I declare that this thesis entitled:

"A Systems Based Approach to Neutrophil Gene Expression"

is entirely my own work except where indicated in the text.

Candidate: **Huw Batten Thomas**

Supervisors: **Professor S.W. Edwards**
Institute of Integrative Biology
University of Liverpool

Professor R. J. Moots
Institute of Ageing and Chronic Disease
University of Liverpool

Contents

Acknowledgements

This work was funded by the BBSRC in collaboration with the pharmaceutical company UCB Celltech.
My sincere thanks go to my supervisors, Professors Steven Edwards and Rob Moots for providing me with the opportunity to carry out this research, and for their support and guidance throughout the last 4 years.

I feel incredibly lucky to have had the opportunity to undertake a PhD, but it would not have been possible without the help, support, and friendship of several people. I owe a debt of gratitude to these people and a brief mention here in no way adequately reflects their importance or my level of appreciation for their part in making me the person I am today.

A huge thanks to the members of the SWE group, past and present, who have always made "work", feel like a second home. In particular, to Luke for making my return to science during my MRes course a thoroughly enjoyable experience and giving me the confidence to believe I was capable of greater things. But above all, to Helen, without which, I would never have achieved so much. During the past 4.5 years, she has played the part of colleague, supervisor, sister, mother, but most of all, a dear friend. My time at Liverpool truly has been the most enjoyable years of my life and I feel incredibly lucky to have done so much, and met so many kind and inspiring people.

However, none of this would have been possible without the support of the love of my life, and best friend, Vicky. Her constant encouragement, patience, and unflinching love gave me the strength and confidence to make the break from a sedentary life and follow my ambitions.

Finally, to my family, in particular my parents, for their constant love and support, but most of all, for believing in me.

Abstract

Neutrophils are the major cellular constituent of blood leukocytes and play a central role in the inflammatory response, expressing an array of destructive molecules and antimicrobial processes that characterise the cells as front-line defenders of the innate immune system, thus neutrophils are crucial to host defence. It is now appreciated that neutrophils produce and respond to a variety of inflammatory signals and are able to regulate both the innate and adaptive immune response. The molecular changes that underlie this regulation are poorly defined, yet represent an attractive area of research to fully elucidate the role and regulatory capacity of neutrophils within the immune response. RNA-Seq provides an accurate and robust mechanism for global characterisation of cellular transcripts.

Neutrophils were isolated from healthy donors and incubated with or without inflammatory cytokines for 1 h. RNA was extracted and analysed by RNA-Seq using the SOLiD or Illumina platforms. Raw data was quantified using a number of software packages which formed a bioinformatic pipeline for data analysis which was developed during the course of the research. Results were validated by a selection of traditional laboratory functional assays.

Priming of neutrophils by GM-CSF and TNFα was found to induce differential gene expression and activation of transcription factors, which led to differential regulation of apoptotic pathways. Stimulation of neutrophils with inflammatory cytokines/chemokines (IL-1β, IL-8, G-CSF, IFNγ) resulted in expression of discrete gene sets and differential activation of signalling pathways. Stimulation of neutrophils with IL-6 did not induce any significant expression of genes but result in activation of STAT signalling. Comparison of gene expression of neutrophils isolated by density gradient and magnetic bead preparation revealed significant differences in gene expression and function, in part attributable to levels of contamination associated with each isolation method. Bead isolation was found to enrich a more heterogeneous neutrophil population including a subpopulation of neutrophils expressing transcripts previously associated with low density granulocytes.

Thus, RNA-Seq and bioinformatic analysis has provided a full characterisation of neutrophil gene expression under inflammatory conditions and identified several new areas of research that could lead to targeted drug design for the treatment of inflammatory disease.

Publications and presentations

Publications

Wright H.L.†, **Thomas H.B.†**, Moots R.J., Edwards S.W. (2013) RNA-Seq reveals activation of both common and cytokine-specific pathways following neutrophil priming. *PLoS One* 2013;8(3):e58598. PMID: 23554905 **(† Joint 1st author)**.

Tamarozzi, F., Wright, H.L., **Thomas, H.B.**, Edwards, S.W., Taylor, M.J. (2014) A lack of confirmation with alternative assays questions the validity of IL-17A expression in human neutrophils using immunohistochemistry. *Immunology Letters*, Vol 162 (2B): 194-198

Wright, H.L., **Thomas, H.B.**, Edwards, S.W., Moots, R.J. (2015) Interferon gene expression signature in RA neutrophils predicts response to anti-TNF therapy. *Rheumatology,* Vol 54 (1): 188-193

Veraldi, N., Hughes, A.J., Rudd, T.R., **Thomas, H.B.**, Edwards, S.W., Hadfield, L., Skidmore, M.A., Siligardi, G., Shute, J.K., Naggi, A.,. Yates, E.A. (2015) Heparin Derivatives for the Targeting of Multiple Activities in the Inflammatory Response. *Carbohydrate Polymers,* Vol 117: 400-407

Thomas H.B., Moots R.J., Edwards S.W., Wright H.L. (2014) The impact of isolation methods and purity on *in vitro* neutrophil function and gene expression. *(In preparation).*

Chiewchengchol D., Wright H.L., **Thomas, H.B.**, Lam C., Roberts K.J., Moots R.J., Beresford M., and Edwards S.W. (2014) Down Regulation of Death Receptor Signalling in Human Neutrophils following TNF Stimulation. *(In preparation).*

Presentations

1. *A systems based approach to neutrophil gene expression – Next generation analysis of the neutrophil transcriptome under conditions of inflammation.* University of Liverpool department of Biochemistry and cell biology – Graduate symposium. Sept 2011 (*Oral presentation)*

2. *Developing a pipeline for whole transcriptome sequencing of inflammatory neutrophils.* Presented at The Neutrophil in Immunity Conference, Québec, Canada. June 2012. (*Poster presentation*)

3. *Gene expression profiling of cytokine stimulated neutrophils in vitro.* Presented at The British Society of Immunology (BSI) Conference, Liverpool. December 2013. (*Poster presentation*)

4. *Digitising the neutrophil – RNA-Seq reveals activation of both common and cytokine-specific pathways following neutrophil priming.* Presented at The University of Liverpool Musculoskeletal Biology Science Day. February 2013. (*Poster presentation*)

5. *Gene expression in primed neutrophils.* Presented at The University of Liverpool Postgraduate Poster Presentation Day. April 2013. (*Poster presentation*)

6. *A systems based approach to neutrophil gene expression.* Presented at the UCB PhD research day, Royal College of Surgeons, London. Oct 2014. (Oral presentation).

Abbreviations

ACTB	β-Actin
AKT	Protein kinase B
ALOX15	Arachindonate 15-lipoxygenase
AP-1	Activator protein-1
APAF-1	Apoptotic protease-activating factor-1
APS	Ammonium persulphate
APRIL	A proliferation-inducing ligand
ARL4C	ADP-Ribosylation Factor-Like 4C
ASCII	American standard code for information
ATP	Adenosine triphosphate
AZU1	Azurocidin 1
B2M	β_2 microglobulin
BAD	B-cell leukaemia-2 associated death promoter protein
BAK	B-cell leukaemia-2 homologous antagonist/killer protein
BAFF	B-cell activating factor
BAM	Binary sequence alignment map
BAX	B-cell leukaemia-2 associated protein-X
BCF	Binary variant calling file
BCL11B	B-cell CLL/lymphoma 11B
BCL2	B-cell leukaemia-2 protein
BCL2A1	B-cell leukaemia-2 related protein A1
BFL-1	B-cell leukaemia-2 related protein A1
BGI	Beijing Genomics Institute
BH	B-cell leukaemia-2 homology domain
BID	BH3 interacting domain death antagonist protein

BIM	B-cell leukaemia-2 like protein-11
BPI	Bactericidal/Permeability-Increasing Protein
CAMP	Cathelicidin antimicrobial peptide
Caspase	Cysteine-aspartic acid protease
CCL	Chemokine (C-C motif) ligand
CCR	Chemokine (C-C mitif) receptor
CD	Cluster of differentiation
CD96	T-cell activation, increased late expression
CEACAM	Carcinoembryonic antigen-related cell adhesion molecule
CEBP	CCAAT/enhance binding protein
CGD	Chronic granulomatous disease
CGR	Centre for genomic research
ChIP-Seq	Chromatin immunoprecipitation sequencing
C/EBP	CCAAT/enhancer-binding protein
CFLAR	Caspase 8 and FADD-like apoptosis regulator
ChIP	Chromatin Immunoprecipitation
CLC	Charcot-Leyden crystal galectin
CLU	Clusterin
COPD	Chronic Obstructive Pulmonary Disease
CORE	Committee for research ethics
CRISP3	Cystein-rich secretory protein
Ct	Cycle threshold
CTSA	Cathepsin A
CTSG	Cathepsin G
CXCL	Chemokine (C-X-C Motif) Ligand
CXCR	Chemokine (C-X-C Motif) Receptor
CYB5D1	Cytochrome B5 Domain Containing-1
DAMP	Damage associated molecular pattern
DE	Differentially expressed
DEFA4	Defensin α-4

DISC	Death inducing signalling complex
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotidetrisphosphate
dsDNA	double stranded DNA
DSN	Duplex specific nuclease
DUSP5	Dual-specific phosphatase 5
ECL	Enhanced chemiluminescence
ELANE	Neutrophil elastase
EMR4P	Egf-like module containing, mucin-like, hormone receptor-like 4
ERK	Extracellular signal-regulated kinase
FADD	Fas-aassociated via death domain
FASLG	Fas ligand
FcγR/FCGR	Fragment crystallisable gamma receptor
FDR	False discovery rate
FITC	Fluorescein isothiocyanate
fMLP	formyl-methionyl-leucyl-phenylalanine
FOS	FBJ Murine Osteosarcoma Viral Oncogene Homology
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAS	Interferon-γ-activated site
GB	Gigabyte
GBP	Guanylate binding protein
GBP1P1	Guanylate binding protein-1, Interferon-Inducible Pseudogene 1
G-CSF	Granulocyte-colony stimulating factor
GM-CSF	Granulocyte macrophage-colony stimulating factor
GPR	G protein-coupled receptor
H_2O_2	Hydrogen peroxide
HBA2	Heamoglobin, alpha 2
HBA1	Heamoglobin, alpha 1

HBB	Heamoglobin, beta
HBSS	Hanks' balanced salt solution
HIF1α	Hypoxia-inducible factor 1 alpha
HOCL	Hypochlorous acid
HRP	Horseradish peroxidase
HPRT1	Hypoxanthine phosphoribosyltransferase
ICAM	Intercellular adhesion molecule
IDO1	Indoleamine 2,3-Dioxygenase 1
IER3	Immediate early response-3
IFN-γ	Interferon-alpha
IFN-γ	Interferon-beta
IFN-γ	Interferon-gamma
IFN-γ	Interferon-lambda
IgG	Immunoglobulin G
IGV	Intergrative genomics viewer
IκB	Nuclear factor of kappa light polypeptide gene in B cells inhibitor
IKK	Inhibitor of nuclear factor kappa B kinase
IL	Interleukin
IL7R	Interleukin 7 receptor
IL-1RN	IL-1 receptor agonist
INDEL	Insertions and/or deletions
INS	Insulin
IPA	Ingenuity pathway analysis
ISRE	Interferon-stimulated response elements
ITGAL	Integrin, alpha L
ITGAM	Integrin, alpha M
ITGB7	Integrin, beta 7
ITK	IL2-inducible T-cell kinase
JAK	Janus kinase
JNK	C-Jun N-terminal kinase

JSLE	Juvenile systemic lupus erythematosus
JUN	Jun proto-oncogene
kDa	kilo Dalton
LCN2	Lipocalin 2
LDGs	Low-density granulocytes
LEF1	Lymphoid enhancer-binding factor
LPS	Lipopolysaccharide
LTF	Lactotransferrin
MAC-1	Macrophage-1 antigen
MAPK	Mitogen-activated protein kinase
MDS	Multi-dimensional scaling
MeV	Multiple experiment viewer
MCL-1	Myeloid cell leukaemia
MCL -1L	Mcl-1 long-form
MCL -1S	Mcl-1 short-form
MCL -1ES	Mcl-1 extra-short-form
MMP8	Matrix metalloproteinase 8
MMP9	Matrix metalloproteinase 9
MPO	Myeloperoxidase
mRNA	Messenger ribonucleic acid
miR	micro-RNA
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NAMPT	Nicotinamide Phosphoribosyltransferase
ncRNA	non-coding RNA
NDG	Normal density granulocytes
NGAL	Neutrophil gelatinase-associated lipocalin
NGFR	Nerve growth factor receptor
NGS	Next generation sequencing
NETs	Neutrophil extracellular traps
NF-κβ	Nuclear factor kappa beta
NK-cells	Natural killer cells

PAF	Platelet activating factor
PAMP	Pathogen-associated molecular pattern
PBMC	Peripheral blood mononucleated cell
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pDC	plasmacytoid dendritic cell
PEST	Proline-glutamic acid-serine-threonine
PFA	Paraformaldehyde
PI	Propridium iodide
piRNA	piwi-interacting RNA
PMA	Phorbol 12-myristate 13-acetate
PRR	Pattern recognition receptor
PRSS33	Protease, serine, 33
PSGL-1	P-selectin glycoprotein ligand-1
PU.1	Spleen Focus Forming Virus Proviral Integration Oncogene
PVDF	Polyvinylidene fluoride
qPCR	quantitative polymerase chain reaction
rRNA	Ribosomal ribonucleic acid
RA	Rheumatoid arthritis
RAM	Random access memory
RAMP	Resolution associated molecular pattern
RIN	RNA intergrity number
RNA	Ribonucleic acid
RNASE2	Ribonuclease 2
RNASE3	Ribonuclease 3
RNA-Seq	Ribonucleic acid sequencing
RPKM	Reads per kilobase per million
S1PR1	Sphingosine-1-phosphate receptor 1
SAM	Sequence alignment map
SD	Standard deviation

SDF-1	Stromal derived factor -1
SDS	Sodium dodecyl sulphate
SEM	Standard error of mean
SIGLEC8	Sialic acid binding Ig-like lectin 8
SLE	Systemic lupus erythematosus
SMPD3	Sphingomyelin Phosphodiesterase 3
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
SOCS3	Suppressor of cytokine signalling-3
STAT	Signal transducer and activator of transcription
TB	Terabyte
TBC1D4	TBC1 domain family, member 4
TCR	T-cell receptor
TEMED	Tetramethylethylenediamine
TGF	Tumor growth factor-beta
THBS1	Thrombospondin 1
Th-cells	T-helper cells
TLR	Toll-like receptor
TMEM170B	Transmembrane protein 170B
TNF-α	Tumour necrosis factor-alpha
TRAIL	TNF-related apoptosis-inducing ligand
TRAT1	T-cell receptor associated transmembrane adaptor 1
tRNA	transfer ribonucleic acid
TSS	Transcriptional start site
TYK2	Tyrosine kinase 2
UV	Ultra violet
VCF	Variant calling file
ZMW	Zero mode waveguide

Table Ab.1 Bioinformatic file formats and file extension abbreviations

File format	File extension	Summary
Sam file	*.SAM*	Sequence Alignment Map – output format from high-throughput mappers such as Bowtie, Tophat, or BWA
Bam file	*.BAM*	Binary format of a SAM file – smaller file size and can be processed faster than an equivalent SAM file.
General feature format / General transfer format	*.GFF/.GTF*	Reference file listing various sequence features and attributes such as gene boundaries and coding frame
FASTA	*.FASTA*	Text based file for nucleotide sequence data
Colorspace Fasta	*.CSFASTA*	Similar to FASTA but nucleotide format is encoded in colorspace
Fasta Quality file	*.QUAL*	List of quality values for nucleotide sequence to accompany a FASTA format file
Fastq	*.FASTQ*	FASTA format file where nucleotide sequence and quality value are encoded into a single ASCII character to decrease digital footprint
Variant calling file	*.VCF*	Text file containing data on individual genomic positions, used primarily to list SNPs and indels
Binary variant calling file	*.BCF*	Binary format of a .vcf used for faster processing

Chapter 1: Introduction

1.1 Project Overview

Neutrophilic polymorphonuclear leukocytes (neutrophils) were first described by Paul Ehrlich in the late nineteenth century using cell-staining techniques to investigate the subpopulations of white blood cells [1]. Ehrlich discovered three sub-types of cells each featuring polymorphous nuclei that could be distinguished from each other by their individual staining properties. Whilst eosinophils and basophils were named due to their cytoplasm staining positively with eosin and basic dyes respectively, the third subtype exhibited a tendency to retain neutral dyes and was so termed the neutrophil [2].

Neutrophils are the most abundant leukocyte found in circulating blood, accounting for 40-60% of the total white blood cell population [3]. They form the major cellular constituent of the innate immune system and are indispensable for defence against invading bacterial and fungal pathogens due to their ability to phagocytose cells and micro-organisms, release lytic enzymes from internal granules, and produce reactive oxygen metabolites with antimicrobial potential [4,5]. Their highly-conserved mechanisms of anti-microbial activity, coupled with a characteristic short life span, have historically defined the neutrophil as a one dimensional effector cell with little capacity to influence the more complex, adaptive arm of the immune system, predominantly regulated by T-cells and B-cells.

However, in recent years, this view of neutrophils has been profoundly altered. Neutrophils are now known to produce and release numerous cytokines, chemokines and angiogenic/fibrogenic factors [6]. They have also been shown, following cytokine stimulation, to express MHC Class II molecules and present antigen-to and activate T-cells [7].

The perceived role of neutrophils in inflammatory disease has also been altered in recent years. Neutrophil dysregulation has been associated with the pathogenesis of a variety of chronic inflammatory diseases such as rheumatoid arthritis (RA) [8], juvenile (and adult) systemic lupus erythematosus (SLE) [9], chronic obstructive pulmonary disease (COPD) [10], asthma [11] and alcoholic hepatitis [12]. This current view of neutrophils places them central to the immune system with a significant capacity to regulate, influence and affect both the innate and adaptive response in health and disease. Despite a greater appreciation for neutrophil involvement in the immune response, relatively little work has focused on the underlying mechanisms of neutrophil activation and regulation in the context of inflammation, instead focusing more on the traditionally associated mechanisms of functions such as chemotaxis, phagocytosis and apoptosis.

Quantifying the transcriptional output of a cell has often been used by researchers to unpick the mechanisms by which a cell behaves under normal conditions, or adapts and responds to a changing environment or signal. The sensitivity of quantification of transcriptomes has grown exponentially over the past 2 decades, firstly with the development of

array-based technologies and real-time quantification by polymerase chain reaction (PCR). Subsequently, the successful development of micro-array technology provided the first means to quantify the entire population of transcripts within a cell or tissue population. However, the greatest technological advancement has come in the area of massively high-throughput sequencing (HTS). Today, whole genome/exome/transcriptome sequencing can now be performed over the course of a few days/weeks within a research lab for a few thousand pounds. This represents a minute fraction of the cost and time expended on the first large scale genome project, the Human Genome Project, which over the course of its 13 years was estimated to have cost over $3 billion [13].

This thesis aims to define a set of modern tools and bioinformatic software packages that can consistently, accurately and robustly quantify the gene expression profile of neutrophils using RNA-Seq. This pipeline will then be utilised to investigate the changes in gene expression following stimulation of neutrophils with inflammatory cytokines. Finally, the differences in two commonly used neutrophil isolation methods and the influence of inherent cellular contamination will be analysed, including their effect on neutrophil function and gene expression of both *in vitro* stimulated and unstimulated neutrophils.

1.2 Neutrophils

Examples of cells with phagocytic abilities are found in organisms as distantly related as the slime mould *Dictyostelium discoideum* [14] and the African clawed frog *Xenopus laevis* [15]. Indeed, such is the level of evolutionary conservation between species, that both zebrafish and rodents are often used as model organisms for *in vivo* studies of neutrophil function despite their neutrophils comprising a much lower proportion of the leukocyte population (15-20%) than in higher mammals [2]. In humans, neutrophils comprise up to 60% of all leukocytes and are produced at a rate of $5 \times 10^{10} – 30 \times 10^{10}$ cells /day [16]. The vast majority are retained in the bone marrow, and only a small fraction is released into the circulating peripheral blood, such that the concentration of neutrophils in blood is approximately $3\text{-}5 \times 10^{6}$/mL.

1.2.1 Neutrophil production and maturation

Granulocyte-colony stimulating factor (G-CSF) is essential for regulating the production and release of neutrophils from the bone marrow, whilst the chemokine receptors 4 (CXCR4) and 2 (CXCR2) are central to regulating neutrophil retention and release from the bone marrow, respectively. CXCR4 acts with its ligand, stromal derived factor -1 (SDF-1), which is produced by bone marrow stromal cells, to retain neutrophils in the bone marrow. Conversely, growth regulated protein (Gro)α and Groβ released by stromal cells acts through CXCR2 to increase neutrophil release [17].

Levels of SDF-1 are tightly regulated by G-CSF, which in turn is regulated by levels of Interleukin (IL)-17A produced by neutrophil-regulatory T-Cells. IL-17A release is dependent on levels of IL-23 released by tissue macrophages and dendritic cells. Uptake of apoptotic neutrophils by macrophages and dendritic cells results in a decrease in IL-23 levels. Thus, a reciprocal negative feedback loop exists such that as levels of neutrophils in tissue increases, levels of G-CSF decrease, ultimately leading to a decrease in neutrophil release from the bone marrow [18] (Fig 1.1-inset).

Neutrophil maturation is largely dependent on the transcription factors CCAAT/Enhancer binding protein alpha – zeta (C/EBPα-ζ) and PU.1. Terminal differentiation into either a monocyte or a granulocyte lineage is ultimately decided by the balance of expression between these two transcription factors [19].

1.2.2 Neutrophil granules

Neutrophil granules and granule proteins are produced sequentially during neutrophil maturation. Granule proteins serve mainly as antimicrobial proteins aiding pathogen killing during phagocytosis. Granules have historically been classified into 3 types based on their protein content: azurophilic (primary) granules, which contain myeloperoxidase (MPO); specific (secondary) granules, which contain lactoferrin; and gelatinase (tertiary) granules, which contain matrix metalloproteinase 9 (MMP9). However, more recently, neutrophil granules have been further classified into several sub-types. For example, azurophilic granules can be

differentiated into defensin^{hi} and defensin^{low} [20], whilst specific granules can be sub-divided into at least 4 sub-types; lactoferrin^{hi}, cysteine-rich secretory protein 3 (CRISP3)^{hi}, gelatinase^{hi} and ficolin-1^{hi} [18,21–23]. The mechanism underlying the existence of separate granules has been termed "targeting by timing of biosynthesis"[24] whereby proteins expressed at similar stages of maturation are localised to similar granule subtypes. This allows for differences in the mobilization of proteases, with the granules formed last during maturation being most likely to be released first. It is also important that some of these granule proteins are localised to different compartments as they can antagonise the activities of each other: for example neutrophil elastase (found in azurophil granules) can digest neutrophil gelatinase-associated lipocalin (NGAL) (found in specific granules) [18,23]. Neutrophils also contain secretory vesicles that contain proteins and receptors associated with neutrophil adhesion and migration. Upon neutrophil activation, vesicles containing adhesion molecules such as $\beta_2$integrins are incorporated into the neutrophil surface membrane where they facilitate neutrophil migration into tissues.

1.2.3 Activation, rolling, adhesion and extravasiation

Neutrophil migration from peripheral blood is mediated by interactions with vascular endothelium, predominantly at postcapillary venules [18]. Neutrophils move to the site of inflammation along a chemotactic gradient. Expression of L-selectin on neutrophils and both E-selectin and P-selectin on endothelial cells mediate tethering and lumal rolling of neutrophils by

binding of P-selectin glycoprotein ligand-1 (PSGL-1) and other glycosylated ligands [25]. Neutrophils are further activated by chemokines and proinflammatory agents presented on the surface of endothelial cells, such as CXCL8 (IL-8) and CXCL2 (MIP-2). This in turn, activates integrins such as β_2 integrin (CD18) leading to high-affinity binding with integrin ligands expressed on the surface of endothelial cells, such as intracellular adhesion molecules 1 and 2 (ICAM-1 and ICAM-2). The interaction of integrin alpha L (ITGAL) and integrin alpha M (ITGAM) with ICAM1 is important for neutrophil adhesion and intraluminal crawling, allowing neutrophils to move to endothelial borders in preparation for extravasation [26]. Following firm adhesion to the endothelial layer, neutrophil movement into tissues can occur in one of two ways; paracellular migration, where neutrophils squeeze between endothelial cells, or less-commonly, transcellular migration, where neutrophils pass through an individual endothelial cell. Neutrophil transmigration is facilitated by several neutrophil surface molecules, including CD54 (ICAM-1) [27], CD31 (platelet/endothelial cell adhesion molecule 1 – PECAM-1) [28], CD44 [29], and CD47 (integrin associated protein – IAP) [30]. Once in the tissue, neutrophils continue to move along a chemotactic gradient towards the source of inflammation where they can carry out other immune processes, such as cytokine release, phagocytosis and NETosis (Fig 1.1).

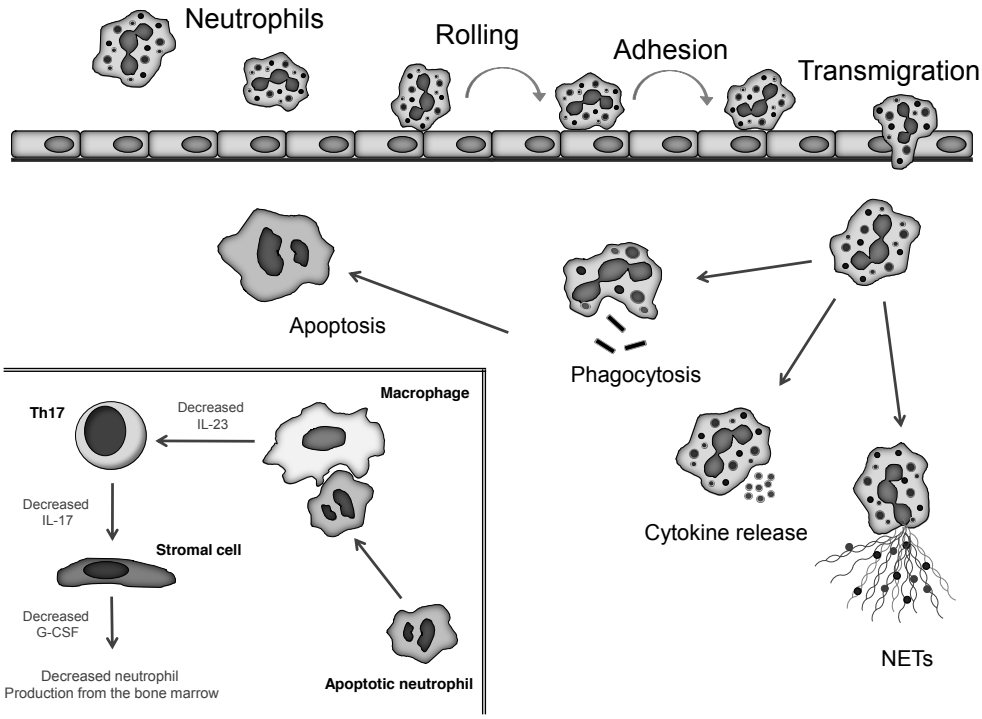


Fig 1.1 Overview of neutrophil production and function. Peripheral blood neutrophils migrate to site of inflammation via chemotactic signalling and transmigrate into tissue following a number of sequential processes (rolling, adhesion and transmigration) where they can carry out their key functions, such as cytokine release, phagocytosis or NETosis. Neutrophils subsequently become apoptotic during inflammation resolution. (Inset) Neutrophil release from the bone marrow is regulated by levels of G-CSF expressed by stromal cells. As levels of apoptotic neutrophils increase in tissue, G-CSF levels are decreased through the down-regulation of IL-23 and IL-17 release by macrophages and Th17 cells, respectively.

1.3 Neutrophil function

1.3.1 Priming

Neutrophils exist in one of three activation states; quiescent (also known as unprimed), primed or activated [31]. Under normal conditions, neutrophils patrol the vascular system in large numbers in a dormant (or unprimed) state. In the absence of any stimulating factors, peripheral blood neutrophils rapidly undergo constitutive apoptosis resulting in a short lifespan, typically less than 24 h [32].

Upon exposure to a variety of inflammatory stimuli (such as chemokines, cytokines, bacterial peptides or by adhesion), neutrophils become primed. Priming involves several rapid molecular changes; intracellular granules containing pre-formed receptors are mobilised to the plasma membrane, receptor affinity is altered, and components of the Nicotinamide Adenine Dinucleotide Phosphate (NADPH) oxidase complex are assembled at the plasma membrane [33]. Additionally, priming involves other, less rapid molecular changes, for example, increased gene expression resulting in the production of cytokines or chemokines, and stabilisation of proteins involved in apoptosis, ultimately extending the life of a primed neutrophil [3]. The overall consequence of priming is a neutrophil which is capable of a rapid and increased response to subsequent activation signals. A major advantage of priming is as a regulatory step to ensure peripheral blood neutrophils are not inappropriately or non-specifically activated, leading to unregulated release of the neutrophils toxic armoury and unnecessary localised tissue damage. Interestingly, studies using platelet activating factor (PAF) have revealed that neutrophil priming can be reversed [34], suggesting that in the absence of an activation stimulus a mechanism exists to allow neutrophils to be returned to a quiescent state for subsequent recycling at the inflammatory site [34].

1.3.2 Pattern recognition receptors (PRR)

The primary role of neutrophils is to identify and destroy pathogenic organisms from the body. In order to achieve this, neutrophils must first be

able to differentiate between foreign antigens (for example bacterial peptides), host antigens (for example immunoglobulins and complement factors) and host (self) proteins before initiating an appropriate immune response. Consequently, neutrophils express a variety of pattern recognition receptors (PRR) such as Toll-like receptors (TLRs), Nod-like receptors (NLRs) and C-type lectin receptors [5,35]. PRRs can be activated by microbial structures, more commonly-known as pathogen associated molecular patterns (PAMPs) or endogenous signals produced by host cells in response to trauma, ischemia or tissue damage, more commonly known as damage-associated molecular (DAMPs) [35]. More recently, a new class of pattern molecules has been proposed. Due to their anti-inflammatory actions, members of the heat shock protein (HSP) family (among others) have been termed resolution associated molecular patterns (RAMPs) [36]. This highlights the importance of pattern recognition receptors in both the activation and resolution of neutrophil function.

1.3.3 Phagocytosis

Following neutrophil priming and chemotaxis towards a source of pathogenic insult, clearance of foreign microbes is achieved by phagocytosis. Phagocytosis is triggered by the interactions of opsonised particles with specific receptors on the surface of neutrophils. For example, complement receptors (CR) such as CR1 and CR3 bind to particles opsonised with complement, whereas particles opsonised with γ–

immunoglobulin (IgG) are phagocytosed via interaction with Fcγ receptors (FcγR) (which bind the Fc portion of immunoglobulins). Ultimately, microorganisms are internalised by the neutrophil into a membrane-bound vacuole, known as a phagosome.

Neutrophils express 3 classes of FcγR. The low-affinity binding receptors FcγRIIA (CD32) and FcγRIIIB (CD16) are both expressed on unprimed neutrophils. On the other hand, expression of the high-affinity receptor FcγRI (CD64), although present in synovial neutrophils from rheumatoid arthritis (RA) patients [37] is only expressed on healthy neutrophils following activation by IFNγ [38].

The mechanism of pathogen internalisation is dependent on the specific interactions between the neutrophil and the microorganism, for example interactions may be direct through activation of PRR by PAMPs, or may be mediated by opsonins [2]. Phagocytosis of opsonised particles is largely mediated by either the FcγRs or complement receptors (CR). Engulfment is initiated by the localised clustering of phagocytic receptors (for example FcγRs) following ligation with their cognate ligand. Subsequently, extended membrane structures, or pseudopodia, engulf the particle forming an intracellular phagosome. It can take less than 20 s to internalise a microorganism. Interestingly, CR-mediated phagocytosis does not result in the formation of pseudopodia, but instead, the opsonised particle appears to sink into the neutrophil [39].

Microbial killing within the phagosome is achieved via a two step process; firstly, internal granules fuse with the phagosome membrane, releasing their content into the phagosomal lumen. Simultaneously, reactive oxygen species (ROS) are produced via assembly of the NADPH oxidase complex on the phagosomal membrane [2]. These processes produce a hostile intra-phagosomal environment that leads to the destruction of the engulfed microorganism.

In addition to an effective mechanism of pathogen clearance – since neutrophils can express MHCII molecules [40] - phagocytosis also serves to provide antigens to the adaptive immune system. This highlights the important role that neutrophils play as both an innate effector, but also an immune-regulator of both the innate and adaptive immune system.

1.3.4 NETosis

In the last decade, evidence has emerged that neutrophils are capable of a specialised form of programmed cell death, mechanistically distinct from either necrosis or apoptosis. Under conditions of increased stimulation or overwhelming numbers of bacteria, neutrophils are able to extrude neutrophil extracellular traps (NETs) which act as an additional antimicrobial mechanism in the neutrophil armoury. NETosis is characterised by NADPH-oxidase-dependent dissolution of the nucleus and intracellular membranes, followed by the rupture of the plasma membrane and expulsion of decondensed chromatin [41] (Fig 1.2). NETs are decorated with a variety of cellular proteins, such as granule proteins (which form the greatest

proportion of NET-associated proteins), nuclear proteins (such as histones) and cytoplasmic proteins (such as calcium-binding proteins). The advantages of NET formation in host defence are numerous. Firstly, NETs provide a physical barrier, preventing microbial spread. Secondly, the concentration of anti-microbial proteins that decorate NETs increases their efficiency and promotes synergistic actions of proteases. Thirdly, the restricted dispersion of proteases limits the amount of localised tissue damage induced by localised NETosis [42].

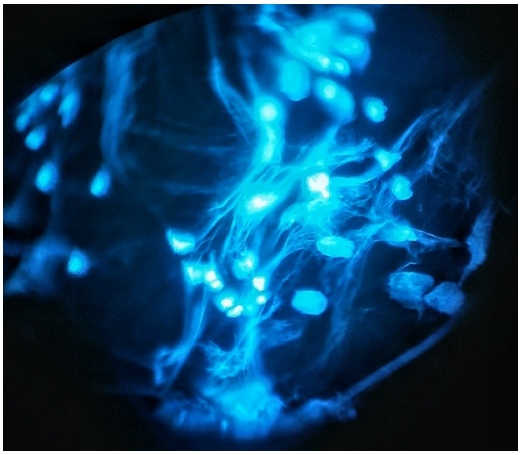


Fig 1.2 Neutrophil NETs. Neutrophils stimulated with PMA for 2h to induce NET formation. Cells stained for DNA using DAPI. Viewed under fluorescent microscope (x40 magnification)

Following the formation of NETs, it is generally assumed that the neutrophil will die, as this is considered a specialised form of cell death. However, evidence suggests that, under certain conditions, NETs can be formed from mitochondrial DNA alone, without decreasing the potential life-span of the neutrophils [43].

Importantly, NETs are also known to have a role in signalling, by activation of plasmacytoid dendritic cells (pDC) via their Toll-like receptor 9 (TLR9) [44], which is triggered by DNA. NETs have also been implicated in CD4+-cell priming via a TLR9 independent mechanism [42]. Further evidence to support the notion that NETs are an evolutionary-conserved process is the discovery of several bacterial strategies for evading NETs, for example, *Streptococcus pneumonia* is capable of altering its capsule charge, thereby decreasing its binding-affinity to NETs [45]. Bacteria are also capable of expressing nucleases on their surface [46] or releasing endonucleases [47] which can degrade NETs.

Despite a clear host advantage for NETosis formation, numerous studies have implicated NETs in the pathogenesis of several diseases [48], cystic fibrosis [49], periodontitis [50], preeclampsia [51] and autoimmune conditions such as systemic lupus erythematosus (SLE) [9,44,52–54] and rheumatoid arthritis [55]. The formation of NETs in autoimmune conditions can lead to the release of nuclear and protein neo-antigens as a potential mechanism for auto-antibody production.

1.3.5 Apoptosis

In order to balance the high rate of neutrophil production and release into peripheral blood, neutrophils must also regulate the rate at which they undergo apoptosis and are removed from circulation. Apoptosis is an evolutionary-conserved mechanism of cell death that is tightly-regulated

and ensures that the destructive proteases and lytic enzymes contained in neutrophil granules are not released into the surrounding environment, thereby minimising their potential for localised tissue damage [56]. In the absence of any stimulating factors, neutrophils can undergo constitutive apoptosis within 8 h [57]. Apoptosis leads to several morphological changes that are characteristically distinct from changes that occur during necrosis. These include: condensation of nuclear chromatin; nuclear fragmentation, leading to the loss of the characteristic neutrophil multi-lobed nucleus; DNA and protein degradation; retention of organelles and plasma membrane disruption and blebbing [56,58,59]. Apoptotic neutrophils also exhibit molecular alterations on their cell surface, either by down-regulation of cell-surface receptors such as FcγRs (CD16 and CD32), complement receptors (CD35 and CD88), or TNF receptor (CD120b) [60] or by the exposure of new cell surface molecules such as phosphatidylserine (PS). Under normal conditions, PS is held on the inner leaflet of the plasma membrane by the actions of the enzyme flippase [61]. However, during the early stages of apoptosis, PS translocates to the outer surface of the plasma membrane where it facilitates the recognition of apoptotic neutrophils by macrophages [62]. This molecular switch is the basis of an apoptosis assay whereby fluorescently-labelled annexin V binds to exposed PS on the surface of apoptotic neutrophils and can be quantified by flow cytometry.

Neutrophil apoptosis can be activated via two pathways, dependent on whether the activating signal originates internally or externally to the cell.

1.3.5.1 The intrinsic apoptosis pathway

The intrinsic apoptosis pathway is regulated by members of the B-cell leukaemia-2 (Bcl-2) family of proteins, and in neutrophils is activated by an absence of anti-apoptotic factors, for example during constitutive apoptosis or in response to stress signals originating from within the cell, such as DNA damage. Interactions of the Bcl-2 family members (of which there are both pro- and anti-apoptotic members) ultimately leads to the de-polarisation and permeabilization of the mitochondrial membrane and the release of cytochrome-c into the cytosol [63].

1.3.5.2 The extrinsic apoptosis pathway

Exposure of neutrophils to pro-apoptotic signals results in apoptosis via the receptor-mediated extrinsic pathway. Molecules, including tumour necrosis factor-alpha (TNFα) Fas-ligand (FASLG) and TNF-related apoptosis-inducing ligand (TRAIL) bind to one of several "death receptors" expressed on the surface of neutrophils. These include FAS-receptor (FASR), TNF-receptor super-family member 10 (TNFRSF10) and nerve growth factor receptor (NGFR). These receptors are members of the larger TNF receptor super-family and are characterised by a cytoplasmic region of approximately 80 amino acids, known as a "death domain" [64]. Following activation of receptors via attachment with corresponding ligand, several cytoplasmic proteins are recruited to the death domain forming a death-inducing signalling complex (DISC). These proteins include Fas-associated via death domain (FADD), and pro- cysteine-aspartic proteases-8 (pro-

caspase-8). Formation of the DISC initiates a signalling cascade leading to activation of caspase 8 and other downstream caspases [65,66].

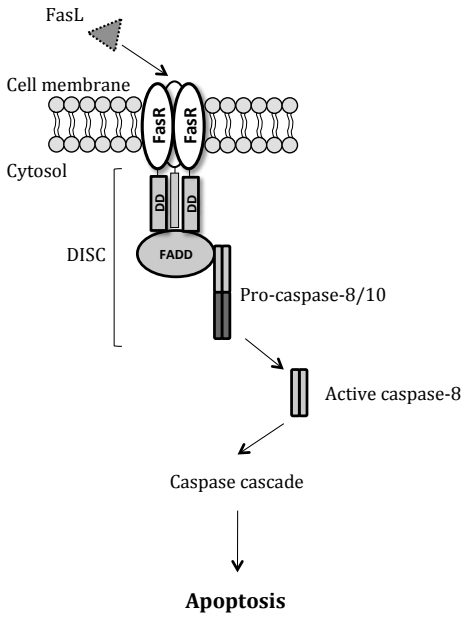


Fig 1.3 Schematic summary of the extrinsic apoptosis pathway in neutrophils. Activation of death receptors for example Fas-receptor (FasR) by its ligand Fas-ligand (FasL) induces formation of the death inducing signalling complex (DISC) comprising of the receptor death domain (DD), Fas-associated via death domain (FADD) and pro-caspase-8. Activation of caspase 8 leads to the initiation of a caspase cascade ultimately leading to apoptosis. Adapted from [64,67].

Ultimately, activation of either apoptosis pathway leads to the activation of initiator caspases, which sequentially cleave their inactive pro-caspase targets into shorter, active forms during a caspase cascade. The action of the caspase cascade results in the cleavage and degradation of a number of target cellular proteins which culminates in the disassembly of the cell and the exposure of "eat me" signals (such as PS) on the neutrophil surface leading to their clearance by phagocytic cells, such as macrophages or activated neutrophils [68].

1.3.6 Inflammation resolution

Appropriate activation of neutrophil apoptosis is crucial to limit any collateral damage by neutrophil granule proteins. Apoptotic neutrophils can also regulate the immune response, for example, by the secretion of annexin-1. The rapid release of annexin-1 from tertiary granules can induce increased uptake by macrophages [69]. Furthermore, uptake of apoptotic neutrophils results in the release of several anti-inflammatory mediators from phagocytic cells, such as IL-10, tumour growth factor-β (TGF-β) and prostaglandin-E_2 (PGE_2) which are crucial for the resolution of inflammation [70,71].

1.4 Neutrophils and disease

1.4.1 Neutrophil impairment

Given the indispensable role of neutrophils in host defence, it is unsurprising that disorders relating to impaired neutrophil function share a common phenotype of increased incidence of infection and overall immune deficiency in the patient [72]. For example, impairment of neutrophil adhesion and extravasation due to defective integrin or selectin expression is characteristic of leukocyte adhesion disorder (LAD) [73] , while an estimated 1 in 200,000 people suffer from chronic granulomatous disease (CGD), a group of inherited diseases caused by a defect in any of 4 subunits of the phagocyte NADPH oxidase complex [75].

Neutrophil myeloperoxidase (MPO) deficiency, is a disorder affecting approximately 1 in 4000 individuals [76]. Deficiency is often due to point mutations in the MPO gene, resulting in defective post-translational processing of the MPO precursor protein. Deficiency of MPO prevents the formation of hypochlorous acid (HOCl) from chloride and hydrogen peroxide (H_2O_2) [72]. Interestingly, despite the fact that neutrophils may be deficient of MPO, they do not have impaired bacterial killing ability *in vitro*, patients are usually asymptomatic unless presenting with other clinical disorders such as diabetes, whereby patients exhibit increased fungal infections [72]. This observation suggests a level of redundancy in neutrophil bacterial killing mechanisms.

1.4.2 Neutrophils and inflammatory disease

In contrast to diseases where neutrophil function is impaired, inappropriate or over-activation of neutrophils is associated with a number of inflammatory diseases. For example, neutrophils have been implicated in the pathogenesis of systemic lupus erythematosus (SLE) [77], Behçet's Disease [78], COPD [10] and RA [3]. Inappropriate release of neutrophil proteases and other anti-microbial proteins can lead to localised tissue damage, which characterises the pathology of each of the above diseases.

1.4.3 Low density granulocytes

Heterogeneity in granulocyte density has been known for almost two decades. In 1986, Hacbarth and Kajdacsy-Balla first described a sub-set of granulocytes that they termed "low buoyant density granulocytes" in the peripheral blood of patients with systemic lupus erythromatosis (SLE), rheumatoid arthritis (RA) and acute rheumatic fever, which correlated with disease activity [79]. More recently, low density granulocytes (LDGs) have been additionally associated with disease activity in a number of conditions such as HIV-1 [80] and SLE [81], whilst their function and phenotype have been further characterised.

LDGs are a sub-population of granulocytes which, due to their lower density, sediment in the PBMC fraction following density-gradient centrifugation of whole blood. They have been shown to have increased capacity for type I interferon (IFN) production, increased secretion of pro-inflammatory cytokines [81], decreased ability to phagocytose bacteria, but an increased tendency to form NETs compared to mature neutrophils [52]. LDGs are $CD15^{high}$ and $CD14^{low}$ (compared to monocytes which are $CD15^{low}$ and $CD14^{high}$), whilst also expressing the mature granulocyte markers CD10 and CD16 [32]. Conversely, the transcriptional profile of LDGs (in addition to their low density properties) suggests they are a sub-population of immature granulocytes [52]. The precise function of LDGs remains unclear

but understanding their function and precise role in inflammatory conditions is currently of great interest and importance.

1.4.4 Treatment of inflammatory disease

A hallmark of inflammatory disease is elevated cytokine levels at sites of inflammations. This is best demonstrated by the efficacy of anti-cytokine (or anti-cytokine receptor) therapies in treating a wide range of inflammatory diseases. Whilst multiple cytokines are elevated in these diseases, often, specific cytokine-blockade is highly successful at alleviating symptoms and decreasing disease activity. Indeed, therapies of this nature represent the front-line treatment for several debilitating inflammatory diseases [82]. However, responses of patients to specific therapies are often heterogeneous, and will often need to switch therapies before disease activity is controlled. This suggests that inflammatory diseases have heterogeneous pathology and different cytokines are responsible for driving the inflammatory response in different patients. Table 1.1 lists several examples of cytokines (or their cognate receptors) targeted by anti-inflammatory disease therapies, and the diseases they are administered for. Despite having a central role in several inflammatory diseases, direct targeting of neutrophils is unfeasible as a form of therapy due to the resulting neutropenia and increased risk of developing infections. However, modifying specific neutrophil functions by directly (or indirectly) targeting

cytokine signalling may lead to much more efficient regulation of neutrophils in inflammatory disease without compromising host-defence.

Table 1.1 Current (or clinical trial phase) cytokine-targeting-therapies for inflammatory diseases: Vasculitis (V); Inflammatory bowel disease (IBD); Rheumatoid arthritis (RA); Cancer (C); Multiple sclerosis (MS); Acute gout (AG); Still's disease (SD); juvenile idiopathic arthritis (JIA); ulcerative colitis (UC); Crohn's disease (CD); Psoriatic arthritis (PA); psoriasis (P); Ankylosing spondylitis (AS); Behçet's disease (BD).

Drug name	Target cytokine (or cytokine receptor)	Disease	Refs
Tocilizumab	IL-6 receptor	V/IBD/RA/C	83
MOR103 (phase II) **Mavrilimumab**	GM-CSF GM-CSF receptor	MS/RA	84,85
Anakinra	IL-1 receptor	AG/SD/JIA/RA	3,86–88
Infliximab (Remicade®) **Golimumab** **Adalimumab (Humira®)** **Etanercept (Enbrel®)** **Certolizumab-pegol (Cimzia®)**	TNFα	RA/UC/CD/IBD/PA/P/AS/BD	89–94
Fontolizumab	IFNγ	CD	95
Secukinumab	IL-17A	RA/P	96

1.5 Systems biology

1.5.1 Implementation

Systems biology is a relatively recent area of biological research. The development of high powered computing and analytical approaches which generate large amounts of data (such as proteomics, genomics and metabolomics) have necessitated a new approach to scientific research. The fundamental goal of systems biology is to integrate comprehensive biological data sets from diverse systems in an attempt to understand complex interactions at the molecular level, thus providing a mechanism of predicting phenotype changes in a biological system following a defined stimulus [97].

This approach has been applied to several areas of biology, such as analysing the entire kinase population (also known as the kinome) of *Drosophila melanogaster* during cell-cycle [98] or temporal analysis of gene-promoter activity of amino-acid biosynthesis genes in *Escherichia coli* [99].

In addition to individual systems based projects studying specific cells or pathways, broader, multi-institutional research consortiums are attempting to integrate data from multiple bioinformatic projects in an attempt to tackle wider conceptual problems such as inflammatory disease or patient drug-response. For example, a recently funded consortium led by the Queen Mary hospital, London – in collaboration with the charity Arthritis Research UK – will attempt to apply a stratified approach to understanding patient heterogeneity and drug response in rheumatoid arthritis (RA) [100].

This example of a multi-faceted research approach highlights how systems biology (or more specifically bioinformatic data) can successfully be up-scaled to undertake much broader scientific questions [101].

1.5.1 Transcriptomics

Genomic studies attempt to quantify the entire genome of an organism. This information is encoded by a cell's DNA and is virtually identical in all somatic cells of an organism, for example the information encoding the insulin gene (INS) is present in cells as disparate as pancreatic beta-cells and dendritic cells. Whilst, there may be variations in the copy number or epigenetic features, the hard-coded DNA sequence will be identical in both cell types, yet only in pancreatic beta-cells will the information be translated to produce the insulin protein.

In contrast, transcriptomics focuses solely on the information contained with the RNA population. The population of RNA transcripts is both cell specific and dynamically regulated and can be thought of as the functional intent of a cell. Furthermore, since an estimated 92-94% of eukaryotic genes are subject to alternative splicing [102] an RNA population may not only be different between two different populations of cells but also between similar cells stimulated differently.

In simplistic terms, transcriptomics attempts to define a level of gene expression at a specific time or under specific conditions by quantifying the

abundance of mRNA transcripts pertaining to each gene in an organism's genome [97].

1.5.1.1 Sanger sequencing

The first successful method of sequencing DNA fragments was developed in 1977 by Fred Sanger and colleagues using chain-terminating inhibitors [103]. Nucleotide-specific sequence fragments were created using terminating dideoxynucleotides for each of the 4 bases. Populations relating to each base-type were then placed in separate lanes of a gel and subject to electrophoresis. Since fragments were fluorescently labelled, and the terminating base-type was known, fragments could then be "read" by virtue of their migration distance through the gel [104]. This method of sequencing became the gold-standard in genetic research for over 25 years, until several technological advancements were made during the undertaking of the human genome project between 1990-2003 [105]. This technique is still employed for targeted-sequencing, but for larger projects and genome/transcriptome wide analysis it has been superseded by more high-throughput approaches [101].

1.5.1.2 SAGE

Early attempts to define an entire transcriptome utilised serial analysis of gene expression (SAGE) technology to sequence small unique fragments (15-20 bp) of cDNA transcripts (reverse transcribed from the RNA population) relating to each gene [106]. SAGE uses traditional Sanger technology to sequence short sequence tags which can be uniquely

associated with the original mRNA fragment to which it relates. The abundance of tags relating to each gene can then be used to get a quantification of gene expression. However, this technology is relatively expensive and often the short fragments cannot be uniquely associated to a specific gene. In addition, since only a small fragment of the parent mRNA is analysed, any isoforms of the same gene are often indistinguishable from each other and cannot be quantified [107]. Consequently, this technology is now more often used as a way of sequencing smaller, specific areas of a genome or transcriptome rather than a wider, global approach to gene expression.

1.5.1.3 Microarrays

Although initially developed as far back as the early 1980's to screen a small subset of genes in tumour cells [108], microarray technology has been consistently developed and improved, such that it became the first technology to be thought of as a truly high-throughput transcriptomic technology. Until recently, microarrays were the most established and popular method for studying nucleotide sequences on a massive scale. Microarrays contain thousands of single-stranded sequences of DNA (or RNA), known as probes, which are attached to specific location of a glass- or polymer-slide. Typically the RNA (or DNA) sample being measured is converted into a population of complementary fragments (cRNA/cDNA) and is labelled with a fluorescent dye. The sample is subsequently washed over the microarray, enabling labelled sequences with high complementary

sequence-similarity to bind to the microarray probes. Relative abundance of sequences can then be estimated by optical fluorescent measurement of each specific probe location on the microarray [109].

The hybridisation of labelled sequences to well characterised probes is a fundamental aspect of all microarray experiments. However, several different technological and conceptual variations of microarray technology exist providing huge versatility in the analysis of nucleotide sequences. For example, high-density oligonucleotide microarrays employ a dual probe system whereby one probe is designed to include a single "mismatched" nucleotide at the centre of the probe in comparison to the "perfect match" paired-probe. This serves as an internal control for hybridisation estimation and improves the accuracy of the system [110]. Elsewhere, microarray technology has been adapted to investigate protein-binding DNA-sequences by chromatin immunoprecipitation (ChIP) (a process more commonly known as ChIP-chip) [111], epigenetic studies [112], and quantification of non-coding RNA such as microRNA [113].

Although microarrays are a well-established and relatively inexpensive technology, they do present several limitations. For example, *a priori* knowledge of probe sequences is required to quantify expression levels of known gene transcripts, such that quantification of novel transcripts is not possible. Additionally, non-specific hybridisation of sequences to partially complementary probes increases the inherent background noise of a microarray experiment affecting the quantification accuracy [114].

Furthermore since hybridisation efficiencies can differ between probes, comparison of hybridisation results of different probes within a single experiment may not be as accurate as comparison of results for a single probe across multiple experiments [115]. Importantly, since quantification of probe hybridisation relies on an analogue measurement of fluorescence, the dynamic range of quantification is often limited to a few orders of magnitude, meaning that sequences present in low abundance are poorly defined whilst measurements of sequences with extremely high abundance can often become saturated [107].

1.5.2 RNA-Seq

Sequencing of RNA molecules using modern, so called "next generation" sequencing technology platforms (RNA-Seq) has, in recent years grown in popularity. Upwards of millions of fragments of DNA or RNA are sequenced in parallel using precise sequencing chemistry. Following sequencing, the use of powerful bioinformatic techniques enables the researcher to determine where each of the millions of reads originated from within the genome. This in turn builds up a density profile of mapped reads which, when cross-referenced with known gene locations can give a digital representation of the transcript abundance in the original sample for each gene. The rapid increase in speed and capacity, coupled with a dramatic decrease in cost has led to RNA-Seq superseding microarrays as the principal technology in transcriptomics. RNA-Seq offers several

advantages over microarray technology. Firstly, unlike microarray-based technologies, RNA-Seq is not limited to detecting transcripts that correspond to previously known sequences, making it suitable for studies involving non-model organisms [107]. Secondly, RNA-Seq has a much higher dynamic range of detection than microarrays, (shown to be approximately 5 orders of magnitude) [116]- meaning that low and highly expressed transcripts are equally well detected.

However, the greatest advantage of RNA-Seq technology over microarrays is the variety of data available from a single experiment. For example, quantification of gene expression of both coding [117] and non-coding RNA populations [118], splice-variant usage [119], single nucleotide polymorphism (SNP) discovery and allele specific expression [120] can all be extracted from the raw data of a single RNA-Seq experiment. Conversely, to gather such information using microarrays would require multiple bespoke arrays independently analysed, resulting in much higher costs and the introduction of unwanted technical variance.

1.5.3 Next generation sequencing

Inasmuch as Sanger sequencing is regarded as the first generation of sequencing technology, the 3 sequencing platforms provided by Roche (454 sequencing), Applied Biosystems' Sequencing by Oligonucleotide Ligation and Detection (SOLiD) and Illumina (Genome analyser I/II and HiSeq) represent the first iteration of the 2nd generation sequencers (also

known as next generation sequencing – NGS). Each were developed and released to market within months of each other but crucially, differed in their underlying mechanisms and sequencing chemistry. However, in the most simplistic terms, each of the technologies rely on the addition of deoxynucleotides (dNTPs) to a template DNA strand complementary to the DNA sequence fragment being sequenced, the platforms differ in their method of measurement and quantification of this addition. A short summary of the 3 platforms is detailed below.

1.5.3.1 Library preparation

Prior to sequencing by any of the sequencing platforms, sample DNA/RNA is subject to several preparation steps eventually producing a library of read fragments. Although, the specifics of each step are subtly different dependent on the eventual sequencing platform – for example read lengths differ greatly between sequencing platforms – the processes are largely similar.

1.5.3.2 RNA enrichment

The first step in processing an RNA sample is to enrich the population of RNA to be sequenced from a sample of total RNA, for example the mRNA transcripts or the microRNA population. This improves signal strength and avoids unnecessary sequencing of unwanted transcript populations. For standard RNA-Seq experiments enrichment can be achieved in a number of ways:

a) Terminator exonuclease treatment

Treatment of total RNA with a terminator exonuclease results in the removal of sequences without a 5'-cap, a physical feature of mRNA transcripts but not non-coding RNA (ncRNA). Thus ribosomal RNA (rRNA), tRNA and microRNA populations are depleted.

b) Ribosomal RNA depletion

Ribosomal RNA accounts for approximately 80% of a total RNA sample [121], hence strategies to deplete ribosomal RNA can efficiently enrich the remaining populations. This method is particular popular in studies that wish to quantify the microRNA populations – such as piwi-interacting RNA (piRNA) or small nucleolar RNA (snoRNA) – in addition to mRNA transcripts.

c) Duplex-specific-nuclease (DSN)

Enrichment and normalisation of low abundant transcripts within a population can be achieved by DSN treatment. DSN is a nuclease originally found in the Kamchatka crab *Paralithodes camtschaticus* [122]. The nuclease digests double-stranded DNA (dsDNA) with high specificity. During library preparation, following conversion of RNA to dsDNA (see later section for more details of library preparation) the sample is briefly denatured and then incompletely re-natured and DSN digested. Sequences which are highly abundant anneal more rapidly than less abundant sequences, thus digestion with DSN enriches low abundant transcripts and removes the most highly abundant population such as rRNA and transferRNA (tRNA) [122].

d) Poly-adenylated (Poly-A) tail selection.

In addition to a 5'-cap, mature mRNA transcripts feature a poly-A tail. This feature can be exploited by magnetic bead selection to positively select the mRNA population from a total-RNA sample. This method provides the most effective way of enriching mRNA transcripts and is the most widely used technique for sample enrichment. However, final yield of mRNA can be affected by this enrichment method, in particular if the RNA sample is degraded since Poly-A tail integrity is most likely to be affected during RNA degradation [123]

If the sample purity is of great importance and enough total RNA is available a combination of the above treatments can be carried out to increase purity – albeit at the cost of a lower yield of enriched sample [117]. The effect of these different treatments on the eventual sensitivity of the sequencing experiment is non-trivial and in addition to affecting which transcripts are detected in a sample, can also influence the amount of reads that ultimately map to intronic regions of the genome, since the population of immature mRNA (which still contain intron sequences) can be more- or less-enriched by the above treatments [124].

1.5.3.3 Fragmentation and cDNA conversion

Enriched RNA is subsequently converted into complementary DNA (cDNA) by reverse transcriptase and fragmented, typically by DNAse I treatment or

sonication [107]. Fragment lengths are dependent on the sequencing platform being used but can vary between 35 and 700 nucleotides long.

1.5.3.4 Adapter ligation and amplification

Single strands of cDNA are synthesised into double stranded cDNA, treated to generate blunt ends of sequences, and adapter sequences are ligated to both ends to aid in downstream sequencing protocols. Similar to read lengths, the size and sequence of the adapter sequences are both platform-specific. Finally sequences are amplified by PCR and fragments are size-selected by gel extraction.

1.5.4 Roche 454 sequencing

The 454 sequencing platform employs a "sequencing by synthesis" approach, quantifying the incorporation of dNTPs to a DNA template by the indirect measurement of pyrophosphate (PPi) release (Fig 1.3). Sample DNA is amplified, fragmented and hybridised with a sequencing primer in a reaction volume. Subsequently, each of the 4 dNTPs (adenine, guanine, cytosine or thymine) are added sequentially to the reaction mix. DNA polymerase catalyses the addition of a complementary dNTP to the template strand resulting in the release of PPi (which is proportional to the amount of nucleotide incorporation). The presence of ATP sulfurylase and adenosine 5´ phosphosulfate (APS) in the reaction volume results in the conversion of PPi to ATP. The ATP can then drive a luciferase-mediated conversion of luciferin to oxyluciferin, a reaction that generates visible light

proportionally to the amount of ATP generated – which can be measured in real-time. Finally the addition of a nucleotide-degrading enzyme - apyrase, removes any unincorporated dNTPs in preparation for a subsequent cycle of dNTP addition. Progressive iterations of nucleotide incorporation allows single-nucleotide resolution sequencing of reads by measurement of the signal peaks produced during addition [125].

The 454 platform produces the largest sequence reads of any of the 3 platforms discussed (200-700 nucleotides long). This is of great importance for downstream quantification of sequence reads, either when mapping reads to a reference genome or using reads for *de novo* assembly of a genome where a reference sequence is not available, for example in non-model organisms. The larger reads make the assembly process much less computationally demanding and require much less sequencing depth than would be necessary with much shorter reads [126]. However, since dNTPs are added sequentially, sequences with areas of successive nucleotides (homopolymers) are poorly quantified and subject to sequencing errors [125].

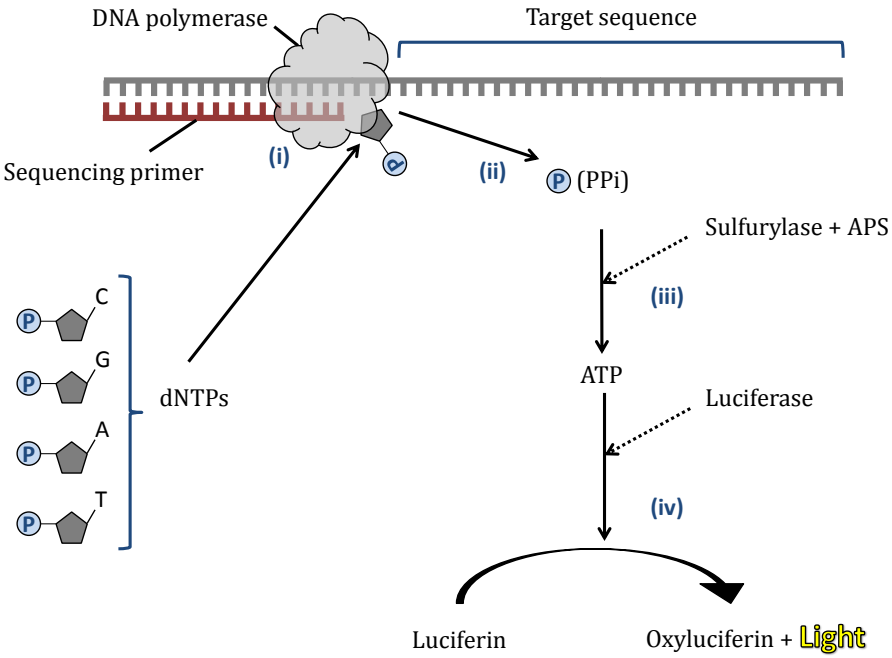


Fig 1.3 A schematic representation of Roche's 454 sequencing technology. dNTPs are sequentially added to a reaction volume containing millions of copies of target sequence with annealed sequencing primer. (i) DNA polymerase catalyses the elongation of the sequencing primer. (ii) incorporation of a nucleotide results in the release of pyrophosphate (PPi). (iii) PPi and APS are converted to ATP by the actions of Sulfurylase. (iv) Luciferase and ATP catalyse the conversion of luciferin to oxyluciferin which produces light in the visible range. Detection of light produced allows quantification of nucleotide addition. Adapted from [127,128].

1.5.5 Illumina sequencing (Genome analyser I/II and HiSeq)

Illumina sequencing technology makes use of reversible dye terminators to sequentially measure each nucleotide position by measurement of fluorescent emission (Fig 1.4). As with pyrosequencing, sample RNA is enriched, fragmented and adapter sequences ligated. However, sequences are attached to a glass slide by their adapter sequences and amplified *in situ* in a process called bridge amplification [129], this increases the platforms throughput capacity over other platforms [130]. A primer sequence is annealed to fragment reads and sequencing proceeds in a single base

synthesis fashion. Unlike pyrosequencing, all four dNTP types are added together during sequencing cycles to ensure competitive binding of nucleotides – this increases sequencing accuracy of homopolymers . Each of the 4 dNTPs are modified to include a terminating group which inhibits sequence extension and are labelled with a different dye for identification. During each cycle, (due to the terminating group) a single nucleotide is incorporated to the primer sequence. Following the incorporation of a dNTP the remaining unincorporated nucleotides are washed away and the newly incorporated nucleotide identity is determined by laser excitation of the dye. Both the terminating group and labelled dye are cleaved from incorporated nucleotide and the cycle is repeated for the addition of the next nucleotide. Extension of template sequence continues for the entire length of sequence read.

