University of Southampton

Faculty of Medicine

Clinical and Experimental Sciences

Investigating the crosstalk between structural and immune components of human skin to understand pathogen sensing in cutaneous health and inflammation

by

Kalum Clayton

Thesis for the degree of Doctor of Philosophy

March 2021

University of Southampton

Abstract

FACULTY OF MEDICINE

Clinical and Experimental Sciences

Doctor of Philosophy

INVESTIGATING THE CROSSTALK BETWEEN STRUCTURAL AND IMMUNE COMPONENTS OF HUMAN SKIN TO UNDERSTAND PATHOGEN SENSING IN CUTANEOUS HEALTH AND INFLAMMATION

by Kalum Clayton

The epidermal environment of local cytokine, immune and microbial events are often mediated by keratinocyte-centric responses and dysregulation can lead to inflammatory skin disease. High dimensional technologies (microarrays and RNA-sequencing) allow interrogation of the global transcriptome in complex multi-cell tissues, but the transcriptomic changes from specific cellular subsets are difficult to isolate from the bulk signal. Computational approaches utilising unsupervised and hypothesis-driven methods resolved bulk skin gene expression into key immune and keratinocyte-related compartments underscoring inflammatory skin disease. Using an *in-silico* algorithm called CIBERSORT, bulk skin data were resolved into 19 cell signatures of purified populations from publicly available datasets, including 11 from keratinocytes stimulated with disease-associated cytokines.

Microarray data from whole skin biopsy samples were deconvoluted to observe shifting keratinocyte responses to local T_H-specific inflammation. Atopic samples, regardless of biopsy lesion status, showed a T_H2 immunophenotype, termed KC2. Psoriatic lesions and chronic atopic lesions, in contrast, resolved an IL-17 (KC17) induced signature. Atopic dermatitis and psoriasis treatment responses further validated the provenance of these KC2 and KC17 fractions.

The cutaneous microbiome is also important for maintenance of cutaneous health. Models of stratified keratinocyte-only epidermis were challenged with pathogenic *S. aureus* or commensal *S. epidermidis*. Bulk data from models indicated a strong transcriptomic response to *S. epidermis*. Applied as a reference signature for deconvolution from biopsy samples from atopic dermatitis, correlations of this fraction to KC2 and KC17 fractions, and barrier expression, strongly suggested a role for an altered commensal microbiome in the pathogenesis of atopic dermatitis.

Keratinocyte responses to *Staphylococcal spp.* challenge were further investigated at the single-cell level using DropSeq. This revealed highly stratified epidermal responses to challenge. Pathogen-challenged models indicated an inflammatory spinous population, which could also be resolved from psoriatic lesions and was ameliorated by etanercept treatment. Commensal and pathogen challenge of the model system induced distinct basal population of keratinocytes, compared together and to sterile controls. Basal keratinocytes from commensal-challenged models indicated a subtle alteration to a remodelling phenotype, which was found to be intensified by pathogen-challenge in addition to inflammatory MMP and IL-1 alarmin expression.

The potential of this experimental pipeline for modelling keratinocyte transcriptomic responses coupled with an *in-silico* deconvolution approach allows for elucidation of key pathogenic mechanisms and cellular-driven disease processes. This work underscores the importance of keratinocyte-centric responses to the maintenance of cutaneous health and their role during inflammation and dysbiosis.

Table of Contents

Table of Tables

Table of Figures

Table of Figures

Table of Figures

Research Thesis: Declaration of Authorship

Print name: **Kalum Clayton**

Title of thesis: **Investigating the crosstalk between immune and structural components of human skin to understand pathogen sensing in human health and inflammation**

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:- [please list references below]:

Clayton, K., Vallejo, A., Sirvent, S., Davies, J., Porter, G., Reading, I. C., Lim, F., Ardern-Jones, M. R. and Polak, M. E. (2021) 'Machine learning applied to atopic dermatitis transcriptome reveals distinct therapy-dependent modification of the keratinocyte immunophenotype*', British Journal of Dermatology, 184(5), pp. 913–922. doi: 10.1111/bjd.19431.

Signature: ..... Date:.....

{Important note:

The completed signed and dated copy of this form should be included in your print thesis.

