



**University of
Nottingham**

UK | CHINA | MALAYSIA

CHARACTERIZATION OF DEFA1A3 GENE EXPRESSION IN HUMAN BLOOD CELLS

IGNACIO RODRÍGUEZ PASTRANA

Email: mbxir2@exmail.nottingham.ac.uk

Supervised by

PROFESSOR JOHN ARMOUR

Email: plzjaa@exmail.nottingham.ac.uk

THESIS SUBMITTED TO THE UNIVERSITY OF NOTTINGHAM FOR THE DEGREE OF
MASTER BY RESEARCH IN MOLECULAR GENETICS

September 2019

ACKNOWLEDGEMENT

I would first like to thank and express my special appreciation to my supervisor Professor John Armour for his endless support, encouragement and patience and for his priceless guidance and advices during the last year. I am extremely grateful for the opportunity of being part of his research group during my master. I would also like to thank Nzar Shwan, Reem Alhamidi and Assel Arginbekova as part of the research group for their help in John's absence.

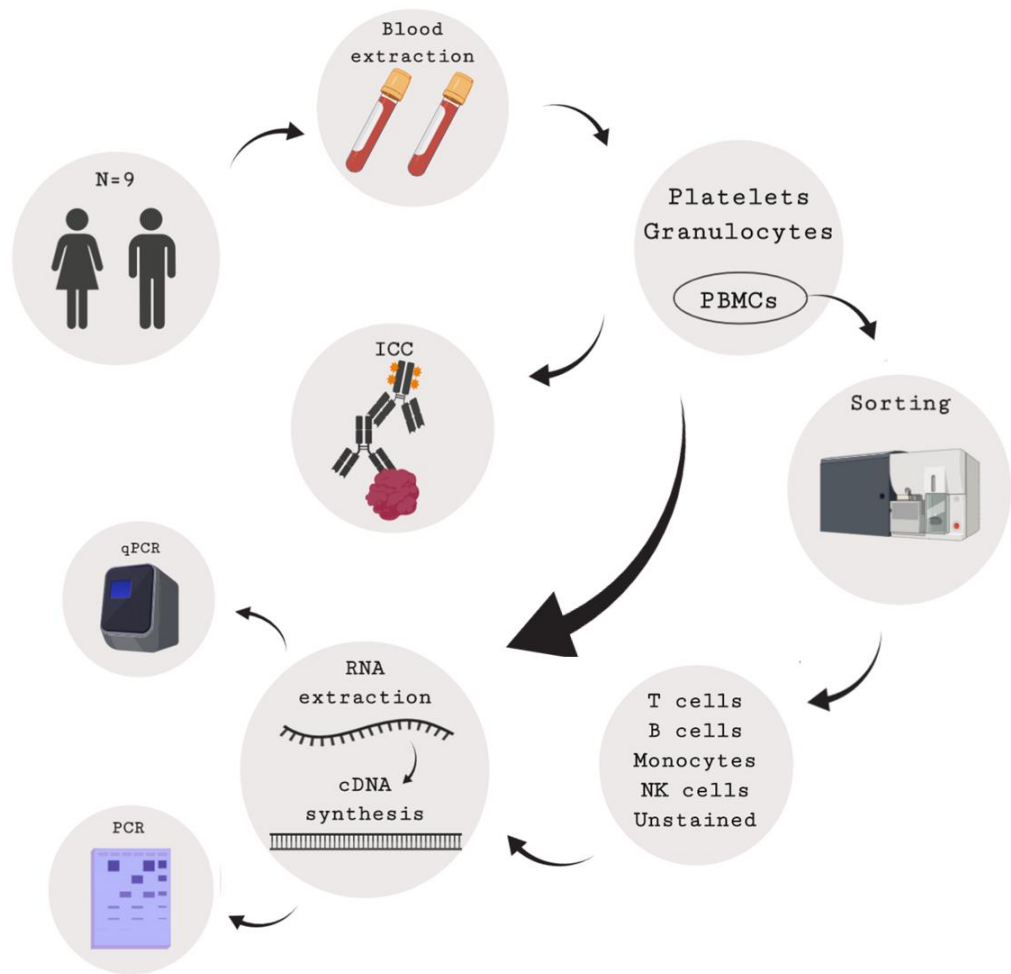
I would also want to thank my second supervisor Luisa Martínez Pomares and Sirina Muntaka for their tremendous assistance during the project. Their empathy and hard-working personalities are contagious, and their laboratory was always open whenever I had a question about my research.

I would also want to thank Lucy Fairclough who without her participation and guidance in the NK cell world, this project could not have been successfully finished. Moreover, I would like to thank David Onion and Nicola Croxall from the flow cytometry facility for their irreplaceable help during the cell sorting procedure.

Finally, a special thanks to my family. Words cannot describe how grateful I am for all the sacrifices they have done on my behalf during this year. I owe my deepest gratitude for their extremely incalculable love, patience, encouragement and support. Without them finishing this master would have not been possible.

ABSTRACT

Alpha-defensins are small cationic peptides important in the immune response as they have antibacterial and virus properties and regulate the production of proinflammatory cytokines. Alpha-defensin 1-3 peptides are encoded in the copy number variable *DEFA1A3* gene at chromosome 8p23.1 and are believed to be possible contributors to various inflammatory and autoimmune diseases. Here, we have answered the previously unresolved biological questions of establishing in which blood cell populations *DEFA1A3* is expressed, assessing the *DEFA1A3* mRNA expression levels and correlating the presence of *DEFA1A3* mRNA with the presence of alpha-defensin 1-3 peptides. This project used a platelet-enriched plasma and density-gradient centrifugation protocols to separate platelets, granulocytes and peripheral blood mononuclear cells (PBMCs). PBMC subpopulations – T cells (CD3⁺), B cells (CD19⁺), monocytes (CD14⁺), natural killer cells (NK cells, CD56⁺) and other minor PBMC subsets – were further purified by a fluorescence activated cell sorting (FACS) procedure. Our results from *DEFA1A3*-specific polymerase chain reaction (PCR), quantitative-real time polymerase chain reaction (qRT-PCR) and immunocytochemistry, reconfirmed the presence of *DEFA1A3* mRNA in granulocytes, which contributed to 71.4% of the whole blood *DEFA1A3* gene expression levels, and their corresponding alpha-defensin 1-3 peptides respectively. We have also demonstrated the presence of *DEFA1A3* mRNA, its contribution to the whole blood *DEFA1A3* gene expression levels and the presence of alpha-defensin 1-3 peptides in platelets (1.4%) and in PBMCs (27.2%) – distributed between T cells (0.8%), B cells (1.0%), monocytes (3.4%), NK cells (83.8%) and the unstained fraction (11.0%). In conclusion, *DEFA1A3* mRNA is present in all unstimulated white blood cells and platelets and can be associated with the presence of alpha-defensin 1-3 peptides. In whole blood, granulocytes are the major contributors to the *DEFA1A3* gene expression levels; however, surprisingly the minor CD56⁺ NK cell subpopulation accounts for the majority of the *DEFA1A3* gene expression levels attributed to PBMCs.



CONTENT

1. INTRODUCTION.....	6
1.1. Circulating immune blood cells.....	6
1.1.1. Lymphoid development process.....	7
1.1.2. Myeloid development process.....	9
1.2. Online blood datasets and new haematological studies.....	15
1.3. <i>DEFA1A3</i> locus structure.....	19
1.4. Alpha-defensin peptides and their role in the immune response.....	25
1.5. Alpha-defensin 1-3 peptides in disease.....	28
2. BIOLOGICAL QUESTIONS AND OBJECTIVES.....	32
3. RESULTS.....	34
3.1. Characterization of <i>DEFA1A3</i> mRNA in blood samples.....	34
3.2. <i>DEFA1A3</i> mRNA levels are low in purified platelets.....	37
3.3. Peripheral blood mononuclear cells have <i>DEFA1A3</i> mRNA and alpha-defensin 1-3 peptides.....	42
3.4. Fluorescence activated cell sorting of peripheral blood mononuclear cell subpopulations.....	44
3.5. Characterization of <i>DEFA1A3</i> mRNA in T cells, B cells and monocytes.....	48
3.6. NK cells contribute to the majority of the <i>DEFA1A3</i> mRNA from peripheral blood mononuclear cells.....	50
4. DISCUSSION AND CONCLUSIONS.....	55
5. MATERIALS AND METHODS.....	68
6. REFERENCES.....	75

INTRODUCTION

Circulating immune blood cells

Blood is one of the principal human fluids and its main function is to transport nutrients and oxygen to the cells in the tissues and eliminate their metabolic waste. In vertebrates, blood is composed of two main components – the plasma and the blood cells. The plasma is the non-cellular part of the blood and represents 55% of the total volume. It is mainly composed of water in which different components are dissolved such as proteins (70%), glucose, clotting factors, electrolytes and hormones (30%). The remaining 45% of the volume is represented by the blood cells each of which delivers a specific function such as oxygen transport, immune response or blood clotting¹.

All blood cells are formed from common precursors located in the bone marrow called pluripotent haematopoietic stem cells. The development of each population depends on the stimuli these pluripotent haematopoietic cells receive, the cell-to-cell interactions (specifically with stromal cells which generate a haematopoietic-inducing microenvironment) and their movement across different niches through the bone marrow. The differentiation process is called haematopoiesis and is summarised in Figure 1. Briefly, pluripotent haematopoietic stem cells initially differentiate into two more specialised ancestors called the common lymphoid progenitor and the common myeloid progenitor which could generate the blood cell populations².

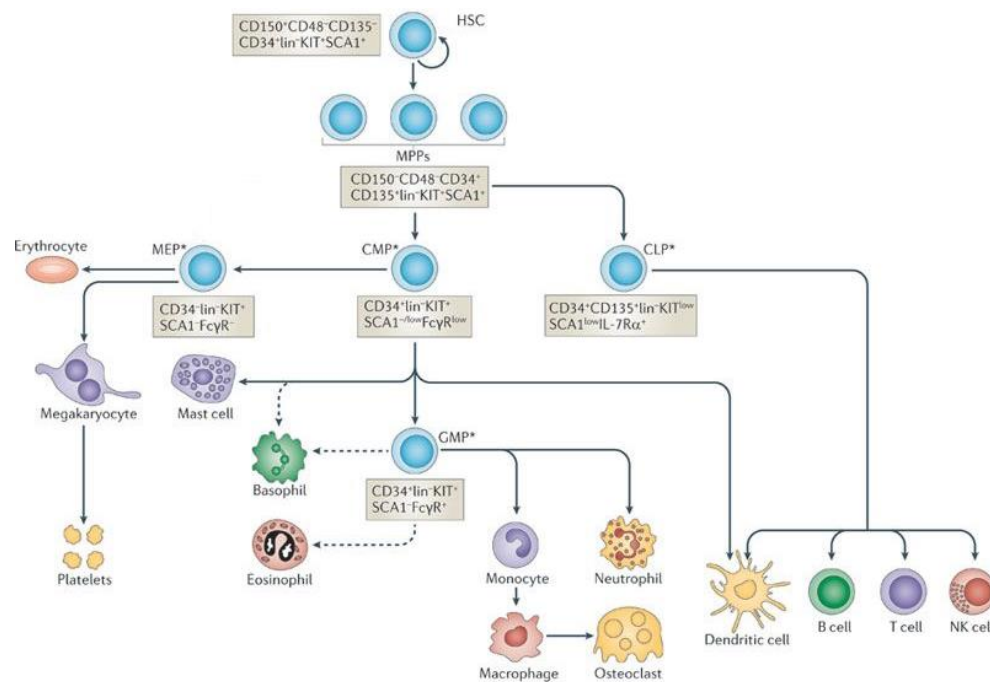


Figure 1 | Schematic-tree representation of the haematopoiesis process. Pluripotent hematopoietic stem cells from the bone marrow (top branch of the tree) initially differentiate into the common lymphoid and the common myeloid progenitors presented in the third branch of the tree. The common lymphoid progenitor will subsequently differentiate into lymphocytes (right segment of the tree) whilst the common myeloid progenitor will develop into granulocytes, monocytes, mast cells, megakaryocytes which generate platelets, and erythrocytes (left segment of the tree). **HSC:** hematopoietic stem cells, **CLP:** common lymphoid progenitor and **CMP:** common myeloid progenitor. Image obtained from Wang and Wagers' publication³.

Lymphoid development process

Lymphocytes are a population of blood cells with a spherical shape, 7-15µm in diameter, with prominent, condensed nuclei and variable cytoplasmic content from scarce (the most common representation formed by a small blue-stained ring surrounding the nuclei) to abundant (Figure 2C-D). In the cytoplasm, it is common to detect the presence of granules which store molecules important for the immune response such as alpha-defensin peptides as will be discussed later. Circulating lymphocytes represent 25-33% of the total white blood cell count and can be mainly classified into three populations – T cells, B cells and natural killer cells (NK cells). The common lymphoid progenitor can differentiate into those cell populations in a process called lymphopoiesis².

In the development of T lymphocytes, the common lymphoid ancestor initially differentiates into an immature thymocyte ($CD3^-CD4^-CD8^-$)⁴, which then migrates from the bone marrow to the thymus⁵ and activates the *Wnt* pathway forming a double positive intermediate ($CD3^+CD4^+CD8^+$)⁴. Subsequently, this double-positive thymocyte is submitted to a positive selection in the cortex area⁶ and a negative selection in the medulla⁵, finally generating mature, single positive T lymphocytes ($CD3^+CD4^+CD8^-$ or $CD3^+CD4^-CD8^+$)⁴. By contrast, the development of B cells is essentially coordinated by the presence of interleukin-7 (IL-7)⁷ and the cell-to-cell interactions between stromal cells from the bone marrow and the B cell precursors⁸. B cell development begins with the rearrangement of the immunoglobulin heavy chain to obtain the late pro-B cell status⁹, which is then followed by the rearrangement of the immunoglobulin light chain reaching the pre-B cell stage¹⁰. Finally, immature B cells are obtained when there is a surface expression of fully-rearranged immunoglobulins – IgD and IgM¹¹. After a double selection process, similar to the one described in T cell development¹⁰, mature B cells are released into the bloodstream². NK cell development requires a specific microenvironment in which NK cell progenitors ($CD34^+Lin^-$) are exposed to growth factors such as flt3 (FL) and c-kit (KL) ligands to differentiate into NK cell precursors ($CD34^+IL-15R^+$). Then, these precursors can migrate to the secondary lymphoid organs and develop into $CD56^{bright}CD16^{dim/neg}$ by the interaction with IL-15 and secreted IL-2 from activated T cells¹³. The mechanisms of production of $CD56^{dim}CD16^{bright}$ NK cells are still unknown, but research suggests that this population could be either derived from $CD56^{bright}CD16^{dim/neg}$ NK cells in the lymph nodes¹⁴, have a unique precursor not yet identified or be under the control of additional cytokines during development such as IL-21¹³.

Lymphocytes need three signals to become activated and elaborate a complete immune response. The first signal is based on the recognition of the antigenic epitopes which is followed by a co-stimulation and ended by a cytokine control in order to promote lymphocytes' clonal expansion

and final activation². NK cells have a special activation process called immunological synapsis based on the dynamic equilibrium theory of their cell surface receptors. This theory proposes that when the target cell is infected, it modifies its cell surface molecules increasing the activation signals to NK cells triggering the destruction of the target cell; otherwise NK cells receive more inhibitory stimuli suppressing the destruction of the target cell¹⁵.

After their activation, T cells and B cells participate in the adaptive immune response whilst NK cells have an essential role during the innate immune response. T lymphocytes can be divided into two subpopulations including helper T lymphocytes and cytotoxic T lymphocytes². The first secrete different cytokines that participate in multiple immunological processes such as B cell antibody class switching and immunoglobulin secretion¹⁶, growth and activation of cytotoxic T lymphocytes and the increase of macrophage phagocytosis². Meanwhile, cytotoxic T lymphocytes and NK cells kill infected cells by secreting lytic molecules such as perforins and granzymes¹⁷. Moreover, B cells participate in the humoral immune response based on the expression, production and secretion of immunoglobulins against specific epitopes¹⁸. On their own, immunoglobulins can activate multiple processes such as phagocytosis, opsonization, complement pathway response or enzymatic activity neutralization².

Myeloid development process

The common myeloid progenitor differentiates into two intermediates called granulocyte/macrophage progenitor and megakaryocyte/erythrocyte progenitor which are able to generate the rest of the blood cells².

Looking at the megakaryocyte/erythrocyte progenitor, this intermediate can be directed to an erythropoiesis process generating erythrocytes¹⁹;

or to a megakaryopoiesis/thrombopoiesis process generating megakaryocytes that ultimately form platelets²⁰. These processes are represented schematically in Figure 1.

In mammals, erythrocytes (commonly known as red blood cells) commonly are anucleate cells with a biconcave disc shape, 5-7.5µm in diameter (Figure 2A). They are the most abundant cells in the blood representing 90% of the total blood cell fraction and their function consists primarily of transporting oxygen to the cells in the tissues and eliminating the residual carbon dioxide (CO₂). On average, healthy individuals have about 5×10^9 erythrocytes per litre of blood¹, which are destroyed after an average of 100-120 days in a process called haemolysis which is based on a controlled macrophage phagocytosis in the spleen and liver¹⁹. Erythropoiesis is mainly controlled by two factors, the genetic activation and expression of erythropoietin²¹ and hepcidin¹⁹ and environmental conditions which include hypoxic environment²¹ and iron uptake¹⁹. These factors initiate the differentiation pathway from the megakaryocyte/erythrocyte progenitors to proerythroblasts. After two intermediate stages, which are characterised by the accumulation of haemoglobin, the proerythroblasts form the normoblasts which lose their nuclei by extrusion forming reticulocytes²² that after some maturation days in the bone marrow are finally liberated to the bloodstream as mature erythrocytes²³.

Platelets (also called thrombocytes) are small, irregularly-shaped cytoplasmic fragments, 2-3µm in size and without nuclei (Figure 2B). Healthy adults can produce 10^{11} platelets daily and they can persist 8-11 days in the bloodstream²⁰. Platelets have three principal functions including haemostasis maintenance²⁴, production and secretion of growth factors²⁰ and secretion of clotting factors and thrombus formation reducing uncontrolled bleeding²⁵. Similar to other blood cells, platelets are formed in the bone marrow in which the megakaryocyte/erythrocyte progenitors are directed into a megakaryopoiesis process in which they differentiate into mature megakaryocytes. This first differentiation is

regulated by the presence of thrombopoietin and IL-3²⁰ and the activation of GATA-2, NF-E2 and stem cell leukaemia (SCL) transcription factors²⁶. Later, platelets are formed by fragmentation of the megakaryocytes' pseudopodia in a pathway called thrombopoiesis which starts with a cytoskeleton and cytoplasmic organelle rearrangement forming special pseudopodial projections called protoplatelets²⁷. When the organelles are blocked in these new regions, the megakaryocytes activate an apoptotic pathway²⁸ that produces a second cytoskeleton rearrangement which ends up in multiple fragmentation of the protoplatelets forming the mature platelets²².

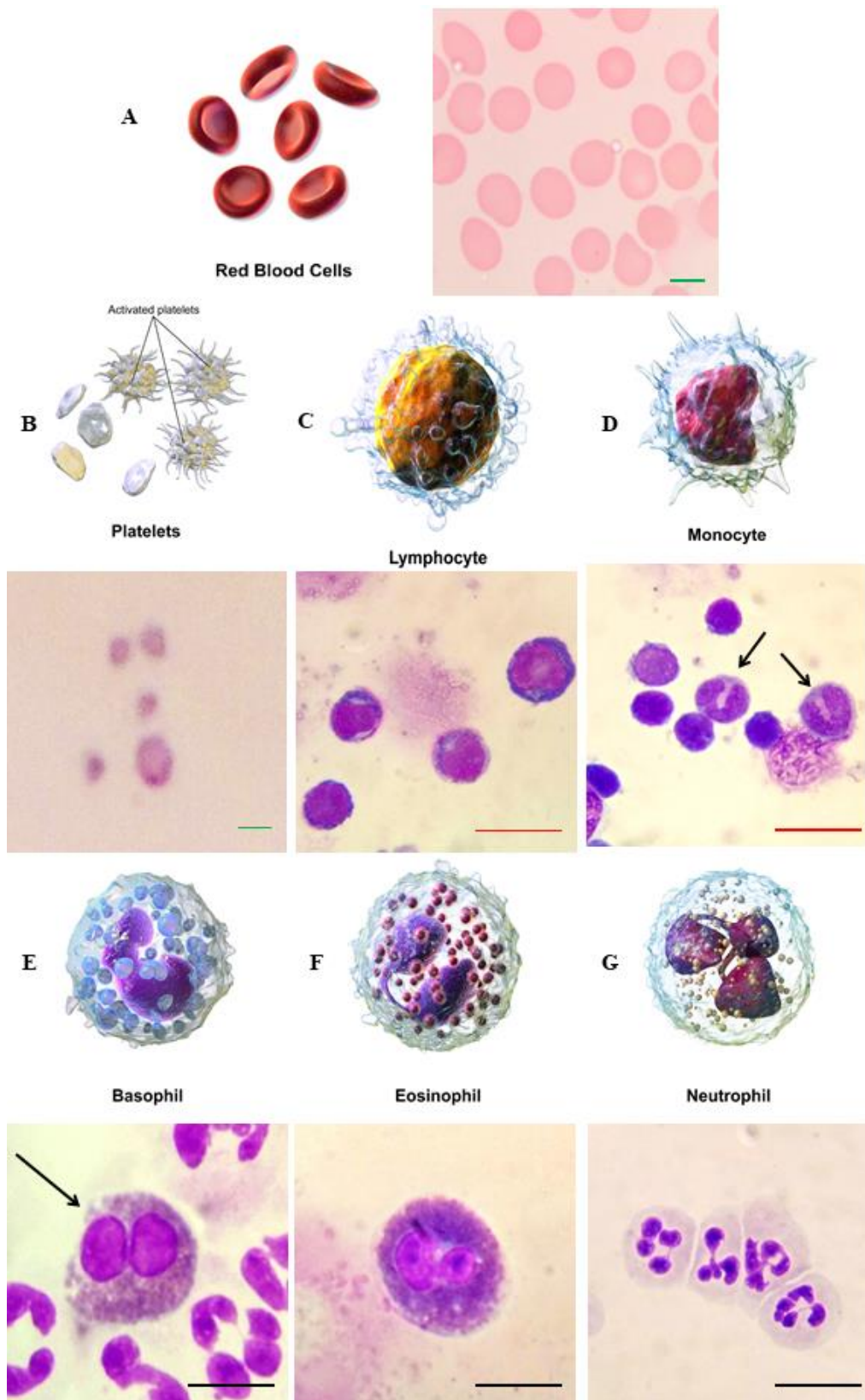
In the other branch of development, the granulocyte/macrophage progenitors can be directed to differentiate into granulocytes or monocytes². These processes are schematically represented in Figure 1. Granulocytes can be divided into three subpopulations, namely basophils, eosinophils and neutrophils (Figure 2E-F).

