrophil production to meet the increased demand for effector cells[97,98].

During severe infections, inflammatory cytokines and growth factors, such as IL-6 and G-CSF, are released to activate emergency granulopoiesis which results in the rapid production of neutrophils[97-101]. It was thought that initiators of emergency hematopoiesis were primarily the HSCs and leukocytes at the site of infection[99,102]. HSCs express pattern recognition receptors, including the Toll-like receptor (TLR) family, allowing them to directly detect and respond to infectious threats by increasing production of myeloid cells[99,102]. Meanwhile, macrophages and other leukocytes at the site of infection would be activated and express IL-6, further stimulating myelopoiesis[103]. Evidence has emerged that endothelial cells also play a critical role in initiating emergency granulopoiesis; they detect pathogens with TLRs and produce G-CSF in response[98]. This finding suggests that endothelial cells act as sentries, setting off the alarm to tell HSCs to increase neutrophil production when they detect invading pathogens[98]. When these signals reach

the bone marrow, they trigger a transcriptional switch in HSCs and progenitors[97]. Inflammatory cytokines downregulate the expression of *CEBPα,* the master regulator of steady state hematopoiesis, and upregulate CCAAT/enhancer binding protein beta (*CEBPβ)* which is the primary transcription factor in emergency granulopoesis[97]. The downregulation of C/EBPα reduces its inhibitory effects on cell proliferation while C/EBPβ promotes granulocytic differentiation; together these changes allow for increased production of neutrophils[97].

While emergency granulopoiesis is essential in the immune response against infections, Meisel and colleagues suggested that it may also contribute to leukemogenesis in *TET2*-deficient hosts[103]. *Tet2*-deficient mice exhibited a dysregulated emergency response to infection, leading to a CMML-like phenotype with excessive myeloproliferation and splenomegaly[103]. The mechanism appears to involve over-stimulation of the IL-6 signalling pathway in granulocyte-monocyte progenitors (GMPs): *Tet2*-deficient GMPs are sensitized to IL-6 due to increased surface expression of the IL-6 receptor and this is compounded by the elevated levels of IL-6 associated with *Tet2*-deficiency[75,103]. The dysregulated proliferation of inflammatory *Tet2*-deficient leukocytes may exacerbate inflammation, potentially contributing to the development of comorbid diseases.

After the neutrophils are produced, they migrate to the site of infection. G-CSF, and the chemokines C-X-C motif ligand (CXCL) 1 and CXCL2 are involved in this mobilization process[104,105]. The bone marrow expresses the chemokine CXCL12 to control the release of neutrophils into the system, but during infections, G-CSF downregulates expression of the CXCL12 receptor in neutrophils, enabling them to leave the bone marrow[104]. Then the neutrophils navigate to the site of infection through CXCL1 and CXCL2 gradients[104,105]. As neutrophils approach the site of infection, they encounter inflammatory cytokines including

TNFα and IL-1β which activates the neutrophils and increases expression of integrins and adhesion molecules involved in extravasation[95,106]. After extravasation, the neutrophils home in on the precise location of the pathogen for phagocytosis, but they need to navigate through multiple chemotactic gradients, including those produced by endothelial cells, other leukocytes, and pathogens[107-109]. Mechanistic studies have shown that neutrophils prioritize pathogen-specific chemotactic signals such as N-formyl-methionyl-leucyl-phenylalanine (fMLP), over host-derived chemokines like CXCL2[107-109]. The phosphatidylinositol 3-kinase and p38 mitogen-activated protein kinase signalling pathways are involved in translating the chemotactic signals to directional migration[107-109].

Depending on the severity of the infection, neutrophils will respond using different means[95,110,111]. Neutrophils tend to use degranulation and phagocytosis to fight pathogens, but during severe infections, neutrophils can resort to neutrophil extracellular trap release (NETosis) whereby they expel fibrous structures of decondensed chromatin coated with microbicidal proteins[110,111]. These structures help to contain the infected region by ensnaring and degrading bacteria and toxins[110]. While NETs have an important role in immunity, they have also been implicated in the pathogenesis of numerous diseases including atherosclerosis[25], thrombosis[52], and acute kidney injury[112]. Recently Wolach *et al*. demonstrated NETosis was dysregulated in individuals with *JAK2*^{V617F}-mutant CHIP and this increased their thrombosis risk[52]. Neutrophils isolated from these patients were primed for NETosis and epidemiological analysis showed that 25% of individuals with *JAK2*^{V617F}-mutant CHIP had thrombotic events[52]. This was the first study that reported the involvement of neutrophils in the CHIP disease process, and it showed that neutrophil dysregulation can contribute to disease development. However, *JAK2*^{V617F} only

accounts for a small fraction of CHIP cases, while the effects of common mutations, like *TET2*, on neutrophils have not yet been characterized.

Although *TET2*-deficiency has not been directly studied in neutrophils, there are reports of neutrophil dysfunction in MDS[27,28], a hematological malignancy where *TET2* is frequently mutated[10]. Neutrophils isolated from MDS patients show impaired chemotaxis and bactericidal function; these defects in neutrophil function are thought to contribute to the increased risk of infection associated with MDS[29,30]. Given that *TET2*-deficiency associates with epigenetic dysregulation which likely disrupts neutrophil function[31], it is possible that inactivating *TET2* mutations may contribute to the dysfunctional neutrophil phenotype observed in MDS patients.

2.7 Hypothesis and experimental aims

Overall, the literature suggests that *TET2* mutations may perturb neutrophil biology, potentially predisposing to disease. Epigenetic mechanisms important in regulating neutrophil development and function are predicted to be disrupted in *TET2*-CHIP[31,32]. Furthermore, although neutrophils are essential in innate immunity and play a role in many cardiovascular and inflammatory diseases, they have not been characterized in *TET2*-CHIP. This study addresses a gap in the literature by investigating the effects of *TET2*-deficiency on neutrophils. We hypothesized that *TET2*-inactivating mutations dysregulate neutrophil gene expression and function. The experimental aims are as follows: **Aim 1:** Validate the neutrophil granule signature by performing qRT-PCR and serum ELISA on samples from CHIP subjects and controls, **Aim 2:** Assess functional differences between *Tet2*-deficient and control murine neutrophils, **Aim 3:** Assess transcriptomic differences between *Tet2*-deficient and control murine neutrophils.

Chapter 3: Materials and Methods

3.1 CHIP subject cohort

Four hundred and seventeen adults over the age of 65, that did not have hematological malignancies at the time of recruitment, were recruited from Baycrest and Sunnybrook Hospitals in Toronto, Ontario. Informed written consent was obtained from all participants and the study received approval from research ethics boards at Queen's University, Baycrest Hospital, and Sunnybrook Hospital (ethics approval number: 6014336, Appendix A). Upon enrollment, PB, serum, and clinical data including complete blood counts were collected from each participant. PB samples were collected in PAXGene DNA and RNA tubes (Qiagen, Hilden, Germany). Serum samples were obtained from aliquots of blood collected for routine blood tests, separated by centrifugation.

To determine the CHIP status of the participants, a custom, pan-myeloid Ion Torrent DNA sequencing panel was used to detect mutations in 48 CHIP genes[1,113] (ThermoFisher Scientific, Waltham, MA). Participants were classified by mutation status: No CHIP, *TET2*-CHIP, *DNMT3A*-CHIP, and Other CHIP. No CHIP was defined as having no detectable deleterious CHIP gene variants, or they were present at VAF < 2% in the PB. *TET2*-CHIP and *DNMT3A*-CHIP were defined as having a deleterious mutation in the respective gene at a VAF ≥ 2% in the PB, while Other CHIP was defined as having a deleterious CHIP mutation other than *TET2* or *DNMT3A* at a VAF ≥ 2% in the PB.

To validate the enriched neutrophil granule signature[1], PB samples from 18 individuals not in the RNA-seq cohort were selected for qRT-PCR analysis (Appendix Table S1). Serum samples from 64 individuals were analyzed with ELISA (Appendix Table S2), with 23 of the 64 individuals from the RNA-seq cohort and 16 from the qRT-PCR cohort.

3.2 RNA extraction and quantification of neutrophil gene expression

RNA was extracted from PB samples using a PaxGene Blood RNA kit (Qiagen, Hilden, Germany) following the manufacturer's instructions and stored at -80°C. 60ng of RNA was converted to complementary DNA (cDNA) using the M-MuLV Reverse-Transcriptase kit (New England BioLabs, Ipswich, MA) following the manufacturer's protocol and the cDNA products were stored at -20°C. THP-1 cells (American Type Culture Collection, Manassas, VA) were selected as calibrator samples for the qRT-PCR experiments because they express *LTF* and *MPO* at moderate levels[114]. RNA was extracted from a sample containing approximately 1 million cells using the RNeasy Mini Kit, following the manufacturer's protocol (Qiagen, Hilden, Germany). cDNA was made and stored as described above.

LTF and *MPO* mRNA expression in PB samples were quantified using qRT-PCR. qRT-PCR was performed using the Fast SYBR Green qPCR kit (Applied Biosystems, Foster City, CA) on a StepOne machine (Applied Biosystems, Foster City, CA) following a standard protocol (Appendix Table S3). Each 20 μL reaction contained: 1 μL cDNA, 10 μL SYBR Green Master Mix, 1 μL forward primer, 1 μL reverse primer, and 7 μL distilled water. Primers used are listed in Appendix Table S4.

Fold change was calculated using the delta-delta cycle threshold ($\Delta\Delta C_T$) method[115]. The housekeeping genes beta actin (*ACTB*) and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) were used to establish reference levels of mRNA expression and control for differences in RNA quantity between samples. THP-1 cells were included in each batch and the fold changes in *MPO* and *LTF* expression were calculated relative to this calibrator. All samples were run in duplicate and average fold change was reported.

3.3 Serum ELISA for MPO and LTF proteins

The Human MPO Instant ELISA Kit (Invitrogen, Carlsbad, CA) and Human LTF ELISA Kit (Sigma, St. Louis, MO) were used to quantify MPO and LTF serum protein titers. Assays were performed according to the manufacturer's instructions except for the MPO ELISA where samples were diluted 1 to 25 instead of 1 to 50. The VersaMax microplate spectrophotometer (Molecular Devices, San Jose, CA) was used to quantify the colorimetric reaction (450nm target wavelength and 650nm background). All samples were run in duplicate and the average optical density was compared to standards to infer the protein concentration.

3.4 Tet2-deficient mouse model

With ethics approval from the Queen's University Animal Care Committee (Protocol number 2017-1724, Appendix B), the *Tet2*$^{f/f}$ *Vav-iCre* conditional murine knockout model was used to investigate the effects of *Tet2*-deficiency on neutrophils. C57BL/6 mice with LoxP sites flanking the 3rd exon of the *Tet2* gene (*Tet2*$^{f/f}$) (stock number 017573, Jackson Laboratory, Bar Harbor, ME) were crossed with C57BL/6 mice expressing Cre recombinase from the *Vav1*-promoter (stock number 008610, Jackson Laboratory, Bar Harbor, ME) to create *Tet2*$^{f/f}$ Vav-iCre$^{+/-}$ mice (*Tet2*-KO). In *Tet2*-KO mice, all hematopoietic cells are predicted to be *Tet2*-deficient because exon 3, which encodes half the Tet2 protein, has been excised from the *Tet2* gene[5]. This knockout model shares some similarities with *TET2*-CHIP in humans because most inactivating mutations occur within exon 3 of the human *TET2* gene[5]. Age and sex-matched *Tet2*$^{f/f}$ Vav-iCre$^{-/-}$ mice were chosen as the *Tet2*-sufficient controls (*Tet2*-FF).

Ear punches collected for genotyping were digested in 150 μL of TNES plus proteinase K (Qiagen, Hilden, Germany) lysis solution in a 54°C water bath overnight. 100 μL of the digested ear punch solution was mixed with 5 μL of 5M NaCl and 73.5 μL of isopropanol, then DNA was

pelleted by centrifugation. After washing with 70% ethanol, the DNA was re-suspended in 50 μL TE buffer (Invitrogen, Carlsbad, CA). Genotyping was done using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA), following the manufacturer's protocol (details in Appendix Table S5). Genotyping primers recommended by Jackson Laboratories were used (Appendix Table S6). PCR parameters are described in Appendix Table S7. PCR products were run on a 2% agarose gel for 40 minutes at 120 volts and visualized using the Amersham Imager 600 (GE Life Sciences, Marlborough, MA). Genotypes were assessed as follows: *Tet2*-floxed-allele product is 430bp while the wildtype allele is 250bp; *Vav-iCre* positive allele is 236bp while the wildtype allele is 324bp.

3.5 Murine neutrophil isolation and serum collection

Bone marrow was extracted from the femurs and tibias of experimental animals (2 – 3 months old) using a protocol modified from Amend and colleagues[116]. Experimental animals were euthanized by CO_2 asphyxiation, followed by cardiac puncture and cervical dislocation in accordance with a protocol approved by Animal Care Services at Queen's University (Protocol number 2017-1724). PB collected from the cardiac puncture was centrifuged at 1000 xg for 10 minutes at 4°C to isolate serum which was then stored at -80°C.

Femurs and tibias were extracted, cleaned and placed in ice cold Ca^{2+}, Mg^{2+}-free phosphate-buffered saline (PBS) solution (GE Life Sciences, Marlborough, MA) with 1% penicillin/streptomycin (GE Life Sciences, Marlborough, MA). The epiphysis of each bone was removed and the marrow was pelleted by centrifuging at 10,000 xg for 10 seconds at 4°C. Then bone marrow cells were resuspended in ice cold Ca^{2+}, Mg^{2+}-free PBS solution with 2% fetal bovine serum (FBS) (Sigma, St. Louis, MO) and 1mM ethylenediaminetetraacetic acid (EDTA) (Sigma, St. Louis, MO). The suspension was filtered through a 70 μm strainer (ThermoFisher

Scientific, Waltham, MA) and kept on ice. A 200 μL aliquot of the suspension was taken for flow cytometry analysis on the MacsQuant Analyzer 10 (Miltenyi Biotec, Bergisch Gladbach, Germany).

Bone marrow neutrophils were isolated using a mouse neutrophil negative immunomagnetic selection kit (catalog number 19762, Stem Cell Technologies, Vancouver, Canada), according to the manufacturer's instructions. In brief, an antibody cocktail targeting all bone marrow populations except neutrophils was added to the bone marrow suspension. Biotinylated secondary antibodies and magnetic beads were added, then the suspension was exposed to a magnet leaving an enriched population of neutrophils in suspension. The neutrophil isolation process takes about 4 hours on average.

Flow cytometry was performed to assess purity and viability. Shortly after isolation, neutrophils were stained with CF-Blue-labelled Annexin-V and propidium iodide (PI) (catalog number: ab214485, Abcam, Cambridge, U.K.) to assess apoptosis. The viability assay was performed according to the manufacturer's instructions. Briefly, 100,000 cells were resuspended in 100 μL of Annexin-V binding buffer and stained with 5 μL of PI and 5 μL of Annexin-V CF Blue labelled antibody for 15 minutes at room temperature in the dark. An unstained sample and a sample heated for 2 minutes at 85℃ were used as the negative and positive controls, respectively. Results were analyzed using Flowing Software 2 (Turku Centre for Biotechnology).

To assess purity, neutrophils were stained with allophycocyanin (APC) labelled antibodies against integrin subunit alpha M (CD11b) (clone M1/70.15.11.5, Miltenyi Biotech, Bergisch Gladbach, Germany) and fluorescein isothiocyanate (FITC) labelled antibodies against lymphocyte antigen 6 complex locus G6D (Ly6G) (clone REA526, Miltenyi Biotech, Bergisch Gladbach). Ly6G is a neutrophil-specific surface marker[52,117], while CD11b is expressed by

many innate immune cells including macrophages and neutrophils[118]. CD11b and Ly6G are often used in combination to detect neutrophil populations in flow cytometry[119] and was recommended by the manufacturer of the mouse neutrophil isolation kit. To verify flow cytometry results, the neutrophil suspension was smeared on a slide, stained with hematoxylin and eosin (H & E), then analyzed by a hematopathologist.

3.6 Evaluating phagocytosis and motility through time-lapse imaging

Phagocytic function and motility of neutrophils from *Tet2*-KO and *Tet2*-FF mice was assessed through an *in vitro* phagocytosis assay adapted from a published protocol[120]. Immediately following isolation, an aliquot of approximately 1 million neutrophils was taken and centrifuged at 400 xg, for 10 minutes at room temperature with slow acceleration. The pellet was resuspended in 1 mL of clear RPMI (ThermoFisher Scientific, Waltham, MA) with 10% mouse serum, 150 nM Mitotracker Deep Red to stain mitochondrial DNA (catalog number: M22426, ThermoFisher Scientific, Waltham, MA), and 2 drops of Nuc Blue to stain nuclear DNA (catalog number: R37605, ThermoFisher Scientific, Waltham, MA) yielding a final concentration of $1x10^6$ neutrophils/mL of media. Neutrophils were incubated in the imaging solution at room temperature for 10 minutes, then 500 μL of the suspension was loaded onto a 35mm glass bottom imaging plate (catalog number: P35G-1.5-14-C, Mattek corporation, Ashland, MA). $5 x10^6$ green fluorescent protein (GFP) labelled *Staphylococcus aureus* were added to the neutrophil suspension to achieve a 10:1 bacteria:neutrophil ratio. Time lapse imaging was performed using the SP8-X confocal scanning microscope (Leica, Wetzlar, Germany). Images were captured from 4 independent fields of view every 15 seconds for 30 minutes. Samples were imaged in a randomly determined order (i.e. *Tet2*-KO vs. *Tet2*-FF first).

Analysis was performed using the FIJI and TrackMate extensions for ImageJ v1.52p[121,122]. Nuc Blue staining highlighted the multi-lobular morphology of the nucleus and complemented the flow cytometry data in verifying that the cells were neutrophils. Since the neutrophil cytoplasm stained strongly for the MitoTracker Deep Red, the fluorescent signal was used to approximate the neutrophil's cell boundaries. Colocalization of GFP labelled bacteria with the Red MitoTracker fluorescence was interpreted as a phagocytosis event (Figure 5). The semi-automated process of quantifying co-localization involved three stages: identifying the area occupied by neutrophils, identifying GFP-signals from individual bacteria, and evaluating whether the signals overlap. More details can be found in Appendix C. To normalize data between experiments, the number of bacteria phagocytosed per neutrophil was divided by the ratio of *S. aureus* to neutrophils since this ratio was positively correlated with phagocytosis rates (Appendix Figure S1).

Motility analysis was performed using the TrackMate module[122]. Briefly, the module searches through each time frame and identifies neutrophils based on size and MitoTracker fluorescence intensity. Each neutrophil was tracked through adjacent time frames based on their positions in the current and previous frame, thus proposing a trajectory over time (Figure 6). Then average velocity was calculated for each neutrophil.

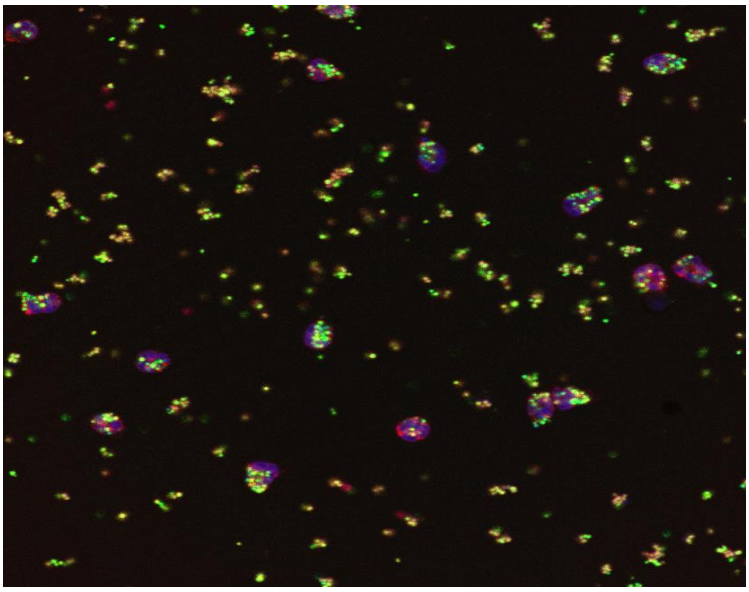


Figure 5. Representative image of neutrophils from *Tet2*-FF mice phagocytosing GFP-labelled *S. aureus*
This is an image that shows the merged Nuc Blue, MitoTracker Deep Red and GFP fluorescence. The neutrophils fluoresce red and blue from MitoTracker and Nuc Blue respectively, while *S. aureus* express GFP from a construct and fluoresce green. *S. aureus* associated-GFP signals can be seen colocalizing with fluorescence from neutrophils, and this was interpreted as phagocytosis.

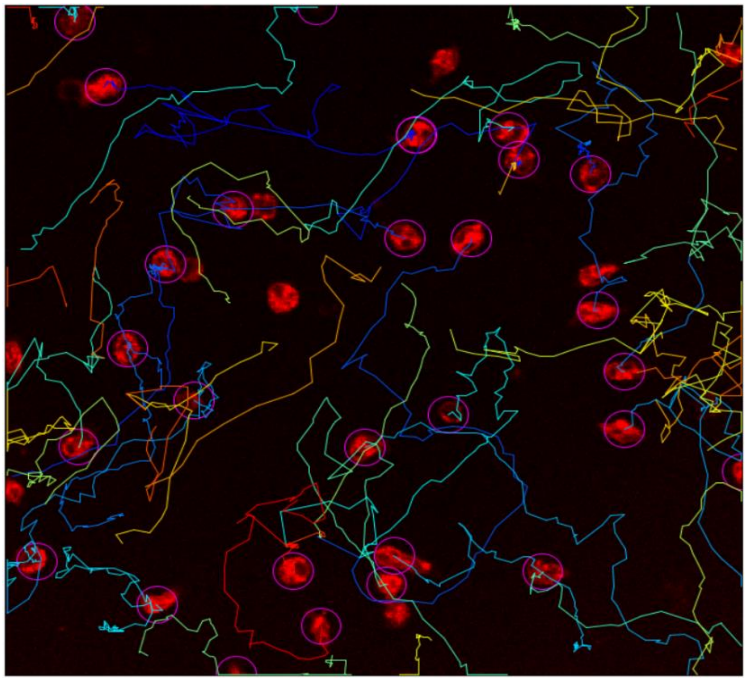


Figure 6. Representative motility analysis image
Neutrophils fluoresce red from the MitoTracker dye. The TrackMate module identified neutrophils (highlighted with a purple circle) based on fluorescence intensity and shape parameters and tracked their movement from frame to frame. Each line represents the trajectory of a neutrophil over the 30 minute experiment.

3.7 Stimulation of neutrophils with S. aureus-derived lipoteichoic acid

To explore the transcriptomic differences between *Tet2*-KO and *Tet2*-FF neutrophils in their activated state, *S. aureus*-derived LTA (catalog code: tlrl-pslta, InvivoGen, San Diego, CA) was selected as a stimulus since it simulates the activation of neutrophils during the *in vitro* phagocytosis assay previously done. LTA is a cell wall component in Gram-positive bacteria and a TLR2 agonist that activates the NF-κB pathway and triggers an inflammatory phenotype[123-125].

Previous studies have used LTA to stimulate human, but not mouse neutrophils[123-126]. The concentration of LTA used ranged from 100 ng/mL to 10 μg/mL, with higher concentrations showing more pronounced differences in expression[123,124,127]. To determine the optimal conditions for LTA stimulation in mouse neutrophils, pilot experiments were performed. *Il1b* and *Tnfa* expression were measured by qRT-PCR and used (primers listed in Appendix Table S8) as an indicator of activation, since they are well characterized target genes of the nuclear factor kappa B (NF-κB) pathway[92]. Stimulation with 10 μg/mL of LTA induced significant upregulation of inflammatory cytokine expression suggesting robust activation, thus this concentration was used for subsequent experiments (Appendix Figure S2).

The duration of LTA stimulation was also optimized, using 1, 3, 6, 9, and 12 hour time points. Previous studies showed that longer stimulation yielded greater changes in expression[123], but this needed to be balanced with the neutrophils' short life spans[128]. For each time point, apoptosis was assessed through flow cytometry using Annexin-V PI staining and activation was assessed through *Il1b* and *Tnfa* mRNA expression and Cd11b surface expression. Cd11b translocates to the surface upon stimulation with inflammatory cytokines or microbial products[52,124,129]; since this could be detected with flow cytometry, it was a fast and

complementary method to verify neutrophil activation in addition to inflammatory cytokine expression.

The 1 hour time point was ruled out because the neutrophils failed to show upregulation of inflammatory cytokines *Tnfa* and *Il1b* which suggests that not all activation-associated transcriptional changes have occurred yet (Appendix Figure S3). The 9 and 12 hour time points were ruled out because over 30% of the population was in early or late apoptosis (Appendix Figure S4). The 3 and 6 hour time points both showed evidence of neutrophil activation based on inflammatory gene expression (Appendix Figure S3), and Cd11b surface expression (Appendix Figure S5), and viability at both time points was similar (Appendix Figure S4). Taken together, the data suggests that the 3 and 6 hour time points were both good options, however the 6 hour time point was chosen since other studies analyzing LTA stimulation in human neutrophils have treated neutrophils for 6 hours or 8 hours[123,125,127].

3.8 Bulk RNA-seq sample collection

Four *Tet2*-KO and *Tet2*-FF pairs were selected for the experiment. To account for age and sex as confounding variables, only 2.2 month old male mice were selected. However, validation qRT-PCR will be done in both females and males. Neutrophils were isolated using the previously described negative immunomagnetic selection protocol (see section 3.5), then divided into three treatment groups: baseline, LTA stimulation or vehicle (Figure 7). Baseline was evaluated as the neutrophil physiological resting state and RNA was isolated immediately following isolation.

The LTA stimulation group were treated with PBS with 2% FBS and 10 μg/mL of *S. aureus* LTA for 6 hours prior to extracting RNA. To account for transcriptional changes due to the 6 hour treatment time, a vehicle group was included where neutrophils were treated with PBS

with 2% FBS and 1% water for 6 hours prior to RNA extraction. A total of 24 samples were collected with 4 replicates per treatment group. Viability was assessed through flow cytometry with Annexin-V, PI staining, and purity was assessed through Cd11b, Ly6G flow cytometry and H & E staining.

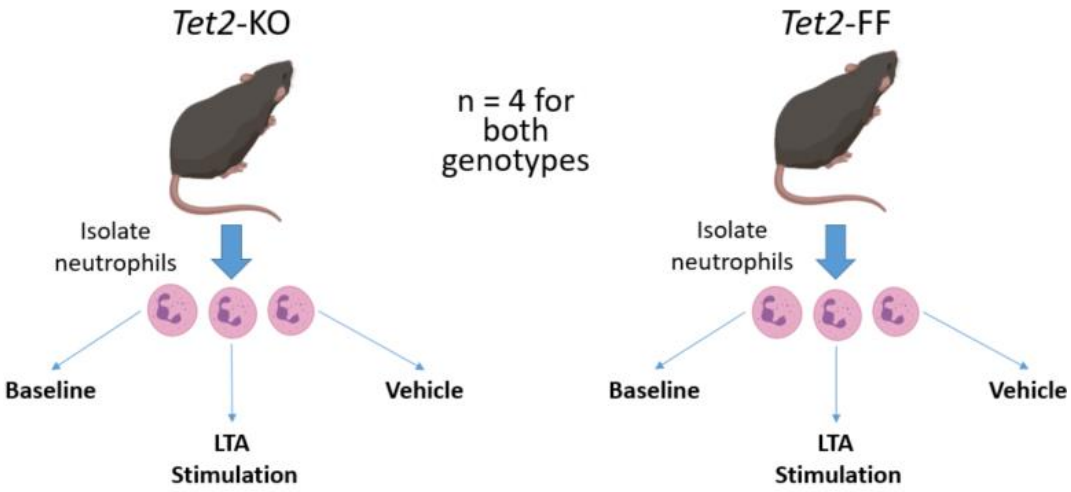


Figure 7. Bulk RNA-seq experimental design
We selected 4 age and sex-matched pairs of *Tet2*-KO mice for the experiment. Neutrophils were isolated and divided into 3 treatment groups: baseline, LTA stimulation and Vehicle. In the baseline group, we extracted RNA immediately after isolating neutrophils. In the LTA Stimulation group we treated neutrophils with PBS with 2% FBS and 10ug/mL of *S. aureus* LTA for 6 hours before extracting RNA. In the vehicle group, neutrophils were treated with PBS with 2% FBS and 1% water for 6 hours before RNA extraction.

Neutrophil RNA was collected in TRIzol reagent (Invitrogen, Carlsbad, CA) and stored at -80℃ until extraction. An aqueous/organic phase extraction with chloroform (ThermoFisher Scientific, Waltham, CA) was performed to extract the RNA from Trizol and the aqueous layer containing the RNA was processed using the RNeasy Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA aliquots were made for quality control checks using qRT-PCR and the Bioanalyzer RNA 6000 Nano kit (Agilent, Santa Clara, CA).

3.9 Library preparation and sequencing

Libraries were prepared using the QuantSeq 3' mRNA-Seq library prep kit for Illumina (Lexogen, Vienna, Austria) with 300ng of RNA input. A 3' poly-A tail capture approach for library generation was chosen to facilitate quantification of mRNA transcripts. A limitation was that non-coding RNAs may not be captured, however it is better suited for detecting shorter and low abundance transcripts[130]. In this approach, instead of using random hexamer primers, oligo-deoxythymidine (oligo-dT) primers were used for first strand synthesis, thus 3' bias was predicted in the sequencing coverage[130].

Single-end, 75bp read length, bulk RNA-seq with 130 million reads was performed on the Illumina NextSeq platform at the Queen's Cardiopulmonary Unit core facility. Twenty-four 3' mRNA libraries were sequenced with an average of 6 million reads per library. Unfortunately, the indexing barcode malfunctioned for one sample: only 8000 reads were assigned to that library and 6 million reads were left unassigned. The failed library and unassigned reads were dropped from analysis.

3.10 Alignment of reads and creating gene count tables

The sequencer output was de-multiplexed, fastq files were generated, reads were aligned, and a gene count table was created. Read alignment was performed on the Queen's Center for Advanced Computing computer cluster. Raw fastq files were analyzed using the FastQC module to check read quality (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, accessed March 19th, 2020). Read quality was high in all libraries, but the FastQC results suggested that the poly-A tails of transcripts were captured in the reads (Appendix Figure S6). Since potential adaptor sequences were detected in some samples (Appendix Figure S6C), raw fastq files were

processed through Cutadapt, an adaptor trimming software[131]. After trimming the reads, the suspected adaptor sequences were removed.