The entire process is carried out on an 8-lane flow cell, which enables massively-parallel sequencing of millions of reads [131].

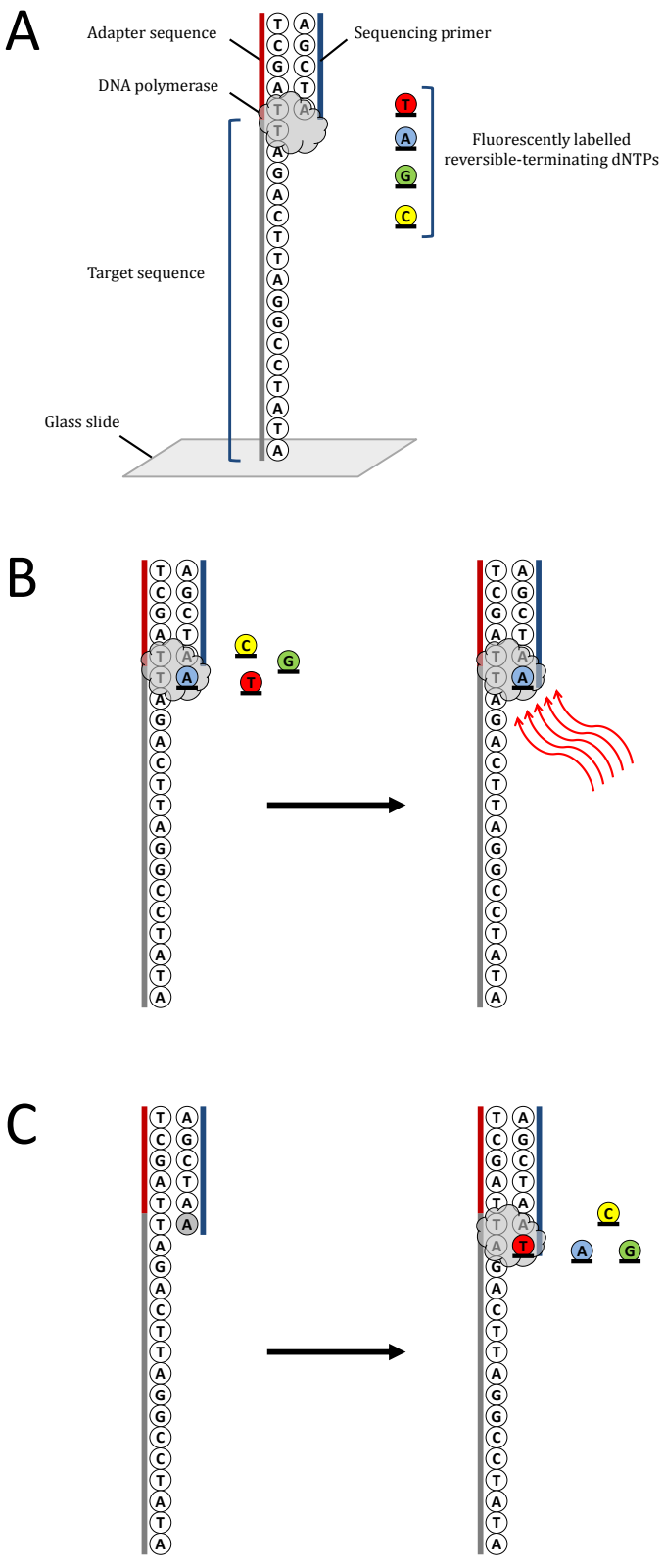


Fig 1.4 Schematic representation of Illumina sequencing technology. (A) Target sequence is attached to flow-cell and sequencing primer is annealed. (B) DNA polymerase catalyses the addition of 1 of 4 possible dNTPs. Incorporated dNTP is identified by laser excitation. (C) Terminating group and dye portion is removed from incorporated dNTP and the process is repeated for full length of target sequence.

1.5.6 SOLiD™ sequencing by Applied Biosciences

Similarly to the Illumina platform, the SOLiD™ sequencing platform sequences millions of fragmented reads in parallel on a flow cell following a standard set of read library preparation steps (enrichment, fragmentation, conversion to cDNA, adapter ligation and size selection). However, the SOLiD™ platform has several features which distinguish it from both the Illumina and 454 platforms. Firstly, unique read fragments are attached to a micro-bead (approximately 1 μm in diameter) and amplified by PCR (in a process known as emulsion-PCR) to produce a monoclonal bead with thousands of identical reads attached to it, this amplifies the signal when detecting nucleotide additions during sequencing. The beads are subsequently attached to the flow cell for sequencing [129]. Secondly, rather than employing a polymerase for template extension, the SOLiD™ platform relies upon ligation of labelled probes for strand extension [132]. Thirdly, raw sequencing data is encoded using a "colorspace" encoding system which increases accuracy of base calling by deciphering between sequencing errors and legitimate SNP events in the read fragments [133,134].

1.5.6.1 Di-base probes

Rather than the sequential addition of single dNTPs by a polymerase as seen in the 454 and Illumina platforms, the SOLiD™ platform uses a "sequencing by ligation" approach whereby fluorescently labelled di-base

probes are ligated to the primer strand by the enzyme ligase [131]. Probes are 8 nucleotides long and are labelled with one of 4 fluorescent dyes. The sequences of the probes differ in the first two bases of the 3′ end, thus each possible combination of two bases is represented by 16 different probes (4 probes of each colour). The remaining portion of the probe is made up of random hexamers. For each di-base, the reverse sequence (for example CA and AC), the complementary sequence (for example CA and GT) and the reverse complementary sequence (for example CA and TG) are always represented by the same colour (Fig 1.5).

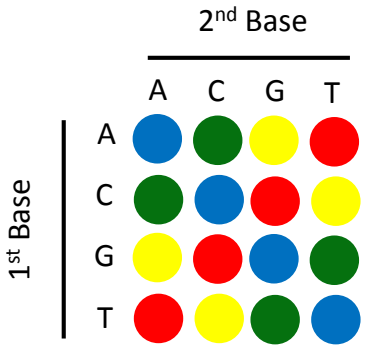


Fig 1.5 Representation of the 2 base colorspace encoding system employed by the SOLiD™ sequencing platform di-base probes. Each of the 16 different combinations of 2 bases are represented by one of 4 colours. Pairs of the same colour are either reversed, complementary or reversed-complementary of each other. Adapted from [135].

1.5.6.2 SOLiD™ sequencing steps

Sequencing begins with the ligation of di-base probes to the primer sequence under competitive conditions (all 16 possible probes are present) (Fig 1.6A). Following ligation, sequences are imaged and the fluorophore portion of probe is removed such that a 5 nucleotide portion is

incorporated into primer sequence. The process is repeated for the entire length of the sequence fragment (Fig 1.6B) (for example a 35 nucleotide sequence fragment requires 7 ligation cycles). Since only information on 2 nucleotides of each probe is available, only 2 in every 5 nucleotides in a sequence-fragment are interrogated per ligation cycle (Fig 1.7). Furthermore, since each colour probe relates to one of 4 possible nucleotide pairings, it is not possible to determine the base sequence from a single interrogation. Thus, following ligation cycles, the newly extended primer sequence (and primer) are removed and a new primer sequence is added which is offset by 1 nucleototide (n-1) (Fig 1.6C). Ligation cycles are repeated for new primer round and a further 3 primer rounds (n-2,n-3 and n-4) (Fig 1.6D). Following a total of 5 primer rounds, each nucleotide will have been interrogated twice on two separate primer rounds (Fig 1.7), providing sufficient information to decode the sequencing information from colorspace into basespace (raw nucleotide information).

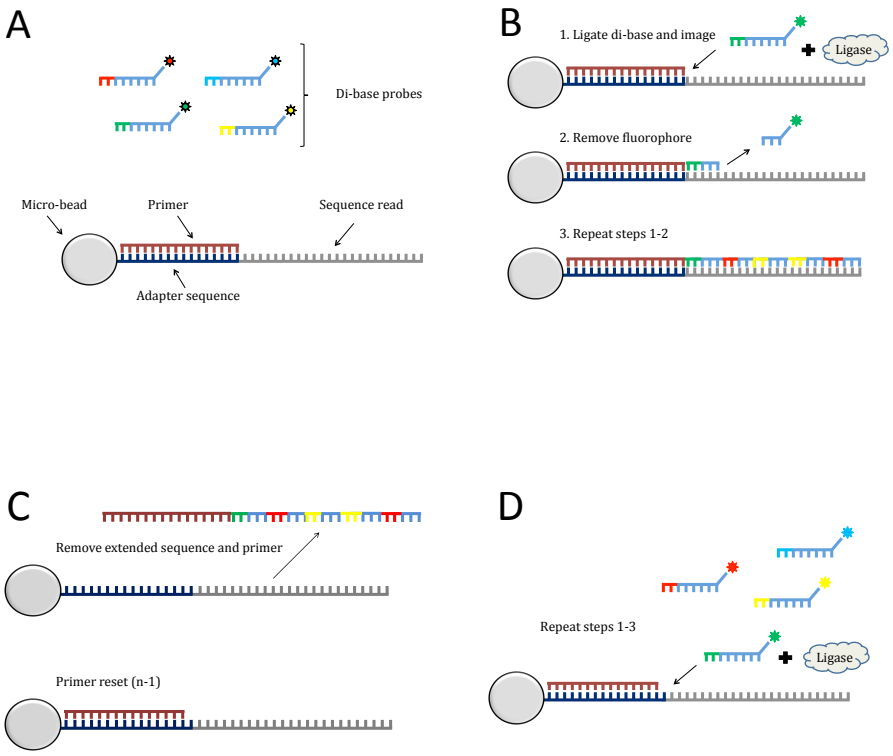


Fig 1.6 SOLiD sequencing methods. A) Read library-fragments are attached to a micro-bead and annealed with a primer sequence in preparation for sequencing. B) 16 Di-base probes labelled with one of 4 coloured dyes competitively bind to primer sequence extending template strand, ligation catalysed by ligase enzyme. Di-base probe incorporation is imaged by laser excitation. Fluorophore portion of probe is removed and process of probe ligation is repeated for entire length of read-fragment. C) Newly formed extended sequence is removed and primer sequence is reset using a 1 nucleotide shorter primer (n-1). D) Probe incorporation steps are repeated for 4 additional primer rounds. Adapted from [135].

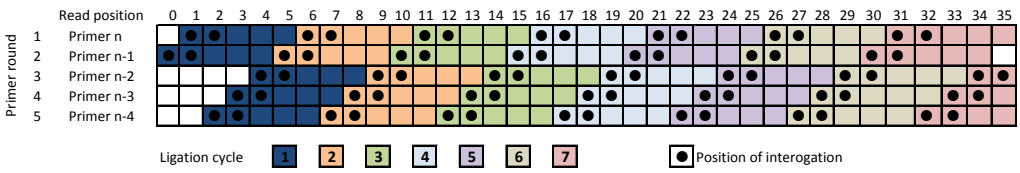


Fig 1.7 Representation of the dual interrogation of read positions by multiple ligation cycles and primer rounds in SOLiD sequencing platform. Di-base probe incorporation during SOLiD sequencing extends a primer sequence by 5 bases at a time, interrogating the first two bases on each cycle. Multiple ligation and primer rounds result in each read position being measured twice on separate primer rounds. Colorspace encoded information can them be used to accurately determine the original nucleotide sequence. Adapted from [135].

1.5.6.3 Colorspace

The bioinformatic advantage of the colorspace encoding system is two-fold. Firstly, having each read position interrogated twice improves the accuracy of base calling. Secondly, since each read position is represented by two colours, if an appropriate reference sequence is available, colorspace encoding system provides a means to distinguish between sequencing errors and true single nucleotide polymorphisms (SNPs) that may exist between the reference sequence and the sample sequence. For example, due to the structure of the encoding system, SNPs are identified by both colours of a particular read position differing from the reference sequence, whereas a sequencing error would result in a single change in colour to the reference (Fig 1.8) [136].

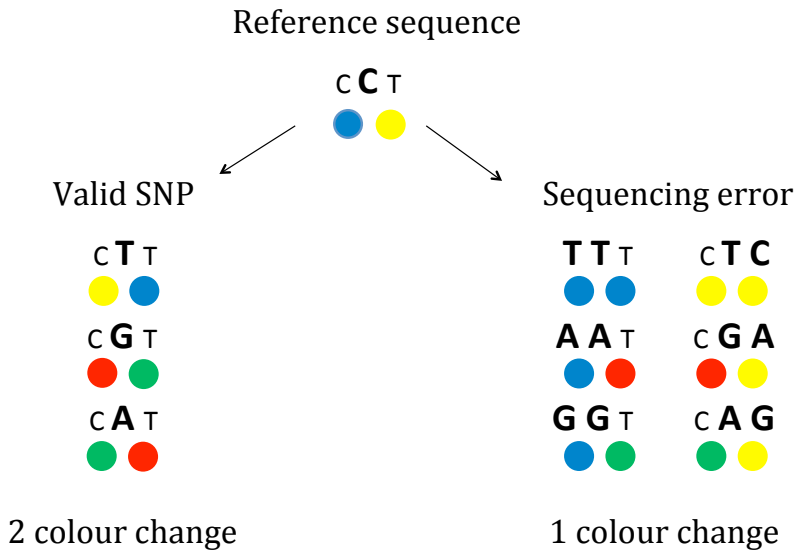


Fig 1.8 Example of colorspace encoding in the SOLiD™ platform. Encoding enables downstream identification of valid single nucleotide polymorphisms (SNP) or sequencing errors. There are 3 possible single nucleotide polymorphism of the central nucleotide in a 3 nucleotide codon, each change results in a two colour change in colorspace. Conversely, an error in sequencing would cause a single change in colour and consequently a change in 2 nucleotides when converted back to basespace. Adapted from [135].

1.5.7 Indexing/barcoding

An added benefit of all of the above technologies is the ability to "index" read fragments (also known as barcoding). Since the adapter sequences can be modified to include an additional bespoke sequence (usually 6-10 nucleotides long), read fragments from a common RNA sample can be indexed. This means that up to 96 RNA samples from separate experiments, time points or even organisms can be run in parallel on the same flow chip without losing information on which sample the reads originated from, the raw sequencing data is re-grouped into sample data sets following sequencing[129,137].

1.5.8 Paired-end sequencing

An alternative approach to sequencing a read library which is available on each of the discussed platforms is paired-end sequencing. Rather than sequencing a read in a single direction as per single-end sequencing. Paired-end sequencing incorporates an additional sequencing step initiated at the opposite end of the read, thus two portions of the same fragment are sequenced in opposite directions. The benefit of this approach is not only the additional information that can be collected from a single sequencing run but that since both portions of sequencing data are known to originate from the same read fragment this information can be used to more accurately map both reads to the reference genome in the bioinformatic mapping stage. For example a short read containing an

area of nucleotide repetition may map to multiple areas of the genome due to its low sequence complexity. However, paired reads originating from the same area of the genome which are separated by an estimated distance are much less likely to map to multiple genome locations and are therefore much easier to assign a mapping location. Variations of paired end sequencing include mate-pair sequencing where two portions of a single read are sequenced in the same direction and the resulting distance between mate-pairs is much larger (2k-5k nucleotides) [104].

1.5.9 Future generation sequencing

The fundamental mechanics of each of the 3 platforms described above have not been altered in almost a decade. Constant development of the technology has meant that successive iterations of the hardware and software in each platform have seen improvements in sequencing speed, capacity and accuracy and a reduction in costs of both the sequencing machinery and biological reagents. However, future generations of sequencing technology are emerging and could provide a significant increase in speed and usability such that whole genome sequencers could become more commonplace in the laboratory or clinical environment. For example nanopore-technology is a method of sequencing a strand of DNA or RNA by passing it through a nanopore where an electrical current exists across the pore. As individual bases pass through the pore they can be identified by the amount to which they disrupt the current across the pore

[101,138]. Alternatively, Pacific Biosciences technology employs nucleotides which are fluorescently labelled on their phosphate group. Using nano-visualisation chambers known as zero mode waveguides (ZMWs) which have a volume of 20 zeptolitres (10^{-21} litre), the incorporation of nucleotides can be detected against the high background signal of surrounding unincorporated nucleotides [128]. The advantage of both these systems is that nucleotide incorporation or movement across a nano-pore can be measured without the need for interruption, thus sequencing can be measured in real time with the only limiting factor being the rate at which a DNA/RNA polymerase can operate [139]. The successful development of a 3rd generation of sequencers will decrease the amount of time taken to sequence a human genome to a matter of minutes rather than days [140].

1.6 Bioinformatic software

The primary function of any high-throughput sequencing technology is to turn a biological sample of DNA or RNA into digital data. A single sequencing run can produce upwards of 60GB of raw data [141] which, when typed equates to approximately $22x10^6$ sheets of A4 paper (size 12 font, single line spacing). In addition to being computationally demanding due to its size, data from a next generation sequencing platform is also produced in a raw format. Following a sequencing run, data is filtered to remove incomplete reads and adapter sequences. Indexed data can be separated into sample-associated sets, finally, the raw data is then collated into a single file (or pair of files dependent on sequencing platform) containing only the read data and an accompanying quality value for each read position. Thus, a raw data file from a single sample in an RNA-Seq experiment may simply contain > $40x10^6$ lines of data each 50 characters long (for a 50bp single-end experiment).

For RNA-Seq data, several computational processes are required to extract meaningful data from the raw dataset produced by a next generation sequencer. These include: assembly of transcriptome, read mapping to reference sequence and expression quantification. Each process requires a specific set of software which often allows running parameters to be manually altered to improve performance depending on the type of input data provided and/or output data required.

1.6.1 Transcriptome assembly

If an adequate reference sequence is not available (for example in transcriptome studies using non-model organisms), quantification of gene expression can be achieved by *de novo* assembly of a transcriptome using sequenced reads. Successful assembly relies on sufficient depth of coverage from sequence reads and is aided by larger read lengths such as those produced by the 454 platform. Software programs such as Trinity [142], Trans-ABySS [143] and Rnnotator [144] are able to *de novo* assemble transcriptomes using RNA-Seq data, whilst software such as Cufflinks [145], Scripture [146] and ERANGE [116] are able to use a reference sequence to build upon in a process known as *ab initio* assembly [147]. The use of either *de novo* or *ab initio* assembled transcriptomes provides a way of identifying and quantifying novel transcripts that may be present in a cell specific manner.

1.6.2 Read mapping

Quantification of raw sequence data begins with the process of mapping reads back to a reference sequence (also known as read alignment). Mapping involves the determination of genomic origin for each read within a dataset [148]. Reads can be mapped to a reference genome or transcriptome sequence. Mapping to a genome sequence provides a more comprehensive quantification of reads since reads which map (either partially or entirely) to non-coding areas (for example intronic regions) are

also quantified. Ideally, mapping of reads would result in the assignment of each raw read to a single location within the reference sequence producing a mapping profile which can be transformed into an expression profile for each location (based on the number of reads which are assigned to that location). However, in reality, several factors impact the level of mapping achievable. Firstly, the reference sequence is often not a perfect representation of the biological source of RNA that was sequenced. For example, sample (or cell-specific) SNPs and insertion-deletion events (indels) represent areas of variation between samples which cannot be represented by a ubiquitous reference sequence. *De novo* assembly of a reference sequence using raw reads can limit this level of inherent variation but often requires a much greater read-depth coverage, ultimately increasing the overall cost of a sequencing experiment for very little added accuracy [149]. Secondly, the use of sequencing platforms which produce relatively short reads (35-50bp) such as SOLiD™ or Illumina increases the chances of reads mapping to multiple locations, in particular for areas of high repetition [107]. Thirdly, high levels of sequencing errors (as might be expected from low integrity RNA samples) can heavily impact the mapping rates [150]. Thus, the percentage of reads mapped during the mapping process can inform on both the success of the sequencing experiment and original quality of the RNA sample.

Mapping software programs take a raw sequencing file (such as a *.fastq* or a *.csfasta* for colorspace reads) as input, and where necessary an

accompanying quality file (*.qual*). Often the reference sequence to be used for mapping against will be heavily indexed by the mapping software prior to a mapping run such that the large volume of data included in a reference genome (or transcriptome) can be computationally managed, ultimately this makes the mapping process much quicker and decreases the amount of random access memory (RAM) required by the software. Reads are aligned to the reference sequence based on their entire sequence but often, to increase mapping speed, a small portion of the read (known as the seed) is used to initially map the read to a number of locations, once the seed sequence is aligned, the remaining portion of the read is aligned and the number of possible mapping locations decreased. Mapping builds up a profile of alignments against the reference sequence that can subsequently be quantified by downstream software (Fig 1.9A-B). When mapping to a reference genome reads that align entirely within transcribed portions of the genome (that is, exons) will be mapped easily by any mapping software. However, reads that originate from transcripts that were subject to splicing events are much harder to align to a reference genome. Indeed, for several mapping programs – such as Bowtie [151], Burrows-Wheeler alignment tool (BWA) [152] and short oligonucleotide alignment program (SOAP) [153] – reads relating to spliced transcripts will fail to map to a reference genome, thereby reducing the overall percentage of reads mapped (Fig 1.9C). To avoid this, several mappers have been designed to be able to handle reads that span splice junctions (Fig 1.9D), these

programs include Blat [154], Tophat [155], GSNAP [156], SpliceMap [157] and MapSplice [147,158].

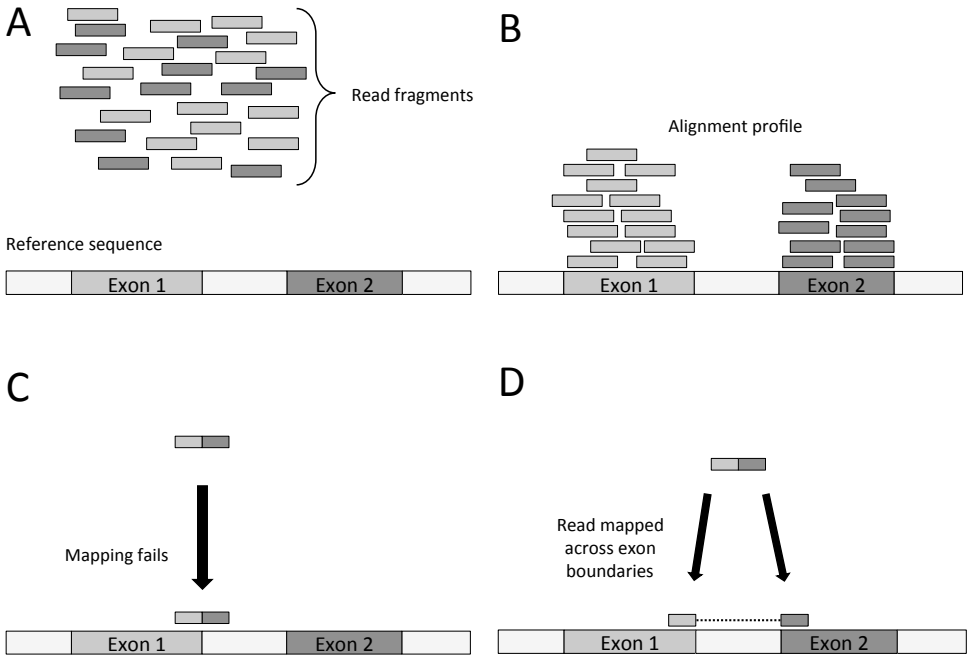


Fig 1.9. Schematic representation of read mapping of NGS data. A) Read fragments and reference sequence are provided to mapping software. B) Reads are aligned to reference sequence to produce a virtual alignment profile. C) Within a read library, a proportion of the reads will originate from spliced transcripts, standard mapping software will fail to align reads that span splice junctions. D) Specialized mapping software such as Tophat [155] can align read portions independently thus allowing mapping of reads that span splice junctions.

Of the multiple different mapping software packages available, many employ different mathematical algorithms when aligning reads. This variation in software is highlighted by the vast difference in speed and accuracy achievable by each of the mappers. For example, a comparison in 2011 of the most commonly used mappers available revealed that the percentage reads mapped ranged from 17.7% - 88.7% and time taken varied between 1.5 h – 145 h amongst 8 different mappers – despite each

software being provided with identical raw data [148]. A common feature of most mapping software is the ability to manually alter the mapping parameters and levels of confidence in aligning reads. For example, the number of allowed mismatched bases between reads and reference in a valid alignment or the number of multiple mapping locations permissible before excluding a read from alignment can both be altered to increase the number of reads mapped – often at the cost of accuracy [155].

1.6.3 Expression quantification

Following mapping, reads must be annotated and assigned to a biologically meaningful unit for example, exons, transcripts or genes. Since mapping of reads produces a set of genomic (or transcriptomic) coordinates for each read, annotation is reliant on a suitable set of reference coordinates relating to the genomic feature to which the reads are to be quantified. For example, the simplest approach is to quantify the number of reads that align to known genes within the genome, thereby arriving at an expression value for each gene. However, this approach makes no allowances for genes that overlap and also eliminates the possibility of quantifying novel transcripts. Alternatively, each exon can be quantified independently, allowing for quantification of: whole genes (by aggregation of exon expressions); exons, or individual splice variants of a specific gene. The choice of software for quantification and annotation is therefore crucial to the type of quantification required.

1.6.3.1 Sequence alignment map (SAM file)

The output file produced from the previously described mapping/alignment software is predominantly a sequence alignment map (*.sam* file) or an equivalent binary version (*.bam*). Both of these file types are compatible with most annotation/quantification software. *Sam/bam* files contain information on: sequenced reads, including quality values for each base position; alignment information, such as the number of locations a read aligned to in the reference sequence; and number of mismatches and/or indels within read sequence (compared to reference sequence provided during mapping stage).

1.6.3.2 Count-based quantification

Broadly, there are two popular approaches to quantifying RNA-Seq data, the first of these is count based quantification. Software programs such as edgeR [159], DESeq [160] and DEGSeq [161] quantify the number of reads which align to any portion of a genes within a reference sequence (including intronic and untranslated regions). This number is then normalised to the size of the entire read library for that sample, such that values for a single gene can be compared between separate samples. This method of quantification is often used for broad analysis of differential expression of genes since no calculations of individual exons are made [162]. However, since reads are only normalised to their library size, no allowance is made for the size of each gene. For example, given two hypothetical genes (Gene A and Gene B), assuming Gene B was twice as long as Gene A, and

both genes had equal expression within a sample, Gene B would have twice as many reads mapping to it due to its size alone. Consequently, count based quantification would incorrectly identify Gene B as having double the expression value of Gene A. Thus, direct comparison of expression values for two genes within the same sample set is not achievable using the count based approach.

1.6.3.3 Fragment-based quantification

Alternatively, the software program Cufflinks [145] is able to quantify expression values for individual exons of a transcripts, based on a set of splice locations provided to the software or by *de novo* discovery of alternative spice junctions during mapping stage. Consequently, due to individual quantification each exon of a gene, in addition to gene expression analysis, Cufflinks can be used to quantify expression of alternative isoforms of genes and novel transcripts. However, since alternative isoforms of a gene often share large portions of common sequence, quantification is achieved using a high degree of statistical assumption and modelling of the variation within a single transcript [149].

1.6.3.4 Reads per kilobase of exon model per million mapped reads (RPKM)

Importantly, Cufflinks employs an additional normalisation step and produces the expression metric, reads per kilobase of exon model per million mapped reads (RPKM) [116], which is defined as:

$$RPKM = \frac{\text{No. of mapped reads x 1 kilobase x 1 million mapped reads}}{\text{Length of transcript x No. of total reads}}$$

Therefore, in addition to normalising to the total number of reads in each sample, RPKM is also normalised to the size of the transcript. This means that expression values for two genes within the same sample can be compared. Equally, expression of a single gene in multiple separate experiments can be compared.

1.6.4 Differential expression testing

Following quantification of gene expression by either count-based or RPKM-based techniques, a variety of statistical tests can be performed between multiple data sets to determine if genes are differentially expressed (DE) between samples. Each of the software programs available for annotation and quantification of gene expressions provide a method of determining DE genes, but vary in the way the data is modelled. Cufflinks used a negative binomial distribution to model the data for differential analysis whereas several of the count based software programs use a Poisson distribution (for example edgeR, DEGSeq and DESeq). Whilst the distribution of read fragments across a genome in distinct locations (genes) can be thought of as a Poisson distribution a negative binomial modelling is considered more appropriate when modelling variation that arises from multiple biological replicates since Poisson-based analyses are prone to

high levels of false positives due to an underestimation of the biological variance [149,163].

1.6.4.1 False discovery rate

The traditional threshold for significance in a biological experiment is $p< 0.05$ – that is, there is less than a 5% chance of the observed quanta occurring by chance alone. When dealing with a large number of variable such as during a global gene expression analysis, this threshold no longer provides a satisfactory measure of significance. For example, if comparing the expression values for every gene in the human genome (approximately 23,000 genes) for genes that are DE between two samples, using a $p< 0.05$ would result in > 1000 genes being false positives. Hence, it is more appropriate with transcriptome and genomic studies to use a transformed p-value (known as a q-value) which applies a false discovery rate (FDR) of 5% to the data. Therefore, $q< 0.05$ implies that there is less than 5% chance the gene(s) in question are differentially expressed by chance alone allowing for a 5% false discovery rate. FDR is controlled by Benjamini-Hochberg correction [164].

1.7 Summary

Neutrophils play a key role in host defence via a number of evolutionarily conserved anti-microbial mechanisms[2,23,165]. Rather than acting merely as a front-line defensive cell, neutrophils are now recognised as central to the inflammatory response through the production of several inflammatory mediators regulating both the innate and adaptive immune systems [5]. The global molecular changes that underlie neutrophil priming and immune regulation are poorly defined, yet represent an attractive area of research to fully elucidate the role and regulatory capacity of neutrophils during the immune response. The speed, accuracy, and robust nature of RNA-Seq as a quantification platform, and the ability to extract data relating to multiple genetic features from a single sequencing experiment have all contributed to RNA-Seq superseding micro-array analysis as the gold-standard method of transcriptome analysis. However, a global approach to analyse neutrophil gene expression using RNA-Seq has yet to be undertaken. A greater understanding of the molecular properties of both quiescent and primed neutrophils will not only expand our knowledge of neutrophil biology but will inform on how neutrophils contribute to many inflammatory diseases, ultimately providing new areas of research into neutrophil regulation in health and disease.

1.8 Hypothesis and aims

The hypothesis that this project sought to test was:

Different cytokines lead to changes in neutrophil phenotype, but do so via different signalling pathways that lead to the switching on or off of different sets of genes.

The aims of this project were to:

1. Develop and define a robust pipeline of bioinformatic software programs and protocols to accurately quantify and analyse the neutrophil transcriptome under conditions of simulated inflammation by *in vitro* stimulation.

2. Quantify the changes in gene expression in neutrophils following priming with inflammatory cytokines.

3. Identify changes in gene expression profiles between neutrophils stimulated with different inflammatory mediators.

4. Analyse the effects of neutrophil isolation methods and neutrophil purity on gene expression and function.

Chapter 2: Materials and Methods

2.1 Materials

Table 2.1 Cell isolation and culture

Cell isolation and culture	
Materials	**Supplier**
Lithium heparin vacuette	Greiner Bio-one, Thermo Fisher Scientific (Gloucestershire, UK)
Safety butterfly needles	
RPMI 1640 (+ 25mM HEPES with L-glutamine)	Gibco, Life technologies (Paisley, UK)
Dulbecco's phosphate buffered saline	
Polymorphprep ™	Axis-shield (Cambridge, UK)
Isoton diluent	Beckman Coulter Inc.
Ammonium Chloride lysis buffer - Ammonium Chloride (NH_4Cl) - Potassium hydrogen carbonate ($KHCO_3$) - Ethylenediaminetraacetic acid (EDTA)	Sigma (Poole, UK)
Foetal bovine serum	Invitrogen, Life technologies (Paisley, UK)
Lymphoprep ™	Stemcell technologies (Grenoble, France)
EasySep® Human Neutrophil enrichment kit	
HetaSep®	
Rapid Romanowsky stain	HD Supplies (Aylesbury, UK)

Human AB serum	Sigma (Poole, UK)
Paraformaldehyde (PFA)	
Dimethyl sulphoxide (DMSO)	

Table 2.2 Cytokines/chemokines/stimulants/inhibitors

Cytokines/Chemokines/Stimulants/Inhibitors	
Materials	**Supplier**
Recombinant human G-CSF	Sigma (Poole, UK)
Recombinant human IL-6	
Recombinant human IL-1β	
Recombinant human IL-8	Invitrogen, Life technologies (Paisley, UK)
Recombinant human IL-10	
Recombinant human IFN-α	
Recombinant human IFN-β	
Recombinant human GM-CSF	Roche diagnostics (East Sussex, UK)
Recombinant human IFN-γ	
Recombinant human TNF-α	Calbiochem (Nottingham, UK)
fMLP	Sigma (Poole, UK)
PMA	
LPS	Source BioScience (Nottingham, UK)
MALP	Enzo Life Sciences (Exeter, UK)
JAK-1 inhibitor	Calbiochem (Nottingham, UK)
Wedelolactone	

Table 2.3 Primary and secondary antibodies

Primary and Secondary antibodies	
Materials	**Supplier**
anti-human STAT1 (#9175)	Cell Signalling (Massachusetts, USA)
anti-human-phosphorylated STAT1 (#9177)	
anti-human STAT3 (#9138S)	
anti-human-phosphorylated STAT3 (4904P)	
anti-human ERK (#9102)	
anti-human-phosphorylated ERK (#9106S)	
anti-human p38 (#9212)	
anti-human-phosphorylated p38 (#9216S)	
anti-human-AKT (#9272)	
anti-phosphorylated-AKT (#4060S)	
anti-human-IKβα (#9242)	
anti-human-phosphorylated NFκβ (#3033S)	
Rabbit-anti-IL-1β (#12703)	
Mouse anti-human β-actin (#8226)	Abcam (Cambridge, UK)
Mouse anti-human GAPDH (G8795)	Sigma (Poole, UK)
HRP-conjugated-sheep anti-mouse (A5906)	
HRP-conjugated donkey-anti-rabbit (25005179)	GE healthcare Life Sciences (Buckinghamshire, UK)
FITC-conjugated mouse anti-human CD11b (FAB16991P)	R&D Biosystems (Oxfordshire, UK)
FITC-conjugated mouse anti-human CD4 (FAB3791F)	

FITC-conjugated mouse anti-human CD15 (F0830)	Dako UK (Cambridgeshire, UK)
FITC-conjugated mouse anti-human CD64 (FAB12571F)	R&D Biosystems (Oxfordshire, UK)
FITC-conjugated mouse anti-human CD16 (555406)	BD Biosiences (Oxfordshire, UK)
FITC-conjugated mouse anti-human L-Selectin (BBA33)	R&D Biosystems (Oxfordshire, UK)
FITC-conjugated mouse anti-human IgGI isotype (SC-2855)	Santa cruz biotechnology (Heidleberg, Germany)

Table 2.4 Samples preparation and western blot

Sample preparation and western blot	
Materials	**Supplier**
Laemmli buffer -Glycerol -Sodium dodecyl sulphate (SDS) -Tris	Thermo Fisher Scientific (Gloucestershire, UK)
- Dithiothreitol (DTT) -Bromophenol blue	Sigma (Poole, UK)
Hydrochloric acid (HCl)	VWR International (Leicestershire, UK)
Glycine	Sigma (Poole, UK)
Sodium Chloride	
Ammonium persulphate (APS)	
Isopropanol	
Methanol	
Tween-20	
Tetramethylethylenediamine (TEMED)	

Sodium azide	
Hydrogen peroxide (H_20_2)	
Whatman filter paper	
Polyacrylamide	Geneflow (Staffordshire, UK) (Kiddeminster, UK)
Biotinylated protein ladder detection pack	Cell signalling (Massachusetts, USA)
BLUeye prestained protein ladder	Geneflow (Staffordshire, UK)
Phosphatase inhibitor cocktail II	Calbiochem (Nottingham, UK)
Marvel non-fat (< 1%) dried milk powder	Home Bargains (Liverpool, UK)
Ponceau S	Millipore (Hertfordshire, UK)
Bovine serum albumin	
Kodak® photographic Fixer and Developer	
Immobilon Western Chemiluminescent HRP Substrate	
Polyvinylidene fluoride (PVDF) membrane	GE Healthcare Life Sciences (Buckinghamshire, UK)
Enhanced chemiluminescence hyperfilm	

Table 2.5 Neutrophil chemiluminescence and apoptosisaa

Neutrophil chemiluminescence and apoptosis	
Materials	**Supplier**
Hank's balanced salt solution (HBSS)	Gibco, Life technologies (Paisley, UK)
Luminol	Sigma (Poole, UK)
Propidium iodide	
Alexa Fluor 488-conjugated Annexin-V (#A13201)	Life technologies (Paisley, UK)

Table 2.6 RNA isolation cDNA synthesis and qPCR

RNA isolation, cDNA synthesis and qPCR materials	
Materials	**Supplier**
Trizol®	Gibco, Life technologies (Paisley, UK)
Chloroform	Sigma (Poole, UK)
2-Propanol (molecular grade)	
Ethanol (molecular grade)	
RNase-free DNase set	Qiagen (Crawley, UK)
RNeasy mini kit	
Quantitech SYBR green PCR kit	
Superscript III first strand cDNA synthesis kit	Invitrogen, Life technologies (Paisley, UK)
RNase OUT	
Random primers	Promega (Southampton, UK)
Specific primers	Eurofins (UK)

2.2 Methods

2.2.1 Ethical approval

Ethical approval for the study of neutrophils from adult healthy controls was granted by the University of Liverpool Committee for Research Ethics (CORE). All participants gave written, informed consent.

2.2.2 Leukocyte isolation

Blood was collected from healthy volunteers by venupuncture into lithium heparin-coated vacuettes and processed immediately. Blood taken for RNA-Seq processing was taken from volunteers at a similar and consistent time of the day (9 am - 11 am) to mitigate variation from the innate immunity circadian rhythm as recently described [166].

2.2.2.1 Magnetic bead isolation of neutrophils by negative selection

The EasySep® Human Neutrophil enrichment kit was used, following the manufacturer's instructions. Whole blood was gently mixed with HetaSep solution at a ratio of 1:5 (1 part HetaSep to 5 parts Blood) and incubated at 37 °C for 20-30 min until plasma/erythrocyte interphase was at approximately 60% of the total volume. The leukocyte-rich plasma layer was carefully removed and washed in a 4-fold volume of recommended media (Mg^{2+} and Ca^{2+} -free PBS, + 2% FBS and 1 mM EDTA). Cells were centrifuged at 500 g for 5 min and resuspended in a 4-fold volume of recommended media. Cells were centrifuged at 120 g for 10 min to

remove platelet contamination and resuspended at 5 x 10^7 nucleated cells per mL.

A volume of between 0.5 and 2 mL of nucleated cells at 5 x 10^7/mL was used in each neutrophil bead purification process, dependent on the number of purified cells required. 50 µL of EasySep® neutrophil enrichment cocktail, containing a mix of tetrameric antibody complexes produced from monoclonal antibodies directed against the cell surface antigens CD2, CD3, CD9, CD19, CD36, CD56 and glycophorin A – whilst being bi-specific for dextran – was added per 1 mL of nucleated cells and incubated for 10 min at room temperature. 100 µL of EasySep® dextran-coated nanoparticle beads were added per 1 mL of nucleated cells and incubated for a further 10 min at room temperature. The cell/antibody/bead solution was adjusted to a total volume of 2.5 mL with recommended media and placed into an EasySep® magnet for 5 min at room temperature. Unbound neutrophils were poured off and placed into EasySep® magnet for a further 5 min. Highly-pure, unbound neutrophils were briefly centrifuged and resuspended in RPMI 1640 media plus 25 mM HEPES to a concentration of 5 x 10^6/mL. Following the erythrocyte sedimentation step, the neutrophil isolation procedure from whole blood was performed at room temperature and would typically take less than 100 min to complete.

2.2.2.2 Polymorphprep™ isolation of neutrophils

Neutrophils were isolated by single-step centrifugation of whole blood onto Polymorphprep™ density gradient as per the manufacturer's

recommendation. Briefly whole blood was layered onto Polymorphprep™ at a ratio of 1:1 and centrifuged at 500 *g* for 35 min. Granulocytes were carefully removed and resuspended in RPMI 1640 media plus 25 mM HEPES and centrifuged at 500 *g* for 5 min to remove any remaining Polymorphprep™. Cells were resuspended in media and contaminating erythrocytes were removed by hypotonic lysis by the addition of ammonium chloride (13.4 mM $KHCO_3$, 155 mM NH_4Cl and 96.7 mM EDTA in 500 mL distilled water) at a ratio of 1:9 media:lysis buffer for 3 min. Platelet contamination was removed by further centrifugation of the cells at 150 g for 3 min. Neutrophils were counted using a Beckman Coulter Multisizer 3 and the suspension volume adjusted to give a final concentration of 5 x 10^6/mL. Cells were incubated with gentle agitation at 37 °C and supplemented with 10% (v/v) human AB serum for incubations >4 h. The entire neutrophil isolation procedure from whole blood was performed at room temperature and would typically take less than 90 min to complete.

2.2.2.3 Lymphocyte isolation

Lymphocytes were isolated by carefully removing the upper PBMC band following whole blood centrifugation over Polymorphprep™ (as described in previous section). Cells were layered onto Lymphoprep™ density gradient, centrifuged for 20 min at 600 g and resuspended in RPMI (+ HEPES) media at 1x10^6/mL.

2.2.3 Cytospins

Purity of neutrophil isolations was determined by cytospin and visual identification of stained cells. 20 µL of cells at 5 x 10^6 cells/mL was added to 180 µL of PBS plus 10 mM EDTA. The suspension was placed into a single cytology funnel (VWR) and centrifuged onto a glass slide (VWR) at 30 *g* for 5 min using a Shannon 3 cytospin and immediately stained with Rapid Rowmanowsky stain. Neutrophil purity was typically >96% and >98% following Polymorphprep™ or magnetic bead isolation, respectively.

2.2.4 Flow cytometric analysis of neutrophil cell surface markers

For the measurement of cell surface markers by flow cytometry, 20 µL of neutrophils at 5 x 10^6 cells/mL (equating to 1 x 10^5 neutrophils) were washed in 100 µL PBS plus 0.1% (w/v) BSA, centrifuged at 1000 *g* for 3 min, resuspended in 10 µL PBS plus 0.1% (w/v) BSA and incubated at 4 °C in the dark with 2-5 µL of fluorescently-conjugated antibody (as indicated in the text). Following 30 min incubation cells were fixed in 2% paraformaldyhyde (PFA) for no more than 15 min at room temperature in the dark. Cells were centrifuged at 1000 *g* for 3 min and resuspended in 200 µL of PBS plus 0.1% (w/v) BSA and stored at 4 °C prior to analysis. A minimum of 5000 gated events were analysed using a Guava Easycyte flow cytometer (Millipore). Where applicable a suitable isotype control was used to control against non-specific staining.

2.2.5 Flow cytometry measurement of neutrophil apoptosis

For measurement of neutrophil apoptosis/necrosis by flow cytometry, 10^5 cells were incubated in 100 µL HBSS with 10 µg/mL of FITC-conjugated annexin V in the dark, following 15 min, 1 µg/mL propidium iodide (PI) was added and samples were measured immediately on a Guava Easycyte flow cytometer. A minimum of 5000 gated events were collected per sample.

2.2.6 Preparation of protein lysates

Neutrophils were centrifuged at 1000 *g* for 3 min following culture under appropriate conditions (as indicated in the text), the supernatant carefully aspirated, and the pellet immediately lysed in boiling Laemmli buffer containing 10% (v/v) glycerol, 100mM DTT, 3% (w/v) SDS, 1 M Tris-HCL (pH 6.8) and 0.001% (w/v) bromophenol blue to a final concentration of 5 x 10^4 cells/µL. For incubations >4 h neutrophils were additionally washed with PBS to remove supplemented human AB serum prior to lysis. For measurement of protein phosphorylation, Laemmli buffer was additionally supplemented with phosphatase inhibitors (phosphatase inhibitor cocktail II, Calbiochem, Nottingham, UK). Samples were boiled for 5 min with occasional vortexing and stored at -20 °C.

2.2.7 Western blotting

Whole cell extracts were briefly boiled and centrifuged. 10-25 µL of protein lysate (equivalent to 5-15 x 10^4 cells) was loaded per well of an 8-15% polyacrylamide gel (dependent on size of protein under investigation) using a 4.5% stacking gel. 10µL of biotinylated and/or pre-stained molecular weight ladders were also loaded to allow for molecular weight determinations. Samples were electrophoresed at a constant 180-200V for approximately 60 min using a BioRad Mini Protean III electrophoresis kit until proteins were suitably resolved. Proteins were transferred to a PVDF membrane by electrophoresis in a BioRad mini Protean III transfer kit at a constant 100V for 60-90 min, depending on thickness of gel. Successful transfer of proteins was confirmed by briefly staining the membrane in Ponceau S (0.01% (w/v) in 5% (v/v) acetic acid) stain.

To decrease non-specific binding of antibodies, membranes were incubated at room temperature for at least 1h in blocking buffer. Membranes were subsequently incubated overnight with primary antibody in wash buffer containing either 5% (w/v) non-fat dried milk or 5% (w/v) bovine serum albumin (BSA) at 4 °C with gentle agitation. Following washing of membranes for 3 x 5 min in washing buffer, membranes were incubated at room temperature for at least 1h with an appropriate horseradish-peroxidase (HRP)-conjugated secondary antibody in wash buffer containing 5% (w/v) non-fat dried milk (for specific concentrations see Table 2.7). Following further washing of the membrane in wash buffer

(3 x 5 min), the bound antibodies were detected using enhanced chemiluminescence (ECL) reagents and careful exposure of the membrane to hyper-film in a dark-room. Quantification of western blots by densitometry was carried out using the AQM Advanced 6 Imaging Software (Windows) and a digital film scanner.

2.2.8 Measurement of neutrophil respiratory burst

Neutrophil respiratory burst was measured by luminol-enhanced chemiluminescence. Luminol is a membrane permeable molecule and is therefore suitable for quantification of both intra and extracellular reactive oxygen species (ROS). Luminol is oxidised by the enzymatic products of myeloperoxidase and NADPH oxidase during the respiratory burst, resulting in a release of energy in the form of light which can be quantified using a plate reader.

Cells at 5 x 10^6/mL were incubated in RPMI 1640 media with (or without) a priming cytokine for 30 min at 37 °C with gentle agitation. 2 x 10^5 cells (40 µL) were added to an opaque 96-well microtitre plate and volume adjusted to 100 µL with media. 100 µL of HBSS containing 100 µM luminol, and either 1 µM fMLP or 0.2 mg/mL PMA were added to cells and immediately measured on a FLUOstar Omega plate reader at 37 °C plus 5% CO_2.

Table 2.7 Western blot antibody concentrations

Western blot antibody concentrations	
Primary antibody	**Secondary antibody**
STAT 1	Rabbit (1:20,000)
Phosphorylated STAT1	Rabbit (1:20,000)
STAT 3	Rabbit (1:20,000)
Phosphorylated-STAT 3	Rabbit (1:20,000)
ERK	Rabbit (1:20,000)
Phospho-ERK	Rabbit (1:20,000)
p38	Mouse (1:10,000)
phosphophorylated-p38	Mouse (1:10,000)
AKT	Rabbit (1:20,000)
Phosphorylated-AKT	Rabbit (1:20,000)
IκBα	Rabbit (1:20,000)
Phosphorylated-NFκB	Rabbit (1:20,000)
Beta actin	Mouse (1:10,000)
IL-1B	Rabbit (1:20,000)
GAPDH	Mouse (1:10,000)

2.2.9 Extraction and isolation of neutrophil RNA

Neutrophils at 5 x 10^6/mL were centrifuged at 1000 *g* for 3 min, the supernatant aspirated and the pellet lysed in 1mL of TRIzol® per 5x10^6 cells. Cells were further lysed by pipetting the solution several times through a 20 gauge needle and syringe, to obtain a homogenized solution, which was incubated at room temperature for 5 min. 200 µl of chloroform was

added per 1mL of Trizol® and the solution was vigorously mixed for 15 seconds, incubated at room temperature for 2-3 min, before being centrifuged at 10,000 *g* for 15 min at 4 °C. The upper aqueous layer, containing RNA, was carefully removed and added to an equal volume of molecular grade isopropanol. This was stored for at least 24h at -20 °C to ensure complete precipitation of the RNA. Samples were then centrifuged at 10,000 *g* for 30 min at 4 °C, washed in 70% (v/v) ethanol, pelleted and re-suspended in 100 µL of RNase-free water. The RNA was further purified using a Qiagen RNeasy kit according to the manufacturers protocol, which included a 15 min DNase digestion step to eliminate any contaminating genomic DNA. RNA-Seq samples were prepared into a final elution of 30 µl of RNase-free water and stored at -150 °C. All RNA samples were transported by courier on dry ice to either the Centre for Genomic Research (University of Liverpool) or BGI International, Hong Kong. In each case, transportation took less than 60 h and samples were quality-checked both after arrival, and following library construction.

2.2.10 cDNA synthesis for PCR

cDNA was synthesized from total RNA using the Superscript III First Strand cDNA synthesis kit (Qiagen), as per the manufacturers protocols. The total amount of RNA per sample within each experiment was adjusted to an equal amount prior to cDNA synthesis. RNA from each sample was added to 1 µL of random primers (250 ng), 1 µL dNTPs and the reaction volume

was adjusted to 13 µL by adding RNase-free water. Samples were heated to 65°C for 5 min in a heat-block and rapidly cooled on ice for at least 1 min to allow primer annealing. A master-mix of 4 µL first-strand buffer, 1 µL RNaseOUT (RNase inhibitor), 1 µL (0.1M) DTT and 1 µL (200 units/µL) Superscript III reverse transcriptase, were added to each sample before incubation in a Thermo PX2 thermal cycler. cDNA synthesis was initiated at 25 °C for 5 min, completed at 50 °C for 60 min and the reaction was terminated by a final 15 min at 70 °C. Samples were cooled to 4 °C before being stored at -20 °C until further use.

2.2.11 Quantitative (real-time) PCR

Transcript levels were quantified using the Quantitect SYBR Green qPCR detection kit (Qiagen), following the manufacturer's protocol. Briefly, 1 µL of cDNA was added to 0.8 µL of each forward and reverse primer (10 pM) and 10 µL of QuantiTect, in a total reaction volume of 20 µL. Each sample was prepared in duplicate or triplicate and run on a Roche LightCycler 480 qPCR machine in an opaque 96-well microtitre plate using the cycling protocol shown in (Table 2.8). Relative amounts of target transcripts were quantified by normalising their Ct values against those of a suitable housekeeping gene (as indicated in the text) using the Pfaffl method [167]. Primer sequences shown in Table 2.9.

Table 2.8 qPCR cycling parameter

Cycling stage	Step	Temperature (°C)	Time (min)	No. of cycles
1	Taq Activation	95	15	1
2	Denaturation	95	1	45
	Primer Annealing	55	0.5	
	Elongation	72	0.5	
3	Melt Curve Analysis	60	0.5	1

Table 2.9 Primer sequences and PCR product size

PCR primers		
Gene of Interest	**Primer Sequence (5' to 3')**	**Product size (bp)**
FADD-f	CACAGACCACCTGCTTCTGA	176
FADD-r	CTGGACACGGTTCCAACTTT	
FOS-f	CTCCGGTGGTCACCTGTACT	137
FOS-r	GTCAGAGGAAGGCTCATTGC	
ICAM1-f	AGCTTCGTGTCCTGTATGGCCC	128
ICAM1-r	ACACTTGAGCTCGGGCAATGGG	
IL1B-f	CACTACAGCAAGGGCTTCAGGC	98
IL1B-r	TTCTCCTGGAAGGTCTGTGGGC	
IL8-f	AAAAGCCACCGGAGCACTCCAT	143
IL8-r	AGAGCCACGGCCAGCTTGGA	
JUN-f	TGGCAGAGTCCCGGAGCGAA	121
JUN-r	CGAAGCTGAGCGCACGTCCT	
NAMPT-f	GCCAGCAGGGAATTTTGTTA	100
NAMPT-r	TGTCACCTTGCCATTCTTGA	
SOCS3-f	CTGGTCCCCTCCCGGTTGGT	112
SOCS3-r	TGTTGGCGGCCGTGAAGTCC	

TNF-f	CAGAGGGCCTGTACCTCATC	219
TNF-r	GGAAGACCCCTCCCAGATAG	
ACTB-f	CATCGAGCACGGCATCGTCA	211
ACTB-r	TAGCACAGCCTGGACAGCAAC	
B2M-f	ACTGAATTCACCCCCACTGA	114
B2M-r	CCTCCATGATGCTGCTTACA	
GAPDH-f	CTCAACGACCACTTTGTCAAGCTCA	106
GAPDH-r	GGTCTTACTCCTTGGAGGCCATGTG	
PPIA-f	GCTTTGGGTCCAGGAATGG	60
PPIA-r	GTTGTCCACAGTCAGCCATGGT	

2.2.12 Statistics

Statistical analysis of RNA-Seq data was performed by Cufflinks bioinformatic software [145], incorporating a false discovery rate (FDR) of 5% using Benjamini-Hochberg correction for multiple testing [164,168]. Unless otherwise stated, all other data was judged for significance using the Student's t-test for either paired or independent samples, as necessary. Significance was calculated using GraphPad/Prism version 6.0. (GraphPad software, San Diego, CA. USA). Error bars represent SEM unless otherwise stated and differences were considered significant if $p < 0.05$.

2.2.13 Bioinformatic software

Several bioinformatic software programs were used to analyse both raw and annotated RNA-Seq data, (Table 2.10) provides details on the release versions used for analysis of data presented. Further details of individual bioinformatic software used can be found in Chapter 3.

Table 2.10 Release versions of bioinformatic software

Bioinformatic software versions	
Bioinformatic software	**Version**
Bowtie	2.0.07
Tophat	1.4.1-2.0.4
Cufflinks	2.02
Samtools	0.1.18
IPA	n/a
IGV	2.2.7
Microsoft Office	2011 edition
R	2.15.2
EdgeR	3.0.8
DESeq	1.10.1
cummeRbund	2.6.2

Chapter 3: Defining a bioinformatic pipeline for analysis of neutrophil gene expression

Results presented within this chapter were included in a publication in which I was co-lead author:

Wright HL†, Thomas HB†, Moots RJ, Edwards SW (2013) RNA-Seq Reveals Activation of Both Common and Cytokine-Specific Pathways following Neutrophil Priming. PLoS ONE 8(3): e58598. doi:10.1371/journal.pone.0058598

3.1 Introduction

Neutrophils are the most abundant white blood cell in the circulation, representing between 40-60% of the leukocyte population. On average, up to 10^{10} neutrophils are released from the bone marrow per day [169]. These terminally-differentiated cells patrol the vascular system in search of inflammatory signals arising from pathological insult or localised cellular injury. Armed with a variety of anti-microbial enzymes and the ability to rapidly release reactive oxygen metabolites, neutrophils are the major cellular component of the innate immune system.

Historically, neutrophils were regarded as one-dimensional innate cells, with little influence on surrounding immune cells owing to an incapacity for *de novo* gene expression and a short half-life of between 6-8 h. It is now well established that neutrophils play a central role in initiating and

propagating both the innate and adaptive immune responses by the production and release of numerous cytokines and chemokines [5].

Despite the increased appreciation of neutrophil gene expression, few studies have focused on how neutrophil gene expression profiles change in response to external stimuli. Moreover, at the time of commencement of this project, the human neutrophil transcriptome had yet to be quantified by RNA-Seq despite similar studies on other cell types [117,170–172] There are several benefits offered by RNA-Seq over more established technologies such as micro-arrays and quantitative PCR as a method of quantifying the gene expression profile of a cell (as discussed in section 1.5.2).

The past 20 years has seen an exponential increase in sequencing technologies. Both sequencing speed and accuracy have increased dramatically, whilst the financial cost of the instruments and reagents has decreased. These improvements can largely be attributed to the efforts of the human genome project, which brought about the collaboration of several research institutes in an attempt to comprehensively sequence the human genome using the best available technology at the time. The success of the 13 year project has led the way for further improvements and greater decreases in costs of sequencing. By way of example, Fig 3.1 shows how the cost of sequencing a genome has decreased by over 4 orders of magnitude in the space of a decade. These costs declined modestly over the first 5 years, but more rapidly post-2008, via the introduction of the then-called "next generation sequencing" (NGS)

technology, which consisted of Applied Biosystem's SOLiD™ platform, Roche's 454 pyrosequencer technology and Illumina's Genome analyser system [173].

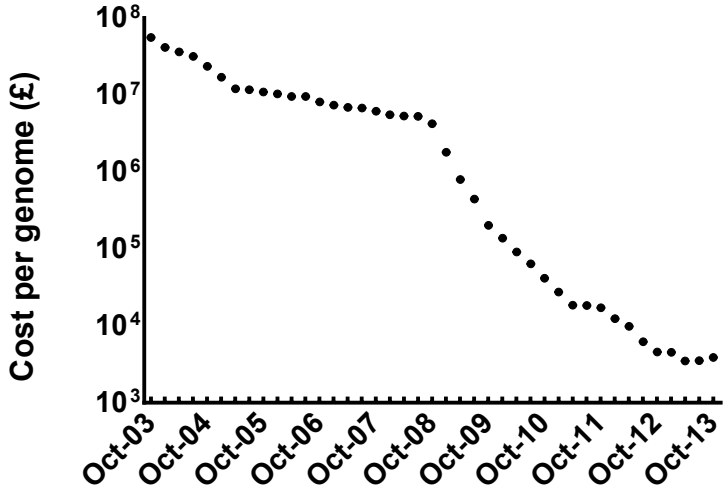


Fig 3.1 The cost of sequencing a genome from 2003-2013. The cost of sequencing has decreased by more than 4 orders of magnitude in the space of 10 years. The current cost (Jan 2014) is approximately £2300. Adapted from, Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcosts. Accessed [23-5-14].

In addition to a decrease in sequencing costs, the digital output of a sequencing run has equally accelerated exponentially during the last decade. The rate of development (or fall in cost) of many technological platforms can often be shown to correlate well with Moore's law, which states that on average, the number of transistors in an integrated circuit doubles every 24 months. One such area of technology that has fallen in line with Moore's law is that of data storage. However, the increase seen in NGS capacity for data production far outstrips that predicted by Moore's law [174]. This disparity in cost between storage and production of NGS data

could ultimately lead to sequence data being more expensive to store than to produce [175].

Whilst the relative improvements in sequencing technologies compared to other technological fields is of interest (and concern), perhaps of greater importance is the disproportionate improvements in sequencing technologies with sequencing data analysis software. As the cost of sequencing decreased to accessible levels, the data analysis software available to analyse the raw data remained largely either inadequate, too technical to operate, or unfeasibly expensive for small-scale studies by scientists.

More recently, a large number of open-access software programs have been developed for biology- users who have basic informatics skills. These allow free distribution and adaptation of programs that are designed to be compatible with the major sequencing platforms and raw data formats, but require the user to be familiar with a basic command line interface and non-standard operating systems, such as Linux or Unix. Additionally, software development of this kind is often small-scale and with limited features. Consequently, there are now a large number of open-source bioinformatic analysis software packages offering a range of utilities, advantages and features over other software packages. While these may not have the breadth of features available or ease of use to appeal to a dedicated bioinformatics research lab, they are usable to researchers with modest bioinformatic skills.

Perhaps unsurprisingly, many of the more successful and popular software packages are those developed by large commercial companies such as the CLCbio Genomics Workbench by Qiagen, or ERGO™ by igenbio, which are software suites offering all the most popular features of other software packages (such as read-mapping, annotation and differential analysis) in one software suite, with the added benefit of offering full user support and guidance on how best to utilise the software and extensive troubleshooting documentation. Software suites of this kind are an attempt to standardise the process of quantifying large datasets but they do so at fairly high cost. Whilst the intuitive graphical interface, full support, and industry-wide recognition is of great importance for large research centres, the often high commercial-licence cost (and annual subscription costs) can price out the smaller laboratories. Consequently, there is little agreement within the NGS community as to what is considered to be 'best practice' when it comes to data annotation and analysis [176].

In summary, It is clear that a global approach to studying neutrophil gene expression by RNA-Seq is both potentially achievable, and of great interest. It will allow the accurate measurement of transcriptional changes in neutrophils under tightly controlled conditions and could ultimately lead to the development of a predictive model to determine the functional consequences of these changes. Whilst the available high-throughput sequencing technology is both adequate and affordable to undergo a large scale study into gene expression, the necessary software and parameters

required are poorly defined. The data presented here aim to define a robust pipeline of software packages, protocols and methods that can be employed to quantify the neutrophil transcriptome under conditions of stimulated inflammation.

3.2 Aims

The overall aims of this Chapter were to establish a pipeline for the analysis of transcriptome data obtained by RNA-Seq of neutrophils stimulated in vitro by inflammatory activators. The specific aims were:

1. To define a pipeline of methods, software programs and settings for compilation and analysis of RNA-Seq data.

2. To determine if changes in gene expression are consistently detectable and if they correlate with a change in phenotype, as measured by laboratory based functional assays.

3. To define a set of downstream analysis techniques to further analyse RNA-Seq gene expression data.

3.3 Methods

The data herein describe the optimisation of a bioinformatics pipeline (of protocols and software packages) to define the neutrophil transcriptome under stimulated conditions of inflammation (*in vitro*). The data is presented in such a way as to be informative and clear to an interested reader of the best methods to adopt in addition to various pit-falls to avoid. The pipeline described here will form the basis of bioinformatic analysis undertaken in later Chapters. For the sake of brevity, many of the comparative and optimisation analyses carried out could not be included, hence much of the results described represent the optimal settings for each software program under comparison.

3.3.1 Sample preparation

Neutrophils were isolated by Polymorphprep™ from healthy donors (as described in section 2.2.2.2). Neutrophil RNA was extracted by TRIzol®-chloroform precipitation and analysed by RNA-Seq using paired-end sequencing on the SOLiD™ 4.0 platform or single-end sequencing using the Illumina HiSeq 2000 platform, as indicated in the text.

3.3.2 Computational processing

All of the bioinformatic software discussed in this Chapter can be run on a stand-alone desktop computer using a Unix/Linux or Mac OS operating system providing there is adequate processing power and random access memory (RAM) available. Technical specifications of the processing unit used (analysis-Mac) in the bioinformatics analysis can be found in Appendix Table A.2.

For bioinformatic analyses using multiple data sets, to allow multiple analyses to be run simultaneously and in a shorter time than using analysis-Mac, it was necessary to utilise a high-powered multi-core processing cluster provided by the University of Liverpool Computer Services Department (CSD). Consequently all mapping, annotation and differential-expression testing processing was carried out using a Gigabit Ethernet cluster (44-core, per core; 2.2 GHz AMD core, 8GB RAM, 72GB disk space, Linux 9.3 SuSE). This cluster was accessed via the analysis-Mac through the University intranet. Raw data was firstly uploaded to cluster servers and bioinformatic analysis software was run via a short command script (*.txt* file). Software commands listed in this Chapter only include the basic commands (and optimised parameters) needed to run the software, and file paths are summarised. An example of the additional scripting required for cluster-based analyses is provided in Appendix Fig A.1.

Raw data from cluster-run analyses was backed up both internally and manually using University of Liverpool CSD backup storage facility.

3.4 Results

3.4.1 RNA quantity and quality

3.4.1.1 Quantity

RNA-Seq technology requires high integrity RNA, with minimal contamination by genomic DNA. Typically, sequencing service providers require a minimum of 5 µg total RNA (per sample) for library construction (which also provides enough material for a second attempt at library construction). The numbers of neutrophils isolated from peripheral blood is donor-dependent and can vary between $1x10^6$-$5x10^6$/mL blood [177]. Furthermore, neutrophils contain considerably less RNA than other leukocytes such as peripheral blood mononuclear cells (PBMCs) [178].