Basophils are the least common cells in the blood, represented by only 0.5-1% of the cells. They have a spherical shape, 12-17 μ m in diameter, with a bilobed nucleus and a big cytoplasm filled with azurophilic-basophilic granules¹. Basophils are formed in the bone marrow in which the granulocyte/macrophage progenitor increases the expression of *GATA-2* and decreases the expression of CCAAT-enhancer-binding protein alpha (*C/EBP α*) maturing into a basophil/mast cell intermediate. Then, this intermediate re-activates the expression of *C/EBP α* developing into a mature basophil²⁹. When basophils are activated, they expel their granules, mainly secreting interleukins, heparin and histamine which activate other immune cells and coordinate different processes such as phagocytosis and inflammation. The unregulated secretion of basophil granules has been implicated in hypersensitivity, anaphylactic shock, atopic dermatitis and allergic rhinitis².

Eosinophils represent 1-3% of the total white blood cells. They have a spherical shape with a diameter of 12-17 μ m, a bilobed nucleus and a big cytoplasm filled with crystalloid-acidophilic granules. Eosinophils can persist in the circulating blood for 8-12 hours before migrating into the tissue where they can survive for an additional 8-12 days without stimulation¹. During the developing process, the granulocyte/macrophage progenitor increases the expression of *GATA-2* and *C/EBP α* maturing into an eosinophil intermediate, which will finally increase the expression of granulocyte colony-stimulating factor (G-CSF), IL-5 and IL-5R developing into a mature eosinophil³⁰. In the tissues, activated eosinophils participate in the immune response against different infections and multicellular parasites by liberating their granules and other cellular components in a process called cytolysis². The unregulated secretion of eosinophil granules has been related to asthma and allergies³⁰.

Neutrophils are the most abundant cells in the white blood cell fraction representing between 40-70% of the total cell count. They have a spherical shape with an 8-12 μ m diameter, presenting a multilobed nucleus and a big cytoplasm filled with granules – important to this project are the primary or azurophilic granules in which alpha-defensin peptides are stored. The granulocyte/macrophage progenitor initiates the differentiation process forming an intermediate called the promyelocyte¹ which develops to mature inactive neutrophils in the presence of the G-CSF³¹. Finally, these neutrophils are liberated into the bloodstream persisting between 6-8 hours before migrating to the tissue, where they can survive between 1-2 days without stimulation³². Neutrophil responses can be mainly classified into three processes including degranulation and antimicrobial activity which consist in liberating their granules maintaining an inflammatory environment and recruiting new immune cells, opsonization and reactive oxygen species mediated pathogen elimination² and the formation of neutrophil extracellular traps which are created by a net of DNA and proteases that snare and eliminate pathogens³³.

Finally, monocytes are a population of mononuclear cells which represent 2-10% of the total white blood cells (Figure 2D). They can present a spherical or amoeboid shape with an ellipsoidal, kidney-shape nucleus¹. They can be classified into three groups depending on the expression levels of their surface receptors – classical (CD14⁺⁺CD16⁻), non-classical (CD14⁺CD16⁺⁺) and intermediate (CD14⁺⁺CD16⁺)³⁴. During the developing process, the granulocyte/macrophage progenitors differentiate into intermediates called monoblasts, which finally develop to mature monocytes in the presence of cytokines such as monocyte colony stimulating factor (M-CSF) and IL-34 and activating different transcription factors – PU.1, IRF8 and KLF-4³⁵. Mature monocytes can elaborate three main immune responses including macrophage differentiation; direct phagocytosis and opsonization² and differentiation into inflammatory dendritic cells³⁶.



Online blood datasets and new haematological studies

Documentation of haematological studies and blood cell diversity was previously confined to publications in a wide variety of journals. As information grew, it became necessary to compile and organise this information. In the last ten years, the growth of online resources and bioinformatic tools made it apparent that online databases are the most convenient place to gather, store and present this information. The main online blood datasets and the most recent technologies used in haematological studies are reviewed in this section.

Genotype-Tissue Expression (GTEx)

The GTEx project is a tissue-specific gene expression online database using samples in genomic and transcriptomic analyses³⁸. Whole-genome sequencing is used as the standard technique for genomic tests describing sample-specific single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) and comparing the sequences with the hg19/GRCh37 reference genome from the UCSC Genome Browser³⁹. Transcriptomic analyses are performed by RNA sequencing (RNA-seq), obtained reads are compared against the reference genome and exons are analysed and optimised by computational algorithms. Genomic and transcriptomic data from the same individual is available describing the sample-specific genetic landscape that could influence gene expression patterns. Some disadvantages include that whole blood was used in the GTEx study in which gene expression patterns cannot be accurately distributed between blood cell populations. Furthermore, the software assigns the reads to a unique transcript losing variants which could be related with undetected protein isoforms and the selective thresholds were set up excluding minor cell populations and low expressed mRNAs⁴⁰.

Human Protein Atlas (HPA)

The HPA includes 44 human tissues and the data is collected by RNA-seq and/or microarrays for analysing gene expression motifs and immunohistochemistry for protein quantification. This dataset combines transcriptomic and proteomic analyses from the same tissue or cell line which allows the correlation between the RNA expression levels and the concentration of its related protein. The main disadvantage is that whole blood is not monitored in this dataset only focusing on immune-derived cells from the bone marrow. Moreover, as described in the GTEx study, selective thresholds exclude minor populations and low expressed mRNAs⁴¹.

Universal Protein Resource (UniProt)

UniProt is a protein dataset divided in two levels – annotation data and validation. During the first stage, the online resource accumulates and administrates public available protein studies that will be pushed forward into the second stage in which the Swiss Institute of Bioinformatics (SIB) validates these published results by *in vitro* analyses developing SwissProt – sub-dataset in UniProt which includes information of the target protein such as post-translation modifications, domains, cellular location, topology and biological functions. The main disadvantage is the lack of gene expression analyses combined with this highly exhaustive protein dataset which can be used to relate sample-specific transcriptome profiles with protein expression patterns⁴².

BLUEPRINT-Epigenome

BLUEPRINT-Epigenome dataset includes epigenetic and gene expression studies from more than 100 primary cells from healthy and neoplastic counterparts as well as 8 primary cells from patients with

chronic autoimmune diseases such as type I diabetes. It is one of the most used datasets for haematological studies as blood samples are collected from three different batches including capillary, cord and venous blood and the different populations – common, rare and precursor cell types – are separated by fluorescence activated cell sorting (FACS) allowing a more precise distribution of the gene expression levels between blood cell populations. Furthermore, RNA-seq results are normalised taking in consideration the number of mapped reads and the length of the transcript avoiding possible gene expression misinterpretations. Also, the computational algorithm has a threshold which includes minor cell populations and low expressed mRNAs⁴³.

The development of sequencing techniques and analytical packages have increased the number of detailed molecular haematological studies in the last five-year period. Single-cell studies are the new analytical techniques to describe cellular characteristics such as genome, transcriptome and proteome at a single cell level in a heterogeneous population that could be hidden when studying a bulk sample. Single cells are obtained by a FACS procedure in which cell-specific markers are stained and used to separate single-cell droplets into individual wells in a plate⁴⁴. Whole-genome sequencing is generally performed after multiple displacement amplification (MDA). MDA approach uses random primers and high-fidelity DNA polymerases (DNAPol) such as Φ 29 amplifying long DNA fragments that are sequenced and mapped against the hg19/GRCh37 reference genome to describe single-cell genetic signatures – SNPs and CNVs⁴⁵. Transcriptomic analyses are mainly done by a combination of single-cell RNA sequencing and qRT-PCR in which RNA is extracted, sequenced, mapped and quantified to obtain gene expression profiles which allow to describe single-cell typing and dynamics in heterogeneous populations⁴⁴. Finally, proteomic analyses are conducted by mass cytometry (CyTOF) in which proteins are stained with antibodies conjugated with rare metal isotopes, ionised through an argon plasma and nebulised and analysed by mass spectrometry – each

protein presents a characteristic time-of-flight creating a cell-specific protein expression pattern⁴⁶. These single-cell approaches are providing a high-resolution tool to analyse blood samples taking into consideration the heterogeneity of each populations and they have already been used to analyse monocytes⁴⁷, CD4⁺ T cells⁴⁸, CD8⁺ T cells⁴⁹ and NK cells⁵⁰.

Single-cell approaches and haematological databases analyse and store, respectively, the information from the whole genome that has been sequenced using next generation sequencing techniques, which are based on short reads that must be mapped back against a reference sequence. Approximately, 50% of the human genome is constituted by repetitive regions, which from a computational perspective create ambiguities in alignment and assembly because they can collapse on top of one another and can cause complex, misassembled rearrangements producing errors when interpreting results⁵¹. Structural genomic variations and specifically copy number variations are one of the most important consequences of repetitive regions as will be detailed in the next section. The majority of the algorithms deal with these ambiguously reads using three strategies. The first one is to ignore them (score = 0), meaning that all reads from repetitive units are discarded which limits the analysis to unique regions of the genome and including some false deletions⁵². The second option uses the best matched reads to report the alignment, if there are multiple alignments with the same score, the algorithm chooses one at random or reports all of them. Finally, the third approach is reporting all alignments⁵¹.

DEFA1A3 locus structure

Genomic variation is a term used to define differences in the DNA between individuals. Structural variations range in size and scale, from large – larger than 3Mb – which usually refers to abnormalities in chromosome size or number and can be monitored microscopically, to small – shorter than 3Mb – which commonly are single nucleotide polymorphisms and copy number variations and require more specialised analytical techniques⁵³. Relevant to this project we will focus on copy number variations.

Copy number variations are deletions, insertions or duplications of more than 1000bp of DNA leading to a variable number of copies of a gene (or other DNA sequence) when it is compared to a reference sequence⁵³ and can be catalogued into two classes – biallelic or multiallelic. Biallelic copy number variations are formed by a single structural variation event resulting in the repetition or absence of a genomic region, whereas multiallelic copy number variations arise when multiple deletions or duplications occur at the same locus resulting in a range of possible copy number outcomes⁵⁴.

Predominantly, copy number variations are formed by non-allelic homologous recombination (NAHR)⁵⁵. NAHR happens between two DNA segments with high sequence similarity located at the same chromosome (a few interchromosomal NAHR can also occur)⁵⁶. The process has been summarised in Figure 3. Briefly, the DNA-repair mechanism of the NAHR pathway is initiated by a double stranded DNA break which produces unfolding in the chromatin next to the broken area (Figure 3A)⁵⁷. Then, Exo1/DNA2 exonucleases (5'→3' activity) cut the blunt ends forming single stranded DNA intermediates with a 3'-OH sticky end in a process called resection (Figure 3B). One NAHR pathway arises when there is a misalignment between the chromosomes re-localising their structure to the closest homology region (second row in Figure 3B). Then, the 3'-OH

sticky intermediates are recognised by Rad51 epistasis protein complex which helps searching for homologous sequences commonly known as donors (minimum length of 250nt), invading the complementary strand forming a D-loop (Figure 3C)⁵⁶. Inside the D-loop, the 3'-OH end of the invading strand acts as a primer for a DNA polymerase which starts the DNA synthesis of the template duplex DNA, forming a Holliday junction (Figure 3D). Holliday junctions can be mainly resolved by two pathways – double-strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA)⁵⁷. Copy number variation events use the DSBR pathway in which a double Holliday junction is symmetrically cut by GEN1 in the intersection of the four strands at 1nt to the 3'-side of the junction point (Figure 3E)⁵⁸. GEN1 dimers cut one non-crossing and the opposite crossing strands forming a cross-over that can culminate in deletions or duplications⁵⁸ which may produce copy number variations (Figure 3F)^{57,58}.

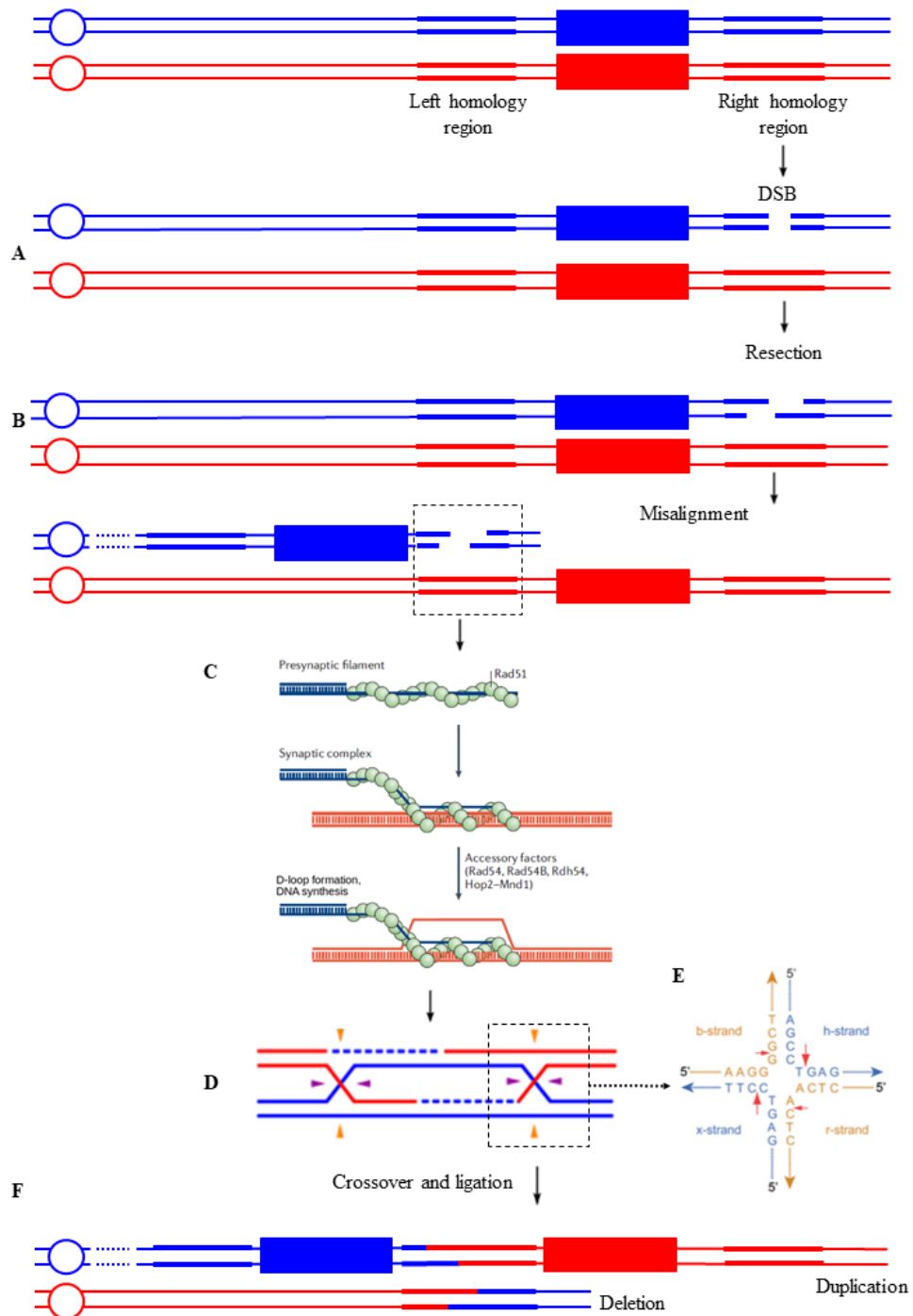


Figure 3 | Non-allelic homologous recombination pathway (NAHR). **A)** A double stranded DNA break activates the DNA-repair machinery that initiates the NAHR cascade. After the break, the chromatin located close to the broken area is unfolded which allows the recruitment of exonucleases. **B)** During the resection process, the exonucleases cut the blunt ends forming single stranded DNA intermediates with a 3'-OH sticky ends. **C)** These intermediates are recognised by Rad51 epistasis protein complex which interacts with the complementary-uncut DNA helix and helps searching for homologous sequences forming a D-loop. **D)** The 3'-OH end of the invading strand acts as a primer for a DNA polymerase which starts the DNA synthesis of the template duplex DNA, ultimately leading to a Holliday junction. **E)** This Holliday junction is resolved by GEN1 using a double-strand break repair mechanism which cuts at 1nt to the 3'-side of the junction point (red arrows). **F)** GEN1 cuts one non-crossing and the opposite crossing strands forming a cross-over that could generate structural genomic variations. **DSB:** Double stranded DNA break. Figure adapted from Sung et al.⁵⁷ and Lee et al.⁵⁹ publications.

The *DEFA1A3* locus is an excellent example in which multiallelic copy number variation events can occur. It is located at chromosome 8p23.1, close to an unstable telomeric region that exhibits accelerated rates of duplications and rearrangements that can contribute to the formation of copy number variations⁶⁰.

The *DEFA1A3* locus has a complex structure in which each haplotype is organised in three main segments – the telomeric and centromeric edges and the central area (Figure 4A). The central area is composed of a variable number (including zero) of full 19kb tandemly repeated units as represented in the two haplotypes in Figure 4B. Each full repeat contains one copy of the *DEFA1A3* gene and one copy of the *DEFT1P/DEFT10P* theta-defensin pseudogenes (inactive in humans, gorillas and chimpanzees)⁶¹. In contrast, the telomeric and centromeric edges are formed by one partial repeat (~10kb) which the centromeric region always contains one copy of the *DEFA1A3* gene as represented in Figure 4B^{61,62}.