A completed and dated but unsigned copy should be included in your e-thesis}

Acknowledgements

My foremost thanks go to my supervisor Dr Marta Polak for the constant academic and personal support offered throughout my PhD for which I am extremely grateful. To Dr Michael Ardern-Jones, also, for his guidance and support over the last four years, similarly without whom this work would not have been possible. I have been incredibly lucky to be able to draw on Marta and Mike's invaluable experience and guidance, culminating in this thesis and my development as a researcher and scientist.

I am also hugely indebted to Dr Sofia Sirvent and Dr Andres Vallejo Pullido for being so willing to share their knowledge, time and patience with me – I truly could not have benefitted from more able and inspiring Post-Docs! Thanks also to the rest of the Systems Immunology Group, past and present: Gemma Porter, Ying Teo, Gabriela Barbetta, and William Haw. Particularly to, *now Dr*, James Davies for the close friendship we shared over the years, making the PhD journey as fun and full of laughter as humanly possible.

I am also grateful for the support and input from our collaborators at Unilever; Dr Fei Ling Lim, Dr Jenny Pople and, formerly, Dr Rebecca Ginger.

This thesis, and surviving a PhD, would also have not been possible without the enormous support of friends and family. To my parents, who have always been so encouraging and proud – thank you. To my former housemates: Luke and Alex (& James) – thanks for the many laughs and memories, and the camaraderie during the Covid-19 Lockdowns.

And, finally, to Tom – my eternal champion and hero.

Definitions and Abbreviations

AAL ..... Acute lesional skin from an atopic dermatitis patient

AD ..... Atopic dermatitis

AL ..... Lesional (acute or chronic) from an atopic dermatitis patient

ALI ..... Air-liquid interface

AMP ..... Anti-microbial peptide

ANL ..... Non-lesional skin from an atopic dermatitis patient

ANOVA ..... Analysis of variance

BBKNN ..... Batch balanced k-nearest neighbour

BCL ..... Base call files; raw data files generated by Illumina sequencing.

BCM ..... Biofilm-conditioned media

BH ..... Benjamini-Hochberg (adjusted p-value)

BSA ..... Bovine serum albumin

CAL ..... Chronic lesional skin from an atopic dermatitis patient

CD ..... Cluster of differentiation

cDNA ..... Complimentary DNA

CFSE ..... Carboxyfluorescein succinimidyl ester, a fluorescent cell staining dye

CFU ..... Colony forming units

CIBERSORT ..... Cell-type identification by estimating relative subsets of RNA transcripts

CLA ..... Cutaneous lymphocyte-associated antigen

CLR ..... C-type lectin receptor

DAMP ..... Damage-associated molecular pattern

DC ..... Dendritic cell

DEG ..... differentially expressed gene

Derm22 ..... Panel of 22 cell types used for deconvolution in Félix Garza *et al.*, 2019

DMSO ..... Dimethyl sulfoxide

DNA ..... Deoxyribonucleic acid

dNTP ..... Deoxyribonucleotide triphosphate

DTT ..... Dithiothreitol/Cleland's reagent

EDC ..... Epidermal differentiation complex

EDTA ..... Ethylenediaminetetraacetic acid

Exo I ..... Exonuclease I

FBS ..... Foetal bovine serum

FC ..... Fold change

FDR ..... False-discovery rate

GCRMA ..... Guanine cytosine robust multi-array analysis

GEO ..... Gene Expression Omnibus

GO ..... Gene ontology

HaCaT ..... Cultured immortalised human keratinocyte

hBD ..... Human beta defensin

HH ..... Healthy skin from patients used as healthy controls

HPEK ..... Human primary epidermal keratinocyte

HTO ..... Hashtag oligonucleotide

HVG ..... Highly variable genes

IFN ..... Interferon

IgG ..... Immunoglobulin G

IL ..... Interleukin

IRF ..... Interferon regulatory factor

KC1 ..... Keratinocyte immunophenotype in response to T_H1 cytokine stimulation

KC17 ..... Keratinocyte immunophenotype in response to T_H17 cytokine stimulation

KC2 ..... Keratinocyte immunophenotype in response to T_H2 cytokine stimulation

KC22 ..... Keratinocyte immunophenotype in response to T_H22 cytokine stimulation

KGF ..... Keratinocyte growth factor

KGM2 ..... Keratinocyte growth medium 2 cell culture media (PromoCell, UK)