Following extraction by TRIzol®-chloroform precipitation, total RNA from 10^6 and 10^7 neutrophils (isolated by Polymorphprep™) and 10^6 PBMCs (isolated by Lymphoprep™) was measured from 4 separate donors using a nanodrop spectrophotometer (Thermo Fisher scientific, Gloucestershire, UK) (Fig 3.2). Mean level of RNA from 10^6 neutrophils = 139.5 ng/mL ± 15.75, 10^7 neutrophils =1797.8 ng/mL ± 91.75 and 10^6 PBMCs = 807.0 ng/mL ± 117.43. These data suggest that PBMCs have approximately 5-times more RNA than neutrophils on a cell basis, and that a minimum of $30x10^6$ neutrophils is required to achieve the recommended 5 µg of total RNA necessary for sequencing.

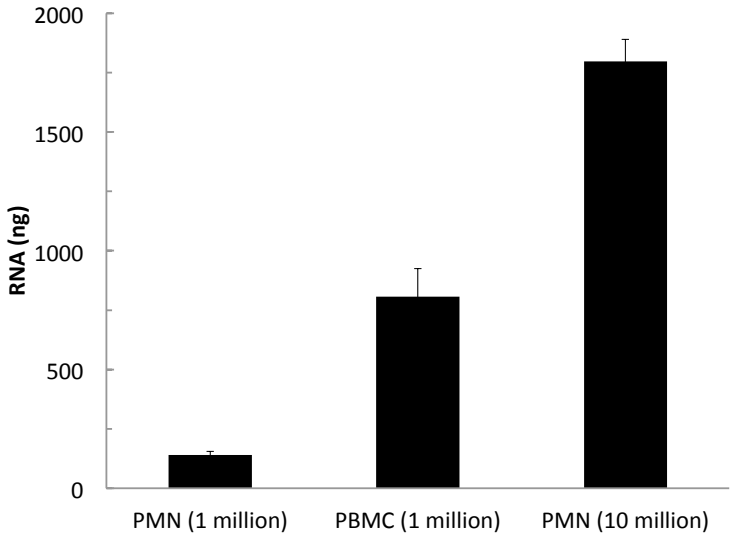


Fig 3.2 Amounts of RNA extracted from 10^6 and 10^7 neutrophils (PMN) and 10^6 peripheral blood mononuclear cells (PBMC). Bars represent mean of 4 separate experiments. Error bars represent SEM.

3.4.1.2 Quality

i) RNA integrity

The lengthy isolation process required for purifying neutrophils from whole blood (60-90 min) can often affect the quality of the final purified RNA. While the contamination of neutrophil preparations by other leukocytes must be considered, the greater number of steps required to isolate highly-pure neutrophils can lead to longer isolation times and opportunities to inadvertently activate neutrophils or compromise the quality of RNA that is recovered.

The overall integrity of an RNA sample is such a crucial determinant of the eventual success of an RNA-Seq experiment that sequencing service-providers often request that all samples be accompanied by an accurate

measurement of integrity. Furthermore, an additional measurement of integrity is often made just prior to the samples being sequenced. A common method of measuring RNA integrity (RIN) is using the Agilent BioAnalyser [179,180]. RNA integrity is measured on a scale of 1-10 where 1 refers to RNA with the lowest integrity and 10 indicating no degradation of RNA. By way of example, output from the Agilent Bioanalyser relating to a low, and high-integrity RNA sample is shown in appendix Fig A.2.

RNA from 10^6 PBMCs and 10^7 neutrophils was extracted by TRIzol®-chloroform precipitation and RNA integrity was measured by Agilent 2100 BioAnalyser (Fig 3.3). Levels of integrity were consistently measured between 7.0 – 8.6 RIN. The mean values across 4 samples were identical in PBMCs and neutrophils (7.9 RIN).

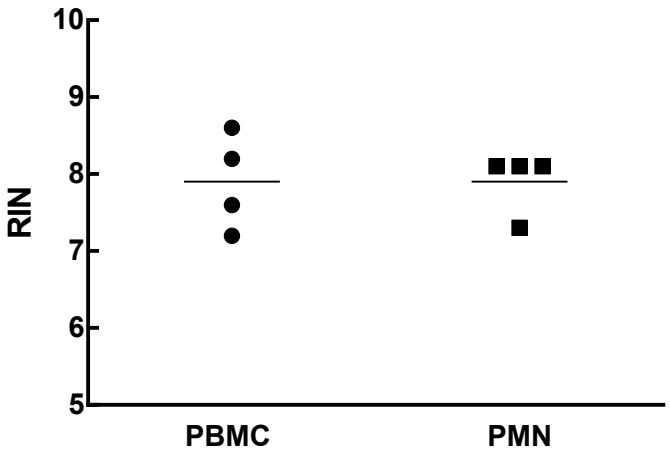


Fig 3.3 RNA integrity number (RIN) values for RNA samples from PBMCs and neutrophils (PMN). Horizontal bars represent mean value from 4 separate experiments/Donors.

ii) RNA purity

RNA purity is of upmost importance for RNA-Seq. Residual solvent from extraction protocols or contaminating genomic DNA can have a significant effect on the success and accuracy of the resulting RNA-Seq experiment. Neutrophil RNA extraction performed in the previous experiments was using TRIzol®-chloroform precipitation, followed by on-column cleanup by Qiagen RNeasy kit including a 15 min the DNase digest step. The efficiency of TRIzol®-chloroform extraction and DNase digestion step was compared to an on-column extraction method using a Qiagen column, with or without a DNase digestion step. RNA from TRIzol®-chloroform extraction and Qiagen on-column extraction was converted to cDNA using the Superscript III first strand cDNA synthesis kit. Primers for MCL-1 (full length) were used to amplify cDNA by PCR. Amplified cDNA from each extraction method was analysed on an agarose gel to assess levels of contamination by genomic DNA. Samples prepared by on-column extraction were found to contain higher levels of contaminating genomic DNA than samples extracted by TRIzol-cholorform precipitation. Additionally, where contaminating genomic DNA was present, a 15 min DNAse digest step was sufficient to eliminate any genomic DNA signal (see Appendix Fig A.3).

These data suggest that neutrophil isolation by Polymorphprep™ and RNA extraction by TRIzol-chloroform precipitation, plus RNA cleanup and DNase digestion steps, provides an appropriate method for RNA sample

preparation with high integrity (> 8.0 RIN) and low contamination by genomic DNA, which are suitable for RNA-Seq experiments.

3.4.2 Sample preparation

Several stimuli are known to induce *de novo* gene expression in neutrophils, such as low oxygen tension (hypoxia), immune complexes and cytokine stimulation. Of these stimuli, the effects of cytokines on neutrophil gene expression are perhaps the most widely studied and established (often by micro-arrays and qPCR), and so these agonists were used to validate transcriptome analysis by RNA-Seq.

3.4.3 Cytokine stimulation and time point selection

Both GM-CSF and TNFα are commonly-used neutrophil priming agents that regulate neutrophil gene expression *in vitro* [181,182]. Neutrophils incubated with previously established priming concentrations of GM-CSF (5 ng/mL) and TNFα (10 ng/mL) [182–184] were prepared as previously described (Methods section 2.2.9) for RNA-Seq analysis.

Priming of neutrophils with either GM-CSF or TNFα has previously been shown to rapidly increase transcription of several genes, such as IL-8 and IL-1β [10,185,186]. First, it was necessary to determine the time course of activation of gene expression following addition of these cytokines. Fig 3.4 shows the relative expression of IL-8, IL-1β, CCL3 and ICAM1 over 2 h following neutrophil incubation with (or without) GM-CSF measured by

qPCR. Time points beyond 2 h were not considered due to the likelihood of *de novo* cytokine production leading to autocrine signalling [186–189].

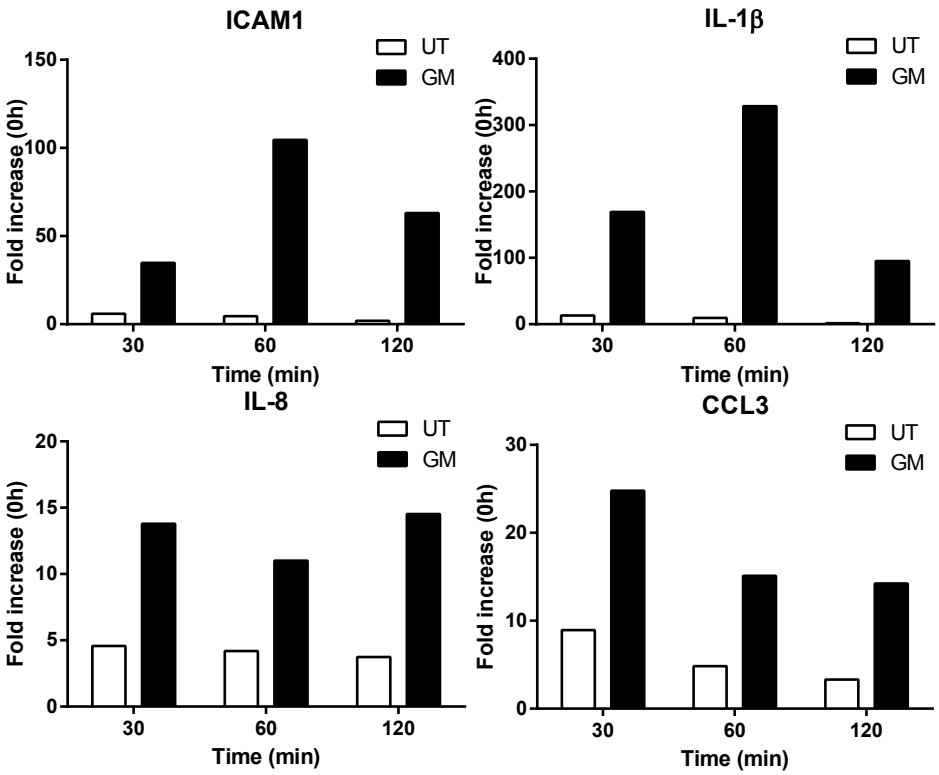


Fig 3.4 Relative expression of mRNA following stimulation with GM-CSF (GM) (5 ng/mL) (black bars), or remaining untreated (UT)(white bars) over 2 h measured by qPCR. Target gene expression was normalised to B2M housekeeping gene[167] and is expressed as fold increase compared to levels of expression at 0h.

Levels of expression were highest at 1 h for ICAM1 and IL-1β, levels of CCL3 were highest at 30 min whilst levels of IL-8 remained high for entire 2 h time course. Whilst it is impossible to have a single time point for RNA-Seq analysis which is optimal for all genes in the transcriptome, these data suggest that a 1 h time point is adequate to identify rapid changes in gene expression following neutrophil stimulation whilst also avoiding any risk of

measuring secondary activation of neutrophils by autocrine signalling following *de novo* synthesis of cytokines.

3.4.4 RNA-Seq pipeline development

3.4.4.1 Platform selection

The most popular platforms for NGS share many similarities, such as sensitivity, capacity and accuracy, but differ in a number of parameters such as read length, output data format, speed and cost (described in more detail in sections 1.5.4-1.5.6) The choice of a suitable platform must consider each of these factors, in addition to availability and access.

At the outset of this project, the Centre for Genomic Research (CGR) located in-house at the University of Liverpool provided next generation sequencing of RNA by both Roche's 454, and Applied Biosystems SOLiD™ 4.0 platforms. In addition, sequencing by the SOLiD™ platform could be carried out using single- (50 bp) or paired-end (50 + 35 bp) reads. The short reads and deep coverage of the SOLiD platform is better suited to gene expression studies of a well annotated genome (e.g. the human genome), than the 454 platform which utilises much larger read lengths and is more appropriate for *de novo* assembly of non-sequenced genomes (or transcriptomes) where a reference sequence is not available. Furthermore, the ability to carry out paired-end sequencing on the SOLiD platform can increase the amount of raw data produced from a single sequencing run by approximately 70% whilst also improving the accuracy of read mapping. Consequently, neutrophil RNA samples were initially

sequenced by paired-end technology on the SOLiD 4.0 platform at the CGR.

3.4.4.2 SOLiD™ 4.0 paired-end sequencing

Whole blood from a healthy donor was prepared by Polymorphprep™. Neutrophils ($3x10^7$/sample) were incubated at 37 °C for 1 h with (or without) GM-CSF (5 ng/mL) or TNFα (10 ng/mL). RNA was extracted as previously described. RNA integrity was analysed by Agilent bioanalyser 2100 and measured at 8.5, 7.5 and 7.0 RIN for the untreated, GM-CSF and TNF samples respectively.

Total RNA was enriched for mRNA transcripts by terminator exonuclease treatment at the CGR (as described in section 1.5.3.2.a). Enriched samples were processed for paired-end sequencing which produced upwards of $6x10^7$ 50 bp 5′→3′ forward (F3) and paired 35 bp 3′→5′ reverse (F5) transcripts per sample (see section 1.5.8 for details of paired end sequencing, and Appendix Table A.4 for number of raw reads per sample).

3.4.4.3 Quality control analysis of paired end sequence data

For each sample sequenced on the SOLiD™ 4.0 platform, 2 data files are produced; a raw data file (filtered to remove adapter sequences and fragment reads) in *.csfasta* file format (i.e. a *fasta* sequence file with the data in colorspace, rather than base space), and an accompanying quality file *(.qual*) which provides a Phred quality score for each base in each read. The Phred score refers to the likelihood that the base calling at each base

position is correct. For example, a Phred score of 40 indicates that there is a 99.99% chance that at that position the sequencing software has correctly called the base nucleotide whereas a score of 10 indicates that there is a 90% chance that the base nucleotide has been called correctly [190]. Quality scores can be used to assess the success of a sequencing run and can be included in the mapping software protocols to improve read mapping rates.

Raw reads were analysed by the Java program Quality Assessment [191] to quantify the number of reads in each data set that had a mean Phred score of >20 (i.e < 1 incorrectly called base per 100 bases). In each sample, less than 50% of reads had a mean Phred quality score of >20, with the smaller F5 reverse-fragments consistently having a lower score than the F3 forward-fragments (Fig 3.5). These values are lower than expected and could impact on successful mapping due to the low quality of the data.

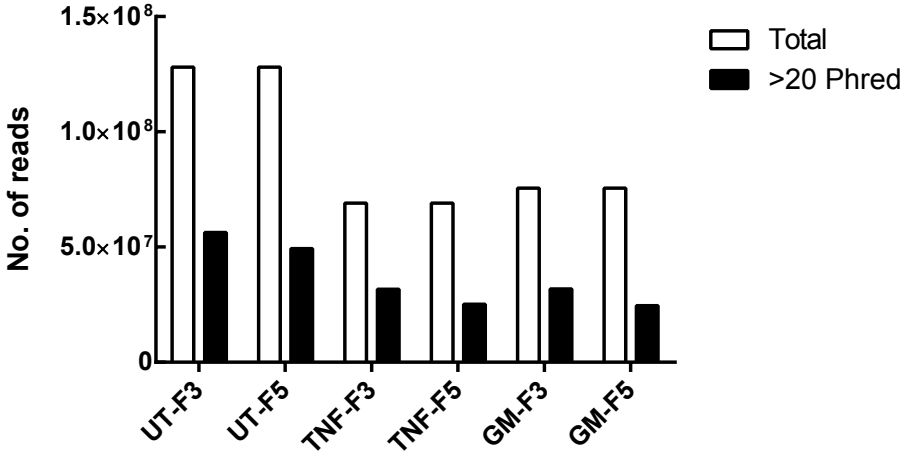


Fig 3.5 Total number of reads (open bars) and number of reads with >20 Phred quality score (black bars) in untreated (UT), GM-CSF (GM) and TNFα (TNF) treated samples, for forward (F3) and reverse (F5) fragments of paired-end raw SOLiD sequencing reads. Reads quantified by Quality Assessment software [191].

3.4.4.4 High throughput mapping of paired-end sequence data by Bowtie/Tophat

The high throughput mapper Bowtie is an ultra-fast, short sequence-read aligner that can operate on modest computational hardware utilising a heavily indexed (Burrows-Wheeler compression† [192]) reference sequence and multi-core processing (where available) to map upwards of $25x10^6$ reads per hour [151]. Bowtie can take single- or paired-end reads in either base space or colorspace as input, such that it can process raw data from any of the 3 major sequencing platforms (454, SOLiD or Illumina). However, Bowtie cannot map reads that span splice junctions.

The closely-related software package Tophat [155] utilises the basic mapping features of Bowtie, while applying an additional mapping algorithm to unmapped reads to enable the accurate mapping of reads that span splice

†Burrows-Wheeler compression is a method of indexing large amounts of data. It is particularly useful for data with large portions of repetition (such as genomic data), and is also a reversible form of compression without the need for additional data.

junction sites that would otherwise be unmapped by Bowtie (covered in more detail in section 1.6.2). Initiation of Tophat first uses Bowtie to map reads not spanning splice junctions, subsequently, Tophat attempts to map remaining reads. Provision of a *.gtf* file (see Abreviations Table Ab.1 for details) to Tophat can improve mapping rates by providing a list of known splice junctions within the reference sequence.

Several different parameters and settings options exist for both software programs which affect how read-data are processed and handled, prior to, or during the mapping process. For example, reads may be trimmed to eliminate areas of low quality prior to mapping. Alternatively, if a read can be aligned to multiple locations in the reference sequence, settings within the Bowtie program determine if the read is eventually positioned at the highest-quality location, spread amongst all sites equally, or removed from the mapping process entirely since no definitive mapping location can be determined. These features exist to enable the user to improve rates of mapping, while also improving the quality of mapping by a decrease in the number of false positives, false negatives or sequencing/mapping artefacts. A summary of the most important parameters are found in Table 3.1, which also highlights the parameters that were altered from default settings during the mapping process to achieve optimal mapping rates.

Table 3.1 Summary of Bowtie options and parameters available. <int> refers to additional parameter required as an integer. Settings shaded grey were assigned in neutrophil sequencing pipeline.

Bowtie settings	Function
Input options	
-q	Input file is in FASTQ format
-f	Input file is in FASTA format
-r	Input file is in RAW format
-c	Input via command line
-C/--color	Input is interpreted in colorspace
-Q/--qual	Input file is a quality file
-s/--skip <int>	Skip <int> no. of reads from input
--solexa1.3-quals	Input quals are in ASCII format, appropriate for Illumina pipeline version > 1.3
Alignment	
-v <int>	Report alignments with no greater than <int> mismatches
-e <int>	Maximum permitted total of quality values of all mismatched read positions throughout entire read alignment
-l <int>	Seed length, i.e the number of reads from the high quality end used to begin alignment
-n <int>	Number of mismatches permitted in the "seed" (0-3, default: 2)
-I/--minins <int>	Minimum length of reads+insert for paired end reads
-X/--maxins <int>	Maximum length of reads+insert for paired end reads
--fr/--rf/--ff	The upstream/downstream orientation of paired-end reads relative to the forward reference strand
--chunkmbs <int>	The amount of memory (in Mbs) assigned to read alignment per thread (default: 64)
Reporting	
-k	Report up to <int> no. of alignments per read (or read pair)
-a/--all	Report all valid alignments per read (or read pair) (default: off)
-m <int>	Suppress all subsequent alignments if a read has more than <int> possible alignment locations
-M <int>	As with –m , supress alignments with greater than <int> valid locations, but assign read to one location at random

--best	Where multiple alignments for a single read occur, use number of mismatches and quality of reads to assign to best alignment location, or where multiple alignments are permitted, list alignments in order of best to worst
--strata	As with –best, where multiple alignment locations occur that fall into multiple stratum, only use alignments from the best strata. (Redundant in later versions of Bowtie)
Output	
-t/--time	Print the amount of time taken to complete each phase of mapping
--quiet	No update text whilst software is running, only output alignments
--al <filename>	Output all reads that successfully aligned into a new <filename> as they appear in the raw input file
--un <filename>	Output all reads that failed to align into a new <filename> as they appear in the raw input file
--max <filename>	Output all reads that failed to align due to exceeding limit of alignment locations into a new <filename> as they appear in the raw input file
--suppress <int>	Suppress column no. <int> from the output file
-S/--sam	Output in SAM format
Performance	
-p/--threads <int>	Run alignment in parallel on <int> no. of processors/cores

3.4.4.5 Optimisation of Bowtie/Tophat settings

Settings such as seed length, number of alignment threshold and read trimming were optimised to increase rates of mapping without increasing the amount of non-specific alignments (see Table 3.1 for explanation of each parameter). Due to the low quality values for F5 paired fragments in each of the SOLiD sequenced samples, F5 reads were omitted from the mapping stage. Furthermore, low quality 3' ends of the F3 fragments were

trimmed (by 8 bases) by Bowtie prior to alignment, as this improved overall rates of mapping.

Due to a software design feature, several of the Bowtie parameters were not available when initiating mapping by Bowtie through running Tophat. Consequently, mapping was completed via a three-stage process. Firstly, reads were mapped to the genome using Bowtie and unaligned reads were outputted into a separate file, using the command:

bowtie -p 8 -S -C -l 20 --trim3 8 -e 100 --un unaligned --chunkmbs 4000 -k 1 -m 1 --best --strata /path/to/Bowtie/reference/genome –f path/to/raw/reads_data –q path/to/raw/quality/values > output.sam

Secondly, unaligned reads were inputted into Tophat and realigned to the reference sequence (human reference hg19), using the command:

tophat -p 8 --color --quals -r 200 --mate-std-dev 30 -a 5 --library-type fr-secondstrand --segment-length 20 –bowtie-n --max-multihits 1 -G /path/to/reference/annotation_file.gtf /path/to/Bowtie/reference/genome /path/to/unaligned_reads –q path/to/unaligned_reads/quality/values

Finally, all aligned reads were merged into a single output file (*.Bam*) and sorted by the java program Picard [193], using the commands:

java -Xmx2g -jar/path/to/picard-merge.jar INPUT=bowtie.bam INPUT=tophat.bam OUTPUT=merged.bam SORT_ORDER=coordinate"

"java -Xmx2g -jar /path/to/Picard-sort.jar INPUT=/path/to/merged.bam OUTPUT=merged-sort.bam REFERENCE=/path/to/reference/hg19.fa

Fig 3.6 summarises the mapping strategy used to map reads to human hg19 genome, using Bowtie and Tophat and output files merged with Picard.

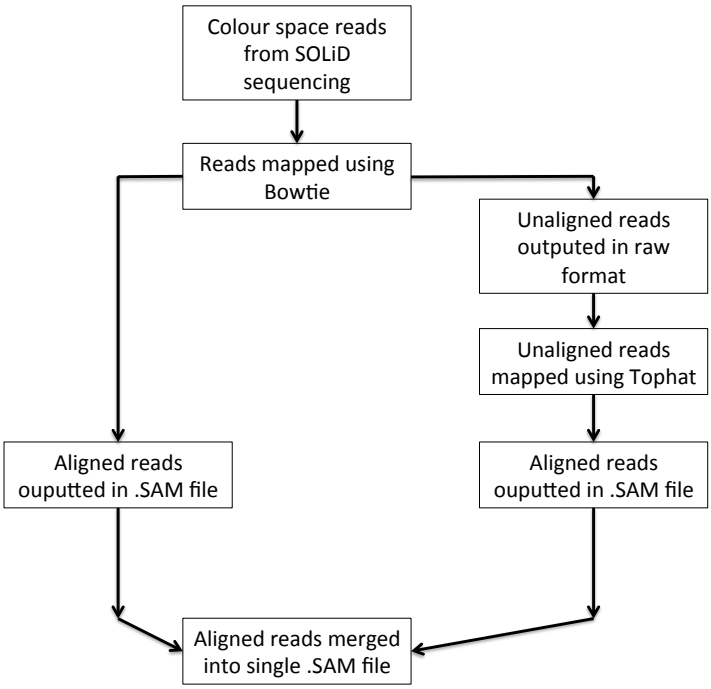


Fig 3.6 Flow chart of mapping process using the high throughput aligners Bowtie and Tophat. Output files were merged using Picard software [193].

3.4.4.6 Mapping results of SOLiD data

Table 3.2 list the maximal mapping percentages achieved using the optimised mapping pipeline shown in Fig 3.6. Mapping rates could not be increased above 35% in any samples. This was likely due to the low quality scores for reads across all samples, as seen in Fig 3.5.

Table 3.2 Percentage reads mapped for each dataset using optimal mapping strategy summarised in Fig 3.6

Dataset	Reads mapped (%)
Untreated	34.9
GM-CSF	32.7
TNFα	31.0

To evaluate the quality of reads that were successfully mapped by Bowtie/Tophat the final *.Sam* file was analysed using the software program FastQC [194]. FastQC is a Java based program that takes *.Bam/.Sam* mapping files as input and performs a series of quality control tests on the raw data. Among the output is a mean quality score (Phred value) for each base position among all reads that were successfully mapped. Fig 3.7 shows the FastQC output for the untreated neutrophil dataset. Mapped reads were found to have highest quality values at the 5′ end (base positions 1-10) but quality values consistently decreased towards the 3′ end, such that the mean values for position 49 was approximately half of the mean value for base position 1 (base position 1 – mean Phred score-60; base position 49 – mean Phred score – 31). The variability in quality scores (i.e. the inter quartile range 25-75 percentile) also increased towards the 3′ end of mapped reads. Similar results were obtained for GM-CSF and TNFα treated samples.

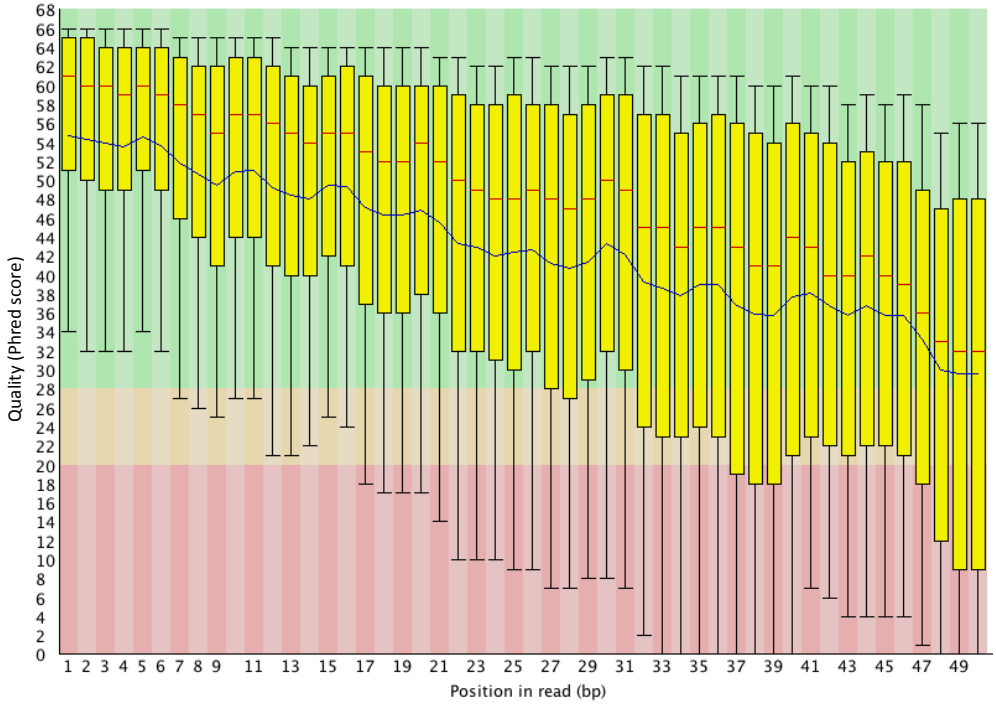


Fig 3.7 FastQC[194] analysis of Bowtie/Tophat mapped reads from neutrophil untreated dataset. Graph shows the mean quality values (Phred score) (y-axis), for each base position within the mapped reads (x-axis). Scores are represented by: inner quartile range (25-75 percentile) (Yellow boxes); median Phred score (Red line); mean value (Blue line); and 10-90 percentile (whisker plots).

These data reveal that only the highest quality reads were able to be mapped (30-35%). Furthermore the quality of mapped-reads was poor at the 3′ end and reveals why removing the 8-bases from the 3′ end (by using the trimming command in bowtie) improved mapping rates. It is likely that the poor percentage of mapped reads and the average quality of reads which successfully mapped will impact on the quantification of data in downstream analyses.

3.4.5 Illumina HiSeq2000 – Single end sequencing

Due to low quality scores for raw data and low mapping rates obtained from the SOLiD paired end sequencing run, further neutrophil samples (untreated, GM-CSF and TNFα, N=1) were subsequently sequenced using

single-end sequencing (50bp) on the Illumina HiSeq2000 platform, by BGI International (Hong Kong). RNA was extracted and prepared using identical methods and couriered on dry ice to BGI. Sample processing was subject to identical QC analysis and any variation in sequencing protocols were due to platform-specific requirements. Quality control analysis (as provided by BGI International) of sequenced reads revealed that > 98% of reads exhibited a Phred score of > 20 (data not shown), representing a marked improvement in read quality over SOLiD datasets.

Downstream analysis of SOLiD platform data (annotation and quantification) was performed using the methods described below and results are later used in validation and comparison analysis vs Illumina platform and qPCR results included in this chapter, (as indicated in the text).

3.4.5.1 Illumina mapping strategy

Due to higher quality values in raw data from the Illumina platform, it was not necessary to trim the 3' end of the reads prior to mapping. In addition, improvements to the Tophat software (release version 1.4), meant that the Illumina reads could be mapped in a single-step process by Tophat using the command:

```
tophat -p 8 --solexa1.3-quals --max-multihits 1 -o ./_output_with_gtf \
--transcriptome-index=/path/to/transcriptome/index \
/path/to/reference/genome path/to/raw.data
```

Improved read quality and mapping strategy resulted in an increase in the percentage reads that mapped to reference sequence. Mapping percentages for Illumina sequenced datasets are listed in Table 3.3. Mapping rates were determined by TopHat version 1.4.

Table 3.3 Percentage reads mapped for Illumina sequenced datasets, following mapping with Tophat 1.4. Percentage mapped reads calculated by Tophat during processing.

Dataset	Reads mapped (%)
Untreated	94.5
GM-CSF	94.7
TNFα	94.9

3.4.5.2 Illumina read quality of mapped reads

To assess the quality of the 94% mapped reads from Illumina sequenced datasets, output files from Tophat (*.Bam* files) were analysed by FastQC (Fig 3.8). Reads had a markedly higher Phred score than those assessed form the SOLiD platform. Values were > 30 throughout the entire length of reads with an increase at the 3′ end. The inner quartile range was also much lower than with previous SOLiD samples. These data verify the high percentage mapping achieved with Illumina sequenced data and confirm that Illumina sequencing-platform (and sequencing by BGI International) is sufficient to produce high quality datasets of neutrophil RNA when mapped using Tophat.

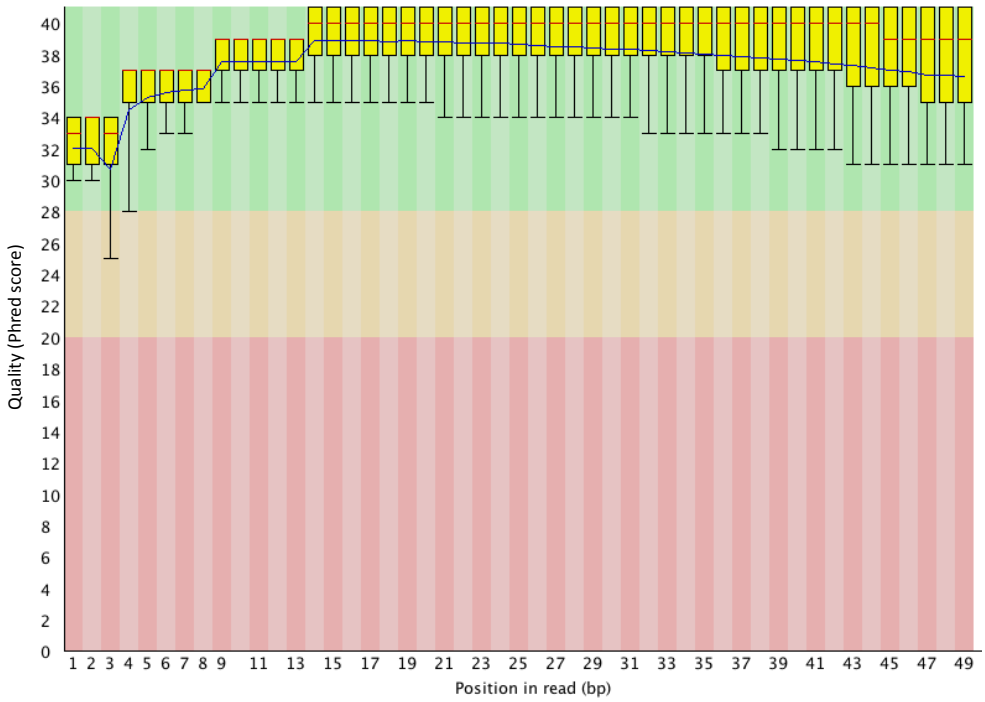


Fig 3.8 FastQC [194] analysis of Bowtie/Tophat mapped reads from neutrophil untreated datasets sequenced on the Illumina platform. Graph shows the mean quality values (Phred score) (y-axis), for each base position within the mapped reads (x-axis). Scores are represented by: inner quartile range (25-75 percentile) (Yellow boxes); median Phred score (Red line); mean value (Blue line); and 10-90 percentile (whisker plots).

3.4.5.3 Mapping-annotation and quantification

Several programs are available for annotating and quantifying NGS data against a reference sequence. Broadly speaking software packages differ in the way they quantify gene expression and model variation within the population, relying on either an absolute count of reads mapping to a specific loci (in the case of edgeR and DESeq) or using a normalised metric for gene expression RPKM (in the case of Cufflinks) (covered in more detail in section 1.6.4). To evaluate the relative effectiveness of annotation software, 3 widely-used, open source software packages (Cuffdiff- a subroutine of the Cufflinks package; DESeq; and edgeR) were compared

for their ability to quantify genes that are significantly differentially expressed (DE) between neutrophil samples incubated with either GM-CSF, TNFα, or untreated (Fig 3.9). DESeq and edgeR were run through the R software environment using a bespoke runscript adapted from the standard operating vignette [159,160]. Cuffdiff was run with default settings providing both a reference genome (*.fa* file) and reference transcriptome (*.gft* file) to improve annotation accuracy. A significance value of $q < 0.05$ (applying a 5% FDR) was applied in all 3 software packages.

Of the three software packages used, Cuffdiff was found to be the most conservative in calling significantly differentially-expressed (DE) genes (UT:GM-CSF=110, UT:TNFα=82, GM-CSF:TNFα=151), whereas DESeq identified a similar number of genes as significantly DE (UT:GM-CSF=167, UT:TNFα=74, GM-CSF:TNFα=201). However, edgeR was the least conservative when calculating significantly DE genes, consistently quantifying approximately 3 times as many genes as the other two software programs (UT:GM-CSF=407, UT:TNFα=234, GM-CSF:TNFα=538) (Fig 3.9). The increased number of significant genes following DESeq and edgeR analysis is likely due to the poisson distribution utilised by these software packages. Poisson distribution of gene expression data is known to increase the number of false positives discovered by virtue of the fact that biological variation is not sufficiently estimated within the population [149,163]. Determination of gene expression by Cufflinks/Cuffdiff provides an added benefit of several downstream software packages which are compatible

with the output format of Cufflinks. For example, cummeRbund [168] – an R based program specifically designed to perform downstream analyses and produce graphical output using Cufflinks/Cuffdiff data. In summary, Cufflinks was determined as the most suitable annotation and quantification software of neutrophil RNA-Seq data and was utilised in all subsequent analyses of datasets.

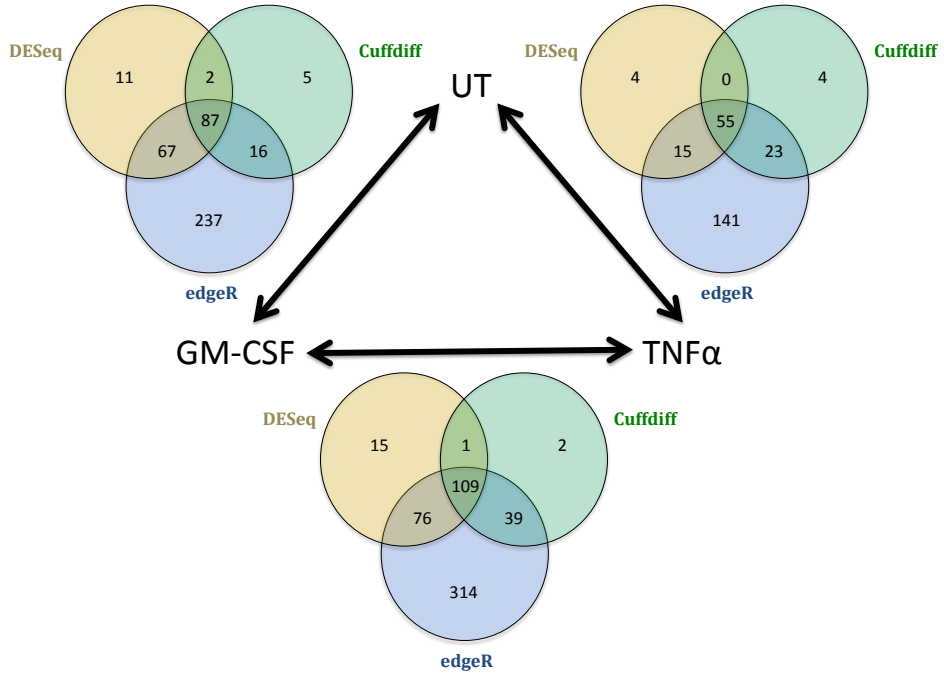


Fig 3.9 Venn diagrams showing the number of differentially-expressed genes between neutrophil samples incubated with GM-CSF (5 ng/mL), TNFα (10 ng/mL) or untreated (UT) for 1 h. Determination of significance calculated using quantification software; DESeq (orange), Cuffdiff (green) or edgeR (blue), q<0.05 (5% FDR) N=3.

3.4.6 Analysis of platform, Donor and experimental variation

To assess the level of variation between platforms, donors and also within a single experiment, RPKM levels of a subset of genes were analysed for correlation (Fig 3.10). To measure the levels of variation seen between

sequencing platforms, biological donors and sample replicates, correlation of the top 1000 most expressed genes between untreated neutrophil samples from several different experiments were compared. Firstly, two datasets of RNA samples from the same donor but sequenced on either the SOLiD or Illumina platforms were compared: this analysis would identify any differences in gene expression due to the different technology platforms (Fig 3.10A). Secondly, samples from 2 different donors, both sequenced on the Illumina platform were compared with each other (Fig 3.10B): this analysis will give insights into donor variability and patterns of neutrophil gene expression. Thirdly, two samples from a single donor (prepared on separate days) sequenced on Illumina platform were compared to each other (Fig 3.10C). This would inform on the intra-donor variability in gene expression. Finally, data from a single experiment on the Illumina platform (but sequenced on different lanes of the sequencing flow cell) were compared to each other to measure the level of intra-experimental variation (Fig 3.10D).

Samples from the same donor sequenced on different platforms (SOLiD and Illumina) (Fig 3.10A), showed a much lower correlation than samples from the same donor sequenced on the same platform (Fig 3.10C), r_s=0.656 and r_s=0.9188 respectively. Indeed, levels of correlation between donors on the Illumina platform were equally high r_s=0.9204 (Fig 3.10B). Intra-experimental (technical variation) correlation was extremely high (r_s=0.9993) (Fig 3.10D).

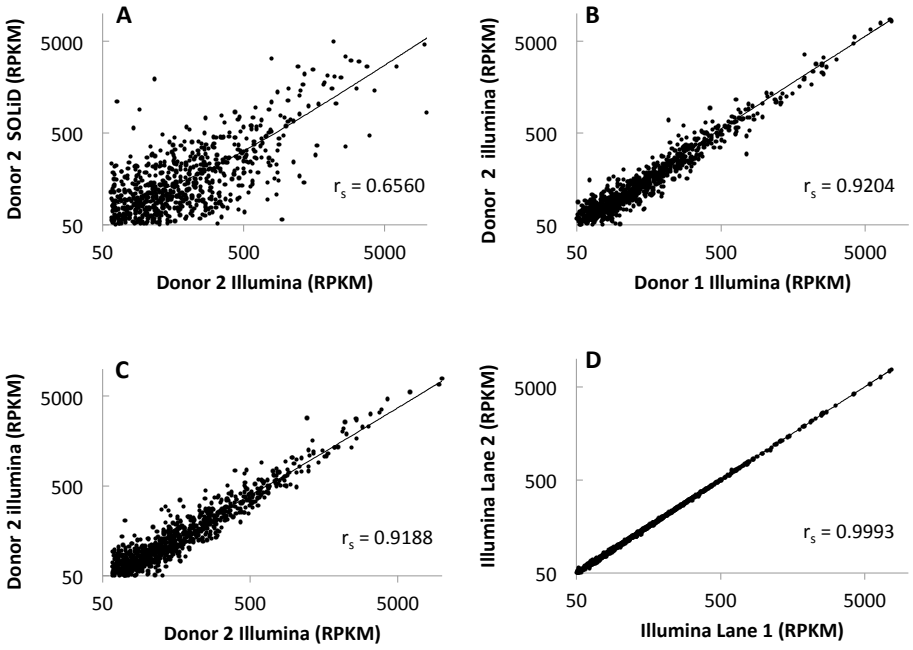


Fig 3.10 Correlation of gene expression values (RPKM) between RNA-Seq samples. Gene expression values were compared between (A) SOLiD and Illumina platforms, (B) 2 biological replicates (on Illumina platform) (C) 2 technical replicates (same Donor on Illumina platform) and (D) 2 lanes from a single sequencing experiment on Illumina platform. In each analysis 1000 genes with the highest expression were used in correlation analysis (Spearman correlation metric (r_s) p<0.0001).

3.4.7 Validation of RNA-Seq data by comparison to qPCR

RNA-Seq analysis methodologies are still regarded as a relatively new technique. As with any new assay or protocol, the most important measure of success, accuracy and applicability is how well the results correlate with the currently used "gold-standard" technique. In the case of gene expression studies the most established techniques remain either micro-array or quantitative PCR (qPCR). RNA-Seq expression values for a selection of genes from 3 Donors measured by both SOLiD and Illumina platforms were compared with those obtained from qPCR for validation. A sample set of genes was selected that included genes that had high levels (>3000

RPKM) of expression: Interleukin-8 (IL-8), nicotinamide phosphoribosyltransferase (NAMPT), and suppressor of cytokine signalling-3 (SOCS3); median levels (50-3000 RPKM) of expression: FBJ murine Osteosarcoma viral oncogene homolog (FOS), intercellular adhesion molecule-1 (ICAM1), and interleukin-1β (IL-1β); and low levels (<50 RPKM) of expression: Fas-associated via death domain (FADD), Jun proto-oncogene (JUN), and TNFα.

Firstly, absolute values of gene expression (RPKM) were compared between SOLiD samples and 3 biological replicates sequenced by Illumina (Fig 3.11A-C). As seen previously, gene expression levels were most similar in samples from the Illumina platform but values from SOLiD sequenced samples were also largely in line with the Illlumina sample replicates.

Secondly, the fold change in gene expression for each gene was compared between the SOLiD samples, Illumina samples and qPCR data (Fig 3.12A-B). Here, the fold change in gene expression for each gene was highly similar across all 3 platforms. This suggests that while the absolute values can vary between platforms (and replicates) the relative change in gene expression correlate well between independent platforms and validation methods.

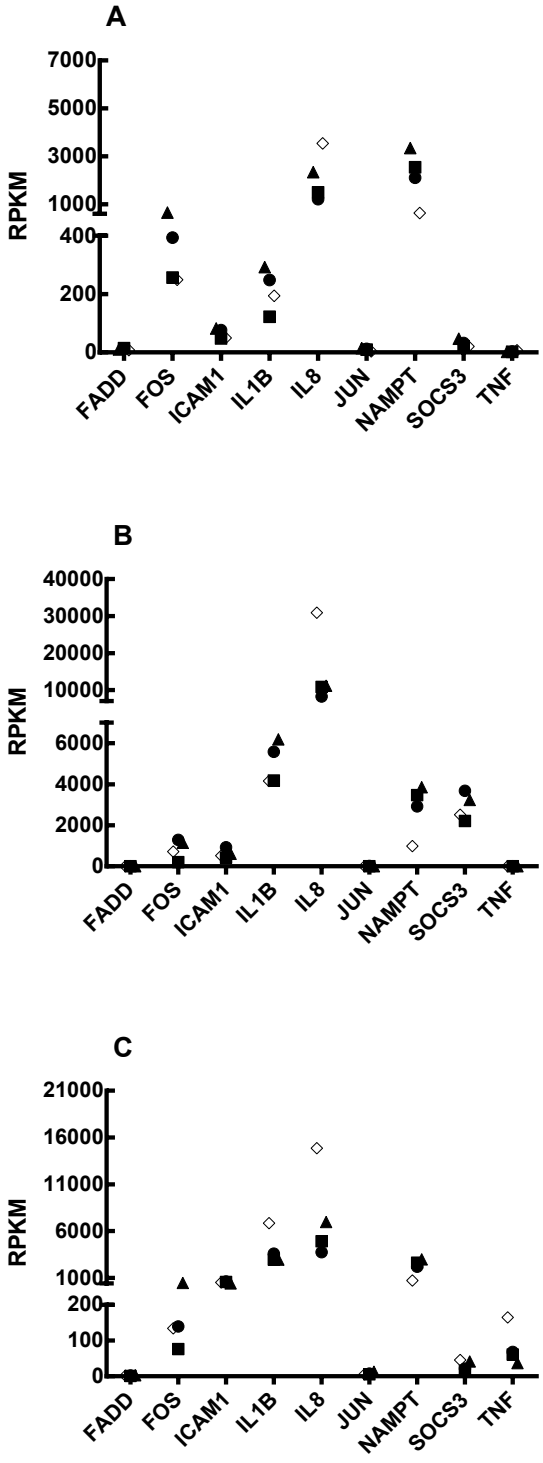


Fig 3.11 Comparison of gene expression levels (RPKM) between RNA-Seq platforms. Expression levels of 9 genes expressed in neutrophils following 1h incubation in (A) absence, or presence of (B) GM-CSF (5 ng/mL) or (C) TNFα (10 ng/mL). Expression levels measured by SOLiD (◇) or by Illumina (▲●■) sequencing platform and calculated by Cufflinks. (SOLiD n=1, Illumina n=3).

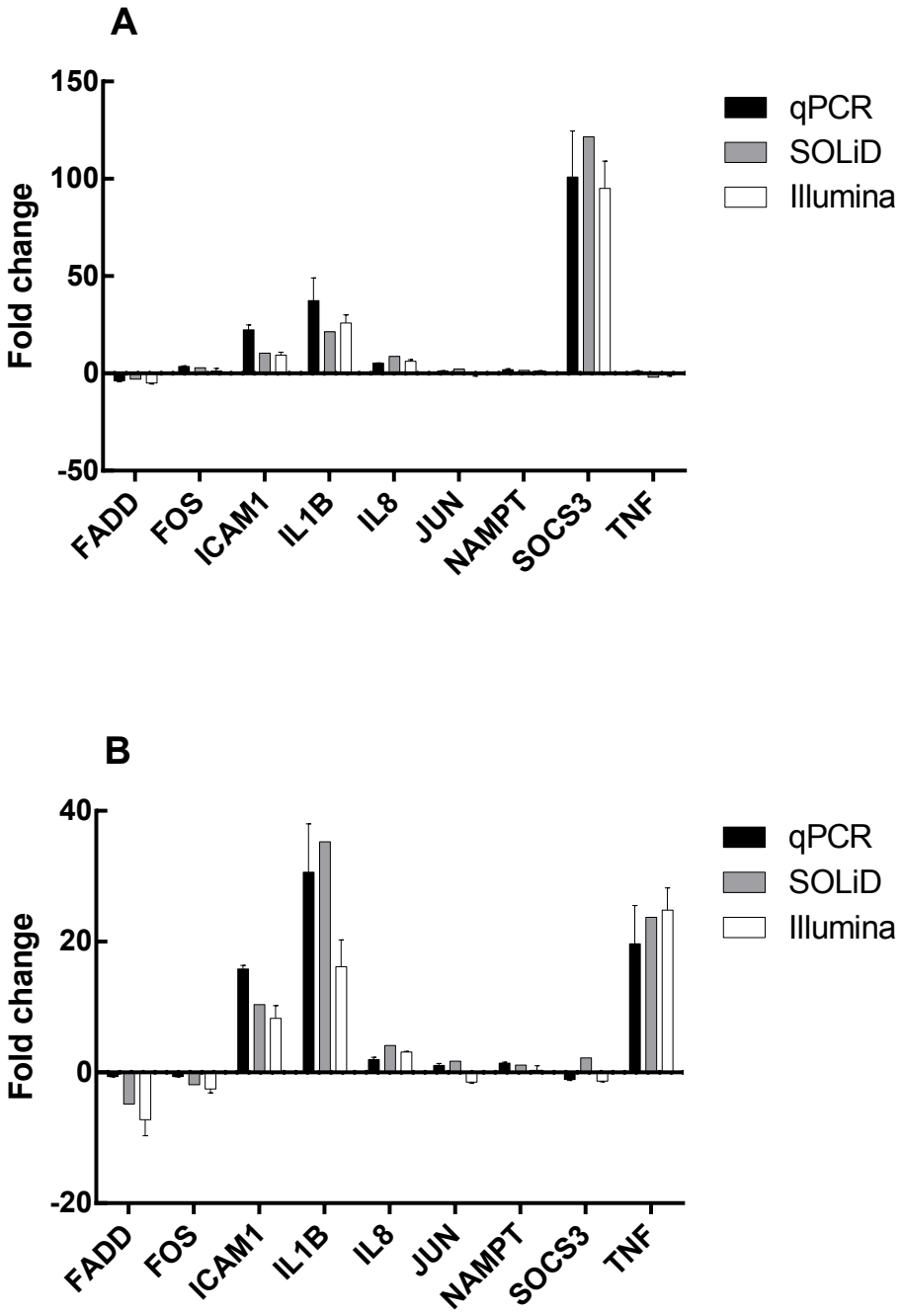


Fig 3.12 Fold change in expression of genes in (A) GM-CSF and (B) TNF-α-treated neutrophils compared to unstimulated cells, measured by qPCR (black bars, n = 3), SOLiD sequencing (grey bars, n = 1) or Illumina sequencing (white bars, n = 3). Error bars represent SEM.

Whilst it is important with any new technique to validate results with more established methods of measurement it is equally important to realise the limitations of the established technique. When comparing data from RNA-Seq and qPCR experiments, there are several factors that need to be considered. Firstly, RNA sample preparation should be consistent between quantification techniques. Secondly, PCR primers should be designed to amplify all variants of the gene under analysis and not miss out alternative splice variants, since RNA-Seq experiments can provide a quantification of all variants of a gene. Most importantly, the effect of normalisation method used for qPCR data must be appreciated.

Quantification of transcript abundance from qPCR data is calculated using the number of PCR cycles required for transcripts to achieve exponential amplification (C_t value). This value can be transformed into a mean normalised expression (MNE) by normalising these values against a housekeeping gene [167,195]. For neutrophil studies, a variety of housekeeping genes can be used, but Standford et al [196] have previously shown many of the commonly-used housekeeping genes used in other cell types vary in expression in neutrophils. This can lead to differences in the calculated relative expression values for a gene of interest, depending on which housekeeping gene is used to normalise the data. By comparison, RNA-Seq data can be normalised to the entire size of the read library and the length of each gene independently (RPKM), and so does not necessitate normalisation to a pre-selected gene or set of genes [116].

3.4.9 Comparison of normalisation method in RNA-Seq and qPCR

mRNA levels of TNFα are known to increase rapidly in neutrophils following stimulation with TNFα [183]. RNA from untreated and TNFα treated neutrophils (10 ng/mL) was prepared and mRNA expression of TNFα was measured using qPCR. Threshold (Ct) values for TNFα mRNA were normalised against several different housekeeping genes that are often used in neutrophil gene expression; β-$_2$microglobulin (B2M), glyceraldehyde-3-phosphate dehyrodgenase (GAPDH), β-actin (ACTB) and hypoxanthine phosphoribosyltransferase 1 (HPRT1) [196]. In parallel, the fold change in TNFα expression for untreated neutrophils versus TNFα treated neutrophils was measured by RNA-Seq and RPKM values calculated by Cufflinks. RNA was prepared using identical RNA preparation methods and neutrophils from the same donor.

The fold changes in expression levels in the TNFα stimulated cells versus the control varied from 10-fold to 22-fold, dependent on which housekeeping gene was selected for normalisation (Fig 3.13). The RNA-Seq value was found to correlate well with qPCR data normalised to GAPDH or B2M, but not to values normalised to ACTB or HPRT1. This highlights the care needed when analysing and interpreting qPCR data, and suggests that consideration should be given to the use of multiple housekeeping genes for normalisation of qPCR data, especially in cells such as neutrophils where the expression of genes used to normalise the data may be subject to

regulation during neutrophil activation. These data suggest that RPKM values are a more suitable and accurate metric for use in gene expression analysis and that this method of quantitation can be compared to appropriately controlled qPCR data.

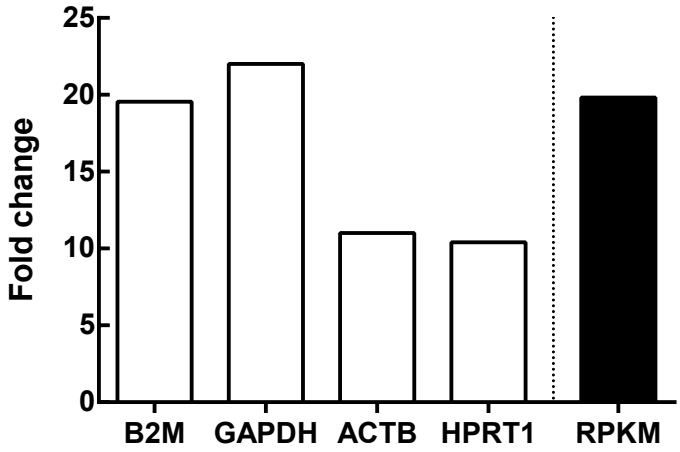


Fig 3.13 Fold change in expression of TNFα mRNA measured by qPCR (open bars) and RNA-Seq (black bar). qPCR Ct values for TNFα expression normalised to values for commonly-used housekeeping genes; β-microglobulin (B2M), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), β-actin (ACTB), hypoxanthine phosphoribosyltransferase 1 (HPRT1). RNA-Seq value normalised to read-library size and gene length (RPKM).

3.4.10 Correlation of qPCR Ct values with RNA-Seq RPKM values

Accurate detection of low-abundance transcripts in samples if often difficult to distinguish from experimental noise and DNA contamination. It is generally considered that a transcript requiring >30 cycles for detection in a qPCR experiment (Ct value > 30,) is likely to be a false positive. Likewise,

care must be taken when setting a RPKM threshold for gene expression in order to balance the number of false positives with the number of false negatives. A study by Ramskold *et al* [197], attempted to characterise transcriptomes across different tissue types in both humans and mice and determined that an RPKM value of 0.3 was an appropriate threshold, above which it can be concluded that the transcript is genuinely expressed.

To confirm these findings and ascertain whether 0.3 RPKM was an appropriate threshold for neutrophil studies, RNA was extracted (as previously described) from 1 h untreated neutrophils and assayed using qPCR to determine the Ct value of 84 genes which had a range of expression levels. Values were compared to RNA-Seq data from 1h untreated neutrophils from the same biological donor. Ct and RPKM values were found to significantly negatively correlate (r_S=-0.91, $p<0.0001$, Spearman correlation). Importantly, data convergence was observed at 30-cycles in qPCR and 0.3 RPKM in RNA-Seq (Fig 3.14). These data confirm that the pre-defined threshold of 0.3 RPKM is entirely appropriate and adequate to determine whether a transcript is expressed in neutrophils.

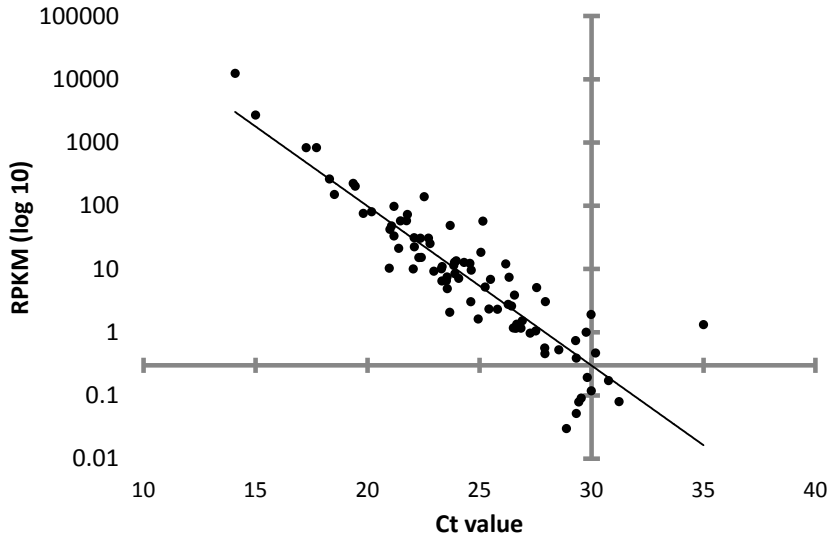


Fig 3.14 Correlation of gene expression values measured by qPCR and RNA-Seq. RNA from untreated neutrophils from the same biological donor was assayed using qPCR and RNA-Seq. Cycle threshold (Ct) values for qPCR were found to correlate significantly with RPKM gene expression values (r=-0.91, p<0.0001, Spearman correlation). qPCR data collected by Dr C. Lam and analysed and reproduced here with permission.

3.4.11 Downstream analysis of RNA-Seq data

The above methods describe the software and protocols necessary to produce accurate gene expression data from a starting sample of total RNA. This pipeline can be utilised to produce normalised absolute-values (RPKM) of gene expression using Cufflinks, or a list of DE genes between multiple samples via Cuffdiff. These data are of great use and can be manually curated to extract information on genes of interest. However, to analyse larger portions of the data, (or compare entire datasets with each other) further downstream bioinformatic techniques are required. In general, bioinformatic analysis of gene expression data aims to simplify large amounts of data by identifying smaller gene-sets with similar expression patterns, or quantify levels of dissimilarity between whole

datasets/samples. This can be achieved in a variety of ways. A selection of software and bioinformatic analyses for downstream analysis of RNA-Seq data are described below. These methods complement the above bioinformatic RNA-Seq pipeline to form a comprehensive set of bioinformatic techniques which form the complete bioinformatic pipeline used in future results chapters (Chapters 4-6). Further details of specific downstream analysis are provided in the results sections of Chapters where they are employed.

3.4.11.1 cummeRbund

cummeRbund is a software package developed by the Cufflinks group (Trapnell et al) [168]. It utilises the output files of Cufflinks and runs in the R software environment to produce several graphical representations of the gene expression data. These include: global visualisation of data in terms of quality, dispersion, or gene expression distribution; multi-dimensional scaling (MSA) for 2-dimensional representation of whole data sets; and gene clustering via heat maps [168].

3.4.11.2 Ingenuity Pathway Analysis (IPA)

The most powerful method of extracting meaningful data from large data sets is to model the data against large databases of canonical biological data. These are provided by a variety of online resources such as the Kyoto encyclopaedia of genes and genomes (KEGG) [198] or CLCbio by Qiagen. By far the most comprehensive database of biological pathways and interactions is held by Ingenuity systems [199]. The Ingenuity pathway analysis

(IPA) software is a licensed online resource which allows the uploading of large datasets of NGS data, the data can then be modelled against their database of canonical pathways and referenced biological interactions. Subsequent analysis of the data can then be carried out to identify pathways or networks of genes which are significantly enriched with DE genes. Additionally, since the database of signalling pathways is so comprehensive, the software can be employed to identify upstream regulators and transcription factors based on the gene expression values provided. This is a powerful technique for predicting activation of signalling pathways and associated networks of genes from raw RNA-Seq gene expression values.

3.4.11.3 Gene ontology

An alternative method of summarising large gene lists is by Gene Ontology analysis. Genes are categorised by their universal gene ontology annotation (as defined by the Gene Ontology consortium [200]). This provides a method of identifying sets of DE genes which share a common biological process, molecular function or cellular component, and identifying relationships between genes that would otherwise be impossible to achieve using a manual approach.

3.4.12 Final bioinformatic pipeline for neutrophil gene expression studies

By way of summary, the components and work flow of the final bioinformatic pipeline developed for neutrophil gene expression analyses is shown in Fig 3.15. This represents a complete workflow for the production, quantification and analysis of neutrophil RNA-Seq data.

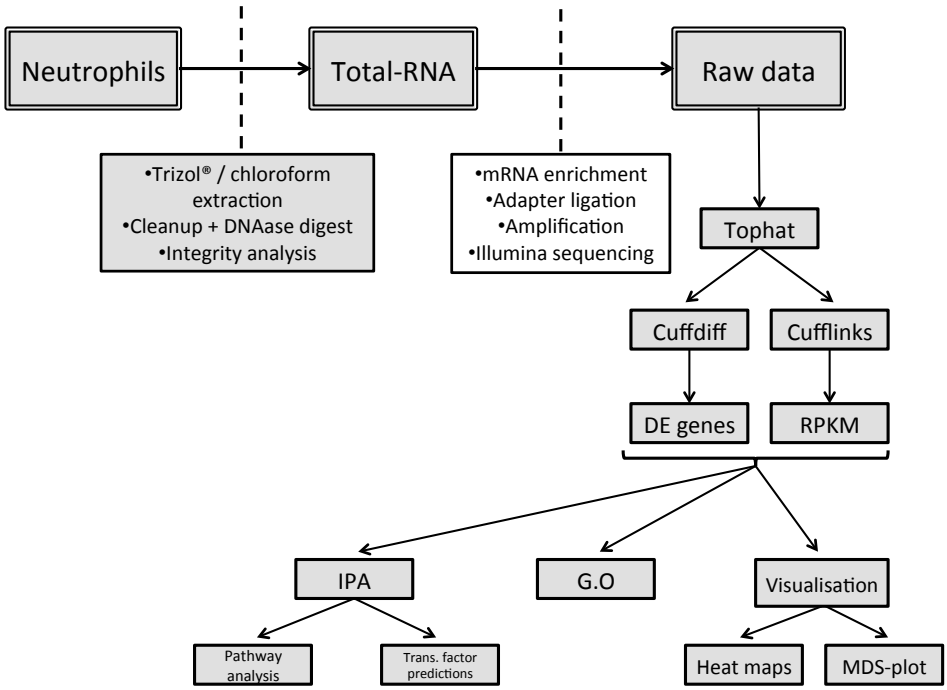


Fig 3.15 Bioinformatic pipeline for production quantification and analysis of neutrophil RNA-Seq data. Flow diagram of sample preparation processes and software incorporated into complete bioinformatic pipeline. Processes in white box were completed by 3rd party (BGI International).

3.5 Summary

Modern sequencing technology provide a method of accurately sequencing millions of DNA or RNA fragments on a massively parallel scale producing huge amounts of raw data. Whilst the efficiency and usability of sequencing technology has improved in recent years, the same cannot be said of NGS analysis software (in particular software that is open source). The bioinformatic community is saturated with various kinds of analysis software, each offering specific benefits over each other with respect to different aspects of the analysis process. Furthermore, there is no community wide agreement on best practices when it comes to analysing NGS data. The eclectic nature of bioinformatic methods in the literature attests to this. Hence before undertaking a study into neutrophil gene expression, it was first necessary to compile a robust set of software and protocols to accurately analyse the raw data produced by RNA-Seq. This pipeline could then be implemented for further studies into neutrophil gene expression.

Neutrophils naturally express less mRNA than other leukocytes [178], thus it was necessary to quantify the amount of total RNA that could be extracted from a whole blood sample, whilst also assessing the integrity and quality of RNA following neutrophil isolation, RNA extraction and cleanup. Standard isolation of $30x10^6$ neutrophils by Polymorphprep™ and extraction using TRIzol® (including a DNA digest step) was found to be

sufficient to produce 5 µg of high integrity (>8 RIN) purified RNA, which is often a prerequisite of 3rd party sequencing service providers.