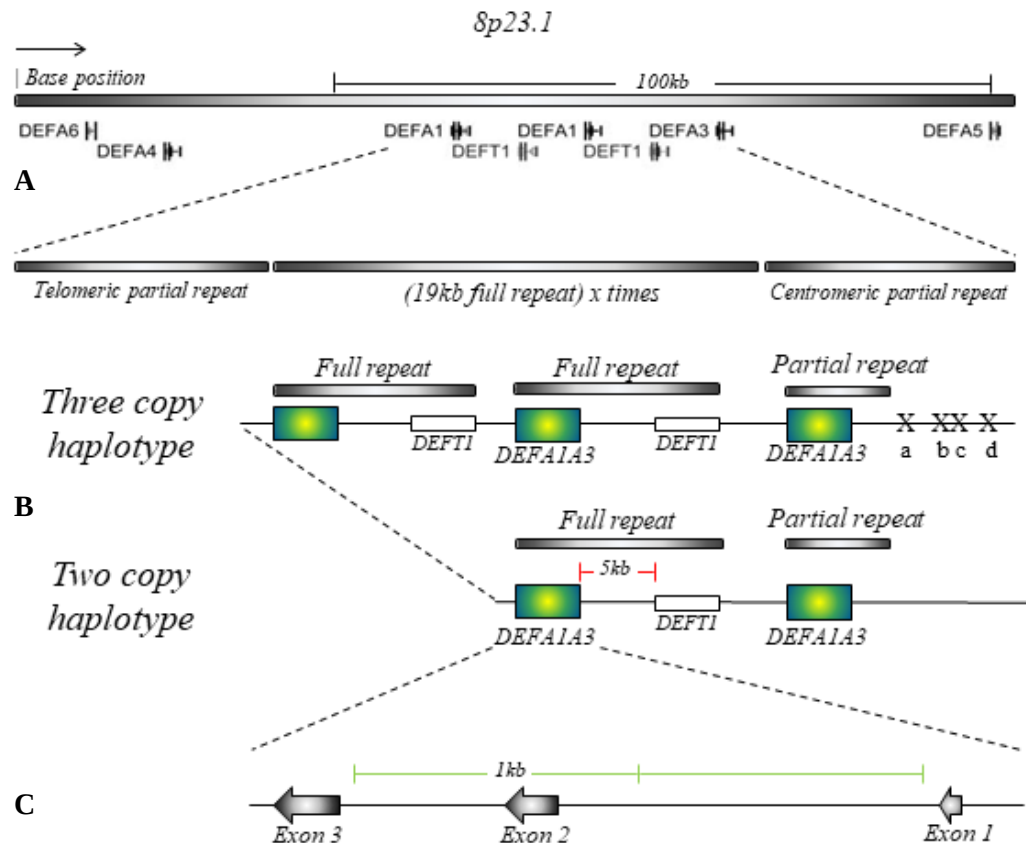


Figure 4 | Example of the *DEFA1A3* locus structure from a five-copy genotype sample. A) The top row represents the *DEFA1A3* location at chromosome 8p23.1 and showing that its transcripts are made from the bottom strand (i.e., from right to left). At the bottom, there is a schematic representation of the full locus structure divided in the three main segments – the telomeric and centromeric edges and the central area. **B)** Distribution of the five copies of *DEFA1A3* presented in a three-copy and a two-copy haplotype. a-d represent the four flanking SNPs that describe the 5 haplotype classes – Reference, Class 1, Class 2, Exchange 1 and Exchange 2. **C)** Representation of the three exons that constitute the transcription sequence of the *DEFA1A3* gene. Black scale bar = 100,000nt, red scale bar = 5,000nt and green scale bar = 1,000nt

The *DEFA1A3* notation was presented in 2005 by Aldred and colleagues after identifying that each full 19kb repeat could contain either *DEFA1* or *DEFA3* genes, which are interchangeable, and only differ in one nucleotide (C3400A) which generates an amino acid change in the first amino acid (N-terminal) of the alpha-defensin peptide sequence, also known as human neutrophil peptides (HNPs). Their phylogenetic analysis also demonstrated that *DEFA3* is human specific and could have arisen from an ancestral *DEFA1*⁶³.

The complexity of the locus is not only illustrated by the interchangeable presence of *DEFA1* or *DEFA3*, but multiple structural-genomic studies have also demonstrated that the number of copies can vary between different populations as well as from individual to individual inside a defined group⁶⁰. In 2013, Khan et al. described that Europeans can have four to eleven copies of *DEFA1A3* per diploid genome⁶² – *DEFA1* was present in a range from three to nine copies with an average of five copies, whereas *DEFA3* presented an interval from zero to four copies with a modal value of one or two copies⁶³ – whilst Cheng and colleagues described that three to eleven copies and four to sixteen copies could be found in Chinese and Japanese populations respectively⁶⁴. Interestingly, Ballana et al. described that 10% of Europeans and Asians completely lacked *DEFA3* whilst this absence increased to 37% in African populations⁶⁵. Surprisingly, even though *DEFA1* and *DEFA3* only differ in one nucleotide, *DEFA3* gene expression levels appear to be 2-fold higher per copy than *DEFA1* which could be attributed to a compensation for its low activity or to their function acting as precursors for the more active HNP-2 peptide⁶³. Interestingly, both genes – *DEFA1* and *DEFA3* – have the same 200bp promoter region, sufficient for their expression⁶⁶. Further investigation must be done to describe the aftereffects of the gene expression differences as well as possible immune consequences in the individuals lacking *DEFA3*.

Additionally, Black et al. demonstrated that different haplotypes in Europeans could be categorised into five main classes based on modes of inheritance using 4 SNPs (Table 1), located in a 4.1kb flanking region from the centromeric area as represented in Figure 4B (marked with a-d). The sequences grouped in the same haplotype class shared identical or highly similar flanking nucleotides making them distinct to other haplotype classes. The five mentioned groups are: Reference, Class 1, Class 2, Exchange 1 and Exchange 2. Reference groups the sequences that are similar to the GRCh37/hg19 human reference assembly sequence, Class 1 and Class 2 both present unique sequence differences compared to the original reference assembly, and Exchange

1 and Exchange 2 present replacement polymorphisms in which a fragment of the partial repeat from the centromeric region is replaced with the equivalent sequence from the full repeat with lengths of 140bp and 1990bp in Exchange 1 and Exchange 2 respectively⁶⁷.

Table 1. Combination of single nucleotide polymorphisms for each characterised haplotype class

	rs4300027	rs7826487	rs7825750	rs62487514
Haplotype class	(chr8:7,023,053, T>C)	(chr8:7,023,403, A>G)	(chr8:7,022,281, T>C)	(chr8:7,022,572, C>A)
Reference	T	A	T	C
Class 1	C	G	T	C
Class 2	C	A	C	C
Exchange 1	T	A	T	A
Exchange 2	C	A	T	C
Restriction enzyme	<i>HinfI</i> (G [^] ANTC)	<i>SspI</i> (AAT [^] ATT)	<i>Fnu4HI</i> (GC [^] NGC)	<i>Tsp509I</i> ([^] AATT)

The first row shows the name of each single nucleotide polymorphism, chromosome location and nucleotide change. Rows 2-6 present the combination of single nucleotide polymorphisms that characterised each haplotype class. The bottom row shows the restriction enzyme used in each test to find the nucleotide change and the sequence required for the cut.

Alpha-defensin peptides and their role in the immune response

This *DEFA1A3* locus encodes the alpha-defensin 1-3 peptides. Defensins are small cationic, amphipathic cysteine-rich peptides with a structure stabilized by 3 conserved disulphide bonds. Based on the location of the cysteine residues and the disulphide bonds these peptides can be classified in alpha-, beta- and theta-defensins. The alpha-defensin group can be further subdivided into myeloid (HNP1-4 peptides) which are mainly expressed in neutrophils, and enteric human defensin 5-6 which are presented in the Paneth cells from the gut crypts⁶⁸.

Alpha-defensin 1-3 peptides are synthesized as inactive precursors (pre-proHNPs) and require the proteolytic removal of their inhibitory N-terminal pro-peptide for activation. The initial product is a 94-amino acid precursor that contains a 19-peptide endoplasmic reticulum signal, a 45-amino acid anionic pro-piece, and a 30-amino acid cationic mature peptide⁶⁸. After reaching the endoplasmic reticulum and cleaving the 19-

peptide signal, the 45-residue anionic pro-peptide can act as an intramolecular chaperone interacting with the cationic C-terminal domain helping during the folding of the human neutrophil peptide. After cleaving the 45-residue anionic pro-peptide, mature HNP1-3 peptides have 29 or 30 amino acids (3-5kDa, represented in Figure 5) and they are stored in the azurophilic granules mainly located in neutrophils⁶⁹ but also detected in other blood cell populations as will be described in this project. It is believed that HNP-1 or HNP-3 peptides (or both) are the precursors of HNP-2 peptide (29-amino acid peptide which lacks the first residue of the sequence) due to the absence of a specific *DEFA2* gene⁶⁸.

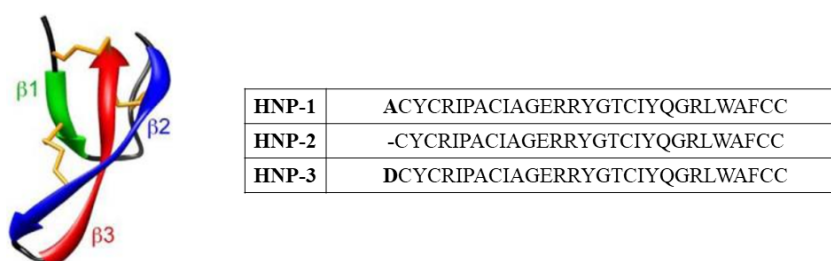


Figure 5 | Alpha-defensin 1-3 peptides. The left image corresponds to the crystal structure of HNP1-3 peptides including the disulphide bonds between cysteine residues, which characterise the structure of alpha-defensin peptides (yellow cords). The right table shows the amino acids that constitute each peptide which are identical for almost the entire sequence except in the first residue, presenting alanine in HNP-1 peptide, absent in HNP-2 peptide and aspartic acid in HNP-3 peptide (bold letters).

Different studies have investigated the correlation between the number of copies of the *DEFA1* and *DEFA3* genes and the amount of alpha-defensin 1-3 peptides. Linzmeier and Ganz (2005) used eight volunteers describing a correlation between the *DEFA1A3* copy number and the levels of HNP1-3 peptides extracted from purified neutrophils and quantified by enzyme-linked immunosorbent assay (ELISA). Their results suggested that, the amount of HNP1-3 peptides is proportional to the number of copies of the *DEFA1* and *DEFA3* genes⁶⁰. During the same year, Aldred and colleagues could not find any correlation between the number of copies of the *DEFA1A3* gene and the amount of *DEFA1A3* mRNA in whole blood ($r^2=0.02$)⁶³. A more recent study conducted by Ai et al. used a cohort of immunoglobulin-A nephropathy (IgAN) patients

and controls reconfirming Aldred and colleagues' results, describing that, although the serum and urine levels of HNP1-3 peptides were higher in IgAN cases, they were not able to find an association between the number of copies of the *DEFA1A3* gene, *DEFA1A3* mRNA expression levels and the extracellular levels of HNP1-3 peptides⁷⁰.

Alpha-defensin 1-3 peptides have a direct role in the innate immune response against bacteria, fungi and enveloped viruses. These peptides can be extracellularly secreted, or the azurophilic granules (organelles where alpha-defensin peptides are stored) can fuse with the phagosome during phagocytosis. Alpha-defensin peptides have a protective activity based on the disruption of the pathogen membranes, in which the positively-charged amino acids are internalised in the membrane by the interaction with the anionic phospholipids. Inside the membrane, the peptides polymerise forming a hexamer creating negative-potential channels that permeabilise the pathogen and deregulate homeostasis ending in the pathogen's death⁶⁸. Moreover, serum HNP1-3 peptides can directly block the interaction between enveloped viruses and target cells acting as competitive inhibitors against the viral surface receptors. They can also act in virus-infected cells blocking downstream pathways such as nuclear import, protein kinase C (PKC) cascade or modifying surface receptor levels⁷¹.

Additionally, alpha-defensin peptides can inactivate the complement pathway response, act as chemoattraction factors and regulate the production of inflammatory cytokines. In the complement pathway, serum HNP1-3 peptides bind to the collagenous tail of C1q and mannose-binding lectin blocking the complement pathway as C1q cannot then interact with C4⁷². Moreover, HNP1-3 peptides act as chemoattraction factors for monocytes, macrophages⁷³, monocyte derived dendritic cells, dendritic cell precursors (CD34⁺) and CD3⁺ T cells – both CD4⁺ and CD8⁺. Interestingly, only naïve CD4⁺CD45RA⁺ T cells can migrate using the HNP1-3 peptides' chemoattraction gradient whereas memory CD4⁺CD45RO⁺ T cells cannot⁷⁴. Furthermore, HNP1-3 peptides increase

the production and secretion of prostaglandin-D2 and histamine in mast cells⁷⁵, or tumour necrosis factor alpha (TNF α) and IL-1 β in monocytes, compared with their ability to block the protein synthesis of IL-6 and IL-8 in white blood cells by sequestering ribosomal proteins and complements inhibiting mRNA translation⁷⁶.

Alpha-defensin 1-3 peptides in disease

Defensins are implicated as especially important in some inflammatory and autoimmune disorders, affecting their predisposition, progress and outcomes.

Sthoeger et al. (2008) used a cohort of 30 healthy and 52 systemic lupus erythematosus (SLE) volunteers to observe a 10-fold increase in the concentration of alpha-defensin 1-3 peptides in the sera from SLE patients. This increase could not be correlated with an increase in neutrophil cell count or renal failure, but they were able to describe in 48% of patients with SLE an over-degranulation of neutrophils. Moreover, they described relatively – but not statistically significant – higher levels of *DEFA1A3* mRNA in SLE volunteers than in healthy controls when monitoring whole blood⁷⁷. These data supported Tamiya and colleagues' work in which, additionally, higher levels of anti-defensin antibodies were found in patients with SLE and returned to normal levels on treatment with high doses of corticosteroids⁷⁸. Additionally, Cheng et al. performed a genetic association study using an Asian cohort of 163 patients with SLE and 137 age and sex matched healthy controls in which they were not able to show differences in the number of copies of the *DEFA1A3* gene between patients and controls (mean of 7 copies), nor any correlation between the number of copies of the *DEFA1A3* gene and the amount of *DEFA1A3* mRNA from whole blood or other SLE manifestations such as photosensitivity, oral ulcers or anti-defensin antibodies⁶⁴.

In 2012, Yu et al. published the first Genome Wide Association Study (GWAS) using a Chinese cohort of 4137 immunoglobulin-A nephropathy (IgAN) cases and 7734 healthy controls to describe the role of *DEFA1A3* locus in IgAN. Their data presented the first association of SNP rs2738048 (chr8:6965263, A>G), located in the region flanking the copy number variable *DEFA1A3* locus, with IgAN ($p < 5 \times 10^{-8}$)⁷⁹. Li et al.'s study (2015) provided the first evidence of three new independent associations between SNPs at the *DEFA1A3* locus and IgAN – rs2738058 (chr8:6964095, T>C, $p = 1.15 \times 10^{-19}$), rs12716640 (chr8:695376, A>C, $p = 9.53 \times 10^{-9}$) and rs9314614 (chr8:6840209, C>G, $p = 4.25 \times 10^{-9}$). Their linkage disequilibrium analysis also correlated rs2738058 and the previously described rs2738048 confirming Yu et al. results⁸⁰. In the same year, Qi and colleagues used a two-stage association analysis to describe the correlation of SNPs within copy number variations in the *DEFA1A3* locus and their role in IgAN. Their data described a strong relationship between rs2738058 with a lower number of copies ($p = 0.012$) whereas rs9644778 (chr8:7052132, A>C) presented a weaker correlation ($p = 0.124$) and was only observed in Chinese cohorts⁸¹. Also, Qi and colleagues' study (2016) described that alpha-defensin 1-3 peptide sera levels were higher in patients with IgAN when compared with healthy controls ($p = 0.013$). Their results could not find any association between HNP1-3 peptide sera levels and neutrophil count, IgA/IgA1 levels or reduction in glomerular filtration. In contrast, their data showed a strong correlation between patients with IgAN expressing high levels of alpha-defensin 1-3 peptides and low levels of free galactose-deficient IgA (Gd-IgA1) ($p = 0.008$). HNP1-3 peptides could interact with glycosyl residues from Gd-IgA1 which may contribute to the aggregation of IgA1 and deposition in the mesangial regions (in the glomerulus of the kidney) becoming a possible risk factor to IgAN⁸². One of the latest genetic association studies performed by Ai et al. presented three independently-associated variations within the *DEFA1A3* locus related to IgAN, including the number of copies of *DEFA1A3* ($p = 3.99 \times 10^{-9}$), *DEFA3* ($p = 6.55 \times 10^{-5}$) and the 4bp deletion polymorphism at intron 2 ($p = 3.5 \times 10^{-16}$). Further analyses confirmed that cases with lower copy numbers of

the three copy number variations had a 5.9-fold higher susceptibility to IgAN when compared with controls. Moreover, even though they found higher levels of HNP1-3 peptides in IgAN patient sera they could not find any correlation between the number of copies of the *DEFA1A3* gene and the amount of peptide⁷⁰.

Teumer and colleagues (2013) used a cohort of 3365 Germans performing a GWAS describing the genetic background that influenced chronic periodontitis and aggressive periodontitis describing the variant rs2738058 and its correlation as a possible risk allele for chronic periodontitis⁸³. Dommisch et al.'s work (2015) provided the first evidence of an increase in the concentration of HNP1-3 peptides in the gingival crevicular fluid in cases when compared with controls, describing a 15-fold increase in cases with aggressive periodontitis and a 60-fold increase in chronic periodontitis volunteers⁸⁴. Recently, Munz et al. performed the biggest GWAS for periodontitis using 896 aggressive periodontitis cases and 7104 controls from a German and Dutch cohort describing a strong association of two independent SNPs in the *DEFA1A3* locus with aggressive periodontitis. Their results reconfirmed the role of rs2738058 ($p=6.78 \times 10^{-10}$) supporting the results shown in Teumer and colleagues' study and presented a new association with the variant rs2978951 (chr8:6965773, A>G, $p=2.06 \times 10^{-8}$). Moreover, their expression quantitative trait loci (eQTL) data suggested that both *DEFA1A3* SNPs presented cis-regulatory effects on *DEFA4*⁸⁵.

Astle and colleagues' seminal work presented an intracellular phenotype investigation based on flow cytometry revealing novel haematological characteristics and their possible implications in understanding inflammatory and autoimmune diseases⁸⁶. Their results showed a strong association ($p=4.2 \times 10^{-313}$) between variation at the defensin locus and neutrophil granularity⁸⁷.

By contrast with genome wide association studies using SNP data, in which the accuracy of genotyping is not in question, some studies of

inflammatory diseases such as Crohn's disease⁸⁸ or severe sepsis⁸⁹ have used direct analysis of copy number at *DEFA1A3*. For this reason, caution needs to be exercised in evaluating the robustness of their conclusions, especially since their observations have not been independently replicated.

BIOLOGICAL QUESTIONS AND OBJECTIVES

Human neutrophil peptides have been extensively characterised in granulocytes with strong evidence supporting the presence of *DEFA1A3* mRNA and alpha-defensin 1-3 peptides in this blood cell fraction. It is still clear that granulocytes are not able to transcribe new *DEFA1A3* mRNA and accumulate alpha-defensin 1-3 peptides which is believed to be restricted to its bone marrow precursor the promyelocyte⁶⁸. Moreover, these alpha-defensin 1-3 peptides as well as structural variations in the *DEFA1A3* locus have been related to the predisposition, progress and outcomes of inflammatory and autoimmune disorders, but it is still unclear how this correlation occurs. Inconsistent reports in previous literature and online databases led us to ask if other blood cell populations can also have *DEFA1A3* mRNA and/or alpha-defensin 1-3 peptides. If *DEFA1A3* mRNA is found in other blood cell fractions, we also want to establish what are the *DEFA1A3* gene expression levels in each blood cell subset and what is their relative contribution to the whole blood *DEFA1A3* gene expression levels.

To test if other blood cell populations, apart from granulocytes, present *DEFA1A3* mRNA the main blood cell fractions will be separated by density-gradient centrifugation and their corresponding subpopulations will be purified using a fluorescence activated cell sorting (FACS) approach. RNA from purified samples will be extracted, reverse-transcribed into complementary DNA (cDNA) and amplified by PCR using specific primers for *DEFA1A3* mRNA and internal control mRNAs. Amplified cDNA bands will be evaluated in an agarose gel and compared with controls. To quantify the relative *DEFA1A3* gene expression levels in each blood cell population and their contribution to whole blood *DEFA1A3* gene expression levels, cDNAs will be amplified by qRT-PCR using specific TaqMan probes from *DEFA1A3* and internal controls. Results will be analysed by the $2^{\Delta\Delta Ct}$ method and compared with the

reference populations – granulocytes or peripheral blood mononuclear cells (PBMCs). To detect the presence of alpha-defensin 1-3 peptides in the azurophilic granules, purified populations will be used for cytospin preparation and indirect immunocytochemistry staining.

RESULTS

Characterization of *DEFA1A3* mRNA in blood samples

To answer these unresolved biological questions granulocytes and PBMCs (composed of white blood mononuclear cells and platelets) were recovered by a double Histopaque density-gradient centrifugation. Respectively, granulocytes were accumulated in the bottom layer, whereas PBMCs were accumulated in the upper section (Figure 6). Cross-contamination was monitored by counting cells in 40 randomly selected fields of which four are shown in Figure 6, with granulocytes and platelet-depleted PBMCs cytopsin preparations shown in the lower and upper panels respectively (from volunteer IRP0001). Because individual platelet volume is significantly lower than white blood cells, based on Haley and colleagues' study⁹⁰ a mean of 120 platelets was used as an equivalent (by volume) to one white blood cell to standardise the cell count results and allow further comparisons. From the whole set of volunteer samples (9 samples), in the granulocyte layer purity results $\pm$ standard deviation (SD) showed equivalent representations of $99.5\% \pm 11.28$ attributed to granulocytes – distributed between neutrophils ($97.9\% \pm 4.8$), basophils ($0.2\% \pm 0.2$) and eosinophils ($1.9\% \pm 0.7$) – $0.5\% \pm 0.2$ to PBMCs and less than 0.1% to platelets. In the PBMC fraction, the values were $98.9\% \pm 8.7$, $1\% \pm 0.5$ and less than 0.1% in white blood mononuclear cells, granulocytes and platelets respectively. After platelet depletion, white blood mononuclear cells were approximately 99% pure.