LC..... Langerhans cell

LCM ..... Laser capture microdissection

LIMMA..... Linear models for microarray data

LLSR ..... Linear least-squares regression

LM22 ..... Panel of 22 cell types used for deconvolution in Newman *et al.*, 2015

logFC..... Log2 fold change

LPS..... Lipopolysaccharide

LTA..... Lipoteichoic acid

MAPK..... Mitogen-activated protein kinase

MAST ..... Model-based analysis of single-cell transcriptomics

MCL ..... Markov clustering algorithm

MDS..... Multi-dimensional scaling

MHC ..... Major histocompatibility complex

MIAMI ..... Minimum information about a microarray experiment

MINISEQE ..... Minimum information about an RNA-sequencing experiment

MMAD ..... Microarray microdissection with analysis of differences

MMP..... Matrix metalloproteinase

mRNA ..... messenger RNA

MSCRAMM..... Microbial-surface components recognising adhesion matrix molecules

MyD88..... Myeloid differentiation factor 88

NFκB..... Nuclear factor kappa-light-chain enhancer of activated B cells

NHEK ..... Neonatal foreskin epidermal keratinocyte

NLR ..... NOD-like receptor

NMF..... Natural moisturising factor

NUSE..... Normalised unscaled standard error

OD_{600}..... Optical density measured at a wavelength of 600nm

PAGA ..... Partition-based graph abstraction

PAMP ..... Pathogen-associated molecular pattern

PBS ..... Phosphate buffered saline

PC ..... principal component (axis)

PCA ..... Principal component analysis

PCM ..... Planktonic-condition media

PCR ..... Polymerase chain reaction

PDG ..... Peptidoglycan

PERT ..... Perturbation model for gene expression deconvolution

PL ..... Lesional skin from a psoriasis patient

PNL ..... Non-lesional skin from a psoriasis patient

PRR ..... Pattern recognition receptor

QD ..... Quadratic deconvolution

Rcf ..... Relative centrifugal force

REVIGO ..... Reduce and visualise gene ontology

RHE ..... Reconstituted human epidermis

RLE ..... Relative log expression

RLR ..... Robust linear regression

RMA ..... Robust multi-array analysis

RNAseq ..... RNA-sequencing

RT ..... Reverse transcription

SA ..... *Staphylococcus aureus*

SAB ..... *Staphylococcus aureus* biofilm

SCANPY ..... Single cell analysis in python; a toolkit for scalable scRNA-seq pipelines

SCORAD ..... SCORing of Atopic Dermatitis index of lesion and disease severity

scRNA-seq ..... single-cell RNA-sequencing

SDS ..... Sodium dodecyl sulphate

SE ..... *Staphylococcus epidermidis*

SEB .....Staphylococcal enterotoxin B

Serpin .....Serine protease inhibitor

SSC.....Saline-sodium citrate

STAMP .....Single transcriptome attached to microparticle

SVM .....Support vector machine

SVR .....Support vector regression

TE.....10mM Tris pH 8.0, 1mM EDTA buffer

T_H .....T helper cell

TIR .....Toll/IL-1 receptor

TLR.....Toll-like receptor

TNFα.....Tumour necrosis factor alpha

TRIF .....TIR-domain containing adapter-inducing interferon-β

TSA .....Tryptic soy agar

TSO .....Template switch oligo

TSST-1.....Toxic shock syndrome toxin 1

TW Tween-20

UMAP .....Uniform manifold approximation and projection

UMI .....Unique molecular identifier

Chapter 1 Introduction

1.1 Biology of human skin

The skin is the largest organ of the body and the main barrier to the external world. For a long time, the skin was considered an inert organ as a structural barrier to physically protect the rest of the bodily organs. This conventional view of the skin has shifted dramatically over recent decades, where the skin is now appreciated to be an active and highly immunologically active organ. Contemporary understanding holds that the skin is essential in maintaining host homeostasis, regulating complex cellular interactions and cross-talk between cutaneous structural components, the immune system, and the cutaneous microbiome composed of commensal and even pathogenic microorganisms (Chuong *et al.*, 2002; Di Meglio, Perera and Nestle, 2011).

1.1.1 Anatomy of the skin

Human skin is composed of two main layers: the inner dermis and the outer epidermis. While the epidermis is the thinnest layer of the skin, it is composed of the bulk of the cutaneous cells found in this tissue (Kanitakis, 2002). The epidermis is the outermost layer of the skin and is exposed to the external environment. The major cell component of the epidermis, and indeed of the whole skin, are keratinocytes. The layers of the epidermis are defined by local stratified keratinocyte morphology (**Figure 1.1**). This stratification is based on the sequential processes of proliferation in the stratum basale, differentiation in the stratum spinosum and stratum granulosum, terminal differentiation in the stratum corneum, and finally, cell shedding via desquamation (Eckhart *et al.*, 2013).