Sequencing of neutrophil RNA was carried out on 2 of the most popular sequencing platforms, SOLiD and Illumina. The paired-end sequencing on the SOLiD platform suffered from poor read quality values such that the shorter paired fragment could not be used in the mapping stage, the remaining reads mapped poorly to a reference sequence (<35 %) . Whilst paired-end sequencing is a useful method of increasing the total amount of read data from a single experiment, the library preparation protocols are more complex than single-end protocols and thus more likely to suffer experimental error. Furthermore, the ability of paired-end sequencing to improve mapping rates by removing mapping-location ambiguity is less relevant when dealing with samples where a complete and comprehensive reference sequence is available (such as with human samples). In contrast, sequencing carried out on Illumina platform by single-end sequencing was of a much higher quality and consequently resulted in extremely high mapping rates (>94%). Annotation and quantification of mapped reads was carried out by count-based (edgeR and DESeq) and gene normalisation (Cufflinks) techniques. These two approaches are frequently used in RNA-Seq studies. Indeed, rather than one technique becoming preferred or optimal, it is likely that studies in the future will start to incorporate a combination of both approaches [201], for example a count based quantification for absolute gene expression values and a normalised

expression (RPKM) approach for transcript-level quantification. An added advantage of RNA-Seq data is that once reads are mapped adequately, data can subsequently be re-analysed or re-quantified using a different approach without the need for producing new samples.

Following quantification, data were validated against results gained by qPCR.

Despite numerous publications using RNA-Seq as the primary method of gene expression analysis, it is still common for research articles to include comparative qPCR results as validation of the accuracy of the primary RNA-Seq data [202,203]. Neutrophil RNA-Seq data from both platforms (SOLiD and Illumina) showed good correlation with results from qPCR, in particular when comparing fold change in gene expression rather than absolute values. In addition, correlation between datasets from the Illumina platform (either from a single donor or two donors) was extremely high ($r > 9.1$). This highlights the robust and reproducible nature of RNA-Seq data as a method of analysing large amounts of data.

Two aspects of gene expression analyses that hinder studies using micro-arrays or PCR are signal to noise ratios and normalisation of data [107,196,204]. During the development of this bioinformatic pipeline the normalisations techniques and threshold value for positive expression were assessed in RNA-Seq by comparison to traditional qPCR methods. It was found that the normalised metric of RPKM was comparable to results gained by qPCR and since it relies on a defined value of gene length, is less variable than results

gained using multiple reference genes by qPCR. Moreover, it was confirmed that a previously defined [197] value of 0.3 RPKM is an appropriate cut-off value for gene expression and correlates well with a C_t value of 30 in qPCR.

Since analysis of RNA-Seq is often (by virtue of the different software programs required) a multi-step process, compatibility of data with downstream analysis software is of crucial importance. The work flow within the bioinformatic pipeline benefits from being fully compatible at all stages, such that the output files from the annotation steps using cufflinks can seamlessly be inputted into each of the downstream analysis programs, for example cummeRbund or IPA. The greatest benefit of this inter-compatibility is the usability and lack of informatic experience needed to carry out a RNA-Seq experiment using the above described software and settings. Improvements in software during the course of this study (Tophat alone received 12 iterative upgrades between 2010 and 2012) led to both improved performance and streamlining of the pipeline, for example, the ability to map all reads in a single Tophat command was not possible until the release of version 1.4.

In summary the processes described above form a robust and user-friendly pipeline of analyses that can accurately measure the gene expression profile of neutrophils under different conditions. This pipeline will be utilised in subsequent chapters to further define the transcriptional profile of neutrophils in conditions of simulated inflammation.

Chapter 4: RNA-Seq analysis of neutrophil priming by GM-CSF and TNFα

Bioinformatic analyses presented within this chapter were included in a publication in which I was co-lead author.

Wright HL†, Thomas HB†, Moots RJ, Edwards SW (2013) RNA-Seq Reveals Activation of Both Common and Cytokine-Specific Pathways following Neutrophil Priming. PLoS ONE 8(3): e58598. doi:10.1371/journal.pone.0058598

4.1 Introduction

Neutrophil function *in vivo* is regulated by "priming" by inflammatory signals generated during an inflammatory response. Priming induces several rapid (<1 h) functional changes such as the mobilisation of internal granules (containing pre-formed receptors) to the cell surface, phosphorylation of key signalling proteins, and assembly of the NADPH oxidase leading to increased respiratory burst in response to a secondary activating signal [3].

Several agents are able to prime neutrophils, including: lipid mediators such as leukotriene B_4 and platelet activating factor (PAF); hormones and growth factors, such as melatonin and substance P; bacterial products, such as lipopolysaccharide (LPS); and numerous cytokines and chemokines, such as IL-1, -3, -8, G-CSF, IFNγ, GM-CSF and TNFα [205,206]. Stimulation of neutrophils by these agents *in vitro* induces a similar, "primed" phenotype

resulting from the short-term, rapid molecular changes described above. Consequently, priming agents are sometimes used interchangeably in neutrophil studies on the assumption that priming occurs via common mechanisms [207,208]. Furthermore, it is well established that cytokines and other priming agents are able to regulate gene expression in neutrophils [11,181,183,209,210] but few studies have examined the global gene expression profile in neutrophils following priming, and none to date have directly compared the patterns of gene expression induced by different priming agents.

Although regarded as a prerequisite for neutrophil activation, priming also serves an important role as a regulatory mechanism. By requiring a second stimulus for transition from a quiescent neutrophil (for example in peripheral blood), to an activated neutrophil (for example at the site of inflammation), neutrophils are able to regulate their activation state much more specifically. Ultimately, this decreases the possibility of inappropriate or excessive neutrophil activation leading to damage to host tissue from the release of ROS and proteases. Despite this, neutrophil dysfunction and inappropriate activation is often a hallmark of inflammatory and autoimmune diseases. In many cases, the mechanisms leading to initial inappropriate activation of the immune system are poorly understood, but it is clear that elevated levels of inflammatory cytokines are fundamental to the progression and exacerbation of these conditions [2,3,32,211].

The importance of inflammatory cytokines in inflammatory diseases is highlighted by the success of the successful application of anti-cytokine (or cytokine-receptor) drug therapy. Drugs such as Anakinra (IL-1R antagonist), Tocilizumab (anti-IL-6R), Milivimumab (anti-GM-CSFR), secukinumab (anti-IL17) and Belimumab (anti-B-cell activating factor (BAFF)) are routinely used in a variety of inflammatory diseases such as gout, SLE, psoriasis, Crohn's and RA [3,211,212] (See table 1.1 for a more complete list of cytokine-targeting therapies). However, the most successful target for treatment of inflammatory disease is TNFα. Several drugs, such as Infliximab, Adalimumab, Cerolizumab-pegol, Golimumab and Etanercept are considered the front line treatment for conditions such as RA, COPD, Ankylosing spondylitis and Crohn's disease. However, an important feature of these drugs is the varying degree to which patients respond. For example, it is estimated that approximately 30% of patients with RA who are prescribed anti-TNF will not respond [213]. These patients will often have to switch therapies a number of times to alternative anti-TNF drugs, or to drugs such as Rituximab (anti B-cell) or Abatacept (anti-T-cell), before adequate disease control is achieved and maintained. This highlights the heterogeneity that exists in inflammatory diseases such as RA and suggests that different cytokines may be responsible for driving inflammation in different patients. Whilst treating inflammatory diseases using a single anti-cytokine drug is of merit, a comprehensive understanding of the molecular changes induced by inflammatory cytokines in health and disease, and how

this regulation differs between cytokines and individuals, is important to understand immune regulation, but could also lead to a more rationale-based approach to drug treatment.

4.2 Aims

The aims of this chapter were:

1. To utilise the previously-described pipeline of methods and bioinformatic techniques (Chapter 3) to quantify the neutrophil transcriptome following priming with GM-CSF and TNFα

2. To compare the molecular changes in neutrophils induced by priming and identify genes and signalling pathways that are either common or specific to each priming agent.

3. To determine if predictions made by bioinformatic analyses can be validated by functional assays.

4.3 Results

4.3.1 Effect of GM-CSF and TNFα on neutrophil function

Several cytokines are commonly used as priming agents for *in vitro* studies of neutrophil function. To assess whether priming by different agents had any effect on neutrophil function, priming of respiratory burst and levels of apoptosis after overnight incubation were compared in neutrophils treated with either GM-CSF or TNFα.

4.3.1.1 Ability of GM-CSF and TNFα to prime the respiratory burst

Neutrophils were treated for 30 min in the presence or absence of priming concentrations of GM-CSF (5 ng/mL) or TNFα (10 ng/mL). Following incubation, neutrophils were stimulated with fMLP (1 μM) in the presence of luminol (10 μM) and immediately measured every 24 s over a 25 min time course using a plate reader (see Methods, section 2.2.8). Both GM-CSF and TNFα primed neutrophils for an enhanced respiratory burst, peaking at approximately 1 min. This first peak represents extracellular release of ROS [214]. A second peak at 5 min was seen in both primed samples, which corresponds to a delayed intra-cellular activation of reactive oxygen species (ROS), as previously described [214].

The primed response of ROS production was slightly higher in GM-CSF primed samples compared to TNFα primed samples (GM-CSF; 32,2316 relative luminescence units (RLU) p=0.0022, TNFα; 28,5858 RLU p=0.0077, N=4). Unprimed neutrophils exhibited a modest increase in

chemiluminescence compared to unstimulated cells (negative control) (Fig 4.1A-B).

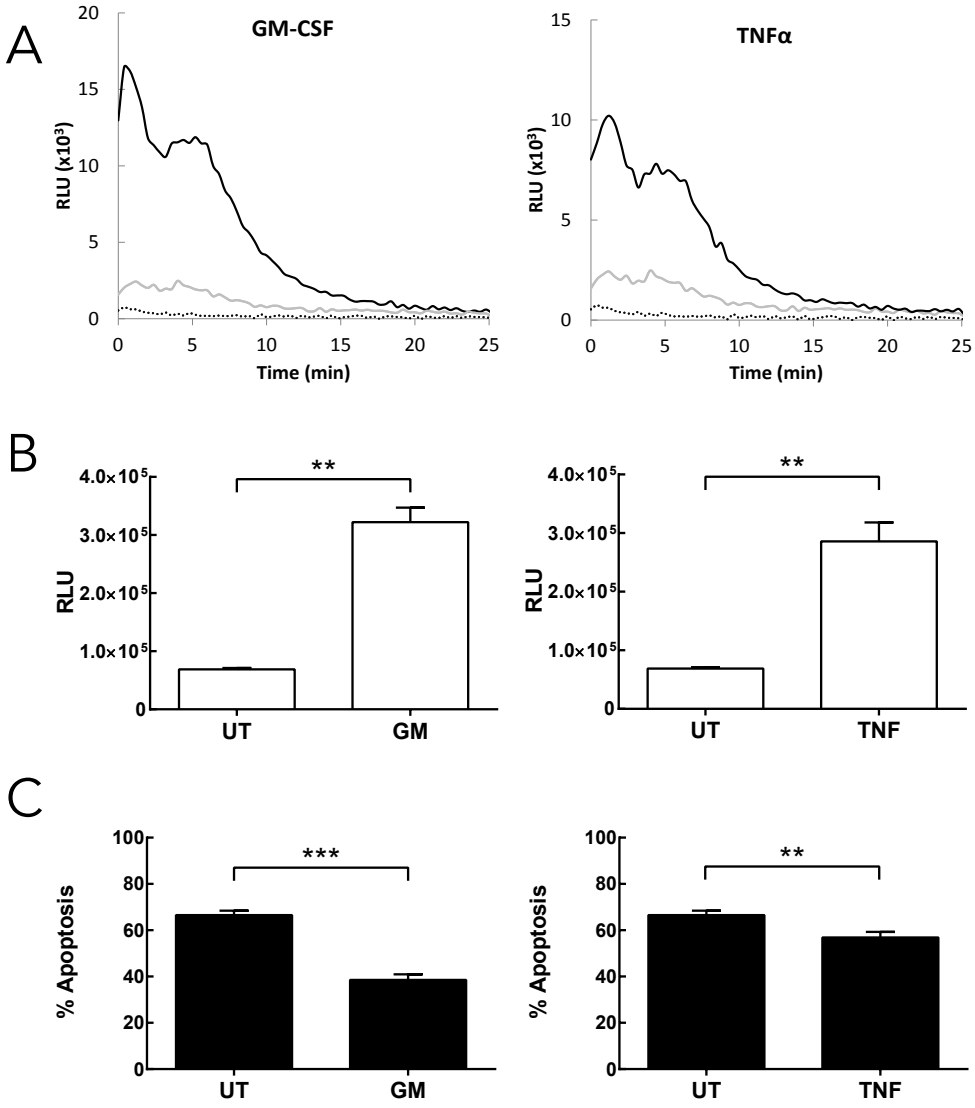


Fig 4.1 GM-CSF (GM) and TNFα (TNF) have similar effects on neutrophil function. (A) Neutrophils were primed with either GM-CSF (5 ng/mL) or TNFα (10 ng/mL), or untreated (UT) for 30 min. The respiratory burst was stimulated with fMLP (1 μM) in the presence of luminol (10 μM). Primed: black line; unprimed: grey line; unstimulated: dotted line. Graph shows representative trace of relative luminescence units (RLU). (B) Results of 4 separate chemiluminescence experiments calculating the RLU by quantifying area under curve in each condition. Error bars represent ± SEM. (p<0.01, Student's t-test). (C) Neutrophils were incubated overnight (18 h) in the absence (UT) or presence of GM-CSF (5 ng/mL) or TNFα (10 ng/mL), and percentage apoptosis was quantified by annexin-V/propidium iodide staining and measured by flow cytometry. N=7. Error bars represent SEM. (*** p<0.001, Student's t-test).**

4.3.1.2 Effect of GM-CSF and TNFα on neutrophil apoptosis

To assess the effect of priming neutrophils with GM-CSF and TNFα on levels of apoptosis, neutrophils were incubated at 37 ºC + 5% CO_2 for 18 h in the absence or presence of GM-CSF (5 ng/mL) or TNFα (10 ng/mL). Apoptosis was quantified by Annexin V/propidium iodide staining and flow cytometry (see Methods section 2.2.5). Levels of apoptosis were significantly lower in both GM-CSF and TNFα treated samples compared to controls ($p<0.01$, paired Student's t-test: untreated 66.5 % ±2.08 %; GM-CSF 38.6 % ±2.55 %; TNFα 56.9 % ±2.61 %).

Taken together, these data indicate that both GM-CSF and TNFα are able to prime neutrophils leading to: elevated respiratory burst (upon stimulation); and delayed apoptosis. However, these effects are not identical between priming agents, suggesting that GM-CSF and TNFα may induce subtle differences in neutrophil signalling during the priming response.

4.3.2 Whole transcriptome sequencing of primed neutrophils

To investigate the molecular changes in neutrophils following priming with GM-CSF and TNFα, whole transcriptome sequencing by RNA-Seq was carried out on mRNA from neutrophils incubated for 1 h with either GM-CSF (designated as GM), TNFα (designated as TNF) or without stimulation (designated as UT).

RNA was extracted by TRIzol®/chloroform precipitation (see Methods section 2.2.9) and sequenced on the Illumina HiSeq 2000 platform (summarised in 1.5.5) using upwards of $40x10^6$ single end reads (see Appendix table A.5 for total no. of reads). Reads where mapped to the human genome (hg 19) by Tophat and annotated using Cufflinks (as previously described in Chapter 3).

4.3.3 RNA-Seq analysis of primed neutrophils

4.3.3.1 Analysis of gene expression by Cufflinks.

Following mapping and annotation (as previously described in Chapter 3), Cufflinks analysis was carried out to determine the number of genes expressed in each of the 3 sample conditions (UT, GM and TNF). Fig 4.2A details the number of genes which are either treatment specific, expressed under multiple conditions, or not expressed in neutrophils in any of the 3 conditions (using a cut-off RPKM of 0.3). In total, 11,201 out of a possible 23,283 genes were expressed in neutrophils in any of the 3 conditions. Of these, 10,056 (89.8%) were expressed in all conditions. Interestingly, some genes were only expressed under certain conditions (condition-specific).

GM treatment resulted in the expression of the most condition-specific genes (229), with marginally less in UT (220) and TNF (193). A total of 12,082 genes were not expressed in any condition, suggesting that under the conditions studied, neutrophils express approximately half of the human transcriptome. This number broadly agrees with previous micro-

array-based studies into neutrophil transcript expression of peripheral blood neutrophils [181,215,216] (Fig 4.2A). If a threshold of expression of RPKM of 10 is applied to the data, to remove genes with low abundance from quantification analysis, the total number of genes expressed (≥ RPKM 10) decreases from 11,201 to 3,574, with the vast majority of filtered genes being those expressed in all conditions (RPKM ≥0.3 – 10,056 genes, RPKM≥ 10 – 2,829 genes). Conversely, filtering out low abundance genes has a lesser effect on the number of condition-specific gene changes. For example, the number of GM-CSF-specific genes decreases from 229 to 187, while TNFα-specific genes decreases from 193 to 109 (Fig 4.2B).

These data reveal that neutrophils express almost half of the human transcriptome and that 90% of these genes are expressed irrespective of GM-CSF or TNFα treatment. However, a large number of these constitutively-expressed genes (a total of 7225 genes) are present at low abundance (< RPKM 10) whereas the majority of condition-specific genes are expressed at a higher abundance (> RPKM 10).

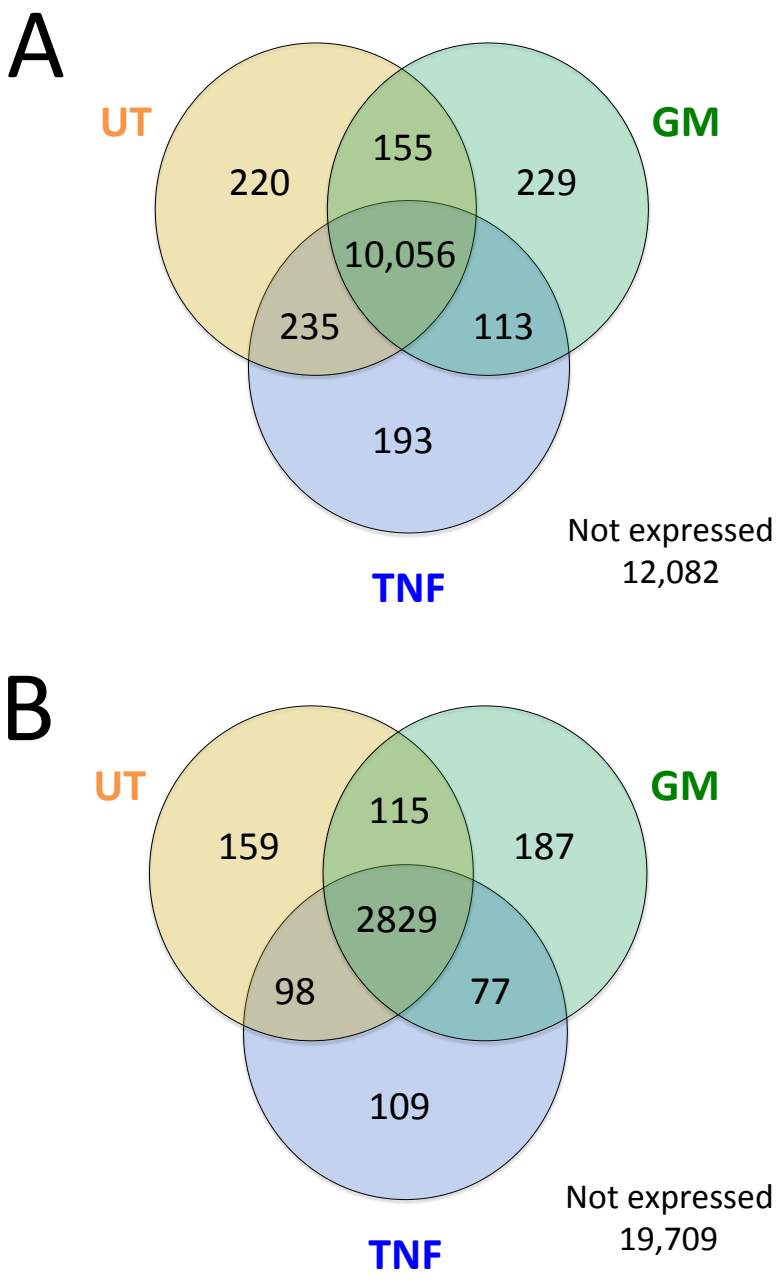
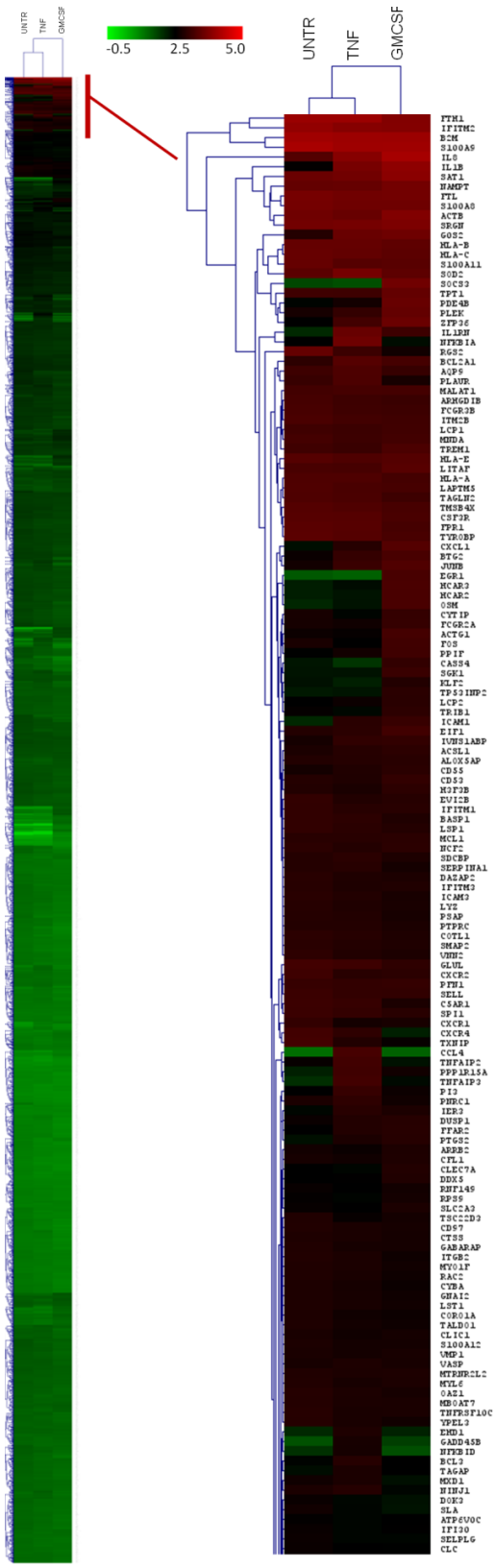


Fig 4.2 (A) Venn diagram showing the number of genes expressed (RPKM ≥ 0.3) in untreated (UT), GM-CSF primed (GM), TNFα primed (TNF) neutrophils, or not expressed in any condition. (B) Increasing the expression threshold to an RPKM of 10 to filter-out low-expression genes decreases the number of expressed genes in all conditions. RPKM values of genes were calculated by Cuffdiff.

4.3.3.2 Hierarchical clustering of highly expressed genes in UT, GM and TNF

To better visualise and identify genes which are either common or specific to neutrophil priming conditions, hierarchical clustering of all genes with an RPKM ≥ 10 (in at least one of the 3 datasets) was performed using Multiple experiment Viewer (MeV)[217](Fig 4.3). An expanded heatmap of the (150) highest expressed genes is also shown in Fig 4.3. These genes are associated with a number of functions and can be broadly categorised as: cell surface receptors; cytokines/chemokines; Interferon-induced genes; Major Histocompatibility Complex (MHC) proteins; calcium-binding proteins; adhesion molecules and apoptosis regulators.



Receptors

complement component 5a receptor 1
colony stimulating factor 3 receptor (granulocyte)
interleukin 8 receptor, alpha
interleukin 8 receptor, beta
chemokine (C-X-C motif) receptor 4
Fc fragment of IgG, low affinity IIa, receptor (CD32)
Fc fragment of IgG, low affinity IIIb, receptor (CD16b)
free fatty acid receptor 2
formyl peptide receptor 1
triggering receptor expressed on myeloid cells 1

Interferon-induced

interferon, gamma-inducible protein 30
interferon induced transmembrane protein 1 (9-27)
interferon induced transmembrane protein 2 (1-8D)
interferon induced transmembrane protein 3 (1-8U)

MHC molecules

beta-2-microglobulin
major histocompatibility complex, class I, A
major histocompatibility complex, class I, C
major histocompatibility complex, class I, B
major histocompatibility complex, class I, E

Calcium Binding Proteins

S100 calcium binding protein A9
S100 calcium binding protein A11
S100 calcium binding protein A12
S100 calcium binding protein A8
S100 calcium binding protein A4

Apoptosis Regulators

immediate early response 3
myeloid cell leukemia sequence 1
BCL2-related protein A1

Adhesion molecules

intercellular adhesion molecule 1
intercellular adhesion molecule 3
integrin, beta 2
selectin L
selectin P ligand

Cytokines / Chemokines

chemokine (C-C motif) ligand 4
chemokine (C-X-C motif) ligand 1
interleukin 1, beta
interleukin 1 receptor antagonist
interleukin 8
oncostatin M

Fig 4.3 Hierarchical clustering of genes expressed (≥ RPKM 10) in untreated neutrophils (UNTR) or following priming by GM-CSF or TNFα. RPKM values are represented on a log_{10} scale where green represents low expression, black median expression and red high expression. Right hand column shows expanded heatmap of the 150 most highly expressed genes. Several of the most highly expressed genes can be categorised into functional groups, such as: cell surface receptors; cytokines/chemokines; Interferon-induced genes; Major Histocompatibility Complex (MHC) proteins; Calcium binding proteins; adhesion molecules and apoptosis regulators.

4.3.3.3 Analysis of differentially expressed genes by Cuffdiff

Following quantification of neutrophil genes expressed in each condition by Cufflinks, Cuffdiff analysis was applied to the data to quantify the number of genes which were significantly differentially expressed (DE) between sample conditions (q< 0.05, 5% FDR). This identifies the genes most affected during neutrophil priming and also identifies which genes were DE in neutrophils during priming by either GM-CSF or TNFα.

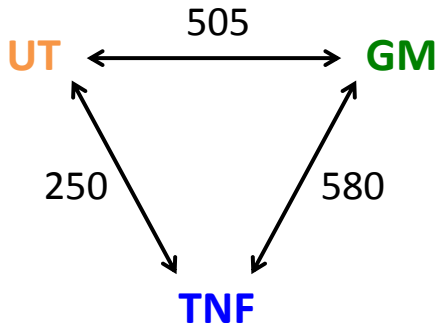


Fig 4.4 Number of significantly differentially expressed genes between neutrophils treated with GM-CSF (GM), TNFα (TNF) or untreated (UT). Significance calculated by Cuffdiff (q<0.05, FDR 5%).

Cuffdiff analysis revealed that neutrophil priming by GM-CSF significantly regulated the expression of 505 genes compared to untreated control. Likewise, priming by TNFα resulted in 250 differentially expressed genes. Surprisingly, the greatest number of DE genes was found to be between

GM-CSF and TNFα (580 genes) (Fig 4.4). It could be hypothesised that since GM-CSF and TNFα exert similar effects on neutrophils phenotypically (as seen in Fig 4.1) their gene expression profile would be similar. However, these RNA-Seq data suggest that there are a greater number of genes differentially expressed between GM-CSF and TNFα than between either GM-CSF or TNFα and untreated controls. These 580 genes are further analysed in a later section (see section 4.3.3.6).

Of the 755 genes which were DE between either GM-CSF or TNFα and UT (UT vs GM, 505 genes; UT vs TNF, 250 genes), 40 genes were up-regulated by at least 10-fold by either GM-CSF or TNFα. These genes are listed in Table 4.1. Interestingly, several genes relating to cytokines/chemokines were differently expressed by the two treatments. For example: Chemokine (C-C motif) ligand 3 (CCL3), CCL4 and the TNFα-gene (TNF) were only significantly upregulated by TNFα treatment, whereas oncostatin M (OSM) was only significantly upregulated by GM-CSF treatment. CXCL1 (also known as Chemokine (C-X-C Motif) Ligand 1) was upregulated approximately 3-fold greater by GM-CSF compared to TNFα (GM, 10.4-fold; TNF, 3.6-fold), whereas CXCL2 was upregulated 6-fold greater by TNF than by GM (TNF,29-fold; GM,4.7-fold). The cytokines interleukin-1A (IL1A), IL-1β and interleukin receptor agonist (IL1RN) were the only genes that were significantly upregulated by more than 10-fold by both treatments (Table 4.1).

Table 4.1 Genes significantly up-regulated at least 10-fold in either GM-CSF or TNFα treated neutrophil samples compared to untreated sample. Values represent fold change in expression relative to control. All values are significant, as calculated by Cuffdiff (q<0.05, 5% FDR) unless stated (NS = not significant).

Gene	GM-CSF	TNFα
CCL3	NS	41.5
CCL4	NS	99.6
CD69	57.5	NS
CISH	102.1	NS
CXCL1	10.4	3.6
CXCL2	4.7	29.0
DUSP2	NS	12.1
EDN1	16.5	NS
EGR1	57.9	NS
EGR2	21.5	NS
GADD45B	NS	15.8
GPR84	NS	74.9
HBEGF	33.2	NS
HCAR2	12.1	NS
HCAR3	12.2	NS
HRH4	32.3	NS
ICAM1	10.6	7.7
IL1A	35.3	67.0
IL1B	22.8	13.8
IL1RN	12.4	31.4
KCNJ2	NS	16.6
MFSD2A	-1.6	11.9
NFKBIA	NS	11.9
NFKBIE	NS	15.8
OLR1	NS	3.2
OSM	15.0	NS
PDE4B	11.8	NS
PLAU	5.8	13.9
PNPLA1	-2.1	10.2
PPP1R15A	3.6	10.2
RHOH	26.3	NS
SLC35B2	NS	10.4
SOCS3	90.2	NS
TARP	13.5	NS
TIFA	6.7	17.7
TNF	NS	25.8
TNFAIP3	2.7	16.1
TNFAIP6	NS	10.6
TRAF1	NS	11.6
ZFP36	11.4	4.7

4.3.3.4 Gene ontology analysis of genes with differential expression in either GM-CSF or TNFα

Following identification of a subset of genes showing differential expression during neutrophil priming with GM-CSF or TNFα, further characterisation of these genes was performed by Gene Ontology (GO) analysis. GO analysis is a method of gene annotation used to collate and categorise large lists of genes (of the type often produced during large genetic studies such as RNA-Seq analyses) on the basis of known functional associations. Genes are assigned to GO terms based on their known function or functional association. GO terms are broadly categorised into 3 hierarchical classes: biological process; molecular function or cellular component [200]. Within each class, additional GO terms are structured in a hierarchical manner such that "high level" (or broadly descriptive) terms would include terms such as "signal transduction" or "cell growth and maintenance". Whereas, more specific "low level" GO terms would include terms such as "pyrimidine metabolism" or "cAMP synthesis" [218]. This allows the summarising of large sets of genes and the identification of common functional properties within groups of co-expressed genes.

Gene Ontology analysis using the online software DAVID [219] revealed that genes DE by either GM-CSF or TNFα led to enrichment of GO terms that were cytokine-specific and common to both treatments (Table 4.2). High level GO categories, such as "inflammatory response" or "response to wounding" were represented in both GM-CSF and TNFα samples, whereas

more specific low level GO categories were represented in only one treatment sample. For example, "regulation of I-kappaB kinase/NF-kappaB cascade" and was only represented in TNFα samples, whilst "positive regulation of nitric oxide biosynthetic process" was only represented in GM-CSF samples (Table 4.2).

4.3.3.5 Pathway analysis of genes with differential expression in either GM-CSF or TNFα

GO analysis is a useful bioinformatic approach to describe biological functions or cellular processes and to discover functional relationships within in a set of genes. However, the process is not sufficiently powerful to accurately predict activation of specific signalling pathways or identify upstream regulation of transcription factors. Therefore, further analyses were carried out on the genes DE by either GM-CSF or TNFα using Ingenuity Pathway Analysis software (IPA) [220] to identify if the genes expressed during neutrophil priming by each priming agent activated common or different signalling pathways and transcription factors.

IPA analysis identified that neutrophils primed with TNFα or GM-CSF significantly regulated a number of intracellular signalling pathways. For example, genes DE following TNFα treatment were found to regulate pathways associated with: TNF- and death-receptor activation; apoptosis; and APRIL (A Proliferation-Inducing Ligand) signalling (Fig 4.5A).

Table 4.2 Gene Ontology analysis of genes with DE during priming with TNFα (TNF) and GM-CSF (GM). * represents which dataset GO-terms were found to be significantly enriched (compared to untreated control dataset).

GO Term	GO Category	GM	TNF
GO:0006954	Inflammatory response	*	*
GO:0009611	Response to wounding	*	*
GO:0006955	Immune response	*	*
GO:0042981	Regulation of apoptosis	*	*
GO:0006952	Defense response		*
GO:0006935	Chemotaxis		*
GO:0043122	Regulation of I-kappaB kinase/NF-kappaB cascade		*
GO:0007243	Protein kinase cascade		*
GO:0031328	Positive regulation of cellular biosynthetic process	*	
GO:0010557	Positive regulation of macromolecule biosynthetic process	*	
GO:0010628	Positive regulation of gene expression	*	
GO:0045321	Leukocyte activation	*	
GO:0010604	Positive regulation of macromolecule metabolic process	*	
GO:0001775	Cell activation	*	
GO:0045766	Positive regulation of angiogenesis	*	
GO:0051789	Response to protein stimulus	*	
GO:0008285	Negative regulation of cell proliferation	*	
GO:0051174	Regulation of phosphorus metabolic process	*	
GO:0019220	Regulation of phosphate metabolic process	*	
GO:0032570	Response to progesterone stimulus	*	
GO:0042325	Regulation of phosphorylation	*	
GO:0045429	Positive regulation of nitric oxide biosynthetic process	*	
GO:0006350	Transcription	*	
GO:0045893	Positive regulation of transcription, DNA-dependent	*	
GO:0051254	Positive regulation of RNA metabolic process	*	
GO:0045859	Regulation of protein kinase activity	*	

However, genes DE by GM-CSF treatment were found to regulate: p38 mitogen-activated protein kinase (MAPK) signalling; JAK/STAT signalling; and protein ubiquitination pathways (Fig 4.5B). The NF-κB pathway was found to be regulated by both GM-CSF and TNFα; however whereas TNFα positively regulated the NF-κB pathway, GM-CSF negatively regulated this (Fig 4.5C-D). IPA analysis reports signalling pathways whose components (genes) have been significantly up- or down-regulated by treatment. The software calculates the significance of pathway regulation based on the proportion of a canonical pathway which is enriched with DE genes. For example, if 18/24 genes in a signalling pathway are DE (versus control) then that pathway is more likely to be defined as significantly regulated than if only 2/24 genes are DE. However, the analysis does not distinguish if the signalling pathway has been (on the whole), up- or down-regulated. To further elucidate the specific regulation within a signalling pathway, the up- or down-regulation of each gene within a signalling pathway can be colour-coded and visualised to identify which portions of the pathway are regulated, and in what way. For example, genes which have increased expression (compared to control) are coloured red, genes whose expression is lower than in control are coloured green), genes with no change are shaded grey and those genes with no data are unshaded (white) (Fig 4.5C-D).

Having identified the NF-κB pathway as being significantly-regulated by both GM-CSF and TNFα, further analysis was carried out to determine if

similar genes within the pathway were regulated in a similar fashion. Fig 4.5C reveals that following TNFα priming, the majority of genes within the NF-κB pathway were upregulated (compared to untreated sample). Conversely, priming with GM-CSF led to a down-regulation of several genes within the NF-κB pathway (Fig 4.5D). This highlights an important difference between the molecular changes induced by both priming agents that would otherwise be overlooked if simply looking at signalling pathways that were significantly regulated.

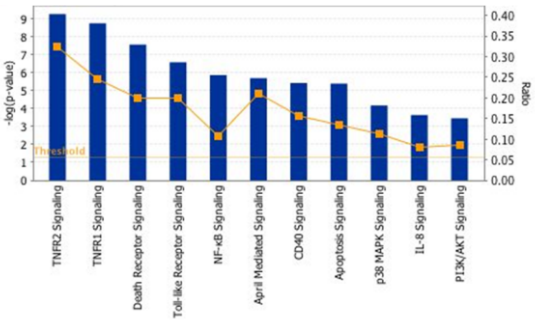
4.3.3.6 Analysis of genes that are differentially-expressed between GM-CSF and TNFα

Cuffdiff analysis identified 580 genes that were DE between GM-CSF and TNFα treated neutrophils (Fig 4.4). These genes therefore represent those that reflect the greatest differences regulated by the two priming agents. Further analysis was carried out on these genes to identify their functional associations.

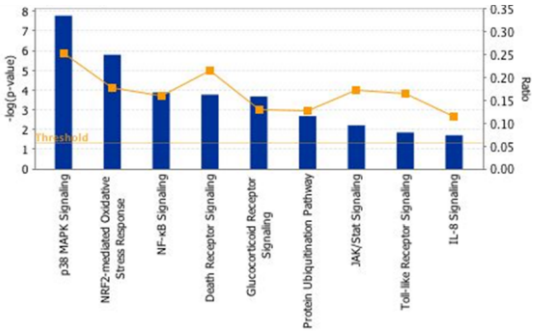
GO analysis revealed that 44 GO-terms were significantly enriched (see Appendix Table A.6 for full list of enriched GO-terms). Of the 44 GO-terms, 11 were directly related to cell death and/or apoptosis, including the most represented GO-term "Regulation of apoptosis" (containing 58 genes from the dataset). A similar result was obtained by IPA analysis of the 580 genes that identified "Apoptosis" as the cellular function with the greatest differential regulation between the two treatments ($p= 6.78^{E-23}$). Table 4.3 lists the RPKM values for each of the 58 genes found to be DE between

GM-CSF and TNFα. IPA analysis was carried out to identify upstream transcription factor activation. A total of 36/58 genes were more highly-expressed in TNFα treated samples, and of these, IPA analysis predicted that 23 were regulated by the NF-κB system ($p=5.77^{E-21}$) (Fig 4.5E). Conversely, a total of 22/58 genes had higher expression in GM-CSF samples, of which, 15 where predicted by IPA analysis to be regulated by the STAT family of transcription factors ($p=2.73^{E-12}$), in particular STAT3 and STAT5 (Fig 4.5E).

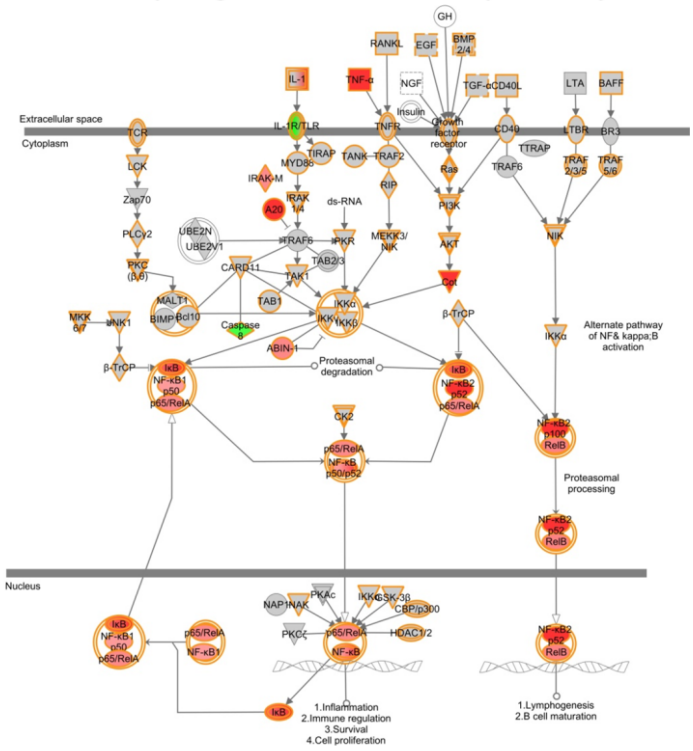
A Pathways regulated by TNF-α priming



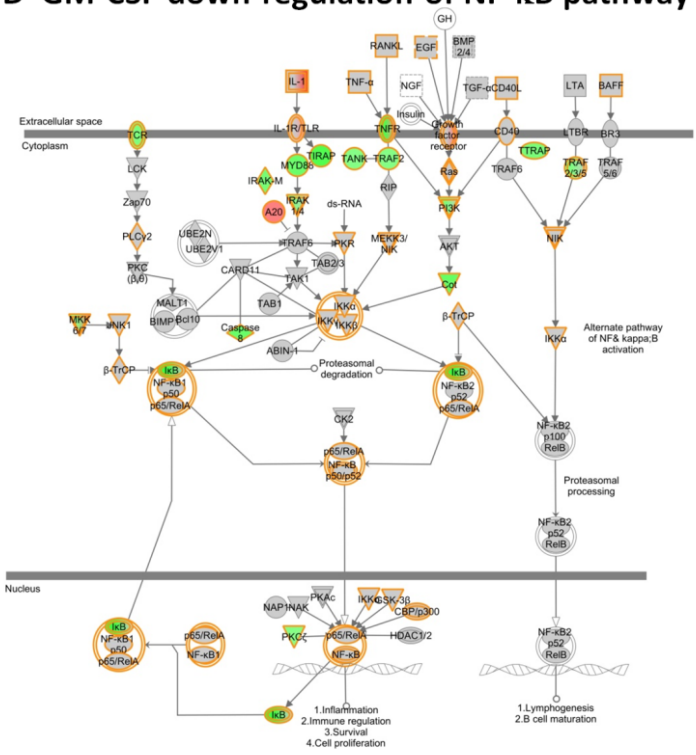
B Pathways regulated by GM-CSF priming



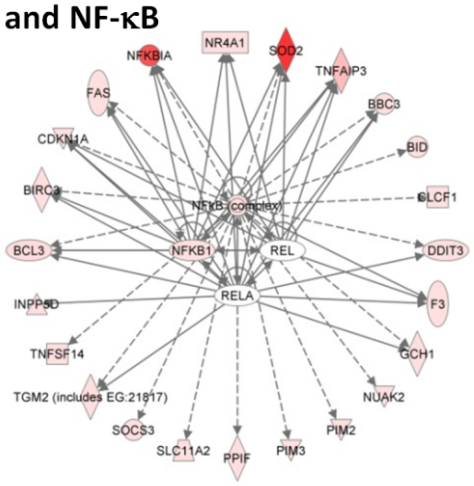
C TNF-α up-regulation of NF-κB pathway



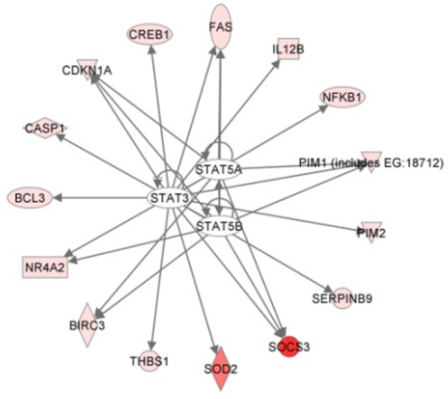
D GM-CSF down-regulation of NF-κB pathway



E Apoptosis regulating genes activated by TNF-α and NF-κB



F Apoptosis regulating genes activated by GM-CSF and STAT



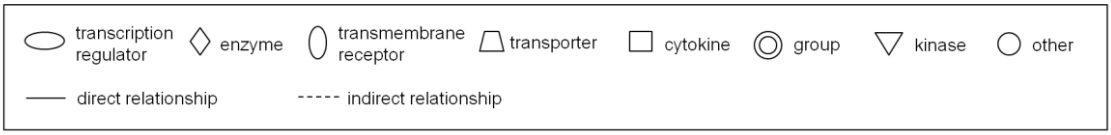


Fig 4.5 Analysis of signalling pathways in TNFα and GM-CSF primed neutrophils. (A-B) Bar graphs show the canonical pathways with the most significant regulation in (A) TNFα- or (B) GM-CSF- treated neutrophils compared to untreated control. Bars represent the probability (p-value) that the enrichment of significant genes within the canonical pathway is due to chance alone. Orange line represents the ratio of number of genes that are enriched in the pathway to the total number of genes in the pathway. (C-D) NF-κB pathway was identified as being significantly regulated in both (C) TNFα- or (D) GM-CSF- treated neutrophils compared to control; genes significantly up-regulated are coloured in red, down-regulated in green and genes with no significant change are shaded in grey. Genes where no data is available are have orange boxes. (E-F) IPA analysis of 58 apoptosis-regulating genes with significant DE between TNFα and GM-CSF treated neutrophils. (E) NF-κB activation was predicted in TNFα- treated neutrophils (p=9.04E-11), whereas STAT activation (F) was predicted in GM-CSF-treated neutrophils (p=2.26E-05). The RPKM values of individual genes are represented by increasing intensity of red.

Table 4.3 Gene expression value (RPKM) of 58 apoptosis-related genes significantly DE in neutrophils treated with GM-CSF or TNFα. RPKM values calculated by Cufflinks and significance of DE calculated by Cuffdiff (q<0.05 5%FDR).

Gene	GM-CSF	TNFα
ANXA1	231.73	82.28
APAF1	20.50	24.50
BBC3	3.58	15.43
BCL3	323.05	764.80
BID	66.72	192.62
BIRC3	14.36	107.60
CARD16	68.22	114.51
CARD6	12.14	2.73
CASP1	65.68	87.50
CDKN1A	44.49	16.55
CDKN2C	0.37	1.09
CHST11	61.81	79.18
CLCF1	0.95	6.73
CREB1	8.22	13.50
DDIT3	240.01	109.21
F3	0.80	2.98
FAS	33.88	62.78
GCH1	12.90	31.89
GHRL	6.28	12.84
HSPD1	20.84	8.29
ID3	2.01	<0.3
INPP5D	68.32	88.53
MAEA	27.03	57.98
NET1	5.76	0.60

NFKB1	21.19	42.05
NFKBIA	225.87	3901.19
NLRP3	16.81	49.76
NR4A1	2.02	21.41
NR4A2	18.14	50.14
NUAK2	5.25	24.99
PIM1	98.05	11.26
PIM2	36.66	260.17
PIM3	74.97	246.79
PLAGL2	4.81	33.72
PPIF	1352.77	488.80
PRNP	17.15	4.24
PROK2	582.65	210.88
PSEN1	39.99	69.94
RIPK2	92.35	36.63
RRM2B	7.70	16.15
SERPINB9	13.36	37.09
SLC11A2	0.78	1.82
SMPD2	2.04	9.13
SOCS2	23.61	2.96
SOCS3	4060.71	36.45
SOD2	2696.87	4619.77
SQSTM1	114.89	280.27
TGM2	4.84	0.68
THBS1	13.16	52.91
TICAM1	9.98	47.41
TNFAIP3	244.30	1451.80
TNFRSF10D	12.54	1.09
TNFSF14	41.61	85.44
TNFSF15	3.59	0.77
TNFSF8	19.93	5.93
TPT1	3025.26	1310.06
UTP11L	5.02	1.39
ZAK	1.40	0.51

4.3.4 Regulation of neutrophil apoptosis by GM-CSF and TNFα via activation of different transcription factors

The above bioinformatic analyses revealed that whilst priming of neutrophils with either GM-CSF or TNFα results in expression of apoptosis-regulating genes, they do so via different signalling pathways and by activation of different transcription factors. Further functional assays were employed to validate these findings, as detailed below.

4.3.4.1 Levels of apoptosis in neutrophils following addition of inhibitors of signalling pathways

Neutrophils were incubated overnight (18 h) with GM-CSF (5 ng/mL) or TNFα (10 ng/mL) in the presence of chemical inhibitors of NF-κB (wedeloactone, 50 μM) and JAK/STAT (JAK inhibitor-1, 10 μM). Levels of apoptosis were measured following annexin-V/PI staining by flow cytometry (see Methods section 2.2.5). Levels of apoptosis in GM-CSF and TNFα treated neutrophils were in-line with previously described experiments (see section 4.3.1.2) (untreated 60.35% ± 6.98%; GM-CSF 36.96% ± 6.56%; TNFα 50.07% ± 5.89%). Inhibition of NF-κB by wedeloactone abrogated the anti-apoptotic effect of TNFα ($p<0.05$, Student's t-test), but had no effect on GM-CSF-delayed apoptosis. Conversely, inhibition of STAT using JAK inhibitor-1 abrogated GM-CSF-delayed apoptosis ($P<0.05$, Student's t-test), and only partially attenuated TNFα-delayed apoptosis (although this was not significant $p>0.05$) (Fig 4.6A).

4.3.4.2 Western blot analysis of neutrophils following addition of signalling inhibitors

Next, activation-states of NF-κB and JAK/STAT pathways were determined by western blotting for phospho-STAT3, phospho-NF-κB and IκB-α. Neutrophils were pre-incubated for 1 h with wedeloactone or JAK inhibitor-1 before addition of GM-CSF (5 ng/mL) or TNFα (10 ng/mL). Following 15 min incubation, protein lysates were made using boiling Laemmli buffer containing a phosphatase inhibitor cocktail (see Methods section 2.2.6). Activation of NF-κB and degradation of IκB-α was seen in TNFα- treated samples, which was abrogated by wedeloactone but not by JAK inhibitor-1. Conversely, GM-CSF treatment did not activate NF-κB but was able to activate STAT3, which was inhibited by JAK inhibitor-1 (Fig 4.6B). These data confirm the predictions made by the bioinformatic analyses: regulation of neutrophil apoptosis by GM-CSF and TNFα is achieved via differential activation of transcription factors.

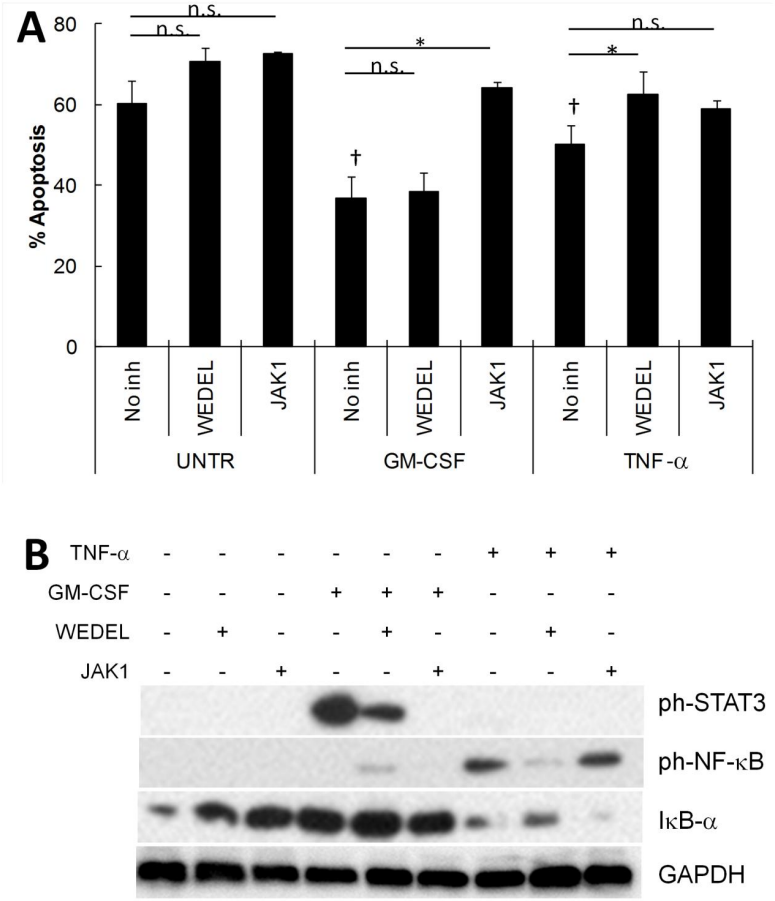


Fig 4.6 Delayed neutrophil apoptosis in GM-CSF- and TNFα- treated neutrophils is regulated by different transcription factors. (A) Overnight (18 h) incubation of neutrophils with GM-CSF or TNFα significantly delayed apoptosis († $p<0.05$, Student's t-test) compared to untreated (UNTR). Inhibition of STAT signalling with JAK inhibitor-1 (JAK1, 10 μM) abrogated the effect of GM-CSF on neutrophil apoptosis (*$p<0.05$) but did not affect TNFα-delayed apoptosis. Inhibition of NF-κB with wedelolactone (WEDEL, 50 μM) abrogated the effect of TNFα (*$p<0.05$), but not GM-CSF, on neutrophil apoptosis. (B) Western blot of NF-κB and STAT3 activation in GM-CSF- and TNFα- treated neutrophils. TNFα induced rapid phosphorylation of NF-κB and degradation of IκBα, which was inhibited by wedelolactone. GM-CSF did not induce phosphorylation of NF-κB or degradation of IκBα, but did induce STAT3 phosphorylation which was inhibited by JAK inhibitor-1. TNFα did not activate STAT3 in neutrophils.

4.4 Discussion

Priming is an important process to regulate neutrophil activation. Appropriate priming ensures that neutrophils arrive at the site of inflammation suitably prepared for their anti-microbial roles. Priming also ensures that neutrophils do not become non-specifically activated at non-inflammatory sites, such as in the bloodstream. The rapid changes induced by priming are well established, but the longer term molecular changes, such as *de novo* gene expression, are poorly defined. Moreover, the specific effects of individual priming agents on gene expression in neutrophils is limited to a few studies [181,182,221,222], and no studies have directly compared the effects of two cytokines using RNA-Seq.

This chapter set out to characterise the molecular changes induced by two inflammatory cytokines, GM-CSF and TNFα. Although both cytokines are known neutrophil-priming agents, and are elevated during inflammation and in inflammatory disease, little was known of any differences that may exist during priming with these cytokines.

Neutrophils were treated with priming concentrations of GM-CSF and TNFα and analysed by RNA-Seq to quantify the global changes in gene expression induced during neutrophil priming. Functional experiments using GM-CSF and TNFα revealed that both agents were able to rapidly prime the respiratory burst and delay apoptosis over the course of 18 h. While GM-CSF was a stronger inducer of these anti-apoptotic effects, the

effects of both priming agents were significantly different when compared to unprimed controls ($p<0.01$).

RNA-Seq and bioinformatic analysis of primed and unprimed neutrophils revealed that approximately half of the human transcriptome was transcribed under the conditions analysed, with the majority of genes being transcribed in all 3 conditions. However, several hundred genes were found to be expressed uniquely in each of the 3 conditions. Analysis of the most highly expressed genes, identified transcripts associated with several functional categories such as cell surface receptors, adhesion molecules and cytokines/chemokines. This suggests that although a large proportion of the human transcriptome is actively transcribed in both primed and unprimed neutrophils, the levels of expression of the majority of these genes are very low and most likely do not play an important role in neutrophil function under the conditions analyses. The identification of specific genes that are only expressed under certain conditions, (the majority of which had expression levels (RPKM) >10), suggest that not only is the neutrophil transcriptome dynamically-regulated, but that the global gene expression profile of primed neutrophils is regulated by the initial priming agent. It is therefore likely that the subsequent phenotype of a primed neutrophil is governed by the priming agent. Importantly, among the most highly-expressed genes in both GM-CSF and TNFα-treated neutrophils were cytokine/chemokine genes. For example, TNFα treatment led to a more than 10-fold increase in expression of TNF, CCL3 CCL4 and

CXCL2. Whereas GM-CSF treatment resulted in a >10-fold expression of CXCL1 and OSM, both treatments upregulated expression of IL1A, IL1B, and IL1RN. Thus, cytokine/chemokine production appears to be differentially-expressed by these two inflammatory cytokines. These findings have important implications for inflammatory disease where high levels of inflammatory cytokines are common, such as TNFα in RA [223] or psoriasis [90].

It is well established that both GM-CSF and TNFα delay neutrophil apoptosis [178,183,184,224,225]. However, RNA-Seq analysis of neutrophils has revealed that the genes regulating apoptosis are DE during priming with these two cytokines. Pathway analysis of 58 apoptosis-related genes found to be DE between cytokine treatments predicted differential activation of two independent transcription factors families. NF-κB was predicted to be activated following TNFα treatment, whereas STAT activation was predicted following GM-CSF treatment.

RNA-Seq followed by bioinformatic analysis is a powerful way to characterise the molecular changes induced by different stimuli or inflammatory conditions. The scale of data produced and the comprehensive databases of canonical biological processes allow accurate predictions of functional mechanisms from the expression profile of associated genes. Nevertheless, where possible, bioinformatic predictions should always be verified by functional assays.

Bioinformatic predictions of differential activation of transcription factors in neutrophils was verified using functional assays measuring neutrophil apoptosis and pathway activation in the presence and absence of specific inhibitors of signalling pathways. Inhibition of NF-κB abrogated the anti-apoptotic effect of TNFα and inhibited TNFα-induced phosphorylation of NF-κB. Conversely, STAT inhibition abrogated the anti-apoptotic effect of GM-CSF and inhibited the GM-CSF-induced phosphorylation of STAT3. Neither wedeloactone nor JAK inhibitor-1 had any significant effects on GM-CSF or TNFα-induced function, respectively. Whilst activation of NF-κB by TNFα and STAT by GM-CSF has previously been shown in neutrophils, their involvement in cytokine-delayed apoptosis is less well studied.

During inflammation, it is likely that neutrophils are exposed to a variety of cytokines at sites of inflammation (including both GM-CSF and TNFα). Therefore, for apoptotic pathways at least, there is a level of redundancy of processes capable of maintaining neutrophil survival. These discrete signalling pathways may allow the effects of each cytokine to act additively to increase the anti-apoptotic effect in situations where multiple cytokines are present.

Among the genes with highest expression following treatment with either cytokine, were several genes associated with suppression or inhibition of signalling. For example, TNFα treatment induced expression of NFkBIA, NFKBIE and TNFAIP3 – inhibitors of NF-κB signalling. Similarly, GM-CSF

increased expression of gene associated with suppression of STAT signalling (CISH and SOCS3). This suggests that in addition to activating different transcription families, priming by GM-CSF and TNFα also induces expression of inhibitors of these transcription factors. This mechanism thus provides a way of controlling the signalling by provision of a negative feedback loop.

In summary, these data suggest that the nature of the initial priming agent during neutrophil priming is crucial in determining the subsequent phenotype and functional capacity of the primed and/or activated neutrophil. The molecular differences that exist between neutrophils primed by different agents have successfully been characterised using a systems-based approach of RNA-Seq technology and current bioinformatic software, and have successfully been validated *post hoc* by traditional laboratory functional assays.

Chapter 5: Activation of gene expression by pro-inflammatory cytokines

Bioinformatic analyses presented within this chapter (section 5.4.2.5) were included in a publication in which I was second author.

Wright, H.L., Thomas, H.B., Edwards, S.W., Moots, R.J. (2015) Interferon gene expression signature in RA neutrophils predicts response to anti-TNF therapy. *Rheumatology*, Vol 54 (1): 188-193

5.1 Introduction

In the previous chapter, neutrophil gene expression was analysed by RNA-Seq following stimulation with two inflammatory cytokines capable of priming neutrophils (GM-CSF and TNFα). Whilst these cytokines are often found at the site of inflammation at high concentrations, neutrophils are unlikely to be exposed to these cytokines alone during an inflammatory response. Epithelial cells and other surrounding activated leukocytes express and release a variety of inflammatory signals during inflammation, each playing a role in regulating and propagating the immune response [5]. Details of how inflammatory signals affect neutrophil function, either individually or in combination are poorly-defined. Indeed, for studies using human neutrophils, there is still contention around whether neutrophils can respond directly to certain cytokines such as IL-6 [206,226] and IL-17 [227]. Thus, a greater understanding of the molecular consequences of stimulation by different inflammatory mediators will provide a means to identify the genes

and proteins most commonly-associated with a particular stimulus (or group of stimuli).

In this chapter, neutrophils were treated with cytokines/chemokines commonly-associated with either sterile or pathogen-driven inflammation, and changes in gene expression were analysed using the bioinformatic pipeline previously described (Chapter 3). Neutrophils were stimulated for 1 h with one of 5 cytokine/chemokines (G-CSF, IFNγ, IL-1β, IL-8, IL-6). In addition, neutrophils were stimulated with GM-CSF and TNFα (dual-treated) to analyse the effects of multiple cytokines compared to activation by these agents used singly.

5.2 Aims

The aims of this chapter were:

1. To utilise a previously-described pipeline of methods and bioinformatic techniques (Chapter 3) to quantify the neutrophil transcriptome following stimulation with a variety of inflammatory mediators (G-CSF, IFNγ, IL-1β, IL-8, IL-6, GM-CSF+TNFα).

2. To identify treatment-specific changes in gene expression and activation of signalling pathways /transcription factors.

3. To analyse whether dual stimulation of neutrophils by GM-CSF and TNFα induces expression of a discrete set of genes when compared to individually treated neutrophils (i.e. GM-CSF alone, TNFα-alone).

4. To validate any treatment-specific findings by functional assays.

5.3 Methods

Neutrophils were stimulated for 1 h in the absence or presence of one or more cytokine/chemokines (Table 5.1). The optimal concentration of each cytokine was previously determined using a range of functional assays (data not shown).

Table 5.1 List of cytokine/chemokine treatments and concentrations, and number of biological repeats per sample.

Inflammatory stimulus	Concentration	No. of biological replicates
G-CSF	10 ng/mL	3
IFNγ	10 ng/mL	3
IL-1β	10 ng/mL	3
IL-8	100 ng/mL	3
GM-CSF+TNFα	5 ng/mL + 10 ng/mL	3
IL-6	10 ng/mL	1

Following 1 h incubation, RNA was extracted and purified as previously described (Section 2.2.9). RNA was sequenced on the Illumina platform (50bp, single-end reads) and analysed using the methods described in Chapter 3. For a list of RNA concentrations, integrity values and number of reads produced per sample library, see Appendix Table A.5.

5.4 Results

5.4.1 Effects of cytokines on neutrophil apoptosis

Activation or priming of neutrophils often results in a delay in neutrophil apoptosis. This mechanism allows neutrophils to carry out their anti-microbial role in inflammatory situations. Consequently, several inflammatory mediators are known to extend the lifespan of neutrophils. Neutrophils were incubated overnight (21 h) in the presence or absence of several inflammatory stimuli (either alone or in combination). Levels of neutrophil apoptosis were analysed by annexin V/PI staining and measurement by flow cytometry (as described in section 2.2.5) (Fig 5.1).

Levels of apoptosis varied considerably between stimuli compared to controls (UT, 69.04%). As previously reported in this thesis, GM-CSF (41.6% $p<0.0001$) and TNFα (56.4% $p<0.05$) significantly delayed neutrophil apoptosis. Similarly, stimulation with IFNγ (55.3% $p<0.01$), G-CSF (34.1% $p<0.001$), IL-8 (55.2% $p<0.01$) or dual stimulation with GM-CSF and TNFα (45.2% $p<0.01$), all significantly delayed neutrophil apoptosis. Conversely, stimulation by IL-1β (66.12% $p>0.05$) and IL-6 (66.81% $p>0.05$) had no effect on neutrophil apoptosis. This suggests that different cytokines/chemokines greatly affect neutrophil function *in vitro* and implies that this difference in phenotype may be explained by the *de novo* gene expression profile induced by each stimulus.