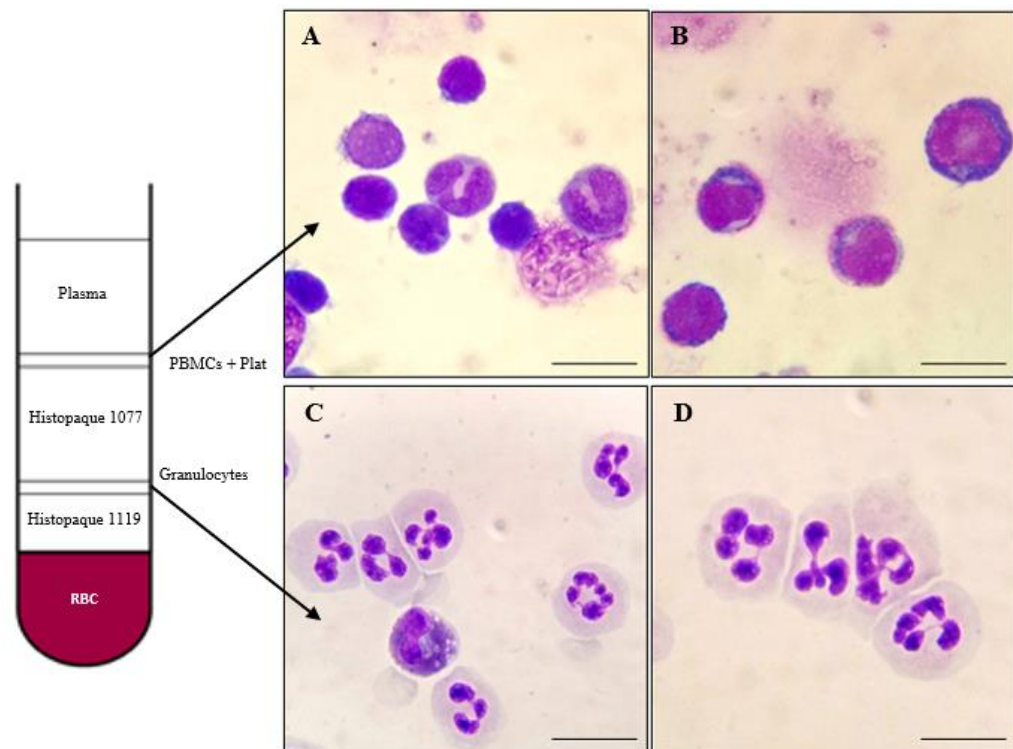


Figure 6 | Double Histopaque density-gradient centrifugation and microscope slides from platelet-depleted peripheral blood mononuclear cells and granulocytes. Peripheral blood mononuclear cells (**A and B**) and granulocytes (**C and D**) were purified by a double Histopaque density-gradient centrifugation as represented on the left. Platelets were eliminated from the mononuclear cell fraction by a double wash, and granulocytes were treated with an erythrocyte lysis buffer to eliminate the presence of possible contaminating erythrocytes. A final concentration of 70,000 cells in 500 μ L solution was used for cytopspin preparations and followed by Hemacolor-staining and slide preparation (represented on the right). Fields were analysed with a 100x oil immersion lens. **PBMC**: peripheral blood mononuclear cells, **Plat**: platelets and **RBC**: red blood cells. Scale bar = 10 μ m.

To analyse the presence of *DEFA1A3* mRNA, cells from purified layers – granulocytes and PBMCs – were recovered and RNA was extracted. RNA samples presented excellent yield, negligible contaminants and high concentration which were evaluated as described in *Materials and Methods* (data not shown). Selected RNAs were reverse-transcribed, cDNAs were amplified by PCR and outcomes were evaluated after 1.5% agarose gel electrophoresis. Figure 7 shows the amplified bands obtained after the PCR from each sample preceded by the bands of a 100bp DNA ladder. The upper row presents the amplified products from *DEFA1A3* mRNA (322bp) followed by a second row with the amplified products from *SDHA* mRNA (306bp). *SDHA* was included as a reference

gene presenting constant expression levels under multiple conditions and in different blood cell populations⁹¹.

The presence of the *DEFA1A3* band in granulocytes, as shown Figure 7, indicated the presence of *DEFA1A3* mRNA in this blood cell fraction which accorded with previous published results⁹². Moreover, recovered PBMCs had *DEFA1A3* mRNA as can be seen in the upper panel of Figure 7. This last result led us to separate the two main components of the PBMC fraction – white blood mononuclear cells and platelets. Respectively, the presence of the amplified *DEFA1A3* bands in wells three and four from Figure 7 indicate the presence of *DEFA1A3* mRNA in both subsets. In addition, the absence of amplified bands in the internal controls, as presented in wells five and six from Figure 7, indicated lack of cross-contamination during the cDNA synthesis and PCR amplification procedures.

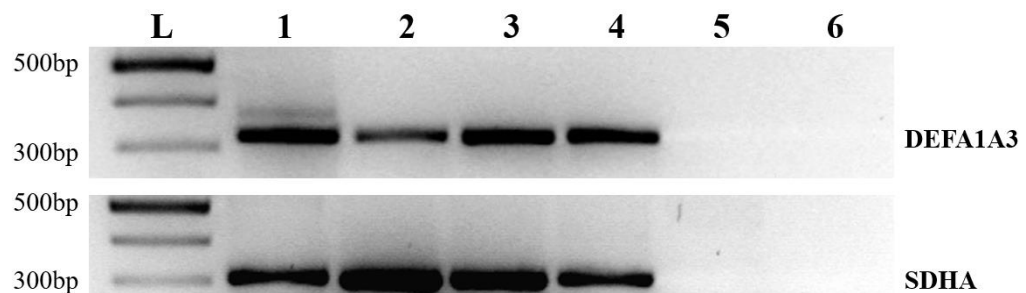


Figure 7 | PCR-amplified products from *DEFA1A3* mRNA and *SDHA* mRNA from recovered granulocytes, peripheral blood mononuclear cells and platelets. From recovered granulocytes, peripheral blood mononuclear cells and platelets, RNA was extracted and reverse-transcribed into cDNA. PCR-amplified cDNAs were evaluated after 1.5% agarose gel electrophoresis. The top panel represents the amplified products from *DEFA1A3* mRNA (322bp) whereas the bottom panel shows the amplified outcomes from *SDHA* mRNA (306bp). Samples included: L – 100bp DNA ladder (500ng), 1 – Granulocyte cDNA, 2 – PBMC cDNA, 3 – cDNA from platelet-depleted PBMCs, 4 – cDNA from Platelets, 5 – cDNA from a control omitting reverse transcriptase and 6 – cDNA from a control without any input.

DEFA1A3 mRNA levels are low in purified platelets

After identifying the presence of *DEFA1A3* mRNA in the PBMC fraction and establishing that both populations – platelets and white blood mononuclear cells – included *DEFA1A3* mRNA (Figure 7), these populations were further purified to accurately describe their *DEFA1A3* gene expression levels and their individual contribution to the whole blood *DEFA1A3* gene expression levels.

Purified platelets were recovered from initial whole blood samples consistent with prior platelet-enriched plasma (PRP) protocols⁹³. Figure 8 shows two platelet sample preparations stained by a Hemacolor protocol (from volunteer IRP0001). The most interesting morphological characteristics shown in these preparations are platelet irregular-shaped morphology (pink eosin staining) and lack of nucleus (absence of blue haematoxylin staining). Purity was analysed as described in the last section. As an example, the cell count and purity results from volunteer IRP0001 are presented in Table 2 in which a total of 7327 platelets, 2 lymphocytes and 1 monocyte were observed. Based on the previously mentioned premise (~120 platelets = 1 white blood cell)⁹⁰, 7327 counted platelets are equivalent in volume to 61.06 white blood cells making a total of 64.06 counted cell equivalents. From volunteer IRP0001, purity results $\pm$ SD indicated equivalent representations of $95.3\% \pm 0.5$, $3.1\% \pm 0.2$ and $1.6\% \pm 0.2$ for platelets, lymphocytes and monocytes respectively. Identical approaches with the whole set of volunteers' samples (9 samples) showed purity greater than 95% [95.3%, 99.5%] by volume in all cases after recovering platelets (data not shown).

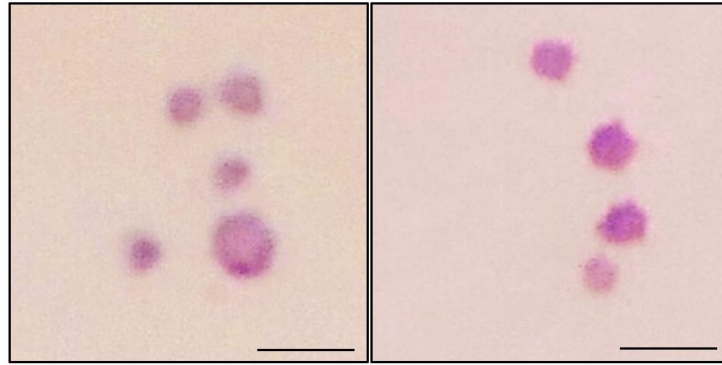


Figure 8 | Microscope slides from Hemacolor-stained platelets. Platelets were concentrated in the plasma with an initial slow spin and subsequently pelleted with a faster spin. Recovered platelets were used for Hemacolor-staining and slide preparation. Fields were analysed with a 100x oil immersion lens. Scale bar = 5 μ m.

Table 2. Cell count and purity results from IRP0001 sample

						<i>Platelets from PRP</i>					
<i>Field</i>	<i>Erythrocytes</i>	<i>Granulocytes</i>	<i>Platelets</i>	<i>Lymphocytes</i>	<i>Monocytes</i>	<i>Field</i>	<i>Erythrocytes</i>	<i>Granulocytes</i>	<i>Platelets</i>	<i>Lymphocytes</i>	<i>Monocytes</i>
1	0	0	184	0	0	21	0	0	180	0	0
2	0	0	86	0	0	22	0	0	158	0	0
3	0	0	260	0	0	23	0	0	94	0	0
4	0	0	85	0	0	24	0	0	120	0	0
5	0	0	312	0	0	25	0	0	88	1	0
6	0	0	272	0	0	26	0	0	136	0	0
7	0	0	107	0	0	27	0	0	224	0	0
8	0	0	248	0	0	28	0	0	168	0	0
9	0	0	232	0	0	29	0	0	180	0	0
10	0	0	200	0	0	30	0	0	120	0	0
11	0	0	120	0	0	31	0	0	156	0	0
12	0	0	79	0	0	32	0	0	160	0	0
13	0	0	176	0	0	33	0	0	168	0	1
14	0	0	256	0	0	34	0	0	244	0	0
15	0	0	318	0	0	35	0	0	256	1	0
16	0	0	160	0	0	36	0	0	252	0	0
17	0	0	132	0	0	37	0	0	168	0	0
18	0	0	188	0	0	38	0	0	156	0	0
19	0	0	232	0	0	39	0	0	124	0	0
20	0	0	284	0	0	40	0	0	244	0	0
	<i>Erythrocytes</i>		<i>Granulocytes</i>		<i>Platelets</i>		<i>Lymphocytes</i>		<i>Monocytes</i>		
Total	0		0		7327		2		1		
Average	0		0		183.175		0.05		0.025		
Cell count (%)	0		0		95.31676857		3.122154286		1.561077143		

The upper segment of the table presents the cell count results from 40 randomly selected fields divided into five blood populations based on their morphological characteristics – erythrocytes, granulocytes, platelets, lymphocytes and monocytes respectively. In the bottom segment, the second row shows the total-counted cell of each population followed by their average – dividing the cell count results by the number of fields. The last row shows the percentage of each population adjusted for volume, obtained by dividing the number of events by the total number of cells or equivalents and multiplied by 100.

Similar procedures to those described in the characterization of *DEFA1A3* mRNA in blood samples were used to determine the presence of *DEFA1A3* mRNA in platelets. Figure 9 shows the amplified products obtained after PCR from each platelet volunteer sample preceded by the bands of a 100bp DNA ladder. The top half of the figure presents the *DEFA1A3* products amplified from cDNA (322bp) followed by the second half with the products from *SDHA* (306bp). The qualitative data presented in Figure 9 indicates that platelets preserved some *DEFA1A3* mRNA (consistent across different volunteers, data from volunteers IRP0084 and IRP0085 are not shown in Figure 9) which was used in further qRT-PCR analyses to quantify the *DEFA1A3* gene expression levels in this blood cell fraction and their contribution to the whole blood *DEFA1A3* gene expression levels.

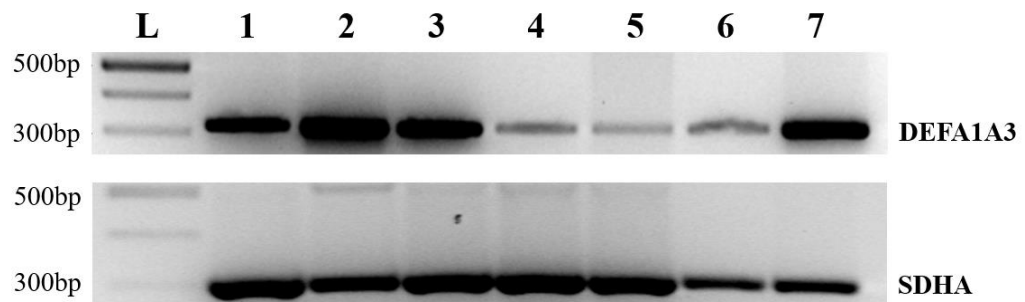


Figure 9 | PCR-amplified products from *DEFA1A3* mRNA and *SDHA* mRNA from recovered platelets. From recovered platelets, RNA was extracted and reverse-transcribed into cDNA. PCR-amplified products were evaluated in a 1.5% agarose gel electrophoresis. The top row shows the *DEFA1A3* products amplified from cDNA (322bp), whereas the bottom row presents the *SDHA* products (306bp). Samples included: L - 100bp DNA ladder (500ng), 1 - IRP0001, 2 - IRP0078, 3 - IRP0079, 4 - IRP0080, 5 - IRP0081, 6 - IRP0082 and 7 - IRP0083.

The gene expression analyses included previously described platelet samples and two controls – granulocytes as a constitutive source of *DEFA1A3* mRNA and platelet-depleted PBMCs. qRT-PCR results were analysed by the $2^{-\Delta\Delta C_t}$ method⁹⁴ and the distribution of gene expression levels between blood cell types was determined as described in *Materials and Methods*.

Table 3 summarises this analysis using the results from volunteer IRP0079 as an example, in which the average threshold cycle (Ct) values $\pm$ SD from granulocytes, platelet-depleted PBMCs and platelets were 26.9 ± 0.6 , 21.1 ± 0.3 and 19.2 ± 1.0 for *DEFA1A3* and 31.2 ± 1.3 , 24.9 ± 0.3 and 22.9 ± 0.2 for *SDHA* respectively. Gene expression levels were obtained following the $2^{\Delta\Delta Ct}$ method normalising the results using *SDHA* as the reference gene and granulocytes as the control population. In volunteer IRP0079, the *DEFA1A3* gene expression levels in whole blood were distributed as: 64.5% in granulocytes, 33.4% in platelet-depleted PBMCs and 2.1% in platelets and the associated ratios of 1.00 : 0.52 : 0.03 (granulocytes : platelet-depleted PBMCs : platelets).

Table 3. Quantitative gene expression results from IRP0079 sample.

	Granulocytes		PBMCs		Platelets	
	<i>DEFA1A3</i>	<i>SDHA</i>	<i>DEFA1A3</i>	<i>SDHA</i>	<i>DEFA1A3</i>	<i>SDHA</i>
<i>Ct values</i>	26.254	32.053	21.253	25.205	18.135	22.994
	27.487	31.905	21.285	24.841	19.710	23.076
	26.986	29.682	20.743	24.543	19.884	22.748
<i>Mean</i>	26.909	31.213	21.094	24.863	19.243	22.939
<i>Standard deviation</i>	0.620	1.328	0.304	0.331	0.963	0.170
ΔCt (<i>DEFA1A3</i> - <i>SDHA</i>)	-4.304		-3.769		-3.697	
$\Delta\Delta Ct$ (Samples - granulocytes)	0.000		0.535		0.608	
$2^{\Delta\Delta Ct}$	1.000		1.449		1.524	

The first rows show *DEFA1A3* (left) and *SDHA* (right) Ct values obtained after qRT-PCR amplification from granulocyte, platelet-depleted PBMC and platelet cDNA and their averages are represented below. The $2^{\Delta\Delta Ct}$ results are attained by a two-step normalization obtaining the initial ΔCt outcomes – *SDHA* Ct values are subtracted from the ones obtained from *DEFA1A3* – and their following $\Delta\Delta Ct$ results – ΔCt from granulocytes is subtracted from the PBMCs or platelets ΔCt results.

Using identical approaches, from the whole set of volunteers' samples (9 samples), the average distribution of the *DEFA1A3* gene expression levels between blood cell populations are represented in Figure 10. The first column constitutes the total expression of *DEFA1A3* in whole blood whereas the second, third and fourth columns show the percentage of *DEFA1A3* gene expression $\pm$ SD attributed to granulocytes, platelet-depleted PBMCs and platelets respectively with the following overall outcome: $71.4\% \pm 16.4$ in granulocytes, $27.2\% \pm 15.1$ in platelet-depleted PBMCs and $1.4\% \pm 2.2$ in platelets.

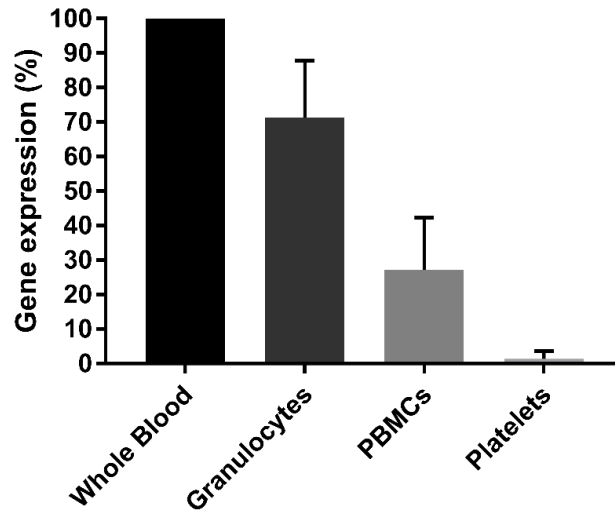


Figure 10 | Distribution of the DEFA1A3 gene expression levels between blood cell fractions. Relative DEFA1A3 gene expression (as a percentage of the total observed expression in whole blood) $\pm$ SD in granulocytes (second column), platelet-depleted PBMCs (third column) and platelets (fourth column) when compared with whole blood (first column) and using granulocytes as the control population.

Due to the low presence of *DEFA1A3* mRNA in platelets (Figure 9) and its negligible contribution to the whole blood *DEFA1A3* gene expression levels (Figure 10), we will not consider platelet populations in further experimental procedures and analyses, focusing our future approaches on platelet-depleted PBMCs (denoted only as PBMCs in the following sections).

Peripheral blood mononuclear cells have *DEFA1A3* mRNA and alpha-defensin 1-3 peptides

PBMCs were purified after a Ficoll-Paque density-gradient centrifugation and platelet depletion as described in *Materials and Methods*. From the whole set of volunteer samples (9 samples), purity result from the PBMC fraction were approximately 99% in all cases and they were used in further analyses as described in the last sections. Using volunteer IRP0001 as an example, *DEFA1A3* and *SDHA* products amplified from cDNA are presented in Figure 7 in wells two and three determining the

presence of *DEFA1A3* mRNA in PBMCs before and after depletion of platelet contaminants respectively. The same results were obtained with the rest of volunteers (data not shown). In addition, as described in the previous section, the relative contribution of PBMCs to the whole blood *DEFA1A3* gene expression levels $\pm$ SD was $27.2\% \pm 15.1$ as represented in the third column in Figure 10.

After establishing that $27.2\% \pm 15.1$ of the *DEFA1A3* gene expression levels from whole blood could be attributed to PBMCs, this blood cell fraction was used in an indirect immunocytochemistry analysis to detect the presence of alpha-defensin 1-3 peptides (as described in *Materials and Methods*). Figure 11 shows the indirect immunocytochemistry results and location of HNP1-3 peptides in granulocytes (Figure 11A) and PBMCs monitoring lymphocytes (Figure 11B) and monocytes (Figure 11C) respectively. The black arrows in Figure 11 point to the putative azurophilic granules – which are the organelles where alpha-defensin peptides are stored – stained with a brown colour after the oxidation of 3,3'-diaminobenzidine (DAB) by the peroxidase conjugated to the secondary antibody.

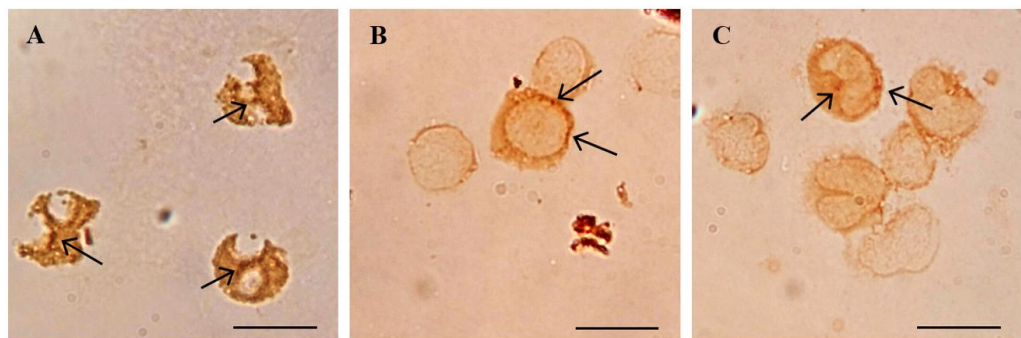


Figure 11 | Immunocytochemistry detection of alpha-defensin 1-3 peptides. Cytospin preparations from recovered granulocytes (positive control population) and platelet-depleted PBMCs were fixed, permeabilised, blocked and incubated with primary monoclonal anti-human alpha-defensin 1-3 peptides and peroxidase conjugated secondary anti-mouse antibodies. Final staining was performed by 3,3'-diaminobenzidine (DAB) leaving a brown deposit. Putative azurophilic granules are pointed out with the black arrows. Samples included: **A)** Granulocytes, **B)** Lymphocytes and **C)** Monocytes. Scale bar = $10\mu\text{m}$.

