ANALYSIS OF GENETIC FACTORS INFLUENCING ALPHA-DEFENSIN GENE EXPRESSION

Paula Eugenia Necsoiu, BSc
Student ID Number: 4289612

THESIS SUBMITTED TO THE
UNIVERSITY OF NOTTINGHAM
FOR THE DEGREE OF MASTER BY RESEARCH
JANUARY 2018

Table of Contents

Acknowledgements

I would like to thank Professor John Armour for the opportunity of achieving my MRes in his research group. I am extremely grateful for his encouragement, advice and patient guidance throughout the last year. I would also like to thank everyone in the JALA lab for their support, especially Nzar Shwan who helped me collect blood in John's absence.

I must express my gratitude to my former supervisors – Reader Dan Cimponeriu, Assistant Professor Alina-Maria Holban and Scientific Researcher Consuel-Severus Ionică - for their constant help during my Bachelor Degree project and the encouragement to pursue this MRes at the University of Nottingham.

Finally, I would like to thank Eric Moise and his mother, Violeta, for their support in the last year; without them, starting this MRes would have not been possible.

ABSTRACT

Background

The *DEFA1A3* locus on chromosome 8p23.1 has been known to exhibit copy number variation (between 3 and 16 copies) and to include either the *DEFA1* gene or the *DEFA3* gene in any repeat. These two genes encode the Human Neutrophil Peptides HNPs 1-3, antimicrobial peptides implicated in the innate immune response.

Objectives

The primary aim of this study was to determine whether there is any significant relationship between alpha defensin copy number and expression levels. The secondary aim was to understand more about 6836028G, a synonymous variant in exon 2 of the *DEFA1A3* locus which affects a valine codon.

Results and Conclusions

Levels of *DEFA3* expression in mRNA are not simply proportional to genomic copy number; in cDNA *DEFA3* appears to be expressed about 6-7 times more per copy than *DEFA1*. A similar relationship is also reflected in the variant 6836028G, which is underrepresented by a factor of about 1.4-fold in cDNA relative to genomic DNA; we assume that this variant might accompany only the *DEFA1* gene, due to low levels of 6836028G in cDNA.

LIST OF ABBREVIATIONS

ACK – Ammonium Chloride Potassium

AgP – Aggressive Periodontitis

BAC – Bacterial Artificial Chromosome

BAT – Basophils Activation Test

BMI – Body Mass index

BPI – Bactericidal/Permeability Increasing Protein

CD – Cluster of Differentiation Molecule

CGH – Comparative Genome Hybridisation

CNV – Copy Number Variation

CR – Complement Receptor

ddH_2O – Double Distilled H_2O

DEPC – Diethylpyrocarbonate

EDTA – Ethylenediaminetetraacetic Acid

FISH – Fluorescence in-situ Hybridisation

fMLP receptors – Formylated Bacterial Peptide Receptors

GWAS – Genome Wide Association Study

hCAP – Human Cathelicidin Antimicrobial Protein

HD – Human Defensin

HiDi formamide – Highly Deionized Formamide

HNP – Numan Neutrophil Peptide

IgAN - Immunoglobulin A Nephropathy

indel - Insertion/Deletion Polymorphism

LD – Linkage Disequilibrium

LPS - Lipopolysaccharide

Mac-1 – Macrophage 1 Antigen

MAPH – Multiplex Amplifiable Probe Hybridisation

MLCN – Maximum Likelihood Copy Number

MLPA - Multiplex Ligation Dependent Probe Amplification

MMP – Matrix Metallopeptidase

MPO – Myeloperoxidase

NADPH oxidase – Nicotinamide Adenine Dinucleotide Phosphate Oxidase

NE – Neutrophil Elastase

NET – Neutrophil Extracellular Trap

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PRT – Paralogue Ratio Test

qPCR – Quantitative Real Time PCR

RBC – Red Blood Cells

RFLP – Restriction Fragment Length Polymorphism

RT – Reverse Transcriptase

SDS – Sodium Dodecyl Sulfate

SDW – Sterile Distilled Water

SLE – Systemic Lupus Erythematosus

SLPI – Secretory Leukocyte Protease Inhibitor

SNP – Single Nucleotide Polymorphism

WBC – White Blood Cells

1. INTRODUCTION

1.1 Neutrophils

Whole blood in humans is composed of a pale, yellow liquid called plasma, which contains water and various organic and inorganic matter, and a cellular fraction which consists of three types of cells: red blood cells (erythrocytes), white blood cells (leukocytes) and platelets (thrombocytes).

The erythrocytes are the most common cells of human blood (about 5 x 10^6 cells per µl, constituting approximately 40% of the total blood volume); they lack a nucleus and most organelles (Figure 1.1). Their main function is to transport oxygen to tissues. These cells are continuously produced through erythropoiesis in the red bone marrow (or the liver in embryos) and have an average lifespan of 115 days (Franco 2012).

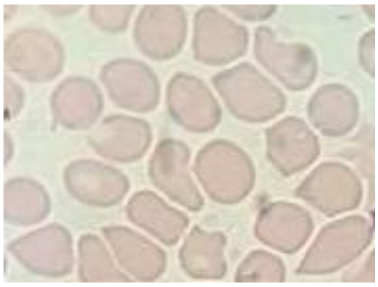


Figure 1.1 - Left: A three-dimensional image of red blood cells (courtesy of the Medical Gallery of Blausen Medical, 2014). Right: Red blood cells on a Leishman-stained blood smear (Paula Necsoiu)

The thrombocytes are numerous (1.5-4 x 10^5 per µl) cytoplasmic fragments derived from megakaryocytes (Machlus et al. 2014); their main role is to stop bleeding by forming clots at the site of an endothelium injury (Laki 1972).

Leukocytes (white blood cells) are part of the immune system, found throughout the body, not just in the blood but also in the lymphatic system (Maton 1997), defending the organism against infections and invaders. The total white cell count varies between individuals (between 5 and 10 x 10^3 cells per µl). The contribution of different white cell classes to the total is also variable, but typically about 60% of all leukocytes are neutrophils, about 30% lymphocytes, 7% monocytes and 1-2% eosinophils and basophils. White blood cells are derived from pluripotent hematopoietic stem cells formed in the bone marrow and can be classified either by structure (granulocytes or agranulocytes) or by lineage (myeloid or lymphoid). Under the influence of various growth factors, the myeloid and lymphoid progenitors evolve through a series of intermediate steps into mature blood cells (Figure 1.2). Based on cell structure, white blood cells are classified as granulocytes (polymorphonuclear leukocytes) or agranulocytes (mononuclear leukocytes) (LaFleur-Brooks 2008).

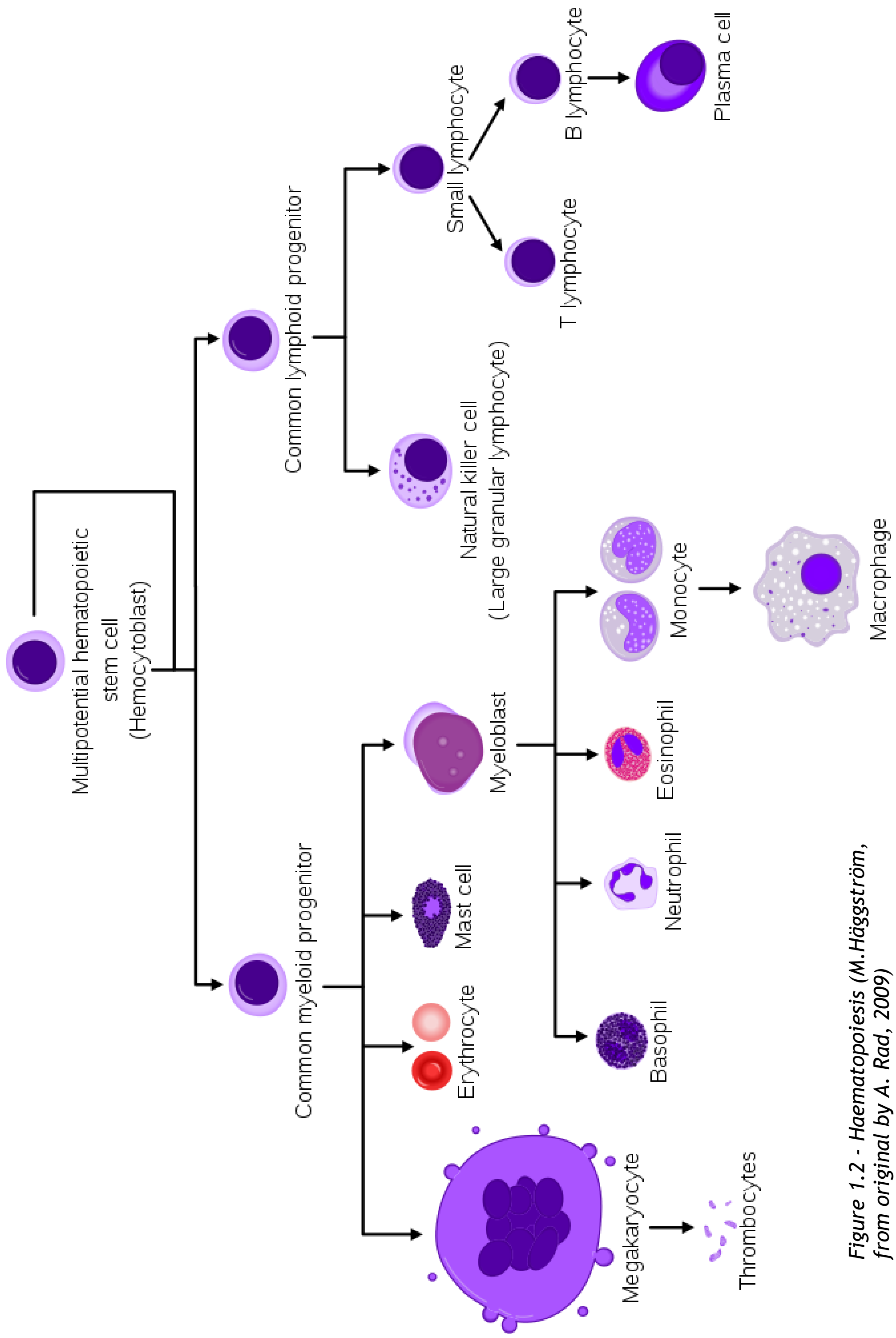
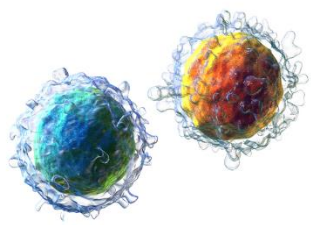


Figure 1.2 - Haematopoiesis (M.Häggström, from original by A. Rad, 2009)

Mononuclear cells are characterised by the shape of the nucleus (one-lobed nucleus) and the absence of granules in their cytoplasm; these can be further classified into lymphocytes and monocytes. Lymphocytes (Figure 1.3) develop from a common lymphoid progenitor cell; following certain pathways they differentiate into three major types: Natural Killer cells, B cells (bursa / bone marrow cells) or T cells (thymus cells). They play a major role in the immune system: T cells participate in cell-mediated immunity, B cells are involved in humoral immunity (antibody production), whereas NK cells are responsible for defending the organism against infected cells and tumours, being a part of the innate immune system (Blumenreich 1990). The hematopoietic stem cells differentiate into the bipotent monoblast cells, which further differentiate into monocytes (Figure 1.4). These can exit and re-enter circulation while retaining their function: in tissues, they are known as macrophages. The monocyte-macrophage system cells phagocytose bacteria, playing an important role in the inflammatory reaction by secreting interleukin, which amplifies the B cell and T cell response (Blumenreich 1990).



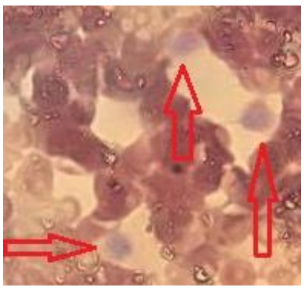


Figure 1.3 – Left: 3D illustration of lymphocytes (Medical Gallery of Blausen Medical, 2014). Right: Lymphocytes on a Leishman-stained blood smear (Paula Necsoiu)

Figure 1.4 - Left: 3D illustration of a monocyte (Medical Gallery of Blausen Medical, 2014). Right: Monocyte on a Leishman-stained blood smear (Paula Necsoiu)

Polymorphonuclear cells or granulocytes owe their name to the presence of multi-lobed nuclei and cytoplasmic granules. These cells can be further classified into eosinophils, basophils and neutrophils. Eosinophils (Figure 1.5) are released from the bone marrow into the circulation, making their way into tissues; unlike monocytes, these cells are not permitted to re-enter the blood stream (Uhm et al. 2012). Recent studies have shown that eosinophils might be involved in inhibiting inflammation and in the defence against multicellular parasites (Rothenberg and Hogan 2006; Shi 2004). Basophils (Figure 1.5) have a similar development and function (Mukai et al. 2013; Khurana 2009); flow-cytometry studies involving Basophils Activation Tests (BAT) demonstrate that, when activated, basophils degranulate to release heparin and histamine, being found in large numbers at sites of parasitic infection or allergies (Böhm et al. 2011).

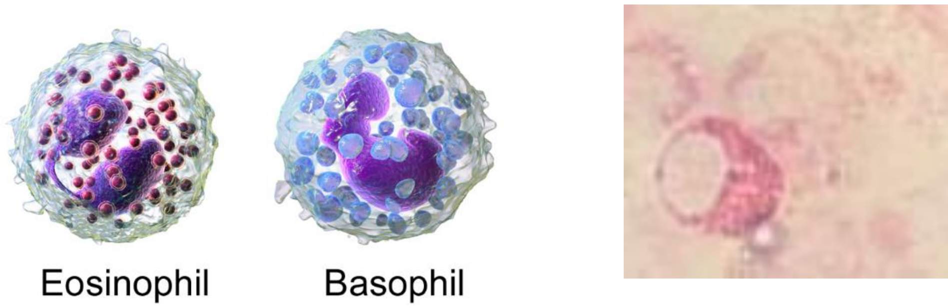


Figure 1.5 - Left: 3D illustration of an eosinophil and a basophil (Medical Gallery of Blausen Medical, 2014). Right: Eosinophil on a Leishman-stained blood smear (Paula Necsoiu)

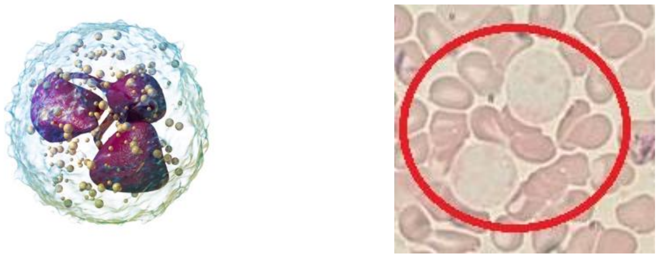


Figure 1.6 - Left: 3D illustration of a neutrophil (Medical Gallery of Blausen Medical, 2014). Right: Neutrophils on a Leishman-stained blood smear (Paula Necsoiu)

In the development of neutrophils (Figure 1.2, Figure 1.6), the first recognisable cell is the myeloblast; as maturation progresses, the myeloblast differentiates into a promyelocyte and later into a myelocyte. Up until this point, the cell growth is exponential; reaching the metamyelocyte stage, the cells are unable to undergo further mitosis but transform into a band cell, which will follow one of two routes: it is either released into circulation where it completes its maturation process (3-5% of WBC) or remains in the bone marrow becoming a segmented neutrophil, later being released in the circulation (Blumenreich 1990). Only 50% of the intravascular neutrophils are circulating, being accounted for in the WBC count. The life-span of a mature neutrophil in circulation has been reported to be, on average, between 5 and 90 hours (Tak et al. 2013), while in the tissues they will enter an apoptotic phase after 1 or 2 days (Stevens et al. 2002).

Being highly motile, neutrophils migrate in a coordinated manner and accumulate at a site of infection, being attracted by cytokines (interleukin-8, C3a, C5a) expressed by activated macrophages; the cycle continues with neutrophils activating and recruiting macrophages to the site of inflammation. Neutrophils express and release cytokines which amplify inflammatory reactions (Ear and McDonald 2008). In addition to the recruitment and activation of other cells, neutrophils are considered to be much more than just “killers” in the front line of defence against pathogens, directly attacking the invaders either by phagocytosis, generation of neutrophil extracellular traps or degranulation (Hickey and Kubes 2009).

First and foremost, neutrophils are highly capable phagocytes, capable of ingesting and killing a wide range of opsonised microorganisms and pathogens via two intracellular pathways: oxygen-dependent or oxygen-independent. The oxygen-dependent or the “respiratory burst” phase involves the formation of a phagosome into which hydrolytic enzymes and reactive oxygen-containing molecules are secreted, with high consumption of oxygen. The oxygen-independent methods are slower than and not as effective as the oxygen-dependent ones; the pathogens are destroyed by charged proteins that damage the membrane, enzymes that break down the invader cell walls (lysozymes), digest invader-specific proteins (proteases and hydrolytic enzymes) and remove essential iron from the bacteria (lactoferrins) (Mihăescu 2001).

Neutrophils produce and release many anti-inflammatory and pro-inflammatory agents often through degranulation, a process that helps to combat infection. There are four main categories of cytoplasmic granules (primary or azurophilic, secondary or specific, tertiary or gelatinase and secretory vesicles) the content of each type having a different role either in cell migration, haematopoiesis, wound healing, tumour metastasis or angiogenesis (Anderson et al. 2008).

The primary or azurophilic granules are found in young neutrophils and contain a wide variety of proteins with anti-microbial properties: NE (neutrophil elastase), MPO (myeloperoxidase), defensins, lysozyme, cathepsin G, BPI (bactericidal/permeability increasing protein), Proteinase 3, azurodicin and vitronecin. The secondary or specific granules are found in mature neutrophils and store a mixture of enzymes and antimicrobial peptides: lactoferrin, NADPH oxidase (nicotinamide adenine dinucleotide phosphate-oxidase), MMP-1 (matrix metallopeptidase 1), MMP-8 (matrix metallopeptidase 1), lysozyme, M-ficolin, neutrophil gelatinase associated lipocalin, hCAP-18 (human cathelicidin antimicrobial protein 18), flavocytochrome b558, SLPI (secretory leukocyte protease inhibitor), Pentraxin 3 and leukolysin. The tertiary or gelatinase granules contain MMP-2 (matrix metallopeptidase 2), MMP-9 (matrix metallopeptidase 9), flavocytochrome b558, arginase-1 and leukolysin. The secretory vesicles consist of β2-integrin CD11b/CD18 (Mac-1, CR3) and formylated bacterial peptide receptors (fMLP-receptors). These peptides together play a key role in both innate and active adaptive immunity by disrupting the cytoplasmic membrane of pathogens, producing toxins and oxidants, inducing the activation of other cells in order to combat infection, promoting or inhibiting apoptosis of neutrophils, tissue remodelling and bactericidal activity (Chistiakov et al. 2015, Yang et al. 2000, Faurschou and Borregaard 2003, Eiserich et al. 1998, Bae et al. 2012, Walzog et al. 1997, Anderson et al. 2008, Selders et al. 2017).

The third mechanism involved in killing pathogens is the release of web-like structures of DNA when neutrophils are activated, known as neutrophil extracellular traps (NETs). Brinkmann and colleagues first observed and described in 2004 that neutrophils can release NETs in two ways. The first pathway is less common and it involves the release of mitochondrial DNA, so that the cell remains intact (Anderson et al. 2008; Kumar et al. 2010; Yousefi et al. 2009). The second pathway refers to NETosis (Figure 1.7), a different form of apoptosis which refers to chromatin decondensation, cell disintegration and the release of NETs coated in antimicrobial peptides (Chistiakov et al. 2015; Smith 1994; Bjornsdottir et al. 2015, Brinkmann and Zychlinsky 2012, Branzk et al. 2014; Fuchs et al. 2007).

This structure is almost 15 times the volume of the neutrophil and traps the pathogen while minimizing damage to the surrounding tissue (Papayannopoulos and Zychlinsky 2009).

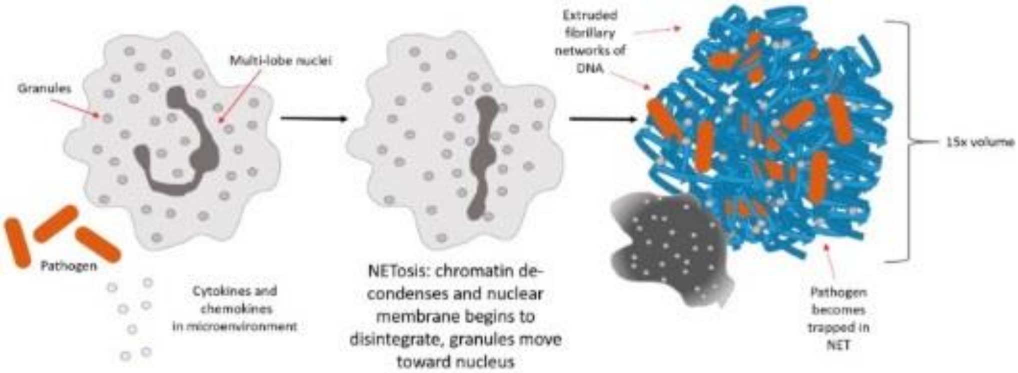


Figure 1.7 - Neutrophil undergoing NETosis (courtesy of Selders et al. 2017)

1.2 Alpha Defensin Variation

Defensins are small cysteine-rich cationic antimicrobial peptides found in plants, invertebrates and vertebrates. They are known to be active during the immune response, with a role to combat infections by disrupting the cytoplasmic membrane of microbes, inducing migration of naïve T cells and immature dendritic cells (Yang et al. 2000) and inducing chemotaxis of CD4+ and CD8+ cells (Faurschou et al. 2003). Defensins are divided into three major classes (alpha, beta, theta), with alpha and beta being highly expressed in humans.