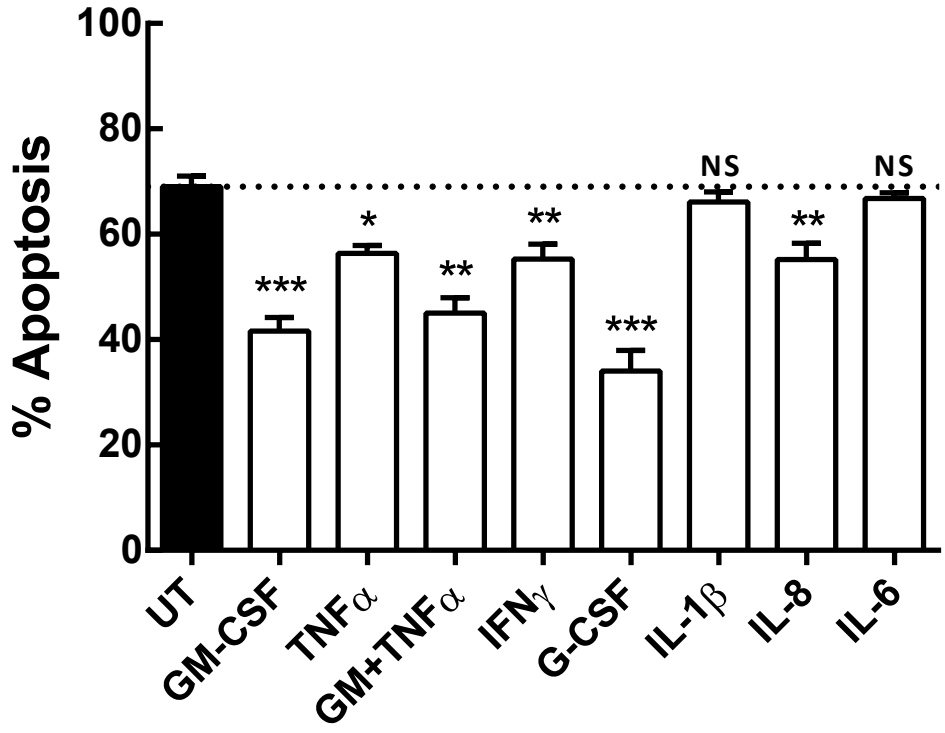


Fig 5.1 Neutrophil apoptosis following overnight (21 h) incubation with several inflammatory cytokines/chemokines. Apoptosis was quantified by annexin-V/PI staining and measured by flow cytometry. N=4. Error bars represent SEM. Significance (* p<0.001, ** p<0.01, * p<0.05, NS= not significant, p>0.05) calculated by Student's t-test versus untreated neutrophils (UT, black bar). Dotted horizontal line represents level of apoptosis in untreated samples.**

5.4.2 Bioinformatic analysis of neutrophils following incubation with inflammatory stimuli

Having identified that levels of neutrophil apoptosis vary depending on the specific inflammatory stimuli, neutrophils were next analysed by RNA-Seq and bioinformatic pipeline software to identify any changes in gene expression between stimuli.

5.4.2.1 Multidimensional scaling of cytokine-induced gene expression profiles.

Following quantification of neutrophil transcriptomes by Cufflinks, the downstream R-based software cummeRbund [168] was utilised to produce multi-dimensional scaling (MDS) plots of each dataset in relation to each other (Fig 5.2). MDS is a method for visualising the relative differences and similarities between samples in a 2-dimensional form. Thus, similar datasets (in terms of their RPKM values across all genes) will be spatially close to each other, whereas datasets with the greatest dissimilarity will be further apart in both spatial dimensions. Nine neutrophil datasets were plotted in 2-dimensional space following MDS (Fig 5.2).

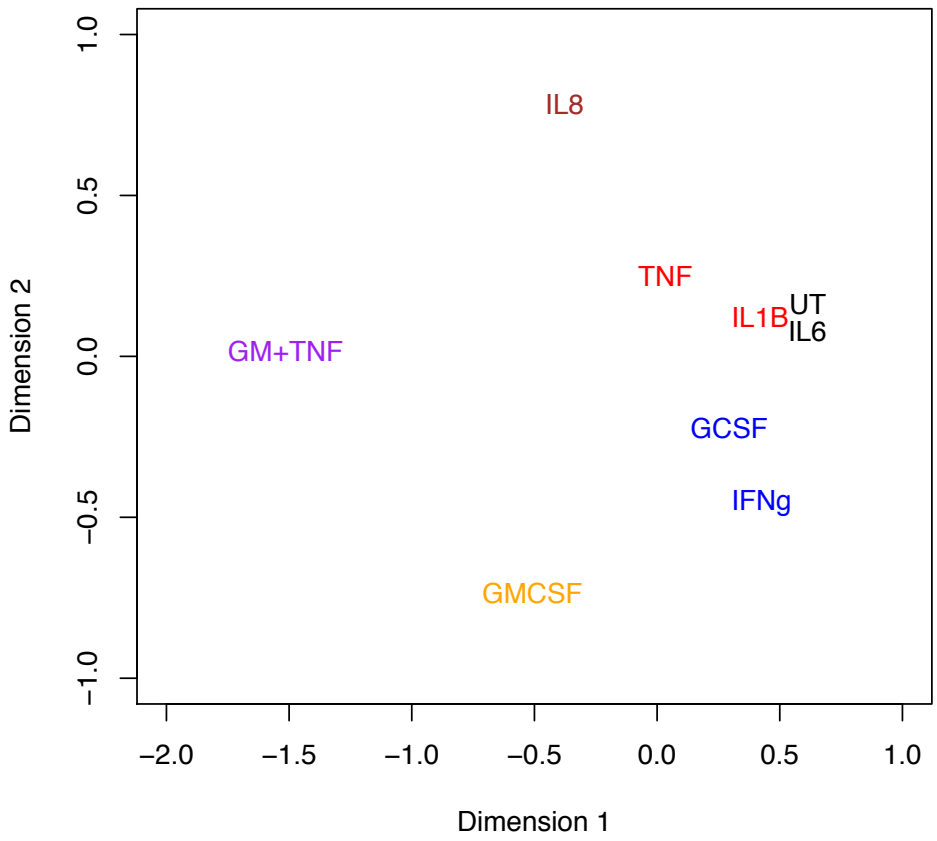


Fig 5.2 Multi-dimensional scaling plot showing the 2-dimensional association of 9 neutrophil transcriptomes under different inflammatory conditions: untreated (UT); IL-6 (IL6); IL-1β (IL1B); TNFα (TNF); G-CSF (GCSF); IFNγ (IFNg); IL-8 (IL8); GM-CSF (GMCSF); and GM-CSF+TNFα dual treatment (GM+TNF). MDS carried out by cummeRbund[168]. Colouring of datasets was arbitrarily applied *post hoc* to highlight dataset groupings.

Datasets most similar were untreated and IL-6, with TNFα and IL-1β forming a closely related pair. G-CSF and IFNγ datasets grouped together whilst IL-8, GM-CSF and GM-CSF+TNFα dual treatment were the most dissimilar data sets. Interestingly, despite GM-CSF and dual treatment GM-CSF+TNFα sharing a common stimulus, their respective datasets were among the two most dissimilar datasets. This suggests that dual stimulation induces expression of discrete gene sets or differential expression of genes that are not explained by a combination of GM-CSF and TNFα expression.

5.4.2.2 Analysis of neutrophil genes significantly DE genes following cytokine/chemokine treatment.

Cuffdiff analysis was used to calculate the number of significantly ($q<0.05$, 5% FDR) DE genes in each treatment compared to control (Table 5.2).

Table 5.2 Number of significantly DE genes (vs untreated). Significance calculated by Cuffdiff (q<0.05, FDR 5%).

Treatment	No. of DE genes
GM-CSF	110
TNFα	81
GM+TNFα	151
IFNγ	110
G-CSF	95
IL-1β	6
IL-8	5
IL-6	0

Similarly to the apoptosis results previously, analysis of DE genes identified a wide range in the number of DE genes following cytokine stimulation. IL-1B, IL-8 and IL-6 treatment resulted in only 6, 5 and 0 genes DE, respectively. Interestingly, these 3 treatments also resulted in the lowest inhibition of neutrophil apoptosis. Moreover, the stimuli that induced the greatest delay in apoptosis also resulted in the greatest number of DE genes. This suggests a correlation between apoptosis levels and number of genes DE. This was confirmed by comparing the number of DE genes with the percentage delay in apoptosis induced by each treatment ($r= 0.719$, $p<0.05$, Pearson correlation coefficient) (Fig 5.2).

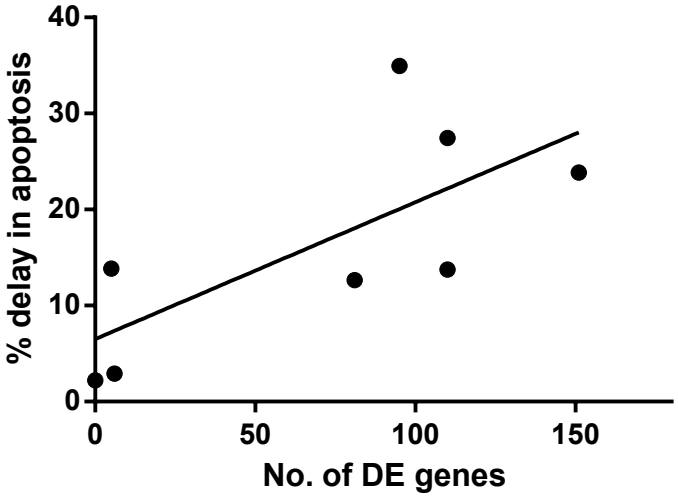


Fig 5.3 Linear regression of the number of DE genes in 8 conditions vs the % delay in apoptosis. Line represents best fit. r= 0.719, p<0.05, (Pearson correlation coefficient).

5.4.2.2 Analysis of gene and protein expression in dual stimulated neutrophils

The greatest number of DE genes was seen following dual stimulation with GM-CSF and TNFα, as might be predicted since previous results (Chapter 4) revealed that each cytokine activated independent transcription factors. Thus, the number of DE genes might represent a combination of both cytokines resulting in the expression of discrete sets of genes. However, further analysis of the genes DE following dual stimulation revealed that only 16 genes were DE in the dual treatment samples and both of the single treatments, with a further 79 genes expressed in dual treatment and one of the single treatments (Fig 5.4). Thus, 56 genes were uniquely DE after GM-CSF/TNFα dual stimulation. The top 30 genes (with the greatest difference in RPKM to untreated samples) of the 56 genes unique to dual treatment are listed in Table 5.3. Genes uniquely DE in dual-treated

neutrophils include several genes which are poorly characterised, for instance: Guanylate binding proteins (GBP) GBP1, GBP3, and GBP1P1; G-protein coupled receptor 132 (GPR132); the heme-binding protein Cytochrome B5 Domain Containing 1 (CYB5D1) and transmembrane coupled receptor 170B (TMEM170B). The gene with the greatest difference in RPKM to untreated is immediate early response-3 (IER3) which is an "anti-death" protein that protects from FASL- or TNFα-induced apoptosis and inhibits ERK dephosphorylation [228]. Conversely, dual treatment also induces expression of dual specificity phosphatase 5 (DUSP5), which is involved in negative regulation of the MAPK/ERK pathway [229].

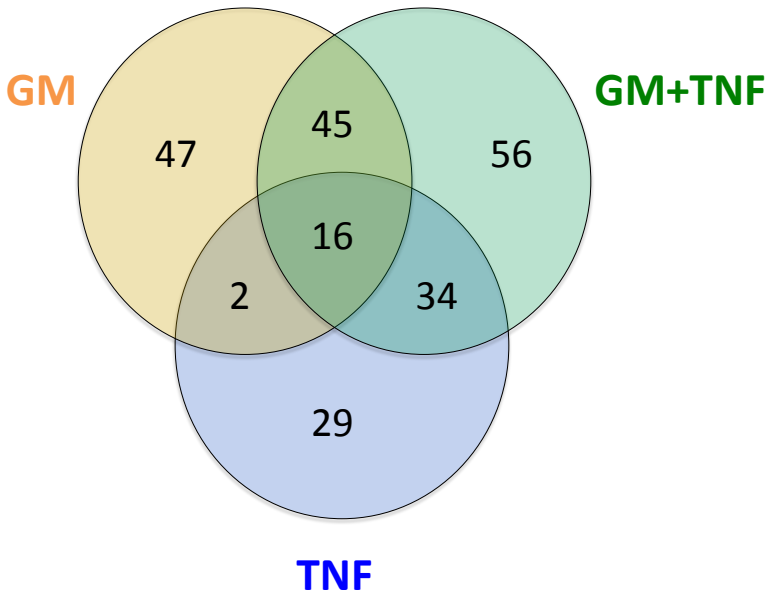


Fig 5.4 Venn diagram showing the number of DE genes in neutrophils following single, or dual stimulation with GM-CSF (GM) and/or TNFα (TNF). Significance calculated by Cuffdiff (q<0.05, 5%FDR).

Table 5.3 The 30 genes with the greatest difference in RPKM compared to untreated (UT) samples from genes that are uniquely DE in GM+TNFα treated samples.7

Gene name	UT	GM+TNF	ΔRPKM
IER3	108.47	1147.56	1039.09
GBP1	22.38	183.91	161.53
CYB5D1	7.55	106.26	98.72
GPR132	10.51	50.44	39.93
TMEM170B	28.01	3.51	-24.50
DUSP5	4.14	24.20	20.05
MRPS24	1.16	16.85	15.69
LOC100506229	0.72	15.97	15.25
GBP1P1	1.73	14.66	12.93
GBP3	2.65	15.12	12.47
LOC257358	1.75	14.17	12.42
KLHDC8B	15.07	2.76	-12.32
C5orf58	2.27	14.36	12.09
HELB	0.75	9.80	9.05
BATF2	1.40	10.36	8.96
TMEM54	0.39	7.96	7.58
PLEKHG2	1.08	8.29	7.22
APLP1	0.88	7.85	6.97
KIRREL2	0.04	6.63	6.58
CCDC85B	0.86	7.33	6.47
LINC00309	0.21	5.98	5.76
LDLR	0.77	6.30	5.53
SPHK1	0.24	5.29	5.05
CAHM	0.76	5.21	4.45
MTVR2	0.61	4.99	4.38
TKTL1	4.08	0.48	-3.61
ENO3	0.14	2.48	2.34

TGM2	0.18	2.48	2.30
PRDM7	0.12	2.28	2.16

Expression of IL-1β was DE in all 3 treatment conditions. However, levels of expression following dual treatment were almost double the combined expression of both individual treatment (GM-CSF 6410 RPKM, TNF 3614 RPKM, GM-CSF + TNFα 17,647) (Fig 5.5). To investigate whether this increased expression resulted in a corresponding high protein expression, expression of IL-1β was assessed by western blot over a 6h period (Fig 5.6). Levels of IL-1β were elevated in all 3 treatment conditions compared to control. Levels were highest in dual treated neutrophils, although the magnitude of the expression in dual treatment was approximately equal to the combined expression of each individual treatment, for example, at 2 h: GM-CSF = 106; TNF = 132; and GM-CSF + TNFα = 207 (arbitrary unit of band intensity, normalised to actin) (Fig 5.5). However, despite protein expression peaking at 2 h in all treatment conditions, individual treatment with either GM-CSF or TNFα resulted in a significant decrease in expression by 4 h which decreased further by 6 h. In contrast, dual treatment resulted in much lower decrease in protein expression, with levels of IL-1β protein being maintained over the 6 h time period. Differential protein expression dynamics were also seen when analysing expression of NFκBIA (also known as IκBA) in dual- and single-stimulated neutrophils (data not shown).

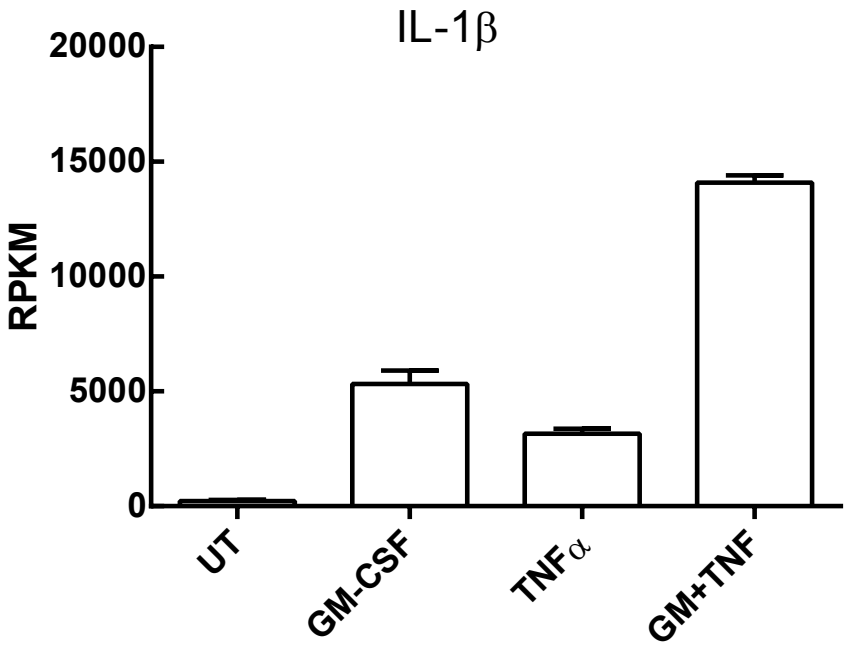


Fig 5.5 Neutrophil gene expression values (RPKM) for IL-1β following stimulation with GM-CSF, TNFα or dual treatment with GM-CSF and TNFα and measurement by RNA-Seq.

Taken together, these data suggest that dual treatment of neutrophils with GM-CSF and TNFα results in the differential expression of discrete genes not otherwise differentially expressed by individual treatments. However, these genes are poorly characterised and provide few clues to their effect on neutrophil phenotype. Additionally, dual stimulation of neutrophils results in different protein-expression dynamics over 6 h.

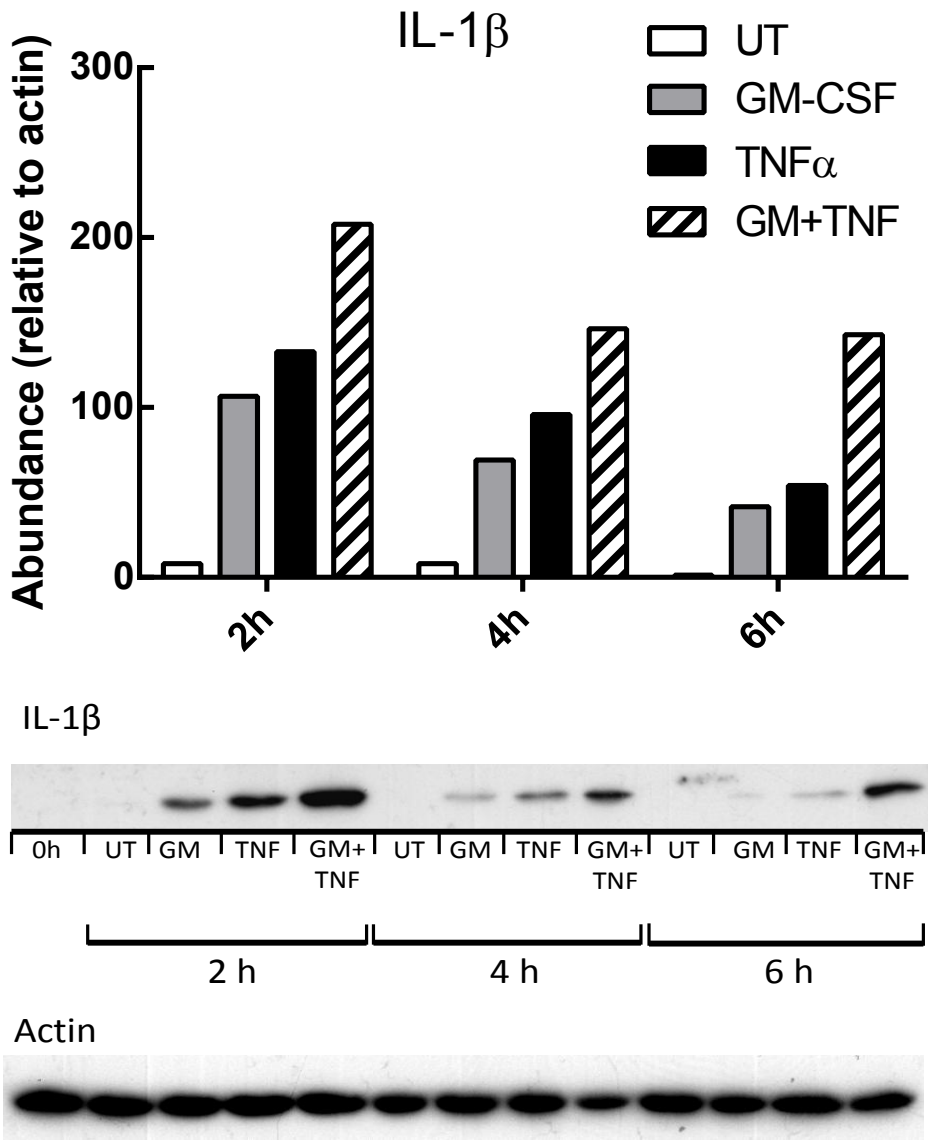


Fig 5.6 Western blot analysis of IL-1β in neutrophils following stimulation with: GM-CSF (grey bars); TNFα, (black bars); GM-CSF and TNFα (striped-bars); or no stimulation (white bars) and representative blot/housekeeping blot. Band intensity normalised to housekeeping control (Actin) N=2.

5.4.2.3 Analysis of signal pathway activation in cytokine/chemokine stimulated neutrophils

To assess whether the differential expression of genes by different cytokines/chemokines resulted from differential activation/regulation of signalling pathways, data was analysed by IPA to predict which upstream

regulators (transcription factors) regulated gene expression in neutrophils incubated under these different conditions.

In order to validate the IPA predictions, neutrophils were stimulated with each of the 9 conditions for 15 min and analysed by western blot for activation of several key signalling pathways: STAT (STAT1/3); NFκB (NFκB); PI3K (Akt); MAPK (ERK1/p38) (Fig 5.7). Results were then compared to those predicted by IPA (Table 5.4).

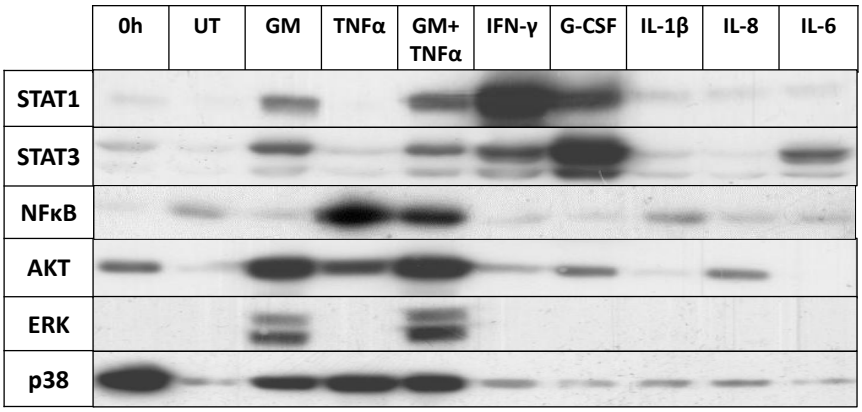


Fig 5.7 Western blot analysis of phospho-proteins in neutrophils treated with inflammatory cytokines/chemokines. Data representative of 3 separate experiments.

Western blot analysis identified differential activation of signalling pathways in each of the conditions analysed. For example, STAT activation was seen in GM-CSF, IFNγ and G-CSF stimulated samples, whereas ERK activation was only seen in GM-CSF stimulated samples (Fig 5.7). IPA analysis of upstream regulators confirmed the results of western blot analysis. Each of the signalling pathways identified by western blot were predicted to be active by IPA analysis ($p<0.05$) (Table 5.4). Interestingly, none of the signalling pathways analysed were activated by IL-1β, although IPA predicted several upstream regulators from the RNA-Seq data; these

include the zinc-finger proteins: CCCTC-binding factor (CTCF) ($p=1.58^{e-05}$) and GATA-binding protein (GATA4) ($p=1.58^{e-03}$). Taken together, these data highlight the predictive power of bioinformatic software and its ability to confirm and complement traditional western blot analysis of signalling pathway activation, which together have identified the differential activation of signalling pathways in neutrophils by different cytokines/chemokines.

Table 5.4 Signalling pathway activation states in neutrophils stimulated with a range of cytokines/chemokines. Table shows results of western blotting and corresponding p-value as calculated by IPA using RNA-Seq data to predict up-stream activation.

Cytokine/chemokine stimulus	Pathway activation (Western blot)	IPA prediction p-value
GM-CSF	ERK	6.52^{e-07}
	p38	4.85^{e-05}
	STAT3	2.31^{e-03}
	STAT1	1.49^{e-02}
	AKT	2.81^{e-02}
TNFα	NFκB	4.56^{e-09}
	AKT	3.59^{e-03}
	P38	7.75^{e-03}
GM+TNF	NFκB	9.29^{e-14}
	ERK	2.89^{e-07}
	STAT3	1.48^{e-05}
	P38	5.56^{e-05}
	AKT	9.66^{e-05}
	STAT1	3.59^{e-03}
IFNγ	STAT3	2.24^{e-04}
	STAT1	8.65^{e-04}
G-CSF	STAT3	3.57^{e-04}
	STAT1	4.07^{e-02}
	AKT	4.65^{e-02}
IL-1B	n/a	n/a
IL-8	AKT	1.42^{e-02}
IL-6	STAT3	2.33^{e-02}

5.4.2.4 Analysis of cytokine/chemokine expression in cytokine/chemokine stimulated neutrophils

In addition to differential activation of signalling pathways, neutrophils also exhibited differential expression of several cytokine/chemokines associated with the inflammatory response.

A gene list was compiled which included all genes defined as a "cytokine" or "chemokine" by their inclusion in the Gene Ontology groups "cytokine activity" (GO-term accession GO:0005125) or "chemokine activity" (GO-term accession GO:0008009). The RPKM values for each gene in all conditions were extracted from the Cufflinks output files and their fold change relative to untreated sample was calculated. The 25 genes with the highest expression were analysed by hierarchical clustering. A heatmap highlighting the differential expression of the cytokine/chemokine genes is shown in Fig 5.8. Samples treated with GM-CSF, TNFα, IL-1β or dual treatment with GM-CSF + TNFα showed the greatest expression of cytokines/chemokines and clustered together. Conversely IFNγ, G-CSF, IL-8 and IL-6 treatment had low expression of these genes, but had the highest expression of CXCL9. The only genes that showed high expression in all conditions (except IL-6) were IL-1B and IL-1RN. Interestingly, IL-1A was only upregulated by GM-CSF, TNFα and dual stimulation, whereas IL-1B was upregulated by all conditions, except IL-6. A table showing the raw RPKM values for each gene is shown in Appendix Table A.7.

These data highlight the ability of neutrophils to express different sets of cytokines/chemokines in response to different stimuli.

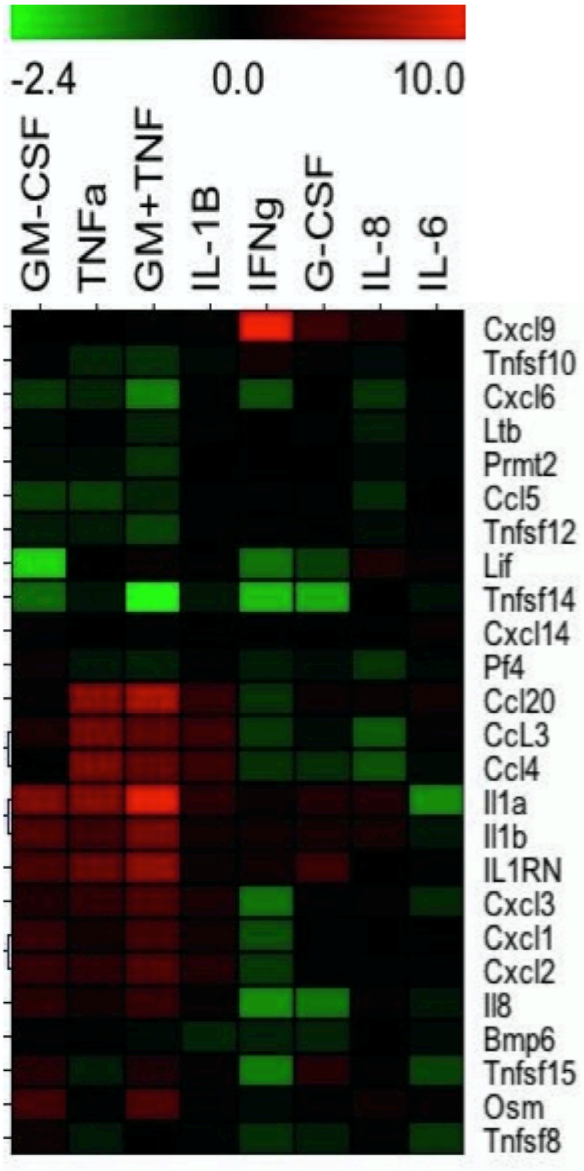


Fig 5.8 Heatmap showing hierarchical clustering of 25 cytokine/chemokine associated genes in neutrophils following stimulation with a range of cytokines/chemokines. Green= low, black = median, red high values (log2 fold change vs untreated samples). Raw values presented in Appendix Table A.7.

5.4.2.5 Modelling of in vitro gene expression data by comparison to ex vivo patient data

[All data in this section relating to patients were collected and analysed by Dr Helen Wright and reproduced here with permission]

An important benefit of producing large quantities of digital data such as those produced by an RNA-Seq experiment is the ability to easily compare and model the data against other datasets. This allows the identification of common features and gene sets within a large amount of data, which would otherwise be impractical and implausible by manual curation.

RNA-Seq analysis of *in vitro* stimulated samples provides a robust set of data that are highly specific to the treatment conditions and exhibit low variance between samples. Thus, gene expression profiles from *in vitro* stimulated samples can be modelled against less well defined datasets, such as those from patients with inflammatory disease, in an attempt to extrapolate meaningful similarities and patterns of association.

To this end *in vitro* stimulated neutrophil datasets were compared to neutrophil gene expression profiles of patients with RA (Fig 5.9).

Neutrophils from 20 RA patients and 6 healthy controls were analysed by RNA-Seq and gene expression profiles produced using the bioinformatic techniques described in Chapter 3. Gene expression analysis of patient data identified high expression of IFN-genes which correlated with the degree of response of patients to anti-TNF (TNFi) treatment (r=0.51 Pearson's correlation, p=0.02) [230]. Data for 59 interferon-response genes

from patient and healthy control datasets were then modelled against *in vitro* stimulated neutrophil datasets (for IFNγ and IFNα) and visualised by heatmap (Fig 5.9).

Hierarchical clustering of *in vitro* and *ex vivo* sample datasets identified that gene expression of the 59 interferon-response genes clustered into two main patient sub-groups: IFN-high (associated with INFγ signalling); and IFN-low (associated with IFNα signalling). Datasets relating to healthy controls clustered into a single group within the IFN-low cluster. IFNα is a type-I interferon which signals via the IFNα receptor complex, whilst IFNγ is a type-II which acts through the IFNγ receptor. These two IFN receptors activate different downstream Janus kinase (JAK)/STAT signalling pathways. For instance, IFN type-I signalling involves the activation of tyrosine kinase 2 (TYK2) and JAK1 leading to STAT1/STAT2 hetero-dimerisation and activation of interferon-stimulated response element (ISRE) transcription factor. In contrast, IFN type-II signalling involves JAK1/JAK2 activation leading to STAT1 homo-dimerisation ultimately resulting in activation of the transcription factor interferon-γ-activated site (GAS) [231]. Thus, the patient cohort can be further sub divided into two groups based on their hierarchical association with type-I and type-II gene expression profiles.

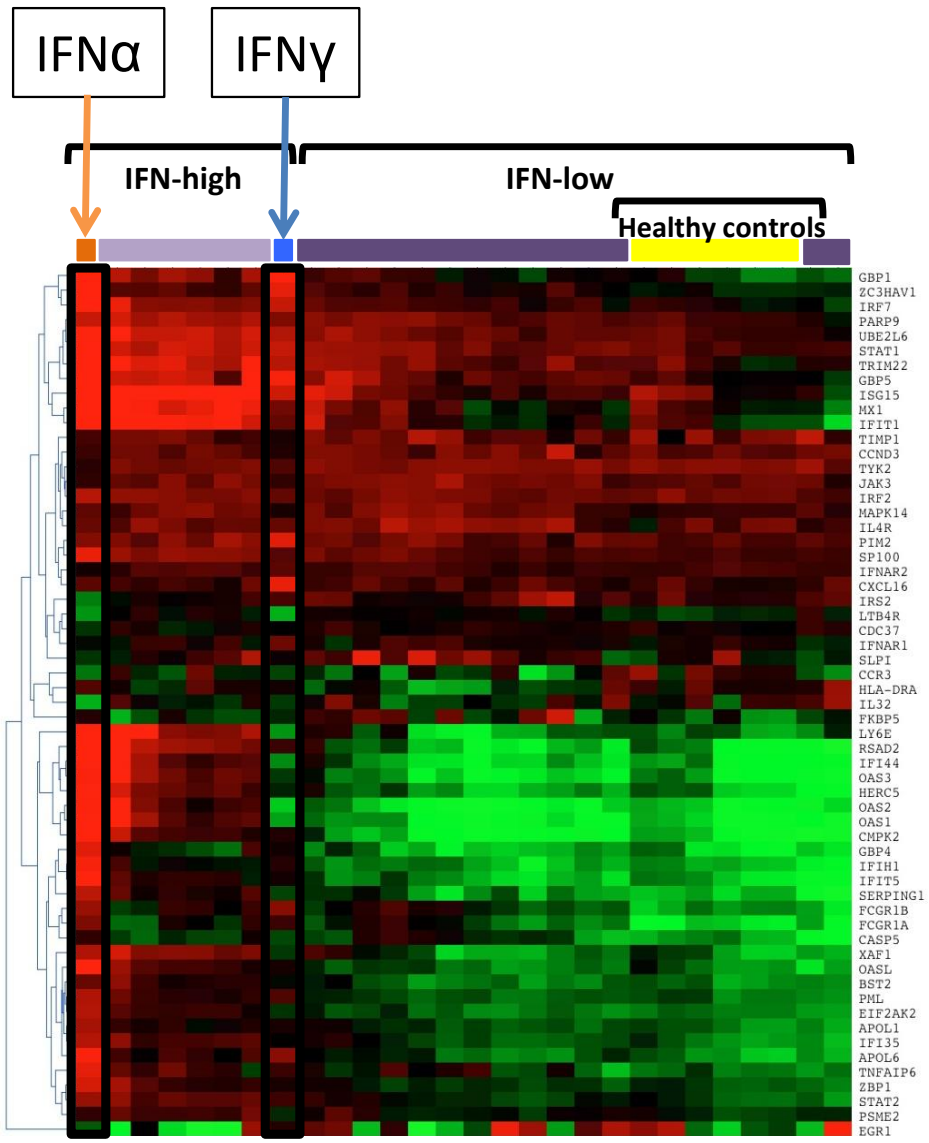


Fig 5.9 Heatmap showing the expression of 59 IFN-regulated genes in 20 RA (purple bars), 6 healthy control (yellow bar), IFNα-treated (orange bar) and IFNγ-treated (blue bar) neutrophils. (Expression level: green=low, black=median, red=high.) Patient data collected and analysed by Dr Helen Wright. Figure adapted from [230].

5.5 Discussion

The aim of this chapter was to measure the global changes in gene expression profiles of neutrophils stimulated with a variety of inflammatory cytokines. Neutrophils were stimulated for 1 h with a range of cytokines/chemokines commonly associated with inflammation and often found at sites of inflammation, those being: G-CSF; IFN γ ; IL-1β; IL-8; and IL-6. In addition, neutrophils were dual stimulated with GM-CSF+TNF α to analyse the effects of multiple cytokine stimulation on neutrophils by comparison to data collected in the previous chapter on neutrophils stimulated by single cytokines. Neutrophil RNA was subsequently sequenced by RNA-Seq and analysed using the previously-described bioinformatic pipeline.

Analysis of the effects of cytokine/chemokine stimulation on neutrophil apoptosis revealed that these agents acted significantly differently to each other. G-CSF had the strongest anti-apoptotic effect, whereas IL-6 and IL-1β had no effect on neutrophil apoptosis. Similarly, the number of DE genes induced by each cytokine correlated with the anti-apoptotic effect of each stimulus. Since neutrophils have a much shorter life span than other leukocytes, regulation of neutrophil apoptosis is an important biological process. This delay in neutrophil apoptosis may be critically linked to activation of gene expression. Cytokine activation of neutrophils leads to the enhanced expression of other chemokines and cytokines, as well as other key molecules such as adhesion molecules. Therefore, neutrophil

extended neutrophil lifespan is essential for these newly-expressed molecules to be expressed and affect the progress of an inflammatory response. These up-regulated genes may themselves be directly involved in apoptosis delay, although it must be pointed out that some regulation of apoptosis is achieved by stabilisation of anti-apoptotic proteins such as Mcl-1, without the requirement of *de novo* gene expression [232].

In the previous chapter, the effects of stimulation with either GM-CSF alone or TNFα alone on neutrophil gene expression profiles were evaluated. In this chapter, the effects of dual stimulation with both cytokines were investigated. Given that previous results showed independent activation of transcription factors and different gene sets by both GM-CSF and TNFα, it may have been predicted that dual stimulation would result in increased DE gene expression and that these genes would be representative of the gene lists seen from both single-stimulation conditions (i.e. additive effects). In addition, since the apoptosis pathway was found to be regulated differentially by both cytokines, it may also have been predicted that dual stimulation would further delay neutrophil apoptosis beyond that achieved by either single cytokine alone. However, neutrophil apoptosis following dual stimulation was lower than TNFα-stimulation levels but higher than GM-CSF-stimulation levels. Dual stimulation also affected the protein expression dynamics of IL-1β and NκFBIA, resulting in an increased magnitude and maintained expression over 6 h. interestingly, over one third of the DE genes following dual stimulation (56 genes from a total of

151 DE genes) were not DE after either individual treatment conditions. Although several of these genes are poorly characterised, in particular their role in neutrophil biology, this raises important questions as to the effects of multiple stimulants by inflammatory mediators on neutrophil function. This is of considerable importance when considering neutrophil function in disease, since elevation of multiple cytokines at inflammatory sites is a hallmark of several inflammatory or auto-immune conditions [3,77,78,211,233].

Analysis of signalling pathway activation revealed that neutrophil stimulation with different inflammatory cytokines/chemokines resulted in differential activation of a range of signalling pathways. Whilst the activation of these pathways by particular cytokines/chemokines is fairly well-documented [230,232,234,235], the ability to also predict such events using RNA-Seq data highlights the powerful predictive nature of RNA-Seq data. Additionally, it is a powerful way of revealing the expression level of each gene within a particular pathway, thus providing a way of estimating the contribution of each gene to the overall activation state of a specific pathway. For example, RNA-Seq data could identify if a particular signalling pathway is activated by up-regulation of cell surface receptors or through the increased actions of a transcription factor.

Analysis of cytokine/chemokine expression by neutrophils is less well defined. The importance of neutrophil-derived molecules during inflammation has long been overlooked in favour of cells from the adaptive immune response (B-cells, and T-cells). Indeed, much of the current

knowledge on neutrophil-derived products is based on studies from non-human species (often mice), such that their production by human neutrophils remains controversial [5,6]. Bioinformatic analysis of cytokine/chemokine genes revealed both similarities and differences in the expression profiles of neutrophils stimulated with different cytokines/chemokines. Perhaps unsurprisingly, conditions which had exhibited greatest ability to delay apoptosis also showed a similar pattern of cytokine expression, especially GM-CSF-stimulation, TNFα-stimulation and dual stimulation. However, several genes showed decreased expression following IFNγ compared to control, for instance, CXCL1, CCL3, CCL4 and CXC12, while IL-6 stimulated neutrophils showed no DE of cytokine/chemokine genes compared to control. Furthermore, of the 6 genes that were identified as having DE following IL-1β treatment, 3 relate to chemokines: CCL3; CCL4; and CCL4L1.

Although, not directly confirmed by data presented here, it is likely that the differential expression of cytokine/chemokine related mRNA under different inflammatory conditions would lead to differential protein translation and cytokine release by neutrophils. This highlights not only the ability of neutrophils to propagate and maintain the immune response by direct activation/signalling with surrounding cells by release of *de novo* inflammatory mediators, but it also reveals the importance of the nature of the stimulating signal to define the phenotype of activated neutrophils.

Among the most unexpected findings during this analysis was the effect of IL-6 on neutrophil function and gene expression. IL-6 is a crucial mediator of inflammation that is produced by several immune cells and can influence numerous cell types with multiple biological actions, such that it is often regarded as a pleiotropic cytokine [83]. It is often found at high concentrations at inflammatory sites such as synovial fluid [236]. Indeed, the importance of IL-6 to the inflammatory response is best highlighted by the success of therapeutic inhibition of IL-6R by Tocilizumab, which is used to treat conditions such as vasculitis, inflammatory bowel disease, RA and cancer [83]. Interestingly, IL-6 can behave as both a pro- and anti-inflammatory regulator, dependent on whether signalling occurs via the membrane bound IL-6R, known as *classical signalling*, or via *trans signalling*, involving soluble IL-6R (sIL-6) and the ubiquitously expressed gp130 receptor, respectively [237]. The effect of IL-6 on neutrophils is less well-defined. For example, previous studies on the effect of IL-6 stimulation on neutrophil apoptosis are conflicting, with anti-apoptotic, pro-apoptotic and no effect on apoptosis all previously reported [238–241]. Under the conditions studied here (10ng/ mL, for 21 h) IL-6 had no effect on neutrophil apoptosis. Moreover, IL-6 did not significantly differentially regulate the expression of any genes in neutrophils. However, the concentration of IL-6 used (10 ng/mL) was found to be biologically functional by the demonstration that it could induce the rapid phosphorylation of STAT3 (<15 min).

Several reasons could explain the lack of differential gene expression following IL-6 treatment, in spite of STAT3 activation. Firstly, the actions of IL-6 on immune cells may be limited to non-neutrophil cells, thus limiting the activation state of neutrophils at sites of inflammation where levels of IL-6 are elevated. Secondly, *de novo* gene expression may be delayed beyond 1 h, and any such changes would not have been measured by the current studies. Finally, full activation of neutrophils by IL-6 may require a secondary signal [242] or epigenetic changes [243] as has been demonstrated for IL-10 signalling in neutrophils.

The modelling of RNA-Seq data against other sources of biological data enables the identification of patterns of expression in genes that would otherwise be difficult to identify. This systems biology approach using RNA-Seq data from *in vitro* stimulated neutrophils to model against data from RA neutrophils has revealed that sub-populations of patients exist within a cohort of RA patients, relating to either a type-I or type-II IFN phenotype. Whilst modern predictive bioinformatic software is capable of identifying these changes based on patient data alone, the comparison with well-defined *in vitro* samples provides an additional level of data validation, ultimately increasing confidence of results.

In summary, analysis of neutrophils under different conditions of simulated inflammation (by cytokine stimulation) using RNA-Seq has revealed that neutrophils express discrete sets of genes in response to different stimuli. Additionally, stimulation by multiple cytokines induces expression of further

unique gene-sets. Analysis of the nature of the genes expressed reveals that several signalling pathways are differentially activated, which is confirmed by western blot analysis. Among the genes expressed are genes relating to cytokines/chemokines which show differential expression among treatment conditions. These data reveal the plasticity of neutrophils under conditions of inflammation and highlight the importance of surrounding signals on the developing phenotype of a neutrophil during different forms of activation.

Chapter 6: The effects of isolation method and purity on neutrophil gene expression and function

6.1 Introduction

In recent times there has been a growing appreciation of the ability of neutrophils to respond to, and influence, immune signalling between several different cell types, acting as a signalling bridge between the innate and adaptive immune systems [5]. Neutrophils are now known to activate monocytes, dendritic cells and T and B-cells, either through direct interactions or via secreted products [4]. But perhaps more importantly, neutrophils can respond to a variety of signals that affect their function and behaviour at the site of inflammation. For example, neutrophils can be activated by a range of stimuli such as, chemokines, ROS, DAMPs and PAMPs. In response, they release a variety of molecules such as pro- and anti-inflammatory cytokines, angiogenic factors and colony-stimulating factors [5]. Neutrophils can also modulate their function in response to different concentrations of apoptotic neutrophils. Contact with, or uptake of, apoptotic neutrophils, by neutrophils, can lead to inhibition of the respiratory burst and decreased release of the pro-inflammatory cytokines

TNFα and CXCL10, whilst also increasing their secretion of IL-8 and CXCL1 [68].

It is therefore clear that activated neutrophils do not merely localise to the site of infection, clear any pathological insult, and then rapidly undergo apoptosis: they are capable of more sophisticated and complex immune regulation. Indeed, they are central to the progression of the immune response and can shape the outcome of inflammation or infection in response to the signals they receive from the local environment [3].

Given the growing appreciation of intercellular signalling in neutrophil function and activity, it is perhaps somewhat paradoxical that the majority of research on human neutrophil activation and function has been achieved by *in vitro* studies, using purified blood neutrophil preparations. Although minimally-contaminated with other types of immune cells, the contribution of such contaminating cells to the overall assay output is often ignored. Moreover, whilst it is now appreciated that surrounding immune- and tissue-cells affect neutrophil function at the site of inflammation, *in vitro* studies of purified neutrophils attempt to remove the influence of such interacting cells.

An extra consideration when isolating neutrophils is their relative sensitivity to physical stimuli that may promote apoptosis in the absence of stimulating factors. Neutrophils are easily activated by shear forces or over-agitation, further complicating isolation methods. It is often of equal importance that the suspension of neutrophils resulting from a purification

protocol contains both high numbers of viable cells and low numbers of contaminating cells.

A variety of neutrophil isolation techniques are in current use, which generally achieve a final neutrophil purity of >95% and viability of >97%. Until recently, all isolation methods exploited the differing size and density of different blood cells to separate cells into distinct populations. These methods require multiple centrifugation steps and overall isolation times of 60-90 min. More recently, several new cell-sorting techniques, often offering higher sensitivity and specificity, have been developed. For example, Fluorescent Activated Cell Sorting (FACS) and Magnetic Activated Cell Sorting (MACS) techniques provide methods for sorting cells following attachment of a fluorescently-conjugated or magnetic particle-tagged antibody respectively, that bind to particular immune cell-specific surface markers. The use of surface markers to immuno-select mixtures of cells is often less perturbing and a more specific method of cell purification. Both the FACS and MACS approaches dramatically decrease the number of centrifugation steps required and achieve a higher final purity of cells, often in a similar time frame to density-gradient techniques. However, the higher purity afforded by these techniques usually comes at a greater financial cost and often a lower yield [244,245].

An additional concern when choosing an antibody-based isolation method is whether it utilises positive or negative selection. Neutrophils express high levels of receptors that bind immunoglobulins, and crosslinking of

these Ig-receptors can prime or even activate neutrophils [214]. Isolation of neutrophils by positive selection requires binding of specific immunoglobulins to neutrophil-specific cell surface molecules, such as CD16b. The major disadvantage of this approach is the increased risk of priming or activating neutrophils during the isolation method, thus inadvertently changing their phenotype and most importantly, their functional capacity [246].

Alternatively, negative selection approaches require a mixture of specific antibodies that recognise cell surface markers on other immune cell types, which can significantly increase the cost and efficiency of the methodology but leads to highly-pure, non-activated neutrophils. A summary of 3 commonly-used neutrophil isolation methods is detailed below.

6.1.1 Neutrophil Isolation methods

6.1.1.1 Dextran ficoll-paque isolation

Whole blood is mixed with dextran (average molecular mass 200-500 kDa), which causes the erythrocytes to aggregate and sediment more rapidly than other cell types (Fig 6.1A) [247]. Following sedimentation under 1 *g*, the leukocyte-rich upper phase is removed and subsequently layered onto Ficoll-paque. The Ficoll-paque provides a discontinuous density gradient which facilitates the separation of the granulocytes (of which neutrophils are the most abundant) from the lymphocytes (B-cells, T-cells, NK cells) and monocytes, following high speed centrifugation (500-*g* for 30 min). The

monocytes and lymphocytes form a distinct band at the plasma/Ficoll-paque interface, whereas the more dense granulocytes (neutrophils, eosinophils and basophils, and any residual erythrocytes) form a pellet at the bottom of the tube (Fig 6.1B) [248]. The remaining contaminating erythrocytes can subsequently be removed by hypotonic lysis.

6.1.1.2 Polymorphprep™ isolation

Polymorphprep™ is an endotoxin-free solution containing 13.8% (w/v) sodium diatrizoate and 8.0% (w/v) polysaccharide; it has a density of 1.113 ± 0.001 g/mL and an osmolality of 445 ± 15 mOsm [177]. Polymorphprep™ solution incorporates a similar experimental approach to neutrophil isolation as the dextran/Ficoll-paque method (density separation of lymphocytes from granulocytes and the sedimentation of erythrocytes) but does so in a single, one-step centrifugation process (Fig 6.1C). Whole blood is layered onto Polymorphprep™ solution and centrifuged at 500-*g* for 35 min. As erythrocytes pass through Polymorphprep™ the higher osmotic pressure of Polymorphprep™ results in water being lost from the erythrocytes, this in turn dilutes the Polymorphprep™, thereby reducing its density. This produces a continuous density gradient and results in the formation of two distinct bands; an upper band containing lymphocytes and monocytes, and a lower band containing granulocytes. Non-nucleated platelets remain in the upper phase and erythrocytes sediment towards the bottom of the tube. The PBMC band is carefully removed first to avoid cross-contamination with the neutrophil layer. The granulocyte layer is

subsequently carefully aspirated and washed. Any remaining erythrocytes in the granulocyte suspension are removed by hypotonic lysis.

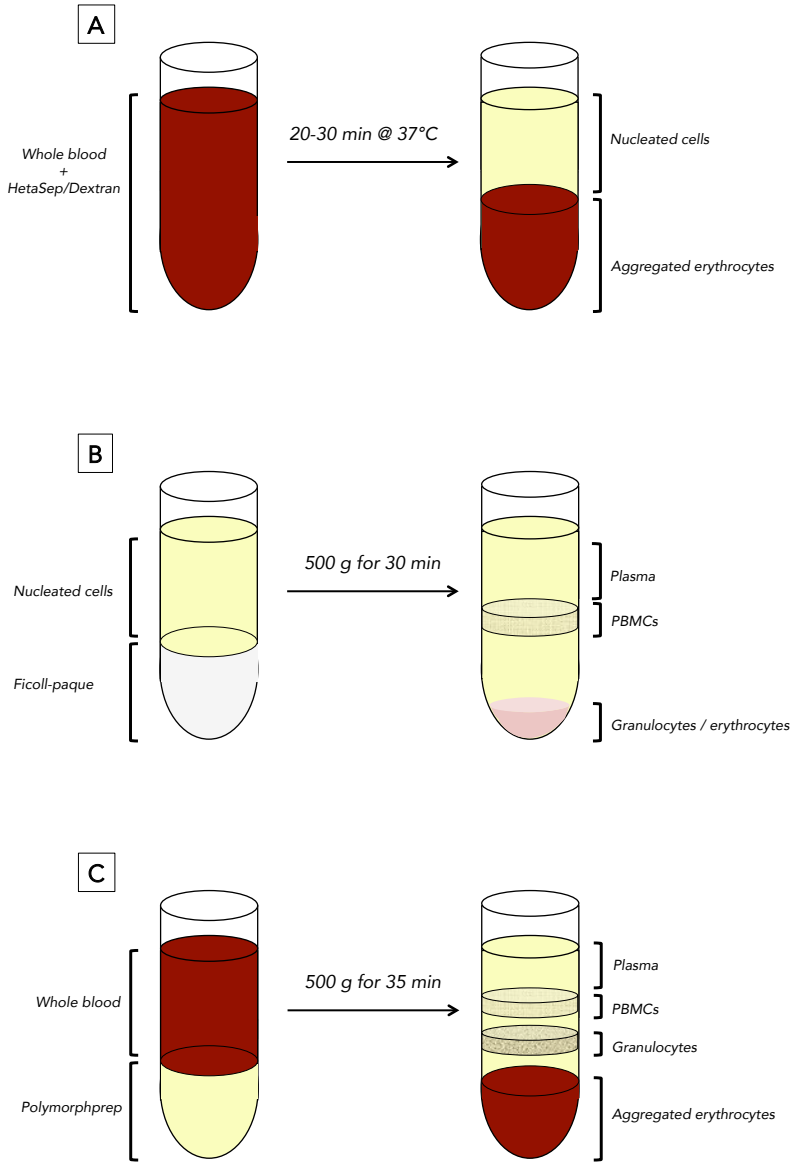


Fig 6.1 Treatment of whole blood prior to neutrophil isolation. A) Whole blood can be depleted of erythrocytes prior to a density gradient separation by the addition of dextran solution, causing erythrocytes to form aggregates and sediment. B) The remaining nucleated cell fraction is layered over Ficoll-paque and centrifuged, separating the granulocytes from the peripheral blood mononucleated cells (PBMCs). C) Alternatively, neutrophil isolation can be achieved by single-step centrifugation of whole blood over an equal volume of Polymorphprep™.

6.1.1.3 Neutrophil isolation by negative magnetic bead isolation

Isolation of neutrophils using magnetic beads is usually carried out using smaller volumes of suspension (0.5 – 2 mL) and begins with a suspension of nucleated cells at a concentration of ≤ 5 x 10^7/mL. This is obtained by sedimentation of erythrocytes from whole blood at 1-*g* using a polysaccharide solution (HetaSep™) which has similar properties to dextran. Cells are then incubated with a cocktail mix of tetra-meric monoclonal antibodies (Fig 6.2). One portion of the antibody complex has specificity for dextran whilst the other portion is specific for one of seven cell-specific markers that are not expressed on neutrophils but are expressed on erythrocytes and other leukocytes (Table 3.1).

Following incubation with this antibody cocktail, dextran-coated beads are incubated with cells before placing the cell suspension in a magnetic field. Antibody-bound cells are retained in the magnetic field whilst unbound neutrophils are decanted (Fig 6.3).

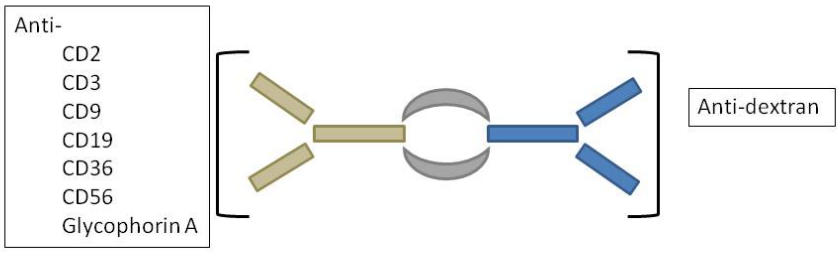


Fig 6.2 Structure and specificity of tetrameric monoclonal antibody complexes. Antibody complexes consist of two antibody molecules connected by a linker. Antibody-complexes possess dual specificity for dextran, and one of 7 different cell surface antigens not expressed on neutrophils.

Table 6.1 Specificity of antibodies recognising antigens on specific cell types, present in a neutrophil magnetic bead isolation kit (StemCell® Genoble, France).

Antibody antigen	Specificity
CD2	T-cells, NK cells
CD3	T-cells
CD9	Eosinophils, Basophils
CD19	B-cells
CD36	Monocytes
CD56	NK-cells
Glycophorin A	Erythrocytes
Dextran	Magnetic beads

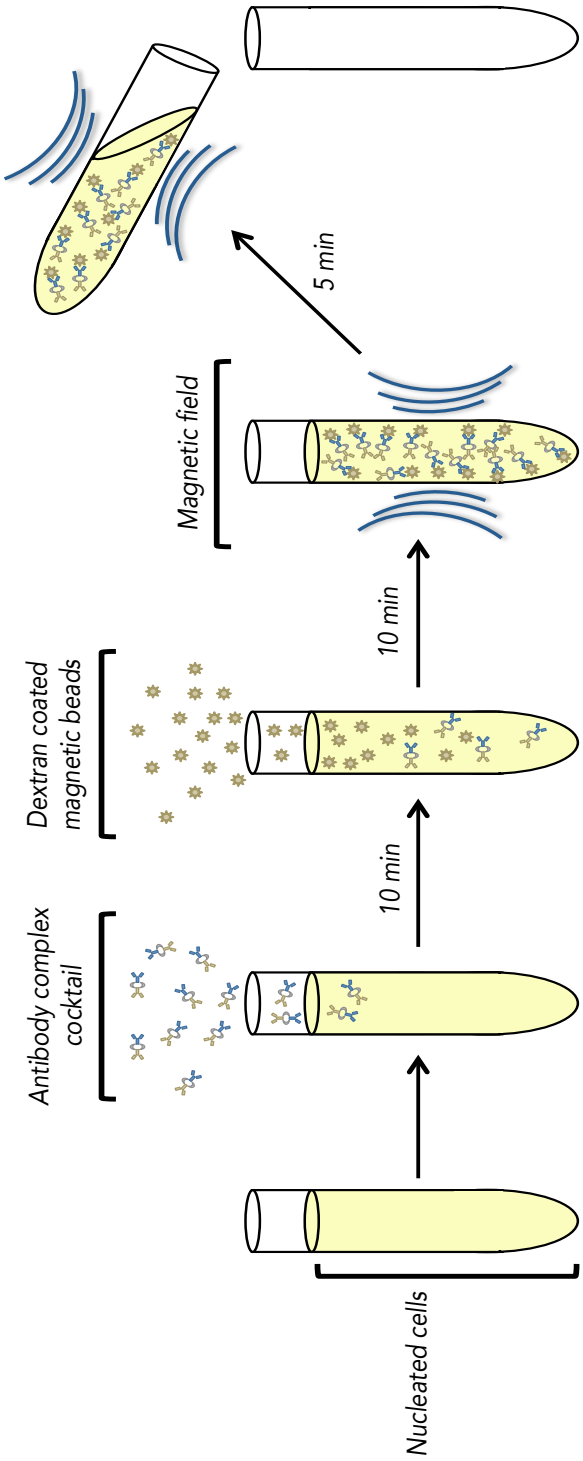


Fig 6.3 Summary of neutrophil isolation using magnetic beads. A mixture of cell-specific antibody complexes are added to a starting suspension of erythrocyte-depleted whole blood. Following incubation, dextran-coated magnetic beads are added and the suspension is placed in a magnetic field. Neutrophils are then decanted, whilst other leukocytes are retained in the tube within the magnetic field

6.1.2 Contaminating cells

The small percentage (1-10%) of contaminating cells often found in neutrophil suspensions following density-gradient isolation methods, is often considered to be an acceptable level of contamination that has minimal effects on the behaviour of neutrophils in the preparations. However, more recently, the impact of contaminating leukocytes in neutrophil suspensions has been questioned. For example, Sabroe and colleagues demonstrated that ultra-pure neutrophils (>99%) behaved differently than those analysed with 5% PBMC contamination, namely that the anti-apoptotic effect of LPS on neutrophils was significantly decreased in the absence of contaminating PBMCs [249]. Whilst it was suggested that monocytes in the PBMCs were responsible for these differences, it is often other granulocytes (in particular eosinophils) that constitute the largest proportion of non-neutrophil cells in most neutrophil preparations. Since neutrophils, eosinophils and basophils have very similar size and density, it is not possible to separate them from each other using approaches based on density-gradient media. Consequently, the level of contamination by other granulocytes is usually donor-dependent, and the overall contamination is also reliant on the technical dexterity of the researcher. It was recently shown that the percentage of contaminating cells in a typical Polymorphprep™ neutrophil isolation can vary between 1-17% across 18 individual blood isolations [177]. This potential for contamination is of particular concern for high-sensitivity experiments such as qPCR, mass

spectrometry or RNA-Seq. Indeed, it has been suggested that eosinophils have a far greater transcriptional capacity than neutrophils, confounding experiments that quantify mRNA in neutrophil suspensions containing high and variable numbers of eosinophils [250].

It is therefore clear that a greater understanding of the technical variations of different neutrophil isolation methods is required, both in terms of the effect of isolation methodologies on function and the contribution of contaminating cells on neutrophil function and gene expression.

6.2 Aims

The aims of this chapter were:

1. To compare the purity of neutrophils isolated by negative selection using antibody cocktails and magnetic beads ("ultra-pure neutrophils") with those isolated by density gradient-centrifugation using Polymorphprep™.

2. To quantify the differences in function and gene expression profiles of ultra-pure neutrophils compared to those of neutrophil preparations following Polymorphprep™ isolation, and evaluate the effects of contamination by non-neutrophil cells.

3. To quantify the changes in function and gene expression profiles of ultra-pure neutrophils and Polymorphprep™ isolated neutrophils following stimulation with inflammatory cytokines.

6.3 Methods

The RNA-Seq data in this chapter were collected using neutrophils from 2 healthy donors of similar age. The donors were chosen on the basis of their consistently high or low levels of eosinophils in neutrophil suspensions following Polymorphprep™ isolation. Neutrophils from these two donors were isolated by both Polymorphprep™ and negative magnetic bead isolation (StemCell®). For a full description of methods see section 2.2.2.

Purified neutrophils (5 x 10^6/mL) were incubated at 37 °C in RPMI media (+ 25 mM HEPES) and, where stated, treated with either 5 ng/mL of GM-CSF or 10 ng/mL of TNFα. Following 1h incubation, total RNA was extracted using Trizol/chloroform separation. RNA was further purified using a Qiagen on-column RNeasy cleanup-kit which included a DNAse digestion step (for more detailed methods see section 2.2.9).

Total RNA was enriched for poly-A mRNA and sequenced using the Illumina Hi-Seq 2000 platform on a single lane producing upwards of 40 million reads per sample (as previously described [251]). For a list of raw reads produced per sample see Appendix Table A.5. All sequencing protocols were carried out by BGI International following shipment of purified total RNA on dry ice.

Purity and integrity of RNA was determined prior to shipping and prior to sequencing using an Agilent bioanalyser (see Appendix, Table A.5 for full quality values and concentrations of all samples). All other methods were as previously described in Chapter 3.

6.4 Results

For these studies, two healthy donors were selected, based on the levels of non-neutrophil leukocytes that consistently contaminated neutrophil preparations obtained following isolation on Polymorphprep™: Donor 1 had consistently low levels of contaminating cells (~1-5%), whereas Donor 2 had consistently high levels of contaminating cells (~10-18% by cytospins). Both donors were otherwise healthy and of similar age. Neutrophil preparations were obtained by Polymorphprep™ and negative selection, using the StemCell® negative selection technique.

6.4.1 Quantification neutrophil purity by cytospin

The levels of contamination of neutrophil preparations following Polymorphprep™ isolation or negative magnetic bead isolation of blood from Donors 1 and 2 were quantified by cytospins and flow cytometry. Cytospins were prepared and the mean number of neutrophils, monocytes, lymphocytes and eosinophils were quantified in 4 separate fields of view (x20 magnification). A minimum of 100 cells was counted in each field (Fig 6.4).

Quantification confirmed previous observations of different levels of contamination in these preparations between the two donors. Similar levels of lymphocyte and monocyte contamination were seen in both donors (Table3.2). However, the greatest differences between donors were detected in the number of eosinophils following Polymophprep™ isolation,

with Donor 1 having 1% eosinophils whereas Donor 2 had ~15% eosinophils (Table 6.2). The overall purity of neutrophils was higher following bead isolation than with Polymorphprep™ (Donor 1; 96% Polymorphprep™, 97.5% beads, Donor 2; 83% Polymorphprep™, 99% beads). Morphological analysis of levels of cell apoptosis immediately after isolation revealed that neither isolation method resulted in significant levels of apoptotic cells (<1%, data not shown).

6.4.2 Quantification of neutrophil purity by flow cytometry

Eosinophils are difficult to distinguish from neutrophils using a coulter counter [252] or by analysis of their forward-scatter and side-scatter profiles using flow cytometry [253] due to their similar sizes and granularity. However, they can be distinguished from neutrophils by flow cytometry based on their levels of auto-fluorescence or cell surface expression levels of CD16 (FcγRIII). Eosinophils have a higher autofluorescence and express lower levels of CD16 on their cell surface then neutrophils.