Fluorescence activated cell sorting of peripheral blood mononuclear cell subpopulations

This study provided evidence that PBMCs had *DEFA1A3* mRNA (Figure 7) contributing to $27.2\% \pm 15.1$ of the whole blood *DEFA1A3* gene expression levels (Figure 10) and accumulated alpha-defensin 1-3 peptides (Figure 11). PBMCs include multiple blood cell types such as T cells, B cells, monocytes and other minor subpopulations. The next logical approach was to individually purify each of the main PBMC subpopulations using a FACS approach in order to establish which of the PBMC subsets have *DEFA1A3* mRNA, their relative *DEFA1A3* gene expression levels and their contribution to the whole *DEFA1A3* gene expression levels from PBMCs.

Our initial analysis monitored the three major cell types from PBMCs – T cells, B cells and monocytes – and the unstained fraction (mainly composed of other minor subpopulations and cross-contamination with unstained T cells, B cells and monocytes) using the gating strategy described below and represented in Figure 12. These populations were separated from contaminant debris using a dot plot of Forward Scatter Height (FSC-H) vs. Side Scatter Height (SSC-H) gated with 'Total PBMCs' (Figure 12A). Moreover, lymphocytes and monocytes were separated based on their size and granularity, using the gates 'Lymphocytes' and 'Monocytes' represented in Figure 12A following Ratts and colleagues' work⁹⁵. Because we were looking for individual cells, total PBMCs gated as in Figure 12A were then plotted as Forward Scatter Width (FSC-W) vs. Forward Scatter Area (FSC-A) obtaining single cells by using the gate 'Singlets' as represented in Figure 12B. A minimum of 15,000 events were acquired in the previously mentioned dot plots.

Before being able to sort the three main PBMC subpopulations, we evaluated the unstained and single-color fluorochrome controls. The first

was used to monitor possible background fluorescence or autofluorescence in each of the fluorochrome channels, whilst the second was used to individually mark each PBMC subpopulation using the fluorochromes CD3-PE (T cells), CD19-BV421 (B cells) and CD14-PECy7 (monocytes). A minimum of 15,000 events per gate and population were monitored in each of the previously mentioned controls.

Then we evaluated the efficiency of the staining process using the dot plots of single fluorescence area (representing fluorescence intensity) against SSC-Height (representing granularity) as shown in Figures 12C, 12F and 12I. Unstained lymphocyte and monocyte populations are in the bottom left section, similar to the ones described in the unstained control in Figure 12A. In contrast, stained populations have an increase in the fluorescence intensity, after the fluorochrome excitation, that moved these populations across the abscissa axis (X axis)⁹⁵, as can be seen in the individual groups gated as 'CD3⁺ T cells' (Figure 12C), 'CD19⁺ B cells' (Figure 12F) and 'CD14⁺ Monocytes' (Figure 12I). To corroborate this, we monitored the differences in the fluorescence intensity between unstained and single-colour stained controls as represented in the merged histograms in Figures 12D, 12G and 12J. From the three monitored volunteers (IRP0001, IRP0082 and IRP0083), the average of the mean fluorescence intensity results $\pm$ SD are summarised in Table 4. Stained cells presented a statistically significant (Student's t-test, $p < 0.05$) increase in the fluorescence intensity compared with the unstained control. This increase allowed us to individually gate unstained and stained populations using these gates in the following FACS approach to purify T cells, B cells and monocytes respectively.

Using this gating strategy, a multicolour sample containing 7×10^6 peripheral blood mononuclear cells was sorted by an initial four-way enrichment process roughly separating T cells, B cells, monocytes and unstained cells. These cell subsets were then used in an individual-sorting purity approach obtaining highly purified samples of the different subpopulations. Results from sorted T cells, B cells and monocytes are

presented in Figures 12E, 12H and 12K, which plot cell count against the single-colour fluorescence area and are gated into unstained (left section) and stained (right section) based on the analysis of single-color fluorochrome controls. The representation of a single peak in the stained fraction corroborate the recovery of individually and purified PBMC subpopulations. Purity results and cell count data from the three monitored volunteers (IRP0001, IRP0082 and IRP0083) are summarised in Table 4.

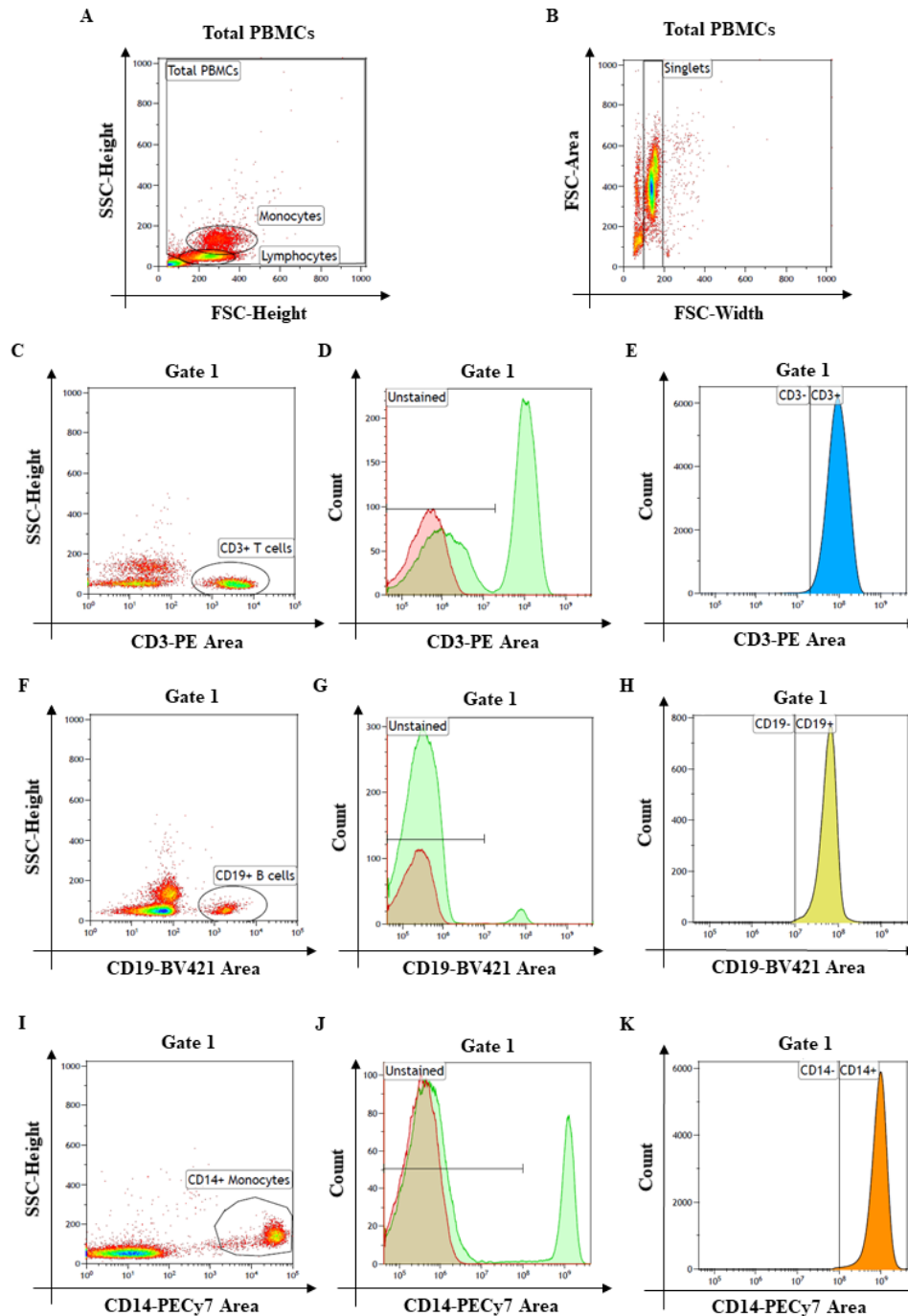


Figure 12 | Gating strategy used during the fluorescence activated cell sorting approach to purify T cells, B cells and monocytes. Recovered PBMCs were stained with the following antibodies CD3-PE (T cells), CD19-BV421 (B cells) and CD14-PECy7 (Monocytes). **A)** Forward Scatter Height (FSC-H) vs. Side Scatter Height (SSC-H) dot plot in which the debris was eliminated using the gate 'Total PBMCs'. Lymphocytes and Monocytes were separated by size and complexity. **B)** Forward Scatter Width (FSC-W) vs. Forward Scatter Area (FSC-A) dot plot for the selection of single cells using the gate 'Singlets'. **C), F) and I)** Fluorochrome area (fluorescence intensity) vs. SSC-Height dot plots to evaluate the efficiency of the staining process. Stained populations were gated as 'CD3⁺ T cells' (C), 'CD19⁺ B cells' (F) and 'CD14⁺ Monocytes' (I). **D), G) and J)** Merged histograms representing fluorochrome area vs. cell count to analyse the increase in the fluorescence intensity between unstained (red) and stained (green) PBMC subpopulations. Unstained cells were gated under the 'Unstained' window. **E), H) and K)** Fluorochrome area vs. cell count histograms showing single peaks in the stained fraction as a representation of the purity results after the sorting process. 'Gate 1' was defined as a combination of 'Total lymphocytes' and 'Singlets' gates.

Table 4. Efficiency of the staining process and cell count results from purified T cells, B cells and monocytes

Staining						
	T cells		B cells		Monocytes	
	Unstained	Stained	Unstained	Stained	Unstained	Stained
MFI (units)	391,628.88	83,988,186.67	284,653.91	63,515,936.00	329,743.81	1,270,841,600.00
SD	17,404.25	20,676,311.89	60,018.05	11,365,422.04	12,220.10	194,940,821.66
p value ($\alpha = 0.05$)	0.02		0.01		0.008	
Cell count						
	T cells		B cells		Monocytes	
% of total PBMCs	53.3 $\pm$ 5.1		5.6 $\pm$ 1.4		19.4 $\pm$ 7.2	
Purity	96.8 $\pm$ 2.4		94.6 $\pm$ 6.8		96.3 $\pm$ 5.2	

The upper half of the table represents the efficiency of the staining process by analysing the differences in the mean fluorescence intensity results between unstained and stained controls in T cells, B cells and monocytes respectively. *p*-values were obtained after a Student's *t*-test establishing a significantly threshold of $p < 0.05$. The bottom segment of the table shows the cell count results from purified T cells, B cells and monocytes after a four-way enrichment and an individual-sorting purity approaches, describing their representation (%) in the whole fraction of PBMCs and their purity results after the fluorescence activated cell sorting.

Characterization of *DEFA1A3* mRNA in T cells, B cells and monocytes

After recovering purified samples of T cells, B cells and monocytes – following the gating strategy described in the last section – RNA was extracted, cDNA was synthesised, amplified by PCR and outcomes were evaluated after 1.5% agarose gel electrophoresis as described in *Materials and Methods*. Figure 13 shows the amplified bands obtained after the PCR from each sample preceded by the bands of a 100bp DNA ladder. The upper half presents the amplified products from *DEFA1A3* mRNA followed by the amplified products from *SDHA* mRNA located in the bottom half. The qualitative data presented in Figure 13 shows that all monitored populations – T cells, B cells, monocytes and unstained cells – had detectable *DEFA1A3* mRNA, which was used in further qRT-PCR analyses to quantify their *DEFA1A3* gene expression levels and their individual contribution to the whole *DEFA1A3* gene expression levels from PBMCs (which were used as the control population).

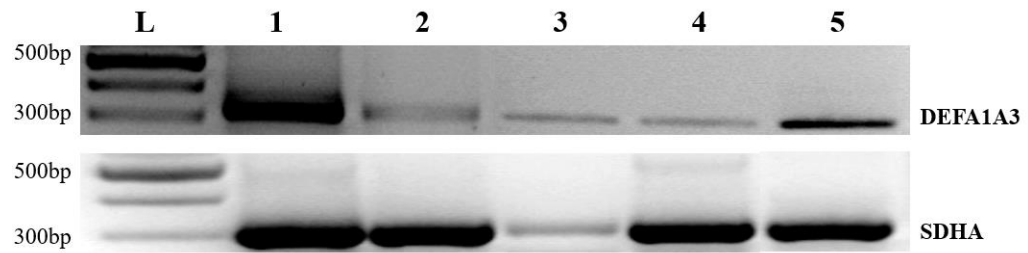


Figure 13 | PCR-amplified products from DEFA1A3 mRNA and SDHA mRNA from purified T cells, B cells and monocytes. From each purified subset, RNA was extracted and reverse-transcribed into cDNA. PCR-amplified cDNAs were evaluated after 1.5% agarose gel electrophoresis. The top panel represents the amplified products from DEFA1A3 mRNA (322bp) whereas the bottom panel shows the amplified outcomes from SDHA mRNA (306bp). Samples included: L - 100bp DNA ladder (500ng), 1 – Whole fraction of PBMCs, 2 – T cells, 3 – B cells, 4 – Monocytes and 5 – Unstained cells.

The average qRT-PCR results from the three monitored volunteers (IRP0001, IRP0082 and IRP0083) are represented in Figure 14 illustrating the percentage of DEFA1A3 gene expression $\pm$ SD distributed between T cells, B cells, monocytes and the unstained fraction respectively with the following overall outcome: $11.5\% \pm 16.7$, $2.4\% \pm 0.8$, $1.9\% \pm 1.8$ and $84.2\% \pm 17.7$.

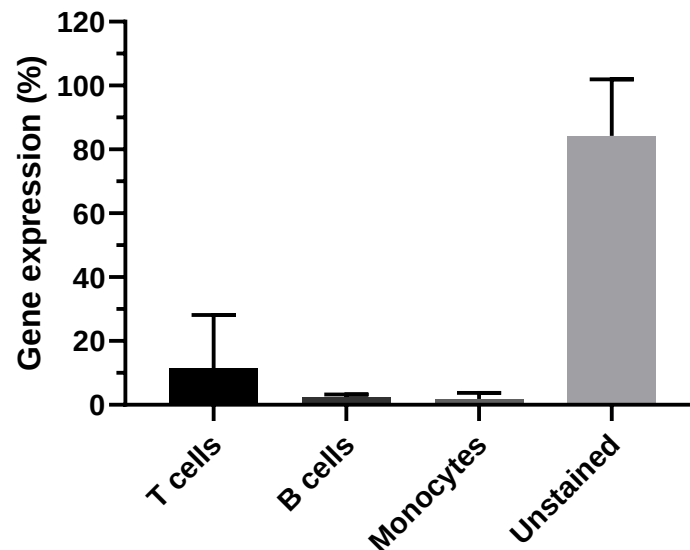


Figure 14 | Distribution of DEFA1A3 gene expression levels from purified T cell, B cells and monocytes. Relative DEFA1A3 gene expression (as a percentage of the total observed expression in PBMCs) $\pm$ SD in T cells (first column), B cells (second column), monocytes (third column) and the unstained fraction (fourth column) when compared with whole PBMCs (used as the control population).

NK cells contribute to the majority of the *DEFA1A3* mRNA from peripheral blood mononuclear cells

Unexpectedly, the gene expression analysis (Figure 14) suggested that $84.2\% \pm 17.7$ of the *DEFA1A3* gene expression from PBMCs was attributed to the unstained fraction, which was mainly composed of minor PBMC subpopulations such as NK cells and dendritic cells, and some cross-contamination of unstained T cells, B cells and monocytes. To discover which of the minor subpopulations from the unstained fraction contribute the most to the *DEFA1A3* gene expression levels from PBMCs, we decided to recover NK cells specifically using a similar gating strategy as previously described.

Briefly, contaminant debris was eliminated, lymphocytes and monocytes were grouped based on their size and granularity as shown in Figure 12A and single cells were selected as represented in Figure 12B. In addition to the previously mentioned fluorochrome panel – CD3-PE (T cells), CD19-BV421 (B cells) and CD14-PECy7 (monocytes) – we included CD56-APC antibody staining to analyse and recover the NK cell subpopulation. As mentioned before, unstained and single-color fluorochrome controls were evaluated before using this staining in the sorting approach.

The efficiency of the staining process was evaluated as previously described obtaining similar results (Student's t test, $p > 0.05$) for T cells, B cells and monocytes (Figure 12C-K). Moreover, results for NK cells are shown in Figure 15. Using the dot plot of Single Fluorescence Area vs. Side Scatter Height we evaluated the efficiency of the staining process as represented in Figure 15A. Unstained lymphocyte and monocyte populations are in the bottom left section, similar to the ones described in the unstained control represented in Figure 12A, whilst stained NK cells have an increase in the fluorescence intensity after the fluorochrome excitation that moved this population across the abscissa axis (X axis)

gated as 'NK cells' (Figure 15A). Furthermore, the differences in the fluorescence intensity between unstained and single-colour stained controls is represented in the merged histogram in Figure 15B. From the three monitored volunteers (IRP0001, IRP0084 and IRP0085), the average of the mean fluorescence intensity results $\pm$ SD are summarised in Table 5. Stained NK cells presented a significant (Student's t-test, $p < 0.05$) increase in the fluorescence intensity compared with the unstained control. This increase allowed us to individually gate unstained and stained populations, that in combination with the gating strategy described for T cells, B cells and monocytes, allowed us to purify four PBMC subpopulations (including NK cells).

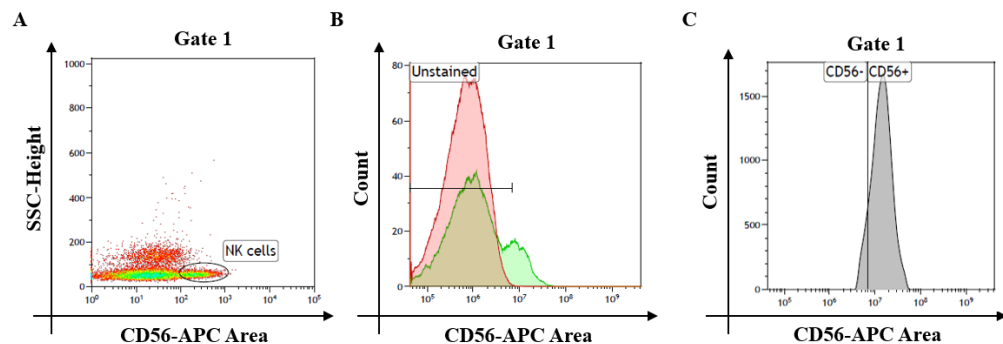


Figure 15 | Gating strategy used during the fluorescence activated cell sorting approach to purify natural killer cells. Recovered PBMCs were stained with the following antibodies CD3-PE (T cells), CD19-BV421 (B cells), CD14-PECy7 (Monocytes) as described in Figure 12 and with CD56-APC (NK cells). **A)** Fluorochrome area (fluorescence intensity) vs. Side Scatter Height (SSC-H) dot plot to evaluate the efficiency of the staining process. Stained NK cells were gated as 'NK cells'. **B)** Merged histogram representing fluorochrome area vs. cell count to analyse the increase in the fluorescence intensity between unstained (red) and stained (green) NK cells. Unstained cells were gated under the 'Unstained' window. **C)** Fluorochrome area vs. cell count histogram showing a single peak in the stained fraction as a representation of the purity results after the sorting process. 'Gate 1' was defined as a combination of 'Total lymphocytes' and 'Singlets' gates.

Then, a multicolour sample containing 7×10^6 peripheral blood mononuclear cells was sorted by an initial five-way enrichment process roughly separating T cells, B cells, monocytes, NK cells and unstained cells which were then used in an individual-sorting purity approach obtaining highly purified samples. Results from sorted T cells, B cells, monocytes and NK cells are represented in Figures 12E, 12H, 12K and Figure 15C respectively. The appearance of a single peak in the stained

fraction corroborates the recovery of individually and purified PBMC subpopulations. From the three monitored volunteers (IRP0001, IRP0084 and IRP0085), purity results and cell count data for NK cells are summarised in Table 5.

Table 5. Efficiency of the staining process and cell count results from purified NK cells

	Staining	
	Unstained	Stained
MFI (units)	676,704.13	17,476,892.00
SD	62,598.47	6,776,537.03
p value ($\alpha = 0.05$)	0.049	
	Cell count	
	NK cells	Unstained cells
% of total PBMCs	10.0 $\pm$ 2.2	11.7 $\pm$ 2.4
Purity	90.9 $\pm$ 3.1	-

The upper section of the table represents the differences in the mean fluorescence intensity results between unstained and stained controls in NK cells to evaluate the efficiency of the staining process. p-values were obtained after a Student's t-test ($p < 0.05$). The bottom segment of the table shows the cell count results of NK cells after a five-way enrichment and an individual-sorting purity approaches and the results from the recovered unstained fraction, describing their representation (%) in the whole fraction of PBMCs and the purity results after the fluorescence activated cell sorting.