Spliced Transcripts Alignment to a Reference (STAR) was chosen as the alignment tool because it performs well with 3' mRNA libraries and was recommended by the manufacturer of the library prep kit[132]. STAR aligned the trimmed fastq files to the current mouse genome assembly, GRCm38[133], using the GENCODE VM23 annotation[133] to produce gene count tables and bam/bai files. Read alignment of *Tet2* was visualized using Integrative Genomics Viewer (IGV)[134]. Read coverage was dense near the 3' end of the transcript, as expected (Appendix Figure S7A). In *Tet2*-KO mice, exon 3 was predicted to be excised and no reads aligned to exon 3 (Appendix Figure S7B).

3.11 Data pre-processing

Prior to analysis, data was inspected for sequencing depth, skewness of distribution, Spearman correlation, outliers, and batch effects (Appendix Figure S8). The libraries had different sequencing depths and the data had a prominent positive skew, but this was not an issue since the DESeq2 analysis program contains algorithms that adjust for these factors[135,136]. One library was flagged as an outlier for having a relatively low Spearman correlation compared to the other samples (Appendix Figure S8D). This sample had poor amplification during the library generation step, so it was excluded from analysis since the difference could be due to a technical issue. No other outliers or batch effects were identified, so the 22 remaining samples were passed to DESeq2 for differential gene expression analysis.

3.12 Differential gene expression analysis

Differential gene expression and pathway enrichment analyses were performed using DESeq2[136] and g:Profiler respectively[137]. Four comparisons were made through DESeq2:

1. Comparing *Tet2*-KO vs. *Tet2*-FF neutrophils while adjusting for the treatment group as a covariate.
2. Subgroup analysis focusing on differences between LTA-treated *Tet2*-KO and *Tet2*-FF neutrophils.
3. Subgroup analysis focusing on differences between "baseline" *Tet2*-KO and *Tet2*-FF neutrophils.
4. Comparing LTA-treated and vehicle-treated neutrophils to verify LTA-induced activation.

For each comparison, lists of significantly upregulated and downregulated transcripts were generated and passed to g:Profiler for functional enrichment analysis[137]. Benjamini-Hochberg multiple testing correction[138] was applied using a significance threshold of adjusted p-value (padj) < 0.10. Differential expression results were visualized using volcano plots. Scripts used can be found in Appendix D.

Chapter 4: CHIP subjects show signs of dysregulated neutrophil granule biology

4.1 MPO and LTF mRNA expression in CHIP and No CHIP subjects

MPO and *LTF* expression, as measured by qRT-PCR, did not differ significantly between CHIP (n = 14) and No CHIP (n = 4) subjects (Figure 8). However, stratifying the CHIP group by mutation subtype (*TET2*, *DNMT3A*, or Other CHIP) revealed that *LTF* expression was significantly higher in the Other CHIP subgroup than the No CHIP controls (p = 0.016, Mann-Whitney) (Figure 8D). Potential confounding variables, age and absolute neutrophil count, were investigated (Appendix Table S1). They were similar between the Other CHIP and No CHIP groups suggesting that they did not confound the association. Although there was a significant difference in one of the subgroup analyses, the predominant negative result in the overall CHIP versus No CHIP comparison was unexpected as a previous bulk RNA-seq pilot study had found that *MPO* and *LTF* mRNA expression were upregulated in CHIP subjects[1]. One limitation of these qRT-PCR results was that it may have been underpowered to detect significant differences in expression given the high variability in *MPO* and *LTF* expression.

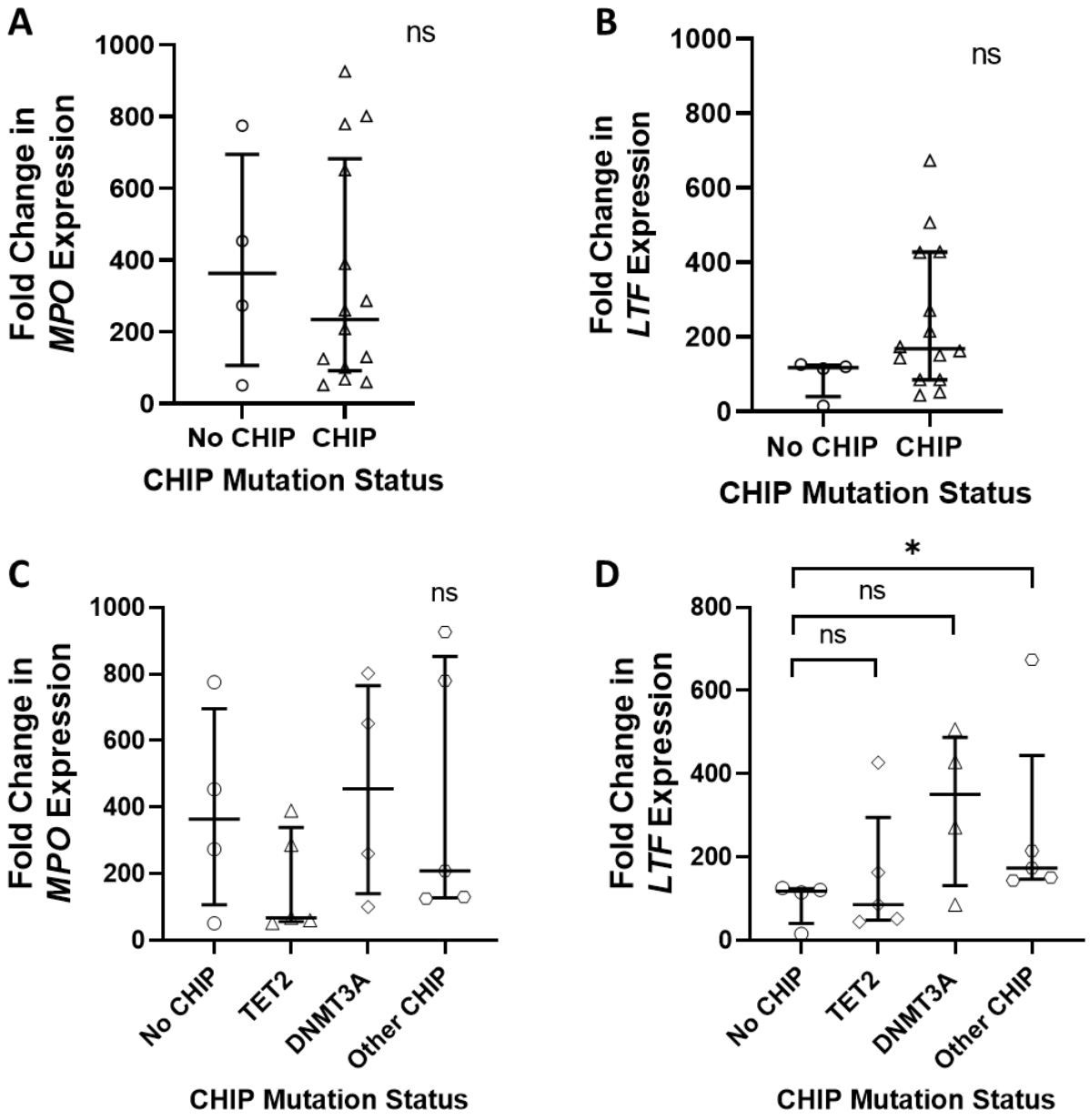


Figure 8. *MPO* and *LTF* expression in CHIP subjects and controls
qRT-PCR was performed on RNA extracted from PB samples collected from 4 No CHIP, 5 *TET2*, 4 *DNMT3A*, and 5 Other CHIP subjects. *GAPDH* and *ACTB* were used as housekeeping genes. Fold change is relative to expression in the THP-1 cell line. **A)** There was no significant difference in *MPO* expression between the CHIP (n=14) and No CHIP (n=4) group. **B)** *LTF* expression did not differ significantly between CHIP subjects and controls. **C)** There were no significant differences in *MPO* expression between the CHIP subgroups and the controls. **D)** Stratifying by mutation subtype, *LTF* was upregulated in the Other CHIP group ($p = 0.016$) while the *TET2* and *DNMT3A* groups did not show a significant differences in *LTF* expression compared to controls. * indicates $p < 0.05$, ns indicates not statistically significant, Mann-Whitney.

4.2 Serum protein titers of MPO and LTF were elevated in CHIP subjects

While *MPO* and *LTF* mRNA expression did not differ significantly, MPO and LTF serum protein titers were significantly higher in CHIP subjects (n = 51) compared to No CHIP controls (n = 13) (p = 0.006 and p = 0.001 respectively, Mann-Whitney) (Figure 9). Median MPO titers were approximately 3 times higher in CHIP subjects than controls while median LTF titers were 2.5 times higher. Stratifying by mutation subtype, *TET2* and Other CHIP subjects had significantly higher MPO serum protein titers than No CHIP controls (p = 0.007 and p = 0.017 respectively, Mann-Whitney), while *DNMT3A* subjects showed a trend approaching significance (p = 0.057, Mann-Whitney) (Figure 9C). Each CHIP subgroup had significantly higher LTF serum protein titers than No CHIP controls: *TET2* vs. No CHIP, p = 0.010, *DNMT3A* vs. No CHIP, p = 0.009, and Other CHIP vs. No CHIP, p = 0.002 (Mann-Whitney) (Figure 9D). Median age and absolute neutrophil count were similar among all groups, thus it was unlikely that they explained the observed difference. The results showed that CHIP mutations were associated with increased serum protein titers of MPO and LTF, but the underlying mechanism was unclear.

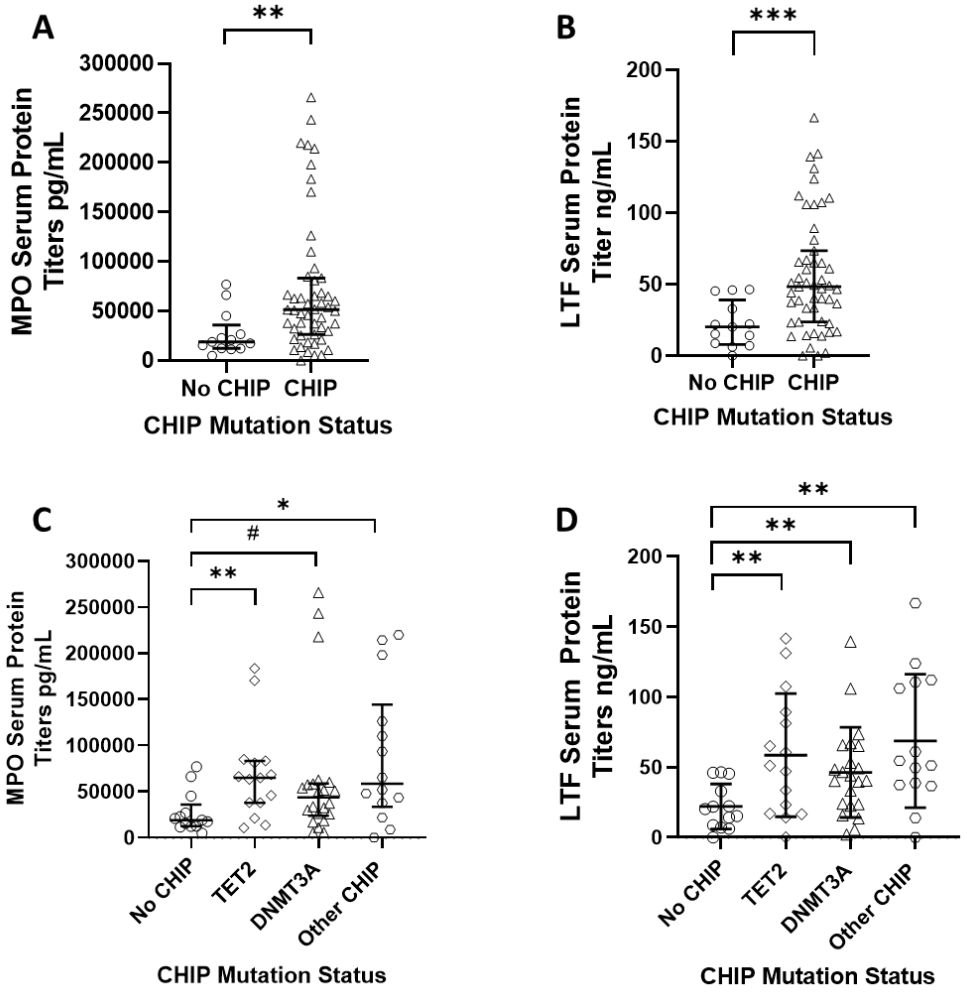


Figure 9. CHIP subjects had greater MPO and LTF serum protein titers than controls
MPO and LTF ELISA was performed on serum samples from 13 No CHIP, 15 *TET2*, 22 *DNMT3A* and 14 Other CHIP subjects. Samples that had titers below the threshold of detection were assigned a protein titer of zero. **A)** MPO titers were significantly higher in the CHIP subjects (n=51) compared to controls (n=11) (p=0.006). **B)** LTF titers were significantly higher in CHIP subjects compared to controls (p=0.001). **C)** Stratifying by CHIP mutation subtype, we found that MPO titers were significantly higher in *TET2*-CHIP and Other CHIP subjects compared to controls (p=0.007 and p=0.017 respectively) while *DNMT3A*-CHIP subjects showed a trend approaching significance (p=0.057) **D)** LTF titers were significantly higher in all CHIP subgroups compared to controls. # indicates $p < 0.10$, * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p \leq 0.001$, Mann-Whitney test.

Linear regression analysis indicated that there was no significant correlation between the mRNA expression and serum protein titers of MPO or LTF (Appendix Figure S9). This was consistent with the "targeting by timing" principle as azurophilic and specific granule proteins

like MPO and LTF were predicted to be synthesized early in development and mature PB neutrophils were not expected to have high *MPO* and *LTF* mRNA expression[88]. In fact, a study profiling the neutrophil proteome and transcriptome found that mRNA expression did not correlate with protein expression of neutrophil granule proteins in mature neutrophils[79]. Furthermore, there were many steps between mRNA transcription and release of granule proteins into PB, thus there could be other factors (i.e. protein stability or granule release) that affect serum protein titers.

Interestingly, there was a positive correlation between LTF serum protein titers and CHIP VAF (n = 51) (Figure 10). Although the correlation was weak ($R^2 = 0.095$), it suggested a dose-dependent effect. Subjects with high MPO or LTF serum protein titers tended to have higher mutation burdens, either carrying multiple CHIP mutations or having a CHIP clone with VAF > 0.10. These findings are consistent with reports that higher CHIP mutation burden associates with more severe functional dysregulation[2,17].

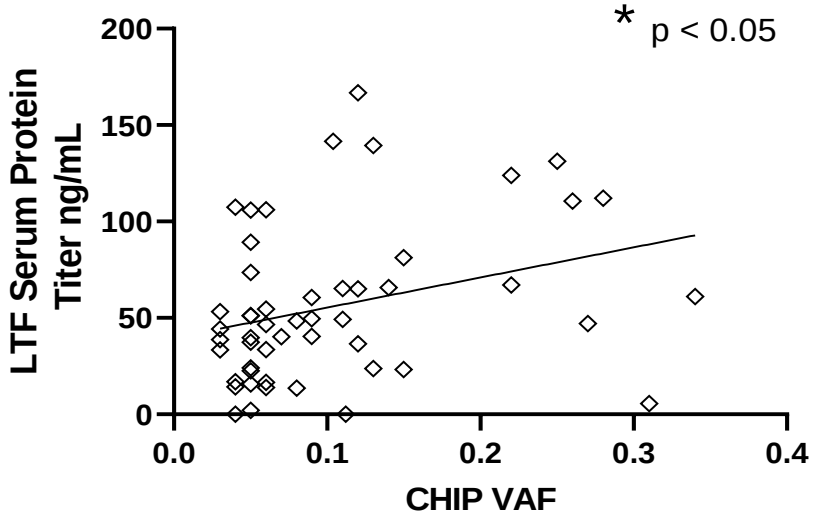


Figure 10. LTF serum protein titers correlated with CHIP clone VAF
51 samples were included in this linear regression model comparing serum protein titers and CHIP VAF. LTF serum protein titers were positively correlated with CHIP VAF ($p = 0.028$). The line of best fit was $Y = 68.7X + 9.03$ ($R^2 = 0.095$).

Taken together, the results suggested that CHIP mutations associated with upregulation of MPO and LTF serum protein titers, although the mechanism remains unknown. These elevated serum protein titers might have contributed to comorbidities associated with CHIP. High MPO serum protein titer is a well-established risk factor for cardiovascular disease and has been used clinically to stratify risk[33,34,139]. Individuals with MPO serum titers above 480 pmol/L (equivalently 72000 pg/mL) were considered high risk for developing cardiovascular comorbidities[34,139]; 27% of CHIP subjects and 8% of No CHIP controls in our cohort met this criterion. Other studies suggested that over-abundance of MPO associates with inflammatory diseases including certain types of vasculitis[86,140]. A potential explanation for why high MPO levels are harmful and associate with many comorbidities is that MPO can trigger an activation cascade: MPO generates reactive oxygen species that activate surrounding neutrophils which then release more MPO creating a cycle that amplifies neutrophil activation, inflammation, and MPO release[33,141,142]. It would be informative to investigate whether high serum protein titers of MPO associated with CHIP exacerbate cardiovascular and inflammatory comorbidities.

One limitation of this cross-sectional study was that disease progression and the development of comorbidities could not be assessed. Since the study only provided a snapshot of the population, subtle effects or those that occur over a long period of time may not have been captured. A potential future direction would be to follow up with the original participants and assess their health status.

Although it was hypothesized that epigenetic dysregulation in neutrophils could be a potential cause of elevated MPO and LTF serum protein titers, differential methylation and hydroxy-methylation has not been confirmed experimentally. One approach could be to use bisulfite sequencing or methylated DNA immunoprecipitation (MeDIP) to quantify and compare

5-methylcytosine (5-mC) or 5-hydroxymethylcytosine (5-hmC) at MPO and LTF promoter and enhancer regions in CHIP subjects and controls. It may also be informative to analyze promoter and enhancer elements associated with C/EBPα, GFI-1, and C/EBPε since they are transcription factors regulating MPO and LTF expression[83,84,93,94]. Furthermore, it was assumed that MPO and LTF in the serum predominantly originated from neutrophils, but macrophages are also capable of releasing MPO and LTF[143,144]. It was not clear what proportion of serum MPO and LTF comes from neutrophils compared to other sources. In the future, serum titers could be evaluated in a mouse model where the *TET2* inactivation is limited to neutrophils.

One strength of this study was that it included analyses on human samples while the majority of studies on CHIP have been done in animal models. The finding that serum protein titers of neutrophil granule proteins are upregulated in CHIP subjects is novel and could have implications on cardiovascular and inflammatory comorbidities. However, independent validation in larger cohorts would be beneficial.

Chapter 5: *Tet2*-deficient neutrophils show impaired motility and phagocytosis

5.1 Isolation and quantification of mouse neutrophils

After processing bone marrow cells with the neutrophil isolation kit, the enriched suspension consisted of approximately 70% mature neutrophils ($Cd11b^{hi}$ $Ly6g^{hi}$) (Figure 11). Extending from the mature neutrophil cluster was a population with high CD11b and a gradient of Ly6G representing 20% of the population (Figure 11C), these were predicted to be neutrophil precursors in varying stages of differentiation[145]. Viability was assessed using flow cytometry, and it was found that 85% of the neutrophils were viable (Annexin-V^{low} and PI^{low}), with 10% and 5% in early and late apoptosis respectively (Figure 12).

Absolute counts and relative proportions of mature neutrophils isolated from the bone marrow of twelve age and sex matched pairs of *Tet2*-KO and *Tet2*-FF mice were collected. *Tet2*-KO mice had higher absolute counts of mature bone marrow neutrophils ($p = 0.015$, paired t-test) (Figure 13A) compared to matched *Tet2*-FF mice. *Tet2*-KO mice also had a greater proportion of mature neutrophils in their bone marrow compared to *Tet2*-FF mice ($p = 0.015$, paired t-test) (Figure 13B). *Tet2* inactivating mutations are known to promote myeloid expansion and these results were consistent with the CMML-like phenotype previously reported in *Tet2*-deficient mice[5,103].

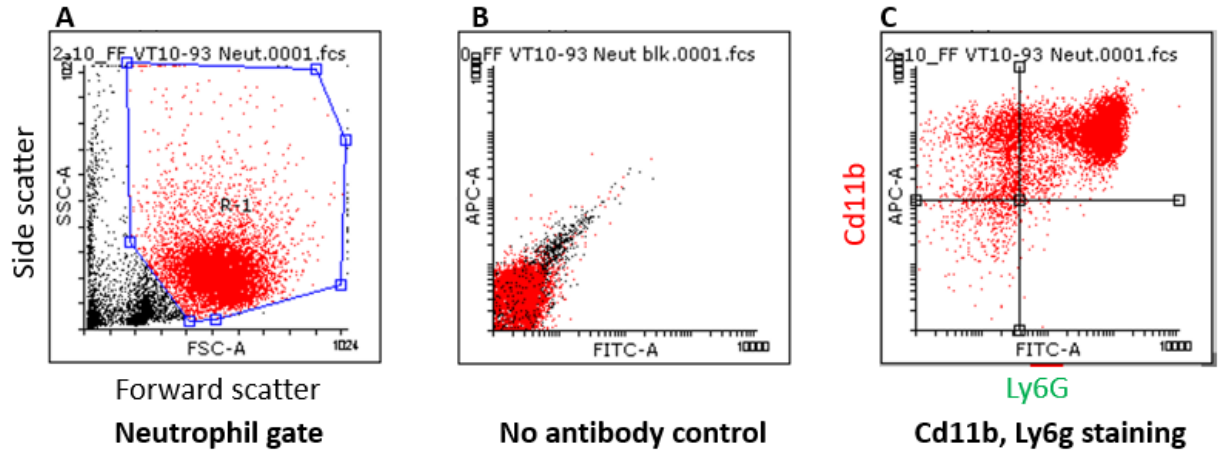


Figure 11. Representative Cd11b Ly6g flow cytometry data
A) We gate for granulocytes based on their predicted forward scatter and side scatter profile. Cells within this gate are coloured red. **B)** The level of background noise was determined by analyzing a sample that has not been stained with antibodies. **C)** Only cells within the granulocyte size gate (red) are included in fluorescence analysis. Allophycocyanin-labeled Cd11b antibodies were used together with fluorescein isothiocyanate-labeled Ly6G antibodies. Mature neutrophils are Cd11b^{hi} Ly6g^{hi} and are represented in the upper right quadrant of the scatter plot. There is a tail coming from the upper right quadrant population and these likely represent neutrophil progenitors that are Cd11b^{hi}, but show a gradient of increasing Ly6G depending on their differentiation state. In this example, 72% of the population are predicted to be mature neutrophils, 20% are in Cd11b^{hi,} Ly6g^{low} tail (likely neutrophil precursors), and 8% are Cd11b^{low}.

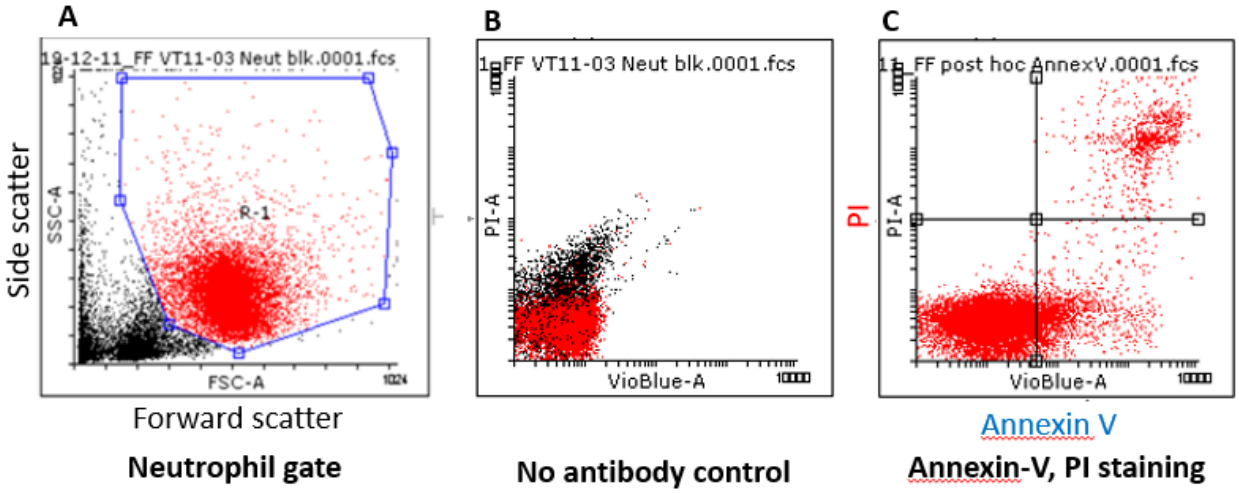


Figure 12. Representative Annexin-V PI flow cytometry data
A) We gate for granulocytes based on their predicted forward scatter and side scatter profile. Cells within this gate are coloured red. **B)** The level of background noise is determined by analyzing a sample that has not been stained with antibodies. **C)** Only cells within the granulocyte size gate (red) are included in fluorescence analysis. CF-blue fluorescence representing Annexin-V is on the x-axis and PI fluorescence is shown on the y-axis. In this example 83.7% of the cells are viable (Annexin-V^{low} and PI^{low}), 10.6% are early apoptotic (Annexin-V^{high} and PI^{low}), and 5.8% are late apoptotic (Annexin-V^{high} and PI^{high}).

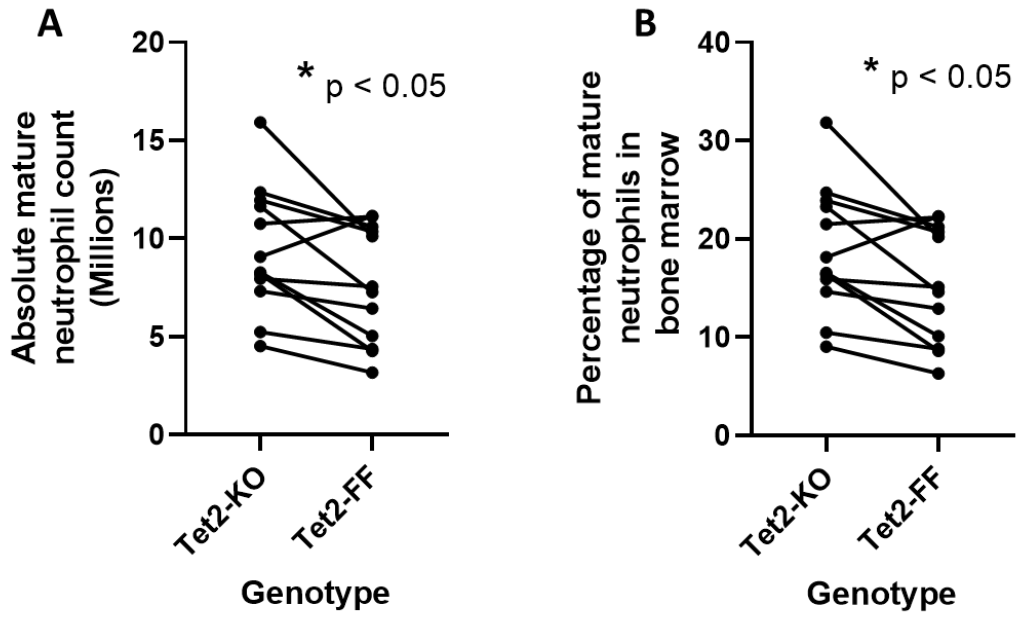


Figure 13. Absolute counts and relative proportions of mature neutrophils isolated from age and sex matched mice pairs
Graphs represent data from 12 matched pairs of *Tet2*-FF and *Tet2*-KO mice **A)** Absolute counts (millions) of mature neutrophils isolated from *Tet2*-FF and *Tet2*-KO mice are presented. *Tet2*-KO mice have significantly more mature neutrophils than *Tet2*-FF mice (p=0.015, paired t-test). **B)** Percentage of mature neutrophils in the bone marrow of *Tet2*-FF and *Tet2*-KO mice are presented. *Tet2*-KO mice have a significantly greater percentage of neutrophils in their bone marrow (p=0.015, paired t-test).

5.2 Tet2-deficient neutrophils showed impaired motility and phagocytosis compared to controls

Six independent *in vitro* phagocytosis experiments were performed (Appendix Figures S9 and S10) and they showed that neutrophils from *Tet2*-KO mice had significantly lower motility (p = 0.004, t-test) (Figure 14) and phagocytic function (p=0.034, t-test) (Figure 15) than neutrophils from *Tet2*-FF mice. Viability was investigated as a potential confounding factor, but early and late apoptotic rates were similar in *Tet2*-KO and *Tet2*-FF neutrophils (Appendix Figure S4), so this was unlikely to explain the difference. The findings suggested that *Tet2*-deficiency

associated with functional defects in mouse neutrophils. Given the importance of neutrophils in innate immunity, *Tet2*-deficient hosts may have increased susceptibility to infection.

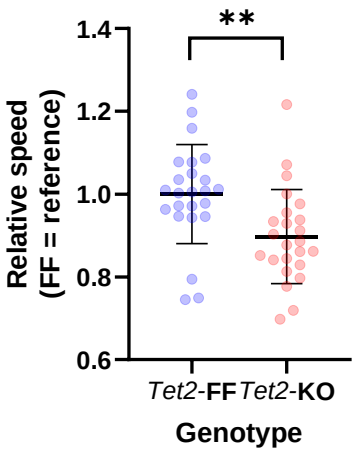


Figure 14. Neutrophils from *Tet2*-KO mice showed reduced motility compared to controls.
There was a significant difference in motility between neutrophils from *Tet2*-FF and *Tet2*-KO mice ($p = 0.004$, t-test). This represents the data from six independent experiments, mean with standard deviation is displayed. ** indicates $p < 0.01$.

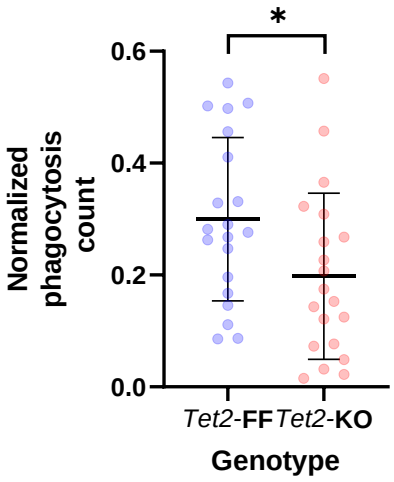


Figure 15. Neutrophils from *Tet2*-KO mice showed impaired phagocytosis compared to controls
There was a significant difference in phagocytosis between neutrophils from *Tet2*-FF and *Tet2*-KO mice ($p = 0.034$, t-test). This represents the data from six independent experiments, mean with standard deviation is displayed. * indicates $p < 0.05$.