Following isolation of neutrophils by density-gradient or magnetic bead isolation, suspensions were stained with FITC-conjugated anti-CD16 monoclonal antibody and analysed by flow cytometry. Cells were first gated on their forward- and side-scatter profiles so that subsequent measurements were made on granulocytic cells (see Appendix Fig A.6 for example of gating used). Gated cells were subsequently analysed by their forward-scatter and CD16 properties (Fig 6.5). Two distinct populations

were detected: cells which were $CD16^{bright}$; and a lower population of cells of similar size that were $CD16^{dim}$. These populations represent neutrophils and eosinophils, respectively (Fig 3.5). Quantification of the eosinophil population revealed that neutrophils isolated by density-gradient separation contained a higher proportion of eosinophils than neutrophils isolated by magnetic bead depletion. This analysis revealed that Donor 1 exhibited 3.5% contamination following density gradient isolation, compared to 2.42% by magnetic bead isolation, whereas Donor 2 had 12% eosinophil contamination following density gradient separation, compared to 2.5% following magnetic bead isolation. These levels of eosinophil contamination correlate well with the levels of purity as assessed by cytospins (Table 6.2).

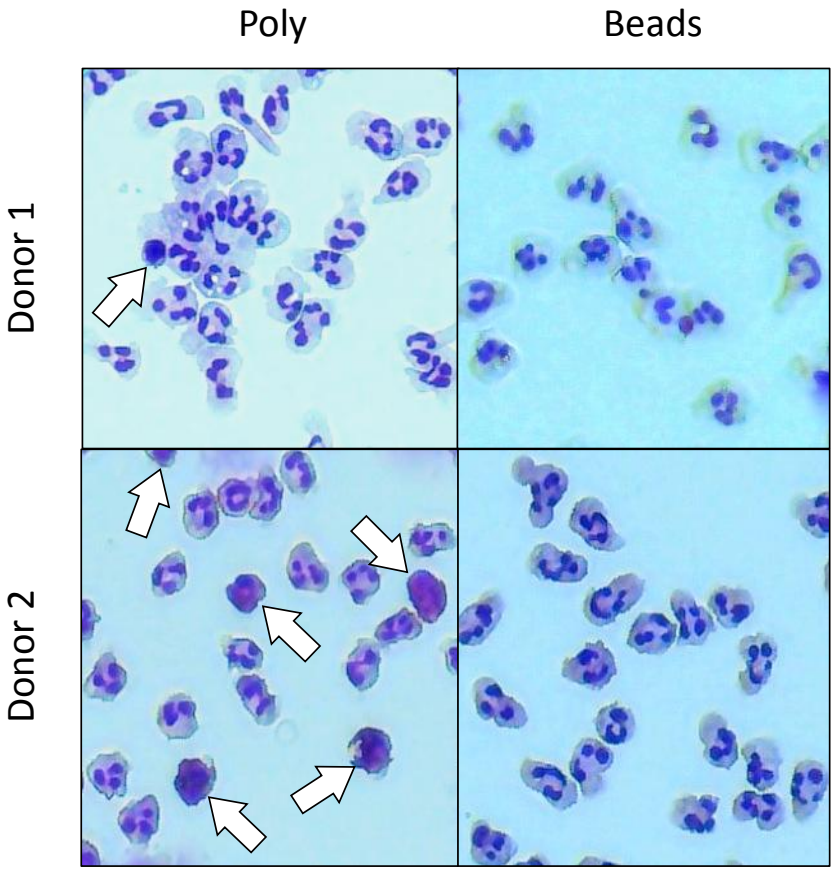


Fig 6.4 Representative cytospins of neutrophil preparations following Polymorphprep™ (left panels) or magnetic bead isolation (right panels) from Donor 1 (top) and Donor 2 (bottom). White arrows highlight non-neutrophil cells. All images at x20 magnification. Quantification of data is summarised in Table 6.2.

Table 6.2 Percentage of leukocytes in each preparation from each donor. Cells quantified by cell morphology and staining properties using cytospins (calculated from 4 separate fields of view, counting > 100 cells per field per donor).

Cell type	Donor 1		Donor 2	
	Isolation method			
	Poly	Beads	Poly	Beads
Neutrophil	96%	97.5%	83%	99%
Eosinophil	1%	<1%	15%	<1%
Monocytes /Lymphocytes	3%	2%	2%	<1%

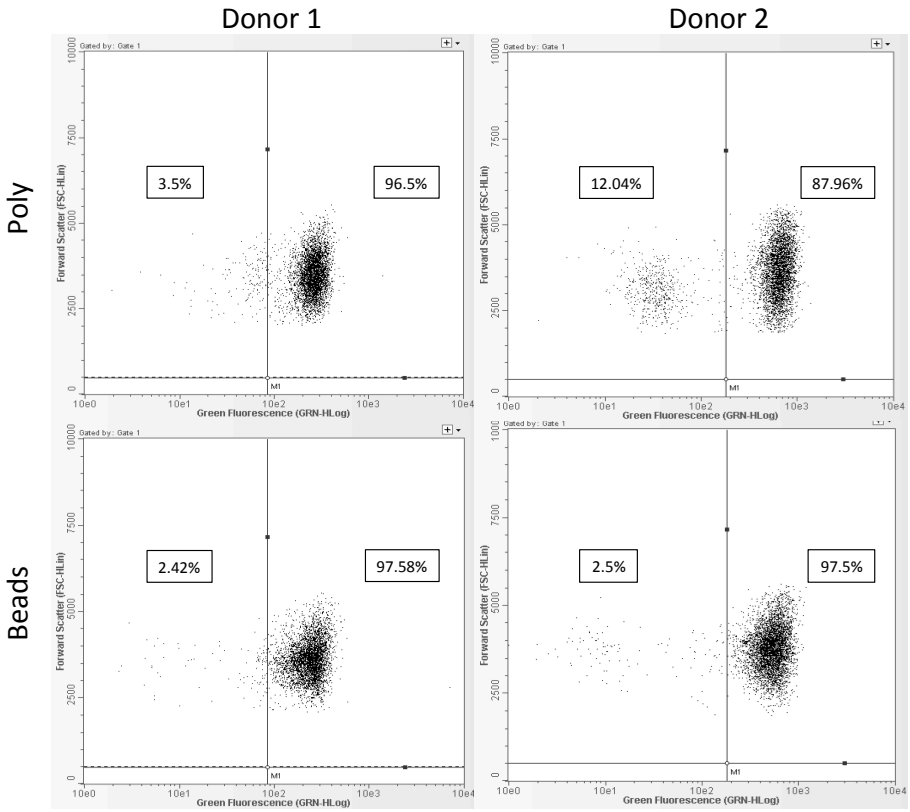


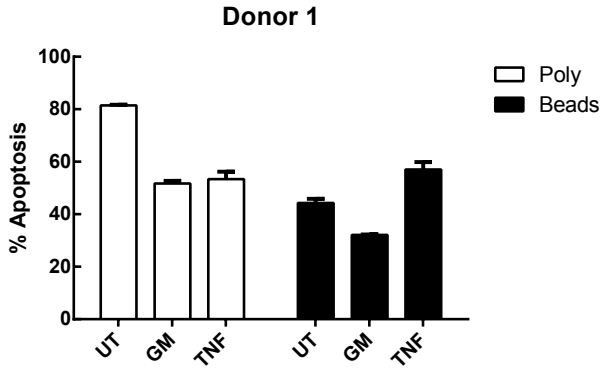
Fig 6.5 Flow cytometry scatterplots of neutrophil preparations by Polymorphprep™ (top panels) or magnetic bead (bottom panels) isolation protocols. Plotted by green fluorescence (CD16 positive, X-axis) and forward-scatter (Y-axis). Donor 1 (left panels) and Donor 2 (right panels). Numbers shown are percentage of cells in each of the two quadrants shown.

6.4.3 Effect of neutrophil isolation method and population purity on neutrophil apoptosis

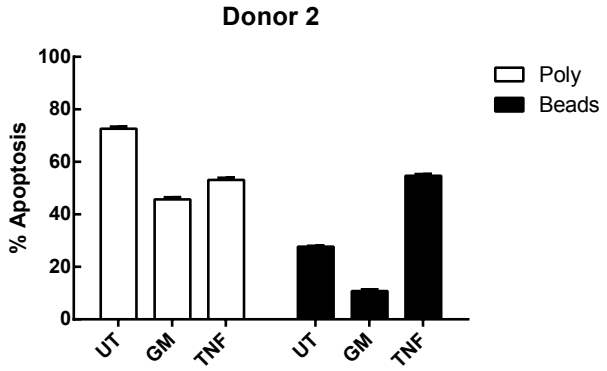
Contaminating cells in neutrophil preparations have been implicated in affecting the behaviour of neutrophils in response to stimulation [225,249,254]. Neutrophils prepared by both isolation methods were therefore incubated overnight (18 h) with (or without) GM-CSF (5 ng/mL) or TNFα (10 ng/mL). Levels of neutrophil apoptosis were subsequently measured by flow cytometric analysis of annexin V/PI staining (Fig 6.6). Mean levels of neutrophil apoptosis in suspensions from Donor 1 following Polymorphprep™ isolation were: untreated 81.5% ± 0.25%; GM-CSF 51.7% ± 1%; TNF 53.4% ± 2.8%, which are in line with previously published data [251]. Neutrophils from Donor 1 isolated by magnetic beads showed lower levels of apoptosis after GM-CSF treatment than control neutrophils: untreated 44.3% ± 1.5%; GM-CSF 32.1% ± 0.2%. However, neutrophils treated with TNFα exhibited a higher level of apoptosis than control neutrophils (TNF 57%, ± 2.9%), suggesting that under these conditions, TNFα is pro-apoptotic to neutrophils. Neutrophils from Donor 2 showed a similar pattern of apoptosis to Donor 1 for both the Polymorphprep™ isolated preparations (untreated 72.6% ± 0.62%, GM-CSF 45.7% ± 0.53%, TNF 53.1% ± 0.61%) and in the bead-isolated preparations (untreated 27.7% ±0.1%, GM-CSF 10.8% ± 0.5%, TNF 54.7% ± 0.5%), including an increase in apoptosis following TNFα treatment of bead-isolated neutrophils. Increasing the number of biological replicates to N=4 resulted

in values for treated samples reaching significance ($p < 0.05$) when compared to their paired, untreated sample (Fig 6.6). These data suggest that firstly, magnetic bead isolated neutrophils have lower levels of constitutive apoptosis than Polymorphprep™ isolated neutrophils, and secondly, that treatment of neutrophils with TNFα has opposing effects on apoptosis levels, dependent on neutrophil isolation method and/or preparation purity.

A



B



C

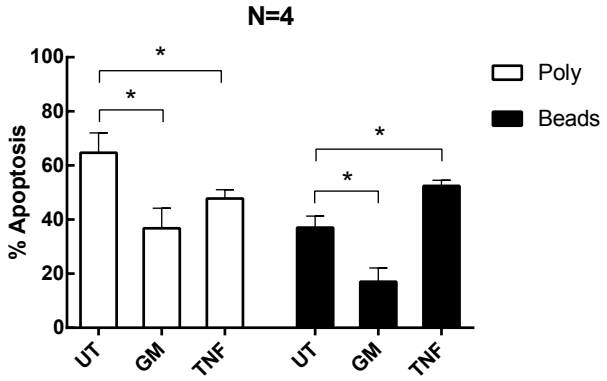


Fig 6.6 Levels of neutrophil apoptosis following overnight incubation (18 h) with inflammatory cytokines. Neutrophils isolated by Polymorphprep™ (Poly) or magnetic beads (Beads) were incubated overnight (18 h) in the presence (or absence (UT)) of inflammatory cytokines GM-CSF (GM) (5 ng/mL) or TNFα (TNF) (10 ng/mL). Levels of apoptosis were measured by annexin V/PI staining. For (A) Donor 1 and (B) Donor 2, bars represent mean value of duplicate measurements (± SD), In (C) bars represent mean value of 4 separate experiments from 4 donors, error-bars represent SEM. * Represents significance ($p<0.05$) as measured by a paired student's t-test.

6.4.4 Cell surface markers of neutrophils prepared by different methods

Appropriate cell surface markers can inform on the relative purity of cell preparations and the activation state of the cells. To assess levels of neutrophil purity and activation following Polymorphprep™ or bead isolation, freshly-isolated neutrophils were stained with FITC-conjugated-monoclonal antibodies for CD16 (FcγIII), CD15, CD11b (ITGAM) and CD64 (FcγRI) and relative fluorescence measured by flow-cytometry (Fig 6.7). Levels of CD16, CD15 and CD11b were slightly lower in neutrophils isolated by magnetic bead isolation whilst levels of CD64 were slightly lower in Polymorphprep™ isolated neutrophils. However, levels were not significantly different between neutrophils isolated using different methods ($p>0.05$ paired student's t-test) suggesting that neutrophil isolation methods and preparation purity has only marginal effects on cell surface expression.

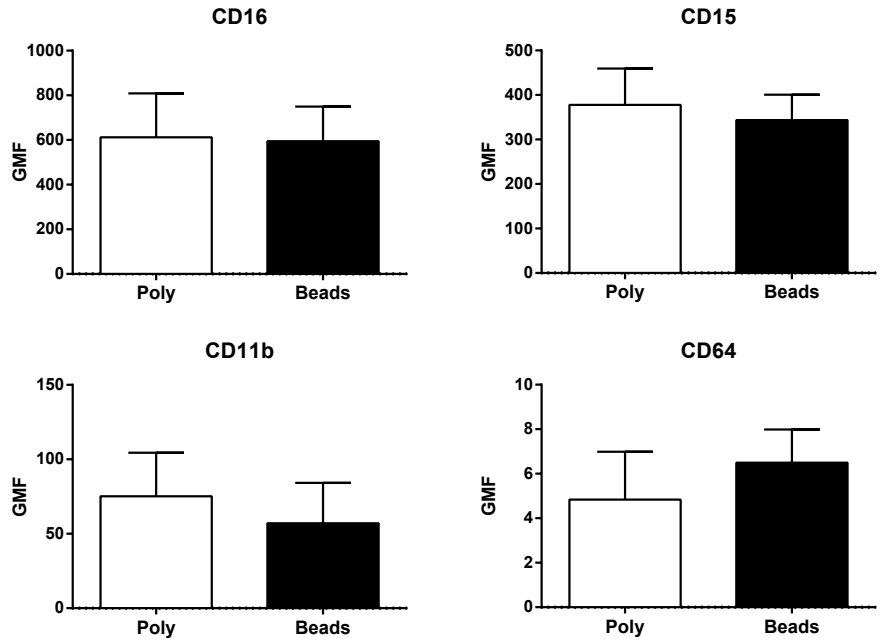


Fig 6.7 Levels of expression of cell surface markers in neutrophils isolated by Polymorphprep™ (Poly) or by magnetic beads (Beads). Geometric mean fluorescence (GMF) of CD16 (N=5), CD15 (N=3), CD11b (N=3) and CD64 (N=4) was measured by flow cytometry and normalised to an appropriate isotype control. Error bars represent SEM. No significant difference in expression found between isolation methods (Student's t-test p>0.2).

6.4.5 Neutrophil yield from whole blood

In addition to purity levels, different isolation methods result in different final yields of neutrophils from whole blood [244,245]. To quantify differences between Polymorphprep™- and magnetic bead- isolated neutrophils, a sample of whole blood from a healthy donor was divided into two and neutrophils were isolated by Polymorphprep™ and magnetic beads in parallel. Final preparations of neutrophils were counted using a coulter counter and the number of neutrophils recovered per mL of whole blood was calculated (Fig 6.8). This process was repeated for 5 different donors. Mean levels of neutrophils after magnetic bead isolation were approximately 40% of those recovered after Polymorphprep™ isolation

using blood from the same donor (Polymorphprep™ 2.81x10^6 /mL whole blood ± 1.02; Beads 1.14x10^6 /mL whole blood ± 0.76). This suggests that a large proportion of whole blood neutrophils are lost during a magnetic bead isolation, and that this proportion would otherwise be retained using a Polymorphprep™ preparation.

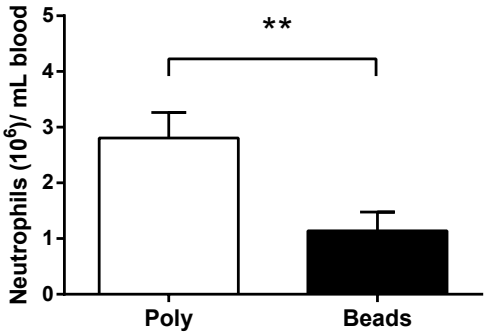


Fig 6.8 Comparison of neutrophil yield from whole blood following isolation by Polymorphprep™ (Poly) or magnetic beads (Beads). ** p<0.01 (paired student's t-test). Paired data from 5 independent experiments. Error bars represent SEM.

6.4.6 RNA-Seq analysis of neutrophils prepared by different methods

Having shown that different donors and different neutrophil isolation procedures generate suspensions containing different numbers of contaminating cells, notably eosinophils, it was then necessary to determine how these contaminating cells and isolation techniques contribute to transcriptome studies. Neutrophil preparations isolated by two different methods (Polymorphprep™ and Beads) from donors with high or low eosinophil contamination levels were incubated with (or without) inflammatory cytokines (GM-CSF or TNFα) for 1 h before total RNA was extracted and processed for high throughput sequencing by RNA-Seq (see Methods). For details of RNA integrity and number of raw reads per sample see Appendix Table A.5.

6.4.7 Transcript levels of antigens targeted by antibodies in the bead kit protocol

To quantify the efficiency of the magnetic bead isolation kit to deplete contaminating cells, transcripts for the cell-specific antigens utilised in the bead kit were analysed (transcripts for genes listed in Table 3.1). Of the 7 antigens, only transcripts for CD9, CD36, CD2, and CD3 were detected in any samples, and these are expressed in eosinophils, monocytes, T-cells/NK cells and T cells, respectively.

Transcripts for CD36, CD3 and CD2 were detected in all samples from both donors following Polymorphprep™ isolation, but were absent (or below the RPKM threshold of 0.3) from all samples isolated by magnetic beads (Fig 6.9). CD9 transcripts (specific to eosinophils) showed the highest levels of expression of all 7 genes, with the highest levels seen in the untreated sample from Donor 2 following Polymorphprep™ isolation (RPKM=28.9). Samples processed by magnetic bead isolation showed much lower levels of CD9 expression than those isolated by Polymorphprep™. However, unlike the other transcripts, levels of CD9 where not entirely absent in samples isolated by beads, and were at or around the 0.3 RPKM threshold (Fig 6.9). The high levels of CD9 in all samples from Donor 2 following Polymorphprep™ isolation confirmed the presence of significant levels of eosinophils, whilst the greatly decreased levels of CD9 transcripts in the samples isolated by magnetic beads confirmed the ability of the magnetic bead kit to effectively deplete eosinophils.

Treatment of neutrophils from each preparation type and from both donors with GM-CSF or TNFα had no significant effect on any bead kit antigen transcripts compared with their corresponding untreated sample. However, it is noteworthy that the largest variation in values between treatments (untreated, GM-CSF and TNFα) was consistently seen in Donor 2 samples following Polymorphprep™ isolation (Fig 6.9), that is, the preparation with the highest levels of non-neutrophil leukocytes.

6.4.8 Expression of other non-neutrophil transcripts

The levels of transcripts encoding each of the 7 bead kit antigen target antibodies correlate well with the overall levels of cellular contamination measured by cytospins and flow cytometry. However, whilst these gene products are known to be unique cell-surface markers on specific cell types, it is possible that they may also be transcribed in other leukocytes, but not translated and expressed. We therefore measured the expression levels of other transcripts, which, according to the literature, are only expressed in non-neutrophil leukocytes, in order to further elucidate the extent that contaminating cells contribute to the transcriptome profile of the neutrophil preparations. Figure 6.10 shows the RPKM levels for the T-and-B-cell specific transcripts CD8a and CD5 (respectively), the monocyte-specific transcripts Chemokine Ligand 2 (CCL2) and CD163, and the eosinophil transcripts, Interleukin 5 Receptor Alpha (IL5RA) and Charcot-Leyden crystal galectin (CLC).

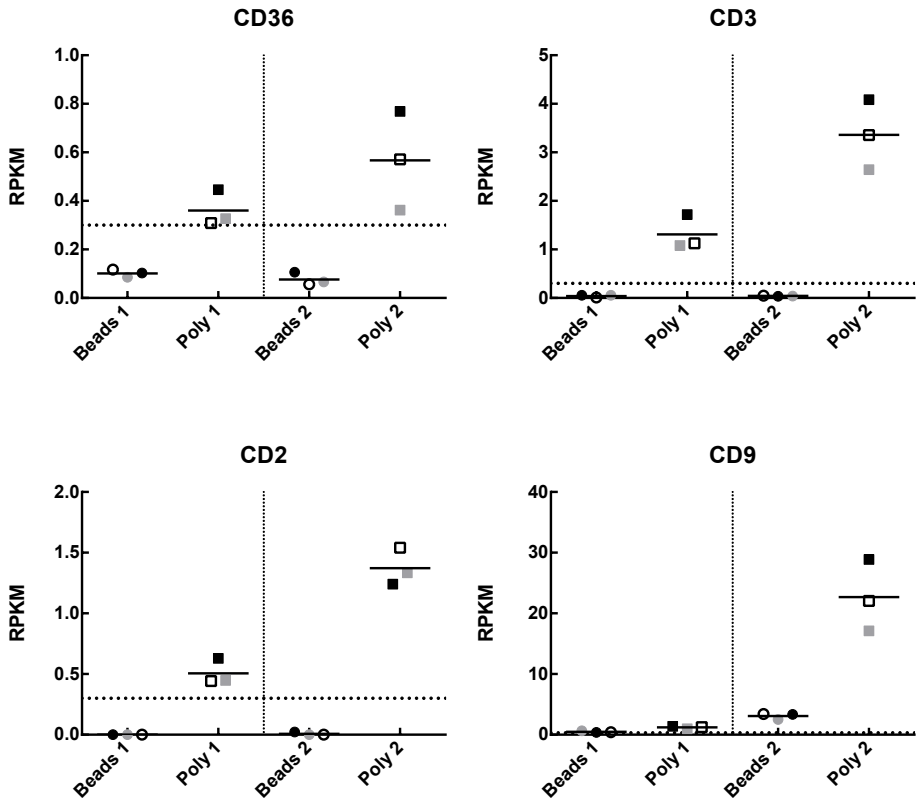


Fig 6.9. RPKM values for non-neutrophil genes of the antigen targets in the magnetic bead isolation kit. The antigens not shown (CD19, CD56 and glycophorin A) were not expressed above the cut off RPKM values (0.3) under any conditions. Neutrophils were either isolated by magnetic beads isolation (Bead) or by Polymorphprep™ (Poly) from Donor 1 (1) and Donor 2 (2) . Neutrophils were treated with 5 ng/mL GM-CSF (●/■), 10ng/mL of TNFα (○/□) or untreated (●/■) for 1h. Horizontal dotted lines represent RPKM expression threshold of 0.3. Horizontal bars represent mean value.

All 6 transcripts were present at very low levels in all samples (RPKM < 2), apart from the eosinophil specific transcript CLC in Polymorphprep™ prepared samples from Donor 2 (RPKM=249.5). As described previously for transcripts for cell-surface markers, the expression levels of transcripts for other cell-specific markers were higher in Polymorphprep™ preparations than in magnetic bead preparations from both donors. Furthermore, with the exception of CD163, all transcripts had highest expression levels in the

sample prepared by Polymorphprep™ from Donor 2. Importantly, where a transcript is expressed at a level >0.3 RPKM (that is, above the expression threshold) in the Polymorphprep™ preparations, the expression level in the paired bead isolation preparation was <0.3 RPKM (with the exception of CD163, RPKM = 0.36), suggesting that most contaminating cells have been removed. The data in Fig 6.9 and 6.10 confirm that the major source of cellular contamination seen in the Polymorphprep™ preparations originates from eosinophils and that the magnetic bead isolation procedure is sufficient to remove this contamination.

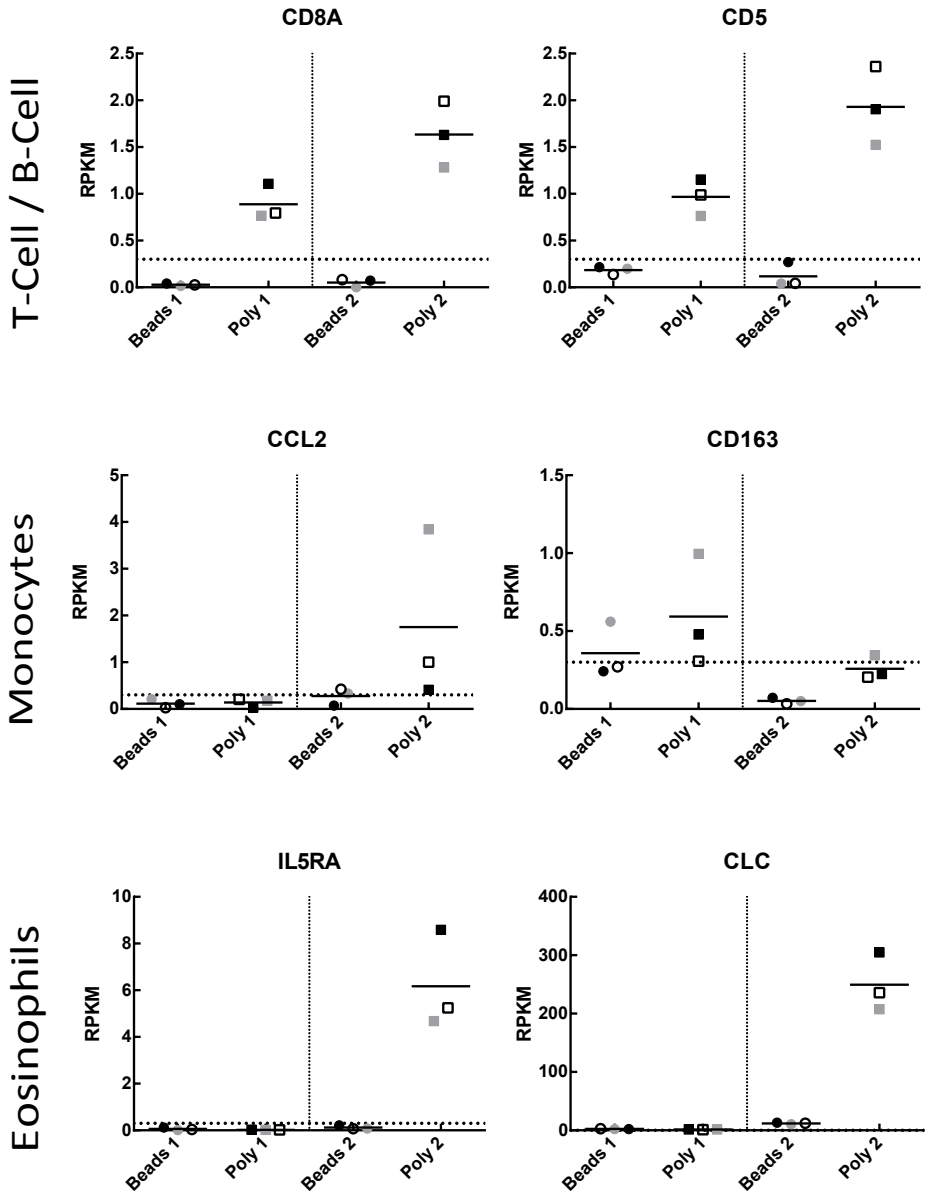


Fig 6.10 RPKM values for non-neutrophil specific genes associated with T and B cells (top) monocytes (middle) and eosinophils (bottom). Neutrophils were either isolated by magnetic beads isolation (Bead) or by Polymorphprep™ (Poly) from Donor 1 (1) and Donor 2 (2) . Neutrophils were treated with 5 ng/mL GM-CSF (●/■), 10ng/mL of TNFα (○/□) or untreated (●/■) for 1h. Horizontal dotted lines represent RPKM expression threshold of 0.3. Horizontal bars represent mean value.

6.4.9 Comparison of differentially-expressed genes between isolation methods.

Having analysed the expression levels of transcripts expressed by non-neutrophil cell types, the open-source high-throughput annotation software Cufflinks [145] was used to perform differential expression tests on the raw data to identify all transcripts within the transcriptome that were DE between isolation methods. All samples from both isolation methods were initially compared with each other (all 6 samples prepared using Polymorphprep™ compared with all 6 bead-isolated samples) and then subsequently, samples were compared with only their paired-treatment sample, for example, untreated samples isolated by Polymorphprep™ (from Donor 1 and Donor 2) were compared with untreated samples prepared by magnetic bead isolation (from Donor 1 and Donor 2). This would determine two things, firstly, the effect of neutrophil isolation method (and typical levels of eosinophil contamination) on gene expression profiles, and secondly, it would identify transcripts that were DE (between isolation methods) following stimulation with inflammatory cytokines.

6.4.9.1 Polymorphprep™ vs Magnetic bead isolation (all samples)

Samples from both donors were analysed by RNA-Seq to identify changes in neutrophil gene expression following either Polymorphprep™ or negative magnetic bead isolation. When comparing all Polymorphprep™ samples against all bead samples from both donors (i.e. 6 Polymorphprep™ samples (two donors, 3 treatments) vs 6 bead samples

(two donors, 3 treatments)). Cuffdiff identified only 16 genes (from a possible 24,934) that were significantly differentially-expressed between isolation methods (Table 6.3). All 16 genes showed a low expression value in the Polymorphprep™ samples (ranging from 0.708-20.548 RPKM) and were not detected in the bead samples (with the exception of ADP-Ribosylation Factor-Like 4C (ARL4C) RPKM=0.471). This suggests that their detection is due to increased cell contamination in the Polymorphprep™ samples since several of the genes can be attributed to non-neutrophil cells, for example HBB and CD3 to erythrocytes and T cells, respectively. However, the absolute expression values of all 16 genes were low (< 21 RPKM) and so contribute little to the overall transcriptome profiles of neutrophil suspensions prepared by Polymorphprep™.

Table 6.3 List of genes with significantly different expression levels in neutrophil suspensions prepared by different isolation methods. Table shows all significantly differentially expressed genes in order of the greatest change in RPKM values (ΔRPKM) between neutrophils isolated using either Polymorphprep (poly) or magnetic beads. Significance (q-value) as calculated by Cuffdiff and adjusted for 5% false discovery rate by Benjamini-Hochberg correction for multiple-testing. Data calculated from two biological replicate sets, as described in text.

Gene Name	Poly (RPKM)	Beads (RPKM)	Fold change (log2)	q-value	ΔRPKM
HBA2	20.548	0.093	-7.786	6.50E-04	20.454
THBS1	12.944	0.205	-5.981	1.53E-03	12.739
HBB	11.680	0.080	-7.197	3.83E-03	11.600
ARL4C	10.754	0.471	-4.513	2.98E-02	10.283
ALOX15	10.068	0.041	-7.935	3.12E-06	10.027
PRSS33	7.281	0.008	-9.806	1.55E-02	7.273
IL7R	5.736	0.295	-4.282	3.97E-03	5.441
EMR4P	5.570	0.158	-5.138	6.64E-03	5.412
S1PR1	4.115	0.216	-4.250	2.02E-02	3.899
CD3E	3.867	0.058	-6.054	6.50E-04	3.809
SMPD3	3.232	0.049	-6.036	8.79E-04	3.183
CCR7	2.958	0.154	-4.263	1.71E-02	2.804
SIGLEC8	2.263	0.036	-5.983	1.87E-03	2.227
GPR114	1.072	0.020	-5.778	2.98E-02	1.052
ITK	1.021	0.053	-4.276	9.06E-03	0.968
BCL11B	0.708	0.018	-5.268	8.76E-03	0.690

6.4.9.2 Polymorphprep™ vs magnetic bead isolation (treatment specific comparison)

The following analyses were performed to determine if the method of isolation had any impact on the patterns of gene expression. Samples of neutrophils treated under similar incubation conditions (N=2), but prepared by the two different methods were compared (for example, untreated Polymorphprep™ vs untreated bead-isolated). This analysis identified a total of 25 genes in all 3 treatment groups whose expression was

significantly different between the two preparation methods. Of these, 23 genes were significantly differentially expressed in the untreated samples, and 9 genes were significantly altered in all 3 treatment pairings (Fig 6.11). As with the genes listed in Table 6.3, all expression levels for genes in Fig 6.11 were higher in neutrophil samples prepared by Polymorphprep™ compared to samples prepared by magnetic beads.

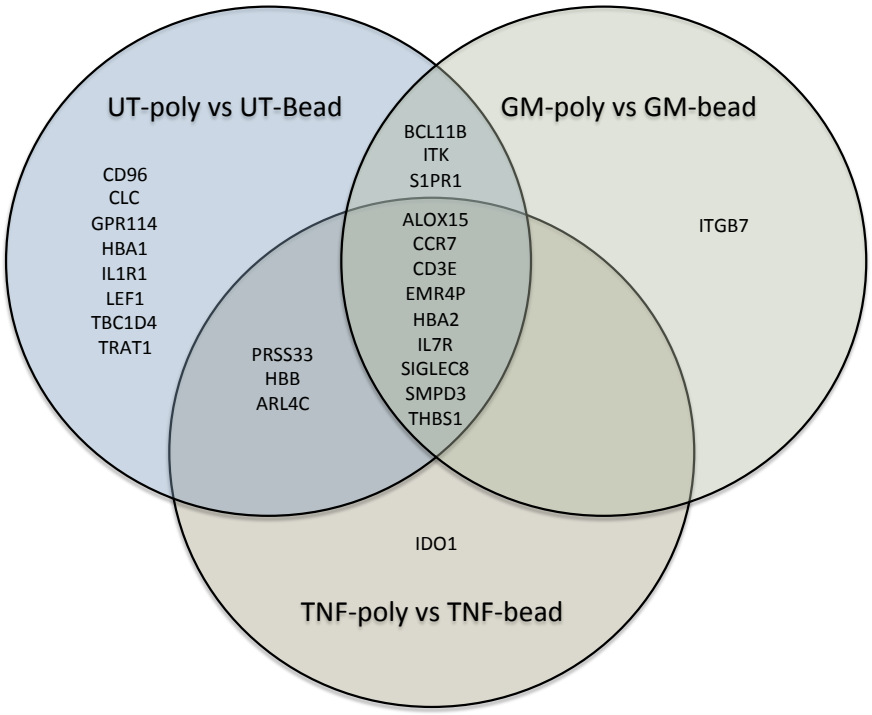


Fig 6.11 Venn diagram showing differentially expressed genes between neutrophil samples prepared by either Polymorphprep™ (poly) or magnetic beads (bead). Comparisons performed by Cufflinks using treatment specific paired-samples from two biological replicates. All genes displayed were significantly differentially expressed due to a higher RPKM in Polymorphprep™ prepared samples. Significance ($q < 0.05$) as calculated by Cuffdiff and adjusted for 5% false discovery rate by Benjamini-Hochberg correction for multiple-testing. See glossary for full gene names.

This independent analytical approach has confirmed the data in sections 6.4.7-6.4.8 that neutrophil suspensions isolated by Polymorphprep™

express only low levels of transcripts that are attributable to contaminating cells. However, neutrophil suspensions isolated by magnetic beads do not express these transcripts. Furthermore, cytokine treatment had very little effect on the number of genes DE between isolation methods. This indicates that the contaminating cells did not respond to these cytokines by alterations in gene expression to these cytokines. Thus, in subsequent experiments, cytokine-regulated changes in gene expression can be attributed to altered activity of neutrophils.

6.4.10 Comparison of gene expression profiles of two different donors following neutrophil isolation using two separate methods

The previous section identified genes which showed significantly higher expression in samples prepared by Polymorphprep™. This section set out to answer a number of different questions:

1) Does the isolation method affect gene expression in neutrophils? For example, can spending different lengths of time at 37°C and/or being subject to different centrifugal forces cause a significant change in gene expression?

2) Are identical sub-populations of neutrophils isolated by different techniques? For example, LDGs (see section 1.4.3) which are present at variable levels in healthy controls or patients with inflammatory disorders, have different density properties and hence sediment at different rates in

density-gradient such as Polymorphprep™. It has been reported that LDGs express different genes compared to normal density granulocytes [52,81]. Additionally, the bead isolation method only yielded 40% of the total number of neutrophils recovered by Polymorphprep™ (Fig.6.8), but the reason for this loss of cells are unknown. It is therefore possible that different sub-populations are isolated by the 2 techniques, and that these different sub-populations may have different gene expression profiles and/or respond differently to cytokines.

Therefore, the analysis of the data in this section was designed to address these questions. This was achieved by comparing changes in gene expression in suspensions purified by either Polymorphprep™ (pooling data from each of the incubation conditions) or purified by magnetic beads. Data for each donor were analysed separately in order to detect changes in expression levels that may be donor-dependent (see appendix A.8 for datasets used in each comparison). Any gene whose expression is lower in the bead-isolated sample is likely to be due to contamination in Polymorphprep™ samples. However, a gene whose expression is higher in the bead-isolated samples could likely be due to either:

a) The physical conditions employed of the isolation method; or

b) A different sub-population of neutrophils.

6.4.11 Donor specific analysis of neutrophil samples prepared by either Polymorphprep™ or magnetic bead isolation

For Donor 1 (low contamination donor), Cuffdiff identified 63 genes DE between isolation methods, whereas a higher number, 282 genes, were detected for Donor 2 (high contamination donor).

Unlike previous analyses using pooled data-sets from both donors, where all significant genes were expressed at higher levels in Polymorphprep™ isolated samples, when analysing each donor independently, a proportion of differentially regulated genes had higher values in the bead-isolated samples. For example, for Donor 1, 10/63 genes showed higher expression in bead-isolated samples than in Polymorphprep™ samples, and in Donor 2 92/282 showed higher expression in bead-isolated samples than in Polymorphprep™ samples. Interestingly, the percentage of significant genes expressed at higher levels in beads compared to Polymorphprep™ in Donor 2 is twice that seen in Donor 1 (32.7% and 15.9% respectively) (see Appendix Table A.9).

6.4.11.1 Donor 1

The 25 genes with the greatest change in RPKM between isolation methods for Donor 1 are listed in Table 6.4. The greatest change in RPKM seen in Donor 1 is for the gene FBJ Murine Osteosarcoma Viral Oncogene Homology (FOS) which has an RPKM of 169 in Polymorphprep™ samples and a significantly higher value of 1148 in magnetic bead isolated samples (q-value = 0.019). The next 3 genes with the greatest change in RPKM

between isolation methods are HBA1, HBA2 and HBB, which encode the alpha and beta subunits of the haemoglobin, an important protein in oxygen transportation in erythrocytes. Despite lacking a nucleus, erythrocytes and reticulocytes (immature erythrocytes) are capable of gene transcription and translation [255]. The presence of these transcripts in Polymorphprep™ samples most likely represents contaminating erythrocytes which were not eliminated by the hypertonic lysis step. Since the bead isolation kit contains an antibody recognising an erythrocyte-specific antigen (glycophorin A), this contamination is predictably absent from all bead-prepared samples. Although the expression values of the remaining genes are judged to be significant by Cuffdiff, the differences in RPKM values between samples are very low, with only 15/63 genes having a change in RPKM of >2 between isolation conditions Table 6.4.

Table.6.4 List of genes significantly regulated between isolation methods in neutrophils from Donor 1 (low contamination). Table shows the top 25 genes with the greatest change in RPKM values (ΔRPKM) between neutrophils isolated using either Polymorphprep™ (Poly) or magnetic beads. Genes expressed at higher levels in bead-isolated samples are shaded grey. Significance (q-value) as calculated by Cuffdiff, adjusted for 5% FDR by Benjamini-Hochberg correction for multiple-testing. Data from 3 paired samples from Donor 1.

Gene Name	Poly (RPKM)	Beads (RPKM)	Fold change (log2)	q-value	ΔRPKM
FOS	169.379	1148.700	2.762	1.90E-02	-979.321
HBA1	26.138	0.016	-10.642	2.29E-02	26.122
HBA2	24.521	0.015	-10.656	2.29E-02	24.506
HBB	16.238	0.040	-8.683	2.95E-05	16.198
CKS2	1.285	10.227	2.993	1.45E-05	-8.943
KLF4	2.642	10.591	2.003	3.51E-02	-7.948
LCN2	1.292	7.993	2.629	2.77E-05	-6.701
LTF	0.799	6.060	2.923	2.10E-06	-5.261
EREG	5.489	0.394	-3.799	1.15E-03	5.094
RPL10A	6.760	1.817	-1.896	3.49E-02	4.943
IL7R	3.758	0.328	-3.517	1.78E-07	3.430
CD3E	2.761	0.029	-6.564	8.10E-07	2.732
THBS1	2.735	0.051	-5.749	0.00E+00	2.684
KIAA0090	0.361	2.880	2.996	8.10E-07	-2.519
CCR7	2.093	0.066	-4.981	4.05E-10	2.027
SERPINB2	1.959	0.286	-2.775	1.06E-02	1.673
CCL5	1.881	0.338	-2.474	9.02E-04	1.543
GIMAP7	1.567	0.263	-2.576	1.47E-03	1.305
VCAN	1.261	0.021	-5.907	9.02E-04	1.240
TCF7	1.421	0.209	-2.768	2.45E-02	1.212
LDHB	1.317	0.137	-3.264	2.14E-04	1.180
ARL4C	1.378	0.227	-2.603	2.50E-04	1.151
S1PR1	1.300	0.150	-3.118	7.56E-06	1.151
MMP8	0.108	1.024	3.245	1.33E-05	-0.916
CD5	1.102	0.192	-2.525	4.52E-04	0.911

6.4.11.2 Donor 2

By comparison, Table 6.5 lists the 25 genes with the greatest RPKM change between isolation methods in samples from Donor 2. Only 3 genes (LCN2, HBA1 and MMP8) feature on both lists. The number of significant genes for Donor 2 is much greater than for Donor 1 and the magnitude of the RPKM differences between isolation methods are also greater in Donor 2. Interestingly, despite the increased purity of neutrophils isolated by magnetic beads (as shown earlier by flow cytometry and cytospins), 15 of the top 25 significant genes exhibit higher expression in bead-isolated samples than in Polymorphprep™ isolated samples. These include the two genes with the greatest change in RPKM (neutrophil defensin – DEFA1 and neutrophil gelatinase-associated Lipocalin LCN2). The genes that have higher expression in Polymorphprep™ samples are those most-commonly associated with other, non-neutrophil cell types, such as CLC (eosinophils), CD52 (lymphocytes) and PLIN2 (epithelial cells). However, the genes with higher expression in bead-isolated neutrophils are known neutrophil genes, for example Carcinoembryonic antigen-related cell adhesion molecule 8 (CEACAM8) and Bactericidal/Permeability-Increasing Protein (BPI): in some cases these are neutrophil-specific genes, for example neutrophil elastase (ELANE) [256] and lipocalin 2 (LCN2) [257]. Consequently, these genes with significantly higher expression in bead isolated samples cannot be attributed to contaminating leukocytes.

Table. 6.5 List of genes significantly regulated between isolation methods in neutrophils from Donor 2 (high contamination). Table shows the top 25 genes with the greatest change in RPKM values (ΔRPKM) between neutrophils isolated using either Polymorphprep™ (Poly) or magnetic beads. Genes expressed at higher levels in bead-isolated samples are shaded grey. Significance (q-value) as calculated by Cuffdiff adjusted for 5% false discovery rate by Benjamini-Hochberg correction for multiple-testing. Data from 3 paired samples from Donor 2.

Gene Name	Poly (RPKM)	Beads (RPKM)	Fold change (log2)	q-value	ΔRPKM
DEFA1	239.933	1082.690	2.174	4.75E-02	-842.757
LCN2	89.952	447.161	2.314	2.45E-02	-357.209
CLC	264.289	15.800	-4.064	6.14E-09	248.489
CAMP	32.519	138.676	2.092	2.93E-02	-106.157
BPI	24.068	117.488	2.287	2.05E-02	-93.420
OLFM4	16.438	87.792	2.417	1.36E-02	-71.354
CD52	56.669	2.804	-4.337	3.79E-04	53.865
CEACAM8	12.433	65.236	2.391	9.63E-03	-52.803
DEFA4	9.968	57.406	2.526	1.26E-03	-47.438
PLIN2	40.458	1.758	-4.524	2.23E-08	38.699
MS4A3	8.837	45.919	2.377	1.54E-02	-37.081
MMP8	9.436	41.732	2.145	3.68E-02	-32.297
AZU1	7.485	38.436	2.360	4.26E-03	-30.951
ELANE	5.851	34.581	2.563	1.00E-03	-28.731
RETN	5.085	28.648	2.494	5.32E-03	-23.563
LGALS12	23.813	0.657	-5.180	7.78E-06	23.157
RGS1	25.601	2.924	-3.130	4.85E-04	22.677
THBS1	22.995	0.367	-5.969	0.00E+00	22.628
ALOX15	22.411	0.047	-8.894	0.00E+00	22.364
ARL4C	22.673	0.775	-4.871	0.00E+00	21.898
HBA1	19.632	0.031	-9.311	4.39E-02	19.601
CEACAM6	5.968	25.207	2.079	2.95E-02	-19.239
H1F0	3.009	21.615	2.845	2.19E-04	-18.605
HBA2	18.605	0.202	-6.529	1.92E-08	18.403
C13orf15	5.171	23.425	2.180	1.16E-02	-18.254

6.4.12 Filtering of gene lists from both donors to enrich for genes with highest expression changes between isolation methods

When a filter is applied to the full list of significant genes to remove genes which have <2 RPKM difference between isolation methods, the number of genes decreases from 63 to 15 for Donor 1 and from 282 to 116 for Donor 2 (Fig 6.12). This filtering also decreases the proportion of genes with a higher value in bead samples to a greater extent in Donor 1 than in Donor 2. This reveals that of the genes DE in either donor, a greater proportion of those in Donor 2 have a considerable difference in RPKM (> 2), whereas the majority of DE genes in Donor 1 have a very small (< 2) difference in RPKM between isolation methods, and although deemed to be significant by Cuffdiff, are unlikely to have a considerable effect on overall gene expression profiles.

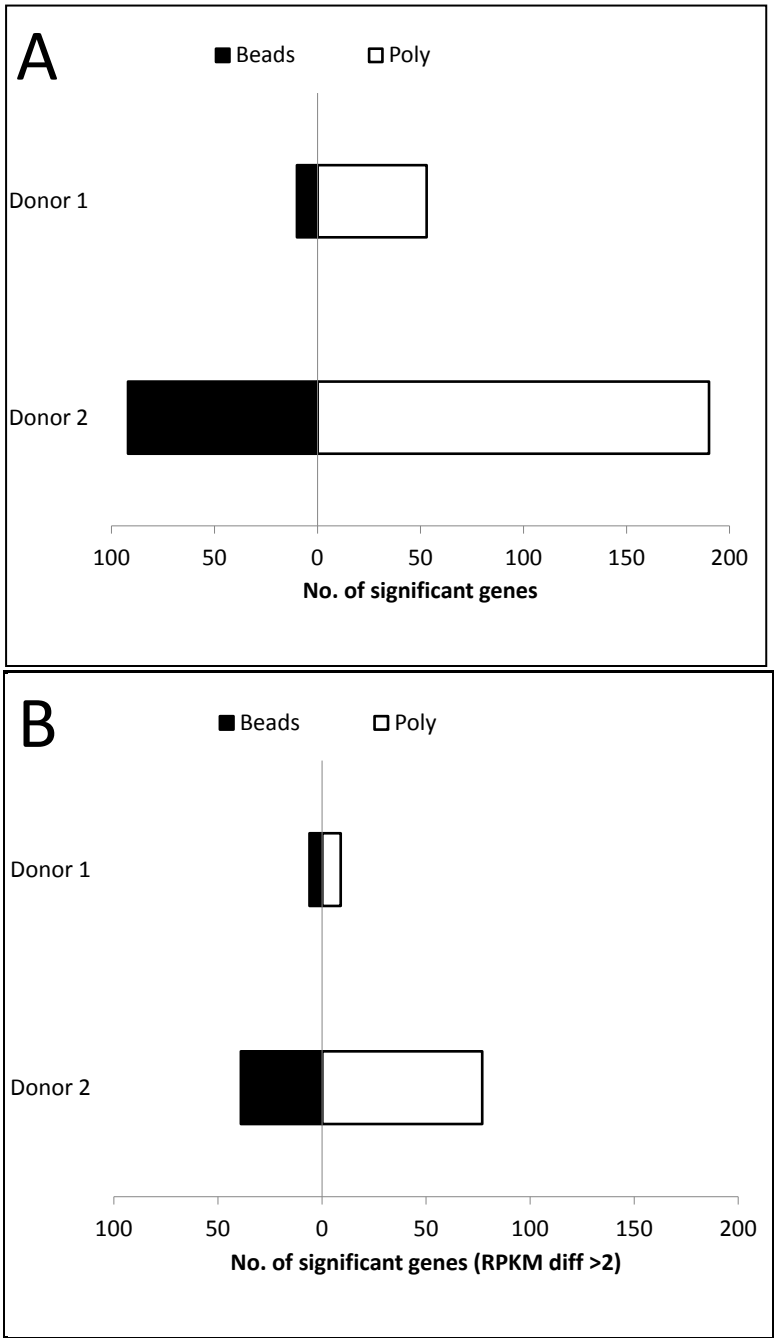


Fig 6.12 Number of genes significantly-regulated in neutrophils is dependent on Donor and neutrophil isolation method. (A) The number of genes significantly differentially expressed (DE) is higher in Donor 2 than Donor 1, and a greater proportion of those genes are higher following Polymorphprep™ isolation (white bars) than magnetic bead isolation (black bars). (B) Filtering significant DE genes for values that only show >2 RPKM difference between isolation method values, dramatically decreases the number of genes in Donor 1 but only halves the number of genes in Donor 2. The proportion of genes that are higher in bead isolation is also much higher in Donor 2 than Donor 1, and is less affected by filtering gene list.

6.4.13 Genes enriched in bead isolated samples

A striking finding in the list of significant genes between Donor 1 and 2 is the proportion of genes that are enriched in the bead-isolated samples. It might be predicted that if any difference existed in gene expression values, this would likely result from the small percentage of contamination from other cell types in the Polymorphrep™ isolated cells. Consequently, these values would be highest in Polymorphrpep™ samples and greatly decreased (or absent) in the bead isolated samples. In fact, of the significantly differentially expressed genes in Donor 2, almost a third of these show higher expression in bead-isolated samples (92/282) (for full list of the 92 genes see Appendix, Table A.10) including 12 of the top 15 genes with greatest changes in RPKM (Table 6.5, grey shading).

Further analysis of these genes identified several encoding neutrophil granule proteins and anti-bacterial peptides. These include (but are not limited to) lactoferrin (LTF), defensin (DEFA1), myeloperoxidase (MPO), neutrophil elastase (ELANE), Bacterial Permeability-Increasing protein (BPI) and azurocidin (AZU1). Neutrophil granule proteins are expressed and compartmentalised prior to maturation and release of mature cells into the peripheral blood [258,259]. Typically, granule protein genes are not transcribed in fully-mature neutrophils [18], suggesting their presence in Donor 2 samples is indicative of the presence of a sub-population of pre-mature neutrophils or progenitor cells, that is otherwise absent or below the threshold of detection in Donor 1 samples.

Recent work by Villanueva and colleagues [52] has sought to define the transcriptional profile of LDGs in SLE. Micro-array studies showed that a total of 18 genes are significantly elevated in SLE LDGs when compared to levels detected in control or SLE normal density granulocytes. Among the 18 genes include those relating to granule enzymes, and molecules associated with ROS production, NET formation and bactericidal activity [52] (Table 6.6).

Analysis of genes significantly expressed between isolation methods in Donor 2 revealed that several of the genes have previously been shown to be associated with immature neutrophils and/or LDGs. This was confirmed by IPA-software analysis (data not shown).

Table 6.6 List of functions of 18 LDG-associated genes as defined by Villanueva et al [52].

Gene symbol	Gene Name	Function
MMP8	Matrix metallopeptidase 8	Gelatinase and collagenase activities
MMP9	Matrix metallopeptidase 9	
CEACAM1	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 1 (CD66a)	Adhesion molecule
CEACAM8	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 8 (CD66b)	
RNASE2	Ribonuclease 2	Non-secretory ribonuclease (pyrimidine specific)
RNASE3	Ribonuclease 3	
CAMP	Cathelicidin antimicrobial peptide (LL37)	Antimicrobial peptide, chemotaxis, inflammatory response regulation
CTSA	Cathepsin A	Lysosomal serine proteases, antibacterial activity (anti-gram-negative)
CTSG	Cathepsin G	
ELANE	Elastase	Sereine protease, elastin degredation, phagocytosis, migration
MPO	Myeloperoxidase	Microbicidal activity, catalyses production of ROS. Granule protein
AZU1	Azurocidin 1	Granule protein, antibacterial, granule protein
DEFA4	Defensin, alpha 4	Microbicidal peptide, granule protein
BPI	Bactericidal/Permeability-Increasing Protein	Bactericidal peptide, LPS binding, granule protein
CRISP3	Cystein-rich secretory protein	Immuno-regulation
LCN2	Lipocalin 2	Iron-sequestering, granule protein
LTF	Lactotransferrin	Iron-binding granule protein, multi-functional
CLU	Clusterin	Extracellular chaperone

Analysis of the expression levels of the 18 LDG genes (defined by Villanueva et al) in both the Polymorphprep™ and bead-isolated samples from Donor 1 and Donor 2 show that all transcripts are elevated in Donor 2 samples compared to Donor 1. Furthermore, when comparing differences in gene expression of LDG genes between isolation methods, 14 of these 18 genes are significantly elevated in bead-isolated samples from Donor 2 compared to paired Polymorphprep™ isolated samples. These findings are summarised and presented in Fig 6.13. Hierarchical clustering of expression values resulted in samples from the same donor grouping together. Taken together, these findings suggest that Donor 2 has elevated levels of LDGs and that bead-isolation of neutrophils enriches for this LDG sub population that could otherwise not be recovered by Polymorphprep™ isolation.

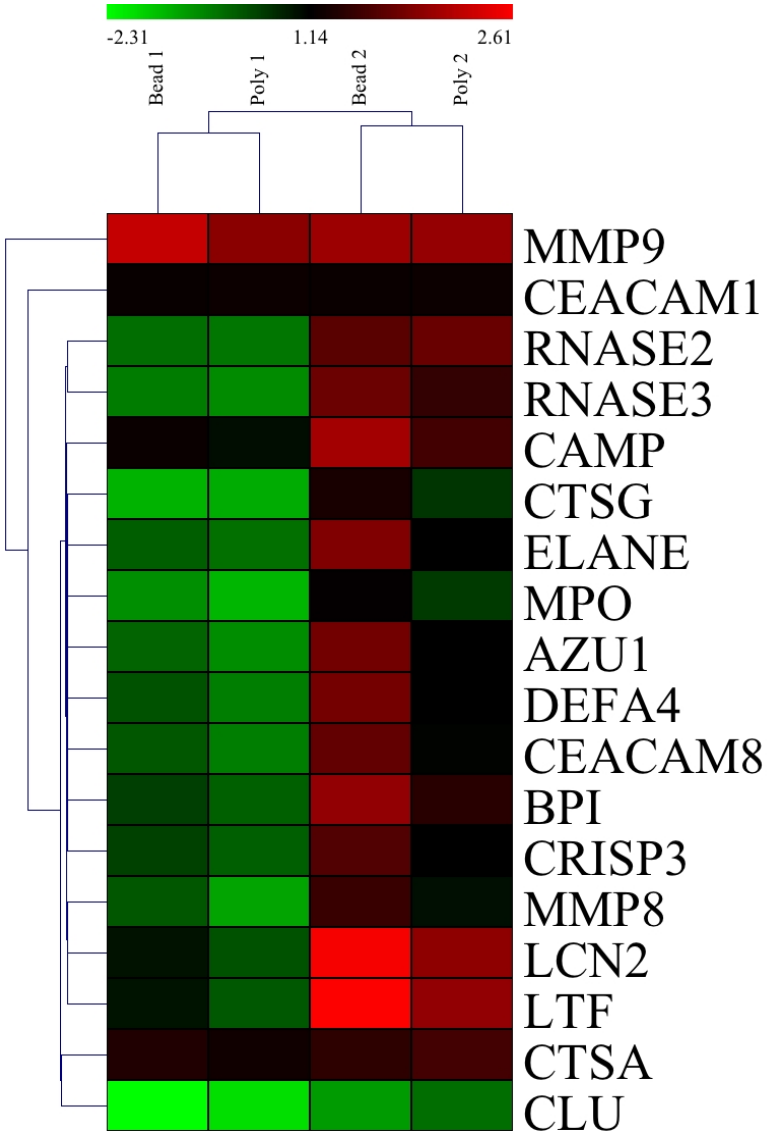


Fig 6.13 LDG genes are elevated in Donor 2 samples and enriched in bead-isolated samples compared to Polymorphprep™ isolated samples. Heatmap of log_2 transformed expression values (RPKM) for 18 LDG associated genes in bead- or Polymoprhprep™ (Poly)- isolated samples from Donor 1 (low contamination) or Donor 2 (high contamination). Lowest expression values are shown in green, median values shown in black and highest expression values shown in red. Hierarchical clustering shows greatest association between paired samples from the same donor and greatest divergence between donor samples. Clustering achieved using Pearson correlation and average linkage by Multiple experiment viewer (MeV).

6.4.14 Comparison of variation between donor following different isolation methods

The previous section identified that different genes are expressed in neutrophil samples isolated by different methods, and that the differences are donor-dependent. However, it is unclear whether the number of genes which are DE between donors is altered by different isolation methods. For example, whether the number of genes differentially expressed between donors following Polymorphprep™ isolation is similar to the number of genes differentially expressed between donors following bead isolation.

To answer this question, Cuffdiff anlaysis was used to identify differentially expressed genes between Polymorphprep™ isolated samples from Donor 1 with Polymorphprep™ isolated samples from Donor 2. Likewise, bead-isolated samples from Donor 1 were compared to bead-isolated samples from Donor 2. This analysis would identify the number of genes that were differentially expressed between donors for each isolation method. Since the purity of neutrophil preparations from both donors was similar following bead isolation (Donor 1 - 97.5%, Donor 2 - 99%) it might have been expected that fewer genes would be significantly differentially expressed between donors after this purification method, than for Polymorphprep™ isolated samples. Likewise, as there is a much greater difference in neutrophil purity levels between the donors following Polymorphprep™ preparation (Donor 1 - 96%, Donor 2 - 83%) it might also be predicted that a greater number of genes were differentially expressed

after preparation by this method than would be in the bead-isolated samples.

However, Cuffdiff identified almost 3 times as many differentially-expressed genes in the Polymorphprep™ samples as were detected in the bead-isolated samples (Table 6.7). This suggests that there is greater heterogeneity in gene expression profiles between donor neutrophil samples if samples are prepared by magnetic bead isolation rather than by Polymorphprep™. Indeed, it is likely the increased number of significantly expressed genes between donors is a consequence of the more heterogeneous neutrophil population selected by magnetic bead isolation, and that the presence of a sub-population of LDGs can significantly contribute to the overall neutrophil gene expression profile.

Table 6.7 Number of significantly differentially expressed (DE) genes between neutrophil samples prepared by either Polymorphprep™ (Poly) or by magnetic bead isolation (Bead) from Donor 1 (1) and Donor 2 (2). Significance (q<0.05) as calculated by Cuffdiff adjusted for 5% false discovery rate by Benjamini-Hochberg correction for multiple-testing. Data calculated from 3 sets of paired replicates.

Samples compared	**Poly-1 vs Poly-2**	**Bead-1 vs Bead-2**
Number of DE genes	531	1544

6.5 Discussion

Recent evidence suggests that the presence of contaminating leukocytes in a neutrophil preparation can alter the behaviour of neutrophils, or generate results that are difficult to interpret [249,254]. Traditional techniques for isolating neutrophils from peripheral blood have relied on centrifugation of whole blood (or erythrocyte-depleted whole blood) over density-gradient media and these typically achieve neutrophil purities of >95%. In recent years, more sophisticated techniques have emerged which can achieve much higher purity levels (>99%), without the need for lengthy centrifugation steps and the ability to automate the whole isolation protocol. The costs of these latter procedures are significantly higher than those based on density-gradient centrifugation, but such methods are reported to mitigate the potential effects of contaminating cells and improve consistency of data obtained from different donors.

Despite recognition of the importance of neutrophil purity for *in vitro* studies, very few studies have focused on the impact of purity and or isolation method on neutrophil behaviour. Indeed, no studies yet have determined the molecular properties of neutrophils isolated by different purification protocols (neither under stimulated or untreated conditions).

In this study, peripheral blood neutrophils from two healthy controls were isolated using two commonly employed methods of neutrophil isolation (density gradient by Polymorphprep™ and magnetic bead negative selection). The two donors were selected because previous work had

identified these as having low levels (Donor 1) or high levels (Donor 2) of contaminating cells, which were mainly eosinophils.

Independent quantification of neutrophil purity by cytospin and flow cytometry confirmed the levels of inherent cellular contamination in neutrophil samples from each donor. Quantification also demonstrated that levels of neutrophil purity achievable by each isolation method were broadly in line with previously published data [227,246,251]. However, and somewhat surprisingly, the often referenced neutrophil isolation purity value of >99% following bead isolation [182,227,260,261] was not achieved in any preparation.

Levels of contamination in leukocyte preparation purity have previously been suggested to affect rates of apoptosis in both neutrophil preparations [249], and eosinophil preparations [262]. However, in both of these studies, contamination by $CD14^+$ cells (monocytes) was shown to be crucial to delaying apoptosis following lipopolysaccharide (LPS) stimulation. Given that levels of PBMCs were equally low (≤3%) in both donors and isolation methods, it is perhaps surprising that (when compared to untreated samples), TNFα incubation was pro-apoptotic in bead-isolated neutrophils and anti-apoptotic in neutrophils isolated by Polymorphprep™. Several reasons could explain this, firstly, only low levels of contaminating cells may be required to alter the affect of TNFα. Secondly, the Polymorphprep™ method involves exposure of cells to greater centrifugal forces that may have an effect on neutrophils perhaps altering their responsiveness to

TNFα. Finally, the apparently pro-apoptotic effect of TNFα on bead-isolated neutrophils could be due to the fact that there are great differences in rates of apoptosis in untreated neutrophils prepared by these two methods (Poly-Untreated 64.71% ± 7.34%, Beads-untreated 37.04% ± 4.21%). Levels of apoptosis in TNFα-stimulated neutrophils are similar regardless of isolation method (Poly-TNF 47.82% ± 3.20%, Beads-TNF 52.46% ± 2.04%). This would suggest that the neutrophils that are "lost" during bead isolation are those with normally high rates of constitutive apoptosis.

TNFα has previously been shown to exhibit bi-modal effects on neutrophils [183,225] and other cell types [263–265], that may be time-dependent [266], or concentration-dependent [265]. Given that the pro-apoptotic effect of TNFα is more pronounced in Donor 2 bead samples than Donor 1 bead samples (Donor 1: +12.68% apoptosis vs control, Donor 2; +27.03% vs control), it is possible that eosinophils may play a role in delaying neutrophil apoptosis. Indeed, eosinophils are capable of expression and release of TNFα under activated conditions [267] and as such may be providing a consistent source of anti-apoptotic paracrine signalling in the Polymorphprep™ isolated samples, thereby delaying apoptosis in these less pure preparations of neutrophils. Additionally, the higher rates of apoptosis seen in all Polymorphprep™ samples (compared to bead isolated samples) may be due to the higher centrifugation steps during the isolation protocol than those employed in a magnetic bead preparation. Indeed, additional

centrifugation of bead-isolated neutrophils on Polymorphprep™ for 30 min increased levels of neutrophil apoptosis to levels similar to those seen in Polymorphprep™ isolated samples (see Appendix Fig A.5).