Theta defensins are small peptides (18 amino acids) that form a single hairpin structure. Until now they have been found only in the white cells of rhesus macaque (Tran et al. 2007) and the baboon *Papio anubis* (Garcia et al. 2008). In humans and other primates, the only theta defensin gene seems to be an expressed pseudogene named *DEFT1P*: the mature peptide is prevented from being produced due to a stop codon in the signal sequence (Aldred, 2004). *DEFT1P* is found on 8p23.1 within the copy-variable alpha defensin cluster.

Beta defensins are widely distributed, being expressed mainly in airways, epididymis epithelia and leukocytes. It has been reported that they present broad antimicrobial properties and are induced upon pathogen invasion. Beta defensin genes form a major copy-variable cluster on 8p23.1 (Linzmeier et al. 1999), and clusters of presumed beta defensin genes have been identified on 6p, 20p and 20q (Schutte et al. 2002). Evolution by duplication followed by adaptive evolution is suggested by the common ancestor each cluster seems to have (Hughes 1999; Morrison et al. 2003).

There are five human alpha defensin genes (*DEFA1*, *DEFA3*, *DEFA4*, *DEFA5*, *DEFA6*) which form a cluster on 8p23.1 (Linzmeier et al. 1999) and encode six proteins (HNP1, HNP2, HNP3, HNP4, HD5, HD6). Human Neutrophil Peptide 2 (HNP2) does not have a corresponding gene, but it is believed to be a cleavage product of peptides encoded by *DEFA1* and/or *DEFA3* (Mars et al. 1995). HNP-1, 2 and 3 are almost identical peptides which differ by the first amino-acid: alanine in HNP-1, aspartic acid in HNP-3 and absent in HNP-2 (Ganz et al 1985). *DEFA1*, *DEFA3* and *DEFA4* encode the human neutrophil peptides HNP1-4 (Ganz et al. 1985); cleaved prodefensins are expressed in the promyelocytes (Valore and Ganz 1992) and stored in the primary azurophilic granules of neutrophils (Rice et al. 1987) whereas in myelocytes they are expressed as the 75 amino-acid uncleaved form (Faurschou and Borregaard 2003). Circulating neutrophils have high levels of stored alpha defensins, but cannot synthesize new defensin peptides, with synthesis restricted to neutrophil precursors, particularly promyelocytes (Valore and Ganz 1992). *DEFA5* and *DEFA6* are expressed in the Paneth cells of the ileum (last section of the small intestine). The synthesised products (HD5, HD6) are stored as prodefensins (precursor form) in the cells and processed after secretion (Cunliffe et al. 2001).

The main function of all defensin classes is antimicrobial activity achieved by disrupting the cytoplasmic membrane of pathogens (Yang et al. 2000). Their mode of action is to form and attach hexamers of defensin dimers to the microbial membrane, creating voltage dependent ion channels (Kagan et al. 1990, Wimley et al. 1994) that will allow other enzymes (lysozyme, MPO, etc.) to invade and kill the microbe (Rice et al. 1987). Due to their positive charge, defensins are highly effective against negative charged membranes (Fujii et al. 1993). Their cationic, hydrophobic properties and lack of specific lipid requirements allows defensins to degrade bacteria (both Gram positive and Gram negative), fungi and enveloped viruses (against non-enveloped viruses defensins are inactive) (Schibli et al. 2002, Lehrer et al. 1989, Daher et al. 1986, Ericksen et al. 2005, Ganz et al. 1985).

In addition to their antimicrobial properties defensins are known to activate the host immune response by regulating the complement system, a process that enhances the ability of antibodies and phagocytic cells to destroy pathogens (Chertov et al. 2000, Yang et al. 2002). Defensins have been shown to induce migration of mast cells, macrophages, naïve dendritic cells and naïve T cells which activate the adaptive immune response through various processes (inflammation, wound healing) (Grigat et al. 2007, Yang et al. 2000, Lillard et al. 1999, Selsted and Ouellette 2005, Hazlett and Wu 2011).

Copy number variation (CNV) has been observed in the α-defensin cluster on 8p23.1, independent of the β-defensin CNV locus (Khan et al. 2013, Aldred et al. 2005, Mars et al. 1995, Linzmeier and Ganz 2005). The version of the human genome assembly used in the present study (February 2009) and comparison with other available sequences shows the structure of the *DEFA1A3* locus: two partial repeats of approximately 10kb flanking a variable number of circa 19kb full repeats (Figure 1.8). Each repeat unit can present either the *DEFA1* or *DEFA3* variant, and thus both genes can vary in copy number (Khan et al. 2013, Aldred et al. 2005, Mars et al. 1995, Linzmeier and Ganz 2005). For each repeat unit there is a copy of either *DEFT1P* or *DEFA10P*, pseudogenes with high sequence similarity (Patil et al. 2004, Ballana et al. 2007). Ganz and colleagues showed in 1990 that about 10% of individuals lack *DEFA3* (Ganz et al. 1990). Recently it has been demonstrated that, depending on the population studied, the percentage of individuals that do not possess DEFA3 can vary between

10% and 37% (Aldred et al. 2005, Ballana et al. 2007). The functional consequences are still unknown (Aldred et al. 2005).

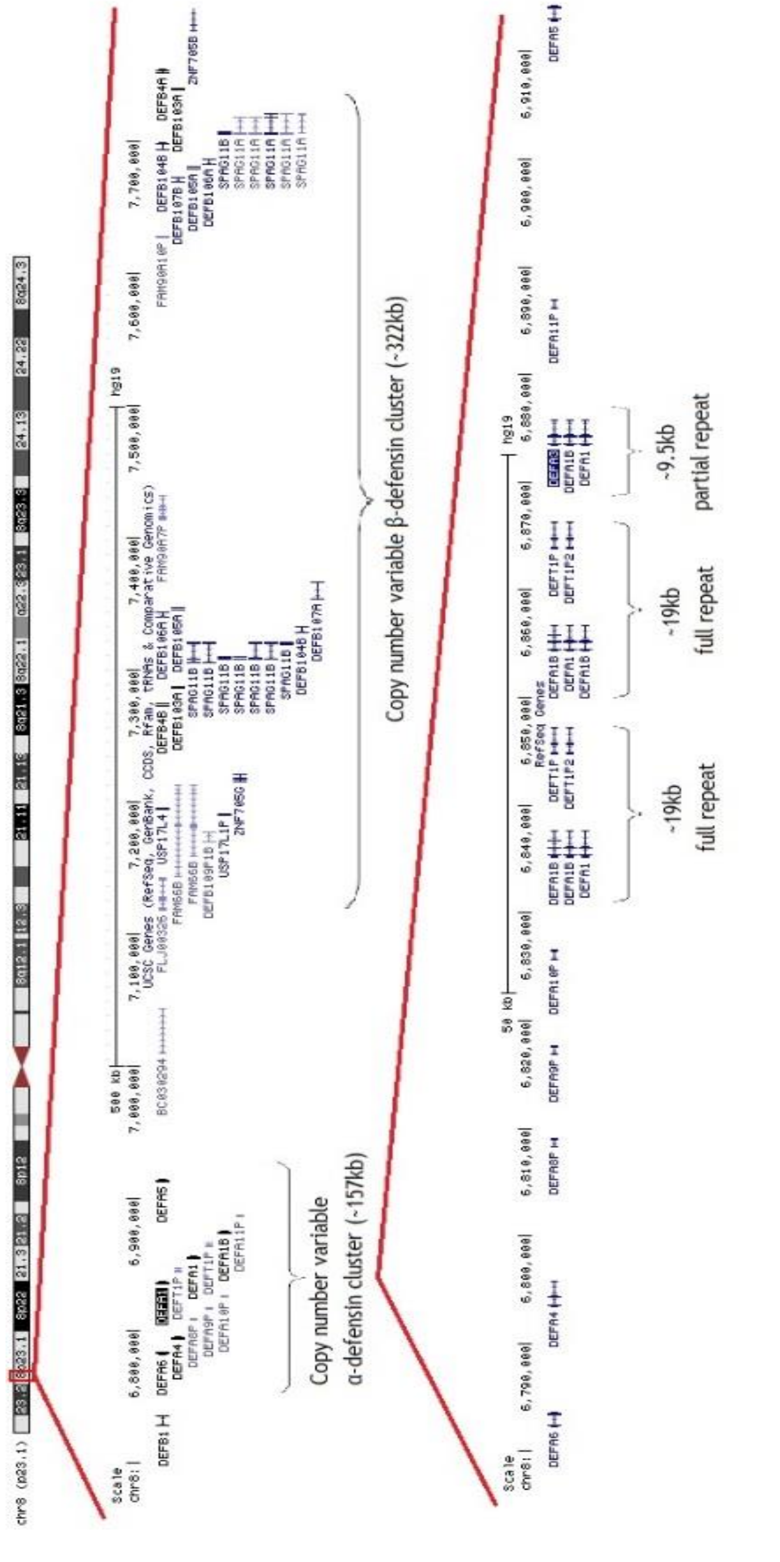


Figure 1.8 - Schematic of DEFA1A3 position on chromosome 8 (Paula Necsoiu, adapted after UCSC Genome Browser)

There are seven commonly used locus-specific techniques that have been used to accurately genotype CNVs in large cohorts: Southern Blotting, comparative genome hybridisation (CGH), multiplex amplifiable probe hybridisation (MAPH), multiplex ligation-dependent probe amplification (MLPA), fluorescence in-situ hybridization (FISH), quantitative real-time PCR (qPCR) and paralogue ratio test (PRT). Southern Blotting of pulsed-field gels is the most direct way to quantify the copy number of a repeat locus, if an enzyme can be found that excises the full repeat block in a single fragment; however, this assay requires a lot of time and very high quality of DNA in order to generate clear results (Aldred, 2004). CGH is a high throughput technique that compares a reference sample to a test sample on a microarray; Mantripragada and colleagues published in 2004 their results showing that the resolution of BAC-array CGH was 75kb so in order to detect a 19kb repeat region a probe smaller than a BAC clone would have to be carefully designed. Both MAPH (Armour et al. 2000) and MLPA (Schouten et al. 2002) are PCR-based methods which rely on the basic principle that after hybridisation of specifically-bound probes to genomic DNA (double stranded for MAPH, single stranded for MLPA), these can be recovered and amplified in order to detect deletions and duplication in relation to specific genetic disorders (Aldred et al. 2005, den Dunnen and White 2006); the high degree of multiplexing made them widely applied as ideal methods for targeted disease-specific approaches (Hollox et al. 2002, White et al. 2002, Hogervorst et al. 2003, Slater et al. 2003, White et al. 2004). FISH is another powerful, widely applied technique that uses fluorescent specifically-bound probes which will only bind to chromosome regions that present high sequence complementarity (Rosenberg et al. 1998); however, the probes used are usually large (40-100kb) and not advisable for detecting small sequences between 1-50 kb, so closely adjacent fluorescent signals will not be easily distinguished (den Dunnen and White 2006). qPCR is a powerful and quick method able to detect very small deletions and duplications; the disadvantage is that it can only analyse one target region per assay (Yau et al. 1996) and has been shown to distort associations at CNV loci (Aldhous et al. 2010, Cantsilieris et al. 2012, Cantsilieris et al. 2014). While previous studies (Linzmeier and Ganz 2005, Chen et al. 2010, Jespersgaard et al. 2011) have used real-time PCR in order to quantify *DEFA1A3* copy number, a combination of four allelic ratios and two PRTs (Armour et al. 2007) provided a highly accurate, reliable technique for copy number measurement (detailed description in Section 2.4 and Section 2.5).

DEFA1A3 copy number commonly varies between 3 and 16 copies in Chinese populations and between 3 and 11 copies in European population samples (Khan et al. 2013, Aldred et al. 2005, Linzmeier and Ganz 2005, Cheng et al. 2013, Nuytten et al. 2009). There are SNPs on either side of the *DEFA1A3* locus displaying high linkage disequilibrium (LD) (International HapMap Consortium 2003, Black et al. 2014) which suggest that, within the locus, crossover events are rare. However, high sequence similarity between repeats in a locus that manifests extensive copy number variation would suggest that copy number mutations occur with a high frequency. Generally, the process by which the *DEFA1A3* locus varies in copy number is still unclear (Black et al. 2014).

Due to the limited ability of a bi-allelic SNP to tag multiple copy number states, the association of copy number with flanking sequence variants is poor at multi-allelic CNV loci (Beckmann et al. 2007). However, in previous studies the SNP rs4300027 has been identified to correlate reasonably well with *DEFA1A3* copy number in European populations, where it can distinguish between 3-6 diploid copy number genotype (CC variant) from those that present 6-8 copy number genotype (CT variant) and 8-16 copy number genotype (TT variant) (Khan et al. 2013). Further work was carried by Black and colleagues in 2014 to try and determine whether the *DEFA1A3* flanking region contains other SNPs that might provide more detailed prediction of haplotype copy numbers and variant composition. Using a combination of four SNPs (rs4300027, rs7826487, rs7825750 and rs62487514), Black and colleagues observed 21 different sequences across a 4.1kb region, which they grouped into five classes, based on sequence similarity: Reference Sequence, Class 1, Class 2, Exchange 1 and Exchange 2. Another haplotype, Class 3, was observed on a single occasion, having a different sequence in comparison to the other classes. SNP rs4300027 is of great importance as it separates Class 1, Class 2 and Exchange 2 from the Reference Sequence and Exchange 1 haplotypes (Figure 1.9).

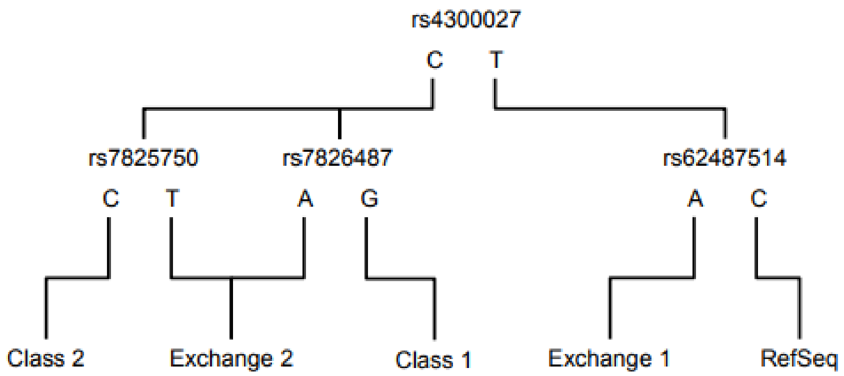


Figure 1.9 - The four SNPs used to tag the haplotype classes at DEFA1A3. rs4300027 genotype profile can be used to indicate the classes of the haplotypes for each sample. (Black, 2014)

In 2005 Linzmeier and Ganz found that the amount of defensin peptide expression was proportional to *DEFA1A3* copy number but used only eight subjects. In a second, larger study, no relationship was observed (Danielle Carpenter and colleagues, unpublished work). Furthermore, analysis of over 150 subjects failed to demonstrate any significant association between *DEFA1A3* copy number and levels of HNP1-3 from serum, urine or LPS-stimulated neutrophils (Ai et al 2016). Another study conducted by Aldred and colleagues in 2005 showed there was no relationship between *DEFA1A3* copy number and total mRNA expression in neutrophils, although the relative *DEFA1* vs *DEFA3* ratio was maintained, suggesting all copies are expressed.

1.3 Defensins in diseases

Defensins have been believed to be possible contributors to various diseases and conditions, especially those of an autoimmune nature, either affecting the progress or the outcome. Β-defensins are known to be active against bacteria (*H.influenzae*, *M.catarrhalis*, *S.pneumoniae*) (Lee et al.1998) and have an anti-infective and immunomodulatory role in the case of psoriasis (Hollox et al. 2008). Various researchers have hypothesized and investigated the alleged role of β-defensins CNV in the case of a number of diseases; no significant association has been found in the case of cystic fibrosis (Hollox et al. 2005), systemic lupus erythematosus (Sthoeger et al. 2009), Crohn's disease (Aldhous et al. 2010, WTCCC, Craddock et al. 2010), or rheumatoid arthritis (WTCCC, Craddock et al. 2010).

A case-control study conducted by Sthoeger and colleagues in 2009 investigated the levels of neutrophil α-defensin and β-defensin in plasma of SLE (systemic lupus erythematosus) patients. Although no significant correlation was found for β-defensins, they discovered higher levels of α-defensins in patients versus controls and a link with severity of disease; whether increased α-defensin levels contribute to SLE severity or not is yet unknown.

Previous studies have demonstrated variations in β-defensin expression levels in both gingival and periodontal inflammatory diseases (Krisanaprakornkit et al. 2000, Chung et al. 2004, Krisanaprakornkit et al. 2002, Feucht et al. 2003, Dommisch et al.

2005, Lu et al. 2004, Dunsche et al. 2002, Lu et al. 2005); β-defensin mRNA was less frequently detected in samples from gingivitis and periodontitis patients in comparison to healthy gingival biopsies (Dunsche et al. 2002). A recent GWAS study using German and Dutch case-control samples conducted by Munz and colleagues in 2017 identified two single nucleotide variants at *DEFA1A3* (rs2738058, rs2978951) as susceptibility loci for aggressive periodontitis (AgP); these results have been validated in German samples of chronic periodontitis and replicated in Turkish samples of AgP.

Recent studies (Yu et al. 2012, Kiryluk and Novak 2014, Li et al. 2015, Ai et al. 2016, Qi et al. 2016) have identified and validated the *DEFA1A3* locus as a strong genetic risk for IgA nephropathy (IgAN). In 2016 Ai and colleagues analysed two independent Chinese Han cohorts and confirmed the association in a Caucasian cohort: low copy-number of *DEFA1A3* CNV manifests strong correlation with IgA nephropathy and renal dysfunction in IgAN patients. Another study conducted in 2016 by Qi and colleagues found that low galactose-deficient IgA1 was correlated with increased α-defensin levels in IgAN patients of Chinese origin. Improving the methods of analysis will push forward the IgAN studies, offering the opportunity to understand in detail how *DEFA1A3* influences the risk of developing IgA nephropathy.

1.4 Aims of this study

In this study, the primary aim was to ask what influence does the CNV of alpha defensin have on the gene function and if there is any significant relationship between alpha defensin copy number and expression levels. We studied the relationship between genomic and cDNA *DEFA1A3* ratios and combined them with results obtained from the flanking variant assays.

Secondly, previous unpublished work predicted that the synonymous 6836028G variant affecting a valine codon is linked to the Reference Sequence haplotype and so in this project we studied this variant to see if there is any association with *DEFA1A3* copy number, gene expression or haplotype class.

2. MATERIALS AND METHODS

2.1 Samples

For this project 77 healthy, unrelated volunteers of European origin (operationally defined for this study as individuals who have at least three of their four grandparents born in Europe) agreed to donate 20ml of blood by signing an informed, written agreement (project approved by the Medical School Ethics committee, BT 10/02/2010). Once taken, the blood samples were only referred to by ID numbers (AN17001-AN17077) and our results include no information tying these samples to any individual. Plasma, DNA and RNA were immediately extracted from the samples, and no blood cells were stored.

Blood collection was performed by Prof.Armour; WBC isolation, DNA and RNA extraction, sample typing and analysis were performed by Paula Necsoiu.

2.2 From blood sample to nucleic acids

2.2.1 WBC isolation

In order to separate WBC from EDTA-anticoagulated whole blood we used the direct separation method (i.e. “buffy coat” collection; centrifuged for 15 minutes at 2000g). Plasma was isolated in 2ml tubes for each sample and stored at -80°C for later use. For a few random samples we made Leishman-stained blood smears after each important step, which we mounted for later visualisation. The buffy coat of WBC was collected into two 15ml Falcon tube; the remaining RBC were lysed twice using ACK lysis buffer (1X Ammonium Chloride Potassium: 155mM NH_4Cl, 10mM $KHCO_3$, 0.1mM EDTA). The cell pellets were washed once with PBS solution (1X Phosphate Buffered Saline: 137mM NaCl, 2.7mM KCl, 8mM Na_2HPO_4, 2mM KH_2PO_4), transferred into a 1.5ml Eppendorf tube and used immediately for nucleic acid extraction.

2.2.2 RNA extraction

Total RNA was obtained using a Trizol® (TRI Reagent: guanidine, thiocyanate, phenol) extraction method (Invitrogen™). After adding 1ml cold Trizol and 0.2ml chloroform on top of the WBC pellet, mixing and centrifuging, the clear upper aqueous phase which contains total RNA was separated from the pink lower organic phase and transferred to a new 1.5ml Eppendorf tube. Total RNA was precipitated using 0.5ml isopropanol. The pellet was washed with 1ml of 75% ethanol and air-dried for 15 minutes at room temperature, before resuspending the pellet in 20µl of SDW (Sterile Distilled Water). The final step was incubating the samples for 5-10 minutes at 60°C and storing them at -80°C.

2.2.3 DNA extraction

DNA extraction from WBC was performed using a standard phenol/chloroform extraction and ethanol precipitation method adapted from the Nucleon DNA extraction kit. The first step of this method is to add 1-1.5ml Solution A (10mM Tris HCl, pH 8.0, 320mM sucrose, 1% Triton X-100, autoclaved) on top of the WBC pellet; this gently lyses cells but leaves white cell nuclei intact. Next steps are to add 0.35ml Solution B (400mM Tris HCl, pH 8.0, 60mM EDTA, 150mM NaCl, autoclave, 1% SDS) to the pellet which will lyse the nuclei, freeing the DNA from protein complexes; adding 0.1ml 5M Sodium Perchlorate will further dissociate proteins from DNA and an aqueous phase containing the nucleic acid was extracted using 0.5ml phenol:chloroform:isoamylalcohol (25:24:1, equilibrated with Tris-HCl pH 8.0). The recovered aqueous phase was transferred to a 1.5ml Eppendorf tube and the DNA precipitated using 0.9ml 100% ethanol. The pellet was washed with 1ml of 80% ethanol, air-dried for 15 minutes, resuspended in 0.1ml H_2O, left to dissolve completely overnight at 4°C and then stored at -20°C.