Notably, a similar neutrophil dysfunction phenotype has been reported in MDS, a malignancy where *TET2* is recurrently mutated[27,28]. Previous studies demonstrated that

neutrophils from MDS patients have impaired migration[28] and microbicidal activity[27], and have suggested that these defects contribute to the increased risk of bacterial infections. Based on the results and previous literature, we hypothesized that *TET2*-CHIP patients might also have increased susceptibility to infection.

However, a caveat to these results is that phagocytosis, but not bactericidal activity was measured. First, the killing of phagocytosed bacteria was not confirmed. Second, neutrophils can kill bacteria without phagocytosing them through mechanisms such as degranulation and NETosis[110]. It raises the possibility that these alternative microbicidal mechanisms could compensate for defects in phagocytosis and motility. Therefore it would be informative to assess bactericidal activity in addition to phagocytosis in subsequent studies.

It is not clear whether *TET2*-deficiency would associate with increased risk of infection as there is evidence for and against this prediction. On the one hand, impaired phagocytosis and motility are predicted to increase infection risk, as seen in MDS patients[27]. On the other hand, the hyper-inflammatory phenotype associated with *TET2*-CHIP[17,53] could protect against infection as the results of the CANTOS trial suggested[76]. It would be informative for further epidemiological studies to investigate the effect that *TET2*-CHIP has on infection risk.

If there is indeed an increased risk of infection due to the impaired ability to clear pathogens, then it could be a factor driving clonal expansion and potentially disease progression. A recent study demonstrated that infection and microbial stimuli could accelerate clonal expansion in *Tet2*-deficient mice[103]. Those mice have a dysregulated emergency granulopoiesis response characterized by hyper-sensitivity to microbial stimuli, rampant inflammation and excessive proliferation of granulocyte-monocyte precursors[103], thus facilitating the rapid expansion of the *Tet2* clone. As clone size increases, the risk of acquiring subsequent oncogenic

mutations, such as neuroblastoma RAS viral oncogene homolog (*NRAS)* or Fms related receptor tyrosine kinase (*FLT3*), through stochastic processes is predicted to increase proportionally, as does the risk of malignant transformation[7]. Furthermore microbial-driven inflammation would exert strong selective pressure on the bone marrow environment exacerbating hematopoietic dysregulation. While wildtype HSCs would undergo apoptosis, *TET2*-mutant HSCs which are resistant to inflammatory stress, would proliferate and take their place[12,44]. This would lead to an increased mutational burden which can also promote leukemogenesis.

Chapter 6: *Tet2*-deficiency affects gene expression in neutrophils

6.1 Neutrophil samples pass viability, purity and quality checks

Flow cytometry for Cd11b and Ly6g showed that on average, samples contained 80% mature neutrophils ($Cd11b^{high}$ $Ly6g^{high}$) (Appendix Table S9). H & E staining confirmed that the majority of the population were mature neutrophils with some neutrophil precursors (Appendix Figure S12). Approximately 85% of the neutrophils in the baseline group were viable (Annexin-V^{low} PI^{low}), while about 80% of neutrophils in the vehicle and LTA groups were viable (Appendix Table S9). All samples had a RNA integrity number (RIN) between 9 and 10 indicating that the samples were excellent quality with minimal or no degradation (Appendix Figure S13).

6.2 Tet2 inactivation and LTA-induced activation were successful

Tet2 expression was quantified through qRT-PCR and inactivation (excision of exon 3) was confirmed in *Tet2*-KO samples (Appendix Figure S14). LTA treated neutrophils showed upregulation of *Tnfa* mRNA expression and increased Cd11b surface expression suggesting that the neutrophils were activated (Appendix Figures S15 and S16). Gene expression between neutrophils treated with vehicle and LTA was compared in DESeq2 and inflammatory genes including *Il1a, Il23a, Cxcl1* and *Cxcl2* were upregulated in LTA treated neutrophils (Appendix Figure S17). Functional enrichment analysis on g:Profiler[137], found that gene ontology terms related to the immune response to bacteria including “defense response”, “inflammatory response”, “response to molecule of bacterial origin”, and “response to bacterium” were enriched in LTA treated neutrophils compared to vehicle treated neutrophils. Furthermore, genes related to apoptosis and NF-κB signalling were significantly upregulated in LTA-treated neutrophils, in

agreement with previous findings analyzing neutrophil activation[123,127,146]. Taken together, the results confirm *Tet2* inactivation and provide evidence of LTA-induced neutrophil activation.

6.3 Tet2-deficient neutrophils show dysregulation of genes related to motility, anti-viral immunity, and neutrophil development

Comparison of *Tet2*-KO and *Tet2*-FF neutrophils in DESeq2 identified 139 differentially expressed transcripts (Appendix Table S10), 40 were significantly upregulated in *Tet2*-KO neutrophils and 99 were significantly downregulated (Figure 16). The predominant effect of *Tet2* inactivation was downregulation of gene expression which was consistent with the predicted effects of Tet2 loss-of-function[31].

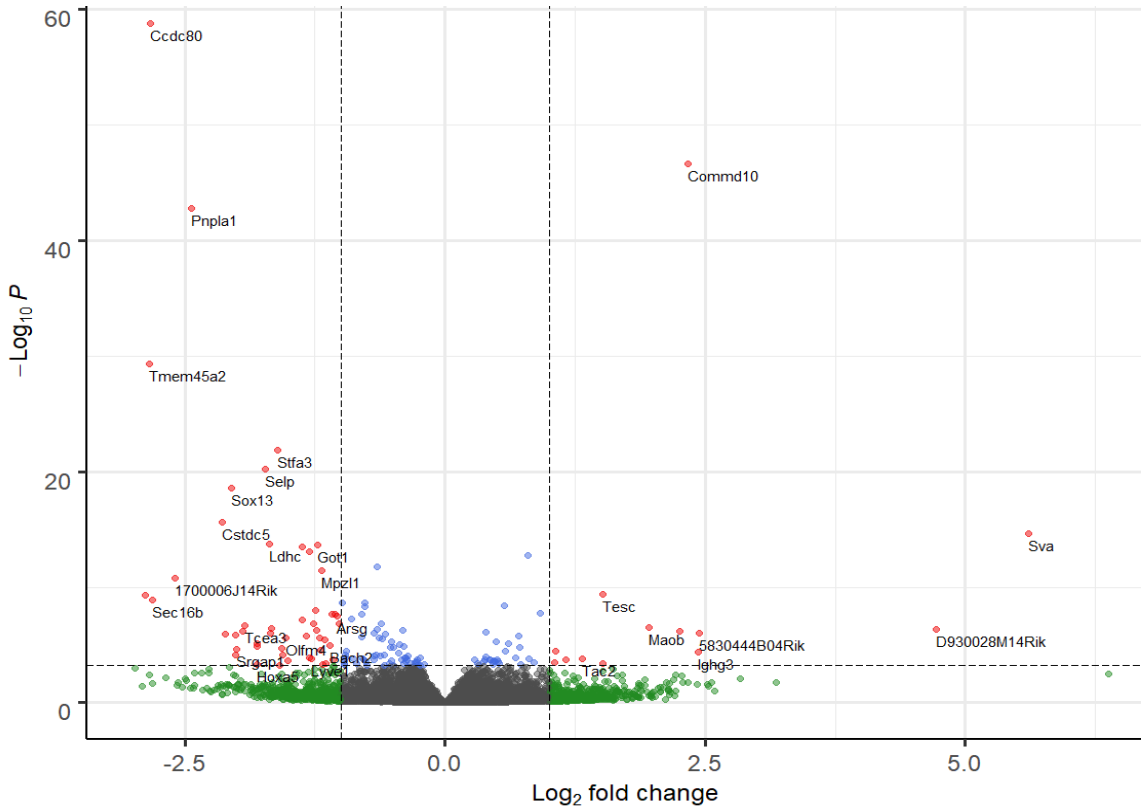


Figure 16. Volcano plot of differentially expressed transcripts between *Tet2*-KO and *Tet2*-FF neutrophils
The threshold of significance was padj < 0.05, and transcripts with an absolute log2 fold change value above 1 were considered to have a large fold change. Transcripts meeting both the significance and fold change thresholds are marked in red, those that only meet the significance threshold are blue, those that only meet the fold change threshold are in green and the rest that meet neither threshold are in grey.

Functional enrichment analysis found that gene ontology terms related to anti-viral immunity such as “response to virus” and “regulation of viral life cycle” were enriched in *Tet2*-KO neutrophils compared to *Tet2*-FF controls. Many genes related to the interferon response, including 2’-5’-oligoadenylate synthetase 3 (*Oas3*), adenosine deaminase RNA specific (*Adar*), interferon induced transmembrane protein 1 (*Ifitm1*), and interferon gamma induced GTPase (*Igtp*)[147], were significantly upregulated in *Tet2*-KO neutrophils. The interferon gene expression signature was unexpected because the stimulus was a bacterial molecule not related to viruses. Interestingly, the anti-viral signature in *Tet2*-KO neutrophils resembles the interferon signature previously reported in MDS patients[148,149]. Since *TET2* is recurrently mutated in MDS, this supports the hypothesis that *Tet2*-deficiency is associated with a dysregulated interferon response.

Consistent with the live cell imaging results, gene ontology terms related to motility were significantly downregulated in *Tet2*-KO neutrophils. Notably, mRNA expression of *Cxcr1*, a chemokine receptor with an important role in neutrophil chemotaxis[150], was more than two times lower in *Tet2*-KO neutrophils than controls (Figure 17). Genes predicted to be involved in actin cytoskeleton rearrangement including Ras guanine exchange factor domain-containing family member 1A (*Rasgef1a*), Cas scaffold protein family member 4 (*Cass4*), Fps/Fes related tyrosine kinase (*Fer*), and protein kinase N2 (*Pkn2*), were also downregulated in *Tet2*-KO neutrophils. Taken together these findings suggest that dysregulation of chemotaxis and cytoskeleton dynamics may contribute to the impaired motility phenotype.

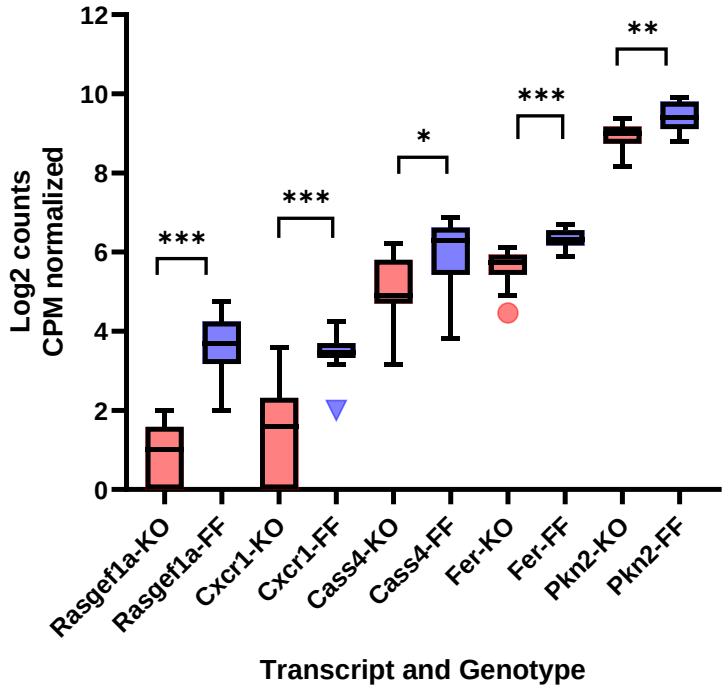


Figure 17. *Tet2*-KO neutrophils showed decreased expression of genes related to motility
Box and whisker plots are used to visualize the expression of five genes related to motility. All five genes showed significantly lower expression in *Tet2*-KO neutrophils compared to *Tet2*-FF neutrophils. Red plots represent expression in *Tet2*-KO samples while blue plots represent expression in *Tet2*-FF samples. Gene counts were normalized through the counts per million method that adjusts for sequencing depth. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$ (t-test).

Genes related to neutrophil development were dysregulated in *Tet2*-KO neutrophils (Figure 18) and this may have contributed to the CMML-like phenotype in *Tet2*-KO mice[5]. *Tet2*-KO neutrophils showed greater Tescalcin (*Tesc)* expression than controls, which is predicted to favour granulocytic differentiation[151]. Furthermore, three members of the homeobox A (*Hoxa*) gene cluster, *Hoxa5*, *Hoxa7*, and *Hoxa9*, were downregulated in *Tet2*-KO neutrophils. A previous study demonstrated that *HoxA*-cluster-haplosufficient mice have increased numbers of neutrophils in their bone marrow and spleen[152] Overall, it appears that the transcriptional changes associated with *Tet2*-deficiency favours the expansion of neutrophils.

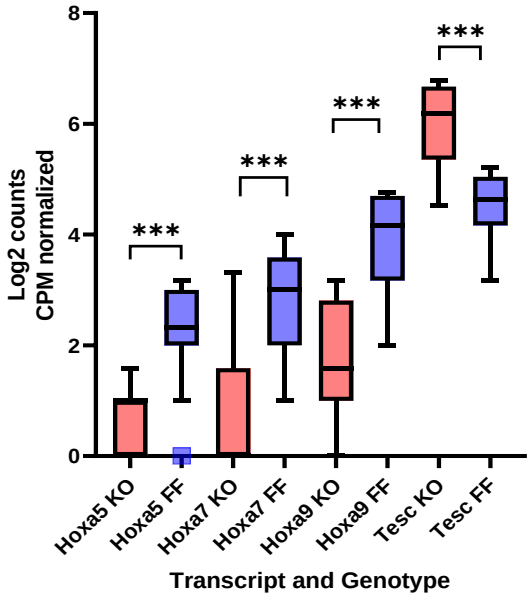


Figure 18. Genes related to neutrophil development are dysregulated in *Tet2*-KO neutrophils
Box and whisker plots are used to visualize the expression of four genes related to neutrophil differentiation. *HoxA* cluster genes were significantly downregulated in *Tet2*-KO neutrophils while *Tesc* was upregulated in *Tet2*-KO neutrophils. Red plots represent expression in *Tet2*-KO samples while blue plots represent expression in *Tet2*-FF samples. Gene counts were normalized through the counts per million method that adjusts for sequencing depth. *** indicates $p < 0.001$ (t-test).

Although we found *Hoxa* genes to be downregulated in *Tet2*-KO neutrophils, *Hoxa* genes are upregulated in myeloid malignancies such as AML[153-155]. The literature suggests that over-expression of *Hoxa* genes in HSCs impairs myeloid and erythroid differentiation and increases production of blasts[155-157]. While it may be predicted that *Hoxa* gene expression should be upregulated in CHIP as it is in AML, one should also expect differences between CHIP and AML since the former is an early disease state while the latter is an advanced disease. In particular, the blast phenotype associated with *Hoxa* over-expression is consistent with aggressive malignancies like AML, but is inconsistent with CHIP.

The finding that transcripts not commonly associated with neutrophils, such as *Sva* and *Ighg3,* were upregulated in *Tet2*-KO neutrophils was unexpected. While the upregulation of immunoglobulin genes may have been due to the presence of lymphoid precursors, it could also indicate lineage infidelity in *Tet2*-KO neutrophils which is plausible given the dysregulation of genes related to the determination of cell fate, such as SRY-box transcription factor (*Sox*) 13, *Sox4*, and the *Hoxa* gene cluster. As inactivating *TET2* mutations are associated with differentiation block in hematopoietic cells[2], it would be informative to explore whether *Tet2*-inactivation is associated with an immature phenotype in neutrophils.

6.4 Tet2-inactivation does not have a major effect on LTA-induced gene expression

Subgroup analysis comparing LTA-treated *Tet2*-KO and *Tet2*-FF neutrophils found 10 significantly downregulated (padj < 0.05) and three significantly upregulated transcripts (Figure 19). It was surprising that so few differentially expressed transcripts were identified given the significant functional differences observed in the phagocytosis assay. Since the comparison only involved four samples in each group, it may have been underpowered to detect significant differences in gene expression. Alternatively, the results may indicate that *Tet2* does not play a major role in regulating the expression of genes induced by LTA. It may be beneficial to perform follow up qRT-PCR experiments assessing expression of genes related to motility in LTA-treated *Tet2*-KO neutrophils and controls to clarify the effects of *Tet2*-deficiency on LTA-induced gene expression.

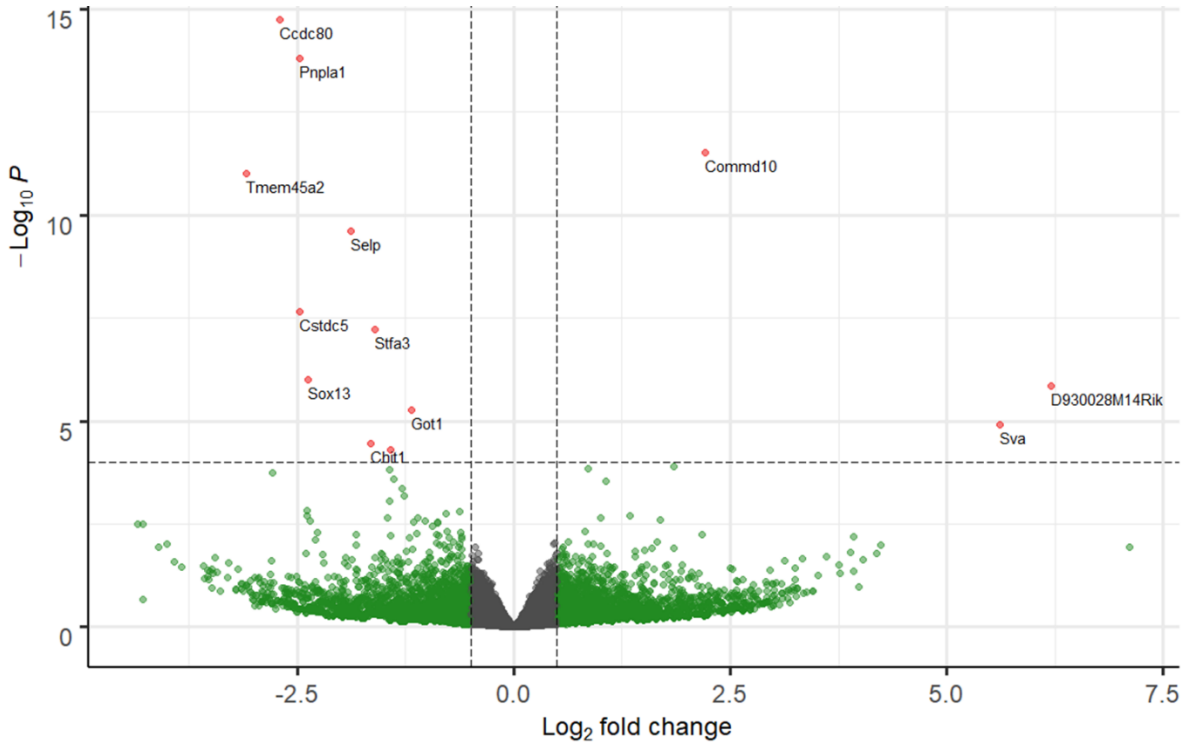


Figure 19. Volcano plot of differentially expressed transcripts between LTA-treated *Tet2*-KO and *Tet2*-FF neutrophils
The threshold of significance was padj < 0.05, and transcripts with an absolute log2 fold change value above 1 were considered to have a large fold change. Transcripts meeting both the significance and fold change thresholds are marked in red, those that only meet the fold change threshold are in green and the rest that meet neither threshold are in grey.

6.5 Tet2-deficient neutrophils at baseline show downregulation of genes associated with motility and adhesion

Subgroup analysis comparing *Tet2*-KO and *Tet2*-FF neutrophils at baseline found 25 significantly downregulated (padj < 0.05) and five significantly upregulated genes (Figure 20). Many of the downregulated genes were associated with motility or adhesion; *Rasgef1a*, CDC42 effector protein 2 (*Cdc42ep2*), and dedicator of cytokinesis 9 (*Dock9*) have predicted involvement in actin cytoskeleton rearrangement, while olfactomedin 4 (*Olfm4*), and thrombospondin 1 (*Thbs1*) have predicted roles in adhesion[158]. Perhaps *Tet2*-deficient neutrophils have a lower baseline level of motility compared to wildtype neutrophils due to

constitutive defects. To test this hypothesis, a potential future direction would be to perform migration and motility assays on neutrophils at baseline or with minimal stimulation.

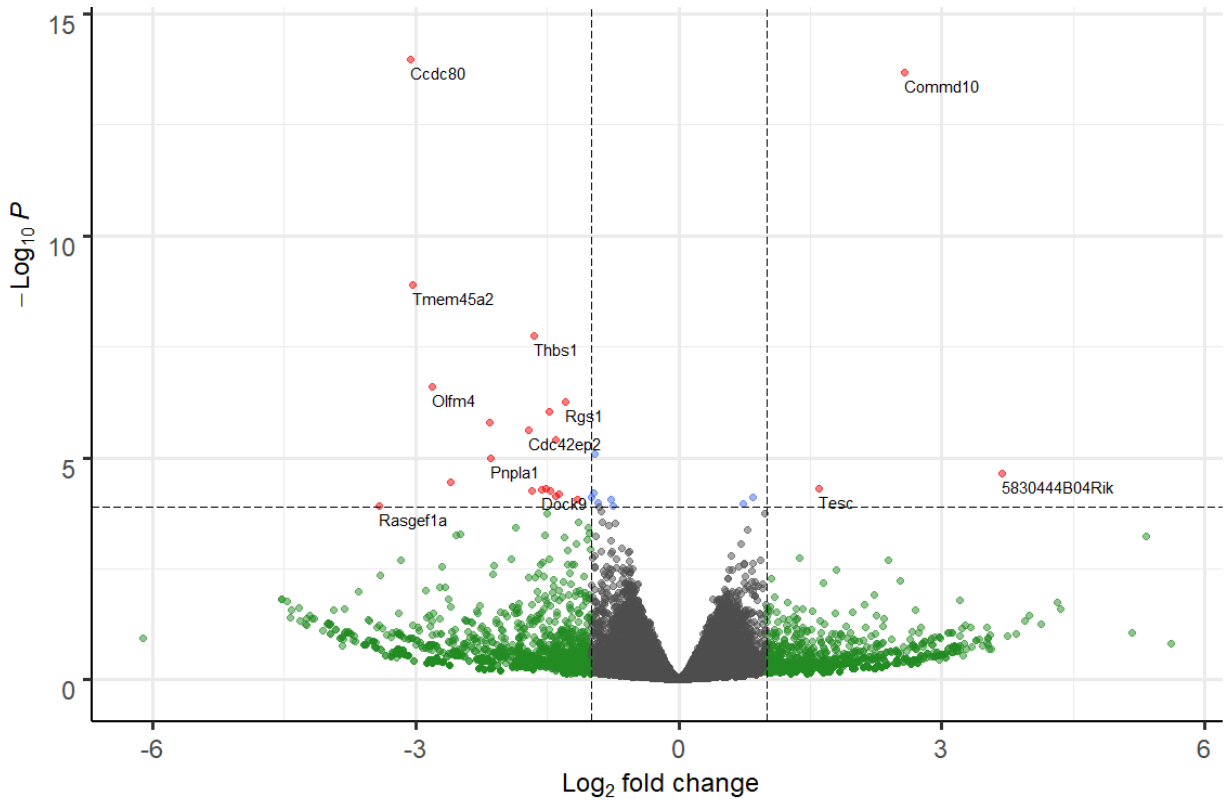


Figure 20. Volcano plot of differentially expressed genes between *Tet2*-FF and *Tet2*-KO neutrophils at baseline
Many genes associated with adhesion and motility were downregulated in *Tet2*-KO neutrophils. The threshold of significance was padj < 0.05, and transcripts with an absolute log2 fold change value above 1 were considered to have a large fold change. Transcripts meeting both the significance and fold change thresholds are marked in red, those that only meet the significance threshold are blue, those that only meet the fold change threshold are in green and the rest that meet neither threshold are in grey.

6.6 Strengths and limitations of the bulk RNA-seq analysis

The effects of *Tet2*-deficiency on the neutrophil transcriptome have not yet been characterized and this study addressed a gap in the literature by investigating gene expression differences between *Tet2*-KO and *Tet2*-FF neutrophils at baseline and after activation with LTA.

While the literature has focused on pathological gain-of-function phenotypes associated with *Tet2*-deficiency, such as increased self-renewal in HSCs[5,43] and the hyperinflammatory phenotype in macrophages[17], this study shows that it may also associate with defects in neutrophil function.

A limitation is that the differentially expressed genes identified by RNA-seq have not yet been validated. Expression of key genes involved in motility (*Cxcr1*, P-selectin (*Selp*), *Cass4*, *Fer*), neutrophil differentiation (*Hoxa5, Hoxa7, Hoxa9, Tesc)*, and the interferon response (*Ifitm1, Igtp, Oas3, Adar*) will be assessed in the future through qRT-PCR, Western blotting, and flow cytometry where applicable. Furthermore, although *Tet2* is predicted to be inactivated through excision of exon 3, loss of Tet2 enzymatic function has not been confirmed. This can be addressed through MeDIP to quantify levels of 5-hmC and 5-mC between *Tet2*-KO and *Tet2*-FF neutrophils.

Bulk RNA-seq was performed on bone marrow samples containing mostly mature neutrophils with some immature progenitors; this had advantages and disadvantages. The disadvantage with the heterogeneous population is that it is unclear whether the gene expression differences were due to the mature neutrophils or progenitor populations. However, including bone marrow progenitors facilitates detection of immature phenotypes or abnormalities in neutrophil development while a homogenous sample of mature neutrophils in the PB is unlikely to reveal subtle differences in differentiation. Our lab is planning to conduct single cell RNA-seq on *Tet2*-deficient leukocytes, which will clarify the cell-specific effects of *Tet2*-inactivation.

A limitation of the 3' poly A tail capture method used in library preparation was that miRNAs and long non-coding RNAs (lncRNA) lacking poly-A tails[159] were not captured. *Tet2* regulates the expression of certain miRNAs[160,161] and it possible that miRNAs may have

mediated some of the gene expression differences between *Tet2*-KO and *Tet2*-FF neutrophils. Hypermethylation and downregulation of miRNAs may be part of the mechanism underlying the upregulation of genes in *Tet2*-deficient neutrophils, but this will need to be investigated in subsequent miRNA sequencing experiments.

Another consideration relating to the experimental design is that the *Tet2*$^{flox/flox}$ *Vav-iCre* mouse model has a documented *Vav*-i*Cre* transgene insertion in the intron of COMM domain containing 10 (*Commd10*)[162]. Differential expression analysis identified *Commd10* as a significantly upregulated transcript in *Tet2*-KO neutrophils, but this might be a consequence of the transgene insertion. *Tet2*-KO mice have the *Vav-iCre* transgene, but *Tet2*-FF mice do not. The active *Vav1* promoter might be driving expression from the *Commd10* gene, thus the *Commd10* finding should be interpreted with caution. For this reason, *Commd10* was not a gene we considered among those plausibly affected by *Tet2*-deficiency.

Chapter 7: Summary and Future Directions

The findings of this study suggest that *Tet2*-deficiency affects neutrophil function and gene expression. CHIP subjects, including the *TET2*-CHIP subgroup, showed upregulation of MPO and LTF serum protein titers which suggests that *TET2* and other CHIP mutations may be associated with dysregulation of neutrophil granule biology. Interestingly, CHIP VAF correlated with LTF serum protein titers, implying a dose-dependent effect. The upregulation of neutrophil granule proteins, particularly MPO, was concerning because elevated serum protein titers are associated with cardiovascular and inflammatory diseases[33,34,86]. In fact, 27% of CHIP subjects have MPO serum protein titers over 72,000 pg/mL, meeting the clinical threshold for high risk of cardiovascular disease[34,139].

Elevated serum protein titers were associated with CHIP, but the underlying mechanism is unclear. One hypothesis was that enhancers and promoter elements associated with neutrophil granule proteins and transcription factors that regulate the granule genes were differentially methylated between CHIP subjects and controls. A future direction could be to perform MeDIP to assess whether the levels of 5-hmC and 5-mC differ at these regulatory elements. Another hypothesis was that the increased serum titers may reflect differences in degranulation rather than differences in expression. To test this, neutrophils isolated from CHIP subjects and controls would be treated with varying concentrations of an activating agent like fMLP to determine their threshold for degranulation. Furthermore, Western blotting and intracellular ELISA assays could be performed to determine whether there are differences in protein abundance of neutrophil granules between CHIP and No CHIP subjects. These experiments could clarify the mechanism by which CHIP mutations contribute to increased serum protein titers of neutrophil granule proteins.

In the *Tet2*$^{flox/flox}$ *Vav-iCre* mouse model, it was shown that *Tet2*-deficient neutrophils have impaired motility and phagocytic function. Complementing the functional findings, bulk-RNA-seq analysis showed that *Tet2*-KO neutrophils had decreased expression of genes related to motility. Since phagocytosis is one of the primary immune mechanisms against bacterial pathogens, these functional defects may contribute to an increased risk of bacterial infection in *TET2*-deficient hosts. However, *Tet2*-deficiency is also associated with a hyper-inflammatory phenotype[74,75] which may protect against infection; therefore the overall effect of *Tet2*-deficiency on infection risk is unclear. To clarify this, a future direction could be to compare infection severity and the rate of infection clearance between *Tet2*-KO and control mice in an *in vivo* model of infection.

The defects in motility and phagocytosis were demonstrated in a knockout mouse model and it would be beneficial to investigate whether the neutrophil dysfunction phenotype generalizes to *TET2*-CHIP in humans. Similar *in vitro* phagocytosis and motility assays can be performed on neutrophils isolated from *TET2*-CHIP subjects and controls. Another caveat is that while all hematopoietic cells are predicted to be *Tet2*-deficient in *Tet2*-KO mice, only a fraction of leukocytes in *TET2*-CHIP subjects are predicted to carry the mutation, thus the wildtype leukocytes may compensate for the loss-of-function in the *TET2*-mutant leukocytes, reducing the risk of infection. Therefore, it would also be informative to perform epidemiological analyses to assess whether *TET2*-CHIP subjects have increased susceptibility to infection.