The dermis is also composed of two layers, each distinctly defined by the differently composition of the dense network of collagen fibres. Thin collagen fibres characterise the stratum papillare (papillary dermis), while thick collagen fibres characterise the stratum reticular (reticular dermis). The main cell type of the dermis is populations of dermal fibroblasts. The characteristic dermal collagen fibres provide a non-cellular mechanical barrier and framework for the structures and cells of the dermis. These include blood vessels, lymphatic drainage, and the cutaneous appendages of sebaceous and sweat glands, hair follicles and shafts; and cellular populations such as dermal dendritic cells and other immune populations.

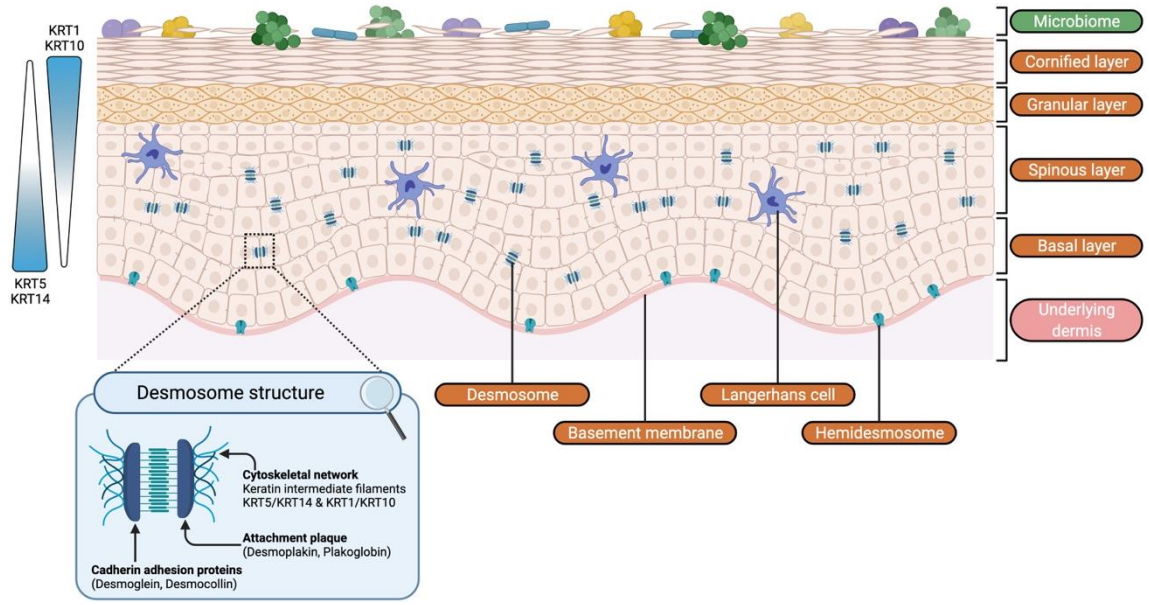


Figure 1.1. Structure of the human epidermis. Created using biorender.com

1.1.2 Proliferation and differentiation of epidermal keratinocytes

The chemical barrier of the epidermis is intrinsically linked to the physical barrier (Madison, 2003). The physical barrier is enhanced by moisture, lipid and acid components to protect the body from a wide variety of potential environmental insults. The physicality of the skin barrier is provided foremost by epidermal keratinocytes. Keratinocytes form the structure of the epidermis, which at the outer most layer form a protective layer of dead but tightly linked keratinocytes. Within the closely packed layers of keratinocytes, these cells form tight junctions which hold together this physical barrier to prevent access by external insults. At the same time, this barrier must remain semi-permeable to facilitate surveillance and tolerogenic immune processes, and secretion of endogenous factors such barrier lipids and anti-microbial peptides (Madison, 2003; Mizutani *et al.*, 2009).