2.2.4 DNA/RNA sample normalisation

In order to make two 96-well plates of DNA and RNA normalised to standard concentrations, Thermo Scientific™ NanoDrop™ 2000c Spectrophotometer was used to quantify the total absorbance of each individual sample.

Based on these results, 100µl aliquots of 10ng/µl were made for each DNA sample. The only exception is sample AN17002 for which a 50µl aliquot of 10ng/µl was made due to low DNA concentration.

Similarly, 20µl aliquots of 100ng/µl were made for each RNA sample. The two exceptions are samples AN17004 and AN17037 for which 20µl aliquots of 50ng/µl were made due to low RNA concentration.

For 7 DNA samples, the PRTs and Ratio Assays failed consistently, indicating that the true concentration of DNA was lower than reported by the Thermo Scientific™ NanoDrop™ 2000c Spectrophotometer, so new aliquots of higher concentration were required. After a few unsuccessful attempts, we had positive results for the following apparent concentrations: 200ng/µl (sample AN17030) and 100ng/µl (samples AN17045, AN17046, AN17056, AN17062, AN17064, AN17074).

2.2.5 cDNA synthesis

Reverse transcription was used to synthesize cDNA, accurately following the instructions given by the manufacturer for the Tetro cDNA synthesis kit (Bioline). All components (DEPC Treated Water, 5x RT Buffer, Oligo dT Primers, dNTP mix, RNase Inhibitor, Tetro Reverse Transcriptase) except input RNA were premixed and dispensed to 77 individual wells before adding RNA, thus ensuring uniformity of conditions for each reaction. The plate was incubated 30 minutes at 45°C, for 5 minutes at 85°C, and then stored at -20°C.

2.3 Qualitative *DEFA1A3* Ratio Assays

This method was used mainly as a control in order to quickly visualise in the 77 samples the presence or absence of *DEFA3*, in both genomic DNA and cDNA. In addition, it was used to check the results obtained from the quantitative *DEFA1A3* Ratio Assays (both genomic and cDNA).

Both for genomic DNA and cDNA, 10μl PCR reactions were made using 1μl DNA/cDNA and 9μl master mix, composed of ddH_2O, 1μl 10x LD mix (H_2O, 500mM Tris HCl pH 8.8, 125mM Ammonium sulphate, 1.25mg/ml BSA, 200μM dATP, 200μM dTTP, 200μM dGTP, 200μM dCTP, 14mM $MgCl_2$, 75mM 2-mercaptoethanol) (Khan et al. 2013), 1μl each 10μM DefRes4F+DefRes4R primers (for sequences, check Appendix 1) and 0.5U Taq DNA polymerase (New England Biolabs). Reaction conditions were 35 cycles of 95°C for 30 seconds, 56°C for 30 seconds and 70°C for 2 minutes.

When the PCR reaction was complete, 5μl of PCR product was digested with 10μl of a mix containing 1.5μl 10x NEBuffer4 (New England Biolabs) and 1U *Hae*III (total volume of 15μl) at 37°C for at least 2 hours.

The digested products were separated on a 2% agarose gel and pictures of the gels were taken using Bio-Rad™ Quantity One®.

2.4 PRTs and Allelic Ratio Assays

Two duplex paralogue ratio tests (PRT), combined with four allelic ratio assays (*Hae*III digestion, indel 5, 7bp duplication, 4bp deletion) were used to accurately measure the copy number of *DEFA1A3*, as well as the individual copy number of *DEFA1* and *DEFA3* for each sample (Khan et al. 2013, Ai et al. 2016).

2.4.1 Paralogue Ratio Tests

The two PRTs are specifically designed to measure the copy number in both centromeric (CEN1) and telomeric (TEL2) regions of chromosome 8, thus allowing us to establish the *DEFA1A3* copy number.

Both CEN1 and TEL2 PRTs are implemented as two duplicate PCR assays: each pair of duplicates share the same unlabelled reverse primer with the difference that one uses a HEX-labelled forward primer and the other one uses a FAM-labelled forward primer. Each PRT (CEN1 or TEL2) is therefore done in duplicate, once with HEX, and once with FAM, to give a total of four PRT measurements per sample. Using two separate dyes allowed us to analyse both PCR products for each PRT for each sample in a single electrophoresis capillary. The average between the two PCR products obtained, both in CEN1 and TEL2, allowed us to determine a single unrounded copy number estimate for each sample.

For each PCR, 1µl of DNA (10ng/µl; for exceptions see Section 2.2.4) was added to a 9µl master mix, which contains ddH_2O, 1µl 10x LD mix (Khan et al. 2013), 0.5µl each 10µM primers and 0.5U Taq DNA polymerase (New England Biolabs).

At the centromeric end, products were amplified using 0.5µl of a common reverse DEFAcen1R primer and 0.5µl of either a FAM-labelled forward FAM-DEFAcen1F primer or a HEX-labelled forward

HEX-DEFAcen1F primer (for primer sequences, see Appendix 1). Reaction conditions were 25 cycles of 95°C for 30 seconds, 58°C for 30 seconds and 68°C for 2 minutes, followed by a hold at 72°C for 40 minutes. The CEN1 PRT will generate amplicons of 333bp for non-centromeric copies of *DEFA1A3* and 444bp for the two reference (centromeric) copies.

At the telomeric end, products were amplified using 0.5µl of a common reverse DEFAtel2R primer and 0.5µl of either a FAM-labelled forward FAM-DEFAtel2F primer or a HEX-labelled forward HEX-DEFAtel2F primer (for primer sequences, see Appendix 1). Reaction conditions were 25 cycles of 95°C for 30 seconds, 66°C for 30 seconds and 63°C for 2 minutes, followed by a hold at 72°C for 40 minutes. The TEL2 PRT will generate amplicons of 262bp for the two reference (telomeric) copies and 266bp for non-telomeric copies of *DEFA1A3*.

2.4.2 Quantitative *DEFA1A3* Ratio Assay

This assay measures the ratio between *DEFA1* and *DEFA3* using a *Hae*III restriction site difference between them (present in *DEFA1* and absent in *DEFA3*). It uses a deliberate mismatch that alters the sequence of the forward primer from GCT to GCC, creating an artificial control site for *Hae*III, allowing us to distinguish between a completely undigested product and the *DEFA3* product. This particular method is adapted from the one used by Khan and colleagues in 2013. The generated amplicons vary in size, making it easy to determine the ratio between products from *DEFA1* (144bp) and *DEFA3* (161bp).

2.4.3 Indel5

This insertion/deletion variant of 5bp is present in many repeats at the *DEFA1A3* locus and was previously used and described by Khan and colleagues in 2013. The generated amplicons are 124bp (deleted) and 129bp (undeleted).

2.4.4 7bp duplication

This duplication variant of 7bp is present in many repeats at the *DEFA1A3* locus and was previously used and described by Khan and colleagues in 2013. The generated amplicons are 275bp (unduplicated) and 282bp (duplicated).

2.4.5 4bp deletion

This 4bp deletion variant assay was developed and described by Ai and colleagues in 2016. The generated amplicons are 211bp (deleted) and 215bp (undeleted).

2.4.6 PCR purification

A 10µl PCR master mix containing 3µl of the inactivated DEFAHae3 digestion product, 1µl from Indel5 PCR, 1µl from 7bpdup PCR, 1µl from 4bpdel PCR, 1µl from Cen1FAM PCR, 1µl from Cen1Hex PCR, 1µl of Tel2Fam and 1µl of Tel2Hex was purified and freed of contaminants following the instructions given by the manufacturer for Agencourt® AMPure® PCR Purification Kit (Agencourt Bioscience by Beckman Coulter).

Subsequently, 1μl of the purified PCR master mix was denatured in a mixture of 10μl HiDi formamide/Genescan™ 500 ROX™ marker (Applied Biosystems by Life Technologies) at 96°C for 3 minutes.

2.4.7 Capillary electrophoresis

The denatured mixture of PCR products/HiDi formamide/ROX was submitted for capillary electrophoresis (injection at 1kV for 15 seconds) using ABI 3130xl Genetic Analyzer at the Nottingham DeepSeq sequencing facility. GeneMapper 4.1 software (Applied Biosystems by Life Technologies) was used to extract the peak areas of the separated PCR products.

2.5 6836028 Variant Assay

A single nucleotide polymorphism at position 6836028 is a synonymous change, turning the bottom-strand codon for Valine from GTG (reference) to GTC (variant), which on the reference strand is CAC>GAC, and destroying a site for *Bsl*I (CCN_5/N_2GG) which cuts at 6836033. In the assembly, all three copies have C. There are overlapping ($CCCN_6GGG$) sites nearby, cutting at 6836015/6, but no others in exons 1, 2 or 3. A genomic assay cannot be made much longer, as there is a further *Bsl*I site almost immediately into intron 1 (at 6836174). This all means that for genomic DNA, the presence of a site at 6836033 would trim the 158bp fragment between 6836016 and 6836174 by 17bp to give 141bp.

The primers designed for this assay (6836028F and 6836028R) worked well for genomic DNA. The forward 6836028F primer was situated in intron 2, so in the case of cDNA the assay would not have worked. Therefore, we decided to replace it with the reverse DefRes4R primer from the qualitative DEFA1A3 Assay. This means that the cDNA uncut amplicon would be 344bp and the restriction fragments will be 196bp, 1bp and 17bp+130bp (C variant) or 147bp (G variant).

2.5.1 Qualitative assay at 6836028

This method was used mainly as a control in order to quickly visualise in the 77 samples the presence or absence of 6836028G variant, in both genomic DNA and cDNA. In addition, it was used to check the results obtained from the quantitative 6836028 Assays (both genomic and cDNA).

For genomic DNA, 10μl PCR reactions were made using 1μl DNA/cDNA and 9μl master mix, composed of ddH_2O, 1μl 10x LD mix, 1μl each 10μM 6836028F+6836028R primers (for sequences, check Appendix 1) and 0.5U Taq DNA polymerase (New England Biolabs). For cDNA, the forward 6836028F primer was replaced with the reverse DefRes4R primer (for sequences, check Appendix 1). Reaction conditions were 37 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 70°C for 30 seconds. When the PCR reaction was complete, 10μl of PCR product was digested with 20μl of a mix containing 3μl 10x NEBuffer4 (New England Biolabs) and 2.5U *BslI* (total volume of 30μl) overnight at 55°C. For genomic DNA, the expected amplicons would be 51bp+1bp+17bp+130bp (C variant) or 51bp+1bp+147bp (G variant). For cDNA, the expected amplicons would be 196bp+1bp+17bp+130bp (C variant) or 196bp+1bp+147bp (G variant). The digested products were separated on a 3% agarose gel and pictures of the gels were taken using Bio-Rad™ Quantity One®.

2.5.2 Quantitative assay at 6836028

The quantitative fluorescent version of the assay proceeded as above but used a FAM-labelled forward primer FAM-6836028F for genomic DNA or FAM labelled reverse primer FAM-DefRes4R (for sequence, see Appendix 1) and 25 cycles only. After the digestion, AMPure was used to purify 10µl (out of the 30µl total); 1µl was mixed in 10µl HiDi formamide/Genescan™ 500 ROX™ and separated using capillary electrophoresis (methods described in Section 2.4.6 and Section 2.4.7).

2.6 Flanking Variants Assays

Assigning *DEFA1A3* haplotype class to each sample was done using a combination of four SNP genotypes previously described in 2013 by Khan and colleagues and in 2014 by Black and colleagues.

2.6.1 Initial PCR

The centromeric region around the four SNPs was initially amplified as a 10µl PCR containing 1µl input DNA, 1µl 10x LD mix, 1µl each of 10µM forward TagSNPsF and reverse RepSpec2RB primers (for sequences, see Appendix 1) and 0.2µl (5U/µl) Taq DNA polymerase. Reaction conditions were 37 cycles of 95°C for 30 seconds, 57°C for 30 seconds and 70°C for 3 minutes. The expected uncut product is 1408bp (hg19 chr8:6879639-6881047).

2.6.2 rs4300027

Designed by Khan and colleagues in 2013 to genotype the SNP rs4300027 (chr8: 6,867,985), this Restriction Length Fragment Polymorphism (RFLP) assay creates a variable *Hin*fI site, in which the uncut T allele lacks a restriction site producing 378bp+130bp+47bp+821bp+36bp fragments, whilst the cut C allele produces 378bp+130bp+47bp+382bp+439bp+36bp fragments. 5µl of initial PCR product was digested with 15µl mix containing 2.5U *Hin*fI and 1x NEBuffer2 overnight at 37°C. The products were separated on a 2% agarose gel.

2.6.3 rs7826487

Designed by Black and colleagues in 2014 to genotype the SNP rs7826487 (chr8: 6,868,335) that tags the Class 1 haplotype, this RFLP assay creates a variable *Ssp*I site, in which the uncut G allele lacks a restriction site producing 679bp+730bp fragments, whilst the cut A allele produces 679bp+611bp+119bp fragments. 5µl of initial PCR product was digested with 15µl mix containing 2.5U *Ssp*I and 1x NEBuffer4 overnight at 37°C. The products were separated on a 2% agarose gel.

2.6.4 rs7825750

An assay was previously designed by Black and colleagues in 2014 to genotype the SNP rs7825750 (chr8: 6,867,213) that tags the Class 2 haplotype. In recent work the results were unsatisfactory, Class 2 homozygotes being hard to distinguish from heterozygotes, so a new assay was designed and used in this study. The new Class 2 assay uses a secondary 10µl PCR containing 0.5µl primary PCR product, 1µl 10x LD mix, 1µl each of 10µM forward TagSNPsF and reverse C2FnuR primers (for sequences, see Appendix 1) and 0.2µl (1U) Taq DNA polymerase. Reaction conditions were 12 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 70°C for 30 seconds. The expected uncut product is 190bp. The RFLP assay creates a variable *Fnu*4HI site, in which the uncut T allele lacks a restriction site producing 158bp+32bp fragments, whilst the cut C allele produces 133bp+33bp+25bp fragments. 5µl of initial PCR product was digested with 15µl mix containing 1U *Fnu*4HI and 1x NEBuffer4 overnight at 37°C. The products were separated on a 2% agarose gel.

2.6.5 rs62487514

This assay is a modification of the assay designed by Black and colleagues in 2014 to genotype the SNP rs62487514 (chr8: 6,867,504) that tags the Exchange 1 haplotype. The X1 assay uses a secondary 10µl PCR containing 0.5µl primary PCR product, 1µl 10x LD mix, 1µl each of 10µM forward Exch1F and reverse Exch1R primers (for sequences, see Appendix 1) and 0.2µl (1U) Taq DNA polymerase.

Reaction conditions were 12 cycles of 95°C for 30 seconds, 50°C for 30 seconds and 70°C for 30 seconds. The expected uncut product is 233bp. The RFLP assay creates a variable *Tsp*509I site, in which the uncut C allele lacks a restriction site producing 45bp+188bp fragments, whilst the cut A allele produces 45bp+162bp+26bp fragments. 5µl of initial PCR product was digested with 15µl mix containing 2.5U *Tsp*509I and 1x NEBuffer1 overnight at 65°C. The products were separated on a 3% agarose gel.

3. RESULTS

3.1 Samples

For this project we gathered 77 healthy, unrelated volunteers who agreed to donate blood. The inclusion criteria we imposed was that they had to be healthy, over 18 years old and of European origin, operationally defined for this study as individuals who have at least three of their four grandparents born in Europe (Appendix 2). Out of 77 volunteers we had a mix of males and females: 43 females with an average age of 32.91 and BMI of 22.58 and 34 males with an average age of 31.47 and BMI of 24.10. With our study taking place in England and the inclusion criteria being restricted to "European", as expected, most of the volunteers have British nationality (57 out of 77), the remaining 20 being either Irish (1), British/Italian (1), Italian (1), Spanish (2), Swiss (1), Dutch (1), German (3), Polish (3), Slovak (1), Romanian (4), Moldovan (1). There was one volunteer of Cypriot origins, whose sample we analysed together with the other 76, but we excluded it from the results. Out of 77 individuals, 8 of them felt like they were unwell that day due to or recovering from a cold. One of our volunteers reported a more serious, long-term condition (i.e. Type I diabetes).

3.2 Genomic DNA Analysis

3.2.1 Copy Number Analysis

3.2.1.1 *DEFA1A3* Qualitative Assay

In preliminary work we looked at a qualitative assay, formerly described by Aldred and colleagues in 2005, for the presence or absence of *DEFA3* (method described in Section 2.3). Out of 76 samples, 13 samples (20.63%) appeared to lack the *DEFA3* gene. Although this is a higher percentage than the 10% reported by Aldred et al, it is not significantly higher (chi-squared $P > 0.1$), and similar to the frequency reported by Khan et al. (2013).

3.2.1.2 PRTs

In this study we have combined two duplex PRTs with four allelic ratio assays (Def*Hae*3, indel5, 7bp duplication, 4bp deletion) and used them to accurately measure the copy number of *DEFA1A3*, as well as the individual copy number of *DEFA1* and *DEFA3* for each sample (Khan et al. 2013, Ai et al. 2016). A set of seven samples for which the copy number was determined by Aldred and colleagues in 2005 using restriction digestion, pulsed field gel electrophoresis and Southern blotting were re-typed with the first three samples gathered for this project (AN17001, AN17002, AN17003) in order to calibrate the ratios obtained from the PRTs and to convert said ratios to copy numbers; these 7 samples were not included in every PRT we performed, as the 3 samples (AN17001, AN17002, AN17003) could act as controls. 76 DNA samples were typed and analysed using the eight assays. One sample (AN17075) was noteworthy in having a higher apparent number of *DEFA3* than *DEFA1* (i.e. 4 copies of *DEFA3* and 1 copy of *DEFA1*, which is unusual). No other variation observed appeared anomalous or unprecedented.

The predicted PCR products from the eight PRTs and allelic ratios, after digestion and capillary electrophoresis, are in the form of well separated peaks, as shown in Figure 3.1. For CEN1 and TEL2 PRTs we used duplicate PCRs, with primers labelled either in FAM (blue) or HEX (green); the ratios are calculated by comparing the non-reference peaks in both FAM and HEX (263bp for TEL2, 333bp for CEN1) with the reference peaks in both FAM and HEX (260bp for TEL2, 444bp for CEN1). The indel5 ratio is the ratio of deleted (124bp) to undeleted (129bp). The DefHae3 ratio is the ratio of *DEFA1* (144bp) to *DEFA3* (161bp). The 4bp ratio is the ratio of deleted (211bp) to undeleted (215bp). The 7bpdup ratio is the ratio of unduplicated (275bp) to duplicated (282bp).

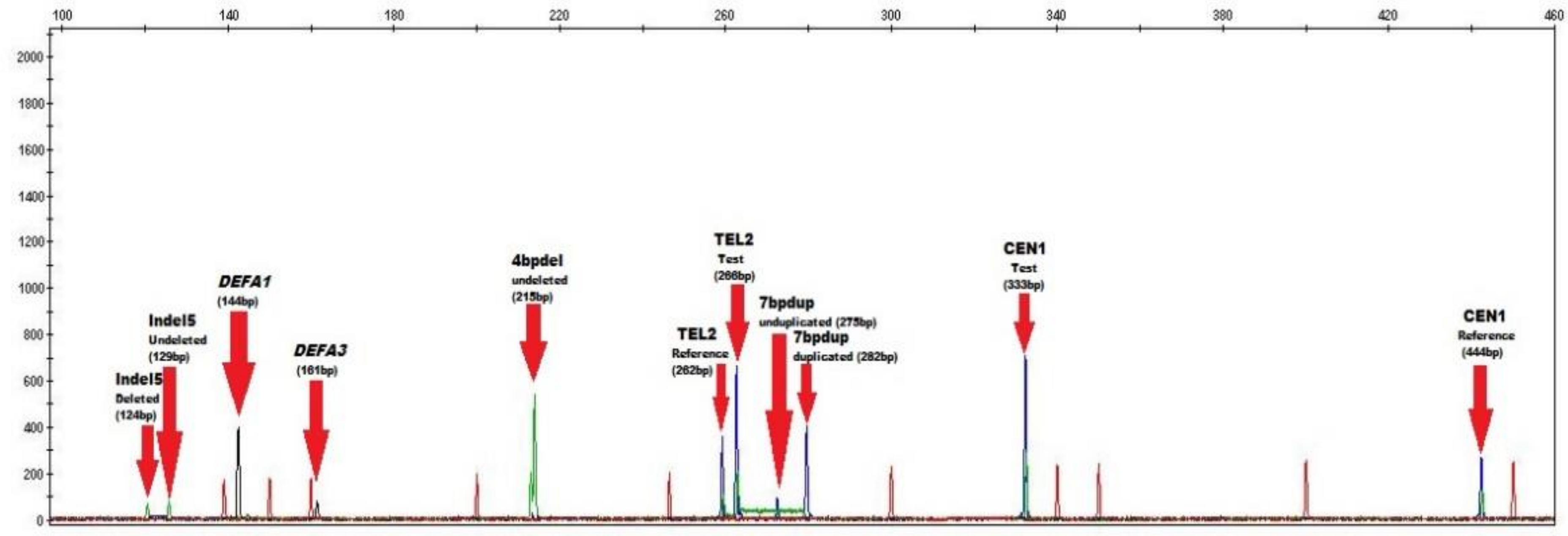


Figure 3.1 – GeneMapper profile of a random sample with a total copy number of 6. Both CEN1 and TEL2 ratios are test to reference (4:2). The indel5 ratio is the ratio of deleted to undeleted (3:3). The DefHae3 ratio is the ratio of DEFA1 to DEFA3 (5:1). The 4bp ratio is the ratio of deleted (0) to undeleted (6). The 7bpdup ratio is the ratio of unduplicated to duplicated (1:5)

The principle on which the PRTs were designed is that each chromosome 8 is assumed to carry exactly one centromeric partial repeat of *DEFA1A3* locus, one telomeric partial repeat and a variable number of full repeats. The CEN1 PRT uses as reference the two copies in the centromeric partial repeat of *DEFA1A3* locus (444bp), with non-centromeric copies giving a product of 332bp; in figure 3.2 the 10-copy sample AN17019 yields these products in amounts corresponding to a copy number ratio of 8 non-centromeric: 2 centromeric. Similarly, the TEL2 PRT will amplify a 2-copy reference product from the telomeric partial repeat (260bp) and a 263bp product from all other copies. Using two separate dyes for each PRT and an agreement between the CEN1 and TEL2 will increase the accuracy of the assay (Figure 3.2).