In addition to the finding of decreased motility, RNA-seq analysis also identified dysregulation of genes related to neutrophil development and an upregulated anti-viral response signature in *Tet2*-KO neutrophils which resembled the interferon signature in MDS patients[148,149]. Given that *Tet2*-CHIP can progress to MDS, this may indicate that dysfunctional

characteristics associated with advanced disease may already be present at early stages like CHIP, which could provide insight into the process of disease progression. Subsequent experiments could analyze whether neutrophils from *TET2*-CHIP subjects show upregulation of interferon related genes. Another potential future direction could be to further explore the dysregulation of neutrophil development and whether neutrophils from *Tet2*-KO mice show lineage infidelity or signs of immaturity. Some studies have performed transcriptomic profiling of neutrophils in different stages of development[32,145], and the bulk RNA-seq results could be compared to those published profiles to evaluate whether *Tet2*-deficient neutrophils resemble immature neutrophils. To address the limitation that the gene expression differences in mature neutrophil and progenitor populations cannot be distinguished in bulk RNA-seq, single-cell RNA-seq will be performed on leukocytes from *TET2*-CHIP subjects. Furthermore, qRT-PCR, Western blot, and flow cytometry validation of RNA-seq targets will be performed.

Beyond investigating the effects of *TET2* deficiency on neutrophils, it would also be informative to characterize how *DNMT3A* deficiency affects neutrophils. Like *TET2*, *DNMT3A* is one of the most commonly mutated genes in CHIP[15,16] and given the prevalence of CHIP among the elderly, it would be important to gain a better understanding of the pathology of *DNMT3A*-CHIP. Furthermore, DNMT3A has a major role in epigenetic regulation[56], thus inactivating mutations are predicted to impact neutrophil function. One of the future directions could be to perform bulk RNA-seq comparing *Dnmt3a*-KO and control neutrophils to explore the transcriptomic effects of *Dnmt3a* deficiency.

In summary, we found that *TET2*-deficiency associated with differences in neutrophil function and gene expression. Follow-up studies are required to gain a better understanding of the underlying mechanisms of these functional changes and whether they contribute to biological

endpoints like risk of infection. Although research on CHIP has been focused on MΦ populations, we are hopeful that these findings will encourage further investigations into the role of neutrophils in CHIP pathology. As CHIP pathogenesis likely involves multiple leukocyte populations, studies on neutrophils and other leukocyte populations are necessary to gain a comprehensive understanding of the disease and thus address the "indeterminate potential" of CHIP.

References

1 Cook, E. K. *et al.* in *61st ASH Annual Meeting and Exposition.* (Blood).

2 Bowman, R. L., Busque, L. & Levine, R. L. Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell stem cell* **22**, 157-170, doi:10.1016/j.stem.2018.01.011 (2018).

3 Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16, doi:10.1182/blood-2015-03-631747 (2015).

4 Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328-333, doi:10.1038/nature13038 (2014).

5 Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer cell* **20**, 11-24, doi:10.1016/j.ccr.2011.06.001 (2011).

6 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).

7 Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2548-2553, doi:10.1073/pnas.1324297111 (2014).

8 Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nature genetics* **44**, 1179-1181, doi:10.1038/ng.2413 (2012).

9 Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *The New England journal of medicine* **360**, 2289-2301, doi:10.1056/NEJMoa0810069 (2009).

10 Cazzola, M., Della Porta, M. G. & Malcovati, L. The genetic basis of myelodysplasia and its clinical relevance. *Blood* **122**, 4021-4034, doi:10.1182/blood-2013-09-381665 (2013).

11 Naqvi, K. *et al.* Clonal hematopoiesis of indeterminate potential–associated mutations and risk of comorbidities in patients with myelodysplastic syndrome. *Cancer* **125**, 2233-2241, doi:10.1002/cncr.32056 (2019).

12 Cai, Z. *et al.* Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. *Cell stem cell* **23**, 833-849.e835, doi:10.1016/j.stem.2018.10.013 (2018).

13 Cimmino, L. *et al.* Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell* **170**, 1079-1095.e1020, doi:10.1016/j.cell.2017.07.032 (2017).

14 Buscarlet, M. *et al.* Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood* **132**, 277-280, doi:10.1182/blood-2018-01-829937 (2018).

15 Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine* **371**, 2488-2498, doi:10.1056/NEJMoa1408617 (2014).

16 Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine* **371**, 2477-2487, doi:10.1056/NEJMoa1409405 (2014).

17 Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *The New England journal of medicine* **377**, 111-121, doi:10.1056/NEJMoa1701719 (2017).

18 Fuster, J. J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science (New York, N.Y.)* **355**, 842-847, doi:10.1126/science.aag1381 (2017).
19 Acuna-Hidalgo, R. *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *American journal of human genetics* **101**, 50-64, doi:10.1016/j.ajhg.2017.05.013 (2017).
20 Sano, S. *et al.* Tet2-Mediated Clonal Hematopoiesis Accelerates Heart Failure Through a Mechanism Involving the IL-1beta/NLRP3 Inflammasome. *Journal of the American College of Cardiology* **71**, 875-886, doi:10.1016/j.jacc.2017.12.037 (2018).
21 Sano, S. *et al.* CRISPR-Mediated Gene Editing to Assess the Roles of Tet2 and Dnmt3a in Clonal Hematopoiesis and Cardiovascular Disease. *Circulation research* **123**, 335-341, doi:10.1161/circresaha.118.313225 (2018).
22 Arends, C. M. *et al.* Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32**, 1908-1919, doi:10.1038/s41375-018-0047-7 (2018).
23 Lawrence, S. M., Corriden, R. & Nizet, V. The Ontogeny of a Neutrophil: Mechanisms of Granulopoiesis and Homeostasis. *Microbiology and molecular biology reviews : MMBR* **82**, doi:10.1128/mmbr.00057-17 (2018).
24 Doring, Y., Soehnlein, O. & Weber, C. Neutrophil Extracellular Traps in Atherosclerosis and Atherothrombosis. *Circulation research* **120**, 736-743, doi:10.1161/circresaha.116.309692 (2017).
25 Franck, G. *et al.* Roles of PAD4 and NETosis in Experimental Atherosclerosis and Arterial Injury: Implications for Superficial Erosion. *Circulation research* **123**, 33-42, doi:10.1161/circresaha.117.312494 (2018).
26 Carlucci, P. M. *et al.* Neutrophil subsets and their gene signature associate with vascular inflammation and coronary atherosclerosis in lupus. *JCI insight* **3**, doi:10.1172/jci.insight.99276 (2018).
27 Fianchi, L. *et al.* Impaired bactericidal and fungicidal activities of neutrophils in patients with myelodysplastic syndrome. *Leuk Res* **36**, 331-333, doi:10.1016/j.leukres.2011.11.012 (2012).
28 Fuhler, G. M., Knol, G. J., Drayer, A. L. & Vellenga, E. Impaired interleukin-8- and GROalpha-induced phosphorylation of extracellular signal-regulated kinase result in decreased migration of neutrophils from patients with myelodysplasia. *Journal of leukocyte biology* **77**, 257-266, doi:10.1189/jlb.0504306 (2005).
29 Toma, A., Fenaux, P., Dreyfus, F. & Cordonnier, C. Infections in myelodysplastic syndromes. *Haematologica* **97**, 1459-1470, doi:10.3324/haematol.2012.063420 (2012).
30 Ali, A. M. *et al.* Patterns of infectious complications in acute myeloid leukemia and myelodysplastic syndromes patients treated with 10-day decitabine regimen. *Cancer medicine* **6**, 2814-2821, doi:10.1002/cam4.1231 (2017).
31 Rasmussen, K. D. *et al.* Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes & development* **29**, 910-922, doi:10.1101/gad.260174.115 (2015).
32 Grassi, L. *et al.* Dynamics of Transcription Regulation in Human Bone Marrow Myeloid Differentiation to Mature Blood Neutrophils. *Cell reports* **24**, 2784-2794, doi:10.1016/j.celrep.2018.08.018 (2018).

33 Khan, A. A., Alsahli, M. A. & Rahmani, A. H. Myeloperoxidase as an Active Disease Biomarker: Recent Biochemical and Pathological Perspectives. *Medical sciences (Basel, Switzerland)* **6**, doi:10.3390/medsci6020033 (2018).
34 Karakas, M. *et al.* Myeloperoxidase is associated with incident coronary heart disease independently of traditional risk factors: results from the MONICA/KORA Augsburg study. *Journal of internal medicine* **271**, 43-50, doi:10.1111/j.1365-2796.2011.02397.x (2012).
35 Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell stem cell* **10**, 120-136, doi:10.1016/j.stem.2012.01.006 (2012).
36 Buscarlet, M. *et al.* DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130**, 753-762, doi:10.1182/blood-2017-04-777029 (2017).
37 Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nature communications* **7**, 12484, doi:10.1038/ncomms12484 (2016).
38 Beerman, I., Seita, J., Inlay, M. A., Weissman, I. L. & Rossi, D. J. Quiescent hematopoietic stem cells accumulate DNA damage during aging that is repaired upon entry into cell cycle. *Cell stem cell* **15**, 37-50, doi:10.1016/j.stem.2014.04.016 (2014).
39 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
40 Osorio, F. G. *et al.* Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. *Cell reports* **25**, 2308-2316.e2304, doi:10.1016/j.celrep.2018.11.014 (2018).
41 Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V. & Chuzhanova, N. A. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Human genomics* **4**, 406-410, doi:10.1186/1479-7364-4-6-406 (2010).
42 Kuranda, K. *et al.* Age-related changes in human hematopoietic stem/progenitor cells. *Aging cell* **10**, 542-546, doi:10.1111/j.1474-9726.2011.00675.x (2011).
43 Ostrander, E. L. *et al.* Divergent Effects of Dnmt3a and Tet2 Mutations on Hematopoietic Progenitor Cell Fitness. *Stem cell reports*, doi:10.1016/j.stemcr.2020.02.011 (2020).
44 Abegunde, S. O., Buckstein, R., Wells, R. A. & Rauh, M. J. An inflammatory environment containing TNFalpha favors Tet2-mutant clonal hematopoiesis. *Experimental hematology* **59**, 60-65, doi:10.1016/j.exphem.2017.11.002 (2018).
45 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine* **363**, 2424-2433, doi:10.1056/NEJMoa1005143 (2010).
46 Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Science translational medicine* **4**, 149ra118, doi:10.1126/scitranslmed.3004315 (2012).
47 Corces, M. R., Chang, H. Y. & Majeti, R. Preleukemic Hematopoietic Stem Cells in Human Acute Myeloid Leukemia. *Frontiers in oncology* **7**, 263, doi:10.3389/fonc.2017.00263 (2017).
48 Desai, P. *et al.* Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nature medicine* **24**, 1015-1023, doi:10.1038/s41591-018-0081-z (2018).

49 De Vita, S. *et al.* Loss of function of TET2 cooperates with constitutively active KIT in murine and human models of mastocytosis. *PLoS One* **9**, e96209, doi:10.1371/journal.pone.0096209 (2014).
50 Ortmann, C. A. *et al.* Effect of mutation order on myeloproliferative neoplasms. *The New England journal of medicine* **372**, 601-612, doi:10.1056/NEJMoa1412098 (2015).
51 Potus, F. *et al.* Novel Mutations and Decreased Expression of the Epigenetic Regulator TET2 in Pulmonary Arterial Hypertension. *Circulation* **0**, doi:doi:10.1161/CIRCULATIONAHA.119.044320.
52 Wolach, O. *et al.* Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. *Science translational medicine* **10**, doi:10.1126/scitranslmed.aan8292 (2018).
53 Cook, E. K. *et al.* Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood Advances* **3**, 2482-2486, doi:10.1182/bloodadvances.2018024729 (2019).
54 Arends, C. M. *et al.* Clonal hematopoiesis in patients with ANCA-associated vasculitis. *Haematologica*, doi:10.3324/haematol.2019.223305 (2019).
55 Mas-Peiro, S. *et al.* Clonal haematopoiesis in patients with degenerative aortic valve stenosis undergoing transcatheter aortic valve implantation. *European heart journal*, doi:10.1093/eurheartj/ehz591 (2019).
56 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
57 Hon, G. C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular cell* **56**, 286-297, doi:10.1016/j.molcel.2014.08.026 (2014).
58 Holz-Schietinger, C. & Reich, N. O. De novo DNA methyltransferase DNMT3A: Regulation of oligomeric state and mechanism of action in response to pH changes. *Biochimica et biophysica acta* **1850**, 1131-1139, doi:10.1016/j.bbagen.2015.02.003 (2015).
59 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).
60 Kunimoto, H. & Nakajima, H. Epigenetic dysregulation of hematopoietic stem cells and preleukemic state. *International journal of hematology* **106**, 34-44, doi:10.1007/s12185-017-2257-6 (2017).
61 Yin, R. *et al.* Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *Journal of the American Chemical Society* **135**, 10396-10403, doi:10.1021/ja4028346 (2013).
62 Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature genetics* **44**, 23-31, doi:10.1038/ng.1009 (2011).
63 Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843, doi:10.1038/nature09586 (2010).
64 Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493**, 561-564, doi:10.1038/nature11742 (2013).

65 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343-348, doi:10.1038/nature10066 (2011).
66 Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature genetics* **46**, 17-23, doi:10.1038/ng.2836 (2014).
67 Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nature reviews. Molecular cell biology* **16**, 519-532, doi:10.1038/nrm4043 (2015).
68 Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**, 295-304, doi:10.1038/nrg2540 (2009).
69 Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770, doi:10.1038/nature07107 (2008).
70 Du, J. *et al.* Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. *Molecular cell* **55**, 495-504, doi:10.1016/j.molcel.2014.06.009 (2014).
71 Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic acids research* **31**, 2305-2312, doi:10.1093/nar/gkg332 (2003).
72 Pronier, E. *et al.* Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood* **118**, 2551-2555, doi:10.1182/blood-2010-12-324707 (2011).
73 Cook, E. K., Luo, M. & Rauh, M. J. Clonal hematopoiesis and inflammation: Partners in leukemogenesis and comorbidity. *Experimental hematology* **83**, 85-94, doi:10.1016/j.exphem.2020.01.011 (2020).
74 Cull, A. H., Snetsinger, B., Buckstein, R., Wells, R. A. & Rauh, M. J. Tet2 restrains inflammatory gene expression in macrophages. *Experimental hematology* **55**, 56-70.e13, doi:10.1016/j.exphem.2017.08.001 (2017).
75 Zhang, Q. *et al.* Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* **525**, 389+ (2015).
76 Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *The New England journal of medicine* **377**, 1119-1131, doi:10.1056/NEJMoa1707914 (2017).
77 Svensson, E. C. *et al.* Abstract 15111: TET2-Driven Clonal Hematopoiesis Predicts Enhanced Response to Canakinumab in the CANTOS Trial: An Exploratory Analysis. *Circulation* **138**, A15111-A15111, doi:doi:10.1161/circ.138.suppl_1.15111 (2018).
78 Kaasinen, E. *et al.* Impact of constitutional TET2 haploinsufficiency on molecular and clinical phenotype in humans. *Nature communications* **10**, 1252, doi:10.1038/s41467-019-09198-7 (2019).
79 Hoogendijk, A. J. *et al.* Dynamic Transcriptome-Proteome Correlation Networks Reveal Human Myeloid Differentiation and Neutrophil-Specific Programming. *Cell reports* **29**, 2505-2519.e2504, doi:https://doi.org/10.1016/j.celrep.2019.10.082 (2019).
80 Cowland, J. B. & Borregaard, N. Granulopoiesis and granules of human neutrophils. *Immunological reviews* **273**, 11-28, doi:10.1111/imr.12440 (2016).
81 Johansen, L. M. *et al.* c-Myc is a critical target for c/EBPalpha in granulopoiesis. *Molecular and cellular biology* **21**, 3789-3806, doi:10.1128/mcb.21.11.3789-3806.2001 (2001).

82 Wang, Q. F. & Friedman, A. D. CCAAT/enhancer-binding proteins are required for granulopoiesis independent of their induction of the granulocyte colony-stimulating factor receptor. *Blood* **99**, 2776-2785 (2002).
83 Gombart, A. F. *et al.* Regulation of neutrophil and eosinophil secondary granule gene expression by transcription factors C/EBP epsilon and PU.1. *Blood* **101**, 3265-3273, doi:10.1182/blood-2002-04-1039 (2003).
84 Verbeek, W., Wachter, M., Lekstrom-Himes, J. & Koeffler, H. P. C/EBPepsilon -/- mice: increased rate of myeloid proliferation and apoptosis. *Leukemia* **15**, 103-111 (2001).
85 Bartels, M. *et al.* Acetylation of C/EBPepsilon is a prerequisite for terminal neutrophil differentiation. *Blood* **125**, 1782-1792, doi:10.1182/blood-2013-12-543850 (2015).
86 Ciavatta, D. J. *et al.* Epigenetic basis for aberrant upregulation of autoantigen genes in humans with ANCA vasculitis. *The Journal of clinical investigation* **120**, 3209-3219, doi:10.1172/jci40034 (2010).
87 Jones, B. E. *et al.* Gene-Specific DNA Methylation Changes Predict Remission in Patients with ANCA-Associated Vasculitis. *Journal of the American Society of Nephrology : JASN* **28**, 1175-1187, doi:10.1681/asn.2016050548 (2017).
88 Rørvig, S., Østergaard, O., Heegaard, N. H. & Borregaard, N. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. *Journal of leukocyte biology* **94**, 711-721, doi:10.1189/jlb.1212619 (2013).
89 Wada, T. *et al.* A Novel In-Frame Deletion in the Leucine Zipper Domain of C/EBPepsilon Leads to Neutrophil-Specific Granule Deficiency. *Journal of immunology (Baltimore, Md. : 1950)* **195**, 80-86, doi:10.4049/jimmunol.1402222 (2015).
90 Marciano, B. E. *et al.* Common severe infections in chronic granulomatous disease. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **60**, 1176-1183, doi:10.1093/cid/ciu1154 (2015).
91 Sengeløv, H., Kjeldsen, L. & Borregaard, N. Control of exocytosis in early neutrophil activation. *Journal of immunology (Baltimore, Md. : 1950)* **150**, 1535-1543 (1993).
92 Ai, Z. & Udalova, I. A. Transcriptional regulation of neutrophil differentiation and function during inflammation. *Journal of leukocyte biology* **107**, 419-430, doi:10.1002/jlb.1ru1219-504rr (2020).
93 Hock, H. *et al.* Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity* **18**, 109-120, doi:10.1016/s1074-7613(02)00501-0 (2003).
94 Wang, W., Wang, X., Ward, A. C., Touw, I. P. & Friedman, A. D. C/EBPalpha and G-CSF receptor signals cooperate to induce the myeloperoxidase and neutrophil elastase genes. *Leukemia* **15**, 779-786, doi:10.1038/sj.leu.2402094 (2001).
95 Kolaczkowska, E. & Kubes, P. Neutrophil recruitment and function in health and inflammation. *Nature reviews. Immunology* **13**, 159-175, doi:10.1038/nri3399 (2013).
96 Zeidler, C., Germeshausen, M., Klein, C. & Welte, K. Clinical implications of ELA2-, HAX1-, and G-CSF-receptor (CSF3R) mutations in severe congenital neutropenia. *Br J Haematol* **144**, 459-467, doi:10.1111/j.1365-2141.2008.07425.x (2009).
97 Hirai, H. *et al.* C/EBPβ is required for 'emergency' granulopoiesis. *Nature immunology* **7**, 732-739, doi:10.1038/ni1354 (2006).

98 Boettcher, S. *et al.* Endothelial cells translate pathogen signals into G-CSF-driven emergency granulopoiesis. *Blood* **124**, 1393-1403, doi:10.1182/blood-2014-04-570762 (2014).
99 Boettcher, S. & Manz, M. G. Regulation of Inflammation- and Infection-Driven Hematopoiesis. *Trends in immunology* **38**, 345-357, doi:10.1016/j.it.2017.01.004 (2017).
100 Schurch, C. M., Riether, C. & Ochsenbein, A. F. Cytotoxic CD8+ T cells stimulate hematopoietic progenitors by promoting cytokine release from bone marrow mesenchymal stromal cells. *Cell stem cell* **14**, 460-472, doi:10.1016/j.stem.2014.01.002 (2014).
101 Walker, F. *et al.* IL6/sIL6R complex contributes to emergency granulopoietic responses in G-CSF- and GM-CSF-deficient mice. *Blood* **111**, 3978-3985, doi:10.1182/blood-2007-10-119636 (2008).
102 Megias, J. *et al.* Direct Toll-like receptor-mediated stimulation of hematopoietic stem and progenitor cells occurs in vivo and promotes differentiation toward macrophages. *Stem cells (Dayton, Ohio)* **30**, 1486-1495, doi:10.1002/stem.1110 (2012).
103 Meisel, M. *et al.* Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature* **557**, 580-584, doi:10.1038/s41586-018-0125-z (2018).
104 Wengner, A. M., Pitchford, S. C., Furze, R. C. & Rankin, S. M. The coordinated action of G-CSF and ELR + CXC chemokines in neutrophil mobilization during acute inflammation. *Blood* **111**, 42-49, doi:10.1182/blood-2007-07-099648 (2008).
105 Nabah, Y. N. *et al.* Angiotensin II induces neutrophil accumulation in vivo through generation and release of CXC chemokines. *Circulation* **110**, 3581-3586, doi:10.1161/01.Cir.0000148824.93600.F3 (2004).
106 Mizgerd, J. P., Lupa, M. M. & Spieker, M. S. NF-κB p50 facilitates neutrophil accumulation during LPS-induced pulmonary inflammation. *BMC Immunology* **5**, 10, doi:10.1186/1471-2172-5-10 (2004).
107 Foxman, E. F., Campbell, J. J. & Butcher, E. C. Multistep navigation and the combinatorial control of leukocyte chemotaxis. *The Journal of cell biology* **139**, 1349, doi:10.1083/jcb.139.5.1349 (1997).
108 Funamoto, S., Meili, R., Lee, S., Parry, L. & Firtel, R. A. Spatial and temporal regulation of 3-phosphoinositides by PI 3-kinase and PTEN mediates chemotaxis. *Cell* **109**, 611-623, doi:10.1016/s0092-8674(02)00755-9 (2002).
109 Liu, X. *et al.* Bidirectional regulation of neutrophil migration by mitogen-activated protein kinases. *Nature immunology* **13**, 457-464, doi:10.1038/ni.2258 (2012).
110 Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science (New York, N.Y.)* **303**, 1532-1535, doi:10.1126/science.1092385 (2004).
111 Fuchs, T. A. *et al.* Novel cell death program leads to neutrophil extracellular traps. *The Journal of cell biology* **176**, 231-241, doi:10.1083/jcb.200606027 (2007).
112 Raup-Konsavage, W. M. *et al.* Neutrophil peptidyl arginine deiminase-4 has a pivotal role in ischemia/reperfusion-induced acute kidney injury. *Kidney international* **93**, 365-374, doi:10.1016/j.kint.2017.08.014 (2018).
113 Sekeres, M. A. *et al.* Randomized Phase II Study of Azacitidine Alone or in Combination With Lenalidomide or With Vorinostat in Higher-Risk Myelodysplastic Syndromes and Chronic Myelomonocytic Leukemia: North American Intergroup Study SWOG S1117. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **35**, 2745-2753, doi:10.1200/jco.2015.66.2510 (2017).

114 Hachiya, M., Osawa, Y. & Akashi, M. Role of TNFα in regulation of myeloperoxidase expression in irradiated HL60 promyelocytic cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1495**, 237-249, doi:https://doi.org/10.1016/S0167-4889(99)00168-8 (2000).
115 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).
116 Amend, S. R., Valkenburg, K. C. & Pienta, K. J. Murine Hind Limb Long Bone Dissection and Bone Marrow Isolation. *Journal of visualized experiments : JoVE*, doi:10.3791/53936 (2016).
117 Hasenberg, A. *et al.* Catchup: a mouse model for imaging-based tracking and modulation of neutrophil granulocytes. *Nature methods* **12**, 445-452, doi:10.1038/nmeth.3322 (2015).
118 Duan, M. *et al.* CD11b immunophenotyping identifies inflammatory profiles in the mouse and human lungs. *Mucosal immunology* **9**, 550-563, doi:10.1038/mi.2015.84 (2016).
119 Sagiv, J. Y. *et al.* Phenotypic diversity and plasticity in circulating neutrophil subpopulations in cancer. *Cell reports* **10**, 562-573, doi:10.1016/j.celrep.2014.12.039 (2015).
120 Surewaard, B. G., van Strijp, J. A. & Nijland, R. Studying interactions of Staphylococcus aureus with neutrophils by flow cytometry and time lapse microscopy. *Journal of visualized experiments : JoVE*, e50788, doi:10.3791/50788 (2013).
121 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
122 Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods (San Diego, Calif.)* **115**, 80-90, doi:10.1016/j.ymeth.2016.09.016 (2017).
123 Hattar, K. *et al.* Lipoteichoic acid (LTA) from Staphylococcus aureus stimulates human neutrophil cytokine release by a CD14-dependent, Toll-like-receptor-independent mechanism: Autocrine role of tumor necrosis factor-[alpha] in mediating LTA-induced interleukin-8 generation. *Critical care medicine* **34**, 835-841, doi:10.1097/01.ccm.0000202204.01230.44 (2006).
124 Lotz, S. *et al.* Highly purified lipoteichoic acid activates neutrophil granulocytes and delays their spontaneous apoptosis via CD14 and TLR2. *Journal of leukocyte biology* **75**, 467-477, doi:10.1189/jlb.0803360 (2004).
125 von Aulock, S. *et al.* Lipoteichoic acid from Staphylococcus aureus is a potent stimulus for neutrophil recruitment. *Immunobiology* **208**, 413-422, doi:10.1078/0171-2985-00285 (2003).
126 Hoogerwerf, J. J. *et al.* Lung inflammation induced by lipoteichoic acid or lipopolysaccharide in humans. *American journal of respiratory and critical care medicine* **178**, 34-41, doi:10.1164/rccm.200708-1261OC (2008).
127 Lee, S. K., Goh, S. Y., Wong, Y. Q. & Ding, J. L. Response of Neutrophils to Extracellular Haemoglobin and LTA in Human Blood System. *EBioMedicine* **2**, 225-233, doi:10.1016/j.ebiom.2015.01.003 (2015).
128 François, S. *et al.* Inhibition of neutrophil apoptosis by TLR agonists in whole blood: involvement of the phosphoinositide 3-kinase/Akt and NF-kappaB signaling pathways, leading to increased levels of Mcl-1, A1, and phosphorylated Bad. *Journal of*

immunology (Baltimore, Md. : 1950) **174**, 3633-3642, doi:10.4049/jimmunol.174.6.3633 (2005).
129 Sengelov, H., Kjeldsen, L., Diamond, M. S., Springer, T. A. & Borregaard, N. Subcellular localization and dynamics of Mac-1 (alpha m beta 2) in human neutrophils. *The Journal of clinical investigation* **92**, 1467-1476, doi:10.1172/jci116724 (1993).
130 Ma, F. *et al.* A comparison between whole transcript and 3' RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC genomics* **20**, 9, doi:10.1186/s12864-018-5393-3 (2019).
131 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).
132 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
133 Yates, A. D. *et al.* Ensembl 2020. *Nucleic acids research* **48**, D682-D688, doi:10.1093/nar/gkz966 (2019).
134 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
135 Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics (Oxford, England)* **35**, 2084-2092, doi:10.1093/bioinformatics/bty895 (2019).
136 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
137 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research* **47**, W191-W198, doi:10.1093/nar/gkz369 (2019).
138 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
139 Heslop, C. L., Frohlich, J. J. & Hill, J. S. Myeloperoxidase and C-reactive protein have combined utility for long-term prediction of cardiovascular mortality after coronary angiography. *Journal of the American College of Cardiology* **55**, 1102-1109, doi:10.1016/j.jacc.2009.11.050 (2010).
140 Yang, J. *et al.* Histone modification signature at myeloperoxidase and proteinase 3 in patients with anti-neutrophil cytoplasmic autoantibody-associated vasculitis. *Clinical epigenetics* **8**, 85, doi:10.1186/s13148-016-0251-0 (2016).
141 Grigorieva, D. V. *et al.* Myeloperoxidase Stimulates Neutrophil Degranulation. *Bulletin of experimental biology and medicine* **161**, 495-500, doi:10.1007/s10517-016-3446-7 (2016).
142 Lau, D. *et al.* Myeloperoxidase mediates neutrophil activation by association with CD11b/CD18 integrins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 431-436, doi:10.1073/pnas.0405193102 (2005).
143 Sugiyama, S. *et al.* Macrophage myeloperoxidase regulation by granulocyte macrophage colony-stimulating factor in human atherosclerosis and implications in acute coronary syndromes. *The American journal of pathology* **158**, 879-891, doi:10.1016/s0002-9440(10)64036-9 (2001).

144 Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science (New York, N.Y.)* **366**, doi:10.1126/science.aax9198 (2019).
145 Evrard, M. *et al.* Developmental Analysis of Bone Marrow Neutrophils Reveals Populations Specialized in Expansion, Trafficking, and Effector Functions. *Immunity* **48**, 364-379.e368, doi:10.1016/j.immuni.2018.02.002 (2018).
146 Ericson, J. A. *et al.* Gene expression during the generation and activation of mouse neutrophils: implication of novel functional and regulatory pathways. *PLoS One* **9**, e108553, doi:10.1371/journal.pone.0108553 (2014).
147 Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology* **32**, 513-545, doi:10.1146/annurev-immunol-032713-120231 (2014).
148 Pellagatti, A. *et al.* Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* **24**, 756-764, doi:10.1038/leu.2010.31 (2010).
149 Pellagatti, A. *et al.* Gene expression profiles of CD34+ cells in myelodysplastic syndromes: involvement of interferon-stimulated genes and correlation to FAB subtype and karyotype. *Blood* **108**, 337-345, doi:10.1182/blood-2005-12-4769 (2006).
150 Fan, X. *et al.* Murine CXCR1 is a functional receptor for GCP-2/CXCL6 and interleukin-8/CXCL8. *The Journal of biological chemistry* **282**, 11658-11666, doi:10.1074/jbc.M607705200 (2007).
151 Levay, K. & Slepak, V. Z. Up- or downregulation of tescalcin in HL-60 cells is associated with their differentiation to either granulocytic or macrophage-like lineage. *Experimental cell research* **316**, 1254-1262, doi:10.1016/j.yexcr.2010.01.007 (2010).
152 Lebert-Ghali, C.-E. *et al.* HoxA cluster is haploinsufficient for activity of hematopoietic stem and progenitor cells. *Experimental hematology* **38**, 1074-1086.e1075, doi:https://doi.org/10.1016/j.exphem.2010.07.006 (2010).
153 Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000-1008, doi:10.1038/leu.2012.356 (2013).
154 Crooks, G. M. *et al.* Constitutive HOXA5 expression inhibits erythropoiesis and increases myelopoiesis from human hematopoietic progenitors. *Blood* **94**, 519-528 (1999).
155 Rawat, V. P. *et al.* Overexpression of CDX2 perturbs HOX gene expression in murine progenitors depending on its N-terminal domain and is closely correlated with deregulated HOX gene expression in human acute myeloid leukemia. *Blood* **111**, 309-319, doi:10.1182/blood-2007-04-085407 (2008).
156 Thorsteinsdottir, U. *et al.* Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood* **99**, 121-129, doi:10.1182/blood.v99.1.121 (2002).
157 Buske, C. *et al.* Overexpression of HOXA10 perturbs human lymphomyelopoiesis in vitro and in vivo. *Blood* **97**, 2286-2292, doi:10.1182/blood.v97.8.2286 (2001).
158 UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* **47**, D506-d515, doi:10.1093/nar/gky1049 (2019).
159 Zhang, Y., Yang, L. & Chen, L. L. Life without A tail: new formats of long noncoding RNAs. *Int J Biochem Cell Biol* **54**, 338-349, doi:10.1016/j.biocel.2013.10.009 (2014).