Analysis of neutrophil cell-surface expression levels revealed no significant changes after purification by different isolation methods (Student's t-test $p>0.2$). Whilst CD15 and CD16 are expressed at high levels on neutrophils, they are also expressed on other leukocytes [268] (with the exception of CD16b which is neutrophil specific). This might explain why levels of expression of these receptors were not significantly different in neutrophils prepared by isolation methods. Since CD11b is known to be upregulated following neutrophil activation [269], the small increase in expression seen in Polymorphprep™ isolated neutrophils may represent an increased level of activation – although levels were not significantly different ($p= 0.27$) – perhaps due to the additional centrifugation time of a Polymorphprep™ isolation method in comparison to a magnetic bead preparation. However, this does not explain the slight increased expression of CD64 in bead-isolated neutrophils (although also not significant $p= 0.33$), which can also be upregulated in neutrophils following activation, albeit by enhanced gene expression in response to cytokines such as IFNγ and LPS [270].

In contrast to the increased neutrophil purity following bead isolation, the absolute yield of neutrophils obtained from whole blood was much lower than can be achieved using Polymorphprep™. It is likely that several steps in the bead isolation protocol contribute to this decreased yield. Firstly,

erythrocyte depletion of whole blood using dextran relies on the aggregated erythrocytes sedimenting at a faster rate than leukocytes, such that an erythrocyte-free upper phase is produced which is removed and processed further. Although this process is effective for depletion of erythrocytes, leukocytes in the lower portion of the tube remain there and will not rise into the erythrocyte-free zone. Secondly, since the isolation method is by negative selection, neutrophils must avoid false-positive selection by each of the 7 specificities of antibodies included in the bead isolation kit. Finally, in the last steps of the bead isolation method, pure neutrophils are decanted into a fresh tube, while the contaminating cells are retained within a magnetic field. This must be performed in a single motion to avoid dislodging contaminating cells from the magnetic field. Consequently, a proportion of this neutrophil suspension is retained within the tube, further decreasing the final yield of pure neutrophils.

Regardless of the factors contributing to a decreased neutrophil yield, this difference in yield between isolation methods, raises considerable concerns for neutrophil studies where large numbers of cells are required (for example RNA studies across a range of time points). Moreover, in situations where the available volume of whole blood is restricted, for example in neutropenic or paediatric patients, this method of neutrophil purification may be unworkable.

RNA-Seq analysis of neutrophil samples revealed that transcripts for contaminating cells (either relating to bead-kit antigens or known cell

specific transcripts) were broadly in line with expected contamination levels measured by flow cytometry of microscopic analysis. Eosinophil transcripts (CD9, IL5RA and CCL2) were at highest levels in Donor 2 Polymorphprep™ samples, with CD9 being the highest non-neutrophil transcript. Interestingly, despite its use as an eosinophil-specific cell surface antigen in the bead-isolation kit, neutrophils are known to express CD9 on their cell surface [271], albeit under disease conditions. However, this could compromise the efficacy of the bead isolation kit in experiments isolating neutrophils from patients with inflammatory diseases.

It is clear from both the data presented here and previously published work that magnetic bead isolation of neutrophils is more efficient at removing contaminating leukocytes than Polymorphprep™. However, despite general acceptance that this method provides more highly-pure neutrophil suspensions, experiments using RNA-Seq have revealed that transcripts unique to contaminating cells are still detectable within the bead-isolated samples. Indeed, levels of the monocyte marker, CD163 are higher in bead-isolated samples from Donor 1 than in Polymorphprep™ isolated samples from Donor 2. This highlights the importance and effects of donor variation on the purity of the samples irrespective of the isolation method.

When assessing the number of genes that were differentially-regulated between the two isolation methods using datasets for both donors, it was found that only a small number of genes were significantly regulated (16 genes). Furthermore, when comparing treated samples from both isolation

methods (for example GM-poly vs GM-bead, N=2) there is only a single additional significant gene in each case (integrin beta 7 (ITGB7) and Indoleamine 2,3-dioxygenase 1 (IDO1) for GM-CSF and TNFα samples respectively), suggesting that differentially expressed genes between isolation methods are not influenced greatly by the presence of inflammatory stimuli.

In contrast, when Polymorphprep™ and bead isolated samples were analysed in a donor specific manner, a greater number of genes were found to be significantly differentially expressed between the two isolation methods (in particular samples from Donor 2). Surprisingly, several genes are found to be elevated in the bead-isolated preparations. Detailed analysis revealed many of these genes are expressed in low density granulocytes, cells that are present in the circulation of patients with SLE [52]. Whilst the exact properties and functions of LDGs remain largely undefined, they have increasingly been associated with abnormal immune responses and in particular, auto-immune disease [52,79–81,272]. It is therefore somewhat surprising neutrophil samples from a healthy control (Donor 2) showed elevated levels of LDG-associated genes, implying high levels of LDGs in the blood of this healthy donor. This suggests that elevated transcription levels of LDG-genes alone are not sufficient to induce or reflect a diseased state. Alternatively, elevated levels of LDGs in peripheral blood may indicate susceptibility to disease. Alternatively, heterogeneity in a leukocyte population may be normal but overlooked as many studies

would not normally use isolation methods that specifically enrich LDGs. It is clear from the experiments described in this thesis that LDGs are not normally isolated in the neutrophil band prepared by Polymorphprep. It is also unclear if the high levels of eosinophils present in Donor 2 blood are related to the increase in LDGs also seen in this donor.

Whether LDGs represent a sub-class of mature, normal density neutrophils (NDGs) or a sub-population of immature neutrophils is still unclear. At present, the transcriptional regulation of neutrophil granule proteins during maturation is poorly understood, but the classical view is of 3 subsets of granules that develop due to temporal expression of granule genes during development. Proteins localised to each granule type are sequentially-transcribed during maturation in the bone marrow, beginning with azurophilic granule proteins. However, more recently, other marker proteins have been discovered with a non-classical transcriptional pattern [20,21,273] suggesting there is much greater granule heterogeneity than previously thought. This is highlighted by the LDG genes expressed in Donor 2 bead-isolated neutrophils, which show elevated levels of transcripts for several azurophilic granule proteins (MPO, DEFA4, BPI, ELANE, AZU1, CTSA, CTSG). The presence of these transcripts is indicative of a transcriptional profile of a neutrophil progenitor cell such as a myelocyte or metamyelocyte [18].

Regardless of the structure and function of LDGs, it is clear that bead-isolation methods for purifying neutrophils provide a method of enriching

the LDG population that might exist. Conversely, since Polymorphprep™ relies on cell density for population separation, any LDGs are excluded from the neutrophil layer and most likely retained in the PBMC layer. Consequently, the level of neutrophil-heterogeneity following Polymorphprep™ isolation is decreased. Indeed, the number of significantly DE genes between datasets of two donors of Polymorphprep™ isolated samples is considerably less than the number of DE genes between 2 datasets of bead-isolated samples (Poly - 531 genes , beads – 1544 genes, $q<0.05$ - 5% FDR). This suggests that there is more heterogeneity between two samples of neutrophils, both with >97.5% purity prepared by bead-isolation than in the two Polymorphprep™ isolated samples, which had 96% and 83% neutrophil purity. This raises important questions for the use of magnetic beads to produce ultra-pure neutrophil samples, where levels of LDGs (or other sub-types of neutrophils) in the population may vary among donors. This is of particular relevance when analysing samples from patients with inflammatory diseases, such as SLE. Additionally, a comprehensive study of LDGs is unlikely using samples of neutrophils isolated by Polymorphprep™.

In summary, whilst neutrophil purity is of significant importance for *in vitro* studies using high-sensitivity assays such as RNA-Seq or mass spectrometry, density-gradient based separation protocols such as Polymorphprep™ solution provide a suitable method of isolating unprimed, viable neutrophils, with an overall purity exceeding 96%. The

major contribution of contamination is from eosinophils. Magnetic bead isolation is effective for increasing neutrophil purity to approximately 98% but does so at the expense of overall yield and increased costs. Furthermore, despite having greater purity levels than Polymorphprep™ samples, bead-isolated neutrophil populations exhibited far greater heterogeneity due to the enrichment of an LDG-like sub-population. These data highlight the mechanistic differences between isolation methods and the inherent variation found between donors that plays an important role in the overall gene expression profile of neutrophils. It is therefore important that the success and reliability of a neutrophil assay be judged on more than the metric of purity, and should take into account several additional factors that ultimately can impact on neutrophil gene expression.

Chapter 7: Future analyses of the bioinformatic pipeline

7.1 Introduction

The pipeline of bioinformatic analyses described in Chapter 3 has enabled an accurate quantification of neutrophil gene expression following stimulation with several cytokines (Chapter 4 and 5), or in neutrophil suspensions of different levels of purity (Chapter 6). In each of these chapters, absolute gene expression values were calculated, and the RPKM metric was used to define relative values between (or within) samples. Subsequent bioinformatic analysis and predictions were made using the expression values and sets of significantly-associated gene lists (as calculated using the expression values). This approach has enabled accurate measurements of neutrophil gene expression under various conditions and has identified important consequences of different neutrophil isolation techniques. However, whilst RNA-Seq provides a suitable platform for measurement of absolute gene expression values, which is broadly comparable to similar analyses using microarrays, the greatest advantage of RNA-Seq over other methods of gene expression is the ability to quantify multiple genetic features from a single sequencing run. For instance, data collected during a single sequencing run can be *post hoc* analysed to quantify splice usage, SNP discovery and indels.

Whilst full analysis of the datasets to extract this genetic information was beyond the scope of this project, during the development of the bioinformatic pipeline, the software and methods necessary for quantifying these additional genetic features were incorporated into the final pipeline. The portions of this final bioinformatic pipeline, which were developed but not used in the analysis of neutrophil samples, are detailed below. In the time available for this project, it was not possible to fully extract all of this information from the datasets.

7.2 Methods

Neutrophil were isolated by Polymorphprep™ isolation from 3 healthy donors (Donors 1-3, see Appendix Table A.3 for further details). RNA samples were collected as previously described (see sections 2.2.2.2 and 2.2.9) and samples were analysed by RNA-Seq using the bioinformatic pipeline described in Chapter 3. The following results demonstrate the downstream analyses available using raw data files obtained from the mapping stage of the bioinformatics pipeline (i.e. *.Bam* files from Tophat) or using the default output files relating to isoform expression, as provided by Cufflinks/Cuffdiff (*isoforms.fpkm_tracking.txt*).

The software programmes and specific command options for each analysis are indicated in the text.

7.3 Results

7.3.1 SNP discovery using RNA-Seq data

Single nucleotide polymorphisms (SNPs) represent the most common form of genetic variation within a genome, occurring around once every 100-300 bases [274]. In recent years, genome-wide association studies (GWAS) have linked many SNPs with human diseases such as RA, diabetes and SLE [275–277]. There are currently more than $62x10^6$ identified SNPs in the human genome [278]. While the vast majority of SNPs (~88%) are located in intergenic or intronic regions of the genome [279], RNA-Seq data provides a source of data for characterising SNPs located within human exomes. It was recently shown that RNA-Seq data with as little as x10 coverage was sufficient to identify 92% of expected SNPs within expressed exons [280]. For comparison datasets analysed here have approximately x40 coverage.

Analyses characterising SNPS in the neutrophil transcriptome before and after stimulation is of little value since the genetic sequence will be unchanged during the time course of an *in vitro* experiment. However, it can be informative when comparing differences between donors or more importantly, patients with inflammatory disease where neutrophil dysregulation is implicated in disease progression, such as RA [3]. For example, a SNP located in a functionally-important region, such as a receptor-binding pocket may represent a locus that confers an increased or decreased response to drug therapy. For instance, responses to the B-cell depleting anti-inflammatory drug Rituximab have been associated with

SNPs in the FcGR genes, the IL-6 gene and the B-Lymphocyte stimulator (BLyS) gene [281]. Furthermore, SNPs in the tumour necrosis factor alpha induced protein 3 (TNFAIP3) gene, which is expressed in neutrophils [181], have been associated with a number of inflammatory conditions, such as SLE [282], psoriasis [283], diabetes [276] and RA [277].

7.3.1.1 SNP quantification in neutrophils using Samtools mpileup

The software package Samtools [284] provides a means of quantifying all SNPs within a dataset by comparison of each nucleotide location with a reference sequence (either transcriptome or genome). The proportion of reads expressing the polymorphism at each location is used to determine if the SNP is significant, a product of sequencing error, or if coverage is too low to conclude either.

RNA-Seq data (*.Bam* file) was used to quantify the total number of SNP identifiable in neutrophils from a single donor. The software Samtools was used to merge 3 *.Bam* files produced by Tophat during mapping into a single file using the command:

```
samtools merge ./output/location path/to/file1 path/to/file2 path/to/file3
```

The Samtools commands "mpileup" and "bcftools" were applied to the merged *.Bam* file using the following command:

samtools mpileup –uf path/to/genome.fa path/to/merged.bam | bcftools view –bvcg - > var.raw.merged.bcf bcftools view var.raw.marged.bcf | perl vcfutils.pl varFilter –D100000 > var.flt.merged.vcf

This command compares the mapping file (.Bam) with the reference genome (genome.fa) to identify SNPs. This command also uses a bespoke Perl script (vcfutils.pl) to filter SNP with extremely high coverage, which are less reliable since they may represent areas of high repetition. The data is then outputted as a *.vcf* (variant calling file).

The *.vcf* file details all variants identified within the merged .Bam file, in addition to several other informative values on each SNP, such as quality (Phred score, see section 3.4.4.3), read depth, allele frequency and significance (p-value). Samtools identified 98,212 SNPs with an average read depth of 42.4 reads and Phred-score quality value of 44.5 (Table 7.1). This shows that that RNA-Seq can effectively identify a large number of polymorphisms that exists between healthy donors, with high fidelity.

Table 7.1 Summary of SNP analysis of neutrophil RNA-Seq data from a single donor using Samtools "mpileup" command and bcftools software [284].

SNP attribute	Value
No. of SNPs	98,212
Average read depth per SNP	42.4
Average quality value (Phred) per SNP	44.5

7.3.1.2 Identification and visualisation of specific SNPs

In addition to the location of each SNP within a dataset, the deep read-coverage potential of RNA-Seq data provides a means of estimating the allele frequency usage of any heterozygous SNPs. The *.vcf* output file produced by the Samtools pipeline (detailed above) includes details of allele frequency for each SNP, but the large volume of data within the *.vcf* file is not suitable for manual curation of small, specific regions of interest. An alternative approach for SNP discovery and characterisation of allele frequency, is using the integrative genomics viewer (IGV) [285,286]. This software package provides a graphical user interface (GUI) for browsing the read mapping data produced by a software mapper (for example, the *.Bam* file produced by Tophat or Bowtie). In addition to graphically representing the reads mapped to each location, single polymorphisms in the data sequence are highlighted. Given sufficient coverage of reads at the SNP location, the number of reads that correspond to each allele type will indicate the overall usage of each allele.

As one example of this, analysis of neutrophil RNA-Seq data (from the Illumina platform) from two separate donors was analysed using IGV. Manual assessment identified a SNP located in the 3'-untranslated region of the β-actin gene (ACTB) that was present in both samples. The coordinates of the SNP corresponded to a known SNP with accession number rs7612 (dbSNP Build 141). This SNP is known to be quad-allelic, that is, example alleles with each of the four possible nucleotides at this

particular location have been identified in the human population. The RNA-Seq data allows us to identify that the donor exhibits a heterozygous SNP at this location whereby 54% of reads are represented by cytosine (C) and the remaining 46% by thymine (T). In contrast, a second donor exhibits a homozygous SNP at this location as 100% of the 12,410 reads which map to that location are represented by thymine (T) (Fig 7.1).

The depth of coverage and sensitivity afforded by RNA-Seq allows accurate measurements of SNP discovery and allele usage that would otherwise be difficult to achieve in conjunction with transcriptome-wide characterisation of genes via other methodologies.

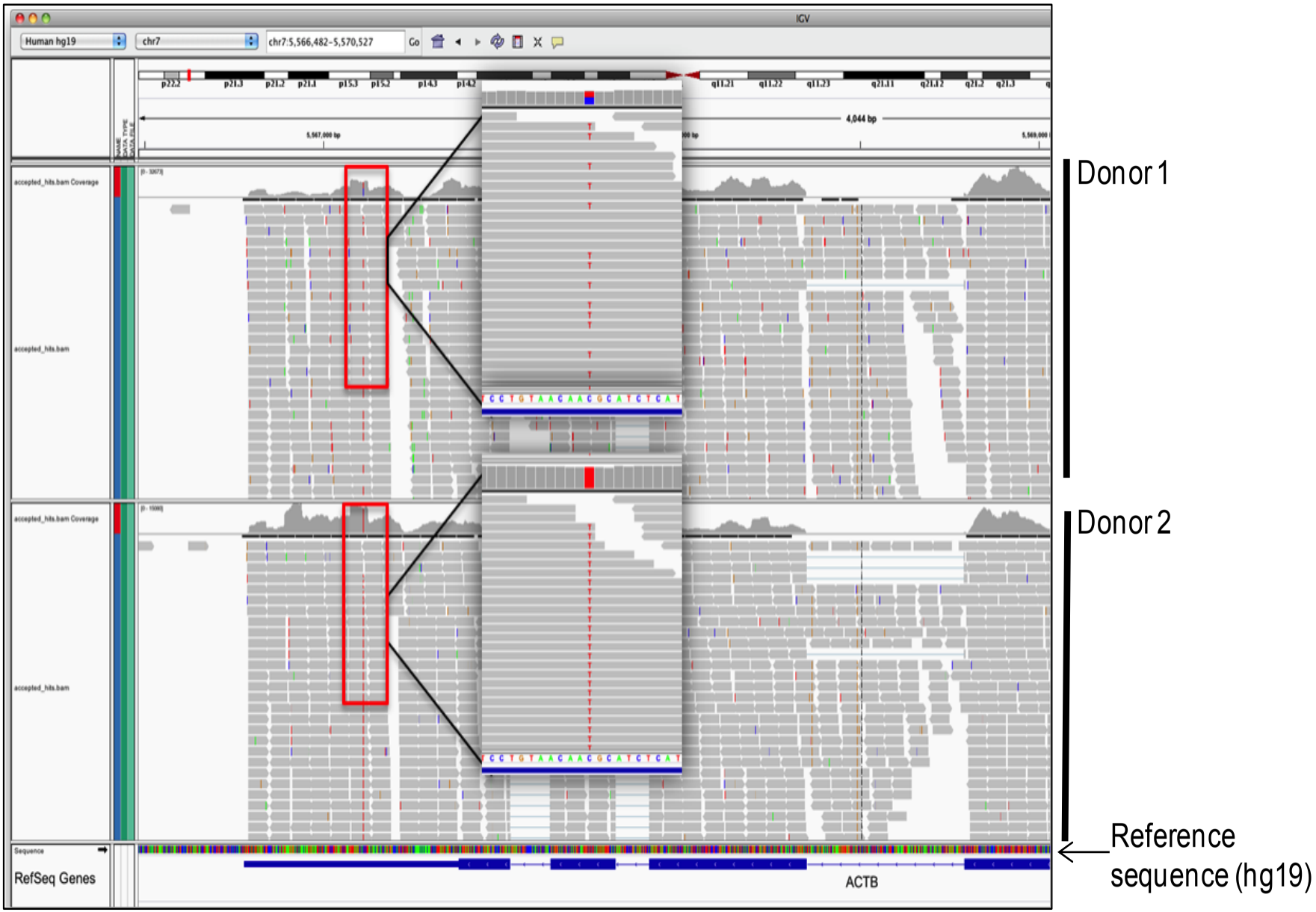

IGV
Human hg19
chr7
chr7:5,566,482-5,570,527
Go
accepted_hits.bam Coverage
accepted_hits.bam
accepted_hits.bam Coverage
accepted_hits.bam
Sequence
RefSeq Genes
ACTB
Donor 1
Donor 2
Reference sequence (hg19)

Fig 7.1 Visualisation of RNA-Seq 50bp read fragments in the Integrative Genomics Viewer (IGV) aligned to human reference genome (hg19). Two samples of untreated neutrophil RNA from two healthy donors show the presence of a single nucleotide polymorphism (SNP) at the same location within the β-actin beta gene. The human reference genome sequence used as reference (hg19-RefSeq) contains a cytosine (C) at this location (which can be seen in the lower frame of the screen shot). The first donor has a heterozygous SNP, with 54:46 of reads containing either cytosine (C), or thymine (T) respectively, at the SNP location. In contrast, the second donor has a homozygous nucleotide whereby all reads contain thymine (T) at this location.

7.3.2 Splice variant discovery

An estimated 92-94% of multi-exonic genes within the human transcriptome are subject to alternate splicing, with ~86% having a minor isoform frequency of >15% [102]. Alternate isoforms of genes not only influence the structure of the translated protein, but also have significant effect on function. For example, the gene myeloid cell leukaemia-1 (MCL-1) plays an important role in neutrophil survival through its actions as an anti-apoptotic member of the B-cell like-2 (BCL-2) family of proteins [232,287]. However, MCL-1 has been shown to undergo alternative splicing to yield 2 possible minor isoforms, MCL-1_S (short) and MCL-1_{ES} (extra-short), both of which are translated into shorter proteins with pro-apoptotic function [288,289].

7.3.2.1 Splice variant discovery in neutrophils using Cufflinks (Cuffdiff)

Since annotation software, such as Cufflinks or DESeq, provide a normalised score for each exon within a gene, it is possible to estimate the relative isoform usage for any particular gene between two or more RNA-Seq samples, provided a suitable reference sequence defining each isoform is inputted into the annotation software. Indeed, as part of its default output, Cuffdiff can quantify all splice variants within a sample

which can be used to estimate the splice variant usage between two or more samples [168].

Following annotation and DE analysis of neutrophil samples (untreated, GM-CSF and TNFα) using Cuffdiff (as previously described), isoform usage was calculated for the gene MCL-1 which has 3 known alternatively spliced isoforms: long form (MCL-1_L, NM0211960); short form (MCL-1_S, NM182763); or extra short form (MCL-1_{ES}, NM001197320). RPKM values for each isoform were extracted from the Cuffdiff output file "isoforms.fpkm_tracking.txt" and percentage usage calculated for each isoform in each treatment (Fig 7.2)

Values for MCL-1_L decreased in both GM-CSF and TNFα treated samples compared to control. Consequently, the relative levels of MCL-1_S and MCL-1_{ES} were increased, with the greatest increase seen in MCL-1_{ES} isoform in GM-CSF-treated neutrophils (untreated 4%, GM-CSF 8.3%).

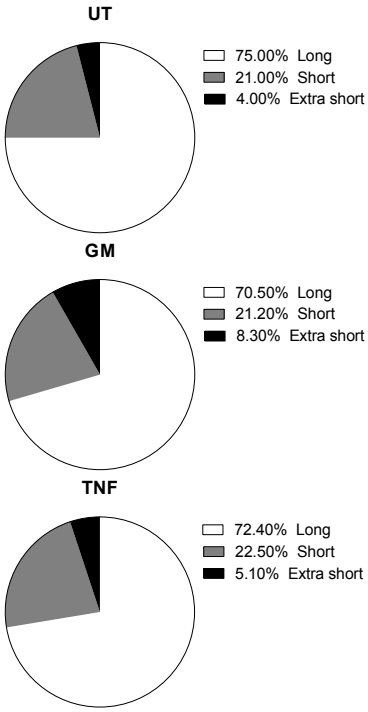


Fig 7.2 MCL-1 isoform usage in neutrophils following cytokine treatment. Neutrophils were treated with (or without) GM-CSF (5 ng/mL) or TNFα (10 ng/mL) for 1 h. RNA was sequenced on the Illumina platform and quantified by Tophat/Cufflinks. RPKM values calculated by Cuffdiff for the 3 isoforms of MCL-1 (long form, short form or extra short form). Values represent mean of 3 biological replicates.

7.3.2.2 Visualisation of isoform usage

As described above, RNA-Seq mapping data can be analysed by GUI based analysis programs, such as IGV [285,286], to provide a visual representation of the mapping results. In addition to SNPs, this method of analysis can also be employed to visualise the isoform usage or splice junction sites at a particular area of interest. Fig 7.3 demonstrates how the reads mapping to the MCL-1 gene can be visualised. Reads that originate from all three MCL-1 isoforms are evident and the depth of coverage at splice junctions can be assessed (Fig 7.3). This method of visualising mapped reads can also be used to identify other transcriptional features,

such as transcriptional start sites (TSS), by assessing the location of reads in relation to the predicted TSS as seen in the reference sequence.

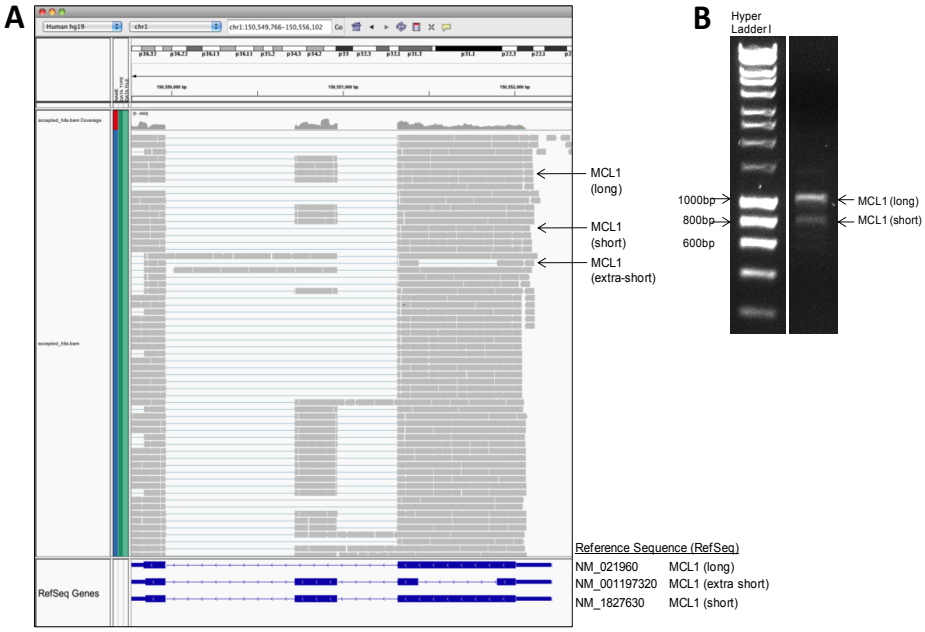


Fig 7.3 Alternative splicing of the MCL-1 gene visualised in the Integrative Genomics Viewer (IGV) and by standard gel PCR. (A) Mapped reads from neutrophil RNA-Seq data (Illumina platform) are displayed using IGV. Reads are represented as grey blocks whilst a blue line connects reads that span exon boundaries. Histogram at top of panel represents the depth of coverage at each base location. Human reference sequence (hg19-RefSeq) isoform structure represented by blue bars at base of panel. (B) PCR analysis of neutrophil RNA using primers designed to detect all three isoforms of MCL-1. PCR bands correspond to MCL-1L (1053bp) and MCL-1S (805bp). No product corresponding to MCL-1ES (594bp) was detected. Gel data is representative of at least 4 biological donors (collected and reproduced with permission by D. Chiewchengchol 2013).

For comparison, gel PCR was used to determine levels of MCL-1_L, MCL-1_S and MCL-1_{ES} in untreated neutrophils. Primers were designed to amplify full length transcripts of all three MCL-1 isoforms. Following cDNA amplification, samples were run on an agarose gel. Levels of MCL-1_{ES} were undetectable in samples from at least 4 biological donors (Fig 7.3B) (data collected and reproduced with permission by D. Chiewchengchol)

The changes seen in isoform abundance in MCL-1 following neutrophil stimulation are small, but highlight the accuracy and sensitivity available by using a bioinformatics approach to detect isoform levels.

7.4 Discussion

The above analyses demonstrate the capacity of raw RNA-Seq data to be re-analysed by different software packages to extract novel results regarding alternative genetic features that complement gene expression values.

SNP discovery can inform on important structural changes in crucial protein-coding areas or even in non-coding areas which are increasingly recognised as important determinants of gene expression profiles, whilst also being implicated in several diseases [290–294]. Indeed, a recent study identified several SNPs located within (or adjacent to) functional elements in human neutrophils from patients with juvenile idiopathic arthritis (JIA) [295]. Whilst a fully comprehensive study of SNP and other polymorphisms in neutrophils would require a different methodology, specifically a genome-wide sequencing approach, the ability to use RNA-Seq data to analyse SNP located in mRNA transcripts provides an additional benefit over array-based analyses.

Similarly, the accuracy of RNA-Seq to quantify absolute values of transcript levels allows gene expression values to be quantified in terms of all associated splice variants. When applied to neutrophil expression of MCL-1, each of the 3 known splice variants were identified and changes to the ratios of expression were found following stimulation of neutrophils. Differential usage of splice variants in neutrophils has previously been

shown for several genes, including 5-lipoxygenase [296], glucocorticoid receptor[297] and the pantetheinase family of genes [298]

However, whilst this approach using Cufflinks is informative and of great use for discovering changes in isoform abundance, it is important to appreciate that this form of analysis is semi-quantitative. Since most isoform sequences differ only in a small portion of their sequence, reads that map to common areas of the reference sequence cannot be definitively assigned to any one isoform, thus only reads that map to an isoform-unique portion of their sequence can be accurately quantified. Paired-end sequencing can decrease the number of reads un-assignable to a particular transcript by providing a pair of read fragments known to originate from the same transcript and lying a known distance from each other. However, ultimately, mapping software must apply some degree of estimation when assigning a level of significance to splicing events [145].

In summary, the software-based analyses described in this Chapter demonstrate an extension to the bioinformatics pipeline presented in Chapter 3. They provide informative data that could complement a concurrent global gene expression analysis. In combination, these methods represent a set of robust analyses for a comprehensive study of the neutrophil transcriptome.

Chapter 8: Conclusions

8.1 Overall conclusions and outcomes

Neutrophils constitute the largest cellular component of the immune system. It is now widely appreciated that neutrophils are central to the immune response and are capable of regulating both the innate and adaptive immune systems through the expression and release of several immune regulators [4,5]. Whilst the functional mechanisms of neutrophils in both health and disease are well characterised, the molecular changes that underlie neutrophil regulation were poorly defined. Transcriptomics represents an attractive, analytical approach to neutrophil gene expression by providing a mechanism of quantifying the entire population of transcripts at a particular time point or following stimulation. A limited number of studies have analysed the global transcriptional profile of neutrophils, with the majority looking at gene expression changes during neutrophil maturation, or over several time points following stimulation by a single cytokine [181,245,258,299]. Moreover, at the outset of this research, the neutrophil transcriptome had yet to be characterised using modern RNA-Seq technology. The benefits of RNA-Seq over established microarray technology are numerous, not least the ability to garner information on various genetic features from a single experimental run. Despite the improvements of RNA-Seq technology, the corresponding bioinformatic software remains eclectic in their functionality and ease of use. Such that,

published bioinformatic methods are often disparate and difficult to compare or judge equally. Indeed, the bioinformatic community has yet to decide on a set of software tools or quantification techniques that represent the best practices for analysis of NGS data.

The aims of this research were to develop a robust, pipeline of methods and bioinformatic analyses using open-source or commercially available software that could accurately measure the gene expression profiles of neutrophils under different inflammatory conditions. This pipeline would then be used to fully quantify neutrophil gene expression following stimulation with a variety of inflammatory mediators, or following two commonly used neutrophil-isolation methods.

Development of the bioinformatic pipeline in Chapter 3 explored the relative merits of both the SOLiD and Illumina sequencing platforms, in addition to paired-end and single-end sequencing techniques. Despite both platforms and sequencing techniques correlating well to qPCR data during validation experiments, the higher mapping rate and read quality achieved by Illumina sequencing platform determined it as the platform of choice for future experiments. Whether the lower quality values achieved with the SOLiD platform were due to: the added complexity associated with paired-end sequencing; technical or human error; or were attributable to the technology platform as a whole, is unclear. However, it is unlikely that a single-end sequencing experiment using SOLiD technology could

improve upon the read quality values and mapping rates achieved by the Illumina platform.

Quantification of mapped reads is predominantly achieved in one of two ways, either using the raw number of reads assigned to each gene in a count-based approach, or by further transforming the raw counts into a value normalised to both the size of the read library, and the length of the gene being quantified (i.e. an RPKM value). The choice of quantification methods for neutrophil gene expression values had a significant effect on the number of DE genes between samples; this effect was even seen between two count-based approaches (DESeq and edgeR). The greater number of DE genes identified following count-based methods was likely due to the techniques used by the software to model the variation within the read population. Although this approach is known to estimate biological variance poorly, and suffer from over-sensitivity [163,300,301], it has been employed for several RNA-Seq studies and remains a popular choice for differential expression studies [302–305]. However, the ability to directly compare two separate genes within the same dataset (or between datasets), and the compatibility with downstream software offered by the Cufflinks quantification route, led to the count-based quantification method being discounted from the final bioinformatic pipeline in favour of Cufflinks/Cuffdiff.

Downstream analysis of gene expression data is an increasingly popular area of bioinformatic analysis [199], since many sequencing service providers

also offer basic gene expression quantification, non-bioinformatically trained research labs are increasingly reliant on downstream analytics to extract meaningful data from large amounts of raw gene-expression data. Many of the most popular software packages utilise large databases of canonical biological data to model the raw data against. Software of this kind are often designed for ease of use by non-bioinformaticians, but are only available via a commercial licence. Whilst this excludes some researchers from the best available software purely on financial grounds, the benefit of commercial analysis-software is two-fold. Firstly, the technical support and usability of software is often far superior to open-source equivalents. Secondly, the professional curation of the information databases is consistently maintained, meaning that the data resources are constantly up-to-date and comprehensive, whilst also providing validation of all canonical interactions with supporting publications. For these reasons, the commercially available pathway analysis software IPA [220] was employed as part of the bioinformatic pipeline. This provided invaluable capacity in downstream analysis that was not feasible using open-source or freely available software. However, other downstream analyses such as hierarchical clustering and heat-map generation were achieved using open-source software. The pipeline described in Chapter 3 therefore represents a robust set of tools for analysing the gene expression profiles of neutrophils that is both comprehensive in capacity and modest in technical-ability requirements.

In Chapters 4 and 5, the bioinformatic pipeline was employed to investigate the effect of inflammatory cytokine stimulation on neutrophil gene expression. The similar priming effects of GM-CSF and TNFα on neutrophils are well characterised, but a global comparison of the molecular changes following priming has not previously been studied. Analysis revealed that despite similar expression in common genes, each cytokine induced expression of discrete gene sets that were as a consequence of differential transcription factor activation. This led to the discovery that the delay in neutrophil apoptosis – seen following stimulation with either cytokine – was regulated by STAT activation in the case of GM-CSF and NFκB activation in TNFα stimulation. The discovery that cytokines regulate neutrophil function via differential expression of genes and activation of signalling pathways has important implications for the study of neutrophil-dysfunction in inflammatory disease. Not least for providing a novel set of biomarkers that can identify the predominant cytokines that may be driving inflammation in different patients.

Analysis of cytokine-induced changes in neutrophils was expanded in Chapter 5 to cover other cytokines associated with inflammation. In addition, the effect of multiple cytokine stimulation was investigated. IL-1β, IL-6 and IL-8 are known inflammatory mediators; but their effect on neutrophils is less defined. IL-6 and IL-1β are crucial activator of many immune cell-types, while IL-8 is a strong chemo-attractant of neutrophils. However, all these cytokines were found to have very little effect on

neutrophil gene expression. Whilst it comes as no surprise that the chemo-attractive capacity of IL-8 is independent of *de novo* gene expression – since localising to the site of inflammation requires rapid execution – the inability of IL-6 and IL-1β to induce gene expression by 1 h was less predicted. In contrast, IFNγ, G-CSF and dual treatment with GM-CSF and TNFα induced significant changes in gene expression, and differential activation of signalling pathways. These results highlight the specific functional and molecular changes induced in neutrophils by similar inflammatory mediators and reveals how stimulation by different cytokines can alter the neutrophil phenotype thus potentially altering how they respond to later stimulation and/or regulate other cells of the immune response.

Fig 8.1 summarises the multiple genetic characteristics that have be quantified using the bioinformatics pipeline developed herein (Fig 8.1A) and the differences in neutrophil phenotypes resulting from different cytokine stimulation (Fig 8.1B).

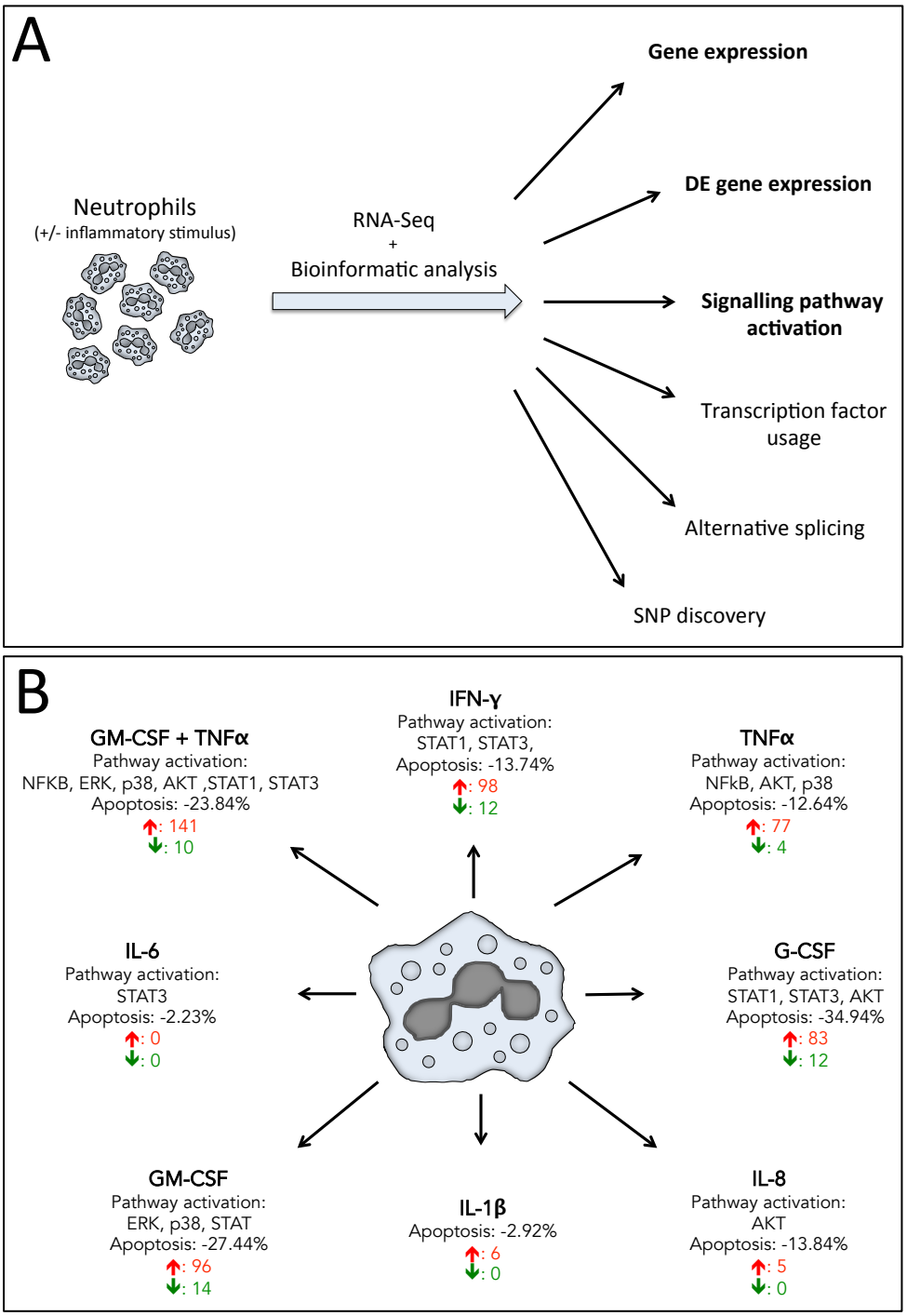


Fig 8.1 Summary of neutrophil bioinformatics pipeline capacity and findings. (A) Analysis of neutrophil RNA by RNA-Seq and neutrophil bioinformatics pipeline has provided a mechanism to quantify several genetic characteristics such as raw gene expression or transcription factor activation from a single RNA-Seq experiment. Characteristics studied in greatest detail are shown in bold. (B) Schematic summary of the phenotypic changes induced in neutrophils by different inflammatory cytokines, including: specific signalling pathway activation; change in levels of overnight apoptosis (% change compared to untreated); and number of genes with a significant increase (red arrow) or decrease (green arrow) in gene expression compared to control (q<0.05, FDR 5%).

With the advent of more sensitive techniques to isolate neutrophils from whole blood, greater emphasis has been placed on the appropriateness of established isolation methods for highly sensitive assays such as proteomics or transcriptomics. In Chapter 6, the impact of neutrophil isolation methods, and levels of non-neutrophil contamination on neutrophil gene expression were investigated. As predicted, magnetic bead isolation resulted in a greater purity of neutrophils, the lack of contaminating cells was confirmed by gene expression analysis for non-neutrophil transcripts which were lower in bead isolations than in samples isolated by Polymorphprep™. The use of two healthy donors with disparate levels of non-neutrophil contamination provided a mechanism for highlighting the impact low and high contamination has on the overall gene expression profile of neutrophils (either untreated or following cytokine stimulation). However, despite a purer population of neutrophils following bead isolation, RNA-Seq analysis revealed that there was greater genetic heterogeneity between donors than when neutrophils were isolated by Polymorphprep™. This difference was largely due to a subpopulation of neutrophils which were enriched in Donor 2 but not Donor 1. These cells were likely the source of LDG-associated transcripts which were also elevated in Donor 2. These results were surprising since LDG associated genes have only previously been identified in the context of inflammatory disease [52,77,81,272]. What is still unclear is if there is any association between the high levels of eosinophils and the high levels of

LDGs both seen in Donor 2. Importantly, this research has identified that bead isolation of neutrophils can exhibit greater heterogeneity between donors due to an enrichment of neutrophil sub-populations. Furthermore, a lower overall yield and greater cost per isolation for magnetic bead isolations suggests that the increased level of purity achievable must be weighed up against these factors when determining a suitable method of isolation.

8.2 Future directions

8.2.1 Future research

Both the methods described here, and the results identified during the research would benefit from further development and analysis. Aspects of the bioinformatic pipeline that were developed but not applied to all datasets were summarised in Chapter 7, these include methods for determining differential splice usage and SNPs discover. Full characterisation of SNPs is likely only useful for situations where healthy neutrophils are being compared to those from inflammatory disease. But an analysis of differential splice usage in neutrophil genes following cytokine stimulation could be very informative.

The benefit of developing a robust bioinformatic pipeline of analysis is that it can readily be applied to several different situations to efficiently analyse the neutrophil transcriptome under these different conditions.

Given the interesting results identified following dual stimulation of neutrophils it would also be of interest to extend this research to investigate the effect of multiple other cytokines on neutrophil gene expression. For example several cytokines have been shown to have synergistic effects when regulating the immune system in combination [306], including TNFα with IFNγ [307,308] or IL-12 [309] but their effect on neutrophils is less known.

Equally, whilst a 1 h time point was entirely adequate for studying initial gene expression in neutrophils following stimulation, results indicate that neutrophil phenotype may be differentially altered dependent on the initial stimulus. Hence, neutrophils ability to respond to a later, secondary signal may be significantly different and would represent an interesting area of research. Indeed a more comprehensive study investigating gene expression over several hours (looking at several time points) would also be of interest and would inform on the speed and magnitude of differential gene expression by different cytokine stimulation.

However, perhaps the most informative of future research would be to investigate the gene expression changes in neutrophils following non-sterile stimulation. Whilst directly incubating neutrophils with microorganisms would be unfeasible for RNA-Seq studies – due to the contamination of neutrophil RNA with microorganism RNA – non-sterile inflammatory conditions could be simulated by direct receptor agonists.

Activation of neutrophils by exogenous molecules is mediated by receptors such as TLRs, NLRs and C-lectin like receptors [5]. Whilst many of the phenotypic changes in neutrophils following activation by bacterial products and inflammatory cytokines are similar, it would be of interest to investigate if the molecular changes induced by bacterial products are distinct from those seen following cytokine stimulation. Indeed, the discovery of specific signalling pathways or target molecules that could regulate neutrophil activation following endogenous stimulation without compromising neutrophils ability to respond to exogenous signals would be of great interest for studies into autoimmune disease where reducing the activation levels of the immune system yet maintaining host defence is of upmost importance.

8.2.2 Future of RNA-Seq

The future direction of RNA-Seq analysis in general is set to increase in magnitude and ubiquity. With the development of 4th generation sequencers reducing cost and increasing the speed at which samples can be sequenced, RNA-Seq analysis will undoubtedly become a standard practice in many research laboratories and clinical environments. Whilst this can only be of benefit to scientific research and modern day healthcare, storage of the vast amounts of data will undoubtedly move towards cloud based solutions. But this raises its own issues such as long term storage costs. All but the largest cloud-based companies are susceptible to commercial failure; whilst the financial ebb and flow nature of scientific

funding may mean that the indefinite, secure, storage of valuable biological data may be a luxury of the past.

In summary this research has gone some way to reveal the molecular changes in neutrophils under different conditions by the development and employment of a robust set of bioinformatic tools. These tools have uncovered a greater regulation in gene expression by neutrophils than was perhaps appreciated. Whilst several avenues of further research have been directly highlighted by this research the methods described here can ultimately be employed for a variety of other studies involving RNA-Seq and continue to uncover important scientific discoveries.

Appendix

Table A.1 List of bioinformatic software and versions.

Bioinformatic software versions	
Bioinformatic software	**Version**
Bowtie	2.0.07
Tophat	1.2.1-1.4.1
Cufflinks	2.02
Samtools	0.1.18
IPA	n/a
IGV	2.2.7
Microsoft Office	2011 edition
R	2.15.2
EdgeR	3.0.8
DESeq	1.10.1
cummeRbund	2.6.2

Table A.2 Details of computer hardware used for analysis.

Computer system (analysis Mac)	**Mac Pro**	**iMac**
Operating system	Mac OSX 10.7-10.8	Mac OSX 10.85
Computer processing unit (CPU)	8 x 2.4GHz Intel Core i5	4 x 3.4GHz Inter Core i7
Random access memory (RAM)	16GB DDR RAM	32GB DDR3 RAM (1600 MHz)
Hard drive memory (HDD)	4 x 1TB – 7200 Serial ATA HDD	3TB-7200rpm Serial ATA HDD
Graphical processing unit	ATI Radeon HD 5770 1GB	NVIDIA GeForce GTX 680MX 2048 MB

```
#!/bin/sh
#$ -cwd -V -pe smp 4 -l h_rt=4:0:00
cufflinks
-b\
/home/hbt/volatile/iGenomes_bowtie2_indexes/Homo_sapiens/UCSC/hg19/Sequence/Bowtie2Index/genome.fa \
-p 4 --max-bundle-frags 100000000 -q -o ./path/to/folder --no-update-check \
-G
/home/hlwright/volatile/Homo_sapiens/UCSC/hg19/Annotation/Genes/genes.gtf
\
$1 \
```

Fig A.1 Example of command script for use on HPC. Script shows commands for running cufflinks to annotate a *.bam* file using a reference genome (*.fa*) and transcriptome (*.gtf*).

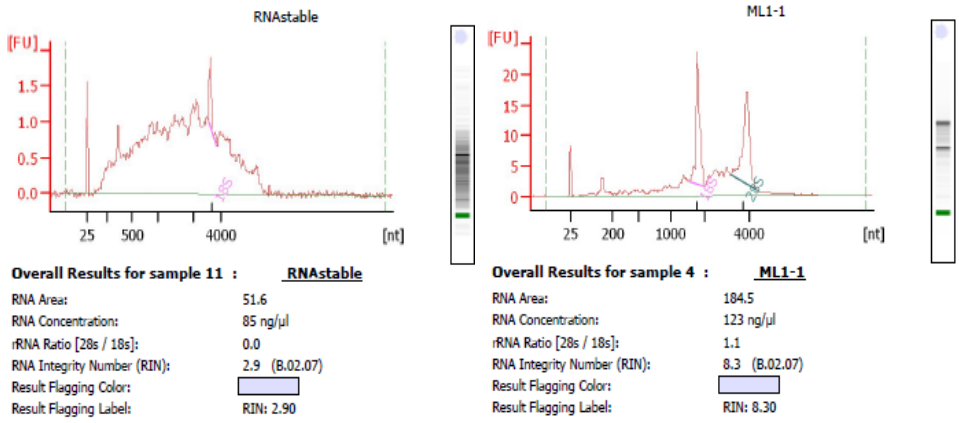


Fig A.2 Example of Agilent Bioanalyser output data. Figure shows two examples of RIN results. (Left) Example of RNA with poor integrity (RIN = 2.9). (Right) example of RNA with high integrity (RIN = 8.3). Major peaks in high-integrity sample relate to 16S and 28S ribosomal RNA population. Integrity is calculated based on the ration between both peaks.

Table A.3 Table listing the details of blood donors and method of neutrophil isolation Polymorphprep™ (P), magnetic bead isolation (B).

Donor No.	Age (approx.)	Sex	Date of donation/sample preparation	Neutrophil isolation method
1	25-35	M	31-1-11 10-12-12 28-8-13	P P P/B
2	55-65	M	30-4-12	P
3	25-35	F	22-1-13	P/B
4	25-35	F	28-1-13	P
5	45-55	M	21-8-13	P
6	25-35	F	17-9-13	P

Table A.4 Table details the RNA integrity and concentration for samples analysed by RNA-Seq. Table also lists the number of raw reads produced per sample during sequencing. All samples sequenced by SOLiD platform (50 + 35 bp paired-end reads).

Donor	Sample name	No. of raw reads	No. of raw reads (reverse)	RIN	RNA (ng/mL)
TBH	Untreated	127885988	127885989	8.2	95
	TNFα	69005645	69005646	7.0	134
	GM-CSF	75544747	75544748	7.5	164

Table A.5 Table details the RNA integrity and concentration for each RNA sample analysed by RNA-Seq. Table also lists the number of reads produced per sample during sequencing. All samples sequenced by Illumina platform (50 bp single-end reads).

Donor	Sample name	No. of raw reads	RNA Integrity number (RIN)	RNA (ng/mL)
TBH	UT	66,552,453	7.3	218
	GM-CSF	64,445,900	8.4	188
	TNFα	65,625,666	8.8	196
EWS	UT	50,794,935	7.6	64
	GM	48,015,235	6.9	82
	TNFα	47,326,637	4.1	37
	IFNγ	48,514,806	6.6	84
	IL-1B	50,565,105	6.0	86
	0h	48,110,305	6.4	76
	G-CSF	47,994,097	7.3	60
MB	GM-CSF+TNFα	54,548,520	7.5	144
	IL-8	58,462,542	8.0	126
	UT-Poly	46,444,696	8.1	124
	GM-Poly	49,516,666	8.0	137
	TNF-Poly	56,689,600	7.8	140
	UT-Bead	59,067,617	7.2	84
	GM-Bead	61,595,651	7.0	117
	TNF-Bead	49,785,361	6.4	88
TBH	UT-Poly	43,862,843	7.7	44
	GM-Poly	44,465,453	8.6	46
	TNF-Poly	44,391,372	7.4	36
	UT-Bead	42,044,643	8.2	40
	GM-Bead	43,095,188	8.6	27
	TNF-Bead	42,044,643	8.2	42
WN	IFNα	63,186,628	8.1	90
CA	UT	69,311,246	5.7	59
	IFNγ	75,916,629	6.6	66
	IL-1B	79,791,533	7.3	74
	G-CSF	74,729,800	6.9	75
	IL-8	69,593,692	6.9	38
	GM-CSF+TNFα	72,530,839	7.0	58
TS	UT	72,168,424	7.5	131
	IFNγ	65,554,973	7.7	121
	IL-1B	121,354,425	7.4	142
	G-CSF	72,434,747	6.9	98
	IL-8	98,504,872	7.1	135
	GM-CSF+TNFα	72,530,839	7.1	102

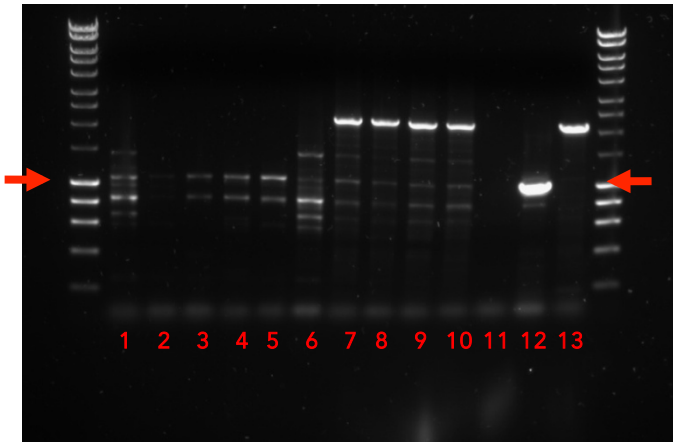


Fig A.3 PCR gel of neutrophil cDNA products following amplification with Mcl-1 primers, samples isolated using either TRIzol® method or RNeasy kit with or without additional DNA digestion step. Mcl-1 band highlighted by red arrows, lane numbers highlighted in red.

Loading legend:

1. **0h Trizol**
2. **0h RNeasy**
3. **2h UT**
4. **2h TNF**
5. **2h GM-CSF**

RNeasy

6. **0h Trizol - Undigested**
7. **0h RNeasy**
8. **2h UT**
9. **2h TNF**
10. **2h GM-CSF**

RNeasy (undigested)

11. **Negative control (H_2O)**
12. **Positive control (K562 - CML cell line)**
13. **Genomic DNA control (HeLa cell line)**

Table A.6 List of 44 GO-terms significantly enriched by genes which are DE between GM-CSF and TNFα treated neutrophils. GO-terms relating to cell death and/or apoptosis are shown in bold. Significance (q-value) calculated using a 5% false discovery rate (FDR)

GO Term	GO Category	No of Genes	FDR (q-value)
GO:0042981	**regulation of apoptosis**	**58**	**3.17E-06**
GO:0043067	**regulation of programmed cell death**	**58**	**4.61E-06**
GO:0010941	**regulation of cell death**	**58**	**5.21E-06**
GO:0006952	defense response	52	2.81E-07
GO:0006955	immune response	52	1.51E-05
GO:0016265	**death**	**51**	**1.84E-04**
GO:0010604	positive regulation of macromolecule metabolic process	51	2.49E-02
GO:0009611	response to wounding	50	1.43E-08
GO:0008219	**cell death**	**50**	**3.72E-04**

GO:0031328	positive regulation of cellular biosynthetic process	49	2.23E-04
GO:0009891	positive regulation of biosynthetic process	49	3.48E-04
GO:0012501	**programmed cell death**	**48**	**1.78E-05**
GO:0006915	**apoptosis**	**47**	**3.16E-05**
GO:0051173	positive regulation of nitrogen compound metabolic process	46	6.20E-04
GO:0010557	positive regulation of macromolecule biosynthetic process	46	9.51E-04
GO:0006357	regulation of transcription from RNA polymerase II promoter	46	1.58E-02
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	42	9.88E-03
GO:0010628	positive regulation of gene expression	41	4.15E-03
GO:0045941	positive regulation of transcription	40	5.06E-03
GO:0006954	inflammatory response	39	3.60E-09
GO:0019220	regulation of phosphate metabolic process	36	6.54E-03
GO:0051174	regulation of phosphorus metabolic process	36	6.54E-03
GO:0007243	protein kinase cascade	35	3.18E-05
GO:0042325	regulation of phosphorylation	35	7.09E-03
GO:0045893	positive regulation of transcription, DNA-dependent	34	2.99E-02
GO:0051254	positive regulation of RNA metabolic process	34	3.63E-02
GO:0043066	**negative regulation of apoptosis**	32	4.10E-04
GO:0043069	**negative regulation of programmed cell death**	32	5.58E-04
GO:0060548	**negative regulation of cell death**	**32**	**5.94E-04**
GO:0001775	cell activation	30	4.67E-05
GO:0045321	leukocyte activation	29	4.09E-06
GO:0001817	regulation of cytokine production	24	1.92E-05
GO:0046649	lymphocyte activation	23	4.93E-04
GO:0006916	**anti-apoptosis**	**22**	**3.49E-03**
GO:0001819	positive regulation of cytokine production	15	1.26E-03
GO:0050867	positive regulation of cell activation	15	1.62E-02
GO:0032496	response to lipopolysaccharide	14	1.20E-03
GO:0002237	response to molecule of bacterial origin	14	4.41E-03
GO:0002696	positive regulation of leukocyte activation	14	4.57E-02
GO:0031349	positive regulation of defence response	12	2.70E-02
GO:0051100	negative regulation of binding	11	2.27E-02
GO:0043433	negative regulation of transcription factor activity	10	1.52E-02
GO:0043392	negative regulation of DNA binding	10	4.41E-02
GO:0032675	regulation of interleukin-6 production	9	2.13E-02

Table A.7 Gene expression values (RPKM) for 25 cytokines/chemokines genes differentially expressed by neutrophils following treatment with a range of cytokines/chemokines, for heatmap of values see Fig 5.8.

Gene name	UT	GM-CSF	TNFα	GM+TNF	IFNg	G-CSF	IL-1B	IL-8	IL-6
Bmp6	2.76	3.70	3.08	2.57	2.07	2.03	1.99	3.21	2.55
Ccl20	0.17	0.27	12.85	28.68	0.11	0.33	1.14	0.32	0.42
CcL3	4.53	13.98	206.16	113.53	2.88	4.04	41.54	2.25	7.73
Ccl4	33.23	38.86	1858.67	980.83	21.61	22.41	281.55	17.10	30.56
Ccl5	14.30	8.58	8.50	10.50	19.06	16.98	18.56	9.98	17.11
Cxcl1	243.48	1869.48	639.56	2804.08	130.20	273.83	660.67	288.21	289.16
Cxcl14	1.86	2.80	1.92	2.74	1.85	2.18	2.15	1.85	3.26
Cxcl2	9.92	69.82	52.27	155.76	6.19	12.83	40.53	13.36	13.88
Cxcl3	0.73	2.77	4.71	10.73	0.31	1.05	2.30	1.16	0.51
Cxcl6	3.97	2.53	2.86	1.55	2.07	4.37	4.74	2.58	4.96
Cxcl9	0.00	0.00	0.00	0.00	2.84	0.02	0.00	0.01	0.00
Il1a	0.83	50.49	56.14	652.64	1.57	3.16	4.18	3.18	0.30
Il1b	369.34	6410.42	3614.27	17674.90	864.85	1071.83	890.18	1103.73	302.74
IL1RN	81.50	1106.43	2599.95	7499.70	191.37	716.42	153.22	82.04	109.22
Il8	2362.95	12539.90	6327.35	17734.70	840.31	961.66	4031.00	3964.57	1905.32
Lif	0.98	0.23	1.02	1.69	0.41	0.59	1.27	3.25	1.76
Ltb	108.54	99.99	136.69	79.53	118.20	106.45	141.29	83.71	155.54
Osm	104.43	1120.90	135.06	1612.75	94.93	163.21	157.90	238.22	167.48
Pf4	1.04	2.24	0.86	0.80	0.81	0.91	1.41	0.68	0.98
Prmt2	14.85	13.26	13.20	9.54	16.11	17.17	15.41	13.09	14.34
Tnfsf10	51.59	73.52	37.17	34.56	126.96	80.16	47.41	47.40	51.52
Tnfsf12	12.76	9.87	9.91	7.42	16.60	15.02	16.13	11.00	17.30
Tnfsf14	76.46	33.72	62.24	14.49	20.78	22.77	61.56	81.74	65.29
Tnfsf15	0.96	4.18	0.74	3.31	0.38	4.37	1.54	1.36	0.55
Tnfsf8	9.08	20.46	7.17	10.85	6.03	7.01	8.71	12.48	5.73

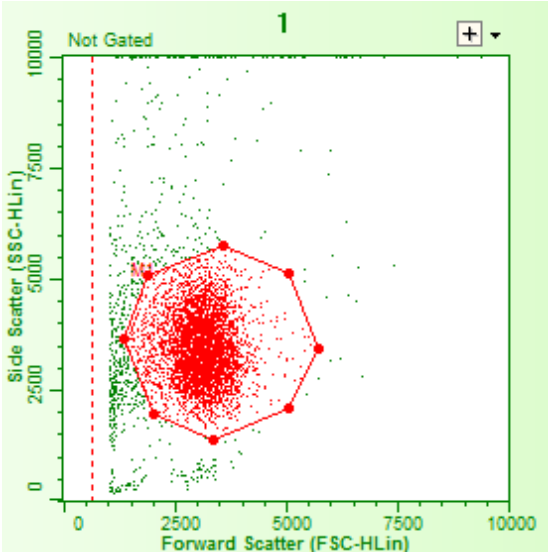


Fig A.4 Example of flow cytometry gating on forward-scatter (x-axis) and side-scatter (y-axis) , used to filter out non-granulocyte cells and cellular debris from further analysis.

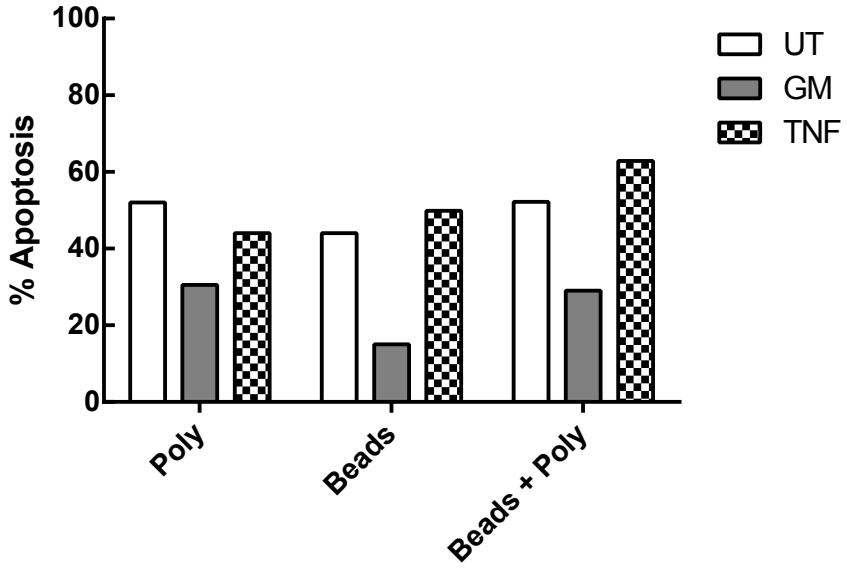


Fig A.5 Neutrophils apoptosis following overnight (18 h) incubations with either 5 ng/mL GM-CSF (GM) (grey bars), 10 ng/mL TNFα (TNF) (checkered bars) or remained untreated (UT) (white bars). Neutrophils were isolated by either Polymorphprep™ (Poly), magnetic bead preparation (Beads) or magnetic bead preparation followed by a 30 min centrifuge on Polymorphprep™ (Bead + Poly). Levels of neutrophil apoptosis were measured by annexin V, propridium iodide staining.

Table A.8 List of data sets used in Cuffdiff comparisons in Chapter 6.

Figure/table of reference	Sample list 1	Sample list 2
Table 6.3	UT-poly-donor-1 GM-poly-donor-1 TNF-poly-donor-1 UT-poly-donor-2 GM-poly-donor-2 TNF-poly-donor-2	UT-beads-donor-1 GM-beads-donor-1 TNF-beads-donor-1 UT-beads-donor-2 GM-beads-donor-2 TNF-beads-donor-2
Fig 6.11	UT-poly-donor-1 UT-poly-donor-2 GM-poly-donor-1 GM-poly-donor-2 TNF-poly-donor-1 TNF-poly-donor-2	UT-beads-donor-1 UT-beads-donor-2 GM-beads-donor-1 GM-beads-donor-2 TNF-beads-donor-1 TNF-beads-donor-2
Table 6.4	UT-poly-donor-1 GM-poly-donor-1 TNF-poly-donor-1	UT-beads-donor-1 GM-beads-donor-1 TNF-beads-donor-1
Table 6.5	UT-poly-donor-2 GM-poly-donor-2 TNF-poly-donor-2	UT-beads-donor-2 GM-beads-donor-2 TNF-beads-donor-2
Table 6.6	UT-poly-donor-1 GM-poly-donor-1 TNF-poly-donor-1	UT-poly-donor-2 GM-poly-donor-2 TNF-poly-donor-2
	UT-beads-donor-1 GM-beads-donor-1 TNF-beads-donor-1	UT-beads-donor-2 GM-beads-donor-2 TNF-beads-donor-2

Table A.9 Poly vs beads in both donors. Number of significant genes in samples from each isolation method.