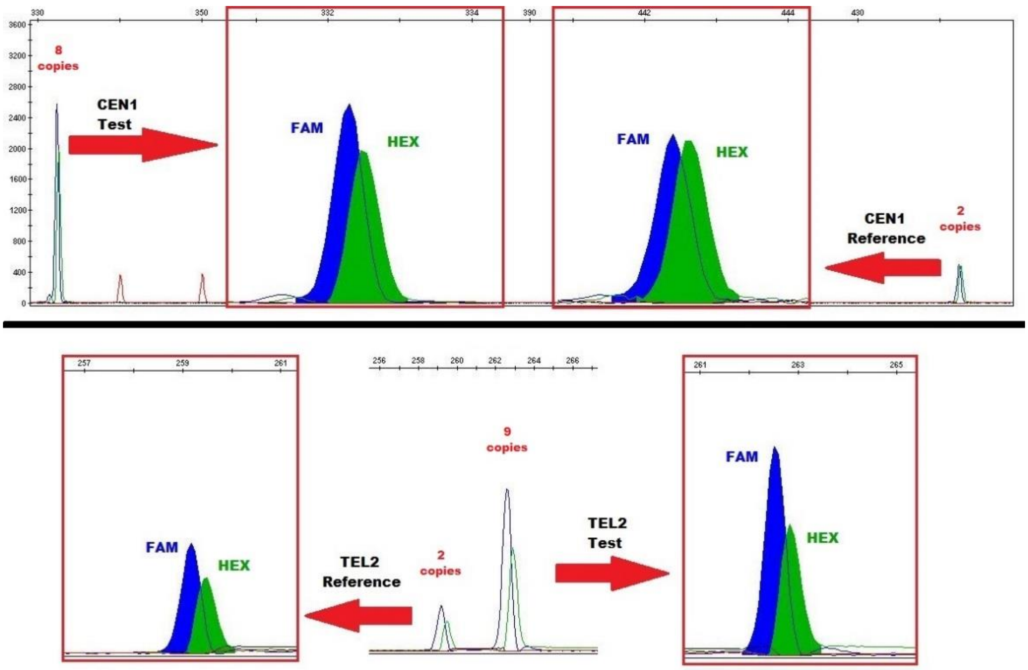


Figure 3.2 - Top: Schematic of predicted PCR products from a CEN1 PRT trace (sample AN17019, 10 copies, 8:2). Bottom: Schematic of predicted PCR products from a TEL2 PRT trace (sample AN17059, 11 copies, 9:2)

Using the DefHae3 assay (which will amplify a specific sequence from the gene) allowed us to determine a copy number ratio between *DEFA1* and *DEFA3* that can be combined with knowledge of the total copy number to identify the exact copy number of each gene. Digesting the PCR products with *Hae*III results in two amplicons which differ in size (144bp, 161bp) (Figure 3.3).

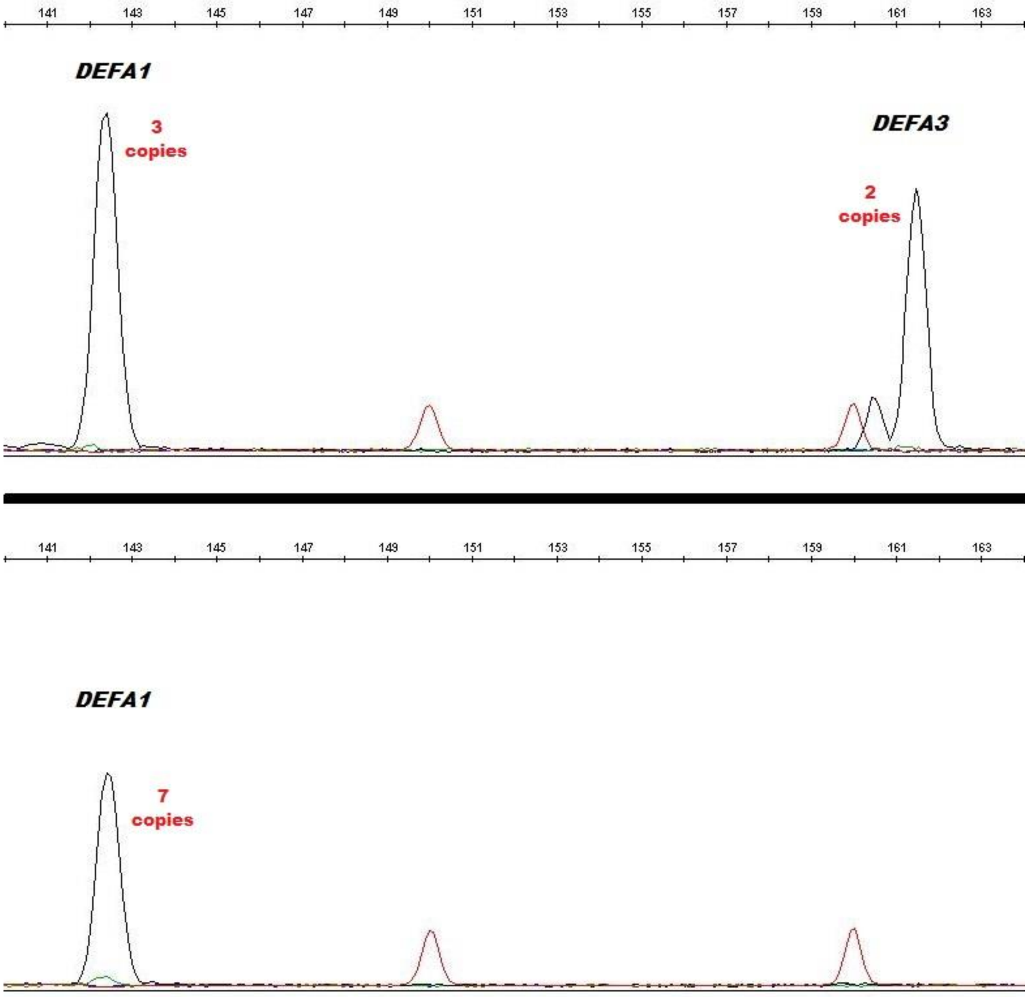


Figure 3.3 - Top: Schematic of predicted PCR products from a DefHae3 Ratio Assay trace (sample AN17002, 5 copies, 3:2). Bottom: Schematic of predicted PCR products from a DefHae3 Ratio Assay trace (sample AN17003, 7 copies, 7:0)

Using the Indel5 assay (which will amplify a specific sequence present in both partial and full repeats) allowed us to determine a ratio between the presence (129bp) and absence (124bp) of the 5bp insertion/deletion polymorphism (Figure 3.4).

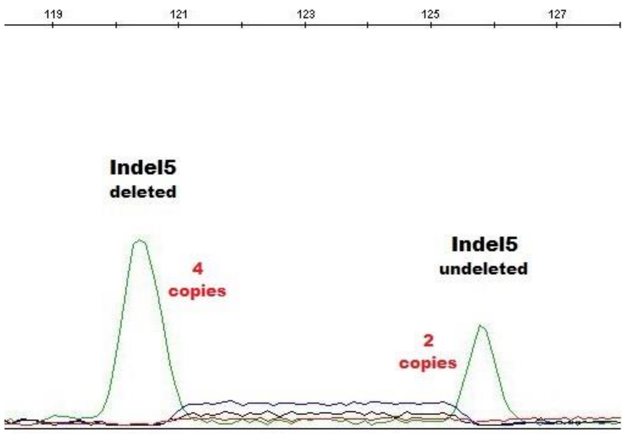


Figure 3.4 - Schematic of predicted PCR products from an Indel5 Ratio Assay trace (sample AN17005, 6 copies, 4:2)

Using the 7bpdup assay (which will amplify a specific sequence from intron 1 of the *DEFA1A3* gene) allowed us to determine a ratio between the presence (282bp) and absence (275bp) of the 7bp duplication polymorphism (Figure 3.5).

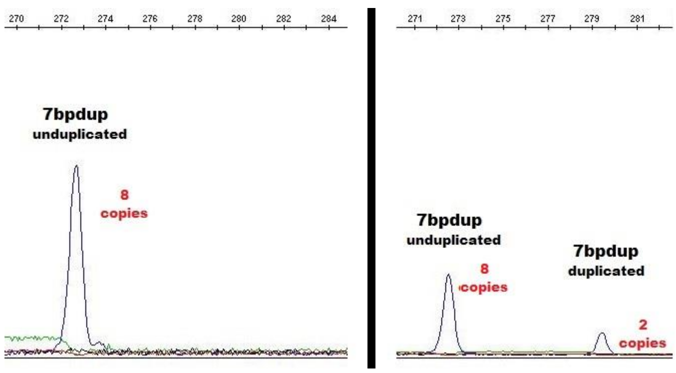


Figure 3.5 - Left: Schematic of predicted PCR products from a 7bpdup Ratio Assay trace (sample AN17012, 8 copies, 8:0). Right: Schematic of predicted PCR products from a 7bpdup Ratio Assay trace (sample AN17021, 10 copies, 8:2)

Using the 4bpdel assay (which will amplify a specific sequence present in intron 2 of the *DEFA1A3* gene) allowed us to determine a ratio between the presence (215bp) and deletion (211bp) of the 4bp deletion polymorphism; double peaks are presumably due to incomplete dATP addition to the 3' end (Figure 3.6).

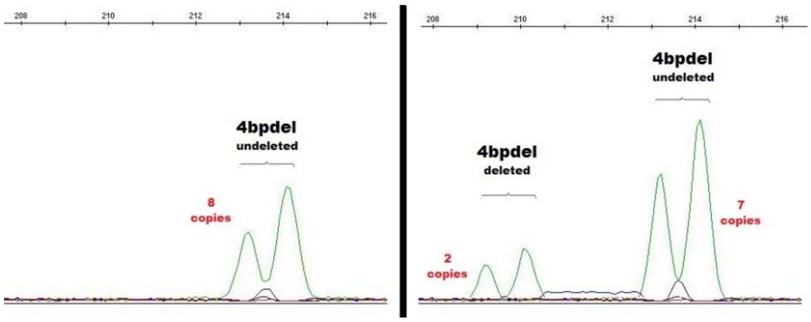


Figure 3.6 - Left: Schematic of predicted PCR products from a 4bpdel Ratio Assay trace (sample AN17012, 8 copies, 8:0). Right: Schematic of predicted PCR products from a 4bpdel Ratio Assay trace (sample AN17013, 9 copies, 7:2)

3.2.1.3 Assigning copy numbers

After typing, analysing and checking for missing values we assigned a copy number to each sample using an application designed by J.Armour, based on Maximum Likelihood (Khan et al. 2013, Black et al. 2014). This application takes a set of measured values and evaluates the probability of getting the PRTs (unrounded copy number), *DEFA1A3* (ratio), indel5 (ratio), 7bp duplication (ratio) and 4bp deletion (ratio) results actually obtained, assuming in turn each possible copy number from 2 to 16; it then combines the likelihood results (by multiplying them) to get the overall probability of the complete set of results for each of the possible copy numbers, with the copy number that maximises the overall probability of the data designated the Maximum Likelihood Copy Number (MLCN). The highest probability in each group is scaled to 1, and all other values expressed relative to that.

For a more specific explanation we have chosen sample AN17063 as an example: given PRT copy number estimates of 7.72591 and 8.23797, indel5 deletion/non-deletion ratio of 6.75289, *DEFA1*/*DEFA3* ratio of 2.46619, 7bp non-duplication/duplication ratio of 6.85387, and a 4bp deletion/non-deletion ratio of 0.13679, the program gave the output shown in Figure 3.7. N=8 gives the highest probability value of 1 for PRT1, PRT2, and 7bp, N=15 is favoured by indel5, N=16 is also favoured by 7bpdup, N=9 gives the highest probability for 4bpdel, and the best single copy number for the *DEFA1A3* ratio is N=11. N=8 is the most likely copy number overall in the "Combined" analysis in the first line. The next highest value in the combined analysis is 4.491×10^{-3} for N=9, and so N=8 has a probability that is more than 200 times higher than N=9. This is what "minimum ratio 222.68" means at the bottom left; taking all the evidence together, the best copy number (N=8) gives a probability that is more than 200 times higher than the next best copy number (N=9). This "minimum ratio" value is therefore a good indicator of confidence in the assigned integer copy number (Figure 3.7).

			best					relative likelihood values for						
			copy no.	analysis	N=2	N=3	N=4	N=5	N=6	N=7	N=8	N=9	N=10	N=11
Sample ID		AN17063	8	combinec	0	0	2.24E-176	2.70E-76	1.02E-21	0.000631	1	0.004491	7.17E-06	2.34E-09
PRT1	7.72591			PRT1	3.28E-178	3.69E-108	1.50E-62	2.87E-29	4.24E-05	0.290712	1	0.222597	0.047688	0.012756
PRT2	8.23797			PRT2	1.73E-211	3.55E-121	4.72E-58	3.11E-28	9.41E-11	0.014577	1	0.481045	0.128266	0.035704
indel5 rati	6.75289			indel5	0	7.45E-98	1.31E-26	4.92E-08	0.024035	0.794259	0.744092	0.17437	0.024576	0.003037
DEFA1A3 r	2.46619			DEFA1A3	1.65E-33	0.043918	0.749463	0.05724	0.043918	0.863976	0.749463	0.238733	0.545257	1
7bp dup ra	6.85387			7bp dup	0	1.54E-89	1.32E-25	2.14E-08	0.008581	0.502542	1	0.497801	0.140961	0.032656
4bp del ra	0.13697			4bp del	5.62E-13	1.58E-09	1.20E-06	0.000248	0.01396	0.213344	0.88603	1	0.306715	0.025565
minimum ratio		222.68												

Figure 3.7 – MLCN output for sample AN17063 (for clarity, the display of the table misses out copy numbers of 12-16 and some other detail in the actual output)

Similar to previous work done by Aldred and colleagues in 2005 and Khan and colleagues in 2013, the total *DEFA1A3* copy number distribution in the 76 sampled Europeans varies between 4 and 12 copies per diploid genome (Figure 3.8). Comparing our data with the results obtained by both Aldred et al. and Khan et al. showed that there is no significant difference and so we can accept them as a result of random sampling in Europeans (chi squared P>0.1).

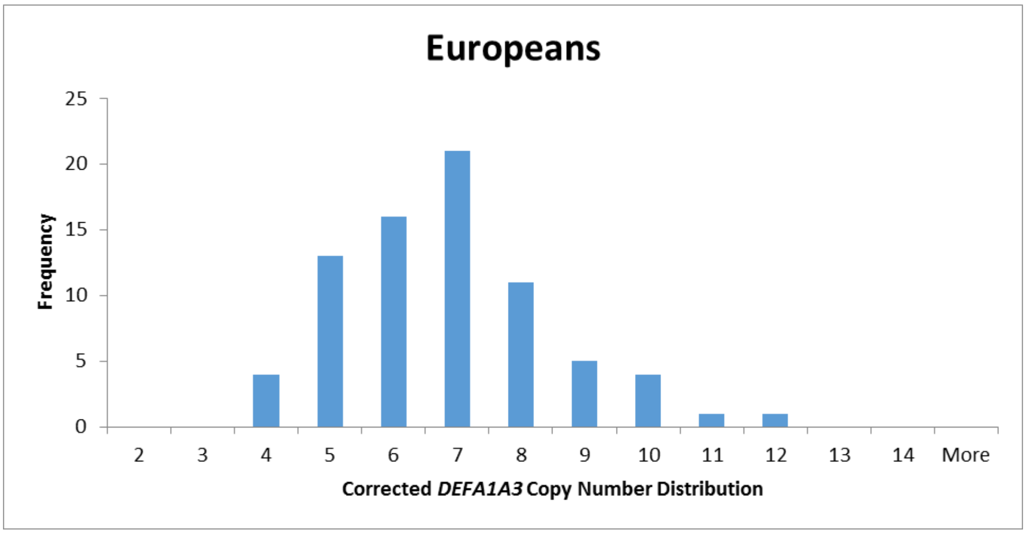


Figure 3.8 – Copy number distribution in 76 unrelated European individuals

The MLCN application takes everything at face value and assumes that the different systems are measuring the same underlying copy number, but the minimum ratio value will indicate the degree of confidence we should have when assigning a copy number: the higher the value, the greater the confidence we can have in accurate designation of the copy number. A minimum ratio greater than 10 was recorded for 57 out of the 76 samples typed (more than 70%) (Figure 3.9).

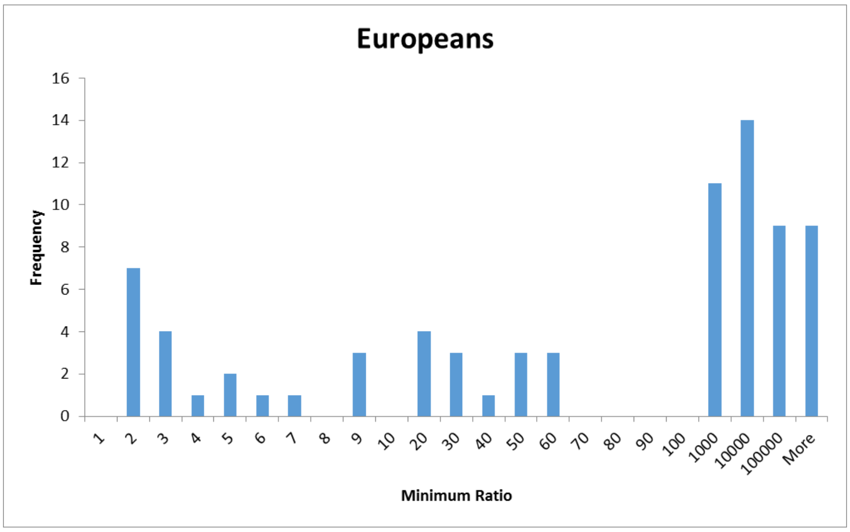


Figure 3.9 – Minimum ratio distribution in 76 unrelated European individuals

The results obtained from the four allelic ratio assays (DefHae3, indel5, 4bpdel, 7pdup) have more than one purpose in our study. Firstly, they validate the unrounded total copy numbers quantified from the PRTs (Table 3.1).

Sample ID	CEN1 PRT (unrounded copy numbers)	TEL2 PRT (unrounded copy numbers)	DefHae3 ratio (DEFA1:DEFA3)			indel5 ratio (deleted:undeleted)		4bpdel ratio (deleted:undeleted)		7bpdup ratio (unduplicated:duplicated)	
			Measured	Corrected	Interpreted	Measured	Interpreted	Measured	Interpreted	Measured	Interpreted
AN17008	4.43517	5.91172	0.93	1.209	3:3	0	0:6	0.00	0:6	4.11377	1:5
AN17014	8.08097	7.40586	2.48601	3.23181	6:2	0	0:8	0.2292	6:2	6.06623	1:7
AN17018	7.34555	8.98428	2.87485	3.73731	7:2	2.19218	3:6	0.15336	8:1	1.85592	3:6
AN17027	5.47474	6.73298	1.67237	2.17408	4:2	5.62672	1:5	0	0:6	2.50253	2:4
AN17039	7.95781	6.89699	2.04697	2.66106	6:2	1.8875	3:5	0	0:8	1.28571	4:4
AN17040	4.36397	6.07102	0.92781	1.20616	3:3	0	0:6	0	0:6	4.7332	1:5
AN17042	7.10017	6.28416	0	0	7:0	0.79614	4:3	0.00	0:7	0.51637	5:2
AN17046	5.20854	4.39677	1.33516	1.73571	3:2	0.00	0:5	0.00	0:5	0.84458	3:2
AN17047	6.26277	8.41208	2.42523	3.1528	6:2	0.52768	5:3	0.00	0:8	1.46767	3:5
AN17054	6.86913	8.03687	2.0468	2.66084	5:2	2.50837	2:5	0.18547	6:1	2.46542	2:5

Table 3.1 - Illustration of the usefulness of the allelic ratios in various samples. In the case of sample AN17027 for example, the unrounded copy numbers from the PRTs are either 6 or 7; using a DefHae3 corrected ratio of 2.1 would support a total copy number of 6 over 7 (4 DEFA1 to 2 DEFA3).

Secondly, we can use the obtained ratios to determine for every polymorphism the exact number of copies containing either the *DEFA1* gene or the *DEFA3* gene, the deleted or the undeleted versions of indel5, the deleted or the undeleted versions of 4bpdel and the duplicated or the unduplicated version of 7bpdup.

Upon getting the results from the DefHae3 assay we observed lower than expected ratios by a factor of about 30%, due to failure of *Hae*III to digest *DEFA1A3* products completely, possibly due to the generation of single-stranded DNA or heteroduplexes (Figure 3.10). The ratios were corrected to account for this error, allowing us to determine (in combination with the total copy numbers measured by PRT) the number of copies containing *DEFA1* or *DEFA3*. *DEFA3* gene was absent in 13 samples out of 76 (20.63%), confirming our previous results obtained via qualitative *DEFA1A3* assay (Figure 3.11). One sample (AN17075) appeared to have a higher copy number of *DEFA3* than *DEFA1* (i.e. 4 copies of *DEFA3* and 1 copy of *DEFA1*). After repeating the result, we have excluded artefacts from (for example) heteroduplex formation. There is only one similar sample known to date (i.e. NA06990) with a haplotype which has more copies of *DEFA3* than *DEFA1*, discussed by Tyson and Armour in 2012. We speculate the possibility that our 5-copy sample (for which we know has a Class2/Reference haplotypes) to have two copies of *DEFA3* and one copy of *DEFA1* on one allele (Class2 haplotype) and two copies of *DEFA3* on the other allele (Reference haplotype).