160 Yang, R. *et al.* Tet1 and Tet2 maintain mesenchymal stem cell homeostasis via demethylation of the P2rX7 promoter. *Nature communications* **9**, 2143, doi:10.1038/s41467-018-04464-6 (2018).
161 Jessop, P. & Toledo-Rodriguez, M. Hippocampal TET1 and TET2 Expression and DNA Hydroxymethylation Are Affected by Physical Exercise in Aged Mice. *Frontiers in cell and developmental biology* **6**, 45, doi:10.3389/fcell.2018.00045 (2018).
162 Goodwin, L. O. *et al.* Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *Genome research* **29**, 494-505, doi:10.1101/gr.233866.117 (2019).

Appendix A: Notice of renewal of the ethics approval for the study on a cohort of CHIP subjects and controls

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD (HSREB)

HSREB Renewal of Ethics Clearance

December 06, 2019

Dr. Michael Rauh
Pathology and Molecular Medicine
Queen's University

TRAQ #: 6014336
Department Code: PATH-147-14
Study Title: "PATH-147-14 Screening for MDS mutations in the aging population for early detection of disease"
Review Type: Delegated
Date Ethics Clearance Effective: December 06, 2019
Ethics Clearance Expiry Date: December 06, 2020

Dear Dr. Rauh:

The Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board (HSREB) has reviewed the application. This study, including all currently approved documentation has been granted ethical clearance until the expiry date noted above.

Prior to the expiration of your ethics clearance, you will be reminded to submit your renewal report through TRAQ. Any lapses in ethical clearance will be documented below.

Lapse in Ethics Clearance: 2019DEC04 – 2019DEC05

Regards,

Albert F. Clark.

Albert F. Clark, PhD
Chair, Queen's University Health Sciences and Affiliated Teaching Hospitals Ethics Board

The HSREB operates in compliance with, and is constituted in accordance with, the requirements of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the international Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Product Regulations; Part 3 of the Medical Devices Regulations, and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is qualified through the CTO REB Qualification Program and is registered with the U.S. Department of Health and Human Services (DHHS) Office for Human Research Protection (OHRP). Federalwide Assurance Number: FWA#: 00004184, IRB#: 00001173

HSREB members involved in the research project do not participate in the review, discussion or decision.

Appendix B: Notice of renewal of the ethics approval from the Queen's University Animal Care Committee

Animals in Science

Your Animal Care Protocol Renewal 2017-1724

Title: Elucidation of immune dysregulation in genetic murine models of myelodysplastic syndromes (MDS)

has now been approved by the Queen's University Animal Care Committee.

Expiration Date: 3/23/2021
Renewal Date: 3/23/2020

Authorized Amount:

Mouse #1
Breeding Only: 74
Category B (GEM): 324

http://topaz.queensu.ca/TOPAZElements/

If you have any questions, please feel free to contact the UACC Coordinator at uacc@queensu.ca.

Thank You

Genevieve Acteau
Office of University Animal Care Committee
Telephone X 78805
uacc@queensu.ca
https://www.queensu.ca/animals-in-science/

Appendix C: Detailed live cell imaging analysis protocol

First, images were filtered to only display red fluorescent signal from the MitoTracker dye. We then applied a function to identify all MitoTracker-positive regions in the field of view that were the size of a cell or larger (Figure C1). This function marks out regions occupied by neutrophils for downstream co-localization analysis. The size restriction distinguishes neutrophils from bacteria and background noise. However if some neutrophils had poor MitoTracker signal, such as if they were out of the focal plane, then they may not be recognized as cells and would be excluded from the analysis.

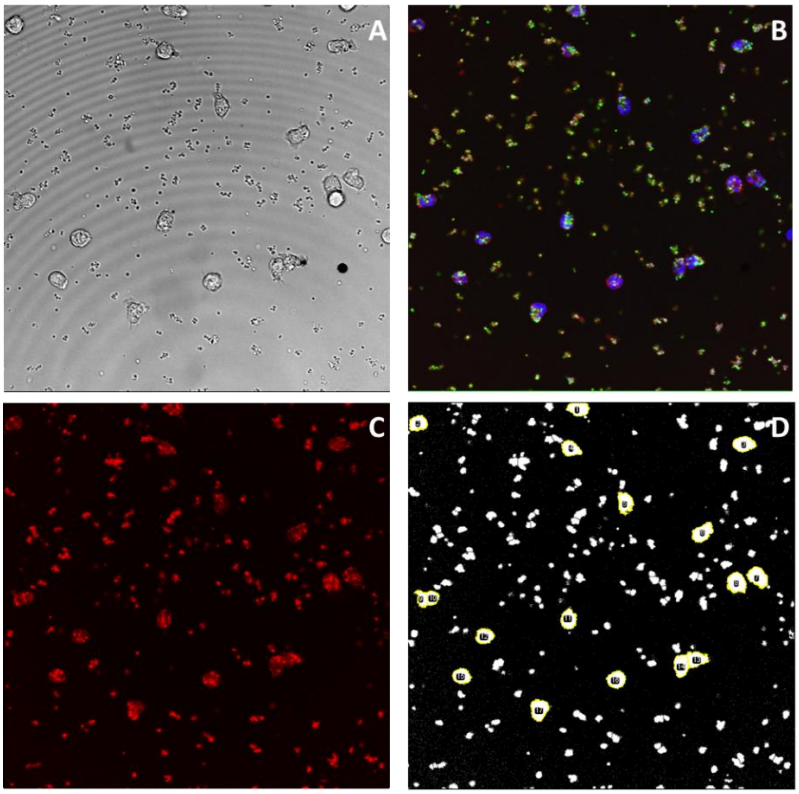


Figure C1. Workflow for identifying neutrophil-occupied regions
A) This bright field image shows the outline of neutrophils and bacteria. **B)** Composite image showing green fluorescence from GFP labelled bacteria, red fluorescence from mitochondria, and blue fluorescence from nuclei. **C)** The red fluorescence filter was applied to only analyze this channel. Both neutrophil and bacterial mitochondria are stained by MitoTracker. **D)** A binary image was made based on the intensity of the red fluorescent signal (i.e. faintly staining regions were excluded from this binary image). To distinguish between neutrophils and bacteria, a size exclusion filter was applied. Only regions of a certain size are counted as neutrophils.

Individual *S. aureus* were identified by their GFP-signals. However phagocytosed *S. aureus* often appear as clusters in neutrophils so it can be challenging to quantify the number of bacteria in the cluster. To address this issue, we applied a function to detect focal points (maxima) with high fluorescence (Figure C2).

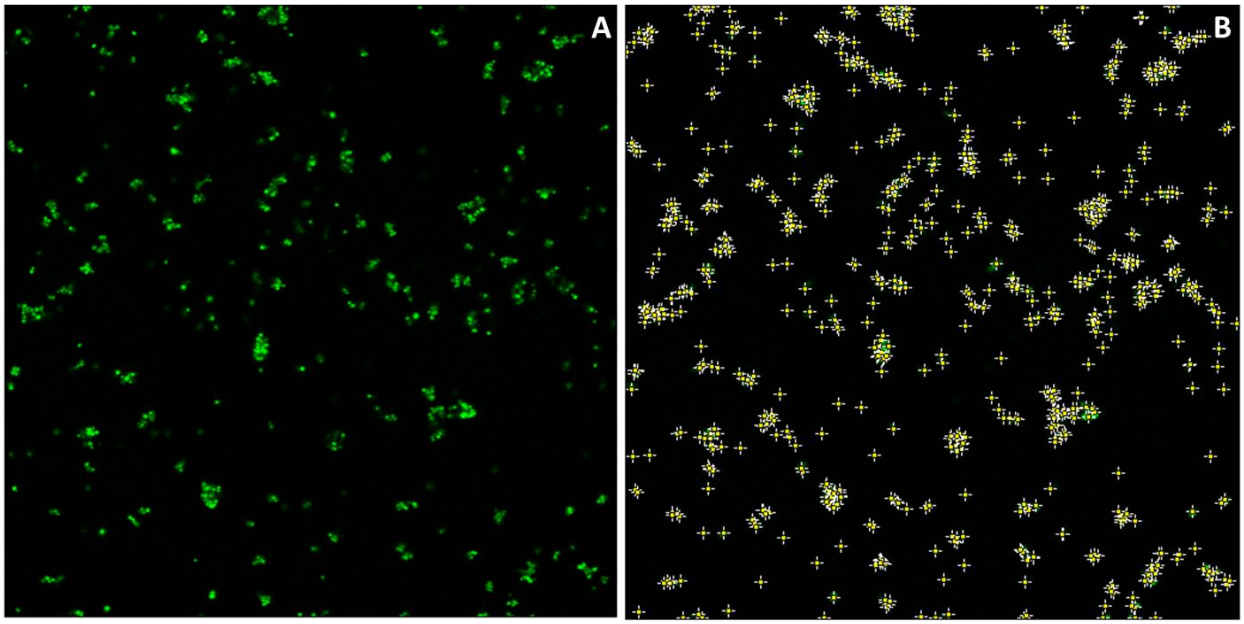


Figure C2. Workflow for identifying individual *S. aureus*
A) The green fluorescence filter was applied to visualize the GFP signal from *S. aureus*. **B)** The find maxima function was used to identify individual bacteria. This method is required to approximate the number of phagocytosed bacteria because they appear as a tightly packed clusters.

Each maxima was translated into a single point on a binary map of the field of view and neutrophil-occupied regions were overlaid onto this map (Figure C3). Each maxima represents a bacterium and this method quantifies the number of bacteria that each neutrophil phagocytosed.

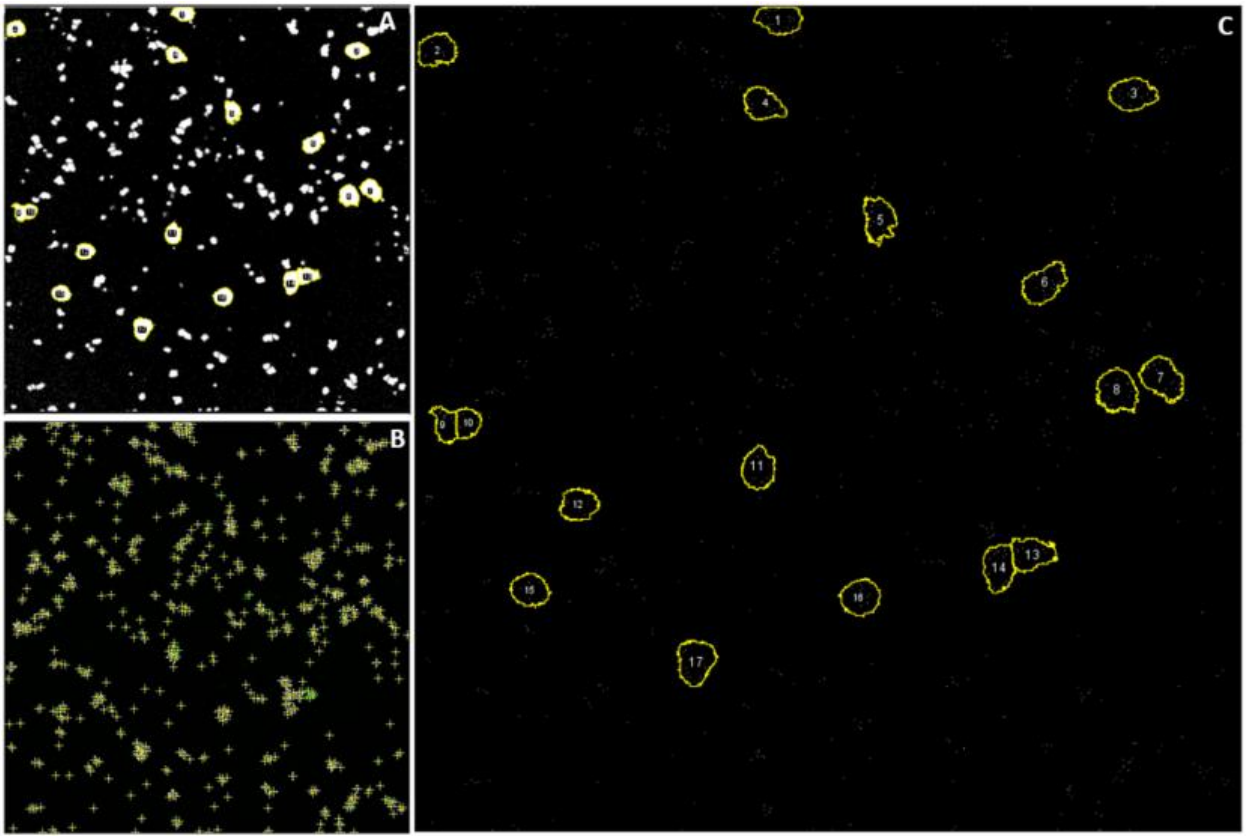


Figure C3. Overlaying *S. aureus* binary map with the map of neutrophil-occupied regions **A)** Map of the regions each neutrophil occupies. **B)** Binary map of the location of each *S. aureus* **C)** Overlay of the two maps, we can see some dots representing *S. aureus* within the neutrophil-occupied regions, these were interpreted as phagocytosis events.

Appendix D: MATLAB Pre-processing and DESeq2 Analysis Code

MATLAB Pre-processing code

```
%% Pre-processing and visualization of the mouse neutrophil
RNAseq data
%{
 Objectives:
1. Visualize data before pre-processing
2. Perform normalization
3. Remove samples with poor correlation, low variability or
low read length
4. Filter out genes with low expression
5. Visualize after pre-processing
%}

%{
Author: Michael Luo
Last modified: March 30 2020
 %}

%% Load the data
% import the .csv file into matlab then save the workspace
% load workspace file
load Star_gene_counts.mat

%% Define variables
sampleIDs = Starcounts(1, 2:end);
GeneIDs = Starcounts(2:end, 1);
gene_counts = cell2mat(Starcounts(2:end, 2:end));
num_rows = size(gene_counts, 1);
num_cols = size(gene_counts, 2);

%% visualize the raw data
% calculate and visualize total counts
total_counts = sum(gene_counts);
plotScatterForData(total_counts, 'Raw Data Total Counts',
'Total Count', sampleIDs);

% calculate and visualize Inter-Quartile Range (IQR)
sample_iqr = iqr(gene_counts);
plotScatterForData(sample_iqr, 'Raw Data IQR', 'Inter-
Quartile Range', sampleIDs);
```

```
% replace zeroes and log-transform, then create boxplots
log2_raw_data = log2(replaceZeros(gene_counts, 'lowval'));

figure
boxplot(log2_raw_data, 'labels', sampleIDs,
'LabelOrientation', 'inline')
xlabel('Mouse Neutrophil Samples')
ylabel('Log2 mRNA Expression')
title('mRNA Expression Across Samples')

% calculate and visualize Spearman Correlations
raw_correlation = SampleCorrelation(gene_counts,
'Spearman');
plotScatterForData(raw_correlation, 'Average Spearman
Correlation',...
    'Spearman Correlation', sampleIDs);

%% Counts per million (CPM) Normalization
% calculate the cpm scaling factor for each column by
dividing the total
% counts by 1 million
cpm_scaling_factor_per_col = total_counts/1000000;

%initialize output matrix with zeroes as placeholders
cpm_counts = zeros(num_rows, num_cols);

% Use a For loop to apply the CPM scaling factor to each
column
for curr_col = 1:num_cols
    for curr_row =1:num_rows
        cpm_counts(curr_row, curr_col) =
gene_counts(curr_row,
curr_col)/cpm_scaling_factor_per_col(1, curr_col);
    end
end

%% visualize the normalized data
% calculate and visualize total counts
cpm_total_counts = sum(cpm_counts);
plotScatterForData(cpm_total_counts, 'CPM Normalized Total
Counts', 'Total Count', sampleIDs);

% calculate and visualize Inter-Quartile Range (IQR)
cpm_iqr = iqr(cpm_counts);
```

```
plotScatterForData(cpm_iqr, 'CPM Normalized IQR', 'Inter-
Quartile Range', sampleIDs);

% replace zeroes and log-transform, then create boxplots
log2_cpm_data = log2(replaceZeros(cpm_counts, 'lowval'));

figure
boxplot(log2_cpm_data, 'labels', sampleIDs,
'LabelOrientation', 'inline')
xlabel('Mouse Neutrophil Samples')
ylabel('Log2 mRNA Expression')
title('mRNA Expression Across Samples')

% calculate and visualize Spearman Correlations
cpm_correlation = SampleCorrelation(cpm_counts,
'Spearman');
plotScatterForData(cpm_correlation, 'Average Spearman
Correlation',...
    'Spearman Correlation', sampleIDs);

%{
sample 2 FF_Base is an outlier with low correlation to the
other samples
we also know that the library amplifcation is weak so we
will remove it
%}
%% removing the sample with poor correlation
lower_outlier_boundary = quantile(cpm_correlation, 0.25)...
    - 1.5*iqr(cpm_correlation);
keepflag = cpm_correlation > lower_outlier_boundary;

% update the gene_count table and the sampleid variable
cpm_counts_no_outlier = cpm_counts(:, keepflag);
sampleIDs_no_outlier = sampleIDs(1, keepflag);

%% visualize the data without the low correlation outlier
% calculate and visualize Inter-Quartile Range (IQR)
cpm_iqr_no_outlier = iqr(cpm_counts_no_outlier);
plotScatterForData(cpm_iqr_no_outlier, 'CPM Normalized IQR
without outlier',...
    'Inter-Quartile Range', sampleIDs_no_outlier);

% replace zeroes and log-transform, then create boxplots
```

```
log2_cpm_counts_no_outlier =
log2(replaceZeros(cpm_counts_no_outlier, 'lowval'));

figure
boxplot(log2_cpm_counts_no_outlier, 'labels',
sampleIDs_no_outlier, 'LabelOrientation', 'inline')
xlabel('Mouse Neutrophil Samples')
ylabel('Log2 mRNA Expression')
title('mRNA Expression Across Samples without outliere of
')

% calculate and visualize Spearman Correlations
cpm_correlation_no_outlier =
SampleCorrelation(cpm_counts_no_outlier, 'Spearman');
plotScatterForData(cpm_correlation_no_outlier,...
    'Average Spearman Correlation without outlier',...
    'Spearman Correlation', sampleIDs_no_outlier);

%% Filter out genes with low counts
%set the filtering threshold to the 90th percentile of
expression and keep
%features that have at least two samples that pass this
threshold.
Filter = FilterLowCounts(cpm_counts_no_outlier, 0.90);

%apply filter to gene count matrix as well as the geneID
list
Filtered_cpm_count_matrix = cpm_counts_no_outlier(Filter,
:);
Filtered_GeneIDs = GeneIDs(Filter, 1);

% The initial dataset contained 46000 genes, the filtered
dataset contains
% around 5700 genes

%% visualize the filtered data
% calculate and visualize Inter-Quartile Range (IQR)
Filtered_cpm_iqr = iqr(Filtered_cpm_count_matrix);
plotScatterForData(Filtered_cpm_iqr, 'Filtered, CPM
Normalized IQR',...
    'Inter-Quartile Range', sampleIDs_no_outlier);

% replace zeroes and log-transform, then create boxplots
```

```
Filtered_log2_cpm_data =
log2(replaceZeros(Filtered_cpm_count_matrix, 'lowval'));

figure
boxplot(Filtered_log2_cpm_data, 'labels',
sampleIDs_no_outlier,...
    'LabelOrientation', 'inline')
xlabel('Mouse Neutrophil Samples')
ylabel('Log2 mRNA Expression')
title('mRNA Expression Across Samples After Filtering')

% calculate and visualize Spearman Correlations
%Filtered_cpm_correlation =
SampleCorrelation(Filtered_cpm_count_matrix, 'Spearman');
%plotScatterForData(Filtered_cpm_correlation,...
%    'CPM Normalized and Filtered Average Spearman
Correlation',...
%    'Spearman Correlation', sampleIDs_no_outlier);

%% END of script
```

***DESeq2 analysis code – Tet2*-FF vs *Tet2*-KO with treatment as a co-variate**

```{r}
#load libraries
library("DESeq2")         # Fold change analysis
library("gplots")         # Plotting functions
library("ggplot2")        # Plotting functions
library("ggrepel")        # prevents overlapping text
library("RColorBrewer")   # Colour palettes
library("biomaRt")        # Download gene annotation
library("EnhancedVolcano")  # volcano plots
```

```{r}
# load data
countData <- read.csv('Star_counts_outlier_removed.csv', header = TRUE, sep = ",")
head(countData)

metaData <- read.csv('mouse_RNAseq_sample_info.csv', header = TRUE, sep =",")
head(metaData)

```

```{r}
# create dds object
ddsFullCountTable <- DESeqDataSetFromMatrix(
  countData = countData,
  colData = metaData,
  design = ~ genotype + treatment, tidy = TRUE)
ddsFullCountTable

dds <- ddsFullCountTable
```

```{r}
# relevel the genotype factor so that the log2 fold changes are calculated releative to Tet2FF
dds$genotype <- relevel(dds$genotype, "Tet2FF")
```

```{r}
# run DESeq2
dds <- DESeq(dds)
resultsNames(dds)
```
```{r plot_PCA,fig.width=12,fig.height=12,eval=T,echo=F}
## Principal component analysis
# Make rlog (regularlized log) transform of counts
logtransform_counts<-rlog(dds)

# Plot PCA, with labelling by genotype, and add text annotation of sample names
plotPCA(logtransform_counts,intgroup="genotype")+geom_text(aes(label=colData(dds)$sample
),
                                hjust=0,vjust=1,size=4)

# Plot PCA with labeling by batch number, and add text annotation of sample names
plotPCA(logtransform_counts,intgroup="batch")+geom_text(aes(label=colData(dds)$sample),
                                hjust=0,vjust=1,size=4)
```

```{r}
#Download Gene Annotations
ens_gene_ids = rownames(dds)
ensembl <- useEnsembl("ensembl", dataset = "mmusculus_gene_ensembl", mirror = "asia")
gene_list <- getBM(attributes = c("uniprot_gn_id", "ensembl_gene_id", "mgi_symbol"),
                   filters = "ensembl_gene_id",
                   values = ens_gene_ids,
```

```
                mart = ensembl)

#Create index of the annotations that we want to keep (filters out annotations not in our data set)
gene_list_index<-match(sapply(strsplit(rownames(dds),"[.]"),"[",1),gene_list$ensembl_gene_id)

```

```{r}
## inspect results produced by default DESeq2 settings

#note in the contrast function, you must specify the three terms in the following order: name of variable, name of level in numerator, name of level in denominator. In this case we are using FF as the denominator so a fold change of 2 means that the KO has 2 times greater epxression than FF

results_KOvsFF <- results(dds, contrast =c("genotype", "Tet2KO", "Tet2FF"), alpha = 0.05)
#head(results_KOvsFF)
summary(results_KOvsFF)
sum(results_KOvsFF$padj<0.05, na.rm = T)

```


```{r}
# apply the apeglm technique (Zhu et al 2018) to generate "shrunken fold changes"
apeglm_results_KOvsFF<-lfcShrink(dds, "genotype_Tet2KO_vs_Tet2FF", type="apeglm")
#head(apeglm_results_KOvsFF, 10)
summary(apeglm_results_KOvsFF)
sum(apeglm_results_KOvsFF$padj<0.05, na.rm = T)
```

```{r}
# apply annotation to both sets of results (default DESeq2 results and after apeglm)
results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]

apeglm_results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
apeglm_results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
apeglm_results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]


#results_KOvsFF_pre_annotate <- cbind(rownames(results_KOvsFF),
data.frame(results_KOvsFF, row.names = NULL))
#names(results_KOvsFF_pre_annotate)[1] <- "ensembl_gene_id"
```

```
#results_KOvsFF_annotated<- merge(results_KOvsFF_pre_annotate, gene_list,
by="ensembl_gene_id")
#logfc_results_KOvsFF_pre_annotate <- cbind(rownames(logfc_results_KOvsFF),
data.frame(logfc_results_KOvsFF, row.names = NULL))
#names(logfc_results_KOvsFF_pre_annotate)[1] <- "ensembl_gene_id"
#logfc_results_KOvsFF_annotated<- merge(logfc_results_KOvsFF_pre_annotate, gene_list,
by="ensembl_gene_id")
#logfc_results_KOvsFF$ensembl_gene_id<-
logfc_results_KOvsFF_annotated$ensembl_gene_id
```

```{r fig.width=12,fig.height=12}
##Volcano plot for results produced using default DESeq2 settings
EnhancedVolcano(results_KOvsFF,
    lab = results_KOvsFF$mgi_symbol,
    x = 'log2FoldChange',
    y = 'pvalue',
    title = 'Tet2KO vs Tet2FF, all treatment groups',
    pCutoff = 0.0006,
    transcriptPointSize = 2.0,
    transcriptLabSize = 4.0)
```

```{r plot_volcanoes,fig.width=10,fig.height=6,eval=T}
## volcano plots (before and after logfc)

## Plot the results before apeglm
with(results_KOvsFF,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="Tet2FF vs Tet2KO Volcano
Plot",pch=20,cex=2.0,col="#99999950"))

## Highlight genes with padj<0.05
with(results_KOvsFF[which(results_KOvsFF$padj<0.05),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))

## plot the apeglm-adjustment results
with(apeglm_results_KOvsFF,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="apeglm adjusted Tet2FFvs Tet2KO Volcano
plot",pch=20,cex=2.0,col="#99999950"))
## Highlight genes with padj<0.05
with(apeglm_results_KOvsFF[which(apeglm_results_KOvsFF$padj<0.05),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))
```

```
# add gene annotations to points of interest
#with(apeglm_results_KOvsFF,
#     text(
#       x=log2FoldChange[log2FoldChange==apeglm_results_KOvsFF$log2FoldChange],
#       y=-log10(pvalue)[-log10(pvalue)== -log10(apeglm_results_KOvsFF$pvalue)],
#       apeglm_results_KOvsFF$mgi_symbol[log2FoldChange ==
apeglm_results_KOvsFF$log2FoldChange],
#       cex = 1,
#       pos = 2,
#       col ="red"
#     ))
```

```{r}
## Write results to file

# All results
df.out<-as.data.frame(apeglm_results_KOvsFF)
write.table(df.out,file="Tet2KO_vs_Tet2FF_all_genes.txt",sep="\t",row.names=F,col.names=T,
quote=F)

# Significant results (padj < 0.05)
write.table(df.out[which(df.out$padj<0.05),],file="Tet2KO_vs_Tet2FF_sig_genes.txt",sep="\t",r
ow.names=F,
   col.names=T,quote=F)

## For g.Profiler analysis, output (a) lists of Ensembl geneIDs (up or down) ranked by fold
change), and
## (b) a list of all Ensembl geneIDs used in the analysis

# Rank df.out by log2FoldChange (low to high)
df.out<-df.out[order(df.out$log2FoldChange),]
sig_down<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange<0),]$ensembl_gene_id

# Rank df.out high to low
df.out<-df.out[order(df.out$log2FoldChange,decreasing=T),]
sig_up<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange>0),]$ensembl_gene_id

#create output files
write.table(sig_up,file="Tet2KO_vs_Tet2FF_up_ensembl.txt",row.names=F,col.names=F,quote
=F)
write.table(sig_down,file="Tet2KO_vs_Tet2FF_down_ensembl.txt",row.names=F,col.names=F,
quote=F)
```

```
write.table(df.out$ensembl_gene_id,file="Tet2KO_vs_Tet2FF_all_ensembl.txt",row.names=F,col.names=F,quote=F)
```

***DESeq2 analysis code* – Baseline subgroup analysis - *Tet2*-FF vs *Tet2*-KO**

```{r}
# load libraries
library("DESeq2")          # Fold change analysis
library("gplots")          # Plotting functions
library("ggplot2")         # Plotting functions
library("RColorBrewer")    # Colour palettes
library("biomaRt")         # Download gene annotation
library("EnhancedVolcano")  # volcano plots
library("dplyr")           # Data manipulation
```

```{r}
# load data
countData <- read.csv('Base_Star_counts_outlier_removed.csv', header = TRUE, sep = ",")
head(countData)

metaData <- read.csv('Base_mouse_RNAseq_sample_info.csv', header = TRUE, sep =",")
head(metaData,9)

```

```{r}
# create dds object
ddsFullCountTable <- DESeqDataSetFromMatrix(
  countData = countData,
  colData = metaData,
  design = ~ genotype, tidy = TRUE)
ddsFullCountTable

dds <- ddsFullCountTable
```

```{r}
# relevel the genotype factor so that the log2 fold changes are calculated releative to Tet2FF
dds$genotype <- relevel(dds$genotype, "Tet2FF")
```

```{r}
# run DESeq2
dds <- DESeq(dds)
```

```
resultsNames(dds)
```

```{r plot_PCA,fig.width=12,fig.height=12,eval=T,echo=F}
## Principal component analysis
# Make rlog (regularlized log) transform of counts
logtransform_counts<-rlog(dds)

# Plot PCA, and add text annotation of sample names
plotPCA(logtransform_counts,intgroup="genotype")+geom_text(aes(label=colData(dds)$sample
),
                              hjust=0,vjust=1,size=4)