Purified PBMC subpopulations (including NK cells) were used in identical gene expression analyses as described during the characterization of *DEFA1A3* mRNA in T cells, B cells and monocytes. Figure 16 shows in the upper row the amplified products from *DEFA1A3* mRNA followed by the amplified products from *SDHA* mRNA located in the bottom row. The qualitative data presented in Figure 16 shows that all monitored populations – T cells, B cells (Figure 13), monocytes, NK cells and unstained cells – had detectable *DEFA1A3* mRNA, which was used to quantify their relative *DEFA1A3* gene expression levels and their individual contribution to the whole *DEFA1A3* gene expression levels from PBMCs.

Figure 16 also shows a second unexpected band at approximately 350bp (marked with *Unexpected*) amplified during *DEFA1A3*-specific PCR amplification. This product from two samples (CD3⁺ T cells from IRP0001 and CD14⁺ monocytes from IRP0085) which only expressed the unexpected band (data not shown) was sequenced as described in *Materials and Methods*. Sequencing results and BLAST analysis demonstrated that the unexpected band was derived from mispriming of *DEFA1A3* primers on sequence from the 18S rRNA (grouped in the 40S rRNA cluster), which will not interfere with further gene expression analyses as the TaqMan probes were specially designed to hybridise with the *DEFA1A3* mRNA sequence. This unexpected band was only detected in samples with low levels of RNA and high number of cycles during PCR amplification that allowed the primers to amplify the rRNA in the absence or low-level presence of *DEFA1A3* mRNA. Only eleven out of the twenty bases from the forward primer match the rRNA (54.5% GC-content), whilst in the reverse primer fourteen out of the nineteen bases presented this complementarity (57.1% GC-content).

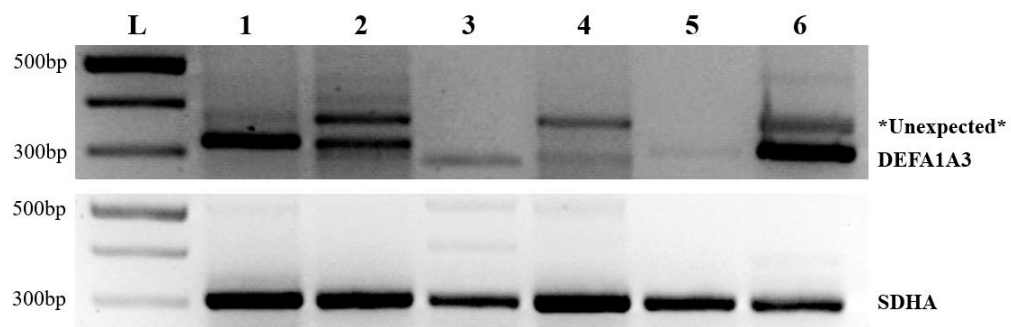


Figure 16 | PCR-amplified products from *DEFA1A3* mRNA and *SDHA* mRNA from purified T cells, B cells, monocytes and natural killer cells. From each purified subset, RNA was extracted and reverse-transcribed into cDNA. PCR-amplified cDNAs were evaluated after 1.5% agarose gel electrophoresis. The top row shows the 18S (*Unexpected*, 341bp) and *DEFA1A3* (322bp) products amplified from cDNA, whereas the bottom row presents the *SDHA* products (306bp). Samples included: L - 100bp DNA ladder (500ng), 1 – Whole fraction of PBMCs, 2 – T cells, 3 – B cells, 4 – Monocytes, 5 – NK cells and 6 – Unstained cells.

Figure 17 shows the average qRT-PCR results from the three monitored volunteers (IRP0001, IRP0084 and IRP0085) describing the percentage of *DEFA1A3* gene expression $\pm$ SD distributed between T cells, B cells,

monocytes, NK cells and unstained cells respectively, and the overall outcomes were $0.8\% \pm 0.2$, $1.0\% \pm 0.2$, $3.4\% \pm 6.5$, $83.8\% \pm 2.4$ and $11.0\% \pm 0.5$.

Differences in the gene expression levels attributed to T cells, B cells and monocytes between the four- and five-way enrichment step of the FACS approach were not significantly different (Student's t test, $p=0.38$, $p=0.10$ and $p=0.7$ respectively), whilst the ones attributed to the unstained fraction were highly distinct (Student's t test, $p=0.007$).

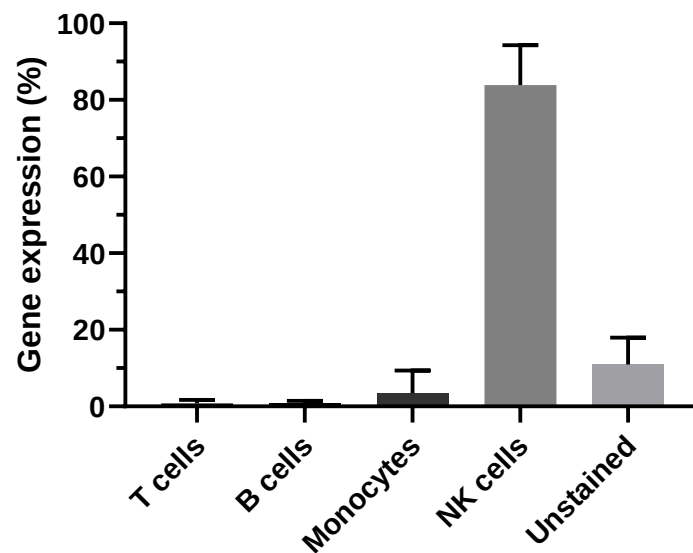


Figure 17 | Distribution of DEFA1A3 gene expression levels from purified T cell, B cells, monocytes and natural killer cells. Relative DEFA1A3 gene expression (as a percentage of the total observed expression in PBMCs) $\pm$ SD in T cells (first column), B cells (second column), monocytes (third column), NK cells (fourth column) and the unstained fraction (fifth column) when compared with whole PBMCs (used as a control population).

DISCUSSION AND CONCLUSIONS

Our results reconfirmed the presence of *DEFA1A3* mRNA (~70% of whole blood *DEFA1A3* gene expression levels) and alpha-defensin 1-3 peptides in granulocytes^{96,97}. In addition, our data showed consistent evidence of the presence of *DEFA1A3* mRNA in platelets (~1% of whole blood *DEFA1A3* gene expression levels), lymphocytes and monocytes (~30% of whole blood *DEFA1A3* gene expression levels) which also present HNP1-3 peptides. Using a FACS approach we purified T cells, B cells, monocytes, NK cells and recovered the unstained fraction, in which all the subsets had detectable *DEFA1A3* mRNA with most of the contribution to the *DEFA1A3* gene expression levels associated with NK cells (~85% of the whole *DEFA1A3* gene expression levels from PBMCs).

Our set of results was obtained using peripheral blood from nine healthy, European volunteers (six men and three women). The small number of samples was due to limitations of the time available for the blood extraction and access to cell-sorting equipment. We restricted the studied group to Europeans in order to maintain the uniformity of the cohort and the number of copies of the *DEFA1A3* gene (usually four to eleven copies per diploid genome) as described in Khan et al.'s work⁶², even though we did not determine the number of copies of *DEFA1/DEFA3*⁶³, the absence of *DEFA3*⁶⁵ or their haplotype classes⁶⁷ which may have varied between individuals. Furthermore, volunteer ages were in a restricted range, between 20-25 (except volunteer IRP0082), because there are no available publications that describe if ageing affects the expression levels of *DEFA1A3* and/or the accumulation of alpha-defensin 1-3 peptides. Future work should build a bigger cohort of sex and age-matched European volunteers monitoring the number of copies of *DEFA1/DEFA3*, which we expect will reproduce the results presented in this project as well as define a correlation between the copy number variations in the *DEFA1A3* locus and the *DEFA1A3* mRNA levels and/or the amount of

alpha-defensin 1-3 peptides. Moreover, building a European cohort can be used as a control group to analyse different outcomes in other populations such as Asians and Africans as it has been described that these populations present differences in the number of copies of the *DEFA1A3* gene⁶⁴, the lack of *DEFA3*⁶⁵ and the haplotype classes⁶⁷.

In this project, ethylenediamine tetra-acetic (EDTA) was used as the anticoagulant method because its carboxylic groups chelate extracellular calcium preventing clot formation⁹⁸. The presence of calcium has been implicated in multiple signalling pathways. Prolonged calcium exposure increases platelet volume, changing their irregular shape morphology to a spherical conformation⁹⁸, increases the synthesis of beta-defensins in gingival epithelial cells⁹⁹ and calcium salts – usually used to reduce cholesterol in the bloodstream¹⁰⁰ – reduce the expression levels of *DEFA1A3* in blood cells reported in a hypercholesterolemia study conducted by Li and colleagues¹⁰¹. Two alternative anticoagulant methods – sodium citrate and heparin – can be used to further evaluate if the absence or presence of extracellular calcium in the anticoagulated blood samples modifies our results. Similar to EDTA, sodium citrate inhibits blood coagulation by sequestering calcium¹⁰², from which we expect to obtain similar outcomes that will reconfirm the results presented in this project using EDTA. In contrast, heparin inhibits clot formation by activating inhibitor antithrombin III which inactivates thrombin, factor Xa and other proteases but leave untouched the serum calcium levels in the blood¹⁰³ which can be used to determine if the presence of extracellular calcium affects the expression of *DEFA1A3* and/or the accumulation of alpha-defensin 1-3 peptides.

To describe the presence of *DEFA1A3* mRNA in blood samples, this project followed standardised available protocols^{93,104,105} to separate the different blood cell populations. Separation of platelet-depleted PBMCs and granulocytes by density-gradient centrifugation worked extremely well as can be seen by their high purity results (Figure 6), whilst the platelet purification protocol was not as efficient (Figure 8). Cross-

contamination between blood cell populations was monitored by counting cells in 40 randomly selected fields with the main problem in evaluation being the differences in the volume between cell types, especially important between leukocytes and platelets. To bypass this problem, based on Haley and colleagues' work⁹⁰, we used a mean of 120 platelets as an equivalent (by volume) to one white blood cell to standardise the results and allow further comparisons, with the proviso that modifying this equivalence can affect the purity results and interpretation of subsequent analyses.

Density-gradient centrifugation has a limited ability for cell separation, only obtaining bulk cell fractions – platelets, PBMCs and granulocytes. Our data suggest that PBMCs had *DEFA1A3* mRNA (Figure 7) and because this fraction groups different cell subpopulations², in order to accurately characterise the distribution of the *DEFA1A3* mRNA levels between different blood cell populations, a FACS approach was used to individually purify T cells, B cells, monocytes, NK cells, and we also recovered the unstained fraction. The staining process and gating strategy (Figure 12 and Figure 15) worked efficiently and allowed us to individually recover each of the subpopulations with extremely high purity results (>95%) in T cells and monocytes (Table 4) and slightly lower (90-95%) in B cells and NK cells (Table 4 and Table 5 respectively). Moreover, the number of recovered cells was significantly higher in the first two. We believed that the marginally lower efficiency in B cells and NK cells can be solved by optimising the antibody staining procedure or monitoring these subpopulations using an individual single-colour enrichment and purity steps during the cell sorting. The FACS approach was selected over other blood cell separation methods because it has the capacity to simultaneously study multiple parameters at a single-cell level which can analyse the heterogeneity of the population; recovered cells usually have high purity (>95%) and sorted cells can be used in subsequent *in vitro* cell culture studies¹⁰⁶. An alternative method based on a positive/negative selection using antibody-conjugated magnetic beads can be used as an enrichment approach pre-selecting a specific

population before the FACS, increasing the number of recovered cells and maintaining low levels of cross-contamination¹⁰⁷.

Due to the short lifespan of granulocytes² and the previously described effects of EDTA in platelets⁹⁸, RNA from these cell fractions was obtained within an hour after the blood extraction, decreasing possible RNA degradation¹⁰⁸. For PBMCs, the results from the 24-hour RNA degradation assay (see *Materials and Methods*) established that within the first eight hours since blood extraction there were no significant differences in the RNA integrity (data not shown). The purification of PBMC subpopulations and subsequent RNA extraction were performed inside this eight-hour window, assuring excellent RNA integrity. Moreover, the high values in the A260/A280 and A260/A230 ratios from the NanoDrop analysis reported that our RNA preparations presented negligible contamination with residual reagents from the RNA extraction procedure (data not shown). Inconsistencies in the RNA integrity results can arise because the comparison between the 28S and 18S rRNA bands from the agarose gel images is subject to human interpretation. To develop a highly reproducible and user-independent RNA integrity analysis, additional studies can include the use of automated software algorithms like the one presented in Schroeder and colleagues' work. Their RNA integrity number (RIN) algorithm use multiple parameters (such as area and height of the peaks in the electropherogram) which provide information about the RNA integrity obtaining a scale from zero to ten in which ten reflects maximum integrity and zero indicates complete RNA degradation¹⁰⁹.

cDNAs for end-point PCR were obtained by a reverse-transcription of the RNAs using oligo(dT) primers which specifically pre-selected polyadenylated RNAs¹¹⁰. This approach can present serious limitations like occasional incomplete reverse transcription due to the location of the poly-A tail at the 3'-end or by partial priming of the oligo(dT) primers in a central area of the RNA which may block the reverse transcriptase. In contrast, cDNAs for qRT-PCR were obtained using random hexamer

primers which have the advantage of amplifying whole RNA which reflects the actual abundance of the target sequence within the environment. Furthermore, random hexamer primers are more suitable for amplifying low quality RNAs and are not restricted to the 3'-end of the molecule which increases the chances of amplifying the 5'-end. The main limitation of using random hexamer primers, however, is the low probability of obtaining full-length cDNA molecules¹¹¹.

To analyse the presence of *DEFA1A3* mRNA we used an end-point, *DEFA1A3*-specific PCR and to establish the gene expression levels, attributed to each blood cell population, we performed a $2^{\Delta\Delta Ct}$ analysis with the 'quantitative' results obtained from a TaqMan-based qRT-PCR. There are several important differences between gel-based PCR and qRT-PCR analyses. Generally, gel-based analyses are cheaper, faster and easier to perform than qRT-PCR. Difficulties arise, however, when an attempt is made to quantify the expression levels or establish small differences in the samples, because of its low sensitivity¹¹¹. By contrast, TaqMan-based qRT-PCR produces more accurate measurements as it involves monitoring fluorescence in real time during the exponential phase of the amplification. Furthermore, the specificity of the TaqMan probes ensures that the fluorescent signal is essentially derived from amplicons of the target sequence¹¹². One major drawback of this approach is that the background fluorescence is not an absolute value, so the threshold cycle (Ct) – which is defined as the cycle when the sample fluorescence exceeds the background fluorescence – is also going to be variable. Since the Ct value is used to determine the gene expression levels – when it is compared with a reference gene or a standard curve – analyses must be interpreted with caution¹¹¹.

In the $2^{\Delta\Delta Ct}$ method⁹⁴ used to establish the gene expression levels attributed to each blood cell population, the representation of each blood cell type was adopted from information summarised in Janeway's textbook² and *SDHA* was selected as the reference gene according to previous studies¹¹³. This approach needs to be interpreted with caution

because in some contexts it does involve potential measurement errors due to variances in the number of white blood cells in the bloodstream² and the normalization of the 'quantitative' results only using one reference gene¹¹⁴. In future investigations, we suggest monitoring (by flow cytometry), whole blood samples from the different volunteers to obtain a more accurate and specific representation of each blood cell population. Moreover, further research should include TATA-binding protein (*TBP*)¹¹³, β 2-microglobulin (*B2M*) or peptidyl-prolyl isomerase B (*PPIB*)¹¹⁵ as additional reference genes to increase the resolution and accuracy of the results.

However, gel-based PCR and qRT-PCR analyses suffer from some serious disadvantages to quantify gene expression levels. To develop a full picture of the *DEFA1A3* gene expression levels in different blood cell populations, additional studies might include RNA-seq analyses. Briefly, some of the advantages of using RNA-seq analyses include that there is no need of previous knowledge about the transcriptome of the target cell, its high-throughput sequencing technology allows us to determine the events behind the alternative splicing that will generate multiple isoforms and define single nucleotide polymorphisms in a transcribed region, it has a wide dynamic range to quantify gene expression level and, theoretically, RNA-seq analyses present a single molecule-based resolution providing specialised information such as the connectivity between exons. Its main disadvantages are the development of strand-specific libraries that are important to analyse the overlapping transcription regions from opposite directions and the limitation in the storage, analysis and interpretation of the data with the available bioinformatic tools^{116,117}.

In addition, to determine the presence of HNP1-3 peptides we used an immunocytochemistry approach. Some limitations of this method include high sample processing, lack of standardised protocols, inconsistencies in the outcomes due to the intense background noise and lack of quantitative results^{118,119}. Standard techniques such as sodium dodecyl

sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western Blotting are not recommended to detect molecules like antimicrobial peptides due to their cationic nature – which necessitates application of voltage in a reverse polarity – and small size (<10kDa) – which increases the chances of losing the peptides during the transfer to the membrane¹²⁰. In future investigations, we suggest using different analytical techniques such as flow cytometry^{107,121} or reverse high-performance liquid chromatography (rHPLC)¹²² which will enhance the resolution and sensitivity to accurately describe the presence of intracellular alpha-defensin 1-3 peptides and will provide quantitative results.

Here, our results reconfirmed the presence of *DEFA1A3* mRNA in granulocytes (Figure 7) and their main contribution to the whole blood *DEFA1A3* gene expression levels (Figure 10). These findings are consistent with those of Cowland et al. (1999) who described the presence of *DEFA1A3* mRNA in promyelocytes and a minimum active expression in mature granulocytes which suggested that the outcomes reported in granulocytes could be considered as 'leftovers' from the *DEFA1A3* mRNA from their precursors⁹⁶. In accordance with these results, Mora-Jensen et al. (2011) described a significant increase in the levels of *DEFA1A3* mRNA in promyelocytes and low but stable levels during the granulocytes' development from myelocytes to mature granulocytes¹²³. These results provided further support for the hypothesis that granulocytes are not able to synthesise new *DEFA1A3* mRNA. Built on the constitutive presence of *DEFA1A3* mRNA in granulocytes, we selected this blood cell fraction as our internal control for the gene expression analyses (see *Materials and Methods*); however, it must be interpreted with caution. Based on the unpublished RNA-seq meta-analysis results from (Lipska I. and Armour J., unpublished), *DEFA1A3* mRNA is mainly present in neutrophils but it can also be detected in eosinophils. In future investigations, it might be possible to use a FACS approach to purify the three subpopulations of granulocytes¹²⁴ – neutrophils (CD45⁺CD66b⁺CD16⁺), eosinophils (CD45⁺CD66b⁺CD16⁻)

and basophils (CD45⁺CD66b⁺CD123⁺) – which will allow the use of purified neutrophils as the reference population in the $2^{\Delta\Delta Ct}$ analysis improving the accuracy of the distribution of the *DEFA1A3* gene expression levels between blood cell populations.

In addition, we were able to determine the presence of alpha-defensin 1-3 peptides in granulocytes (Figure 11) which broadly supports the work of other studies. Valore et al.'s work (1992) investigated the alpha-defensin 1-3 peptide maturation process. Their results showed that granulocytes only contained the mature form of alpha-defensin 1-3 peptides (29-30 amino acids), whereas in promyelocytes, mature and immature peptides (79 amino acids) could be found supporting the hypothesis that granulocytes are pre-loaded with HNP1-3 peptides from their precursor⁹². This also accorded with later observations from Faurschou et al.'s study (2003), which showed that during granulopoiesis, azurophilic granules – organelles where alpha-defensin peptides are stored – were formed and loaded with alpha-defensin 1-3 peptides near the promyelocyte/myelocyte transition⁹⁷.

Our results determined the presence of *DEFA1A3* mRNA at a low level in platelets (Figure 7 and Figure 9) and their negligible contribution to the whole blood *DEFA1A3* gene expression levels (Figure 10). In accordance with these results, Valle-Jimenez and colleagues' work demonstrated the presence of *DEFA1A3* mRNA in circulating platelets and in platelet-like particles derived from cultured megakaryocytes. Their data suggested that megakaryocytes present high levels of *DEFA1A3* mRNA during maturation in the bone marrow which is then transferred to platelets. On the other hand, the accumulation of alpha-defensin 1-3 peptides decreases during megakaryocyte differentiation. Moreover, they provided evidence which suggested that platelets were able to increase their intracellular levels of HNP1-3 peptides by internalising serum alpha-defensin 1-3 peptides instead of by *de novo* protein synthesis¹²⁵. Based on platelets' negligible contribution to the whole blood *DEFA1A3* gene expression levels and the limited support to our findings when reviewing

the literature, we excluded platelets in the detection of their alpha-defensin 1-3 peptides (reviewed in Valle-Jimenez and colleagues' work¹²⁵).