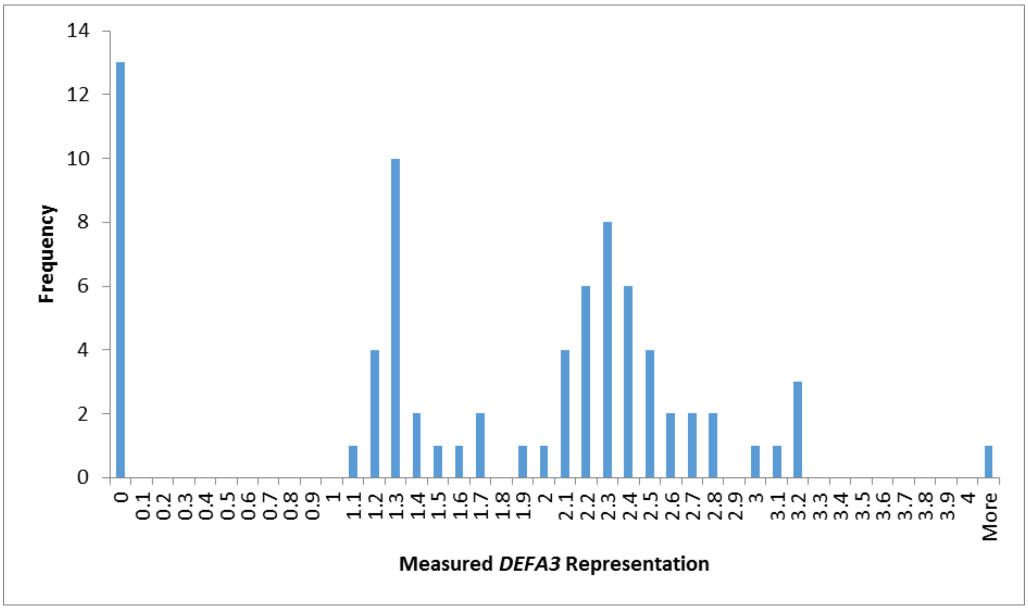


Figure 3.10 – Frequency of DEFA1A3 copies before correction in 76 European samples

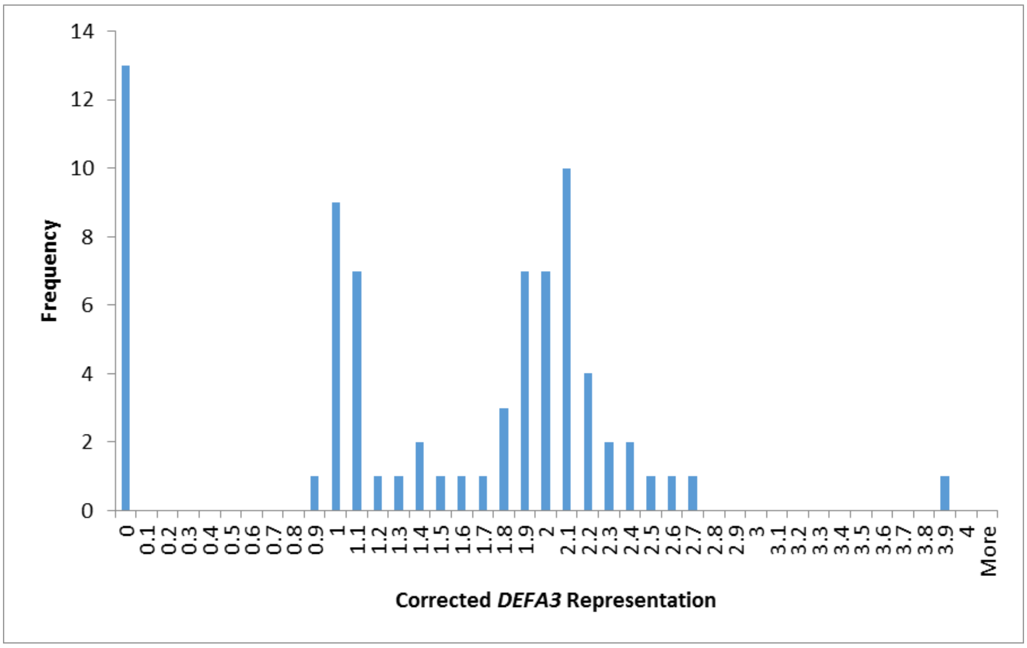


Figure 3.11 – Frequency of DEFA1A3 copies after correction for incompleteness of digestion in 76 European samples

In the case of the Indel5 minor allele in 76 European samples we found up to 4 copies of the deleted variant, results which confirm the observed copy numbers by Khan et al. (Figure 3.12). The clustering of these estimates around integer values suggest that the test is well calibrated.

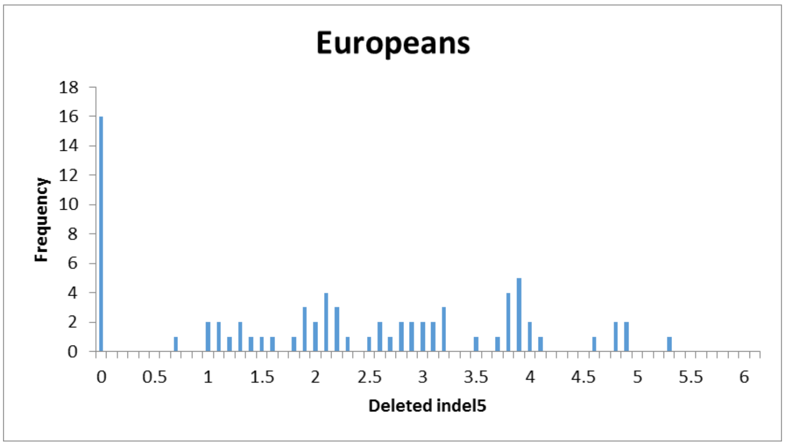


Figure 3.12 - Frequency of deleted Indel5 copies in 76 European samples

In the case of the 4bpdel in 76 European samples we have found up to 3 copies of the deleted variant (Figure 3.13).

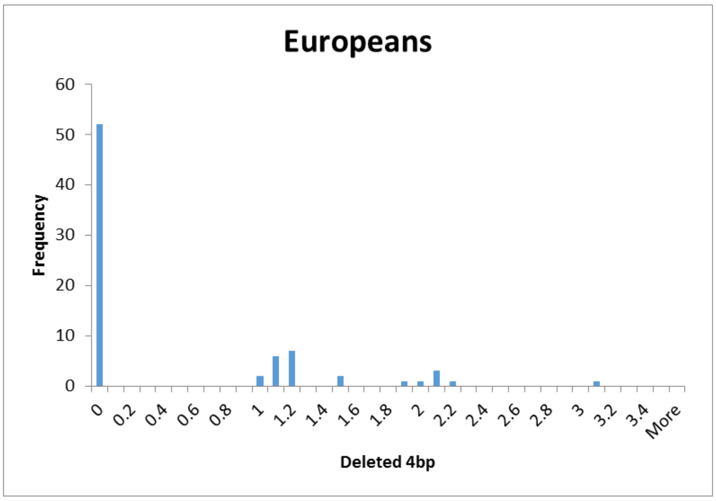


Figure 3.13 - Frequency of deleted 4bp copies in 76 European samples

In the case of 7bpdup in 76 European samples we have found up to 9 copies of the unduplicated variant, with values clustering as expected around integers (Figure 3.14).

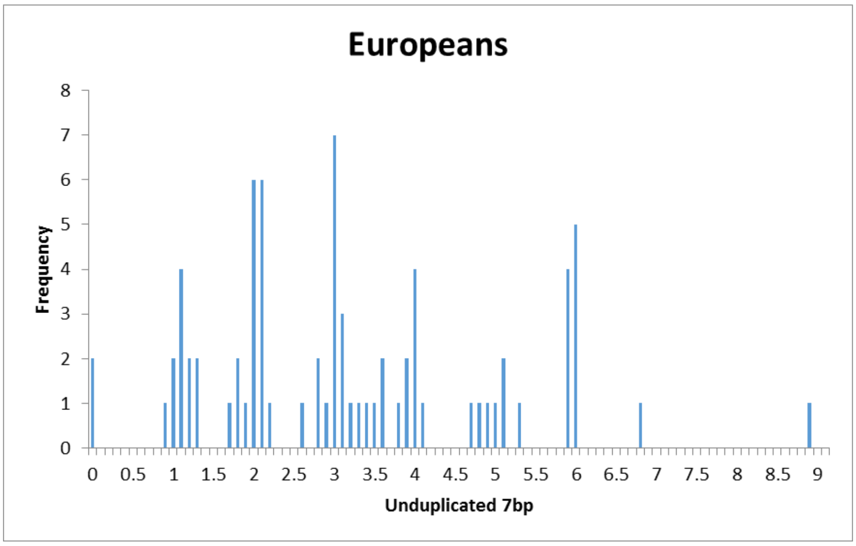


Figure 3.14 - Frequency of unduplicated 7bp copies in 76 European samples

3.2.2 Flanking SNPs and Haplotypes

Four SNPs were genotyped providing haplotype class status for 76 European samples (Black et al. 2014). Combining the genotypes from each flanking SNP assay with total *DEFA1A3* copy numbers in Europeans we confirmed a series of predictions developed by Khan et al. and Black et al. Firstly, our rs4300027 genotype data confirmed the strong associations found by Khan et al. in 2013, which predict that this SNP can be used to distinguish between 3-6 diploid copy number genotype (CC variant) from those that present 6-8 copy number genotype (CT variant) and 8-16 copy number genotype (TT variant) (Figure 3.15, Cochran-Armitage P-value=2.65×10^{-3}). Secondly, our rs7825750 genotype data confirmed the strong associations found by Black et al. in 2014, which predict that Class 2 homozygotes (CC) would have between 4 and 6 copies of DEFA1A3, heterozygotes (TC) between 6 and 8 copies and homozygotes for non Class 2 (TT) 8 and higher copies (Figure 3.15, Cochran-Armitage P-value=2.65×10^{-8}).

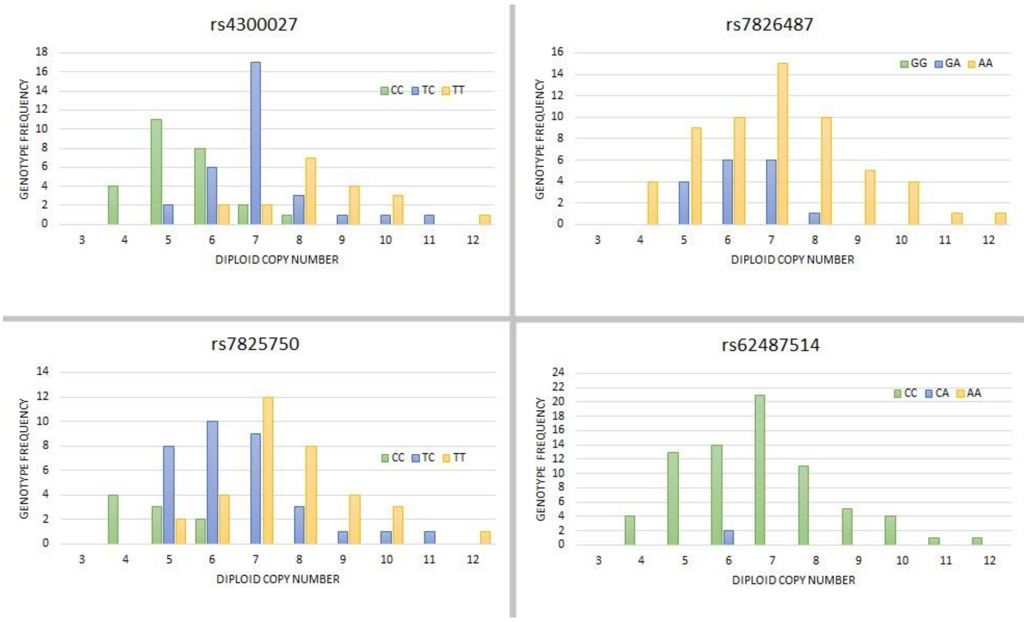


Figure 3.15 – Frequencies of all four SNPs in 76 European samples showing corresponding diploid DEFA1A3 copy numbers. Our data confirms previous predictions of rs4300027 being strongly associated with copy number haplotypes.

Black et al. showed that *DEFA3* has a strong tendency to appear on Reference Sequence or Class 2 haplotypes and to be absent on Class 1, Exchange 1 and Exchange 2 haplotypes. We classified the individuals according to whether they have out of each haplotype 0, 1 or 2+ copies of *DEFA3*. In the case of Reference Sequence and Exchange 2 haplotypes, P-values of 0.111 and 0.264 showed that there appears to be no tendency of *DEFA3* to be either present or absent on these haplotypes. However, in the case of Class 2 haplotype a P-value of $4.73x10^{-3}$ shows there is indeed a significant tendency for *DEFA3* to be present; the same goes for Class 1 and Exchange 1 haplotypes and the tendency of *DEFA3* to be absent (P-values of $9.71x10^{-5}$ and $4.84x10^{-3}$). Based on these results, we then asked if there was a difference in *DEFA3* occupancy in Reference Sequence versus Class 2 haplotypes (Reference Sequence / Class 2 heterozygotes have been excluded); a P-value of $1.47x10^{-2}$ suggests that there is a statistically significant difference and so it looks like Class 2 is a more consistent background for copies of *DEFA3*.

3.2.3 6836028

In preliminary work we have looked at a qualitative assay, for the presence or absence of the 6836028G synonymous variant in exon 2 affecting a valine codon (method described in Section 2.5). Out of 76 samples, 32 samples (42.11%) appeared to present the 6836028G variant.

A quantitative fluorescent assay confirmed our initial findings: 76 DNA samples were typed and analysed using this assay (which will amplify a specific sequence from the gene) allowing us to determine a copy number ratio between 6836028C (130bp) and 6836028G (147/148bp) that can be combined with knowledge of the total copy number to identify the exact copy number of each variant. Digesting the PCR products with *Bsl*I results in two amplicons which differ in size (130bp, 147/148bp); double peaks are due to the closeness of two sites for *BslI* (1bp apart): if the enzyme cuts one site, the other one will be destroyed, and this is done randomly (Figure 3.16).

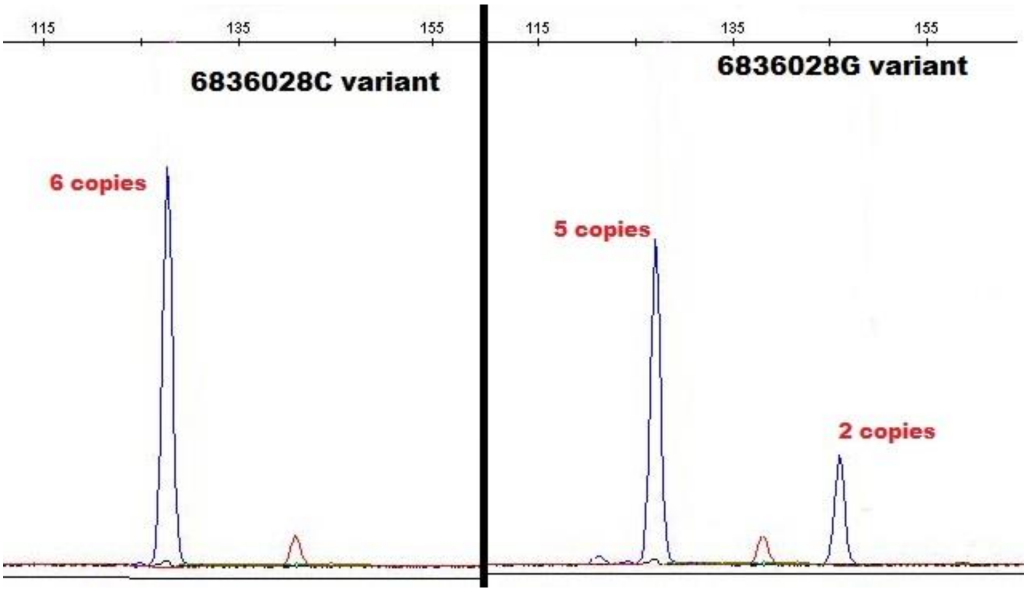


Figure 3.16 - Left: Schematic of predicted PCR products from a 6836028 Ratio Assay trace (sample AN17033, 6 copies, 6:0). Right: Schematic of predicted PCR products from a 6836028 Ratio Assay trace (sample AN17034, 7 copies, 5:2)

The results obtained from this ratio assay have more than one purpose in our study. Firstly, this is a variant that is present also in the cDNA and therefore we can compare both cDNA and genomic DNA. Secondly, we can use the obtained ratios to determine for this polymorphism the exact number of copies containing either the 6836028C variant or the 6836028G variant. Upon getting the results from the 6836028 assay we observed lower than expected ratios by a factor of about 50%, presumably due to failure of *BslI* to digest products completely, possibly due to the generation of single-stranded DNA or heteroduplexes (Figure 3.17). The ratios were corrected to account for this error, allowing us to determine (in combination with the total copy numbers measured by PRT) the number of copies containing 6836028C or 6836028G (Figure 3.18).

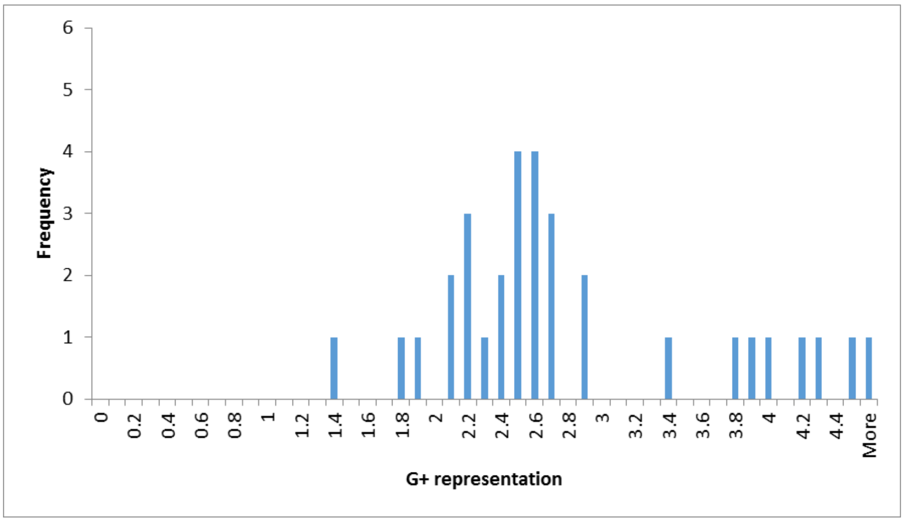


Figure 3.17 - Frequency of 6836028G copies before correction in 32 European samples

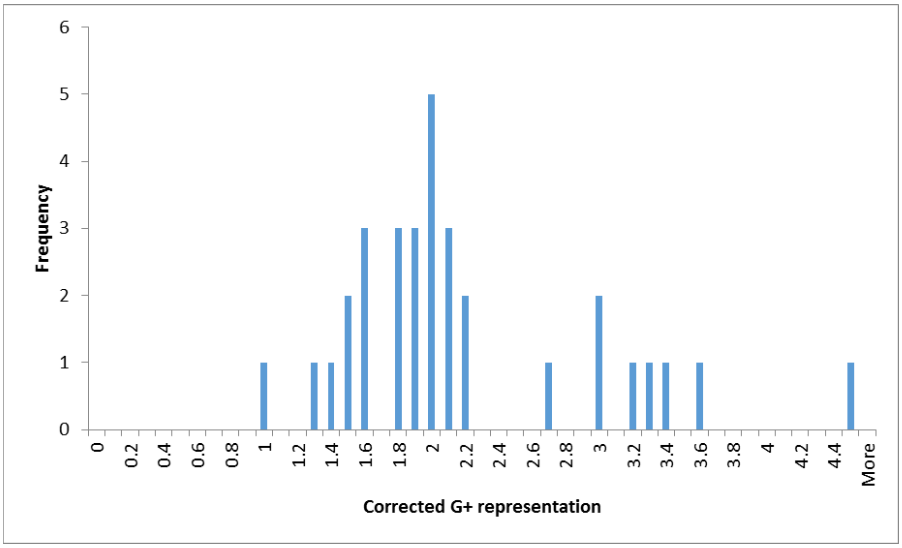


Figure 3.18 - Frequency of 6836028G copies after correction for incompleteness of digestion in 32 European samples

We observed that nearly everyone that has 6836028G variant has at least one Reference Sequence haplotype but not everyone with Reference Sequence haplotype has 6836028G variant, as if 6836028G variant defines a subset of Reference Sequence haplotype. In order to test this hypothesis we compared Reference Sequence heterozygotes with Reference Sequence homozygotes and Non-Reference Sequence homozygotes: a Cochran-Armitage P-value of 1.275×10^{-7} confirms that there is indeed a strong correlation between rs4300027T, representative of a Reference Sequence haplotype, and 6836028G carriers (Figure 3.19).