```

```{r}
#Download Gene Annotations
ens_gene_ids = rownames(dds)
ensembl <- useEnsembl("ensembl", dataset = "mmusculus_gene_ensembl", mirror = "asia")
gene_list <- getBM(attributes = c("uniprot_gn_id", "ensembl_gene_id", "mgi_symbol"),
                   filters = "ensembl_gene_id",
                   values = ens_gene_ids,
                   mart = ensembl)

#Create index of the annotations that we want to keep (filters out annotations not in our data set)
gene_list_index<-match(sapply(strsplit(rownames(dds),"[.]"),"[",1),gene_list$ensembl_gene_id)
```

```{r}
# inspect results produced by default DESeq2 settings
results_KOvsFF <- results(dds, contrast =c("genotype", "Tet2KO", "Tet2FF"), alpha = 0.05)
#head(results_KOvsFF)
summary(results_KOvsFF)
sum(results_KOvsFF$padj<0.05, na.rm = T)

```

```{r}
# apply the apeglm technique (Zhu et al 2018) to generate "shrunken fold changes"
apeglm_results_KOvsFF<-lfcShrink(dds, "genotype_Tet2KO_vs_Tet2FF", type="apeglm")
#head(apeglm_results_KOvsFF, 10)
summary(apeglm_results_KOvsFF)
sum(apeglm_results_KOvsFF$padj<0.05, na.rm = T)
```

```

```{r}
##Add annotations to the volcano plot
# create dataframe
results_dataframe <- as.data.frame(apeglm_results_KOvsFF)

# create volcano plot
volcanoplot <- ggplot(results_dataframe, aes(results_dataframe$log2FoldChange,
                                    -log10(results_dataframe$pvalue))) + geom_point(color = 'black')
+ theme_classic(base_size = 10)




volcanoplot <- ggplot(results_dataframe, aes(results_dataframe$log2FoldChange,
                                    -log10(results_dataframe$pvalue),
                                    label = results_dataframe$mgi_symbol)) + geom_point()

#
volcanoplot + geom_text()


with(apeglm_results_KOvsFF,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="apeglm adjusted Baseline Tet2FF vs Tet2KO Volcano
plot",pch=20,cex=0.7,col="#99999950"))
# Highlight genes with padj<0.05
with(apeglm_results_KOvsFF[which(apeglm_results_KOvsFF$padj<0.05),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))
```



```{r}
# apply annotation to both sets of results (default DESeq2 and apeglm adjustment)
results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]

apeglm_results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
apeglm_results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
apeglm_results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]
```
```

```{r}
## Volcano plots for the apeglm-adjusted results
# plot the apeglm-adjustment results
with(apeglm_results_KOvsFF,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="apeglm adjusted Baseline Tet2FF vs Tet2KO Volcano
plot",pch=20,cex=0.7,col="#99999950"))
# Highlight genes with padj<0.05
with(apeglm_results_KOvsFF[which(apeglm_results_KOvsFF$padj<0.05),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))
```

```{r fig.width=12,fig.height=12}
##Volcano plot for results produced using default DESeq2 settings
EnhancedVolcano(results_KOvsFF,
    lab = results_KOvsFF$mgi_symbol,
    x = 'log2FoldChange',
    y = 'pvalue',
    pCutoff = 0.00013,
    title = 'Baseline Tet2KO vs Tet2FF',
    transcriptPointSize = 2.0,
    transcriptLabSize = 4.0)

```

```{r}
## Write results to file

# All results
df.out<-as.data.frame(apeglm_results_KOvsFF)
write.table(df.out,file="Base_Tet2KO_vs_Tet2FF_all_genes.txt",sep="\t",row.names=F,col.nam
es=T,quote=F)

# Significant results (padj < 0.05)
write.table(df.out[which(df.out$padj<0.05),],file="Base_Tet2KO_vs_Tet2FF_sig_genes.txt",sep
="\t",row.names=F,
   col.names=T,quote=F)

## For g.Profiler analysis, output (a) lists of Ensembl geneIDs (up or down) ranked by fold
change), and
## (b) a list of all Ensembl geneIDs used in the analysis

# Rank df.out by log2FoldChange (low to high)
df.out<-df.out[order(df.out$log2FoldChange),]
sig_down<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange<0),]$ensembl_gene_id


```

```
# Rank df.out high to low
df.out<-df.out[order(df.out$log2FoldChange,decreasing=T),]
sig_up<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange>0),]$ensembl_gene_id

#create output files
write.table(sig_up,file="Base_Tet2KO_vs_Tet2FF_up_ensembl.txt",row.names=F,col.names=F,
quote=F)
write.table(sig_down,file="Base_Tet2KO_vs_Tet2FF_down_ensembl.txt",row.names=F,col.na
mes=F,quote=F)
write.table(df.out$ensembl_gene_id,file="Base_Tet2KO_vs_Tet2FF_all_ensembl.txt",row.name
s=F,col.names=F,quote=F)
```

DESeq2 analysis code – LTA subgroup analysis - Tet2-FF vs Tet2-KO

```{r}
# load libraries
library("DESeq2")          # Fold change analysis
library("gplots")          # Plotting functions
library("ggplot2")         # Plotting functions
library("RColorBrewer")    # Colour palettes
library("biomaRt")         # Download gene annotation
library("EnhancedVolcano")  # volcano plots
library("dplyr")           # Data manipulation
```

```{r}
# load data
countData <- read.csv('LTA_Star_counts.csv', header = TRUE, sep = ",")
head(countData)

metaData <- read.csv('LTA_mouse_RNAseq_sample_info.csv', header = TRUE, sep =",")
head(metaData,9)

```

```{r}
# create dds object
ddsFullCountTable <- DESeqDataSetFromMatrix(
  countData = countData,
  colData = metaData,
  design = ~ genotype, tidy = TRUE)
ddsFullCountTable

dds <- ddsFullCountTable
```

```

```{r}
# relevel the genotype factor so that the log2 fold changes are calculated releative to Tet2FF
dds$genotype <- relevel(dds$genotype, "Tet2FF")
```

```{r}
# run DESeq2
dds <- DESeq(dds)
resultsNames(dds)
```

```{r plot_PCA,fig.width=12,fig.height=12,eval=T,echo=F}
## Principal component analysis
# Make rlog (regularlized log) transform of counts
logtransform_counts<-rlog(dds)

# Plot PCA, and add text annotation of sample names
plotPCA(logtransform_counts,intgroup="genotype")+geom_text(aes(label=colData(dds)$sample
),
                                hjust=0,vjust=1,size=4)

```

```{r}
#Download Gene Annotations
ens_gene_ids = rownames(dds)
ensembl <- useEnsembl("ensembl", dataset = "mmusculus_gene_ensembl", mirror = "asia")
gene_list <- getBM(attributes = c("uniprot_gn_id", "ensembl_gene_id", "mgi_symbol"),
                   filters = "ensembl_gene_id",
                   values = ens_gene_ids,
                   mart = ensembl)

#Create index of the annotations that we want to keep (filters out annotations not in our data set)
gene_list_index<-match(sapply(strsplit(rownames(dds),"[.]"),"[",1),gene_list$ensembl_gene_id)
```

```{r}
# inspect results produced by default DESeq2 settings
results_KOvsFF <- results(dds, contrast =c("genotype", "Tet2KO", "Tet2FF"), alpha = 0.05)
#head(results_KOvsFF)
summary(results_KOvsFF)
sum(results_KOvsFF$padj<0.05, na.rm = T)
```
```

```

```{r}
# apply the apeglm technique (Zhu et al 2018) to generate "shrunken fold changes"
apeglm_results_KOvsFF<-lfcShrink(dds, "genotype_Tet2KO_vs_Tet2FF", type="apeglm")
#head(apeglm_results_KOvsFF, 10)
summary(apeglm_results_KOvsFF)
sum(apeglm_results_KOvsFF$padj<0.05, na.rm = T)

```

```{r}
# apply annotation to both sets of results (default DESeq2 and apeglm adjustment)
results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]

apeglm_results_KOvsFF$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
apeglm_results_KOvsFF$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
apeglm_results_KOvsFF$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]
```

```{r}
## Volcano plots for the apeglm-adjusted results
# plot the apeglm-adjustment results
with(apeglm_results_KOvsFF,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="apeglm adjusted LTA-treated Tet2FF vs Tet2KO Volcano
plot",pch=20,cex=0.7,col="#99999950"))
# Highlight genes with padj<0.05
with(apeglm_results_KOvsFF[which(apeglm_results_KOvsFF$padj<0.05),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))
```

```{r fig.width=12,fig.height=12}
##Volcano plot for results produced using default DESeq2 settings
EnhancedVolcano(results_KOvsFF,
    lab = results_KOvsFF$mgi_symbol,
    x = 'log2FoldChange',
    y = 'pvalue',
    title = 'LTA treated Tet2KO vs Tet2FF',
    FCcutoff = 0.5,
    pCutoff = 0.0001,
    transcriptPointSize = 2.0,

```
    transcriptLabSize = 4.0)

```

```{r}
## Write results to file

# All results
df.out<-as.data.frame(apeglm_results_KOvsFF)
write.table(df.out,file="LTA_Tet2KO_vs_Tet2FF_all_genes.txt",sep="\t",row.names=F,col.names=T,quote=F)

# Significant results (padj < 0.05)
write.table(df.out[which(df.out$padj<0.05),],file="LTA_Tet2KO_vs_Tet2FF_sig_genes.txt",sep="\t",row.names=F,
   col.names=T,quote=F)

## For g.Profiler analysis, output (a) lists of Ensembl geneIDs (up or down) ranked by fold change), and
## (b) a list of all Ensembl geneIDs used in the analysis

# Rank df.out by log2FoldChange (low to high)
df.out<-df.out[order(df.out$log2FoldChange),]
sig_down<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange<0),]$ensembl_gene_id

# Rank df.out high to low
df.out<-df.out[order(df.out$log2FoldChange,decreasing=T),]
sig_up<-df.out[which(df.out$padj<0.05 & df.out$log2FoldChange>0),]$ensembl_gene_id

#create output files
write.table(sig_up,file="LTA_Tet2KO_vs_Tet2FF_up_ensembl.txt",row.names=F,col.names=F,quote=F)
write.table(sig_down,file="LTA_Tet2KO_vs_Tet2FF_down_ensembl.txt",row.names=F,col.names=F,quote=F)
write.table(df.out$ensembl_gene_id,file="LTA_Tet2KO_vs_Tet2FF_all_ensembl.txt",row.names=F,col.names=F,quote=F)
```

DESeq2 analysis code – LTA treated vs Vehicle treated samples

```{r}
#load libraries
library("DESeq2")          # Fold change analysis
library("gplots")          # Plotting functions
```

```
library("ggplot2")         # Plotting functions
library("ggrepel")         # prevents overlapping text
library("RColorBrewer")    # Colour palettes
library("biomaRt")         # Download gene annotation
library("EnhancedVolcano")  # volcano plots
```

```{r}
# load data
countData <- read.csv('Star_counts_LTA_vs_VHL_FF.csv', header = TRUE, sep = ",")
head(countData)

metaData <- read.csv('VHL_vs_LTA_FF_mouse_RNAseq_sample_info.csv', header = TRUE,
sep =",")
head(metaData)

```

```{r}
# create dds object
ddsFullCountTable <- DESeqDataSetFromMatrix(
  countData = countData,
  colData = metaData,
  design = ~ treatment, tidy = TRUE)
ddsFullCountTable

dds <- ddsFullCountTable
```

```{r}
# relevel the genotype factor so that the log2 fold changes are calculated releative to Tet2FF
dds$treatment <- relevel(dds$treatment, "Vehicle")
```

```{r}
# run DESeq2
dds <- DESeq(dds)
resultsNames(dds)
```

```{r plot_PCA,fig.width=12,fig.height=12,eval=T,echo=F}
## Principal component analysis
# Make rlog (regularlized log) transform of counts
logtransform_counts<-rlog(dds)

# Plot PCA, and add text annotation of sample names
```

```
plotPCA(logtransform_counts,intgroup="treatment")+geom_text(aes(label=colData(dds)$sample
),
                              hjust=0,vjust=1,size=4)

```

```{r}
#Download Gene Annotations
ens_gene_ids = rownames(dds)
ensembl <- useEnsembl("ensembl", dataset = "mmusculus_gene_ensembl", mirror = "asia")
gene_list <- getBM(attributes = c("uniprot_gn_id", "ensembl_gene_id", "mgi_symbol"),
                   filters = "ensembl_gene_id",
                   values = ens_gene_ids,
                   mart = ensembl)

#Create index of the annotations that we want to keep (filters out annotations not in our data set)
gene_list_index<-match(sapply(strsplit(rownames(dds),"[.]"),"[",1),gene_list$ensembl_gene_id)

```

```{r}
## inspect results produced by default DESeq2 settings

#note in the contrast function, you must specify the three terms in the following order: name of variable, name of level in numerator, name of level in denominator. In this case we are using FF as the denominator so a fold change of 2 means that the KO has 2 times greater epxression than FF

results_LTAvsVHL <- results(dds, contrast =c("treatment", "LTA", "Vehicle"), alpha = 0.10)
summary(results_LTAvsVHL)
sum(results_LTAvsVHL$padj<0.10, na.rm = T)

```


```{r}
# apply the apeglm technique (Zhu et al 2018) to generate "shrunken fold changes"
apeglm_results_LTAvsVHL<-lfcShrink(dds, "treatment_LTA_vs_Vehicle", type="apeglm")
#head(apeglm_results_KOvsFF, 10)
summary(apeglm_results_LTAvsVHL)
sum(apeglm_results_LTAvsVHL$padj<0.10, na.rm = T)
```

```{r}
# apply annotation to both sets of results (default DESeq2 results and after apeglm)
results_LTAvsVHL$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
```

```
results_LTAvsVHL$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
results_LTAvsVHL$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]

apeglm_results_LTAvsVHL$ensembl_gene_id<-gene_list$ensembl_gene_id[gene_list_index]
apeglm_results_LTAvsVHL$uniprot_gn_id<-gene_list$uniprot_gn_id[gene_list_index]
apeglm_results_LTAvsVHL$mgi_symbol<-gene_list$mgi_symbol[gene_list_index]
```

```{r}
## Write results to file
#df.out<-as.data.frame(apeglm_results_LTAvsVHL)
df.out <- as.data.frame(results_LTAvsVHL)
write.table(df.out,file="VHL_vs_LTA_FF_all_genes.txt",sep="\t",row.names=F,col.names=T,quote=F)

# Significant results (padj < 0.10)
write.table(df.out[which(df.out$padj<0.10),],file="VHL_vs_LTA_FF_sig_genes.txt",sep="\t",row.names=F,
    col.names=T,quote=F)

## For g.Profiler analysis, output (a) lists of Ensembl geneIDs (up or down) ranked by fold change), and
## (b) a list of all Ensembl geneIDs used in the analysis

# Rank df.out by log2FoldChange (low to high)
df.out<-df.out[order(df.out$log2FoldChange),]
sig_down<-df.out[which(df.out$padj<0.10 & df.out$log2FoldChange<0),]$ensembl_gene_id

# Rank df.out high to low
df.out<-df.out[order(df.out$log2FoldChange,decreasing=T),]
sig_up<-df.out[which(df.out$padj<0.10 & df.out$log2FoldChange>0),]$ensembl_gene_id

#create output files
write.table(sig_up,file="VHL_vs_LTA_FF_up_ensembl.txt",row.names=F,col.names=F,quote=F)
#
write.table(sig_down,file="VHL_vs_LTA_FF_down_ensembl.txt",row.names=F,col.names=F,quote=F)
write.table(df.out$ensembl_gene_id,file="VHL_vs_LTA_FF_all_ensembl.txt",row.names=F,col.names=F,quote=F)

# there were no significantly downregulated genes so I leave it out
```

From inspecting results table we need to set the pvalue cutoff to 0.00057 to visualize all transcripts with padj<0.10

```{r fig.width=12,fig.height=12}
##Volcano plot for results produced using default DESeq2 settings
EnhancedVolcano(results_LTAvsVHL,
    lab = results_LTAvsVHL$mgi_symbol,
    x = 'log2FoldChange',
    y = 'pvalue',
    title = 'VHL vs LTA within FF genotype',
    pCutoff = 0.00057,
    transcriptPointSize = 2.0,
    transcriptLabSize = 4.0)
```

```{r plot_volcanoes,fig.width=10,fig.height=6,eval=T}
## Volcano plots for the apeglm-adjusted results

## plot the apeglm-adjusted results
with(apeglm_results_LTAvsVHL,plot(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",
    ylab="-log10(p-value)",main="apeglm adjusted Volcano plot VHL vs LTA within FF
genotype",pch=20,cex=2.0,col="#99999950"))
## Highlight genes with padj<0.05
with(apeglm_results_LTAvsVHL[which(apeglm_results_LTAvsVHL$padj<0.10),],
    points(log2FoldChange,-log10(pvalue),xlab="log2FoldChange",ylab="-log10(p-value)",
        col="#EE2C2C60",pch=20,cex=2.0))

```

Appendix E: Supplemental Tables and Figures

Table S1. Demographic and clinical data of the qRT-PCR validation cohort

Mutation status	Total number	Percentage of females	Median age	Absolute neutrophil count	White blood cell count	Percentage of neutrophils
No CHIP	4	50%	85	4.6	7.3	63%
All CHIP subtypes	14	50%	80	4.8	7.4	64%
TET2	5	20%	81	4.6	7.1	65%
DNMT3A	4	75%	77	4.6	7.5	61%
Other CHIP	5	60%	86	5.0	7.7	65%

Table S2. Demographic data and clinical of the serum ELISA validation cohort

Mutation status	Total number	Percentage of females	Median age	Absolute neutrophil count	White blood cell count	Percentage of neutrophils
No CHIP	13	54%	82.0	4.6	7.5	61%
All CHIP subtypes	51	69%	80.0	4.7	7.2	65%
TET2	15	60%	81.0	4.6	7.0	66%
DNMT3A	22	73%	79.0	4.8	7.4	65%
Other CHIP	14	71%	81.5	4.6	7.3	64%

Table S3. qRT-PCR protocol

The samples were heated at 95°C for 40 seconds to activate the reaction. Then the samples were subject to 40 reaction cycles consisting of denaturation at 95°C for 10 seconds and annealing/extension at 60°C for 30 seconds.

Step	Temp	Time	Cycles
Activation	95	40 sec	1
Denaturation	95	10 sec	40
Annealing/ Extension	60	30 sec	

Table S4. Primers used qRT-PCR on human PB samples
Forward and reverse primers for *MPO*, *LTF*, *ACTB*, and *GAPDH* are listed in this table. All primers were ordered from Invitrogen (Carlsbad, CA).

Primer	**Sequence**
LTF Forward	5' GCC TGC TCT TCA ACC AGA CG 3'
LTF Reverse	5' CTC TCG TTG CTG TTG GGC A 3'
MPO Forward	5' CCT CAC CCT CGG TTC AGA AAC 3'
MPO Reverse	5' AAT TAC ATC TGT CAT GGT TTC TTG GA 3'
ACTB Forward	5' TCT ACA ATG AGC TGC GTG TG 3'
ACTB Reverse	5'GTA CAG GTC TTT GCG GAT GT 3'
GAPDH Forward	5' TGG GTG TGA ACC ATG AGA AG 3'
GAPDH Reverse	5'CCT GTT GCT GTA GCC AAA TTC 3'

Table S5. *Tet2* and *Vav*-iCre genotyping master mixes
The volumes listed are for one sample, where 45 µL of mastermix is mixed with 5 µL of DNA to make a 50uL PCR reaction.

***Tet2* Master Mix**		***Vav-iCre Master Mix***	
Nuclease-Free H_2O	33.25 µL	Nuclease-Free H_2O	33.25µL
10X PCR Buffer	5 µL	10X PCR Buffer	5µL
MgCl2	3 µL	MgCl2	3µL
dNTP	1.5 µL	dNTP	1.5µL
Tet2 Forward Primer	1 µL	*Vav-iCre* Forward Primer	1µL
Tet2 Reverse Primer	1 µL	*Vav-iCre* Reverse Primer	1µL
Platinum Taq	0.25 µL	*Vav-iCre* Control Forward Primer	1µL
		Vav-iCre Control Reverse Primer	1µL
		Platinum Taq	0.25µL

Table S6. Primers used for *Tet2, Vav-iCre* genotyping
All primers were recommended by Jackson Laboratories and ordered from Invitrogen (Carlsbad, CA).

Primer	**Sequence**
Tet2 Forward	5' AAG AAT TGC TAC AGG CCT GC 3'
Tet2 Reverse	5' TTC TTT AGC CCT TGC TGA GC 3'
Vav-iCre Forward	5' AGA TGC CAG GAC ATC AGG AAC CTG 3'
Vav-iCre Reverse	5' ATC AGC CAC ACC AGA CAC AGA GAT C 3'
Vav-iCre – control Forward	5' CTA GGC CAC AGA ATT GAA AGA TCT 3'
Vav-iCre – control Reverse	5' GTA GGT GGA AAT TCT AGC ATC ATC C 3'

Table S7. PCR programs for amplifying *Tet2* and *Vav-iCre*
The following protocols were used to amplify the Tet2 and Vav-iCre products

***Tet2*- PCR**			***Vav-iCre* – PCR**		
Cycles	**Temperature**	**Time**	**Cycles**	**Temperature**	**Time**
1	94 C	3 minutes	1	94C	1:30
30	94 C	30 seconds	36	94C	0:30
	55 C	30 seconds		64C	0:45
	72 C	1 minute		72C	0:45
	4 C	Hold	1	72C	2:00
				4C	Hold

Table S8. qRT-PCR primers for mouse neutrophil targets
All primers were ordered from Invitrogen (Carlsbad, CA).

Primer	**Sequence**
Mouse *Il1b* Fwd	5′ GCC CAT CCT CTG TGA CTC AT
Mouse *Il1b* Rev	5′ AGG CCA CAG GTA TTT TGT CG
Mouse *Tnfa* Fwd	5' AAG CCT GTA GCC CAC GTC GTA
Mouse *Tnfa* Rev	5' GGC ACC ACT AGT TGG TTG TCT TTG
Mouse *Tet2* Fwd	5' AGA GCC TCA AGC AAC CAA AA
Mouse *Tet2* Rev	5' ACA TCC CTG AGA GCT CTT GC

Table S9. Flow cytometry results for the RNA-seq samples
Apoptosis was assessed through flow cytometry with Annexin-V PI staining. The proportion of mature neutrophils ($Cd11b^{hi}$ $Ly6g^{hi}$) was also assessed by flow cytometry with Cd11b Ly6g staining.

Sample	Percent viable	Percent in early apoptosis	Percent in late apoptosis	Proportion of mature neutrophils
D12 FF-Base	79.6	13.4	7.0	71%
D12 FF-LTA	73.7	18.5	7.8	71%
D12 FF-VHL	72.8	17.9	9.3	71%
D12 KO-Base	84.6	6.6	8.8	80%
D12 KO-LTA	79.2	12.9	7.9	80%
D12 KO-VHL	79.7	10.4	9.9	80%
D13 FF-Base	83.6	10.6	5.8	82%
D13 FF-LTA	72.8	21.3	5.9	82%
D13 FF-VHL	69.8	21.8	8.4	82%
D13 KO-Base	83.5	11.7	4.8	81%
D13 KO-LTA	79.2	15.4	5.4	81%
D13 KO-VHL	78.9	15.1	6.0	81%
D14 FF-Base	88.3	7.7	4.0	83%
D14 FF-LTA	85.7	8.2	6.1	83%
D14 FF-VHL	85.2	9.1	5.7	83%
D14 KO-Base	89.8	7.3	2.9	75%
D14 KO-LTA	88.4	7.4	4.2	75%
D14 KO-VHL	88.3	7.4	4.3	75%
D15 FF-Base	82.7	8.9	8.4	84%
D15 FF-LTA	77.8	14.2	8	84%
D15 FF-VHL	76.4	13.9	9.7	84%
D15 KO-Base	85.1	8.5	6.4	85%
D15 KO-LTA	82.5	10.8	6.7	85%
D15 KO-VHL	80.3	13.4	6.3	85%

Table S10. List of differentially expressed transcripts (padj < 0.05) between *Tet2*-KO and *Tet2*-FF neutrophils

40 transcripts were significantly upregulated and 99 transcripts were significantly downregulated in *Tet2*-KO neutrophils. Base mean is the average number of counts across all samples, LfcSE is the standard error of the log 2 fold change. Benjamini-Hochberg multiple testing correction was applied.

Base mean	Log2 fold change	LfcSE	pvalue	padj	MGI_symbol
78.66568	-2.805732621	0.174243	1.57E-59	2.11E-55	Ccdc80
110.693	2.315078355	0.162548	2.20E-47	1.48E-43	Commd10
46.12197	-2.413434752	0.176908	1.61E-43	7.21E-40	Pnpla1
32.14504	-2.799506304	0.250894	4.98E-30	1.67E-26	Tmem45a2
65.58598	-1.573105684	0.166335	1.28E-22	3.45E-19	Stfa3
51.10996	-1.686985841	0.185582	5.76E-21	1.29E-17	Selp
18.30473	-1.995740994	0.230569	2.55E-19	4.90E-16	Sox13
50.40539	-2.074239189	0.265878	2.56E-16	4.30E-13	Cstdc5
4.308913	8.022176753	2.601854	2.10E-15	3.13E-12	Sva
34.7171	-1.628916884	0.22587	1.68E-14	2.25E-11	Ldhc
110.7582	-1.182189316	0.164384	2.16E-14	2.64E-11	Got1
28.63149	-1.325809091	0.184012	3.35E-14	3.75E-11	Chit1
53.08644	-1.250883757	0.177161	7.60E-14	7.85E-11	Brd8dc
94.46565	0.772173976	0.111097	1.88E-13	1.81E-10	Dsp
195.0309	-0.629667836	0.095098	1.71E-12	1.53E-09	Usp46
30.04458	-1.137171661	0.174395	3.39E-12	2.85E-09	Mpzl1
8.728891	-2.471781342	0.392942	1.65E-11	1.30E-08	1700006J14Rik
44.41074	1.437246994	0.252228	4.44E-10	3.31E-07	Tesc
7.734956	-2.71991252	0.472252	4.95E-10	3.50E-07	Rasgef1a
5.965247	-2.651365822	0.472117	1.25E-09	8.39E-07	Sec16b
121.6599	-0.722249174	0.132087	2.07E-09	1.26E-06	Sox4
70.61046	-0.931484938	0.171049	2.06E-09	1.26E-06	Acad11
152.7752	0.540173852	0.100496	3.94E-09	2.30E-06	Itga6
277.3044	-0.726180894	0.136326	4.54E-09	2.54E-06	Acvrl1
161.6331	-1.158377809	0.226149	1.07E-08	5.75E-06	Cdc42ep2
1769.075	0.852692197	0.168533	1.83E-08	9.44E-06	Asprv1
79.16721	-0.746099007	0.147634	2.02E-08	1.00E-05	Mgl2
31.88217	-1.008849503	0.200274	2.24E-08	1.07E-05	Adtrp
68.69248	-0.985787812	0.196568	2.32E-08	1.07E-05	Ebf1
51.86571	-0.965121216	0.195248	3.06E-08	1.37E-05	Arsg
24.32961	-0.836545874	0.172233	5.42E-08	2.35E-05	Plcb3
19.17606	-1.26767695	0.264252	6.33E-08	2.66E-05	Akap12
62.95568	-0.938782754	0.201545	1.35E-07	5.50E-05	Mmp25
1455.671	-0.56734965	0.122196	1.49E-07	5.75E-05	Slfn4
14.10049	-1.159784889	0.249443	1.50E-07	5.75E-05	Gm10030
8.214638	-1.772530597	0.384996	2.10E-07	7.84E-05	Tcea3

6.488757	1.800532925	0.396562	2.87E-07	0.000104	Maob
8.448983	-1.525659595	0.342376	3.88E-07	0.000137	Edaradd
63.67023	-0.60030731	0.136119	4.61E-07	0.000155	Fer
9.508352	4.267144499	1.031101	4.53E-07	0.000155	D930028M14Rik
40.66792	-1.129854071	0.259479	5.42E-07	0.000177	Dock9
342.9689	-0.374499895	0.084876	5.69E-07	0.000182	Ccdc174
9.449944	-1.766674952	0.407879	6.15E-07	0.000192	Hoxa9
15.43371	2.049400934	0.480162	7.00E-07	0.000214	Gm5900
1450.785	0.362989316	0.082848	7.19E-07	0.000215	Zmpste24
45.12541	-0.622140341	0.145957	9.07E-07	0.000265	Nav2
6.965533	-1.529478098	0.357109	9.53E-07	0.000273	Cxcr1
7.280245	2.2127055	0.528316	9.95E-07	0.000278	5830444B04Rik
5.596086	-1.918288865	0.448824	1.05E-06	0.000288	Dnah6
341.0784	-0.522680457	0.124059	1.14E-06	0.000305	Vcam1
7.654711	-1.819652411	0.437003	1.35E-06	0.000357	Prps1l1
76.27699	-1.204181776	0.294442	1.60E-06	0.000413	Krt81
43.78106	0.640612918	0.15571	1.71E-06	0.000433	P2rx1
54.78265	-0.719891213	0.175906	1.87E-06	0.000465	Cass4
88.91044	-0.57511576	0.142226	2.32E-06	0.000558	Lig3
14.53209	-1.084851978	0.269129	2.30E-06	0.000558	Abcb9
43.54228	-1.368689187	0.357374	2.58E-06	0.000607	Olfm4
81.72629	-0.546743751	0.137103	3.04E-06	0.000704	Tnnt1
33.44756	-1.036540992	0.262512	3.23E-06	0.000736	Atg4c
159.7164	0.44429986	0.114495	5.11E-06	0.001144	Serpine2
945.5672	-0.458643472	0.119187	5.51E-06	0.001214	Efhd2
718.7175	-1.606013554	0.430107	6.91E-06	0.001497	Rpl34
54.47462	0.539251998	0.143778	7.87E-06	0.001679	Ifi35
408.6335	-0.384193188	0.103074	9.79E-06	0.002055	Ate1
26.15542	-0.980925586	0.269245	1.03E-05	0.00212	Bach2
118.1871	-0.351249173	0.095797	1.32E-05	0.002692	Erich1
6.02354	-1.593198284	0.441584	1.42E-05	0.002843	Hc
170.5655	-0.438101191	0.122202	1.47E-05	0.002897	Snx11
200.8273	0.62887484	0.177488	1.60E-05	0.003122	Rtp4
301.2317	-0.466748405	0.131528	1.76E-05	0.00338	Arid5b
948.4583	-1.368210391	0.396951	1.96E-05	0.003707	Igha
3.820195	-0.078918288	0.172496	2.33E-05	0.004352	Srgap1
9.381204	-1.039515906	0.30617	2.74E-05	0.005041	Gm37105
108.4895	0.507114829	0.150321	3.14E-05	0.005704	Coro2a
442.1935	0.923555051	0.281239	3.49E-05	0.006246	Ifitm1
16.11678	-0.820544675	0.24704	3.68E-05	0.006514	Vpreb3
110.8544	2.080991763	0.675135	4.06E-05	0.007083	Ighg3