In contrast to these blood cell fractions, previous publications evaluating the presence of *DEFA1A3* mRNA and alpha-defensin 1-3 peptides in peripheral blood mononuclear cells have been contradictory and inconclusive. Agerberth et al. (2000) suggested that *DEFA1A3* mRNA could be found in $\gamma\delta$ -T cells, B cells and NK cells and alpha-defensin 1-3 peptides were detected in these PBMC subpopulations as well as in monocytes¹²⁶. However, Obata-Onai et al.'s study (2002) had challenged some of Agerberth's conclusions, arguing that *DEFA1A3* mRNA was only detected in NK cells and reporting an increase in the *DEFA1A3* mRNA levels after stimulation with IL-15¹²⁷. Chalifour and colleagues' work (2004) provided further evidence of the presence of *DEFA1A3* mRNA and HNP1-3 peptides in CD56^{dim}CD3⁻ and CD56^{bright}CD3⁻ NK cell subpopulations as well as in NKT cells (CD56⁺CD3⁺). Furthermore, an increased accumulation of alpha-defensin 1-3 peptides was reported after KpOmpA stimulation without an association with the *DEFA1A3* mRNA levels¹²⁸. In addition, Rodríguez-García et al.'s study (2007) provided evidence of the presence of *DEFA1A3* mRNA and HNP1-3 peptides in monocytes and monocyte-derived dendritic cells. Their data suggested that immature dendritic cells presented higher levels of *DEFA1A3* mRNA and alpha-defensin 1-3 peptides when compared with monocytes and mature dendritic cells. Moreover, an increased presence of *DEFA1A3* mRNA was reported after stimulation with IL-1 β ¹²⁹. These studies have tended to focus on determining the presence of *DEFA1A3* mRNA in a specific PBMC subpopulation rather than providing a complete view across the different PBMC subsets. Furthermore, one major limitation was the lack of quantitative analyses to describe the *DEFA1A3* gene expression levels attributed to each PBMC subpopulation and their contribution to the whole blood *DEFA1A3* gene expression levels. In addition, in all of these studies, cultured cells were

used for the analyses not considering possible RNA degradation and changes in the *DEFA1A3* mRNA basal levels during the cell culture. Our results simultaneously determined the consistent presence of *DEFA1A3* mRNA (across the five selected volunteers) at detectable levels in all sorted-purified PBMC subpopulations (Figure 13 and Figure 16) and established their relative *DEFA1A3* gene expression levels (Figure 14 and Figure 17). These results made several improvements to the current literature by providing a complete analysis of the PBMC fraction, quantitative outcomes of the *DEFA1A3* mRNA levels in individual PBMC subpopulations as well as the contribution of each PBMC subset to the whole blood *DEFA1A3* gene expression levels. Moreover, this project presented the advantage of using immediately recovered PBMC subpopulations after the FACS which we suggest preserved the unstimulated status of the cells and ensured the recovery of excellent quality RNAs (used in subsequent analyses).

The high purity results from sorted CD3⁺ T cells and CD14⁺ monocytes (Table 4) supported the idea that their low-level contribution to the *DEFA1A3* gene expression levels from PBMCs (Figure 14 and Figure 17) was not due to cross-contamination with other cell types but to low-level expression of *DEFA1A3*. Despite these results, future work should include a new FACS approach which will take into consideration the heterogeneity of T cells and monocytes². We suggest that this new analysis should discriminate between the four monocyte subgroups following Villani et al.'s work (2017) in which we expect to detect the presence of *DEFA1A3* mRNA in 'Mono 1', 'Mono 2' and 'Mono 4' groups based on their cytotoxic description⁴⁷. Furthermore, a distinction between $\gamma\delta$ -T cells¹²⁶, $\alpha\beta$ -CD4⁺ T cells⁴⁸ and $\alpha\beta$ -CD8⁺ T cells¹³⁰ must be included. The discrimination of these subsets could allow us to reconfirm the results from previous studies^{126,129} and provide some new evidence of the presence of *DEFA1A3* mRNA in different monocyte subgroups as well as in CD4⁺ and/or CD8⁺ T cells.

Due to the limited number of recovered B cells and their relatively low purity (Table 4), the presence of *DEFA1A3* mRNA (Figure 13) and their negligible contribution to the *DEFA1A3* gene expression levels from PBMCs (Figure 14 and Figure 17) might even be attributed not to a low-level expression of *DEFA1A3* in this cell population but to the expression derived from contaminating cells of other types. To solve this problem, possible changes in the FACS approach such as an optimization of the antibody staining procedure and a second purity step might increase the number recovered cells with higher purity outcomes which could allow us to establish if the presence of *DEFA1A3* mRNA and their related gene expression levels can be specifically attributed to B cells.

The major unanticipated finding was that NK cells contributed the most to the *DEFA1A3* gene expression levels from PBMCs (Figure 17), but which still needs to be interpreted with caution due to the limited number of recovered cells and their relatively poor purity outcomes (Table 5). The discrepancies between the qualitative and quantitative results might be derived from the poor-quality and low-quantity RNAs recovered from NK cells, which may have impeded the cDNA synthesis only allowing detection by the highly-sensitive TaqMan-based qRT-PCR. We suggest that future work should include a positive selection of NK cells using anti-CD56 conjugated magnetic beads followed by a FACS which should improve the number of recovered cells, their purity outcomes and enhance the RNA characteristics. These suggestions could allow us to reproduce the results presented in this project providing some new qualitative evidence of the presence of *DEFA1A3* mRNA in this cell fraction, discriminate between $CD56^{dim}CD3^{-}$, $CD56^{bright}CD3^{-}$ and $CD56^{+}CD3^{+}$ subsets as well as reconfirm the outcomes from previous studies^{126–128}.

Finally, we have also established the presence of *DEFA1A3* mRNA in the unstained fraction (Figure 13 and Figure 16) with a significant reduction in its *DEFA1A3* gene expression levels between the four- (Figure 14) and five-way (Figure 17) enrichment strategies from the FACS protocol which

might be related to the presence or absence of NK cells in this subset respectively (see *Results*). We hypothesise that the remaining expression associated with the unstained fraction (Figure 17) could be produced by cross-contamination with the previous peripheral blood mononuclear cell types or be attributed to other minor subpopulations such as dendritic cells or hematopoietic stem cells which will support the findings from Rodríguez-García et al.'s study (2007)¹²⁹. Future investigations should purify dendritic cells (CD11c⁺CD85k⁺CD123^{+/-})¹³¹ and hematopoietic stem cells (CD34⁺)¹³² that in combination with the previous recovered cell types, will provide a more accurate distribution of the *DEFA1A3* gene expression levels between PBMC subpopulations.

In addition, our results determined the presence of alpha-defensin 1-3 peptides in circulating PBMCs (Figure 11). This finding was also reported by numerous previous studies^{126,128,129}; however, the major problem with our outcomes is that they failed to establish the presence of HNP1-3 peptides in individual PBMC subpopulations and must be interpreted with caution based on the limitations of the immunocytochemistry approach.

In reviewing the literature, there is no general agreement about whether lymphocytes and monocytes can synthesise new *DEFA1A3* mRNA and/or increase the amount alpha-defensin 1-3 peptides after stimulation. However, based on previous studies we hypothesise that the presence of specific cytokines can increase the levels of *DEFA1A3* mRNA and HNP1-3 peptides in some PBMC subsets^{127,129}. Future studies should include cell culture approaches using the purified PBMC subpopulations to establish if the presence of inflammatory cytokines such as IL-1 β , IL-10 or IL-15 and pathogen-associated molecular patterns (PAMPs) like bacterial flagellin or lipopolysaccharides (LPS) modulate their expression of *DEFA1A3* and/or the accumulation of alpha-defensin 1-3 peptides. Possible bias in detecting alpha-defensin 1-3 peptides needs to be taken into consideration due to the secretion of these molecules. We suggest including a treatment with monensin A – which blocks the transport through the Golgi and inhibits the formation of secretory vesicles¹⁰⁷ – to

describe the level of intracellular alpha-defensin 1-3 peptides, whereas in non-treated cells an ELISA approach can be used to monitor the level of secreted HNP1-3 peptides.

In conclusion, this study has shown that *DEFA1A3* mRNA was present in granulocytes, platelets and PBMCs (Figure 7) and most of the contribution to the whole blood *DEFA1A3* gene expression levels was attributed to granulocytes (Figure 10). However, the remaining expression was almost entirely determined in PBMCs, ignoring the negligible contribution associated with platelets (Figure 10). One unanticipated finding was the detection of *DEFA1A3* mRNA in all purified PBMC subpopulations (Figure 13 and Figure 16), in which most of the contribution to the *DEFA1A3* gene expression levels from PBMCs was attributed to NK cells (Figure 17). In addition, this study reconfirmed the presence of alpha-defensin 1-3 peptides in granulocytes and PBMCs (Figure 11). These findings made several contributions to the current literature, including quantitatively determining the *DEFA1A3* gene expression levels in each individual blood cell population and their relative contribution to the whole blood *DEFA1A3* gene expression levels; however, due to the potential limitations during the project, conclusions must be interpreted with caution. A natural progression of this work will perform RNA-seq analyses in sorted blood cell types, flow cytometric immunophenotyping of HNP1-3 peptides in PBMC subpopulations and *in vitro* cell culture approaches. In addition, we suggest including paralogue ratio tests (PRT) to determine the number of copies of the *DEFA1A3* gene¹³³, that will provide new evidence to describe if there is a correlation between the copy number variations in the *DEFA1A3* locus, the presence of *DEFA1A3* mRNA and the accumulation of alpha-defensin 1-3 peptides.

MATERIALS AND METHODS

Sources of antibodies and reagents

Reagents: AMPure (Beckman Coulter Life Science, USA), Bovine serum albumin (BSA, Sigma-Aldrich, USA), Chloroform (ThermoFisher Scientific, UK), CleanSeq (Beckman Coulter Life Science, USA), Ethanol (Sigma-Aldrich, USA), Ethylenediamine tetra-acetic (EDTA, Sigma-Aldrich, USA), Ficoll-Paque PLUS (GE Healthcare, UK), Haematoxylin (Sigma-Aldrich, USA), Hemacolor® rapid staining of blood smear (Sigma-Aldrich, USA), Histopaque-1077 (Sigma-Aldrich, USA), Histopaque-1119 (Sigma-Aldrich, USA), Human serum (Sigma-Aldrich, USA), Moloney murine leukaemia virus reverse transcriptase (New England Biolabs, UK), Mouse serum (Sigma-Aldrich, USA), Phosphate-buffered saline (ThermoFisher Scientific, UK), SIGMAFAST™ 3,3'-Diaminobenzidine tablets (DAB, Sigma-Aldrich, USA), Taq DNA polymerase (New England Biolabs, UK), TaqMan *DEFA1A3* probes (ThermoFisher Scientific, UK), TaqMan *SDHA* probes (ThermoFisher Scientific, UK), Tetro™ cDNA Synthesis Kit (Bioline, UK), Triton™ X-100 (Sigma-Aldrich, USA) and TRIzol® (Invitrogen, UK).

Antibodies: Mouse PE/Cy7-conjugated monoclonal anti-human CD14 (CD14-PE/Cy7, Clone: RMO52, Beckman Coulter Life Science, USA), Mouse BV421-conjugated monoclonal anti-human CD19 (CD19-BV421, Clone: HIB19, Biolegend, USA), Mouse PE-conjugated monoclonal anti-human CD3 (CD3-PE, Clone: UCHT1, Beckman Coulter Life Science, USA), Mouse APC-conjugated monoclonal anti-human CD56 (CD56-APC, Clone: HCD56, Biolegend, USA), Goat peroxidase conjugated anti-mouse IgG (Clone: A4416, Sigma-Aldrich, USA), Mouse monoclonal anti-human neutrophil peptide 1-3 (Clone: D21, Hycult Biotech).

Volunteers and ethical considerations

The study used healthy European volunteers. All participants were informed of the aims of the project and provided their written informed consent on the day of blood extraction. The project and protocols were approved by the Faculty of Medicine and Health Sciences Ethics Committee of the University of Nottingham in compliance with the Declaration of Helsinki (Approval BT 10-02-2010).

Cell isolation

Blood samples were obtained by venepuncture in EDTA-tubes from healthy European volunteers. Platelet-enriched plasma (PRP) was prepared by a procedure adapted from previously described methods⁹³. Briefly, platelets were concentrated in the plasma with an initial slow spin at 300g for 10 minutes. Then, platelets were pelleted with a faster spin at 800g for 20 minutes. Pellets were resuspended in PBS/EDTA until further use.

Platelet-depleted blood was diluted 1:1 in PBS/EDTA. Granulocytes and peripheral blood mononuclear cells were isolated by Ficoll-Paque PLUS or double Histopaque density-gradient centrifugations according to the manufacturer's directions. Recovered cells were washed twice with PBS/EDTA to reduce possible cross-contamination. Also, the granulocyte fraction was treated with an Ammonium-Chloride-Potassium (ACK) erythrocyte lysis buffer at room-temperature for 15 minutes in order to eliminate erythrocyte contaminants. Pellets were resuspended in PBS/EDTA until further use.

Blood cell populations were subjected to cytocentrifugation – final volume of 500 μ L with a concentration of 140cells/ μ L – at 1500rpm for granulocytes and 500rpm for PBMCs for 5 minutes using a Shandon Cytospin-2 (Thermo Electron Corporation, USA). Slides were air-dried

and further use in HNP1-3 immunocytochemistry detection assay or staining using a standard Hemacolor® method (following manufacturer's directions). Cover slides were prepared with paraffin oil.

Fluorescence-activated cell sorting (FACS)

After PBMCs were isolated using a density-gradient centrifugation, samples were prepared following previously published protocols¹³⁴ with minor modifications. Briefly, cells were pelleted and resuspended in blocking buffer – mouse serum in PBS/BSA - and incubated for 60 minutes at 4°C. After blocking, cells were incubated for 60 minutes at 4°C using antibodies CD3-PE, CD14-PE/Cy7, CD19-BV421 and CD56-APC to stain surface markers of T cells, monocytes, B cells and NK cells respectively. In order to obtain optimal instrument settings to gate and reduce background fluorescence, unstained and single-colour controls were included in the analyses. Cells were counted, enriched and sorted on a MoFlo Astrios EQ Cell Sorter (Beckman Coulter Life Science, USA) from the Flow Cytometry Facility at the University of Nottingham (Queen's Medical Centre, Nottingham). CD3-PE and CD14-PE/Cy7 were excited with a blue laser (488nm) and detected by a 576/21 filter and 795/70 filter respectively, CD19-BV421 was excited with a violet laser (405nm) and detected by 448/59 filter and CD56-APC was excited with a red laser (642nm) and detected by 671/30 filter. Gating strategies are represented in Figure 12 and Figure 15 (see *Results*). Data was analysed by Kaluza analysis software (Beckman Coulter Life Science, USA).

RNA extraction

RNA was extracted from purified granulocytes, platelets, T lymphocytes, B cells, monocytes, NK cells and other minor PBMC subpopulations using TRIzol® following manufacturer's instructions for cells grown in suspension except for an additional room-temperature incubation in

TRIzol for 10 minutes before adding chloroform. Samples were finally eluted in 15µL of RNase-free distilled water (dH₂O) and stored at -80°C. 500ng final concentration aliquots were prepared by diluting RNA in dH₂O. RNA quality was accessed using spectrophotometry with NanoDrop™ (ThermoFisher Scientific, UK) monitoring the outcomes from the A260/A280 and A260/A230 ratios.

RNA integrity assay

PBMC samples were split and resuspended in PBS/EDTA. Samples were kept at room-temperature and RNA was extracted at exponentially spaced time points – 0, 0.5, 1, 2, 4 and 8h – as described in previous studies¹⁰⁸. RNA yield was detected using a degradation analysis of the 28S and 18S rRNAs evaluating the areas of the band pattern after 1% agarose gel electrophoresis and describing high stability in RNA samples when values are constant (ratio of 2 : 1)¹⁰⁹.

cDNA syntheses

1000ng of RNA were reverse-transcribed by Moloney murine leukaemia virus reverse transcriptase using oligo-(dT) primers for *DEFA1A3*-specific PCR amplification or random hexamers primers for qRT-PCR. The procedure was adapted from Tetro™ cDNA Synthesis Kit protocol except that random hexamer primer samples were pre-incubated for 10 minutes at 25°C, the synthesis step was performed at 42°C for 60 minutes and the reverse transcriptase was inactivated at 68°C for 10 minutes. In all cDNA syntheses a control omitting reverse transcriptase was included.

DEFA1A3-specific PCR

cDNA was amplified using Taq DNA polymerase and the PCR profile consisted of an initial pre-denaturation at 95°C for 5 minutes, followed by 30-40 cycles (denaturation at 95°C for 30s, annealing at 58°C for 30s and extension at 72°C for 120s) and ended with an additional elongation step at 72°C for 7 minutes. Primers were *DEFA1A3* forward: 5'-ATCCTTGCTGCCATTCTCC-3', *DEFA1A3* reverse: 5'-TCCCTGTAGCTCTCAAAGCA-3', *SDHA* forward: 5'-GCAACAGAAGAAGCCCTTTG-3' and *SDHA* reverse: 5'-CAGCCACTAGGTGCCAATCT-3'. PCR products were examined after 1.5% agarose gel electrophoresis. Amplified genomic bands had a size of 898/902bp (4bp deletion) and 2133bp for *DEFA1A3* and *SDHA* whereas amplified cDNA bands were 322bp and 306bp respectively. A control without any input was included in all amplification experiments.

Quantitative real-time PCR (qRT-PCR)

cDNA was amplified by qRT-PCR with cycling parameters under standard conditions including an initial step at 50°C for 2 minutes, denaturation at 95°C for 10 minutes and ending with 40 cycles (denaturation at 95°C for 15s and annealing and extension at 60°C for 60s). Amplifications were monitored in real time by TaqMan *DEFA1A3* and *SDHA* probes, in which *SDHA* was used as the reference gene and amplified simultaneously with *DEFA1A3* in separate wells for each cDNA sample. Fluorescence results (Ct values) were analysed using the 7500 Fast software (ThermoFisher, UK) and the relative expression of *DEFA1A3* was calculated following the $2^{\Delta\Delta Ct}$ method as described in Rao et al.'s work⁹⁴. Briefly, qRT-PCR Ct values were normalised obtaining the initial ΔCt outcomes – *SDHA* Ct values were subtracted from the ones obtained from *DEFA1A3* – and their following $\Delta\Delta Ct$ results – ΔCt from granulocytes or PBMCs were subtracted from the ΔCt from the target

blood cell subset. A control without any input was included in all amplification experiments.

Immunocytochemistry

Alpha-defensin 1-3 peptide immunocytochemistry detection was performed according to previously published protocols¹³⁵ with minor modifications. Briefly, cytospin preparations were immunostained after fixation with 4% formaldehyde. Additionally, possible endogenous peroxidase activity was inactivated by a pre-incubation with 3% H₂O₂ in dH₂O for 15 minutes. The permeabilization step was performed with 0.5% Triton™ X-100 for 30 minutes and followed by a blocking step with human serum for 60 minutes. Both antibodies – mouse primary anti-human neutrophil peptide 1-3 (HM2058, Clone: D21, Hycult Biotech) and goat peroxidase conjugated secondary anti-mouse IgG (A4416-1mL, Clone: A4416, Sigma-Aldrich) – were applied following manufacturers' directions with dilutions of 1:100 and 1:200 respectively. Final staining was performed by 3,3'-diaminobenzidine (DAB) leaving a brown stain where the goat peroxidase conjugated secondary anti-mouse IgG was located. A control without the primary antibody was included in all experiments.

Sanger sequencing

PCR amplified products were cleaned by AMPure following the manufacturer's directions. Briefly, PCR-amplified products were mixed with AMPure beads that were collected using a 96R magnet plate (Alpaqua, USA) followed by two washes with 70% ethanol to remove unincorporated primers and dNTPs and resuspended in dH₂O. Clean samples (~50ng) were then used in two DNA Sanger sequencing reactions, following previously published protocols¹³⁶, one with each of the two primers with a profile consisting of 25 cycles (denaturation at

96°C for 30s, annealing at 50°C for 15s and extension at 60°C for 4 minutes). Sequenced products were then purified using CleanSeq, which eliminates unincorporated fluorescent terminators following manufacturer's instructions. Briefly, 5µL of the bead suspension were added to each sequencing reaction and were collected after ethanol precipitation using a 96R magnet plate, followed by two washes with 80% ethanol. DNA was coprecipitated with the dried beads, which were sent to the Deep Seq: Next Generation Sequencing Facility at the University of Nottingham (Centre for Genetics and Genomics at Queen's Medical Centre, Nottingham) and used in a capillary electrophoresis to obtain the electropherogram. Consensus sequences were obtained using MEGA-X (<http://www.megasoftware.net/>) and subsequently analysed with NCBI's sequence search tool (Advanced BLAST search, <http://www.ncbi.nlm.nih.gov/BLAST/>).