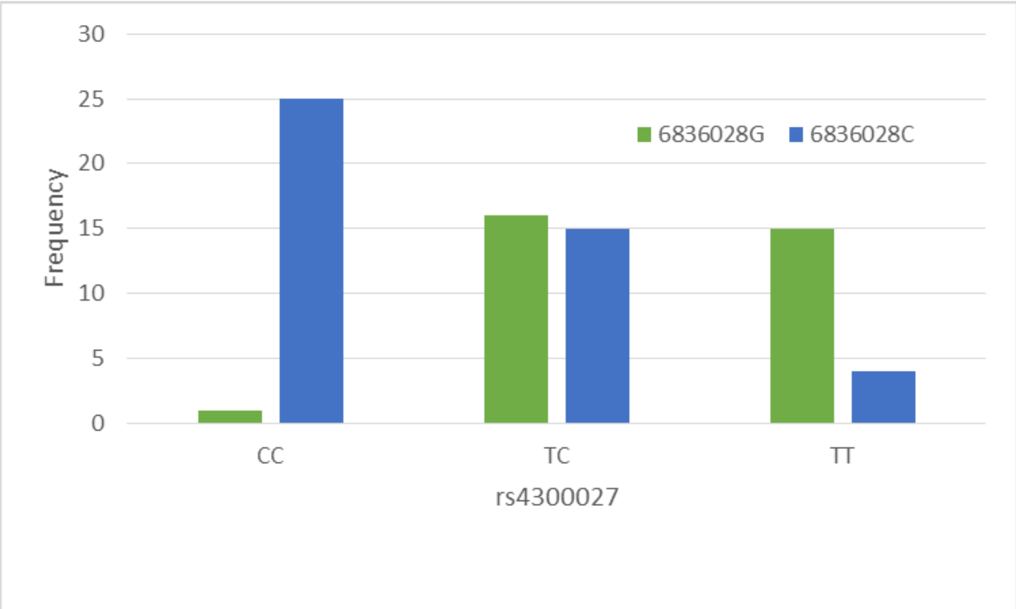


Figure 3.19 - Frequency of 6836028 variant showing corresponding rs4300027 genotype. Of the 32 carriers of the 6836028G only one did not have T. Of the 19 TT homozygotes, 15 were carriers of 6836028G and only 4 did not present the variant.

3.3 cDNA Analysis

3.3.1 *DEFA1A3*

In preliminary work we used a qualitative PCR and *Hae*III digestion assay, for the presence or absence of *DEFA3* in cDNA. Out of 76 samples, 13 samples (20.63%) appeared to lack transcripts from the *DEFA3* gene, confirming our results when analysing genomic DNA with the same assay.

Subsequently we used a quantitative fluorescent assay previously described by Aldred et al. (2005) on cDNA to get ratios which represent the relative levels of expression of *DEFA1* and *DEFA3* mRNA in whole blood. Due to formation of heteroduplexes and single-stranded PCR products, and consequent incomplete digestion by *Hae*III we had to correct the *DEFA1*:*DEFA3* ratios (see Section 3.2.1.3), forcing the *DEFA3* ratios to 0 where the gene is absent. Using regression, a P-value of 9.78×10^{-3} showed there is a significant positive relationship between the genomic ratios and the expression levels of individuals that have both *DEFA1* and *DEFA3*, confirming the results obtained by Aldred et al. (Figure 3.20).

We then asked if these samples have the same relationship between *DEFA1*:*DEFA3* representation in genomic DNA and *DEFA1*:*DEFA3* representation in cDNA. Aldred et al. found that if the *DEFA1*:*DEFA3* ratios were equal in genomic DNA, in cDNA the level of *DEFA1* was about half of *DEFA3*. A best fit line forced through the origin and without outliers gives a gradient of 0.1581 indicating a scaling factor of more than six: if *DEFA1*:*DEFA3* ratios were equal in genomic DNA, in cDNA the level of *DEFA1* would be a at least six times lower than the level of *DEFA3* (Figure 3.20). That suggests that either *DEFA3* is expressed 6 to 7 times more than *DEFA1* per copy, and/or that *DEFA1* and *DEFA3* are synthesised equally from every copy but as promyelocytes mature and become circulating neutrophils, *DEFA3* is more stable than *DEFA1*.

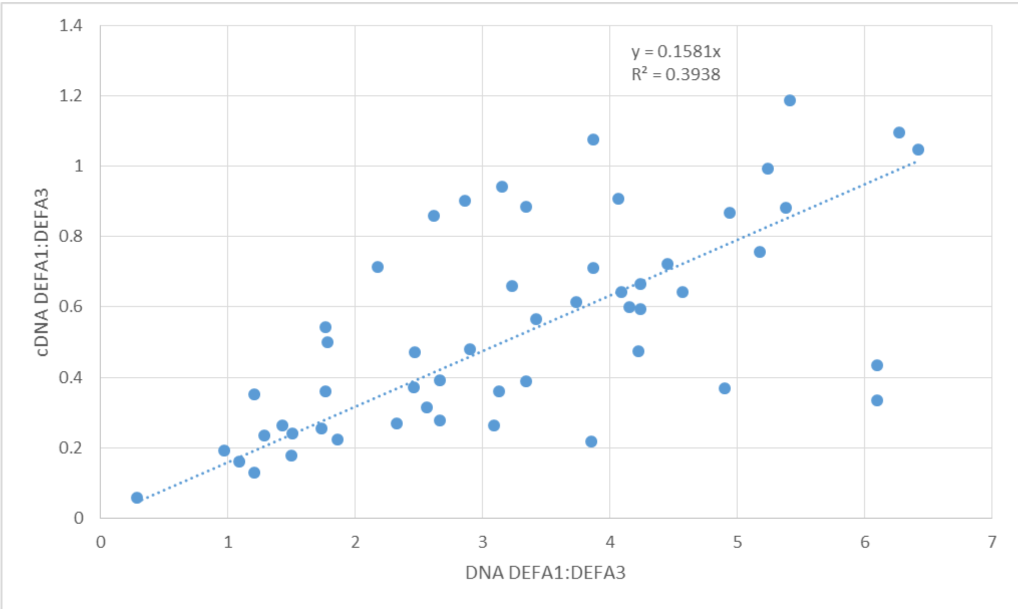


Figure 3.20 - Comparison between genomic DEFA1:DEFA3 and cDNA DEFA1:DEFA3

To ask whether the level of RNA expression in neutrophils might be influenced by flanking haplotype, unpublished work by J.Armour to retrospectively genotype the samples used in Aldred et al. 2005 for flanking variants did not demonstrate any clear patterns, due to a low number of samples. With our dataset of *DEFA1*:*DEFA3* ratios representations in genomic and cDNA, we found no obvious effect that by having one particular haplotype the *DEFA1* levels of expression in cDNA will decrease considerably (Figure 3.21).

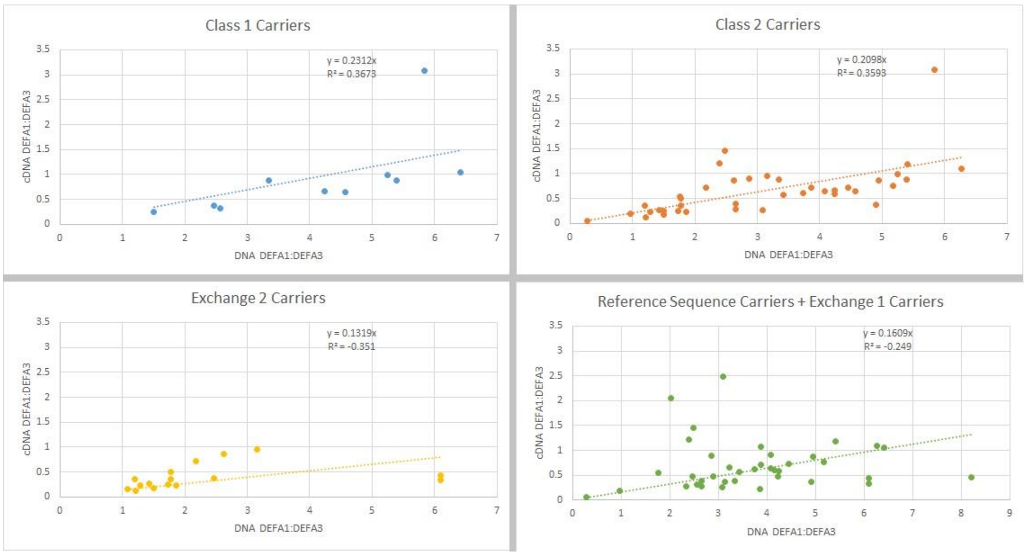


Figure 3.21 – Comparison between DEFA1A3 ratios in genomic and cDNA, divided by haplotype class in 59 European samples (individuals that do not present DEFA3 have been excluded). Our results concerning Exchange 1 haplotype are not consistent with previous reports in Europeans, probably due to random sampling, and so we have combined the Reference Sequence carriers with Exchange 1 carriers.

3.3.2 6836028

In preliminary work we used a qualitative PCR and *Bsl*I digestion assay, for the presence or absence of 6836028G variant in cDNA. Out of 76 samples, 32 samples (42.11%) appeared to have transcripts containing the variant, confirming our results when analysing genomic DNA with the same assay.

Subsequently we used a quantitative fluorescent assay on cDNA to get ratios which represent the relative levels of expression of 6836028G carriers in whole blood. Similarly to the *DEFA1A3* assay, this method uses a restriction enzyme so we had to correct the ratios due to formation of heteroduplexes and single-stranded PCR products, and consequent incomplete digestion by *Bsl*I (Section 3.2.3). Using regression, a P-value of 6.791×10^{-3} showed there is a significant positive relationship between the genomic ratios and the expression levels of 6836028G carriers.

We then asked if our dataset has the same relationship between 6836028G ratios representation in genomic DNA and 6836028G ratios representation in cDNA. A best fit line forced through the origin gives a gradient of 1.4235 indicating a scaling factor of 1.4: in cDNA the level of 6836028G would be about one and a half times lower than the level of 6836028G in genomic DNA (Figure 3.22). That says that 6836028G in genomic DNA is expressed about one and a half times more than 6836028G in cDNA per copy. There was no clear effect of homozygosity or heterozygosity for reference haplotype on 6836028G representation in cDNA (Figure 3.23).

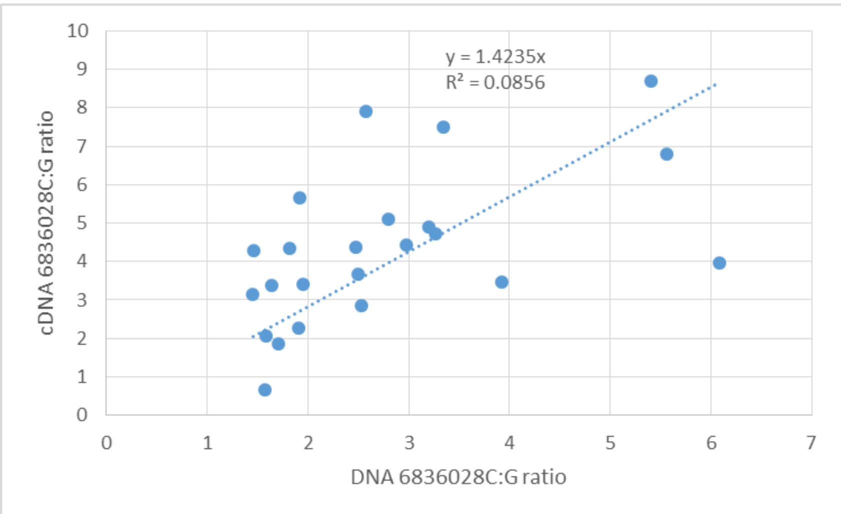


Figure 3.22 - Comparison between genomic 6836028G ratios and cDNA 6836028G ratios in 23 European samples (9 samples failed to produce any sizeable results in cDNA, so these have been excluded)

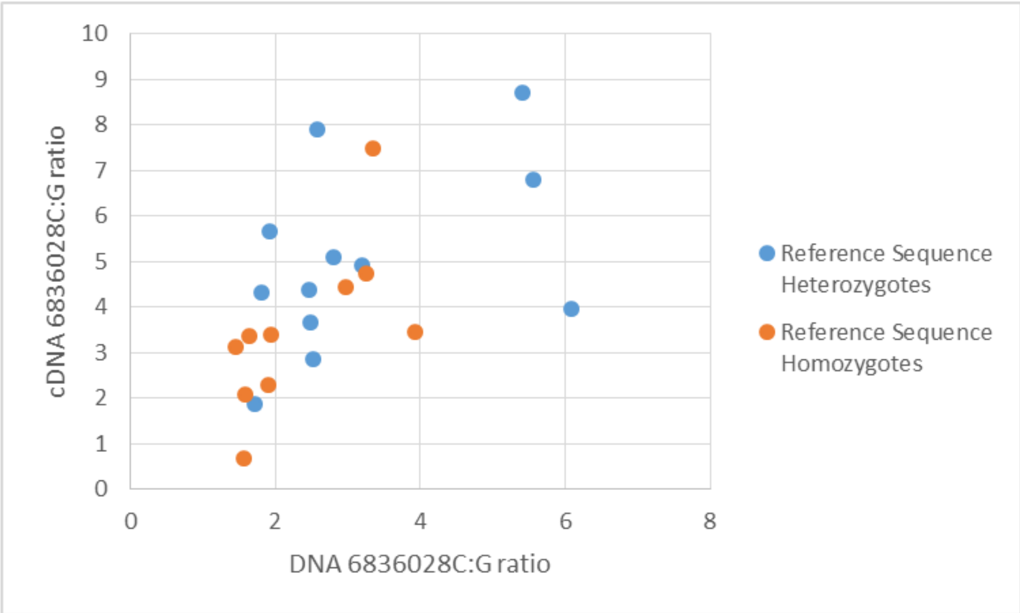


Figure 3.23 - Representation of genomic and cDNA ratios of 6836028G grouped by haplotype

4. DISCUSSION AND CONCLUSION

In order to further investigate the *DEFA1A3* CNV in Europeans we have typed and analysed only 76 willing volunteers, with the restricted numbers mainly due to limitations of the time available for the blood collection period once local ethical approval had been obtained for the study. In association studies larger sample sizes are always better, so replicating these results in a bigger data set (for example, increasing the blood collection phase period to more than 1 month) would be a great starting point.

The protocols used in this study for WBC separation and subsequent DNA and RNA extraction from whole blood have been shown to perform extremely well, but further work involving WBC separation into granulocytes and mononuclear cells and experimenting on both fractions would be the next step in understanding alpha defensin expression in neutrophils in comparison with other cell types.

The results obtained in this study regarding the PRT-based measurement system and the agreement between the four allelic ratios confirm previous statements (Aldred et al. 2005, Khan et al. 2013) and gives even more confidence in the accurate measurement of copy number. Low genomic copy numbers are easy to be sure about (for copy number of 7 or less the minimum ratios obtained are over 10^4), and high copy numbers are hard to be sure about (confidence will decrease with low minimum ratios). However, the main limitation on the genomic data in this study was our inability to put a lot of confidence into the Exchange 1 assay due to our results (2 Exchange 1 carriers out of 76 samples) being inconsistent with what was reported previously. This outcome forced us to analyse these samples together with the Reference Sequence haplotype (in a European population we would expect the frequency of Exchange 1 individuals to be about 15-20% (Black et al. 2014)). Further work on the matter could be to increase the sample size and to distinguish the Exchange 1 and Reference Sequence carriers.

Our analysis confirms the predictions of Black et al 2014, indicating that there is no obvious association between *DEFA1A3* haplotype classes and defensin expression. One possible reason could be the sample size in both studies, thus none or very few homozygotes were observed for each haplotype class and any possible effects were masked by the heterozygous form of the haplotype classes. It is intrinsically difficult to analyse haplotype effects on expression because the same haplotypes can be associated with different copy numbers. For this reason, analysis of repeat types (like *DEFA3* or 6836028G) specific to one haplotype in an individual might act as a proxy for that haplotype in studies of expression.

Comparing the *DEFA1A3* genomic ratios with the *DEFA1A3* levels in cDNA our conclusion differed in detail from the results of Aldred et al 2005: while in the previous study it was shown that if in genomic the amount of *DEFA1* would be equal the amount of *DEFA3*, in cDNA it would be halved, in our study the results are even more extreme, *DEFA3* being expressed in the cDNA 6-7 times more per copy than *DEFA1*. This could have various implications. Firstly, *DEFA3* might be overexpressed in promyelocytes and after an equal amount of decay, what we see in neutrophils is a fair representation of the original levels of expression. Secondly, *DEFA1* and *DEFA3* might be synthesised proportionally but *DEFA3* is more stable and so, when we isolate neutrophils, various levels of gene expression might appear (different experiments might arrive at different conclusions, because of the speed of preparations). Lastly, we have been assuming that all the properties we see from *DEFA1* and *DEFA3* are intrinsic to that type of copy (e.g. every copy of *DEFA3*, whatever its context, will be transcribed at a higher rate and all copies of *DEFA1* will be transcribed at a lower rate). An alternative hypothesis would be that actually Class 2 and Reference Sequence haplotypes are transcribed at a higher frequency than are Class 1, Exchange 1 or Exchange 2; the fact that *DEFA3* is overexpressed is simply because *DEFA3* happens to occupy the haplotypes that are overexpressed. There is nothing, to date, that will reject this hypothesis. What it means is that in addition to being a specific repeat phenomenon, differential expression might actually be a haplotype specific phenomenon, and various complicated modifications could appear (e.g. if the first position of the centromeric end on a Reference Sequence haplotype has a ten fold higher rate of expression than everything else).

Depending on the haplotype composition, we might be able to reject (or say is highly unlikely) a simple haplotype explanation. For example, consider a sample for which 1 copy out of 6 is *DEFA3* and mRNA representation would be equal, the individual being a Reference Sequence / Class 1 heterozygote. The most likely scenario would be Class 1 would have 2 copies and Reference Sequence would have 4 copies, including 1 copy of *DEFA3*. If the increased expression was entirely attributable to the Reference Sequence haplotype, that should also apply to the 3 copies of *DEFA1* on the Reference Sequence haplotype.

For both Class 2 and Reference Sequence backgrounds, they commonly have *DEFA3*. In all examples characterised to date (Black et al. 2014), they have equal or more copies of *DEFA1* than *DEFA3*. If it would apply to the haplotype that *DEFA3* expression is enhanced in *cis*, it should apply to the *DEFA1* copies on that haplotype and there are always equal or more copies of *DEFA1* transcripts than *DEFA3*. In that case there could be a slightly higher mRNA representation of *DEFA3*, but not 6-7 times. In the end, there might be both haplotype and position effects, so that it might not only matter if the gene is *DEFA1* or *DEFA3*, but also where it is.

We assume that the synonymous 6836028G variant might accompany only the *DEFA1* gene, due to low levels of 6836028G in cDNA being potentially associated with low levels of *DEFA1* in cDNA. Unpublished work made by Armour predicts that 6836028G is at the telomeric end on Reference Sequence haplotype, suggesting location within a *DEFA1* gene rather than *DEFA3*. It could be that a recombinant haplotype that starts as a Class 1 haplotype at the centromeric end shows properties of Reference Sequence haplotype at the telomeric end and so the SNPs at the centromeric end might be different from the telomeric end because of crossover events. Further work can be concentrated on the subject and confirm (or disprove) these hypotheses. Analysis of haplotypes in the 1000 Genomes Project (Xu, 2017) suggests that such recombinants between telomeric and centromeric ends of the CNV are quite common.

5. BIBLIOGRAPHY

Aarbiou, J., Ertmann, M., van Wetering, S., van Noort, P., Rook, D., Rabe, K.F., Litvinov, S.V., van Krieken, J.H.J.M., de Boer, W.I., Hiemstra, P.S., 2002. Human neutrophil defensins induce lung epithelial cell proliferation in vitro. Journal of Leukocyte Biology 72, 167–174.

Ai, Z., Li, M., Liu, W., Foo, J.-N., Mansouri, O., Yin, P., Zhou, Q., Tang, X., Dong, X., Feng, S., Xu, R., Zhong, Z., Chen, J., Wan, J., Lou, T., Yu, J., Zhou, Q., Fan, J., Mao, H., Gale, D., Barratt, J., Armour, J.A.L., Liu, J., Yu, X., 2016. Low α-defensin gene copy number increases the risk for IgA nephropathy and renal dysfunction. Science Translational Medicine 8, 345ra88.

Aldhous, M.C., Abu Bakar, S., Prescott, N.J., Palla, R., Soo, K., Mansfield, J.C., Mathew, C.G., Satsangi, J., Armour, J.A.L., 2010. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. Human Molecular Genetics 19, 4930–4938.

Aldred, P.M.R., 2004. Variation in Human Immune Response Genes (PhD Thesis). University of Nottingham.

Aldred, P.M.R., Hollox, E.J., Armour, J.A.L., 2005. Copy number polymorphism and expression level variation of the human α-defensin genes DEFA1 and DEFA3. Human Molecular Genetics 14, 2045–2052.

Anderson, J.M., Rodriguez, A., Chang, D.T., 2008. FOREIGN BODY REACTION TO BIOMATERIALS. Semin Immunol 20, 86–100.

Andiappan, A.K., Melchiotti, R., Poh, T.Y., Nah, M., Puan, K.J., Vigano, E., Haase, D., Yusof, N., San Luis, B., Lum, J., Kumar, D., Foo, S., Zhuang, L., Vasudev, A., Irwanto, A., Lee, B., Nardin, A., Liu, H., Zhang, F., Connolly, J., Liu, J., Mortellaro, A., Wang, D.Y., Poidinger, M., Larbi, A., Zolezzi, F., Rotzschke, O., 2015. Genome-wide analysis of the genetic regulation of gene expression in human neutrophils. Nature Communications 6, 7971.

Armour, J.A.L., Palla, R., Zeeuwen, P.L.J.M., Heijer, M. d., Schalkwijk, J., Hollox, E.J., 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Research 35, e19–e19.

Armour, J.A.L., Sismani, C., Patsalis, P.C., Cross, G., 2000. Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res 28, 605–609.

Bae, H.-B., Zmijewski, J.W., Deshane, J.S., Zhi, D., Thompson, L.C., Peterson, C.B., Chaplin, D.D., Abraham, E., 2012. Vitronectin Inhibits Neutrophil Apoptosis through Activation of Integrin-Associated Signaling Pathways. Am J Respir Cell Mol Biol 46, 790–796.