	Donor 1	Donor 2
Number of gene significantly differentially expressed	63	282
Number of significant genes with higher RPKM in Polymorphprep™ samples	53 (84.1%)	190 (67.3%)
Number of significant genes with higher RPKM in Bead samples	10 (15.9%)	92 (32.7%)

Table A.10 List of genes significantly regulated between isolation methods in neutrophils from Donor 2 (high contamination). Table shows all 92 genes with a significantly higher RPKM values in neutrophils isolated using magnetic bead (Beads) than neutrophils isolated by Polymorphprep™ (Poly) using sample from Donor 2. Significance (q-value) as calculated by Cuffdiff adjusted for 5% false discovery rate (FDR) by Benjamini-Hochberg correction for multiple-testing. N=3 paired technical replicates.

Gene Name	Poly (RPKM)	Beads (RPKM)	Fold change (log2)	q-value	ΔRPKM
DEFA1	239.933	1082.69	2.174	4.75E-02	-842.757
LCN2	89.952	447.161	2.314	2.45E-02	-357.209
CAMP	32.519	138.676	2.092	2.93E-02	-106.157
BPI	24.068	117.488	2.287	2.05E-02	-93.42
OLFM4	16.438	87.792	2.417	1.36E-02	-71.354
CEACAM8	12.433	65.236	2.391	9.63E-03	-52.803
DEFA4	9.968	57.406	2.526	1.26E-03	-47.438
MS4A3	8.837	45.919	2.377	1.54E-02	-37.081
MMP8	9.436	41.732	2.145	3.68E-02	-32.297
AZU1	7.485	38.436	2.36	4.26E-03	-30.951
ELANE	5.851	34.581	2.563	1.00E-03	-28.731
RETN	5.085	28.648	2.494	5.32E-03	-23.563
CEACAM6	5.968	25.207	2.079	2.95E-02	-19.239
H1F0	3.009	21.615	2.845	2.19E-04	-18.605
C13orf15	5.171	23.425	2.18	1.16E-02	-18.254
CTSG	2.296	20.538	3.161	7.90E-06	-18.242
PRTN3	2.929	19.16	2.71	2.88E-04	-16.231
PNP	2.584	16.629	2.686	3.74E-02	-14.045
GJB6	1.717	14.542	3.082	1.06E-03	-12.825
MPO	1.643	12.589	2.938	7.69E-05	-10.946
EHD4	2.231	11.691	2.39	5.17E-03	-9.46
DEFA5	1.824	9.096	2.318	1.45E-02	-7.271
HS3ST3B1	0.901	7.983	3.148	1.56E-03	-7.082
SNAPC1	2.382	9.253	1.958	4.28E-02	-6.871
RAB13	2.113	8.268	1.968	3.50E-02	-6.155
PTGES	0.828	5.651	2.772	1.81E-02	-4.824
C20orf27	1.399	6.109	2.127	1.45E-02	-4.711
ZNF277	1.57	6.216	1.985	4.10E-02	-4.646
BEX1	0.947	5.21	2.459	3.06E-03	-4.262
RRM2	1.113	4.687	2.074	3.27E-02	-3.573
CD177	1.113	4.621	2.053	3.12E-02	-3.508
SEMA6B	0.971	4.275	2.138	3.33E-02	-3.304
SERPINB10	0.784	4.077	2.379	4.53E-03	-3.293

CHIT1	0.868	4.034	2.217	2.93E-02	-3.166
ABCA13	0.776	3.866	2.317	8.56E-03	-3.09
PAPSS2	0.647	3.297	2.349	4.50E-03	-2.65
SCD	0.74	3.174	2.1	1.63E-02	-2.434
ETHE1	0.563	2.742	2.284	1.97E-02	-2.179
PRRT4	0.394	2.567	2.703	1.60E-03	-2.173
COL17A1	0.485	2.296	2.242	7.05E-03	-1.81
MCM4	0.351	1.897	2.434	1.56E-03	-1.546
CLEC11A	0.299	1.829	2.613	6.45E-03	-1.53
LOC285758	0.287	1.738	2.601	4.85E-02	-1.452
HS3ST3A1	0.217	1.639	2.916	1.21E-03	-1.422
MKI67	0.307	1.661	2.435	3.81E-02	-1.354
FSTL3	0.301	1.565	2.376	9.63E-03	-1.264
C19orf77	0.224	1.356	2.596	1.71E-02	-1.132
CDT1	0.265	1.394	2.394	5.20E-03	-1.129
MTSS1	0.24	1.233	2.363	3.66E-03	-0.993
MYBL2	0.296	1.288	2.12	1.68E-02	-0.992
FAM108C1	0.277	1.202	2.12	2.93E-02	-0.926
ZNF367	0.257	1.164	2.178	1.16E-02	-0.907
ITGA9	0.148	1.052	2.826	5.75E-04	-0.904
TPX2	0.162	0.992	2.616	1.28E-03	-0.83
ANLN	0.166	0.992	2.579	1.90E-02	-0.826
PCOLCE2	0.16	0.983	2.622	6.76E-03	-0.823
TPSB2	0.055	0.878	3.987	1.43E-02	-0.823
LOC200772	0.023	0.775	5.055	9.28E-04	-0.752
STX1A	0.07	0.82	3.548	9.96E-04	-0.75
KIF2C	0.135	0.846	2.653	2.84E-03	-0.712
KIF11	0.196	0.872	2.151	1.26E-02	-0.676
GJB2	0.055	0.713	3.701	5.65E-04	-0.658
CDK1	0.09	0.749	3.049	1.45E-02	-0.658
TCTEX1D1	0.131	0.776	2.569	1.09E-02	-0.645
E2F8	0.055	0.672	3.617	8.91E-05	-0.617
GLB1L	0.15	0.711	2.246	2.64E-02	-0.561
SHB	0.123	0.678	2.469	4.17E-03	-0.556
NEIL3	0.085	0.616	2.85	4.71E-03	-0.53
CDC45	0.093	0.597	2.677	3.19E-02	-0.503
LAPTM4B	0.106	0.608	2.524	2.33E-02	-0.502
KCNH4	0.021	0.465	4.489	2.88E-03	-0.445
BUB1B	0.12	0.54	2.167	4.50E-02	-0.419
TRIM16	0.063	0.45	2.831	1.14E-02	-0.387
NPR3	0.022	0.395	4.189	6.54E-05	-0.373
CPXM1	0.024	0.372	3.953	2.86E-02	-0.348
EXO1	0.081	0.426	2.389	3.50E-02	-0.345

C1orf106	0.053	0.373	2.806	3.29E-02	-0.32
PROM1	0.027	0.318	3.55	1.97E-02	-0.291
IQGAP3	0.037	0.319	3.116	7.68E-04	-0.282
DEPTOR	0.017	0.275	4.043	2.48E-02	-0.258
ASPM	0.084	0.332	1.983	2.93E-02	-0.248
ZNF711	0.024	0.271	3.496	4.59E-03	-0.247
CIT	0.063	0.282	2.158	2.37E-02	-0.219
ZNF521	0.015	0.219	3.902	1.49E-03	-0.204
MYOF	0.047	0.249	2.413	1.40E-02	-0.202
DNAH10	0.048	0.248	2.36	5.32E-03	-0.2
DIAPH3	0.018	0.217	3.571	2.60E-02	-0.199
B4GALT6	0.026	0.212	3.017	1.53E-02	-0.185
PRKG2	0.011	0.171	3.967	3.86E-02	-0.16
KCNQ5	0.03	0.172	2.505	3.46E-02	-0.142
KIAA1211	0.011	0.144	3.689	9.38E-03	-0.133
HSPG2	0.009	0.12	3.76	8.21E-04	-0.111

References

1. Ehrlich P, Lazarus A. Histology of the Blood. Normal and Pathological. Cambridge, MA, Cambridge; 1900.

2. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil function: from mechanisms to disease. *Annu. Rev. Immunol.* 2012;30:459–89.

3. Wright HL, Moots RJ, Bucknall RC, Edwards SW. Neutrophil function in inflammation and inflammatory diseases. *Rheumatol.* 2010;49(9):1618–1631.

4. Nathan C. Neutrophils and immunity: challenges and opportunities. *Nat. Rev. Immunol.* 2006;6(3):173–82.

5. Mantovani A, Cassatella M a, Costantini C, Jaillon S. Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat Rev Immunol*. 2011;11(8):519–531.

6. Cassatella MA. Neutrophil-derived proteins: selling cytokines by the pound. *Adv. Immunol.* 1999;73:369–509.

7. Fanger NA, Liu C, Guyre PM, Wardwell K, O'Neil J, Guo TL, Christian TP, Mudzinski SP, Gosselin EJ. Activation of Human T Cells by Major Histocompatability Complex Class II Expressing Neutrophils: Proliferation in the Presence of Superantigen, But Not Tetanus Toxoid. *Blood*. 1997;89(11):4128–4135.

8. Cross A, Bakstad D, Allen JC, Thomas L, Moots RJ, Edwards SW. Neutrophil gene expression in rheumatoid arthritis. *Pathophysiology*. 2005;12(3):191–202.

9. Hakkim A, Fürnrohr BG, Amann K, Laube B, Abed UA, Brinkmann V, Herrmann M, Voll RE, Zychlinsky A. Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107:9813–9818.

10. Oudijk E-JD, Nijhuis EHJ, Zwank MD, van de Graaf E a, Mager HJ, Coffer PJ, Lammers J-WJ, Koenderman L. Systemic inflammation in COPD visualised by gene profiling in peripheral blood neutrophils. *Thorax*. 2005;60(7):538–44.

11. Baines KJ, Simpson JL, Bowden N a, Scott RJ, Gibson PG. Differential gene expression and cytokine production from

neutrophils in asthma phenotypes. *Eur. Respir. J. Off. J. Eur. Soc. Clin. Respir. Physiol.* 2010;35(3):522–31.

12. Bautista AP. Neutrophilic infiltration in alcoholic hepatitis. *Alcohol*. 2002;27(1):17–21.

13. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003;300(5617):286–90.

14. Chen G, Zhuchenko O, Kuspa A. Immune-like phagocyte activity in the social amoeba. *Science*. 2007;317(5838):678–81.

15. Robert J, Ohta Y. Comparative and developmental study of the immune system in Xenopus. *Dev. Dyn.* 2009;238(6):1249–70.

16. ATHENS JW, HAAB OP, RAAB SO, MAUER AM, ASHENBRUCKER H, CARTWRIGHT GE, WINTROBE MM. Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *J. Clin. Invest.* 1961;40:989–95.

17. Martin C, Burdon PC., Bridger G, Gutierrez-Ramos J-C, Williams TJ, Rankin SM. Chemokines Acting via CXCR2 and CXCR4 Control the Release of Neutrophils from the Bone Marrow and Their Return following Senescence. *Immunity*. 2003;19(4):583–593.

18. Borregaard N. Neutrophils, from marrow to microbes. *Immunity*. 2010;33(5):657–70.

19. Dahl R, Walsh JC, Lancki D, Laslo P, Iyer SR, Singh H, Simon MC. Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. *Nat. Immunol.* 2003;4(10):1029–36.

20. Faurschou M, S??rensen OE, Johnsen AH, Askaa J, Borregaard N. Defensin-rich granules of human neutrophils: Characterization of secretory properties. *Biochim. Biophys. Acta - Mol. Cell Res.* 2002;1591:29–35.

21. Rørvig S, Honore C, Larsson L-I, Ohlsson S, Pedersen CC, Jacobsen LC, Cowland JB, Garred P, Borregaard N. Ficolin-1 is present in a highly mobilizable subset of human neutrophil granules and associates with the cell surface after stimulation with fMLP. *J. Leukoc. Biol.* 2009;86:1439–1449.

22. Clemmensen SN, Udby L, Borregaard N. Subcellular fractionation of human neutrophils and analysis of subcellular markers. *Methods Mol. Biol.* 2014;1124:53–76.

23. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* 2013;13(3):159–175.

24. Le Cabec V, Cowland JB, Calafat J, Borregaard N. Targeting of proteins to granule subsets is determined by timing and not by sorting: The specific granule protein NGAL is localized to azurophil granules when expressed in HL-60 cells. *Proc. Natl. Acad. Sci. U. S. A.* 1996;93(13):6454–7.

25. Ley K, Laudanna C, Cybulsky MI, Nourshargh S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* 2007;7(9):678–89.

26. Muller W a. Getting leukocytes to the site of inflammation. *Vet. Pathol.* 2013;50(1):7–22.

27. Diamond MS, Staunton DE, de Fougerolles AR, Stacker SA, Garcia-Aguilar J, Hibbs ML, Springer TA. ICAM-1 (CD54): a counter-receptor for Mac-1 (CD11b/CD18). *J. Cell Biol.* 1990;111:3129–3139.

28. Muller WA, Weigl SA, Deng X, Phillips DM. PECAM-1 is required for transendothelial migration of leukocytes. *J. Exp. Med.* 1993;178:449–460.

29. Khan AI, Kerfoot SM, Heit B, Liu L, Andonegui G, Ruffell B, Johnson P, Kubes P. Role of CD44 and hyaluronan in neutrophil recruitment. *J. Immunol.* 2004;173:7594–7601.

30. Cooper D, Lindberg FP, Gamble JR, Brown EJ, Vadas MA. Transendothelial migration of neutrophils involves integrin-associated protein (CD47). *Proc. Natl. Acad. Sci. U. S. A.* 1995;92:3978–3982.

31. Hallett MB, Lloyds D. Neutrophil priming: the cellular signals that say “amber” but not “green”. *Immunol. Today*. 1995;16(6):264–8.

32. Wright HL, Moots RJ, Edwards SW. The multifactorial role of neutrophils in rheumatoid arthritis. *Nat. Rev. Rheumatol.* 2014;

33. El-Benna J, Dang PM-C, Gougerot-Pocidalo M-A. Priming of the neutrophil NADPH oxidase activation: role of p47phox phosphorylation and NOX2 mobilization to the plasma membrane. *Semin. Immunopathol.* 2008;30(3):279–89.

34. Kitchen E, Rossi a G, Condliffe a M, Haslett C, Chilvers ER. Demonstration of reversible priming of human neutrophils using platelet-activating factor. *Blood*. 1996;88(11):4330–7.

35. Tang D, Kang R, Coyne CB, Zeh HJ, Lotze MT. PAMPs and DAMPs: signal os that spur autophagy and immunity. *Immunol. Rev.* 2012;249(1):158–75.

36. Shields a M, Panayi GS, Corrigall VM. Resolution-associated molecular patterns (RAMP): RAMParts defending immunological homeostasis? *Clin. Exp. Immunol.* 2011;165(3):292–300.

37. Quayle JA, Watson F, Bucknall RC, Edwards SW. Neutrophils from the synovial fluid of patients with rheumatoid arthritis express the high affinity immunoglobulin G receptor, Fc gamma RI (CD64): role of immune complexes and cytokines in induction of receptor expression. *Immunology*. 1997;91(2):266–73.

38. García-García E, Rosales C. Signal transduction during Fc receptor-mediated phagocytosis. *J. Leukoc. Biol.* 2002;72(6):1092–108.

39. Nordenfelt P, Tapper H. Phagosome dynamics during phagocytosis by neutrophils. *J. Leukoc. Biol.* 2011;90(2):271–84.

40. Cross A, Bucknall RC, Cassatella MA, Edwards SW, Moots RJ. Synovial fluid neutrophils transcribe and express class II major histocompatibility complex molecules in rheumatoid arthritis. *Arthritis Rheum*. 2003;48(10):2796–2806.

41. Papayannopoulos V, Zychlinsky A. NETs: a new strategy for using old weapons. *Trends Immunol.* 2009;30(11):513–21.

42. Tillack K, Breiden P, Martin R, Sospedra M. T lymphocyte priming by neutrophil extracellular traps links innate and adaptive immune responses. *J. Immunol.* 2012;188(7):3150–9.

43. Yousefi S, Mihalache C, Kozlowski E, Schmid I, Simon HU. Viable neutrophils release mitochondrial DNA to form neutrophil extracellular traps. *Cell Death Differ.* 2009;16(11):1438–44.

44. Summary E. Neutrophils Activate Plasmacytoid Dendritic Cells by Releasing Self-DNA – Peptide Complexes in Systemic Lupus Erythematosus Roberto Lande et al . 2011;19:

45. Wartha F, Beiter K, Albiger B, Fernebro J, Zychlinsky A, Normark S, Henriques-Normark B. Capsule and D-alanylated lipoteichoic acids protect Streptococcus pneumoniae against neutrophil extracellular traps. *Cell. Microbiol.* 2007;9(5):1162–71.

46. Sumby P, Barbian KD, Gardner DJ, Whitney AR, Welty DM, Long RD, Bailey JR, Parnell MJ, Hoe NP, Adams GG, Deleo FR, Musser JM. Extracellular deoxyribonuclease made by group A

Streptococcus assists pathogenesis by enhancing evasion of the innate immune response. *Proc. Natl. Acad. Sci. U. S. A.* 2005;102(5):1679–84.

47. Buchanan JT, Simpson AJ, Aziz RK, Liu GY, Kristian SA, Kotb M, Feramisco J, Nizet V. DNase expression allows the pathogen group A Streptococcus to escape killing in neutrophil extracellular traps. *Curr. Biol.* 2006;16(4):396–400.

48. Brinkmann V, Zychlinsky A. Neutrophil extracellular traps: is immunity the second function of chromatin? *J. Cell Biol.* 2012;198(5):773–83.

49. Manzenreiter R, Kienberger F, Marcos V, Schilcher K, Krautgartner WD, Obermayer A, Huml M, Stoiber W, Hector A, Griese M, Hannig M, Studnicka M, Vitkov L, Hartl D. Ultrastructural characterization of cystic fibrosis sputum using atomic force and scanning electron microscopy. *J. Cyst. Fibros.* 2012;11(2):84–92.

50. Vitkov L, Klappacher M, Hannig M, Krautgartner WD. Extracellular neutrophil traps in periodontitis. *J. Periodontal Res.* 2009;44(5):664–72.

51. Gupta AK, Hasler P, Holzgreve W, Gebhardt S, Hahn S. Induction of neutrophil extracellular DNA lattices by placental microparticles and IL-8 and their presence in preeclampsia. *Hum. Immunol.* 2005;66(11):1146–54.

52. Villanueva E, Yalavarthi S, Berthier CC, Hodgin JB, Khandpur R, Lin AM, Rubin CJ, Zhao W, Olsen SH, Klinker M, Shealy D, Denny MF, Plumas J, Chaperot L, Kretzler M, Bruce AT, Kaplan MJ. Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. *J. Immunol.* 2011;187(1):538–52.

53. Garcia-Romo GS, Caielli S, Vega B, Connolly J, Allantaz F, Xu Z, Punaro M, Baisch J, Guiducci C, Coffman RL, Barrat FJ, Banchereau J, Pascual V. Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. *Sci. Transl. Med.* 2011;3:73ra20.

54. Yu Y, Su K. Neutrophil Extracellular Traps and Systemic Lupus Erythematosus. *J. Clin. Cell. Immunol.* 2013;4:

55. Sur Chowdhury C, Giaglis S, Walker U a, Buser A, Hahn S, Hasler P. Enhanced neutrophil extracellular trap generation in rheumatoid arthritis: analysis of underlying signal transduction pathways and potential diagnostic utility. *Arthritis Res. Ther.* 2014;16(3):R122.

56. Simon H-U. Neutrophil apoptosis pathways and their modifications in inflammation. *Immunol. Rev.* 2003;193:101–10.

57. Witko-Sarsat V, Pederzoli-Ribeil M, Hirsch E, Hirsh E, Sozzani S, Cassatella M a. Regulating neutrophil apoptosis: new players enter the game. *Trends Immunol.* 2011;32(3):117–24.

58. Savill J, Dransfield I, Gregory C, Haslett C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nat. Rev. Immunol.* 2002;2(12):965–75.

59. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer*. 1972;26(4):239–257.

60. Akgul C, Moulding D a, Edwards SW. Molecular control of neutrophil apoptosis. *FEBS Lett.* 2001;487(3):318–22.

61. Takeda M, Yamagami K, Tanaka K. Role of Phosphatidylserine in Phospholipid Flippase-Mediated Vesicle Transport in Saccharomyces cerevisiae. *Eukaryot. Cell*. 2014;13:363–75.

62. Verhoven B, Schlegel R a, Williamson P. Mechanisms of phosphatidylserine exposure, a phagocyte recognition signal, on apoptotic T lymphocytes. *J. Exp. Med.* 1995;182(5):1597–601.

63. Fulda S, Debatin K-M. Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene*. 2006;25(34):4798–811.

64. Lavrik I, Golks A, Krammer PH. Death receptor signaling. *J Cell Sci*. 2005;118(Pt 2):265–267.

65. Akgul C, Edwards SW. Regulation of neutrophil apoptosis via death receptors. *Cell. Mol. Life Sci.* 2003;60(11):2402–8.

66. Kischkel FC, Hellbardt S, Behrmann I, Germer M, Pawlita M, Krammer PH, Peter ME. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* 1995;14(22):5579–88.

67. Cairrao Domingos, Pedro M., F. Apoptosis: Molecular Mechanisms. *Encycl. Life Sci.* 2010;

68. Esmann L, Idel C, Sarkar A, Hellberg L, Behnen M, Möller S, van Zandbergen G, Klinger M, Köhl J, Bussmeyer U, Solbach W, Laskay T. Phagocytosis of apoptotic cells by neutrophil granulocytes: diminished proinflammatory neutrophil functions in the presence of apoptotic cells. *J. Immunol.* 2010;184(1):391–400.

69. Perretti M, D'Acquisto F. Annexin A1 and glucocorticoids as effectors of the resolution of inflammation. *Nat. Rev. Immunol.* 2009;9(1):62–70.

70. Ravichandran KS. Find-me and eat-me signals in apoptotic cell clearance: progress and conundrums. *J. Exp. Med.* 2010;207(9):1807–17.

71. Scannell M, Flanagan MB, deStefani A, Wynne KJ, Cagney G, Godson C, Maderna P. Annexin-1 and peptide derivatives are released by apoptotic cells and stimulate phagocytosis of apoptotic neutrophils by macrophages. *J. Immunol.* 2007;178(7):4595–605.

72. Dinauer MC. Disorders of neutrophil function: an overview. *Methods Mol. Biol.* 2007;412:489–504.

73. Bunting M, Harris ES, McIntyre TM, Prescott SM, Zimmerman GA. Leukocyte adhesion deficiency syndromes: adhesion and tethering defects involving beta 2 integrins and selectin ligands. *Curr. Opin. Hematol.* 2002;9:30–35.

74. Bogomolski-Yahalom V, Matzner Y. Disorders of neutrophil function. *Blood Rev.* 1995;412:

75. Condino-Neto A, Muscará MN, Grumach AS, Carneiro-Sampaio MM, De Nucci G. Neutrophils and mononuclear cells from patients with chronic granulomatous disease release nitric oxide. *Br. J. Clin. Pharmacol.* 1993;35(5):485–90.

76. Nauseef WM. Insights into myeloperoxidase biosynthesis from its inherited deficiency. *J. Mol. Med.* 1998;76:661–668.

77. Kaplan MJ. Neutrophils in the pathogenesis and manifestations of SLE. *Nat. Rev. Rheumatol.* 2011;7(12):691–9.

78. Neves FS, Carrasco S, Goldenstein-Schainberg C, Gonçalves CR, de Mello SBV. Neutrophil hyperchemotaxis in Behçet's disease: a possible role for monocytes orchestrating bacterial-induced innate immune responses. *Clin. Rheumatol.* 2009;28(12):1403–10.

79. Hacbarth E, Kajdacsy-Balla a. Low density neutrophils in patients with systemic lupus erythematosus, rheumatoid arthritis, and acute rheumatic fever. *Arthritis Rheum.* 1986;29(11):1334–42.

80. Cloke T, Munder M, Taylor G, Müller I, Kropf P. Characterization of a novel population of low-density granulocytes associated with disease severity in HIV-1 infection. *PLoS One*. 2012;7(11):e48939.

81. Denny MF, Yalavarthi S, Zhao W, Thacker SG, Anderson M, Sandy AR, McCune WJ, Kaplan MJ. A distinct subset of proinflammatory neutrophils isolated from patients with systemic lupus erythematosus induces vascular damage and synthesizes type I IFNs. *J. Immunol.* 2010;184(6):3284–97.

82. Maini RN, Taylor PC. Anti-cytokine therapy for rheumatoid arthritis. *Annu. Rev. Med.* 2000;51:207–29.

83. Calabrese LH, Rose-John S. IL-6 biology: implications for clinical targeting in rheumatic disease. *Nat. Rev. Rheumatol.* 2014;2:

84. Nair JR, Edwards SW, Moots RJ. Mavrilimumab, a human monoclonal GM-CSF receptor-α antibody for the management of rheumatoid arthritis: a novel approach to therapy. *Expert Opin. Biol. Ther.* 2012;12(12):1661–8.

85. Deiß A, Brecht I, Haarmann A, Buttmann M. Treating multiple sclerosis with monoclonal antibodies: a 2013 update. *Expert Rev. Neurother.* 2013;13(3):313–35.

86. So A, De Smedt T, Revaz S, Tschopp J. A pilot study of IL-1 inhibition by anakinra in acute gout. 2007.

87. Zeft A, Hollister R, LaFleur B, Sampath P, Soep J, McNally B, Kunkel G, Schlesinger M, Bohnsack J. Anakinra for systemic juvenile arthritis: the Rocky Mountain experience. *J. Clin. Rheumatol.* 2009;15:161–164.

88. Fitzgerald AA, Leclercq SA, Yan A, Homik JE, Dinarello CA. Rapid responses to anakinra in patients with refractory adult-onset Still's disease. *Arthritis Rheum.* 2005;52(6):1794–803.

89. Perrier C, Rutgeerts P. Cytokine blockade in inflammatory bowel diseases. *Immunotherapy*. 2011;3(11):1341–52.

90. Tobin A-M, Kirby B. TNF alpha inhibitors in the treatment of psoriasis and psoriatic arthritis. *BioDrugs*. 2005;19(1):47–57.

91. Samaan MA, Bagi P, Vande Casteele N, D'Haens GR, Levesque BG. An Update on Anti-TNF Agents in Ulcerative Colitis. *Gastroenterol. Clin. North Am.* 2014;43(3):479–494.

92. Nanau RM, Neuman MG. Safety of anti-tumor necrosis factor therapies in arthritis patients. *J. Pharm. Pharm. Sci.* 2014;17(3):324–61.

93. Mazumdar S, Greenwald D. Golimumab. 2009;1(5):422–431.

94. Sfikakis PP, Theodossiadis PG, Katsiari CG, Kaklamanis P, Markomichelakis NN. Effect of infliximab on sight-threatening panuveitis in Behçet's disease. *Lancet*. 2001;358(9278):295–6.

95. Reinisch W, De Villiers W, Bene L, Simon L, Rácz I, Katz S, Altorjay I, Feagan B, Riff D, Bernstein CN, Hommes D, Rutgeerts P, Cortot A, Gaspari M, Cheng M, Pearce T, Sands BE. Fontolizumab in moderate to severe Crohn's disease: A phase 2, randomized, double-blind, placebo-controlled, multiple-dose study. *Inflamm. Bowel Dis.* 2010;16:233–242.

96. Hueber W, Patel DD, Dryja T, Wright AM, Koroleva I, Bruin G, Antoni C, Draelos Z, Gold MH, Durez P, Tak PP, Gomez-Reino JJ, Foster CS, Kim RY, Samson CM, Falk NS, Chu DS, Callanan D, Nguyen QD, Rose K, Haider A, Di Padova F. Effects of AIN457, a fully human antibody to interleukin-17A, on psoriasis, rheumatoid arthritis, and uveitis. *Sci. Transl. Med.* 2010;2(52):52ra72.

97. Kobayashi SD, Deleo FR. Role of neutrophils in innate immunity : a systems biology-level approach. 2009;(December):309–333.

98. Bettencourt-Dias M, Giet R, Sinka R, Mazumdar A, Lock WG, Balloux F, Zafiropoulos PJ, Yamaguchi S, Winter S, Carthew RW, Cooper M, Jones D, Frenz L, Glover DM. Genome-wide survey of protein kinases required for cell cycle progression. *Nature*. 2004;432(7020):980–7.

99. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U. Just-in-time transcription program in metabolic pathways. *Nat. Genet.* 2004;36(5):486–91.

100. QMUL. Matura consortium - Press release http://www.qmul.ac.uk/media/news/items/smd/87799.html. 2012;

101. Gut IG. New sequencing technologies. *Clin. Transl. Oncol.* 2013;15(11):879–81.

102. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456(7221):470–6.

103. Sanger F, Nicklen S. DNA sequencing with chain-terminating. 1977;74(12):5463–5467.

104. Mardis ER. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* 2013;6:287–303.

105. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431(7011):931–45.

106. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science (80-.)*. 1995;270(5235):484–487.

107. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 2009;10(1):57–63.

108. Augenlicht LH, Kobrin D. Cloning and screening of sequences expressed in a mouse colon tumor. *Cancer Res*. 1982;42(3):1088–1093.

109. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* 2006;7(3):200–10.

110. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.* 2006;7(1):55–65.

111. Aparicio O, Geisberg J V, Struhl K. Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr. Protoc. Cell Biol.* 2004;Chapter 17:Unit 17.7.

112. Chernov A V, Baranovskaya S, Golubkov VS, Wakeman DR, Snyder EY, Williams R, Strongin AY. Microarray-based transcriptional and epigenetic profiling of matrix metalloproteinases, collagens, and related genes in cancer. *J. Biol. Chem.* 2010;285(25):19647–59.

113. Yin JQ, Zhao RC, Morris K V. Profiling microRNA expression with microarrays. *Trends Biotechnol.* 2008;26(2):70–6.

114. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008;18(9):1509–17.

115. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy - Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004;20:307–315.

116. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–628.

117. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods*. 2009;6(5):377–82.

118. Huang R, Jaritz M, Guenzl P, Vlatkovic I, Sommer A, Tamir IM, Marks H, Klampfl T, Kralovics R, Stunnenberg HG, Barlow DP, Pauler FM. An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS One*. 2011;6(11):e27288.

119. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 2008;40:1413–1415.

120. Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*. 2009;25:3207–3212.

121. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Processing of rRNA and tRNA. 2000;

122. Christodoulou DC, Gorham JM, Herman DS, Seidman JG. Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease. *Curr. Protoc. Mol. Biol.* 2011;Chapter 4:Unit4.12.

123. Brewer G, Ross J. Poly(A) shortening and degradation of the 3' A+U-rich sequences of human c-myc mRNA in a cell-free system. *Mol. Cell. Biol.* 1988;8(4):1697–708.

124. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*. 2014;15(1):419.

125. Siqueira JF, Fouad AF, Rôças IN. Pyrosequencing as a tool for better understanding of human microbiomes. *J. Oral Microbiol.* 2012;4:

126. Imelfort M, Edwards D. De novo sequencing of plant genomes using second-generation technologies. *Brief. Bioinform.* 2009;10(6):609–18.

127. Ronaghi M. DNA SEQUENCING:A Sequencing Method Based on Real-Time Pyrophosphate. *Science (80-.).* 1998;281(5375):363–365.

128. Kircher M, Kelso J. High-throughput DNA sequencing--concepts and limitations. *Bioessays*. 2010;32(6):524–36.

129. Fuller CW, Middendorf LR, Benner S a, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss J a, Schwartz DC, Vezenov D

V. The challenges of sequencing by synthesis. *Nat. Biotechnol.* 2009;27(11):1013–23.

130. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. 2012;7(2):e30087.

131. Metzker ML. Sequencing technologies - the next generation. *Nat. Rev. Genet.* 2010;11(1):31–46.

132. Sasson A, Michael TP. Filtering error from SOLiD Output. *Bioinformatics*. 2010;26(6):849–50.

133. Ratan A, Miller W, Guillory J, Stinson J, Seshagiri S, Schuster SC. Comparison of Sequencing Platforms for Single Nucleotide Variant Calls in a Human Sample. *PLoS One*. 2013;8(2):e55089.

134. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res.* 2008;18(7):1051–63.

135. Applied Biosystems. Principles of Di-Base Sequencing and the Advantages of Color Space Analysis in the SOLiD System. 2009;2–5.

136. Hormozdiari F, Hach F, Sahinalp SC, Eichler EE, Alkan C. Sensitive and fast mapping of di-base encoded reads. *Bioinformatics*. 2011;27(14):1915–21.

137. Smith AM, Heisler LE, St.Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, Harris AN, Perry KM, Giaever G, Pourmand N, Nislow C. Highly-multiplexed barcode sequencing: An efficient method for parallel analysis of pooled samples. *Nucleic Acids Res.* 2010;38:

138. Rhee M, Burns M a. Nanopore sequencing technology: research trends and applications. *Trends Biotechnol.* 2006;24(12):580–6.

139. Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Single Polymerase Molecules. 2009;(January):133–138.

140. Hayden EC. Human genomes in minutes? *Nat. News.* 2008;

141. McGettigan P a. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* 2013;17(1):4–11.

142. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 2011;29(7):644–52.

143. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, Kamoh B, Prabhu A-L, Tam A, Zhao Y, Moore RA, Hirst M, Marra MA, Jones SJM, Hoodless PA, Birol I. De novo assembly and analysis of RNA-seq data. *Nat. Methods.* 2010;7(11):909–12.

144. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, Sherlock G, Snyder M, Wang Z. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics.* 2010;11:663.

145. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28(5):511–515.

146. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 2010;28(5):503–10.

147. Martin J a, Wang Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* 2011;12(10):671–82.

148. Bao S, Jiang R, Kwan W, Wang B, Ma X, Song YQ. Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet.* 2011;

149. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. *Genome Biol.* 2010;11(12):220.

150. Fleige S, Pfaffl MW. RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.* 2006;27(2-3):126–39.

151. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.

152. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.

153. Li R, Li Y, Kristiansen K, Wang J. SOAP: Short oligonucleotide alignment program. *Bioinformatics*. 2008;24:713–714.

154. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res.* 2002;12(4):656–64.

155. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–1111.

156. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26(7):873–81.

157. Au KF, Jiang H, Lin L, Xing Y, Wong WH. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res.* 2010;38(14):4570–8.

158. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, MacLeod JN, Chiang DY, Prins JF, Liu J. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 2010;38(18):e178.

159. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–140.

160. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.

161. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*. 2010;26(1):136–8.

162. Young MD, Mccarthy DJ, Wakefi MJ, Smyth GK, Oshlack A, Robinson MD. Bioinformatics for High Throughput Sequencing. 2012;

163. Fang Z, Cui X. Design and validation issues in RNA-seq experiments. *Br. Bioinform*. 2011;12(3):280–287.

164. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B*. 1995;57(1):289–300.

165. Segal AW. How neutrophils kill microbes. *Annu. Rev. Immunol.* 2005;23:197–223.

166. Curtis AM, Bellet MM, Sassone-Corsi P, O'Neill L a J. Circadian Clock Proteins and Immunity. *Immunity*. 2014;40(2):178–186.

167. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001;29(9):e45.

168. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 2012;7(3):562–578.

169. Summers C, Rankin SM, Condliffe AM, Singh N, Peters a M, Chilvers ER. Neutrophil kinetics in health and disease. *Trends Immunol.* 2010;31(8):318–24.

170. Fehniger TA, Wylie T, Germino E, Leong JW, Magrini VJ, Koul S, Keppel CR, Schneider SE, Koboldt DC, Sullivan RP, Heinz ME, Crosby SD, Nagarajan R, Ramsingh G, Link DC, Ley TJ, Mardis ER. Next-generation sequencing identifies the natural killer cell microRNA transcriptome. *Genome Res.* 2010;20(11):1590–1604.

171. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, Mckernan KJ, Grimmond SM. Stem cell transcriptome profiling via massive-scale mRNA sequencing. 2008;5(7):613–619.

172. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008;320(5881):1344–9.

173. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 2008;24(3):133–41.

174. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! *Genome Biol.* 2011;12(8):125.

175. Batley J, Edwards D. Genome sequence data: management, storage, and visualization. *Biotechniques*. 2009;46(5):333–4, 336.

176. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat. Rev. Genet.* 2012;13(9):667–72.

177. Degel J, Shokrani M. Validation of the efficacy of a practical method for neutrophils isolation from peripheral blood. *Clin. Lab. Sci.* 2010;23(2):94–8.

178. Moulding DA, Akgul C, Derouet M, White MR, Edwards SW. BCL-2 family expression in human neutrophils during delayed and accelerated apoptosis. *J. Leukoc. Biol.* 2001;70(5):783–92.

179. Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A, Auffray C. Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res.* 2005;33(6):e56.

180. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* 2006;7:3.

181. Kobayashi SD, Voyich JM, Whitney AR, Deleo FR. Spontaneous neutrophil apoptosis and regulation of cell survival by granulocyte macrophage-colony stimulating factor Abstract : Polymorphonuclear leukocytes (PMNs. 2005;

182. Cowburn AS, Summers C, Dunmore BJ, Farahi N, Hayhoe RP, Print CG, Cook SJ, Chilvers ER. Granulocyte/macrophage colony-stimulating factor causes a paradoxical increase in the BH3-only pro-apoptotic protein Bim in human neutrophils. *Am. J. Respir. Cell Mol. Biol.* 2011;44(6):879–87.

183. Cross A, Moots RJ, Edwards SW. The dual effects of TNFalpha on neutrophil apoptosis are mediated via differential effects on expression of Mcl-1 and Bfl-1. *Blood*. 2008;111(2):878–84.

184. Fossati G, Mazzucchelli I, Gritti D, Ricevuti G, Edwards SW, Moulding DA, Rossi ML. In vitro effects of GM-CSF on mature peripheral blood neutrophils. *Int. J. Mol. Med.* 1998;1(6):943–994.

185. Fernandez MC, Walters J, Marucha P. Transcriptional and post-transcriptional regulation of GM-CSF-induced IL-1 beta gene expression in PMN. *J. Leukoc. Biol.* 1996;59(4):598–603.

186. Fujishima S, Hoffman AR, Vu T, Kim KJ, Zheng H, Daniel D, Kim Y, Wallace EF, Larrick JW, Raffin TA. Regulation of neutrophil

interleukin 8 gene expression and protein secretion by LPS, TNF-alpha, and IL-1 beta. *J. Cell. Physiol.* 1993;154(3):478–85.

187. Strieter RM, Kasahara K, Allen RM, Standiford TJ, Rolfe MW, Becker FS, Chensue SW, Kunkel SL. Cytokine-induced neutrophil-derived interleukin-8. *Am. J. Pathol.* 1992;141(2):397–407.

188. Atta-ur-Rahman, Harvey K, Siddiqui RA. Interleukin-8: An autocrine inflammatory mediator. *Curr. Pharm. Des.* 1999;5(4):241–53.

189. Monteseirín J, Chacón P, Vega A, El Bekay R, Alvarez M, Alba G, Conde M, Jiménez J, Asturias J a, Martínez A, Conde J, Pintado E, Bedoya FJ, Sobrino F. Human neutrophils synthesize IL-8 in an IgE-mediated activation. *J. Leukoc. Biol.* 2004;76(3):692–700.

190. Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred . I . Accuracy Assessment. 1998;175–185.

191. Ramos R. Quality Assessment - www.sourceforge.net/projects/qualevaluato/. 2013;

192. Cox AJ, Bauer MJ, Jakobi T, Rosone G. Large-scale compression of genomic sequence databases with the Burrows-Wheeler transform. *Bioinformatics*. 2012;28(11):1415–9.

193. Picard - http://picard.sourceforge.net/. 2014;

194. Andrews S. FastQC: A quality control tool for high throughput sequence data. *babraham Bioinforma.* 2010;1.

195. Simon P. Q-Gene: processing quantitative real-time RT-PCR data. *Bioinformatics*. 2003;19(11):1439–1440.

196. Zhang X, Ding L, Sandford AJ. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. *BMC Mol. Biol.* 2005;6(1):4.

197. Ramsköld D, Wang ET, Burge CB, Sandberg R, Ramsko D. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 2009;5(12):e1000598.

198. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999;27:29–34.

199. Jiménez-Marín A, Collado-Romero M, Ramirez-Boo M, Arce C, Garrido JJ. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proc.* 2009;3 Suppl 4:S6.

200. Gene-ontology-consortium. Gene Ontology : tool for the unifcation of biology. 2000;25(may):25–29.

201. Yendrek CR, Ainsworth EA, Thimmapuram J. The bench scientist's guide to statistical analysis of RNA-Seq data. *BMC Res. Notes*. 2012;5:506.

202. Shi Y, He M. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (Pinctada fucata). *Gene*. 2014;538(2):313–22.

203. Cristino AS, Tanaka ED, Rubio M, Piulachs M-D, Belles X. Deep sequencing of organ- and stage-specific microRNAs in the evolutionarily basal insect Blattella germanica (L.) (Dictyoptera, Blattellidae). *PLoS One*. 2011;6(4):e19350.

204. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010;11:94.

205. Aida Y, Pabst MJ. Priming of neutrophils by lipopolysaccharide for enhanced release of superoxide. Requirement for plasma but not for tumor necrosis factor-alpha. *J. Immunol.* 1990;145(9):3017–25.

206. Swain SD, Rohn TT, Quinn MT. Neutrophil Priming in Host Defense : 2002;4(1):

207. Wittmann S, Rothe G, Schmitz G, Fröhlich D. Cytokine upregulation of surface antigens correlates to the priming of the neutrophil oxidative burst response. *Cytometry. A*. 2004;57(1):53–62.

208. Dewas C, Dang PM-C, Gougerot-Pocidalo M-A, El-Benna J. TNF-α Induces Phosphorylation of p47phox in Human Neutrophils: Partial Phosphorylation of p47phox Is a Common Event of Priming of Human Neutrophils by TNF-α and Granulocyte-Macrophage Colony-Stimulating Factor . *J. Immunol.* . 2003;171 (8):4392–4398.

209. Langereis JD, Franciosi L, Ulfman LH, Koenderman L. GM-CSF and TNFα modulate protein expression of human neutrophils visualized by fluorescence two-dimensional difference gel electrophoresis. *Cytokine*. 2011;56(2):422–9.

210. Dyugovskaya L, Polyakov A, Ginsberg D, Lavie P, Lavie L. Molecular pathways of spontaneous and TNF-{alpha}-mediated neutrophil

apoptosis under intermittent hypoxia. *Am. J. Respir. Cell Mol. Biol.* 2011;45(1):154–62.

211. Rock KL, Latz E, Ontiveros F, Kono H. The sterile inflammatory response. *Annu. Rev. Immunol.* 2010;28:321–42.

212. Goldblatt F, Isenberg DA. New therapies for systemic lupus erythematosus. *Clin. Exp. Immunol.* 2005;140(2):205–12.

213. Koczan D, Drynda S, Hecker M, Drynda A, Guthke R, Kekow J, Thiesen H-J. Molecular discrimination of responders and nonresponders to anti-TNF alpha therapy in rheumatoid arthritis by etanercept. *Arthritis Res. Ther.* 2008;10(3):R50.

214. Fossati G, Bucknall RC, Edwards SW. Insoluble and soluble immune complexes activate neutrophils by distinct activation mechanisms: changes in functional responses induced by priming with cytokines. *Ann. Rheum. Dis.* 2002;61(1):13–9.

215. Holland SM, DeLeo FR, Elloumi HZ, Hsu AP, Uzel G, Brodsky N, Freeman AF, Demidowich A, Davis J, Turner ML, Anderson VL, Darnell DN, Welch PA, Kuhns DB, Frucht DM, Malech HL, Gallin JI, Kobayashi SD, Whitney AR, Voyich JM, Musser JM, Woellner C, Schäffer AA, Puck JM, Grimbacher B. STAT3 mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* 2007;357(16):1608–19.

216. Kobayashi SD, Voyich JM, Buhl CL, Stahl RM, DeLeo FR. Global changes in gene expression by human polymorphonuclear leukocytes during receptor-mediated phagocytosis: cell fate is regulated at the level of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 2002;99:6901–6906.

217. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. TM4 microarray software suite. *Methods Enzymol.* 2006;411:134–93.

218. Blake JA, Harris MA. The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Curr. Protoc. Bioinformatics*. 2008;Chapter 7:Unit 7.2.

219. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009;4:44–57.

220. IPA. Data were analyzed through the use of IPA (Ingenuity® Systems, www.ingenuity.com). 2011;

221. Zhang X, Kluger Y, Nakayama Y, Poddar R, Whitney C, Detora A, Weissman SM, Newburger PE. Gene expression in mature neutrophils : early responses to inflammatory stimuli Abstract : Neutrophils provide an essential de- play a major role in tissue damage during inflam-. 2003;

222. Fessler MB, Malcolm KC, Duncan MW, Worthen GS. A genomic and proteomic analysis of activation of the human neutrophil by lipopolysaccharide and its mediation by p38 mitogen-activated protein kinase. *J. Biol. Chem.* 2002;277:31291–31302.

223. Wright HL, Chikura B, Bucknall RC, Moots RJ, Edwards SW. Changes in expression of membrane TNF, NF-{kappa}B activation and neutrophil apoptosis during active and resolved inflammation. *Ann Rheum Dis*. 2010;

224. Klein JB, Rane MJ, Scherzer J a., Coxon PY, Kettritz R, Mathiesen JM, Buridi a., McLeish KR. Granulocyte-Macrophage Colony-Stimulating Factor Delays Neutrophil Constitutive Apoptosis Through Phosphoinositide 3-Kinase and Extracellular Signal-Regulated Kinase Pathways. *J. Immunol.* 2000;164(8):4286–4291.

225. Van den Berg JM, Weyer S, Weening JJ, Roos D, Kuijpers TW. Divergent effects of tumor necrosis factor alpha on apoptosis of human neutrophils. *J. Leukoc. Biol.* 2001;69:467–473.

226. Wright HL, Cross AL, Edwards SW, Moots RJ. Effects of IL-6 and IL-6 blockade on neutrophil function in vitro and in vivo. *Rheumatology (Oxford)*. 2014;53(7):1321–31.

227. Pelletier M, Maggi L, Micheletti A, Lazzeri E, Tamassia N, Costantini C, Cosmi L, Lunardi C, Annunziato F, Romagnani S, Cassatella M a. Evidence for a cross-talk between human neutrophils and Th17 cells. *Blood*. 2010;115(2):335–43.

228. Arlt A, Schäfer H. Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis. *Eur. J. Cell Biol.* 2011;90(6-7):545–52.

229. Kucharska A, Rushworth LK, Staples C, Morrice NA, Keyse SM. Regulation of the inducible nuclear dual-specificity phosphatase DUSP5 by ERK MAPK. *Cell. Signal.* 2009;21(12):1794–805.

230. Wright HL, Thomas HB, Moots RJ, Edwards SW. Interferon gene expression signature in rheumatoid arthritis neutrophils correlates with a good response to TNFi therapy. *Rheumatology (Oxford)*. 2014;1–6.

231. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* 2005;5(5):375–86.

232. Derouet M, Thomas L, Cross A, Moots RJ, Edwards SW. Granulocyte macrophage colony-stimulating factor signaling and proteasome inhibition delay neutrophil apoptosis by increasing the stability of Mcl-1. *J Biol Chem*. 2004;279(26):26915–26921.

233. Cross A, Edwards SW, Bucknall RC, Moots RJ. Secretion of oncostatin M by neutrophils in rheumatoid arthritis. *Arthritis Rheum.* 2004;50(5):1430–6.

234. Ten Hove W, Houben LA, Raaijmakers JAM, Bracke M, Koenderman L. Differential regulation of TNFalpha and GM-CSF induced activation of P38 MAPK in neutrophils and eosinophils. *Mol. Immunol.* 2007;44(9):2492–6.

235. Suzuki K, Hino M, Hato F, Tatsumi N, Kitagawa S. Cytokine-specific activation of distinct mitogen-activated protein kinase subtype cascades in human neutrophils stimulated by granulocyte colony-stimulating factor, granulocyte-macrophage colony-stimulating factor, and tumor necrosis factor-alpha. *Blood*. 1999;93(1):341–9.

236. Wright HL, Bucknall RC, Moots RJ, Edwards SW. Analysis of SF and plasma cytokines provides insights into the mechanisms of inflammatory arthritis and may predict response to therapy. *Rheumatology (Oxford)*. 2012;51(3):451–9.

237. Scheller J, Chalaris A, Schmidt-Arras D, Rose-John S. The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochim. Biophys. Acta*. 2011;1813(5):878–88.

238. Moore E, Biffl L, Moore A, Barnett C. Interleukin-6 neutrophil suppression concentration of neutrophil dependent of the. 1995;58:6–8.

239. McNamee JP, Bellier P V, Kutzner BC, Wilkins RC. Effect of pro-inflammatory cytokines on spontaneous apoptosis in leukocyte sub-sets within a whole blood culture. *Cytokine*. 2005;31(2):161–7.

240. Afford SC, Pongracz J, Stockley R a, Crocker J, Burnett D. The induction by human interleukin-6 of apoptosis in the promonocytic cell line U937 and human neutrophils. *J. Biol. Chem.* 1992;267(30):21612–6.

241. Ottonello L, Frumento G, Arduino N, Bertolotto M, Dapino P, Mancini M, Dallegri F. Differential regulation of spontaneous and immune complex- induced neutrophil apoptosis by

proinflammatory cytokines . Role of oxidants , Bax and caspase-3 Abstract : Neutrophil apoptosis represents a crucial. 2002;72(July):

242. Elbim C, Reglier H, Fay M, Delarche C, Andrieu V, El Benna J, Gougerot-Pocidalo MA. Intracellular pool of IL-10 receptors in specific granules of human neutrophils: differential mobilization by proinflammatory mediators. *J. Immunol.* 2001;166:5201–5207.

243. Castellucci M, Rossato M, Tamassia N, Gasperini S, Cassatella MA, Bazzoni F. Inhibition of interleukin-8 (CXCL8) and tumor necrosis factor-alpha (TNF-alpha) gene transcription by interleukin-10-induced epigenetic modifications. *Eur. J. Clin. Invest.* 2014;44:31.

244. Hasenberg M, Köhler A, Bonifatius S, Borucki K, Riek-Burchardt M, Achilles J, Männ L, Baumgart K, Schraven B, Gunzer M. Rapid immunomagnetic negative enrichment of neutrophil granulocytes from murine bone marrow for functional studies in vitro and in vivo. *PLoS One*. 2011;6(2):e17314.

245. Lakschevitz FS, Visser MB, Sun C, Glogauer M. Neutrophil transcriptional profile changes during transit from bone marrow to sites of inflammation. *Cell. Mol. Immunol.* 2014;000(000):0.

246. Zhou L, Somasundaram R, Nederhof RF, Dijkstra G, Faber KN, Peppelenbosch MP, Fuhler GM. Impact of human granulocyte and monocyte isolation procedures on functional studies. *Clin. Vaccine Immunol.* 2012;19(7):1065–74.

247. Neu B, Wenby R, Meiselman HJ. Effects of dextran molecular weight on red blood cell aggregation. *Biophys. J.* 2008;95(6):3059–65.

248. Nauseef WM. Isolation of human neutrophils from venous blood. *Methods Mol. Biol.* 2014;1124:13–8.

249. Sabroe I, Prince LR, Dower SK, Walmsley SR, Chilvers ER, Whyte MKB. What can we learn from highly purified neutrophils? *Biochem. Soc. Trans.* 2004;32(Pt3):468–9.

250. Stejskal S. Original Article Isolation of Granulocytes : Which Transcriptome Do We Analyse – Neutrophils or Eosinophils ? 2010;255:252–255.

251. Wright HL, Thomas HB, Moots RJ, Edwards SW. RNA-Seq Reveals Activation of Both Common and Cytokine-Specific Pathways following Neutrophil Priming. *PLoS One*. 2013;8(3):e58598.

252. Weil GJ, Chused TM. Eosinophil autofluorescence and its use in isolation and analysis of human eosinophils using flow microfluorometry. *Blood*. 1981;57(6):1099–104.

253. Gopinath R, Nutman TB. Identification of eosinophils in lysed whole blood using side scatter and CD16 negativity. *Cytometry*. 1997;30(6):313–6.

254. Davey MS, Tamassia N, Rossato M, Bazzoni F, Calzetti F, Bruderek K, Sironi M, Zimmer L, Bottazzi B, Mantovani A, Brandau S, Moser B, Eberl M, Cassatella M a. Failure to detect production of IL-10 by activated human neutrophils. *Nat. Immunol.* 2011;12(11):1017–8; author reply 1018–20.

255. Kabanova S, Kleinbongard P, Volkmer J, Andrée B, Kelm M, Jax TW. Gene expression analysis of human red blood cells. *Int. J. Med. Sci.* 2009;6(4):156–9.

256. Wiedow O, Muhle K, Streit V, Kameyoshi Y. Human eosinophils lack human leukocyte elastase. *Biochim. Biophys. Acta*. 1996;1315(3):185–7.

257. Metso T, Venge P, Haahtela T, Peterson CGB, Sevéus L. Cell specific markers for eosinophils and neutrophils in sputum and bronchoalveolar lavage fluid of patients with respiratory conditions and healthy subjects. *Thorax*. 2002;57(5):449–51.

258. Theilgaard-Mönch K, Jacobsen LC, Borup R, Rasmussen T, Bjerregaard MD, Nielsen FC, Cowland JB, Borregaard N. The transcriptional program of terminal granulocytic differentiation. *Blood*. 2005;105(4):1785–96.

259. Rørvig S, Østergaard O, Heegaard NHH, Borregaard N. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. *J. Leukoc. Biol.* 2013;94(4):711–21.

260. Wardle DJ, Burgon J, Sabroe I, Bingle CD, Whyte MKB, Renshaw S a. Effective caspase inhibition blocks neutrophil apoptosis and reveals Mcl-1 as both a regulator and a target of neutrophil caspase activation. *PLoS One*. 2011;6(1):e15768.

261. Tamassia N, Zimmermann M, Castellucci M, Ostuni R, Bruderek K, Schilling B, Brandau S, Bazzoni F, Natoli G, Cassatella M a. Cutting edge: an inactive chromatin configuration at the IL-10 locus in human neutrophils. *J. Immunol.* 2013;190(5):1921–5.

262. Meerschaert J, Busse WW, Bertics PJ, Mosher DF. CD14(+) cells are necessary for increased survival of eosinophils in response to lipopolysaccharide. *Am. J. Respir. Cell Mol. Biol.* 2000;23(6):780–7.

263. Bertazza L, Mocellin S. The dual role of tumor necrosis factor (TNF) in cancer biology. *Curr. Med. Chem.* 2010;17(29):3337–3352.

264. Tang X, Wang Y, Zhou S, Qian T, Gu X. Signaling pathways regulating dose-dependent dual effects of TNF-α on primary cultured Schwann cells. *Mol. Cell. Biochem.* 2013;378(1-2):237–46.

265. Amadou A, Nawrocki A, Best-Belpomme M, Pavoine C, Pecker F. Arachidonic acid mediates dual effect of TNF-alpha on Ca2+ transients and contraction of adult rat cardiomyocytes. *Am. J. Physiol. Cell Physiol.* 2002;282(6):C1339–47.

266. Murray J, Barbara JAJ, Dunkley SA, Lopez AF, Van Ostade X, Condliffe AM, Dransfield I, Haslett C, Chilvers ER. Regulation of Neutrophil Apoptosis by Tumor Necrosis Factor-alpha : Requirement for TNFR55 and TNFR75 for Induction of Apoptosis In Vitro. *Blood*. 1997;90(7):2772–2783.

267. Costa JJ, Matossian K, Resnick MB, Beil WJ, Wong DTW, Gordon JR, Dvorak AM, Weller PF, Galli SJ. Human Eosinophils Can Express the Cytokines Tumor Necrosis Factor-a and Macrophage Inflammatory Protein-la. 1993;91(June):2673–2684.

268. Theilgaard-Monch K, Knudsen S, Follin P, Borregaard N. The transcriptional activation program of human neutrophils in skin lesions supports their important role in wound healing. *J Immunol*. 2004;172(12):7684–7693.

269. Anderson SI, Hotchin N a, Nash GB. Role of the cytoskeleton in rapid activation of CD11b/CD18 function and its subsequent downregulation in neutrophils. *J. Cell Sci.* 2000;113 (Pt 1:2737–45.

270. Tamassia N, Cassatella M a, Bazzoni F. Fast and accurate quantitative analysis of cytokine gene expression in human neutrophils by reverse transcription real-time PCR. *Methods Mol Biol*. 2007;412(1):455–471.

271. Bisson-Boutelliez C, Miller N, Demarch D, Bene MC. CD9 and HLA-DR expression by crevicular epithelial cells and polymorphonuclear neutrophils in periodontal disease. *J. Clin. Periodontol.* 2001;28(7):650–6.

272. Carmona-Rivera C, Kaplan MJ. Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Semin. Immunopathol.* 2013;35(4):455–63.

273. Udby L, Calafat J, Sørensen OE, Borregaard N, Kjeldsen L. Identification of human cysteine-rich secretory protein 3 (CRISP-3) as a matrix protein in a subset of peroxidase-negative granules of neutrophils and in the granules of eosinophils. *J. Leukoc. Biol.* 2002;72:462–469.

274. General information about Single Nucleotide Polymorphisms. 2005;

275. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, Hirankarn N, Ying D, Pan HF, Mok CC, Chan TM, Wong RW, Lee KW, Mok MY, Wong SN, Leung AM, Li XP, Avihingsanon Y, Wong CM, Lee TL, Ho MH, Lee PP, Chang YK, Li PH, Li RJ, Zhang L, Wong WH, Ng IO, Lau CS, Sham PC, Lau YL. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet*. 2010;6(2):e1000841.

276. Fung EYMG, Smyth DJ, Howson JMM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun.* 2009;10(2):188–91.

277. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PIW, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coblyn J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ, Altshuler D. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* 2007;39(12):1477–82.

278. dbSNP build 141 summary. 2014;

279. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 2009;106(23):9362–7.

280. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. *PLoS One*. 2013;8(3):e58815.

281. Kurkó J, Besenyei T, Laki J, Glant TT, Mikecz K, Szekanecz Z. Genetics of rheumatoid arthritis - a comprehensive review. *Clin. Rev. Allergy Immunol.* 2013;45(2):170–9.

282. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, Shifrin N, Petri MA, Kamboh MI, Manzi S, Seldin MF, Gregersen PK, Behrens TW, Ma A, Kwok P-Y, Criswell LA. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.* 2008;40(9):1062–4.

283. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng B-J, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi SJ, Prahalad S, Guthery SL, Fischer J, Liao W, Kwok P-Y, Menter A, Lathrop GM, Wise CA, Begovich AB, Voorhees JJ, Elder JT, Krueger GG, Bowcock AM, Abecasis GR. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat. Genet.* 2009;41(2):199–204.

284. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.

285. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 2013;14(2):178–92.

286. Robinson J, Thorvaldsdóttir H. Integrative genomics viewer. *Nat. …* . 2011;29(1):24–26.

287. Thomas LW, Lam C, Edwards SW. Mcl-1; the molecular regulation of protein function. *FEBS Lett*. 2010;584(14):2981–2989.

288. Bae J, Leo CP, Hsu SY, Hsueh a J. MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J. Biol. Chem.* 2000;275(33):25255–61.

289. Kim J-H, Bae J. MCL-1ES induces MCL-1L-dependent BAX- and BAK-independent mitochondrial apoptosis. *PLoS One*. 2013;8(11):e79626.

290. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, Dunham I, Elnitski LL, Farnham PJ, Feingold EA, Gerstein M, Giddings MC, Gilbert DM, Gingeras TR, Green ED, Guigo R, Hubbard T, Kent J, Lieb JD, Myers RM, Pazin MJ, Ren B, Stamatoyannopoulos JA, Weng Z, White KP,

Hardison RC. Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 2014;111(17):6131–8.

291. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest ARR, Suzuki H, Hayashizaki Y, Hume D a, Orlando V, Grimmond SM, Carninci P. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 2009;41(5):563–71.

292. Esteller M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* 2011;12(12):861–74.

293. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol.* 2011;21(6):354–61.

294. Chen G, Qiu C, Zhang Q, Liu B, Cui Q. Genome-wide analysis of human SNPs at long intergenic noncoding RNAs. *Hum. Mutat.* 2013;34(2):338–44.

295. Jarvis JN, Jiang K, Liu T, Buck M, Carrier B, Chen Y. A174: JIA-Associated SNPs From Non-Coding Regions Are Located Within or Adjacent to Functional Genomic Elements of Human Neutrophils. *Arthritis Rheumatol. (Hoboken, N.J.)*. 2014;66 Suppl 1:S227.

296. Boudreau LH, Bertin JO, Surette ME. Human Neutrophils produce alternatively spliced variants of 5-Lipoxygenase mRNA. *FASEB J.* 2007;21(6):A975–c–.

297. Van den Akker ELT, Koper JW, Joosten K, de Jong FH, Hazelzet J a, Lamberts SWJ, Hokken-Koelega ACS. Glucocorticoid receptor mRNA levels are selectively decreased in neutrophils of children with sepsis. *Intensive Care Med.* 2009;35(7):1247–54.

298. Nitto T, Inoue T, Node K. Alternative spliced variants in the pantetheinase family of genes expressed in human neutrophils. *Gene.* 2008;426(1-2):57–64.

299. Güngör N, Pennings JL a, Knaapen AM, Chiu RK, Peluso M, Godschalk RWL, Van Schooten FJ. Transcriptional profiling of the acute pulmonary inflammatory response induced by LPS: role of neutrophils. *Respir. Res.* 2010;11:24.

300. Guo Y, Li C-I, Ye F, Shyr Y. Evaluation of read count based RNAseq analysis methods. *BMC Genomics*. 2013;14 Suppl 8:S2.

301. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25.

302. Wu Z, Fu Y, Cao J, Yu M, Tang X, Zhao S. Identification of differentially expressed miRNAs between white and black hair follicles by RNA-sequencing in the goat (Capra hircus). *Int. J. Mol. Sci.* 2014;15(6):9531–45.

303. La Paz JL, Pla M, Centeno E, Vicient CM, Puigdomènech P. The use of massive sequencing to detect differences between immature embryos of MON810 and a comparable non-GM maize variety. *PLoS One*. 2014;9(6):e100895.

304. Soetaert SSA, Van Neste CMF, Vandewoestyne ML, Head SR, Goossens A, Van Nieuwerburgh FCW, Deforce DLD. Differential transcriptome analysis of glandular and filamentous trichomes in Artemisia annua. *BMC Plant Biol.* 2013;13:220.

305. Hamfjord J, Stangeland AM, Hughes T, Skrede ML, Tveit KM, Ikdahl T, Kure EH. Differential expression of miRNAs in colorectal cancer: comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. *PLoS One*. 2012;7(4):e34150.

306. Bartee E, McFadden G. Cytokine synergy: an underappreciated contributor to innate anti-viral immunity. *Cytokine*. 2013;63(3):237–40.

307. Ohmori Y, Schreiber RD, Hamilton TA. Synergy between interferon-gamma and tumor necrosis factor-alpha in transcriptional activation is mediated by cooperation between signal transducer and activator of transcription 1 and nuclear factor kappaB. *J. Biol. Chem.* 1997;272(23):14899–907.

308. Suk K, Kim S, Kim YH, Kim KA, Chang I, Yagita H, Shong M, Lee MS. IFN-gamma/TNF-alpha synergism as the final effector in autoimmune diabetes: a key role for STAT1/IFN regulatory factor-1 pathway in pancreatic beta cell death. *J. Immunol.* 2001;166(7):4481–9.

309. Ahlers JD. Mechanisms of cytokine synergy essential for vaccine protection against viral challenge. *Int. Immunol.* 2001;13(7):897–908.