REFERENCES

1. *Atlas of Clinical Hematology*. (Current Medicine Group, 2008).
2. Janeway's Immunobiology | W. W. Norton & Company. Available at: <http://books.wwnorton.com/books/webad.aspx?id=4294997129>. (Accessed: 30th November 2018)
3. Wang, L. D. & Wagers, A. J. Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nat. Rev. Mol. Cell Biol.* **12**, 643–655 (2011).
4. Petrie, H. T., Hugo, P., Scollay, R. & Shortman, K. Lineage relationships and developmental kinetics of immature thymocytes: CD3, CD4, and CD8 acquisition in vivo and in vitro. *J. Exp. Med.* **172**, 1583–1588 (1990).
5. Surh, C. D. & Sprent, J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).
6. von Boehmer, H. Positive selection of lymphocytes. *Cell* **76**, 219–228 (1994).
7. Kincade, P. W. Molecular interactions between stromal cells and B lymphocyte precursors. *Semin. Immunol.* **3**, 379–390 (1991).
8. Dorshkind, K. & Landreth, K. S. Regulation of B cell differentiation by bone marrow stromal cells. *Int. J. Cell Cloning* **10**, 12–17 (1992).
9. Hamel, K. M., Mandal, M., Karki, S. & Clark, M. R. Balancing Proliferation with Igk Recombination during B-lymphopoiesis. *Front. Immunol.* **5**, 139 (2014).
10. Edry, E. & Melamed, D. Receptor editing in positive and negative selection of B lymphopoiesis. *J. Immunol. Baltim. Md 1950* **173**, 4265–4271 (2004).
11. Tornberg, U. C., Bergqvist, I., Haury, M. & Holmberg, D. Regulation of B lymphocyte development by the truncated immunoglobulin heavy chain protein Dmu. *J. Exp. Med.* **187**, 703–709 (1998).
12. Farag, S. S. & Caligiuri, M. A. Human natural killer cell development and biology. *Blood Rev.* **20**, 123–137 (2006).
13. Cooper, M. A., Fehniger, T. A. & Caligiuri, M. A. The biology of human natural killer-cell subsets. *Trends Immunol.* **22**, 633–640 (2001).
14. Freud, A. G. & Caligiuri, M. A. Human natural killer cell development. *Immunol. Rev.* **214**, 56–72 (2006).
15. Mace, E. M. & Orange, J. S. Multiple distinct NK-cell synapses. *Blood* **118**, 6475–6476 (2011).
16. Parker, D. C. T cell-dependent B cell activation. *Annu. Rev. Immunol.* **11**, 331–360 (1993).
17. O'Rourke, A. M. & Mescher, M. F. Cytotoxic T-lymphocyte activation involves a cascade of signalling and adhesion events. *Nature* **358**, 253–255 (1992).
18. Stavnezer, J. Immunoglobulin class switching. *Curr. Opin. Immunol.* **8**, 199–205 (1996).
19. Kautz, L. & Nemeth, E. Molecular liaisons between erythropoiesis and iron metabolism. *Blood* **124**, 479–482 (2014).
20. Kaushansky, K. Historical review: megakaryopoiesis and thrombopoiesis. *Blood* **111**, 981–986 (2008).
21. Haase, V. H. Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Rev.* **27**, 41–53 (2013).
22. Chasis, J. A. & Mohandas, N. Erythroblastic islands: niches for erythropoiesis. *Blood* **112**, 470–478 (2008).
23. Hartenstein, V. Blood cells and blood cell development in the animal kingdom. *Annu. Rev. Cell Dev. Biol.* **22**, 677–712 (2006).
24. Thomas, M. R. & Storey, R. F. The role of platelets in inflammation. *Thromb. Haemost.* **114**, 449–458 (2015).
25. Franco, A. T., Corken, A. & Ware, J. Platelets at the interface of thrombosis, inflammation, and cancer. *Blood* **126**, 582–588 (2015).
26. Nakahata, T. & Okumura, N. Cell surface antigen expression in human erythroid progenitors: erythroid and megakaryocytic markers. *Leuk. Lymphoma* **13**, 401–409 (1994).
27. Italiano, J. E., Lecine, P., Shivdasani, R. A. & Hartwig, J. H. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. *J. Cell Biol.* **147**, 1299–1312 (1999).

28. Richardson, J. L., Shivdasani, R. A., Boers, C., Hartwig, J. H. & Italiano, J. E. Mechanisms of organelle transport and capture along proplatelets during platelet production. *Blood* **106**, 4066–4075 (2005).
29. Arinobu, Y., Iwasaki, H. & Akashi, K. Origin of basophils and mast cells. *Allergol. Int. Off. J. Jpn. Soc. Allergol.* **58**, 21–28 (2009).
30. Xia, L. *et al.* Eosinophil differentiation in the bone marrow is promoted by protein tyrosine phosphatase SHP2. *Cell Death Dis.* **7**, e2175 (2016).
31. Basu, S., Hodgson, G., Katz, M. & Dunn, A. R. Evaluation of role of G-CSF in the production, survival, and release of neutrophils from bone marrow into circulation. *Blood* **100**, 854–861 (2002).
32. Summers, C. *et al.* Neutrophil kinetics in health and disease. *Trends Immunol.* **31**, 318–324 (2010).
33. Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science* **303**, 1532–1535 (2004).
34. Patel, A. A. *et al.* The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).
35. Terry, R. L. & Miller, S. D. Molecular control of monocyte development. *Cell. Immunol.* **291**, 16–21 (2014).
36. Jakubzick, C. V., Randolph, G. J. & Henson, P. M. Monocyte differentiation and antigen-presenting functions. *Nat. Rev. Immunol.* **17**, 349–362 (2017).
37. Blausen Medical: The source and creator of the world's large... Available at: <https://blausen.com/>. (Accessed: 30th June 2019)
38. eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* **49**, 1664–1670 (2017).
39. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
40. Gallego Romero, I., Pai, A. A., Tung, J. & Gilad, Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* **12**, 42 (2014).
41. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, (2017).
42. UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **35**, D193-197 (2007).
43. Fernández, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst.* **3**, 491-495.e5 (2016).
44. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
45. Spits, C. *et al.* Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.* **1**, 1965–1970 (2006).
46. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780–791 (2016).
47. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, (2017).
48. Kunicki, M. A., Amaya Hernandez, L. C., Davis, K. L., Bacchetta, R. & Roncarolo, M.-G. Identity and Diversity of Human Peripheral Th and T Regulatory Cells Defined by Single-Cell Mass Cytometry. *J. Immunol. Baltim. Md 1950* **200**, 336–346 (2018).
49. van Unen, V. *et al.* Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets. *Immunity* **44**, 1227–1239 (2016).
50. Crinier, A. *et al.* High-Dimensional Single-Cell Analysis Identifies Organ-Specific Signatures and Conserved NK Cell Subsets in Humans and Mice. *Immunity* **49**, 971-986.e5 (2018).
51. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
52. Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S. & Salim, A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* **28**, 2711–2718 (2012).
53. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
54. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
55. Inoue, K. & Lupski, J. R. Molecular Mechanisms for Genomic Disorders. *Annu. Rev. Genomics Hum. Genet.* **3**, 199–242 (2002).

56. Hoang, M. L. *et al.* Competitive Repair by Naturally Dispersed Repetitive DNA during Non-Allelic Homologous Recombination. *PLOS Genet.* **6**, e1001228 (2010).
57. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat. Rev. Mol. Cell Biol.* **7**, 739–750 (2006).
58. Rass, U. *et al.* Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev.* **24**, 1559–1569 (2010).
59. Lee, A. H., Symington, L. S. & Fidock, D. A. DNA repair mechanisms and their biological roles in the malaria parasite *Plasmodium falciparum*. *Microbiol. Mol. Biol. Rev. MMBR* **78**, 469–486 (2014).
60. Linzmeier, R. M. & Ganz, T. Human defensin gene copy number polymorphisms: Comprehensive analysis of independent variation in α - and β -defensin regions at 8p22–p23. *Genomics* **86**, 423–430 (2005).
61. Hollox, E. J., Barber, J. C. K., Brookes, A. J. & Armour, J. A. L. Defensins and the dynamic genome: What we can learn from structural variation at human chromosome band 8p23.1. *Genome Res.* **18**, 1686–1697 (2008).
62. Khan, F. F. *et al.* Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. *BMC Genomics* **14**, 719 (2013).
63. Aldred, P. M. R., Hollox, E. J. & Armour, J. A. L. Copy number polymorphism and expression level variation of the human α -defensin genes DEFA1 and DEFA3. *Hum. Mol. Genet.* **14**, 2045–2052 (2005).
64. Cheng, F., Zhou, X., Zhao, Y., Zhao, M. & Zhang, H. Alpha-defensin DEFA1A3 gene copy number variation in Asians and its genetic association study in Chinese systemic lupus erythematosus patients. *Gene* **517**, 158–163 (2013).
65. Ballana, E., González, J. R., Bosch, N. & Estivill, X. Inter-population variability of DEFA3 gene absence: correlation with haplotype structure and population variability. *BMC Genomics* **8**, 14 (2007).
66. Tsutsumi-Ishii, Y., Hasebe, T. & Nagaoka, I. Role of CCAAT/enhancer-binding protein site in transcription of human neutrophil peptide-1 and -3 defensin genes. *J. Immunol. Baltim. Md 1950* **164**, 3264–3273 (2000).
67. Black, H. A., Khan, F. F., Tyson, J. & Al Armour, J. Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. *BMC Genomics* **15**, 614 (2014).
68. Ganz, T. Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* **3**, 710–720 (2003).
69. Yang, D., Biragyn, A., Hoover, D. M., Lubkowski, J. & Oppenheim, J. J. Multiple roles of antimicrobial defensins, cathelicidins, and eosinophil-derived neurotoxin in host defense. *Annu. Rev. Immunol.* **22**, 181–215 (2004).
70. Ai, Z. *et al.* Low α -defensin gene copy number increases the risk for IgA nephropathy and renal dysfunction. *Sci. Transl. Med.* **8**, 345ra88 (2016).
71. Klotman, M. E. & Chang, T. L. Defensins in innate antiviral immunity. *Nat. Rev. Immunol.* **6**, 447 (2006).
72. Sarma, J. V. & Ward, P. A. The complement system. *Cell Tissue Res.* **343**, 227–235 (2011).
73. Territo, M. C., Ganz, T., Selsted, M. E. & Lehrer, R. Monocyte-chemotactic activity of defensins from human neutrophils. *J. Clin. Invest.* **84**, 2017–2020 (1989).
74. Yang, D., Chen, Q., Chertov, O. & Oppenheim, J. J. Human neutrophil defensins selectively chemoattract naive T and immature dendritic cells. *J. Leukoc. Biol.* **68**, 9–14 (2000).
75. Oppenheim, J. J., Biragyn, A., Kwak, L. W. & Yang, D. Roles of antimicrobial peptides such as defensins in innate and adaptive immunity. *Ann. Rheum. Dis.* **62**, ii17–ii21 (2003).
76. Brook, M. *et al.* Neutrophil-derived alpha defensins control inflammation by inhibiting macrophage mRNA translation. *Proc. Natl. Acad. Sci.* **113**, 4350–4355 (2016).
77. Sthoeger, Z. M., Bezalel, S., Chapnik, N., Asher, I. & Froy, O. High α -defensin levels in patients with systemic lupus erythematosus. *Immunology* **127**, 116–122 (2009).
78. Tamiya, H. *et al.* Defensins- and cathepsin G-ANCA in systemic lupus erythematosus. *Rheumatol. Int.* **27**, 147–152 (2006).
79. Yu, X.-Q. *et al.* A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat. Genet.* **44**, 178–182 (2011).

80. Li, M. *et al.* Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nat. Commun.* **6**, 7270 (2015).
81. Qi, Y. Y. *et al.* DEFA gene variants associated with IgA nephropathy in a Chinese population. *Genes Immun.* **16**, 231–237 (2015).
82. Qi, Y.-Y., Zhou, X.-J., Cheng, F.-J. & Zhang, H. Elevated Plasma α -Defensins (HNP1-3) Levels Correlated with IgA1 Glycosylation and Susceptibility to IgA Nephropathy. *Dis. Markers* **2016**, 8123138 (2016).
83. Teumer, A. *et al.* Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.* **40**, 977–985 (2013).
84. Dommisch, H. & Jepsen, S. Diverse functions of defensins and other antimicrobial peptides in periodontal tissues. *Periodontol. 2000* **69**, 96–110 (2015).
85. Munz, M. *et al.* A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. *Hum. Mol. Genet.* **27**, 941–942 (2018).
86. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).
87. EventPilot Web. Available at: <https://eventpilot.us/web/page.php?page=IntHtml&project=ASHG18&id=180122185>. (Accessed: 31st May 2019)
88. Cerrillo, E. *et al.* Alpha-defensins (α -Defs) in Crohn's disease: decrease of ileal α -Def 5 via permanent methylation and increase in plasma α -Def 1-3 concentrations offering biomarker utility. *Clin. Exp. Immunol.* **192**, 120–128 (2018).
89. Chen, Q. *et al.* Increased gene copy number of DEFA1/DEFA3 worsens sepsis by inducing endothelial pyroptosis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 3161–3170 (2019).
90. Haley, K. M., Loren, C. P., Phillips, K. G. & McCarty, O. J. T. Characterization of Single Platelet Mass, Volume, and Density in Response to Agonist Stimulation. *Blood* **118**, 5261–5261 (2011).
91. Zsóri, K., Muszbek, L., Csiki, Z. & Shemirani, A. Validation of Reference Genes for the Determination of Platelet Transcript Level in Healthy Individuals and in Patients with the History of Myocardial Infarction. *Int. J. Mol. Sci.* **14**, 3456–3466 (2013).
92. Valore, E. V. & Ganz, T. Posttranslational processing of defensins in immature human myeloid cells. *Blood* **79**, 1538–1544 (1992).
93. Etulain, J. *et al.* An optimised protocol for platelet-rich plasma preparation to improve its angiogenic and regenerative properties. *Sci. Rep.* **8**, 1513 (2018).
94. Rao, X., Huang, X., Zhou, Z. & Lin, X. An improvement of the $2^{-\Delta\Delta CT}$ method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinforma. Biomath.* **3**, 71–85 (2013).
95. Ratts, R. B. & Weng, N.-P. Homeostasis of lymphocytes and monocytes in frequent blood donors. *Front. Immunol.* **3**, 271 (2012).
96. Cowland, J. B. & Borregaard, N. The individual regulation of granule protein mRNA levels during neutrophil maturation explains the heterogeneity of neutrophil granules. *J. Leukoc. Biol.* **66**, 989–995 (1999).
97. Faurschou, M. & Borregaard, N. Neutrophil granules and secretory vesicles in inflammation. *Microbes Infect.* **5**, 1317–1327 (2003).
98. Banfi, G., Salvagno, G. L. & Lippi, G. The role of ethylenediamine tetraacetic acid (EDTA) as in vitro anticoagulant for diagnostic purposes. *Clin. Chem. Lab. Med.* **45**, 565–576 (2007).
99. Krisanaprakornkit, S., Jotikasthira, D. & Dale, B. A. Intracellular calcium in signaling human beta-defensin-2 expression in oral epithelial cells. *J. Dent. Res.* **82**, 877–882 (2003).
100. Bell, L., Halstenson, C. E., Halstenson, C. J., Macres, M. & Keane, W. F. Cholesterol-lowering effects of calcium carbonate in patients with mild to moderate hypercholesterolemia. *Arch. Intern. Med.* **152**, 2441–2444 (1992).
101. Li, Y.-X., Lin, C.-Q., Shi, D.-Y., Zeng, S.-Y. & Li, W.-S. Upregulated expression of human alpha-defensins 1, 2 and 3 in hypercholesterolemia and its relationship with serum lipid levels. *Hum. Immunol.* **75**, 1104–1109 (2014).

102. Mann, K. G., Whelihan, M. F., Butenas, S. & Orfeo, T. Citrate anticoagulation and the dynamics of thrombin generation. *J. Thromb. Haemost. JTH* **5**, 2055–2061 (2007).
103. Björk, I. & Lindahl, U. Mechanism of the anticoagulant action of heparin. *Mol. Cell. Biochem.* **48**, 161–182 (1982).
104. English, D. & Andersen, B. R. Single-step separation of red blood cells, granulocytes and mononuclear leukocytes on discontinuous density gradients of Ficoll-Hypaque. *J. Immunol. Methods* **5**, 249–252 (1974).
105. Ferrante, A. & Thong, Y. H. Optimal conditions for simultaneous purification of mononuclear and polymorphonuclear leucocytes from human blood by the Hypaque-Ficoll method. *J. Immunol. Methods* **36**, 109–117 (1980).
106. Basu, S., Campbell, H. M., Dittel, B. N. & Ray, A. Purification of Specific Cell Population by Fluorescence Activated Cell Sorting (FACS). *J. Vis. Exp. JoVE* (2010). doi:10.3791/1546
107. Cossarizza, A. *et al.* Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur. J. Immunol.* **47**, 1584–1797 (2017).
108. Cheneval, D., Kastelic, T., Fuerst, P. & Parker, C. N. A review of methods to monitor the modulation of mRNA stability: a novel approach to drug discovery and therapeutic intervention. *J. Biomol. Screen.* **15**, 609–622 (2010).
109. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
110. Ross, J. mRNA stability in mammalian cells. *Microbiol. Rev.* **59**, 423–450 (1995).
111. Bustin, S. A. & Nolan, T. Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction. *J. Biomol. Tech. JBT* **15**, 155–166 (2004).
112. Shaw, A. E. *et al.* Development and initial evaluation of a real-time RT-PCR assay to detect bluetongue virus genome segment 1. *J. Virol. Methods* **145**, 115–126 (2007).
113. Ledderose, C., Heyn, J., Limbeck, E. & Kreth, S. Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. *BMC Res. Notes* **4**, 427 (2011).
114. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. *J. Appl. Genet.* **54**, 391–406 (2013).
115. Stamova, B. S. *et al.* Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. *BMC Med. Genomics* **2**, 49 (2009).
116. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
117. Costa, V., Angelini, C., De Feis, I. & Ciccodicola, A. Uncovering the Complexity of Transcriptomes with RNA-Seq. *BioMed Research International* (2010). doi:10.1155/2010/853916
118. Ramos-Vara, J. A. Technical aspects of immunohistochemistry. *Vet. Pathol.* **42**, 405–426 (2005).
119. de Matos, L. L., Trufelli, D. C., de Matos, M. G. L. & da Silva Pinhal, M. A. Immunohistochemistry as an Important Tool in Biomarkers Detection and Clinical Practice. *Biomark. Insights* **5**, 9–20 (2010).
120. Porter, E., Valore, E. V., Anouseyan, R. & Salzman, N. H. Detection of antimicrobial (poly)peptides with acid urea polyacrylamide gel electrophoresis followed by Western immunoblot. *Methods Mol. Biol. Clifton NJ* **1225**, 105–115 (2015).
121. Dunphy, C. H. Applications of flow cytometry and immunohistochemistry to diagnostic hematopathology. *Arch. Pathol. Lab. Med.* **128**, 1004–1022 (2004).
122. Wilson, I. D. *et al.* HPLC-MS-based methods for the study of metabonomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **817**, 67–76 (2005).
123. Mora-Jensen, H. *et al.* Technical Advance: Immunophenotypical characterization of human neutrophil differentiation. *J. Leukoc. Biol.* **90**, 629–634 (2011).
124. Cherian, S. *et al.* Evaluation of an 8-color flow cytometric reference method for white blood cell differential enumeration. *Cytometry B Clin. Cytom.* **78B**, 319–328 (2010).
125. Valle-Jiménez, X. *et al.* Human platelets and megakaryocytes express defensin alpha 1. *Platelets* **1–11** (2019). doi:10.1080/09537104.2019.1615612
126. Agerberth, B. *et al.* The human antimicrobial and chemotactic peptides LL-37 and α -defensins are expressed by specific lymphocyte and monocyte populations. *Blood* **96**, 3086–3093 (2000).

127. Obata-Onai, A. *et al.* Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes. *Int. Immunol.* **14**, 1085–1098 (2002).
128. Chalifour, A. *et al.* Direct bacterial protein PAMP recognition by human NK cells involves TLRs and triggers alpha-defensin production. *Blood* **104**, 1778–1783 (2004).
129. Rodríguez-García, M. *et al.* Human immature monocyte-derived dendritic cells produce and secrete alpha-defensins 1-3. *J. Leukoc. Biol.* **82**, 1143–1146 (2007).
130. Szabo, P. A. *et al.* A single-cell reference map for human blood and tissue T cell activation reveals functional states in health and disease. *bioRxiv* 555557 (2019). doi:10.1101/555557
131. Orsini, G. *et al.* Enumeration of human peripheral blood dendritic cells throughout the life. *Int. Immunol.* **24**, 347–356 (2012).
132. Wu, W.-C. *et al.* Circulating hematopoietic stem and progenitor cells are myeloid-biased in cancer patients. *Proc. Natl. Acad. Sci.* 201320753 (2014). doi:10.1073/pnas.1320753111
133. Armour, J. A. L. *et al.* Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. *Nucleic Acids Res.* **35**, e19 (2007).
134. Wouters, K. *et al.* Circulating classical monocytes are associated with CD11c+ macrophages in human visceral adipose tissue. *Sci. Rep.* **7**, (2017).
135. Holterman, D. A. *et al.* Overexpression of α -defensin is associated with bladder cancer invasiveness. *Urol. Oncol. Semin. Orig. Investig.* **24**, 97–108 (2006).
136. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).