Bakar, S.A., Hollox, E.J., Armour, J.A.L., 2009. Allelic recombination between distinct genomic locations generates copy number diversity in human β-defensins. Proc Natl Acad Sci U S A 106, 853-858.

Ballana, E., González, J.R., Bosch, N., Estivill, X., 2007. Inter-population variability of DEFA3 gene absence: correlation with haplotype structure and population variability. BMC Genomics 8, 14.

Barber, J.C., Joyce, C.A., Collinson, M.N., Nicholson, J.C., Willatt, L.R., Dyson, H.M., Bateman, M.S., Green, A.J., Yates, J.R., Dennis, N.R., 1998. Duplication of 8p23.1: a cytogenetic anomaly with no established clinical significance. J Med Genet 35, 491–496.
Beckmann, J., Estivill, X., E Antonarakis, S., 2007. Beckmann, JS, Estivill, X & Antonarakis, SE Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat. Rev. Genet. 8, 639-646.
Beerman, I., Novak, J., Wyatt, R.J., Julian, B.A., Gharavi, A.G., 2007. The genetics of IgA nephropathy. Nature Clinical Practice Nephrology 3, 325–338.
Björnsdottir, H., Welin, A., Michaëlsson, E., Osla, V., Berg, S., Christenson, K., Sundqvist, M., Dahlgren, C., Karlsson, A., Bylund, J., 2015. Neutrophil NET formation is regulated from the inside by myeloperoxidase-processed reactive oxygen species. Free Radical Biology and Medicine 89, 1024–1035.
Black, H.A., 2014. Allelic structures and mechanisms of copy number change at the human DEFA1A3 copy number variable locus (PhD Thesis). University of Nottingham.
Black, H.A., Khan, F.F., Tyson, J., Armour, J.A., 2014. Inferring mechanisms of copy number change from haplotype structures at the human DEFA1A3 locus. BMC genomics 15, 614.
Blumenreich, M.S., 1990. The white blood cell and differential count. Clinical Methods: The History, Physical, and Laboratory Examinations [Internet]. 3rd ed. Boston: Butterworths.
Böhm, I., Speck, U., Schild, H.H., 2011. Pilot study on basophil activation induced by contrast medium: CM basophil degranulation measured by flowcytometry. Fundamental & Clinical Pharmacology 25, 267–276.
Branzk, N., Lubojemska, A., Hardison, S.E., Wang, Q., Gutierrez, M.G., Brown, G.D., Papayannopoulos, V., 2014. Neutrophils sense microbial size and selectively release neutrophil extracellular traps in response to large pathogens. Nat Immunol 15, 1017–1025.
Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., Zychlinsky, A., 2004. Neutrophil extracellular traps kill bacteria. Science 303, 1532–1535.
Brinkmann, V., Zychlinsky, A., 2012. Neutrophil extracellular traps: Is immunity the second function of chromatin? J Cell Biol 198, 773–783.
Brook, M., Tomlinson, G.H., Miles, K., Smith, R.W.P., Rossi, A.G., Hiemstra, P.S., van ’t Wout, E.F.A., Dean, J.L.E., Gray, N.K., Lu, W., Gray, M., 2016. Neutrophil-derived alpha defensins control inflammation by inhibiting macrophage mRNA translation. Proceedings of the National Academy of Sciences 113, 4350–4355.
Campean, V., Karpe, B., Haas, C., Atalla, A., Peters, H., Rupprecht, H., Liebner, S., Acker, T., Plate, K., Amann, K., 2008. Angiopoietin 1 and 2 gene and protein expression is differentially regulated in acute anti-Thy1.1 glomerulonephritis. AJP: Renal Physiology 294, F1174-F1184.
Cantsilieris, S., Baird, P.N., White, S.J., 2013. Molecular methods for genotyping complex copy number polymorphisms. Genomics 101, 86-93.

Cantsilieris, S., Western, P.S., Baird, P.N., White, S.J., 2014. Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. BMC genomics 15, 329.
Carpenter, D., McIntosh, R.S., Pleass, R.J., Armour, J.A.L., 2012. Functional effects of CCL3L1 copy number. Genes and Immunity 13, 374–379.
Chen, Q., Hakimi, M., Wu, S., Jin, Y., Cheng, B., Wang, H., Xie, G., Ganz, T., Linzmeier, R.M., Fang, X., 2010. Increased genomic copy number of DEFA1/DEFA3 is associated with susceptibility to severe sepsis in Chinese Han population. Anesthesiology: The Journal of the American Society of Anesthesiologists 112, 1428–1434.
Cheng, F., Zhou, X., Zhao, Y., Zhao, M., Zhang, H., 2013. Alpha-defensin DEFA1A3 gene copy number variation in Asians and its genetic association study in Chinese systemic lupus erythematosus patients. Gene 517, 158–163.
Chertov, O., Yang, D., Howard, O.M., Oppenheim, J.J., 2000. Leukocyte granule proteins mobilize innate host defenses and adaptive immune responses. Immunological reviews 177, 68–78.
Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., Montgomery, S.B., Battle, A., Conrad, D.F., Hall, I.M., 2017. The impact of structural variation on human gene expression. Nature Genetics 49, 692–699.
Chistiakov, D.A., Bobryshev, Y.V., Orekhov, A.N., 2015. Neutrophil’s weapons in atherosclerosis. Experimental and Molecular Pathology 99, 663–671.
Chung, W.O., Dale, B.A., 2004. Innate Immune Response of Oral and Foreskin Keratinocytes: Utilization of Different Signaling Pathways by Various Bacterial Species. Infect Immun 72, 352–358.
Chung, W.O., Dommisch, H., Yin, L., Dale, B.A., 2007. Expression of defensins in gingiva and their role in periodontal health and disease. Current pharmaceutical design 13, 3073-3083.
Consortium, Gtex, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348, 648–660.
Craddock, N., Hurles, M., Cardin, N., Pearson, R., Plagnol, V., Robson, S., 2010. Genome-wide association study of copy number variation in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464, 713–720.
Cunliffe, R., Rose, F., Keyte, J., Abberley, L., Chan, W., Mahida, Y., 2001. Human defensin 5 is stored in precursor form in normal Paneth cells and is expressed by some villous epithelial cells and by metaplastic Paneth cells in the colon in inflammatory bowel disease. Gut 48, 176-185.
da Silva, F.M., Massart-Leën, A.M., Burvenich, C., 1994. Development and maturation of neutrophils. Veterinary Quarterly 16, 220–225.
Daher, K.A., Selsted, M.E., Lehrer, R.I., 1986. Direct inactivation of viruses by human granulocyte defensins. J Virol 60, 1068–1074.
den Dunnen, J.T., White, S.J., 2006. MLPA and MAPH: sensitive detection of deletions and duplications. Current protocols in human genetics 7-14.

Dommisch H, Acil Y, Dunsche A, Winter J, Jepsen S. Differential gene expression of human beta-defensins (hBD-1, -2, -3) in inflammatory gingival diseases. Oral Microbiol Immunol 2005; 20: 186-90.
Dunsche, A., Açil, Y., Dommisch, H., Siebert, R., Schröder, J.-M., Jepsen, S., 2002. The novel human beta-defensin-3 is widely expressed in oral tissues. European journal of oral sciences 110, 121–124.
Ear, T., McDonald, P.P., 2008. Cytokine generation, promoter activation, and oxidant-independent NF-κB activation in a transfectable human neutrophilic cellular model. BMC Immunology 9, 14.
Eicher, J.D., Landowski, C., Stackhouse, B., Sloan, A., Chen, W., Jensen, N., Lien, J.-P., Leslie, R., Johnson, A.D., 2015. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. Nucleic Acids Research 43, D799–D804.
Eiserich, J.P., Hristova, M., Cross, C.E., Jones, A.D., Freeman, B.A., Halliwell, B., van der Vliet, A., 1998. Formation of nitric oxide-derived inflammatory oxidants by myeloperoxidase in neutrophils. Nature 391, 393.
Ericksen, B., Wu, Z., Lu, W., Lehrer, R.I., 2005. Antibacterial Activity and Specificity of the Six Human α-Defensins. Antimicrob Agents Chemother 49, 269–275.
Fabris, L., Cadamuro, M., Fiorotto, R., Roskams, T., Spirlì, C., Melero, S., Sonzogni, A., Joplin, R.E., Okolicsanyi, L., Strazzabosco, M., 2006. Effects of angiogenic factor overexpression by human and rodent cholangiocytes in polycystic liver diseases. Hepatology 43, 1001–1012.
Faurschou, M., Borregaard, N., 2003. Neutrophil granules and secretory vesicles in inflammation. Microbes and Infection, Forum in Immunology on neutrophils 5, 1317–1327.
Faurschou, M., Sørensen, O.E., Johnsen, A.H., Askaa, J., Borregaard, N., 2002. Defensin-rich granules of human neutrophils: characterization of secretory properties. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1591, 29–35.
Feucht, E.C., DeSanti, C.L., Weinberg, A., 2003. Selective induction of human beta-defensin mRNAs by Actinobacillus actinomycetemcomitans in primary and immortalized oral epithelial cells. Molecular Oral Microbiology 18, 359–363.
Franco, R.S., 2012. Measurement of Red Cell Lifespan and Aging. Transfusion Medicine and Hemotherapy 39, 302–307.
Fuchs, T.A., Abed, U., Goosmann, C., Hurwitz, R., Schulze, I., Wahn, V., Weinrauch, Y., Brinkmann, V., Zychlinsky, A., 2007. Novel cell death program leads to neutrophil extracellular traps. J Cell Biol 176, 231–241.
Fujii, G., Selsted, M.E., Eisenberg, D., 1993. Defensins promote fusion and lysis of negatively charged membranes. Protein Sci 2, 1301–1312.
Gale, D.P., Molyneux, K., Wimbury, D., Higgins, P., Levine, A.P., Caplin, B., Ferlin, A., Yin, P., Nelson, C.P., Stanescu, H., Samani, N.J., Kleta, R., Yu, X., Barratt, J., 2017. Galactosylation of IgA1 Is Associated with Common Variation in *C1GALT1*. Journal of the American Society of Nephrology 28, 2158–2166.
Ganz, T., Selsted, M.E., Lehrer, R.I., 1990. Defensins. European Journal of Haematology 44, 1–8.

Ganz, T., Selsted, M.E., Szklarek, D., Harwig, S.S., Daher, K., Bainton, D.F., Lehrer, R.I., 1985. Defensins. Natural peptide antibiotics of human neutrophils. J Clin Invest 76, 1427–1435.
Garcia, A.E., Osapay, G., Tran, P.A., Yuan, J., Selsted, M.E., 2008. Isolation, Synthesis, and Antimicrobial Activities of Naturally Occurring -Defensin Isoforms from Baboon Leukocytes. Infection and Immunity 76, 5883–5891.
Grigat, J., Soruri, A., Forssmann, U., Riggert, J., Zwirner, J., 2007. Chemoattraction of Macrophages, T Lymphocytes, and Mast Cells Is Evolutionarily Conserved within the Human -Defensin Family. The Journal of Immunology 179, 3958–3965.
Hazlett, L., Wu, M., 2011. Defensins in innate immunity. Cell and Tissue Research 343, 175–188.
Hickey, M.J., Kubes, P., 2009. Intravascular immunity: the host–pathogen encounter in blood vessels. Nature Reviews Immunology 9, 364–375.
Hogervorst, F.B., Nederlof, P.M., Gille, J.J., McElgunn, C.J., Grippeling, M., Pruntel, R., Regnerus, R., van Welsem, T., van Spaendonk, R., Menko, F.H., 2003. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer research 63, 1449–1453.
Hollox, E., Atia, T., Cross, G., Parkin, T., Armour, J., 2002. High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). J Med Genet 39, 790–795.
Hollox, E.J., Armour, J.A.L., Barber, J.C.K., 2003. Extensive Normal Copy Number Variation of a ß-Defensin Antimicrobial-Gene Cluster. Am J Hum Genet 73, 591–600.
Hollox, E.J., Davies, J., Griesenbach, U., Burgess, J., Alton, E.W., Armour, J.A., 2005. Beta-defensin genomic copy number is not a modifier locus for cystic fibrosis. Journal of Negative Results in BioMedicine 4.
Hollox, E.J., Detering, J.-C., Dehnugara, T., 2009. An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. Hum Mutat 30, 477–484.
Hollox, E.J., Huffmeier, U., Zeeuwen, P.L.J.M., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C.M., Traupe, H., de Jongh, G., Heijer, M. den, Reis, A., Armour, J.A.L., Schalkwijk, J., 2008. Psoriasis is associated with increased ß-defensin genomic copy number. Nature Genetics 40, 23–25.
Horn, M., Bertling, A., Brodde, M.F., MüLler, A., Roth, J., Van Aken, H., Jurk, K., Heilmann, C., Peters, G., Kehrel, B.E., 2012. Human neutrophil alpha-defensins induce formation of fibrinogen and thrombospondin-1 amyloid-like structures and activate platelets via glycoprotein IIb/IIIa: Alpha-defensins activate platelets. Journal of Thrombosis and Haemostasis 10, 647–661.
Hughes, A.L., 1999. Evolutionary diversification of the mammalian defensins. Cellular and Molecular Life Sciences CMLS 56, 94–103.
Jespersgaard, C., Fode, P., Dybdahl, M., Vind, I., Nielsen, O., Csillag, C., Munkholm, P., Vainer, B., Riis, L., Elkjaer, M., Pedersen, N., Knudsen, E., Skytt Andersen, P., 2011. Alpha-Defensin DEFA1A3 Gene Copy Number Elevation in Danish Crohn’s Disease Patients.

Jiang, K., Sun, X., Chen, Y., Shen, Y., Jarvis, J.N., 2015. RNA sequencing from human neutrophils reveals distinct transcriptional differences associated with chronic inflammatory states. BMC Medical Genomics 8.

Kagan, B.L., Selsted, M.E., Ganz, T., Lehrer, R.I., 1990. Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. Proc Natl Acad Sci U S A 87, 210–214.

Khan, F.F., 2012. Characterization of the alpha defensin copy number variation in humans (PhD Thesis). University of Nottingham.

Khan, F.F., Carpenter, D., Mitchell, L., Mansouri, O., Black, H.A., Tyson, J., Armour, J.A., 2013. Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. BMC genomics 14, 719.

Khurana, I., 2014. Medical Physiology for Undergraduate Students - E-book. Elsevier Health Sciences.

Kimura, H., 2004. Angiopoietin/tie receptors system may play a role during reconstruction and capillarization of the hepatic sinusoids after partial hepatectomy and liver necrosis in rats. Hepatology Research 29, 51–59.

Kiryluk, K., Novak, J., 2014. The genetics and immunobiology of IgA nephropathy. Journal of Clinical Investigation 124, 2325–2332.

Köttgen, A., Pattaro, C., Böger, C.A., Fuchsberger, C., Olden, M., Glazer, N.L., Parsa, A., Gao, X., Yang, Q., Smith, A.V., O'Connell, J.R., Li, M., Schmidt, H., Tanaka, T., Isaacs, A., Ketkar, S., Hwang, S.-J., Johnson, A.D., Dehghan, A., Teumer, A., Paré, G., Atkinson, E.J., Zeller, T., Lohman, K., Cornelis, M.C., Probst-Hensch, N.M., Kronenberg, F., Tönjes, A., Hayward, C., Aspelund, T., Eiriksdottir, G., Launer, L.J., Harris, T.B., Rampersaud, E., Mitchell, B.D., Arking, D.E., Boerwinkle, E., Struchalin, M., Cavalieri, M., Singleton, A., Giallauria, F., Metter, J., de Boer, I.H., Haritunians, T., Lumley, T., Siscovick, D., Psaty, B.M., Zillikens, M.C., Oostra, B.A., Feitosa, M., Province, M., de Andrade, M., Turner, S.T., Schillert, A., Ziegler, A., Wild, P.S., Schnabel, R.B., Wilde, S., Munzel, T.F., Leak, T.S., Illig, T., Klopp, N., Meisinger, C., Wichmann, H.-E., Koenig, W., Zgaga, L., Zemunik, T., Kolcic, I., Minelli, C., Hu, F.B., Johansson, Å., Igl, W., Zaboli, G., Wild, S.H., Wright, A.F., Campbell, H., Ellinghaus, D., Schreiber, S., Aulchenko, Y.S., Felix, J.F., Rivadeneira, F., Uitterlinden, A.G., Hofman, A., Imboden, M., Nitsch, D., Brandstätter, A., Kollerits, B., Kedenko, L., Mägi, R., Stumvoll, M., Kovacs, P., Boban, M., Campbell, S., Endlich, K., Völzke, H., Kroemer, H.K., Nauck, M., Völker, U., Polasek, O., Vitart, V., Badola, S., Parker, A.N., Ridker, P.M., Kardia, S.L.R., Blankenberg, S., Liu, Y., Curhan, G.C., Franke, A., Rochat, T., Paulweber, B., Prokopenko, I., Wang, W., Gudnason, V., Shuldiner, A.R., Coresh, J., Schmidt, R., Ferrucci, L., Shlipak, M.G., van Duijn, C.M., Borecki, I., Krämer, B.K., Rudan, I., Gyllensten, U., Wilson, J.F., Witteman, J.C., Pramstaller, P.P., Rettig, R., Hastie, N., Chasman, D.I., Kao, W.H., Heid, I.M., Fox, C.S., 2010. New loci associated with kidney function and chronic kidney disease. Nature Genetics 42, 376–384.

Krisanaprakornkit, S., Kimball, J.R., Dale, B.A., 2002. Regulation of Human -Defensin-2 in Gingival Epithelial Cells: The Involvement of

Mitogen-Activated Protein Kinase Pathways, But Not the NF- B Transcription Factor Family. The Journal of Immunology 168, 316–324.
Krisanaprakornkit, S., Kimball, J.R., Weinberg, A., Darveau, R.P., Bainbridge, B.W., Dale, B.A., 2000. Inducible Expression of Human β-Defensin 2 by Fusobacterium nucleatum in Oral Epithelial Cells: Multiple Signalling Pathways and Role of Commensal Bacteria in Innate Immunity and the Epithelial Barrier. Infect Immun 68, 2907–2915.
Kumar, V., Sharma, A., 2010. Neutrophils: Cinderella of innate immune system. Int. Immunopharmacol. 10, 1325–1334.
LaFleur-Brooks, M., 2008. Exploring Medical Language: A Student-Directed Approach. Mosby Elsevier.
Laki, K., 1972. Our ancient heritage in blood clotting and some of its consequences. Annals of the New York Academy of Sciences 202, 297–307.
Ledderose, C., Heyn, J., Limbeck, E., Kreth, S., 2011. Selection of reliable reference genes for quantitative real-time PCR in human T cells and neutrophils. BMC research notes 4, 427.
Lee, B., Doranz, B.J., Rana, S., Yi, Y., Mellado, M., Frade, J.M.R., Martinez-A., C., O'Brien, S.J., Dean, M., Collman, R.G., Doms, R.W., 1998. Influence of the CCR2-V64I Polymorphism on Human Immunodeficiency Virus Type 1 Coreceptor Activity and on Chemokine Receptor Function of CCR2b, CCR3, CCR5, and CXCR4. J Virol 72, 7450–7458.
Lehrer, R.I., Ganz, T., Szklarek, D., Selsted, M.E., 1988. Modulation of the in vitro candidacidal activity of human neutrophil defensins by target cell metabolism and divalent cations. J Clin Invest 81, 1829–1835.
Li, M., Foo, J.-N., Wang, J.-Q., Low, H.-Q., Tang, X.-Q., Toh, K.-Y., Yin, P.-R., Khor, C.-C., Goh, Y.-F., Irwan, I.D., Xu, R.-C., Andiappan, A.K., Bei, J.-X., Rotzschke, O., Chen, M.-H., Cheng, C.-Y., Sun, L.-D., Jiang, G.-R., Wong, T.-Y., Lin, H.-L., Aung, T., Liao, Y.-H., Saw, S.-M., Ye, K., Ebstein, R.P., Chen, Q.-K., Shi, W., Chew, S.-H., Chen, J., Zhang, F.-R., Li, S.-P., Xu, G., Shyong Tai, E., Wang, L., Chen, N., Zhang, X.-J., Zeng, Y.-X., Zhang, H., Liu, Z.-H., Yu, X.-Q., Liu, J.-J., 2015. Identification of new susceptibility loci for IgA nephropathy in Han Chinese. Nature Communications 6, 7270.
Lillard, J.W., Boyaka, P.N., Chertov, O., Oppenheim, J.J., McGhee, J.R., 1999. Mechanisms for induction of acquired host immunity by neutrophil peptide defensins. Proc Natl Acad Sci U S A 96, 651–656.
Linzmeier, R., Ho, C.H., Hoang, B.V., Ganz, T., 1999. A 450-kb contig of defensin genes on human chromosome 8p23. Gene 233, 205–211.
Linzmeier, R.M., Ganz, T., 2005. Human defensin gene copy number polymorphisms: Comprehensive analysis of independent variation in α- and β-defensin regions at 8p22–p23. Genomics 86, 423–430.
Lu, Q., Jin, L., Darveau, R.P., Samaranayake, L.P., 2004. Expression of human beta-defensins-1 and -2 peptides in unresolved chronic periodontitis. Journal of Periodontal Research 39, 221–227.
Lu, Q., Samaranayake, L.P., Darveau, R.P., Jin, L., 2005. Expression of human β-defensin-3 in gingival epithelia. Journal of Periodontal Research 40, 474–481.

Ma, X., Lu, Y.P., Yang, L., Song, J., Luo, G.H., Shi, Y.J., Li, Y.P., 2008. Expressions of Angiopoietin-1, Angiopoietin-2, and Tie2 and Their Roles in Rat Renal Allografts with Chronic Allograft Nephropathy. Transplantation Proceedings 40, 2795–2799.

Machlus, K.R., Thon, J.N., Italiano, J.E., 2014. Interpreting the developmental dance of the megakaryocyte: a review of the cellular and molecular processes mediating platelet formation. British Journal of Haematology 165, 227–236.