322.5441	-0.387110032	0.11702	4.55E-05	0.007843	Nkapd1
32.09057	-0.82145829	0.262932	6.65E-05	0.011302	Pax5
130.0814	-0.568857238	0.182809	6.80E-05	0.011411	Asns
608.8048	-0.503508296	0.162248	7.44E-05	0.012332	Ifi30
4.624158	-1.709153552	0.553416	7.79E-05	0.012749	Hoxa7
5.956487	-1.327020016	0.429978	7.88E-05	0.012749	Rag1
37.17933	-0.582377198	0.188477	8.07E-05	0.012905	Dntt
33.0422	-0.533895803	0.174714	8.91E-05	0.013922	Zfp318
587.9822	-0.30607516	0.096131	8.91E-05	0.013922	N4bp2l2
112.6877	0.338107779	0.108458	9.92E-05	0.015323	Pigyl
470.9928	-0.332032499	0.108978	0.000117	0.017777	Snrnp48
17.95557	-0.809174032	0.275137	0.000118	0.017777	Ccr7
12.37343	-1.091703963	0.375975	0.000125	0.018632	Gm28187
4550.924	-0.20656446	0.06382	0.000128	0.018832	Son
172.5532	-0.317741187	0.104824	0.000129	0.018895	Rabep1
50.49038	0.570724169	0.198852	0.000139	0.020092	Slc6a13
18.1978	-0.663761637	0.229276	0.000142	0.0203	Ppfibp1
26.61383	1.10506451	0.388676	0.000144	0.0203	Tac2
33.68418	-1.069044182	0.376483	0.000146	0.020368	Lyve1
26.13706	-0.732020686	0.259063	0.000161	0.022318	Dtwd1
135.4915	0.669190484	0.237471	0.000166	0.022822	H2-Q7
1092.416	-0.233711468	0.077174	0.000182	0.02319	Ralbp1
3597.839	0.364242767	0.126617	0.000185	0.02319	Fcnb
133.0262	-0.313694195	0.10658	0.000172	0.02319	Acbd6
1640.83	0.242708149	0.080457	0.000179	0.02319	Clec12a
27.22965	-0.641262391	0.233191	0.000182	0.02319	Blnk
65.62676	-0.888948646	0.322545	0.000175	0.02319	Stfa1
95.0987	0.961034648	0.34767	0.000176	0.02319	Rpl17-ps5
139.6983	-0.346733851	0.119934	0.000183	0.02319	E430021H15Rik
157.1449	-0.570387135	0.202199	0.00018	0.02319	C230096K16Rik
200.8156	0.362309526	0.1261	0.000186	0.023198	Oas3
308.6689	0.323088624	0.111276	0.000195	0.023976	Tap1
55.64298	0.421018475	0.149351	0.000205	0.025092	Igtp
224.0762	0.392119593	0.139618	0.000214	0.025859	B4galnt1
62.83644	-0.425994996	0.152195	0.000217	0.026003	Afdn
26.7254	-0.796202303	0.293219	0.00022	0.026108	Rab23
4.681469	-0.065256504	0.126048	0.000244	0.028724	Uaca
584.9936	-0.317383104	0.113661	0.000263	0.030281	Pkn2
75.516	0.422417464	0.155523	0.000261	0.030281	Glod4
835.5485	-0.233784237	0.080566	0.000264	0.030281	Phip
125.8714	0.311021138	0.111318	0.000286	0.032523	Adar

272.1053	0.380755858	0.14005	0.000292	0.033016	Trafd1
78.50333	0.398156359	0.147143	0.000297	0.033023	Xk
172.3749	-0.617128577	0.241755	0.000297	0.033023	Dnah8
15.69282	0.69838305	0.273103	0.000346	0.037229	Nqo1
15.03218	0.856872344	0.33635	0.000345	0.037229	Dmpk
1267.246	-0.242368309	0.086735	0.000346	0.037229	Sltm
323.8297	-0.223926671	0.079018	0.000342	0.037229	Pxylp1
24.46166	-0.681130108	0.271549	0.000396	0.042124	Jup
622.2799	-0.270624666	0.100114	0.000398	0.042124	Zfp292
2385.964	-0.323443983	0.123183	0.000404	0.042435	Ppig
8.594592	-0.917808926	0.370744	0.000409	0.042613	Fabp4
400.7393	0.263787268	0.097941	0.000413	0.042699	Gpc1
313.0034	-0.434182249	0.172583	0.000428	0.043868	Tut4
4.477437	1.212318804	0.488625	0.000436	0.043935	Piwil2
22.38876	0.586433667	0.236126	0.00044	0.043935	H2bc7
2379.813	0.424657822	0.16776	0.000434	0.043935	Ighm
12.19092	-0.077188295	0.158153	0.000442	0.043935	Gm43329
555.4473	-0.171814308	0.059869	0.000459	0.045359	Tmod3
356.2804	0.373068739	0.14734	0.000463	0.045414	Ceacam10
15.62545	-0.767318066	0.320148	0.000482	0.046946	Trib2
20.124	-0.762253191	0.322625	0.000487	0.047031	Ednra
3.39579	-0.040917819	0.088654	0.00052	0.049919	Hoxa5

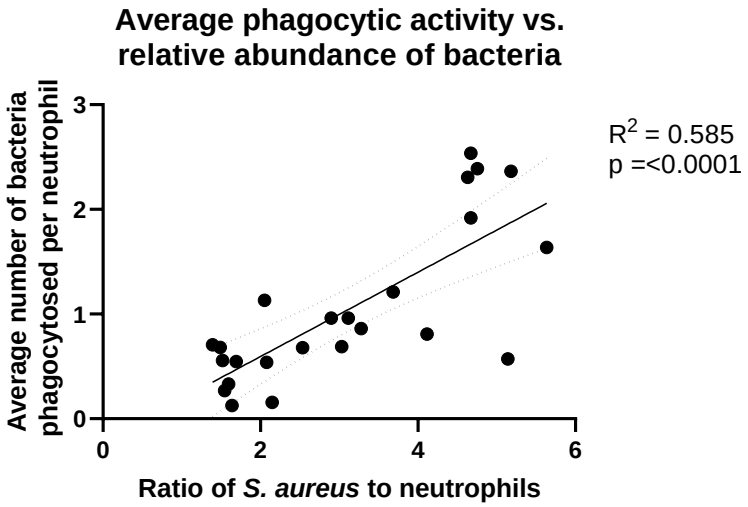


Figure S1. Bacterial abundance affects phagocytic activity
The ratio of bacteria to neutrophils showed a significant positive correlation with the average number of bacteria phagocytosed per neutrophil.

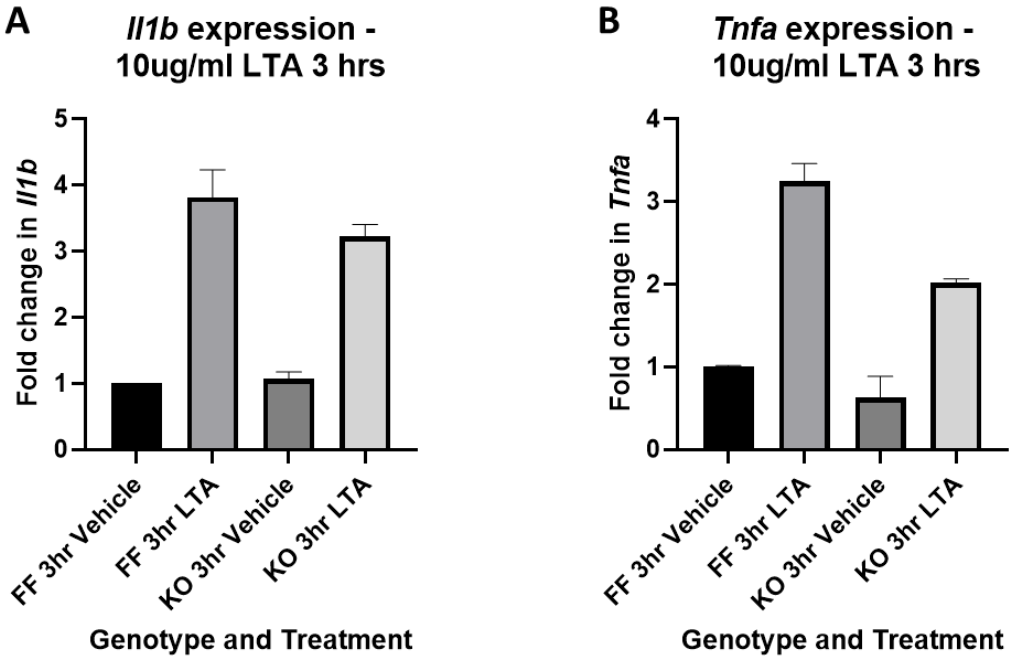


Figure S2. Stimulation with 10µg/mL of LTA leads to upregulation of inflammatory cytokine expression
Neutrophils from *Tet2*-FF (FF) and *Tet2*-KO (KO) mice were treated with either 10µg/mL LTA, 2% FBS and PBS for 3 hours or vehicle. *Tnfa* and *Il1b* expression were measured at each time point through qRT-PCR using *Gapdh* as the housekeeping gene. Fold changes is reported relative to the *Tet2*-FF Vehicle treated sample **A)** *Il1b* mRNA expression after stimulation with 10µg/mL LTA for 3 hours. **B)** *Tnfa* mRNA expression after stimulation with 10µg/mL LTA for 3 hours. Treatment with 10 µg/mL LTA for 3 hours led to upregulation *Il1b and Tnfa* expression.

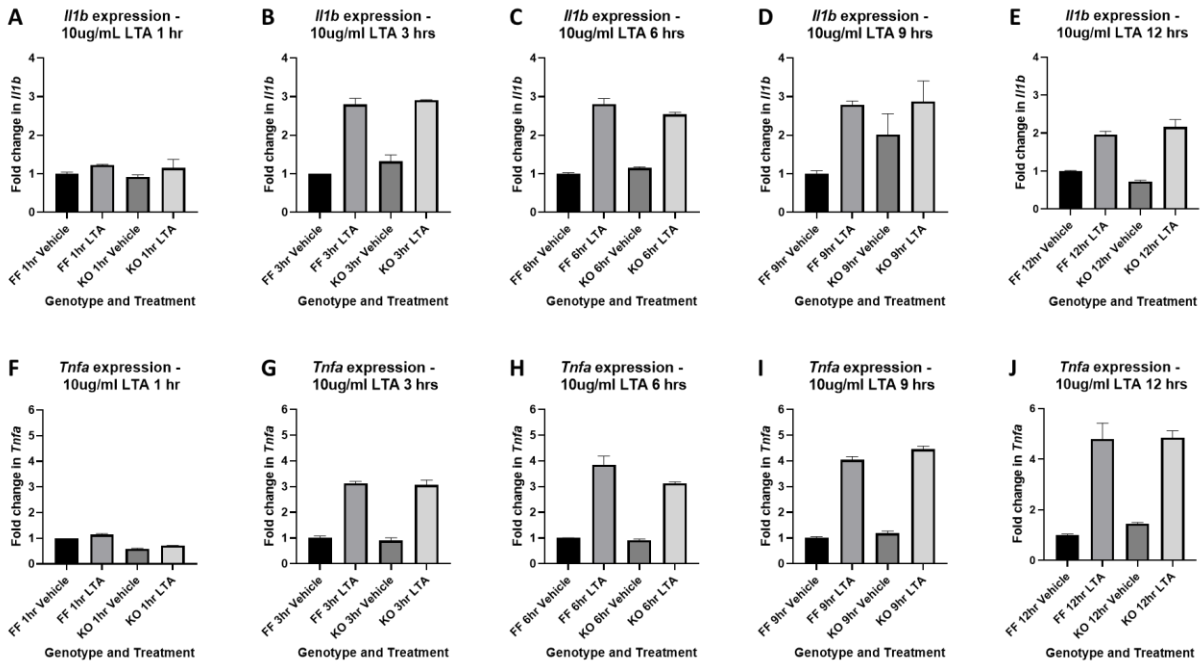


Figure S3. Expression of Inflammatory cytokines at the 1, 3, 6, 9, and 12 hour time points of LTA stimulation

Neutrophils from *Tet2*-FF (FF) and *Tet2*-KO (KO) mice were treated with either 10µg/mL LTA, 2% FBS and PBS for 3 hours or vehicle. *Tnfa* and *Il1b* expression were measured at each time point through qRT-PCR using *Gapdh* as the housekeeping gene. Fold changes is reported relative to the *Tet2*-FF Vehicle treated sample at each time point. *Il1b* expression is depicted in the top row of figures **(A-E)** while *Tnfa* expression is depicted in the bottom row of figures **(F-J).** The 3, 6, and 9 hour time points show prominent upregulation of the two inflammatory cytokines.

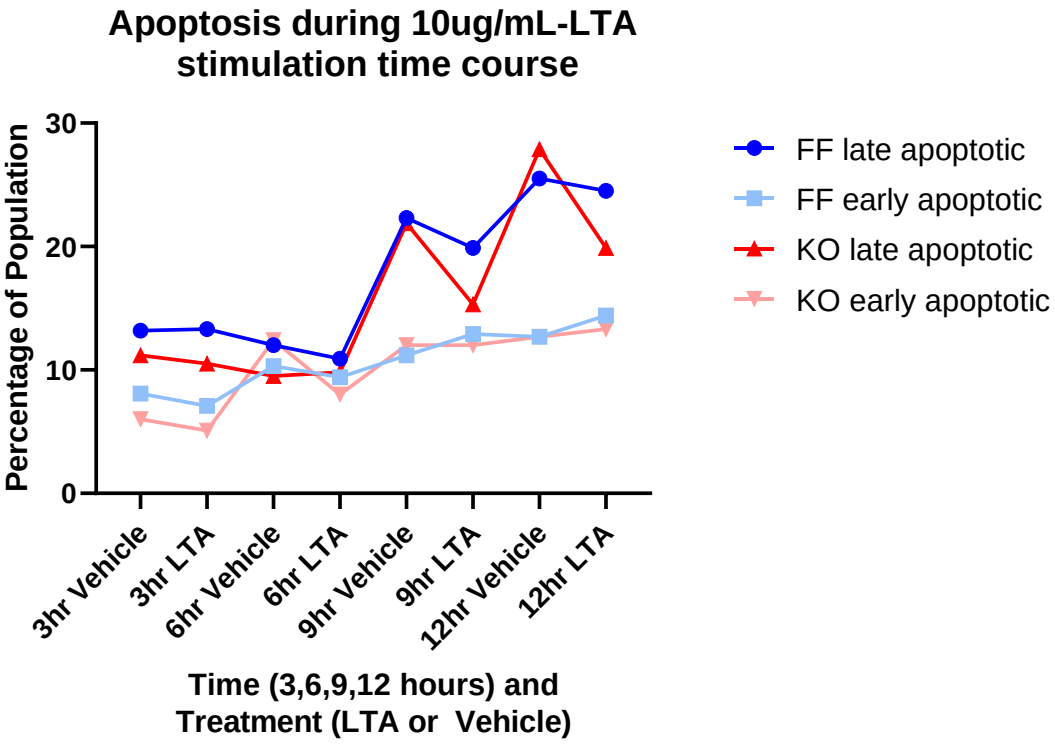


Figure S4. Apoptosis at the 3, 6, 9, and 12 hour time points of LTA stimulation
Neutrophils isolated from *Tet2*-KO and *Tet2*-FF mice were treated with 10µg/mL LTA, 2% FBS in PBS or Vehicle for up to 12 hours. Flow cytometry was performed to assess apoptotic status based on Annexin-V, PI staining. Annexin-V ^{hi} and PI ^{hi} populations were considered late apoptotic, and Annexin-V ^{hi} and PI ^{low} populations were considered early apoptotic. The proportion of early and late apoptotic cells increased as the duration of stimulation increased. At later time points, LTA treated neutrophils showed less apoptosis than Vehicle treated neutrophils. The 6 hour time point appears to have similar apoptotic rates compared to the 3 hour time point.

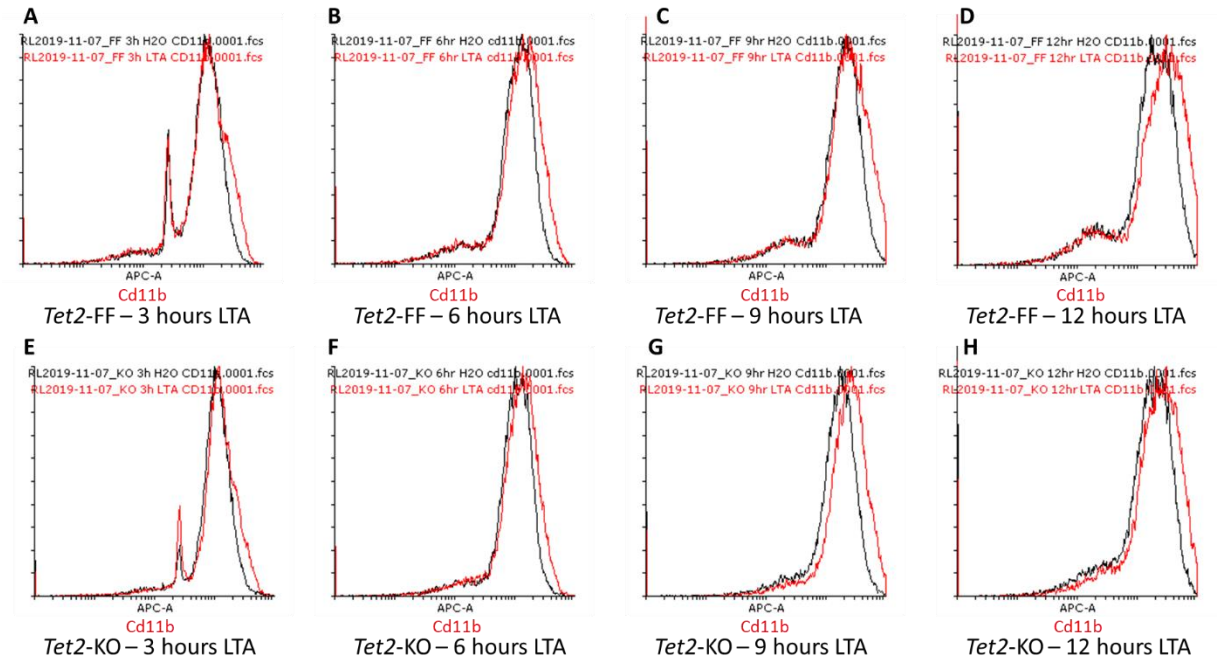
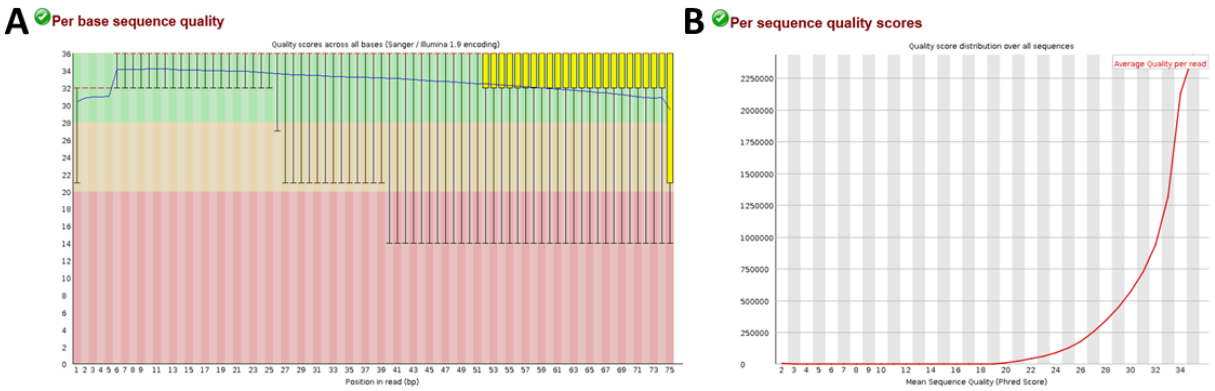


Figure S5. Surface expression of Cd11b at the 3, 6, 9, and 12 hour time points of the LTA stimulation

Neutrophils isolated from *Tet2*-KO and *Tet2*-FF mice were treated with 10µg/mL LTA, 2% FBS in PBS or Vehicle for up to 12 hours. Flow cytometry was performed to analyze the surface expression of Cd11b and overlay histograms were generated to show the difference in cd11b surface expression in LTA treated samples (red) and Vehicle treated samples (black). Cd11b expression of neutrophils from *Tet2*-FF mice over the time course are shown on the top row **(A-D),** while Cd11b expression of neutrophils from *Tet2*-KO mice over the time course are shown in the bottom row **(E-H).** During all time points, LTA treated neutrophils showed greater surface expression of Cd11b compared to vehicle treated neutrophils.



C Overrepresented sequences

Sequence	Count	Percentage	Possible Source
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	8783	0.1536660542535824	No Hit
CGCGACCTCAGATCAGACGTGGCGACCCGCTGAATTTAAGCATATTAGTCAGCGGAGGAAAAAAAAAAAAAAAAA	7574	0.13251357109377584	No Hit

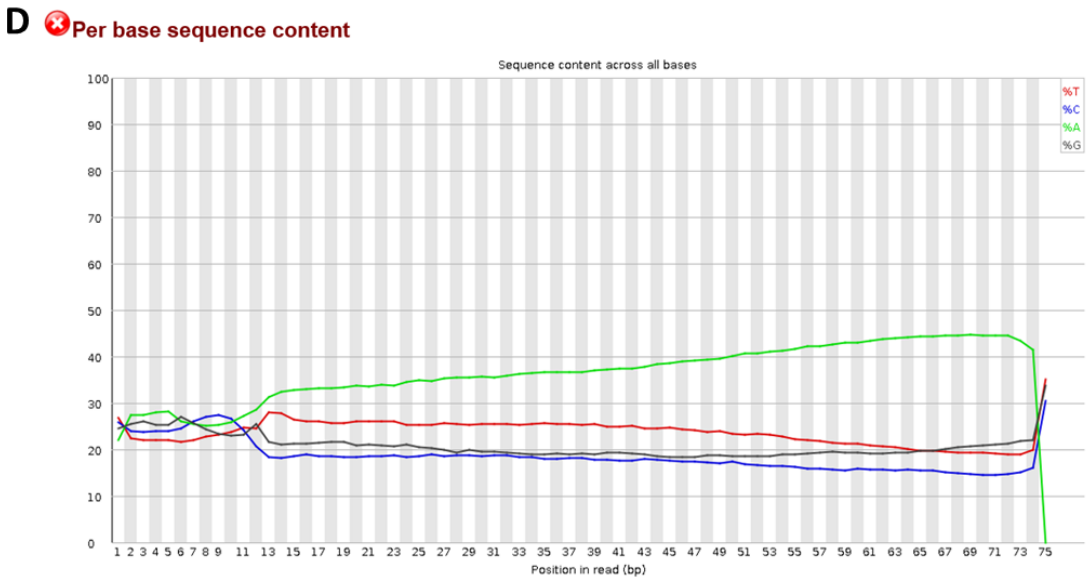


Figure S6. FastQC quality metrics for a representative raw fastq file
A) Chart depicting the sequence quality across different positions of the read. Sequence quality is high overall, but starts to decrease near the end of the read. **B)** Histogram of average sequence quality among all the reads. The majority of reads are very high quality, few reads of moderate quality and almost no low quality reads. **C)** Table of the most overrepresented sequences. We suspect that one sequence is an adaptor sequence, the other sequence appears to have part of its poly A tail captured in the read. **D)** FastQC per base sequence content module report. The module gave an error for all samples because Adenosine is overrepresented in the reads. The abundance of Adenosine in the reads probably reflects parts of the poly A tail being captured in the read.

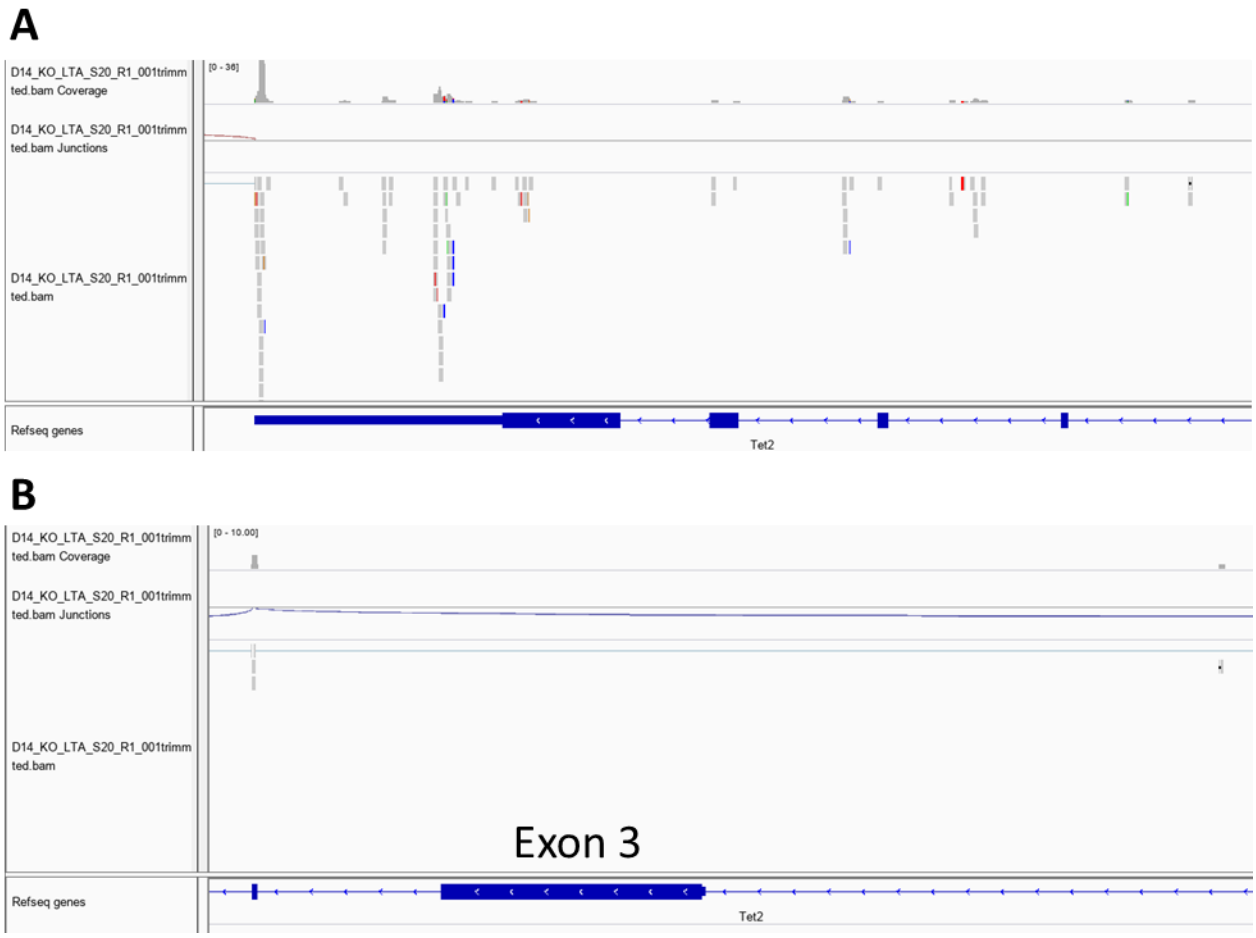


Figure S7. Read alignments on the *Tet2* gene in a representative *Tet2*-KO RNA-seq sample Integrative Genome Viewer software was used to visualize read alignments in bam/bai files **A)** Read alignments near the 3' end of the *Tet2* gene. There is denser read coverage at the 3' end of the gene. **B)** Read alignments near exon 3 on the 5' end of the Tet2 gene. Exon 3 is predicted to be excised in *Tet2*-KO mice and we see no reads aligning to this region. Further more we see read coverage is sparse near the 3' end of the gene with only a few reads aligning to exon 2.

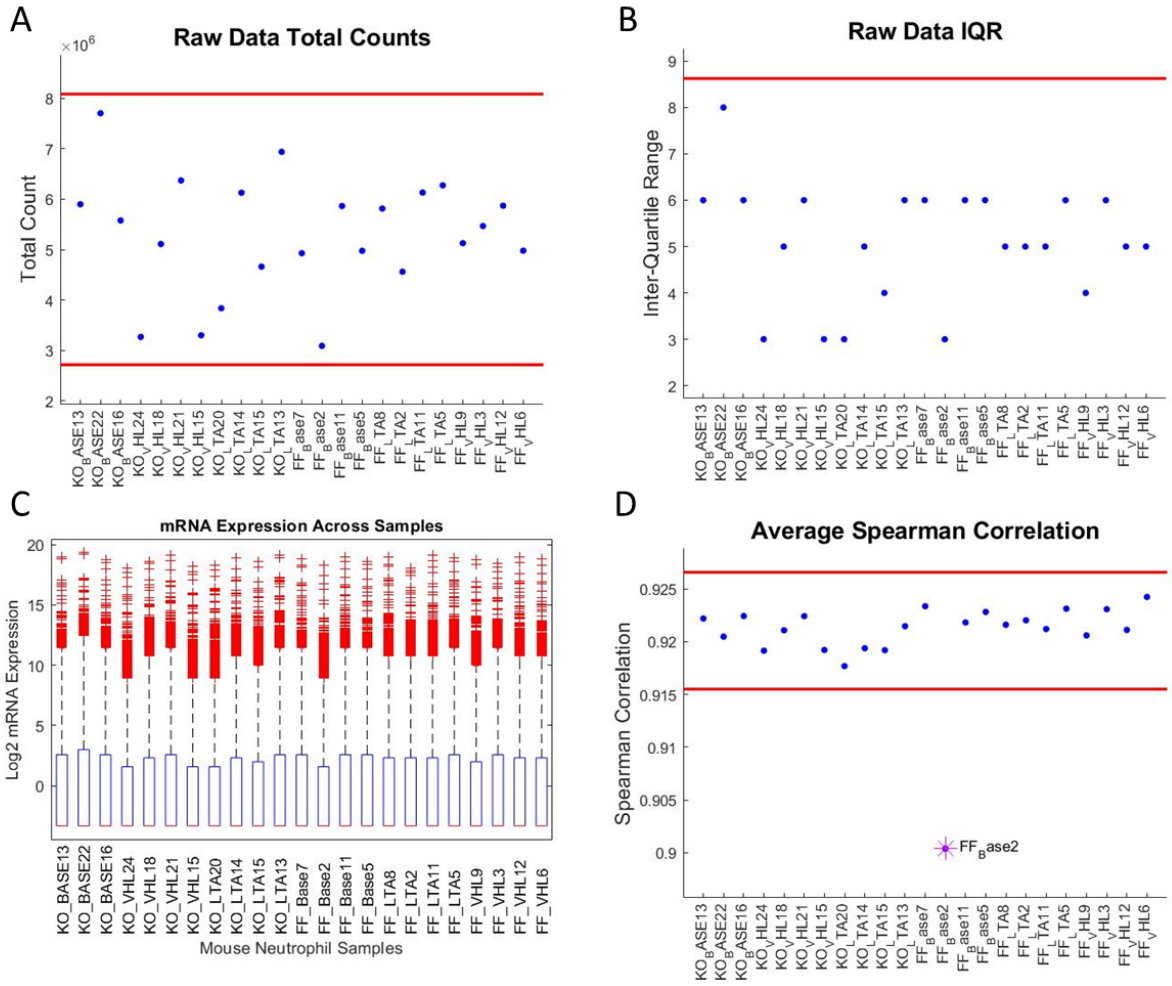


Figure S8. Visualizing RNA-seq data for pre-processing
A) The sequencing depth of each library. **B)** The interquartile range of reads assigned per transcript. **C)** Boxplots of log2-transformed expression of the mRNAs in each library. **D)** The average spearman correlation of each library to all others.