Mantovani, A., Cassatella, M.A., Costantini, C., Jaillon, S., 2011. Neutrophils in the activation and regulation of innate and adaptive immunity. Nat. Rev. Immunol. 11, 519–531.

Mantripragada, K.K., Buckley, P.G., Diaz de Ståhl, T., Dumanski, J.P., 2004. Genomic microarrays in the spotlight. Trends in Genetics 20, 87–94.

Mars, W.M., Patmasiriwat, P., Maity, T., Huff, V., Weil, M.M., Saunders, G.F., 1995. Inheritance of unequal numbers of the genes encoding the human neutrophil defensins HP-1 and HP-3. Journal of Biological Chemistry 270, 30371–30376.

Maton, A., Prentice-Hall, inc, 1997. Human biology and health. Prentice Hall.

Mestecky, J., Novak, J., Moldoveanu, Z., Raska, M., 2016. IgA nephropathy enigma. Clinical Immunology 172, 72–77.

Mihăescu, G., 2001. Imunologie si imunochimie. Editura Universitatii din Bucuresti, Bucuresti.

Morrison, G.M., 2003. Signal Sequence Conservation and Mature Peptide Divergence within Subgroups of the Murine beta-Defensin Gene Family. Molecular Biology and Evolution 20, 460–470.

Mukai, K., Galli, S., 2001. Basophils, in: ELS. John Wiley & Sons, Ltd.

Munz, M., Willenborg, C., Richter, G.M., Jockel-Schneider, Y., Graetz, C., Staufenbiel, I., Wellmann, J., Berger, K., Krone, B., Hoffmann, P., 2017. A genome-wide association study identifies nucleotide variants at SIGLEC5 and DEFA1A3 as risk loci for periodontitis. Human Molecular Genetics 26, 2577–2588.

Naranbhai, V., Fairfax, B.P., Makino, S., Humburg, P., Wong, D., Ng, E., Hill, A.V.S., Knight, J.C., 2015. Genomic modulators of gene expression in human neutrophils. Nature Communications 6, 7545.

Nuytten, H., Wlodarska, I., Nackaerts, K., Vermeire, S., Vermeesch, J., Cassiman, J.-J., Cuppens, H., 2009. Accurate determination of copy number variations (CNVs): Application to the α- and β-defensin CNVs. Journal of Immunological Methods 344, 35–44. Orkin, S.H., Zon, L.I., 2008. SnapShot: Hematopoiesis. Cell 132, 712.e1-712.e2.

Papayannopoulos, V., Zychlinsky, A., 2009. NETs: a new strategy for using old weapons. Trends in Immunology 30, 513-521. Patil, A., Hughes, A.L., Zhang, G., 2004. Rapid evolution and diversification of mammalian α-defensins as revealed by comparative analysis of rodent and primate genes. Physiological Genomics 20, 1–11.

Qi, Y., Zhou, X., Cheng, F., Zhang, H., 2016. Elevated Plasma *α* -Defensins (HNP1–3) Levels Correlated with IgA1 Glycosylation and Susceptibility to IgA Nephropathy. Disease Markers 2016, 1–7.

Ranzani, V., Rossetti, G., Panzeri, I., Arrigoni, A., Bonnal, R.J.P., Curti, S., Gruarin, P., Provasi, E., Sugliano, E., Marconi, M., De Francesco, R., Geginat, J., Bodega, B., Abrignani, S., Pagani, M., 2015. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. Nature Immunology 16, 318–325.

Rice, W., Ganz, T., Kinkade, J. J., Selsted, M., Lehrer, R., & Parmley, R., 1987. Defensin-rich dense granules of human neutrophils. Blood, 70(3), 757-765.

Rodrigues, J.C., Haas, M., Reich, H.N., 2017. IgA Nephropathy. Clinical Journal of the American Society of Nephrology 12, 677–686.

Rosenberg, C., Navajas, L., Vagenas, D.F., Bakker, E., Vainzof, M., Passos-Bueno, M.R., Takata, R.I., Van Ommen, G.J.B., Zatz, M., Den Dunnen, J.T., 1998. Clinical diagnosis of heterozygous dystrophin gene deletions by fluorescence in situ hybridization. Neuromuscular Disorders 8, 447–452.

Rothenberg, M.E., Hogan, S.P., 2006. The eosinophil. Annual review of immunology 24.

Schibli, D.J., Hunter, H.N., Aseyev, V., Starner, T.D., Wiencek, J.M., McCray, P.B., Tack, B.F., Vogel, H.J., 2002. The Solution Structures of the Human β-Defensins Lead to a Better Understanding of the Potent Bactericidal Activity of HBD3 against *Staphylococcus aureus*. Journal of Biological Chemistry 277, 8279–8289.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G., 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30, e57.

Schutte, B.C., McCray, P.B., 2002. β-Defensins in Lung Host Defense. Annual Review of Physiology 64, 709–748. Selders, G.S., Fetz, A.E., Radic, M.Z., Bowlin, G.L., 2017. An overview of the role of neutrophils in innate immunity, inflammation and host-biomaterial integration. Regen Biomater 4, 55–68.

Selsted, M.E., Ouellette, A.J., 2005. Mammalian defensins in the antimicrobial immune response. Nature Immunology 6, 551–557. Shi, H.-Z., 2004. Eosinophils function as antigen-presenting cells. Journal of Leukocyte Biology 76, 520–527.

Slater, H., Bruno, D., Ren, H., Pertile, M., Schouten, J., Choo, K., 2003. Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). J Med Genet 40, 907–912. Smith, J.A., 1994. Neutrophils, host defense, and inflammation: a double-edged sword. J. Leukoc. Biol. 56, 672–686.

Stevens (MRCPath.), A., Lowe, J.S., Young, B., 2002. Wheater’s Basic Histopathology: A Colour Atlas and Text. Churchill Livingstone.

Sthoeger, Z.M., Bezalel, S., Chapnik, N., Asher, I., Froy, O., 2009. High α-defensin levels in patients with systemic lupus erythematosus: High defensin levels in SLE patients. Immunology 127, 116–122.

Stuart, P.E., Hüffmeier, U., Nair, R.P., Palla, R., Tejasvi, T., Schalkwijk, J., Elder, J.T., Reis, A., Armour, J.A.L., 2012. Association of β-Defensin Copy Number and Psoriasis in Three Cohorts of European Origin. Journal of Investigative Dermatology 132, 2407–2413.

Tajuddin, S.M., Schick, U.M., Eicher, J.D., Chami, N., Giri, A., Brody, J.A., Hill, W.D., Kacprowski, T., Li, J., Lyytikäinen, L.-P., 2016. Large-scale exome-wide association analysis identifies loci for white blood cell traits and pleiotropy with immune-mediated diseases. The American Journal of Human Genetics 99, 22–39.

Tak, T., Tesselaar, K., Pillay, J., Borghans, J.A.M., Koenderman, L., 2013. What's your age again? Determination of human neutrophil half-lives revisited. Journal of Leukocyte Biology 94, 595–601.

Tran, D., Tran, P., Roberts, K., Osapay, G., Schaal, J., Ouellette, A., Selsted, M.E., 2008. Microbicidal Properties and Cytocidal Selectivity of Rhesus Macaque Theta Defensins. Antimicrobial Agents and Chemotherapy 52, 944–953.

Tyson, J., Armour, J.A., 2012. Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR. BMC genomics 13, 693.

Tyson, J., Majerus, T.M.O., Walker, S., Armour, J.A.L., 2010. Screening for common copy-number variants in cancer genes. Cancer Genetics and Cytogenetics 203, 316–323.

Uhm, T.G., Kim, B.S., Chung, I.Y., 2012. Eosinophil Development, Regulation of Eosinophil-Specific Genes, and Role of Eosinophils in the Pathogenesis of Asthma. Allergy, Asthma and Immunology Research 4, 68.

Valore, E.V., Ganz, T., 1992. Posttranslational processing of defensins in immature human myeloid cells. Blood 79, 1538–1544.

Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573.

Walker, S., Janyakhantikul, S., Armour, J.A.L., 2009a. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. Genomics 93, 98–103.

Walker, S., Janyakhantikul, S., Armour, J.A.L., 2009b. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. Genomics 93, 98–103.

Walzog, B., Jeblonski, F., Zakrzewicz, A., Gaehtgens, P., 1997. Beta2 integrins (CD11/CD18) promote apoptosis of human neutrophils. The FASEB Journal 11, 1177–1186.

Westra, H.-J., Arends, D., Esko, T., Peters, M.J., Schurmann, C., Schramm, K., Kettunen, J., Yaghootkar, H., Fairfax, B.P., Andiappan, A.K., Li, Y., Fu, J., Karjalainen, J., Platteel, M., Visschedijk, M., Weersma, R.K., Kasela, S., Milani, L., Tserel, L., Peterson, P., Reinmaa, E., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., Homuth, G., Petersmann, A., Lorbeer, R., Prokisch, H., Meitinger, T., Herder, C., Roden, M., Grallert, H., Ripatti, S., Perola, M., Wood, A.R., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Knight, J.C., Melchiotti, R., Lee, B., Poidinger, M., Zolezzi, F., Larbi, A., Wang, D.Y., van den Berg, L.H., Veldink, J.H., Rotzschke, O., Makino, S., Salomaa, V., Strauch, K., Völker, U., van Meurs, J.B.J., Metspalu, A., Wijmenga, C., Jansen, R.C., Franke, L., 2015. Cell

Specific eQTL Analysis without Sorting Cells. PLOS Genetics 11, e1005223.
Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., Zhernakova, A., Zhernakova, D.V., Veldink, J.H., Van den Berg, L.H., Karjalainen, J., Withoff, S., Uitterlinden, A.G., Hofman, A., Rivadeneira, F., 't Hoen, P.A.C., Reinmaa, E., Fischer, K., Nelis, M., Milani, L., Melzer, D., Ferrucci, L., Singleton, A.B., Hernandez, D.G., Nalls, M.A., Homuth, G., Nauck, M., Radke, D., Völker, U., Perola, M., Salomaa, V., Brody, J., Suchy-Dicey, A., Gharib, S.A., Enquobahrie, D.A., Lumley, T., Montgomery, G.W., Makino, S., Prokisch, H., Herder, C., Roden, M., Grallert, H., Meitinger, T., Strauch, K., Li, Y., Jansen, R.C., Visscher, P.M., Knight, J.C., Psaty, B.M., Ripatti, S., Teumer, A., Frayling, T.M., Metspalu, A., van Meurs, J.B.J., Franke, L., 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature Genetics 45, 1238–1243.
White, S., Kalf, M., Liu, Q., Villerius, M., Engelsma, D., Kriek, M., Vollebregt, E., Bakker, B., Van Ommen, G.-J.B., Breuning, M.H., 2002. Comprehensive detection of genomic duplications and deletions in the DMD gene, by use of multiplex amplifiable probe hybridization. The American Journal of Human Genetics 71, 365–374.
White, S.J., Vink, G.R., Kriek, M., Wuyts, W., Schouten, J., Bakker, B., Breuning, M.H., Dunnen, J.T. den, 2004. Two-color multiplex ligation-dependent probe amplification: Detecting genomic rearrangements in hereditary multiple exostoses. Human Mutation 24, 86–92.
Wimley, W.C., Selsted, M.E., White, S.H., 1994. Interactions between human defensins and lipid bilayers: evidence for formation of multimeric pores. Protein Sci 3, 1362–1373.
Xu, X., 2017, Human Alpha defensin CNV haplotype diversity (PhD Thesis). University of Nottingham.
Xie, J., Shapiro, S., Gharavi, A., 2013. Genetic studies of IgA nephropathy: what have we learned from GWAS. Contrib Nephrol 181, 52–64.
Yang, D., Biragyn, A., Kwak, L.W., Oppenheim, J.J., 2002. Mammalian defensins in immunity: more than just microbicidal. Trends in immunology 23, 291–296.
Yang, D., Chen, Q., Chertov, O., Oppenheim, J., 2000. Human neutrophil defensins selectively chemoattract naive T and immature dendritic cells. Journal of leukocyte biology 68, 9–14.
Yang, Y.-N., Bauer, D., Wasmuth, S., Steuhl, K.-P., Heiligenhaus, A., 2003. Matrix metalloproteinases (MMP-2 and 9) and tissue inhibitors of matrix metalloproteinases (TIMP-1 and 2) during the course of experimental necrotizing herpetic keratitis. Experimental Eye Research 77, 227–237.
Yau, S.C., Bobrow, M., Mathew, C.G., Abbs, S.J., 1996. Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. Journal of medical genetics 33, 550–558.
Yousefi, S., Mihalache, C., Kozlowski, E., Schmid, I., Simon, H.-U., 2009. Viable neutrophils release mitochondrial DNA to form neutrophil extracellular traps. Cell Death & Differentiation 16, 1438–1444.

Yu, X.-Q., Li, M., Zhang, H., Low, H.-Q., Wei, X., Wang, J.-Q., Sun, L.-D., Sim, K.-S., Li, Y., Foo, J.-N., Wang, W., Li, Z.-J., Yin, X.-Y., Tang, X.-Q., Fan, L., Chen, J., Li, R.-S., Wan, J.-X., Liu, Z.-S., Lou, T.-Q., Zhu, L., Huang, X.-J., Zhang, X.-J., Liu, Z.-H., Liu, J.-J., 2011. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. Nature Genetics 44, 178–182.

Zhang, X., Ding, L., Sandford, A.J., 2005. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. BMC molecular biology 6, 4.

Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.-J., Bonder, M.J., van Rooij, J., Verkerk, M., Jhamai, P.M., Moed, M., Kielbasa, S.M., Bot, J., Nooren, I., Pool, R., van Dongen, J., Hottenga, J.J., Stehouwer, C.D.A., van der Kallen, C.J.H., Schalkwijk, C.G., Zhernakova, A., Li, Y., Tigchelaar, E.F., de Klein, N., Beekman, M., Deelen, J., van Heemst, D., van den Berg, L.H., Hofman, A., Uitterlinden, A.G., van Greevenbroek, M.M.J., Veldink, J.H., Boomsma, D.I., van Duijn, C.M., Wijmenga, C., Slagboom, P.E., Swertz, M.A., Isaacs, A., van Meurs, J.B.J., Jansen, R., Heijmans, B.T., 't Hoen, P.A.C., Franke, L., 2016. Identification of context-dependent expression quantitative trait loci in whole blood. Nature Genetics 49, 139–145.

6. APPENDICES

6.1 Appendix 1 - Primers used

Assay	Primer	Sequence	Source
DEFA1A3 (DefRes4)	DefRes4F	ATCCTTGCTGCCATTCTCC	Aldred 2004
	DefRes4R	TCCCTGTAGCTCTCAAAGCA	
	FAM-DefRes4R	6-FAM-TCCCTGTAGCTCTCAAAGCA	
DEFA CEN1/TEL 2 PRT	FAM-DEFAcen1F	6-FAM-CCCAGAGAGCTCCTTCATT	J.Armour unpublished
	HEX-DEFAcen1F	HEX-CCCAGAGAGCTCCTTCATT	
	DEFAcen1R	TCCTAGAAAGCTGGTTGCTC	
	FAM-DEFAtel2F	FAM-AGAGCAGCCGTGCACAAAC	
	HEX-DEFAtel2F	HEX-AGAGCAGCCGTGCACAAAC	
	DEFAtel2R	GCATCTYGGGGTCCATTGT	
DefHae3 Ratio	Primer 1	TGTCCCAGGCCCAAGGAAAA	Khan et al. 2013
	Primer 2	NED-TCCCTGTAGCTCTCAAAGCA	
Indel5	Primer 1	HEX-CTGTCCAGGAAGAGGGAGAG	Khan et al. 2013
	Primer 2	CAGCTGGAGGGTCTCTGTTC	
7bpdup	Primer 1	FAM-AGCAAAAATCAAACAACCTGA	Khan et al. 2013
	Primer 2	GCTATGCCTCCAATCTGACC	
4bpdel	Primer 1	TGCTCTCATTTTTGCATTCC	Ai et al. 2016
	Primer 2	FAM-TTTCTCCAAAGACTTGATTCCAA	
Flanking SNPs	TagSNPsF	AAGCTGTCTGTGTGTCTGGT	Khan et al. 2013 Black et al. 2014
	RepSpec2RB	AATGGGTGACGGGGGTTC	
	C2FnuR	CATGTAGTCTGTTGGTCAAAATGCT	
	Exch1F	ATCTCTCTTTGGATGGTG	
	Exch1R	GGTTCATATACAACATCTTCTTAAT	
6836028	6836028F	ATGCCTGCCTCTCTCACCT	J.Armour unpublished
	6836028R	CCCAGCCATGAGGACCC	
	FAM-6836028R	6-FAM-CCCAGCCATGAGGACCC	

6.2 Appendix 2 – Partial questionnaire data

AN17001	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17002	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania
AN17003	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania	White Romanian	Romania
AN17004	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17005	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17006	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17007	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17008	White British	England	White Spanish / Portuguese	Portugal	White Scottish	Scotland	White Portuguese	Portugal	White Spanish	Spain	White Scottish	Scotland	White Scottish	Scotland
AN17009	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain
AN17010	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland
AN17011	White Cypriot	Cyprus	White Cypriot	Cyprus	White Cypriot	Cyprus	White Cypriot	Cyprus	White Cypriot	Cyprus	White Cypriot	Cyprus	White Cypriot	Cyprus
AN17012	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17013	White British	England	White British	England	White British	England	White Irish/Italian	England	White British	England	White British	England	White British	England
AN17014	White British	England	White British	England	White British	England	White British	England	White British	England	9	9	9	9
AN17015	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Poland
AN17016	White British	England	White British	England	White British	England	White Scottish	Scotland	White Scottish	Scotland	White British	England	White British	England
AN17017	White British	England	White British	England	White British	England	9	9	9	9	White British	England	9	9
AN17018	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17019	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17020	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17021	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17022	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17023	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17024	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England

AN17025	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17026	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany
AN17027	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17028	White Italian	Italy	White Italian	Italy	White Italian	Italy	White Italian	Italy	White Italian	Italy	White Italian	Italy	White Italian	Italy
AN17029	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17030	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17031	White British	England	White British	England	White Scottish	Scotland	White British	England	White British	England	White Scottish	Scotland	White Scottish	Scotland
AN17032	White British	England	White British	England	White British	England	White Scottish	Scotland	White British	England	White Indian	India	White British	England
AN17033	White British	England	White British	England	White British	England	White Welsh	Wales	White Scottish	Scotland	White British	9	White British	England
AN17034	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17035	White British	England	White British	England	White British	England	White British	England	White British	England	White German	Germany	White British	England
AN17036	White British	England	White British	India	White British	England	White British	India	White British	England	White British	England	White British	England
AN17037	White British	England	White British	England	White British	England	White British	England	White British	England	White British	India	White British	Burma
AN17038	White British	England	White British	Isle of Man	White British	Isle of Man	White British	Isle of Man	White British	Isle of Man	White British	England	White British	England
AN17039	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany	White German	Germany
AN17040	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17041	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17042	White British	England	White British	England	White German	Germany	White British	England	White British	England	White	9	White German	Germany
AN17043	White British	Singapore	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17044	White Swiss	Switzerland	White Swiss	Switzerland	White Swiss	Switzerland	White Swiss	Switzerland	White Swiss	Switzerland	White Swiss	Switzerland	White Swiss	Switzerland
AN17045	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17046	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17047	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17048	White British	England	White Brazilian	Brazil	White British	England	White Brazilian	Brazil	White Brazilian	Brazil	9	9	9	9
AN17049	White British	England	White Irish	Ireland	White British	England	White Irish	Ireland	White Irish	Ireland	White Italian	Italy	White British	England

AN17050	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17051	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17052	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17053	White British	England	White British	England	White British	Uganda	White British	England	White British	England	White British	England	9	9
AN17054	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain	White Spanish	Spain
AN17055	White British	England	White British	England	White British / Black	England	White British	England	White British	England	White British	England	Black Caribbean	Jamaica
AN17056	White British / White Italian	Germany	White British	England	White Italian	England	White British	England	White British	England	White Italian	Italy	White Italian	Italy
AN17057	White British	Scotland	White Irish	Ireland	White British	England	White British	UK	White British	UK	White British	Wales	White British	UK
AN17058	White Polish	Poland	White Polish	Poland	White Polish	Poland	White Polish	Ukraine	White Polish	Ukraine	White Polish	Poland	White Polish	Poland
AN17059	White Irish	England	White Irish	Ireland	White Irish	England	White Irish	Ireland	White Irish	Ireland	White Irish	Ireland	White Irish	Ireland
AN17060	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17061	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17062	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17063	White German	Germany	White German	Germany	White German	Germany	White German	Germany / Poland	White German	Germany	White German	Germany / France	White German	Germany
AN17064	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania
AN17065	White Slovak	Slovakia	White Slovak	Slovakia	White Slovak	Slovakia	White Slovak	Slovakia	White Slovak	Slovakia	White Slovak	Slovakia	White Slovak	Slovakia
AN17066	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17067	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17068	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17069	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania	White	Romania
AN17070	White Dutch	Netherlands	White Dutch	Netherlands	White Dutch	Netherlands	White Dutch	Netherlands	White Dutch	Netherlands	White Dutch	Netherlands	White Dutch	Netherlands
AN17071	White Moldavian	Moldova	White Moldovan	Moldova	White Moldovan	Moldova	White Moldovan	Moldova	White Moldovan	Moldova	White Moldovan	Moldova	White Moldovan	Moldova
AN17072	White Scottish	Scotland	White Scottish	Scotland	White Scottish	Scotland	White Scottish	Scotland	White Scottish	Scotland	White Scottish	Scotland	White Scottish	Scotland
AN17073	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17074	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17075	White British	England	White British	England	White British	England	White British	England	White British	England	White British	South Africa	White British	England
AN17076	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England	White British	England
AN17077	White British	England	White British	Bahamas	White British	England	White British	England	White British	England	White British	England	White British	England