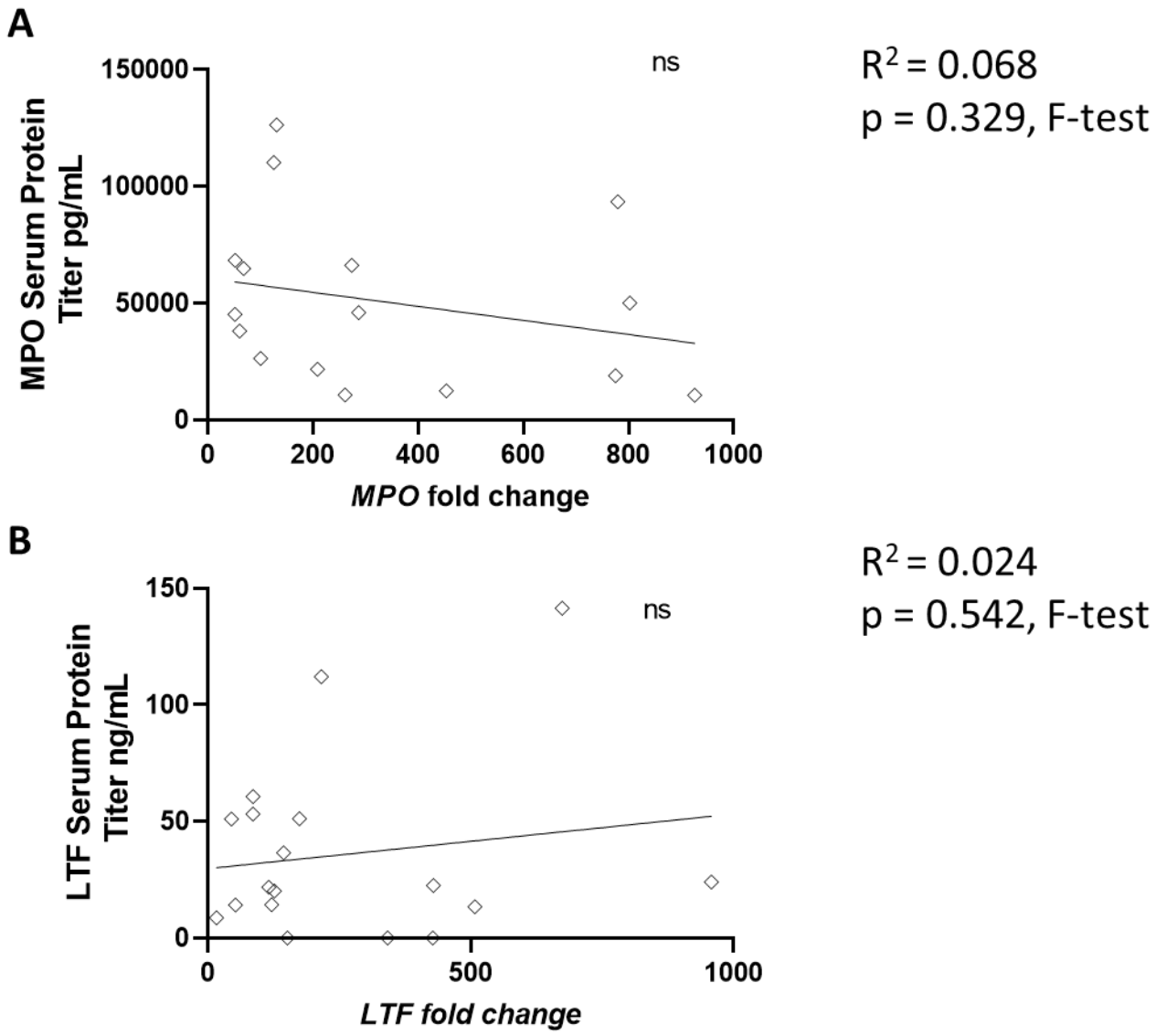


Figure S9. MPO and LTF serum protein titers do not correlate with mRNA expression
16 samples with both qRT-PCR and ELISA results, were analyzed in a linear regression model comparing serum protein titers and mRNA expression. **A)** There was no correlation between MPO serum titers and *MPO* mRNA expression (p=0.319, F-test). The line of best fit was Y = -30.02X + 60628 (R^2 = 0.068). **B)** No significant correlation was found between LTF serum protein titers and *LTF* mRNA expression (p=0.542). The line of best fit was Y = 0.023X + 29.80 (R^2 = 0.024).

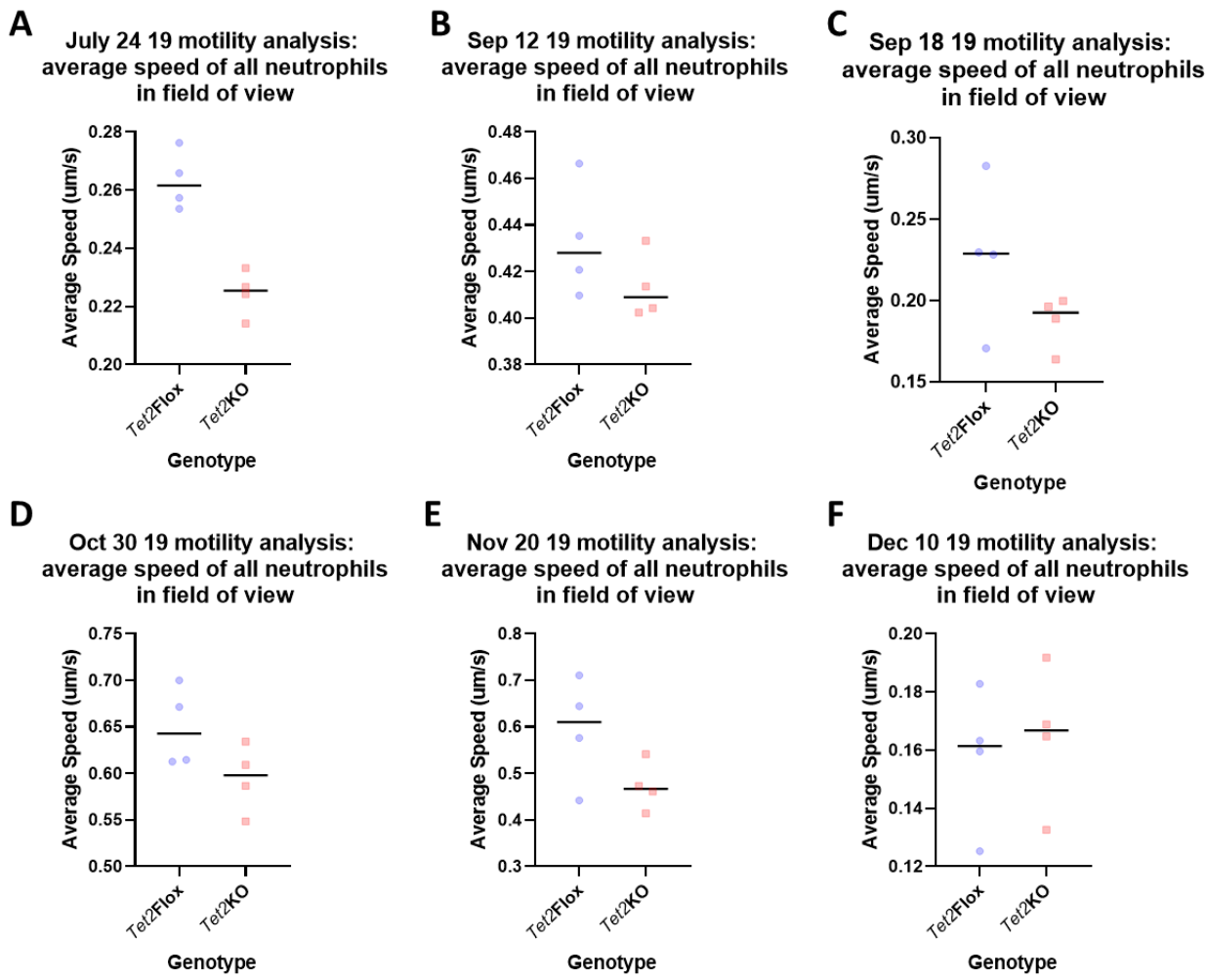


Figure S10. Motility analysis results from six independent experiments
Panels **A-F** each depict the motility results in one of the experiments. In five of the six experiments, neutrophils from *Tet2*-FF mice show increased motility compared to neutrophils from *Tet2*-KO mice.

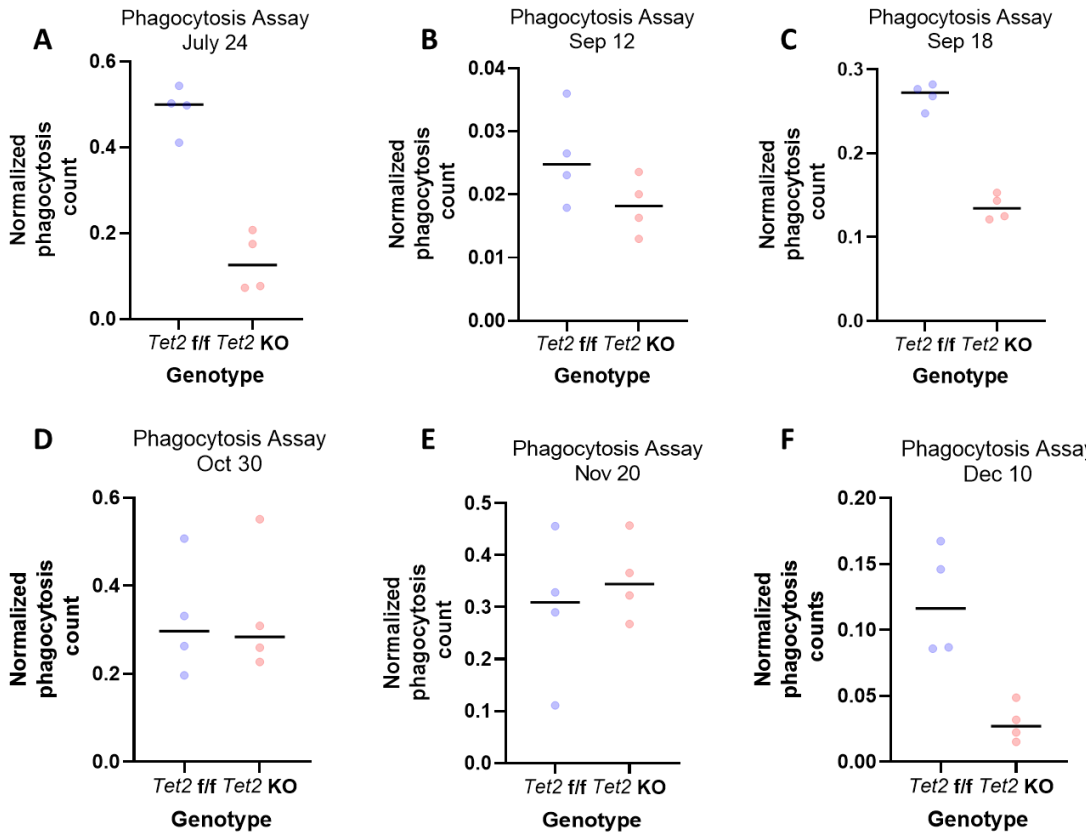


Figure S11. Phagocytosis assay results from 6 independent experiments
Panels **A-F** each depict the phagocytosis results in one of the experiments. In five of the six experiments, neutrophils from *Tet2*-FF mice show increased motility compared to neutrophils from *Tet2*-KO mice.

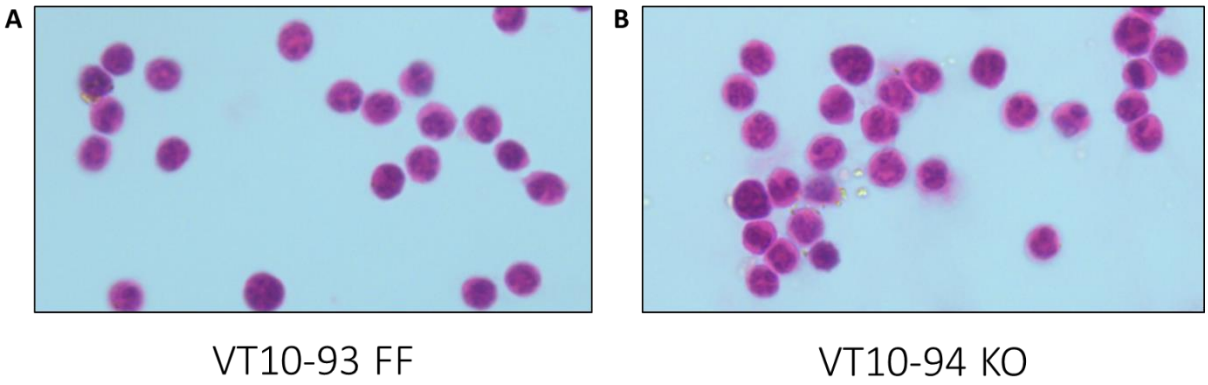


Figure S12. Representative H & E images of neutrophils isolated by negative immunomagnetic selection
A) H & E stain of neutrophils isolated from a *Tet2*-FF mouse. **B)** H & E stain of neutrophils isolated from a *Tet2*-KO mouse. The majority of the cells appear to be neutrophils or neutrophil precursors with multi-lobed nuclei. These images correspond to the flow cytometry data in Table S10.

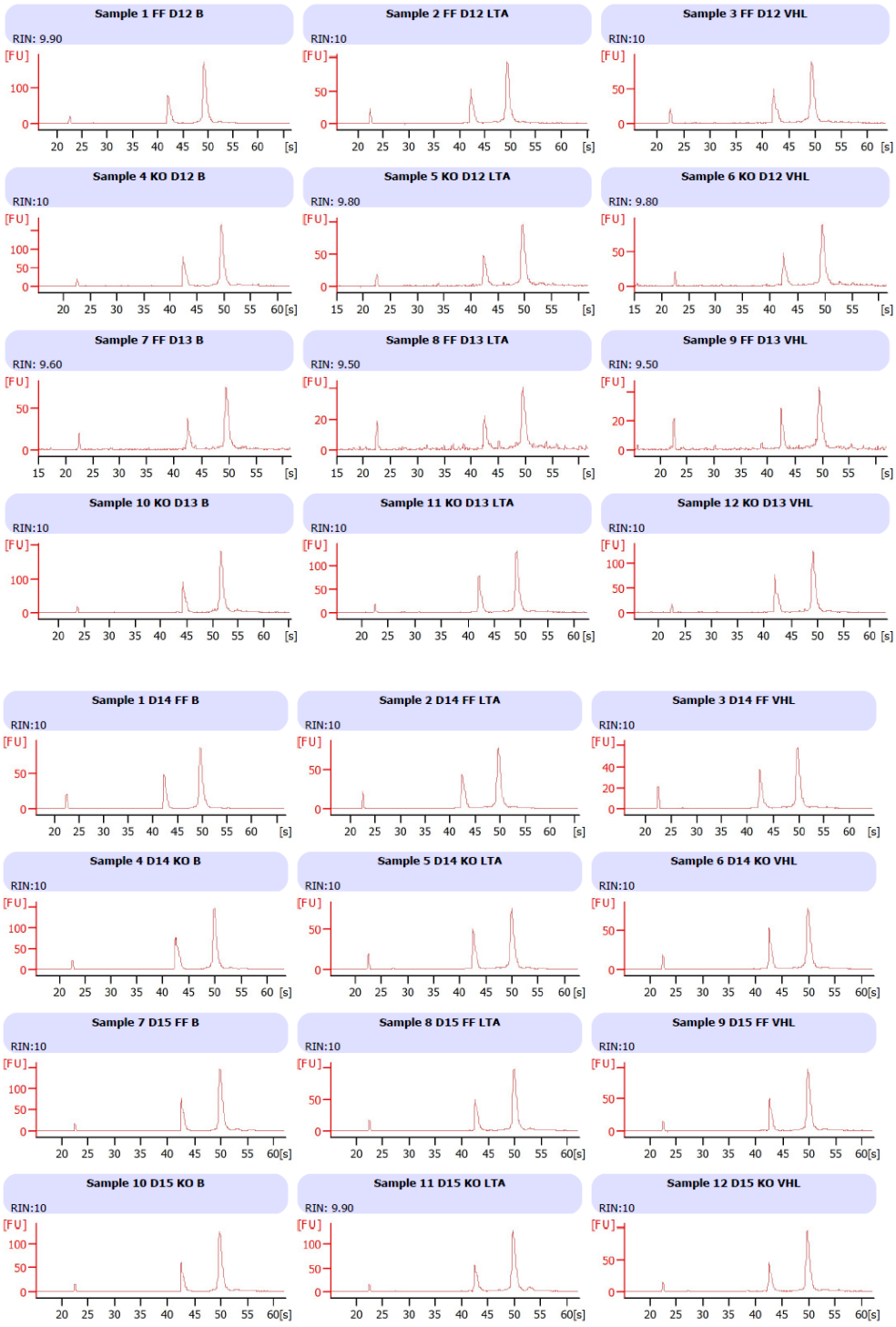


Figure S13. Bioanalyzer results for RNA-seq samples
All 24 samples were analyzed using Bioanalyzer RNA Nano Kit (Agilent, Santa Clara, CA) and their RNA Intergrity Numbers (RIN) are reported. All RNA samples had excellent quality with little to no signs of degradation.

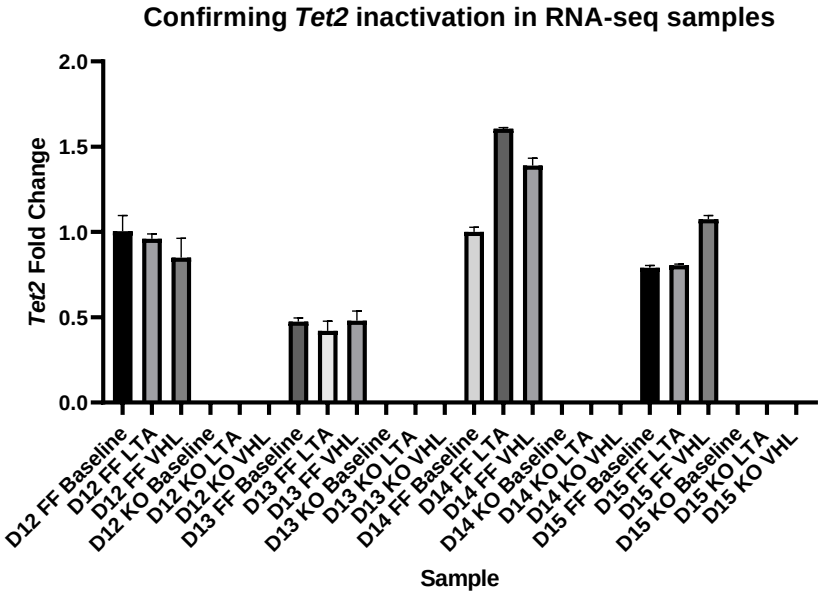


Figure S14. *Tet2* expression in the RNA-seq samples
qRT-PCR was performed on an aliquot of the RNA to assess *Tet2* expression and *Gapdh* was used as the housekeeping gene. Fold change is reported relative to the *Tet2*-FF baseline sample harvested in Dec 12, 2019. As expected, the *Tet2*-KO samples do not show expression of intact *Tet2* mRNA transcripts.

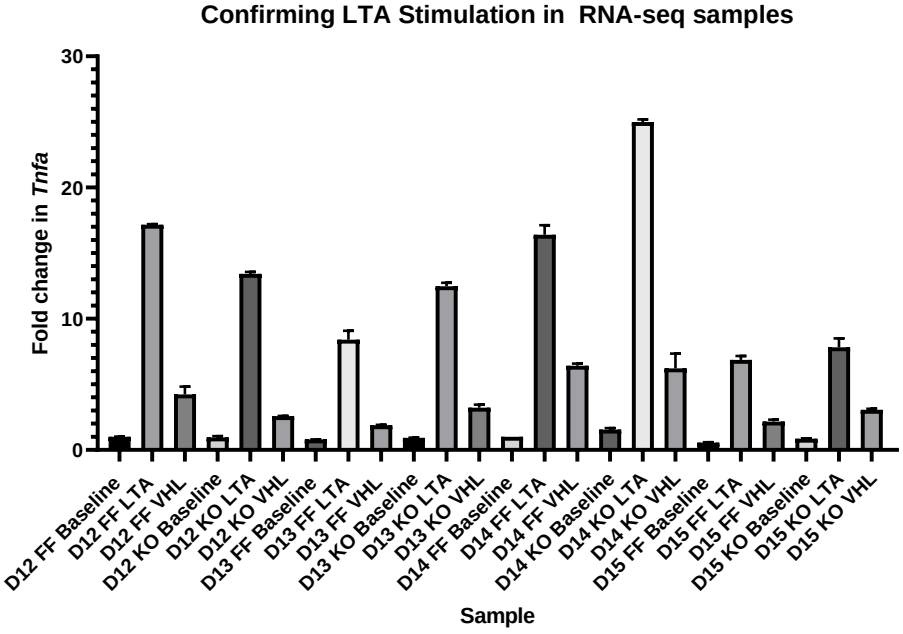


Figure S15. *Tnfa* expression in the RNA-seq samples
qRT-PCR was performed on an aliquot of the RNA to assess *Tnfa* expression. *Gapdh* was used as the housekeeping gene. Fold change is reported relative to the *Tet2*-FF baseline sample harvested in Dec 12, 2019. The LTA treated samples show pronounced upregulation of *Tnfa* expression compared to the vehicle-treated samples suggesting neutrophil activation

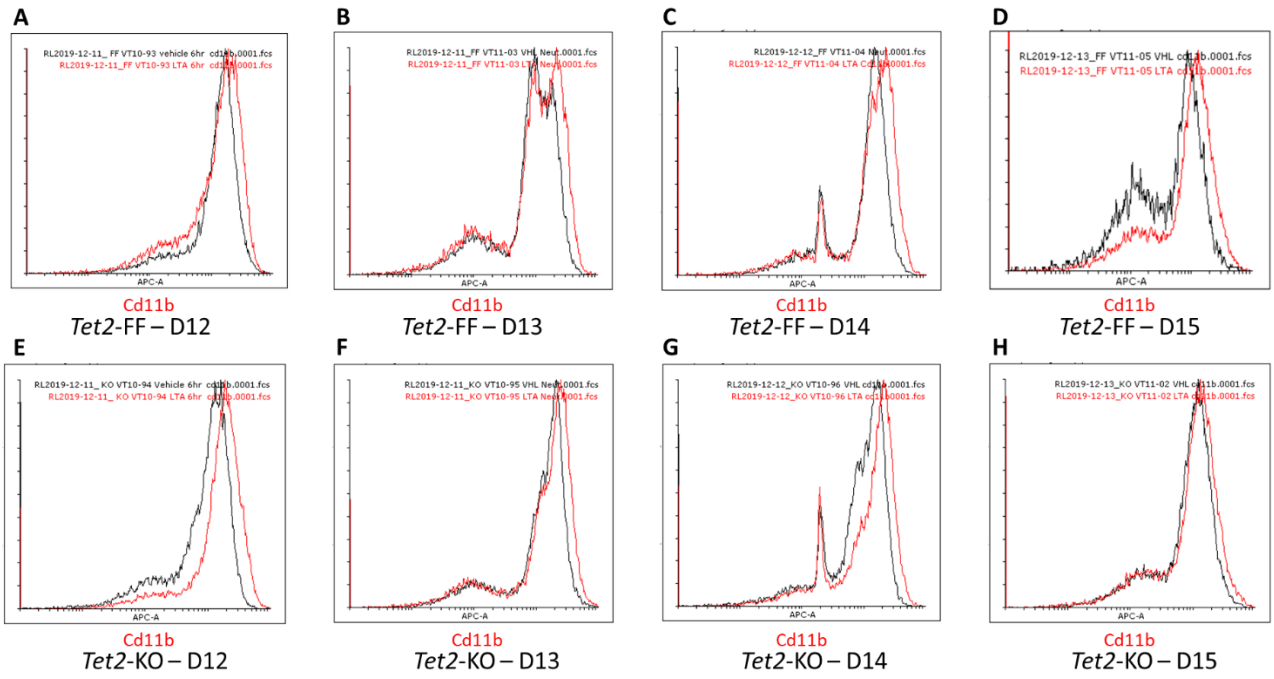


Figure S16. Surface expression of Cd11b in the RNA-seq samples
Flow cytometry was performed to analyze the surface expression of Cd11b and overlay histograms were generated to show the difference in Cd11b surface expression in LTA treated (red) and Vehicle-treated (black) neutrophils derived from the same mouse. Cd11b expression in samples derived from *Tet2*-FF mice are shown on the top row **(A-D)**, while Cd11b expression in samples derived from *Tet2*-KO mice over the time course are shown in the bottom row **(E-H)**. In all samples, Cd11b surface expression is higher in LTA-treated samples than Vehicle treated samples.

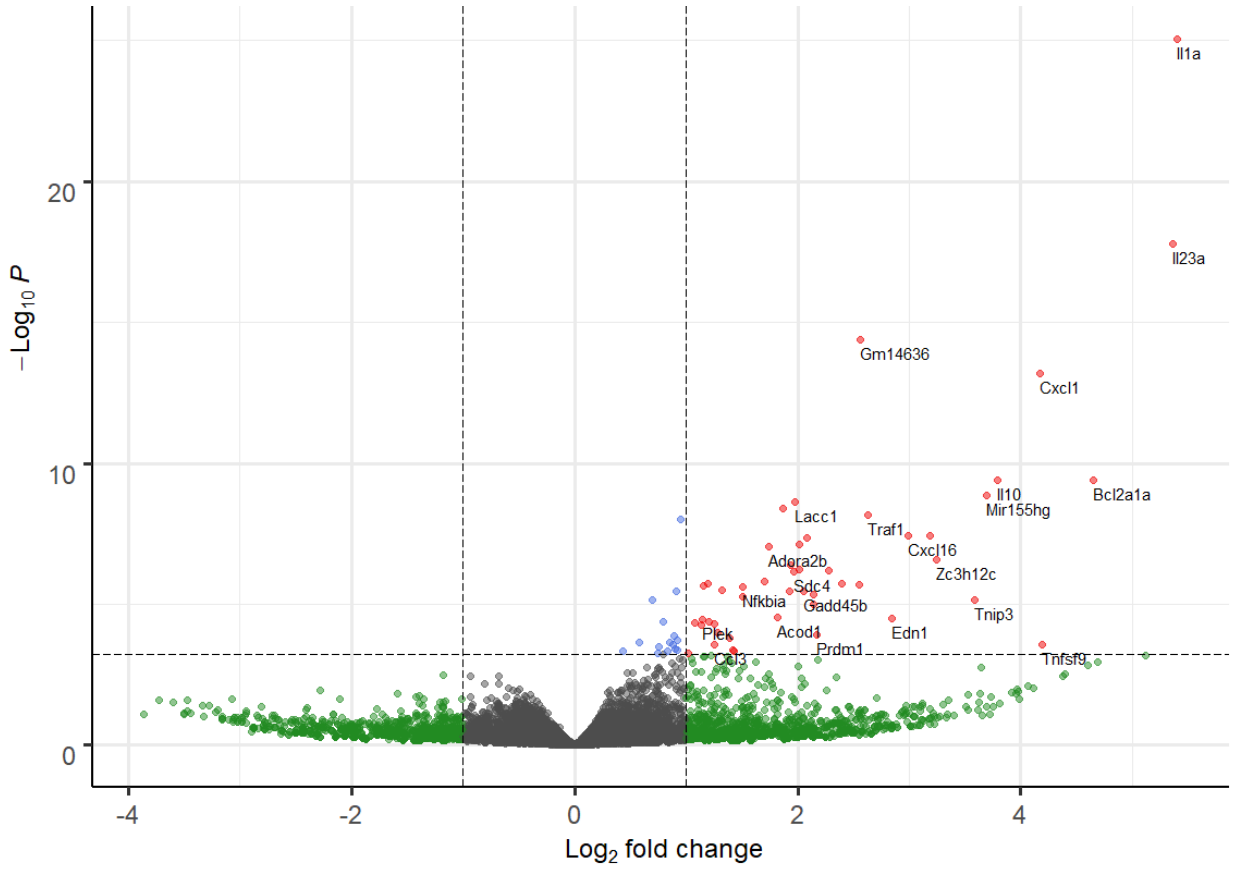


Figure S17. Volcano plot of differentially expressed genes between neutrophils treated with LTA and vehicle
Inflammatory chemokines and cytokines were upregulated in the LTA treated neutrophils. The threshold of significance was padj < 0.10, and transcripts with an absolute log2 fold change value above 1 were considered to have a large fold change. Transcripts meeting both the significance and fold change thresholds are marked in red, those that only meet the significance threshold are blue, those that only meet the fold change threshold are in green and the rest that meet neither threshold are in grey.