The epidermis is marked by continual and dynamic renewal of resident keratinocyte populations. This renewal is underscored by proliferation of basal cells and differentiation of the outer epidermal populations. A distinguishing feature of keratinocytes is the layer-associated keratin intermediate-filaments forming an internal scaffold to provide mechanical strength. In the proliferating basal layer, which is attached to the dermal-epidermal basement membrane via hemidesmosomes, keratin intermediate filaments are typically composed of keratins 5 and 14 (KRT5, KRT14) (**Figure 1.1**). Initiation of differentiation alters the cytoarchitecture of keratinocytes, whereby intermediate filaments are composed of keratins 1 and 10 (KRT1, KRT10) in the spinous layer (**Figure 1.1**) (Moll *et al.*, 1982; Eichner, Sun and Aebi, 1986).Moreover, as keratinocytes begin to differentiate, they

form intercellular desmosome and adherens junctions with neighbouring cells, which are the basis of the physical barrier provided by the skin (Sumigray and Lechler, 2015).

The elements of the cornified barrier are provided by the next layer of the epidermis, the stratum granulosum, so named because of the appearance of the keratinocyte cytoplasm that is filled with keratohyalin granules. These granules contain the profilaggrin and loricrin components necessary for construction of the cornified envelope (Kanitakis, 2002). Intensification of the calcium gradient necessary for regulating keratinocyte differentiation, causes degranulation leading to phosphorylation and proteolysis of profilaggrin to filaggrin (Bikle, Xie and Tu, 2012). The cornified envelope is then formed by flattening of terminally differentiated corneocytes via aggregation of intracellular keratin intermediate filaments with filaggrin, loricrin and other late cornified envelope (LCE) proteins (**Figure 1.2**) (Eckhart *et al.*, 2013).

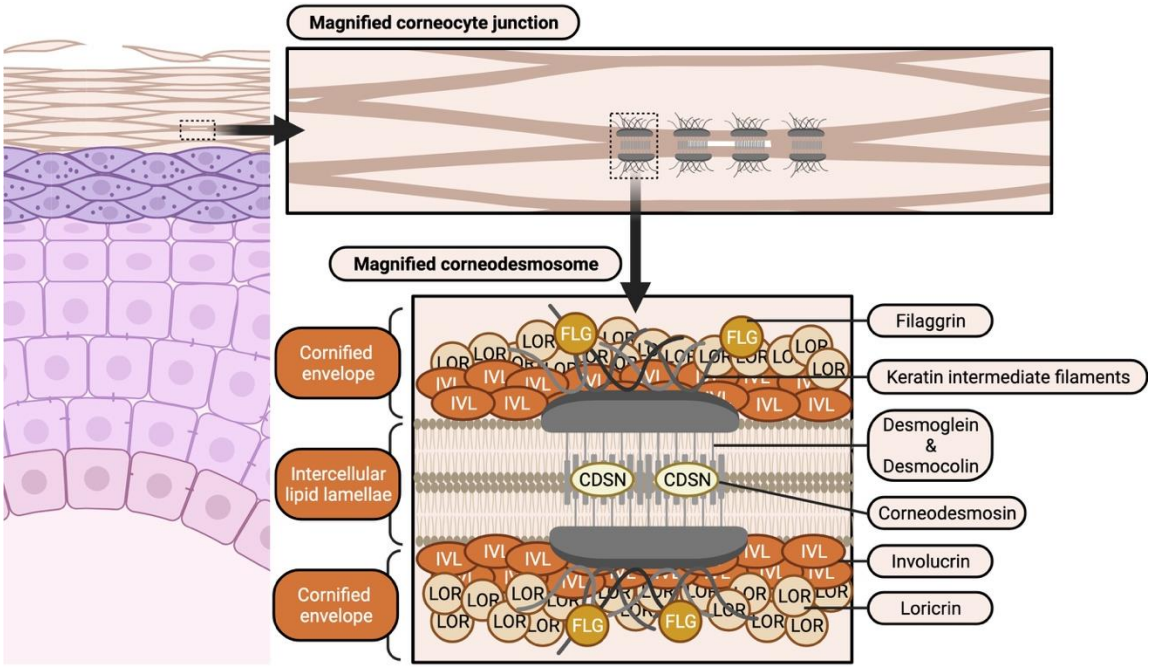


Figure 1.2. Structure of the epidermal cornified envelope. Created using biorender.com

In addition, during terminal differentiation, lipid lamellar bodies fuse with the plasma membrane of cornifying keratinocytes and eject their lipid contents into the extracellular space (Lee and Lee, 2018). The lipid composition of the outer barrier is roughly half ceramides, a quarter cholesterol and one-tenth free fatty acids, and functions as a water impermeable barrier (Madison, 2003; Elias *et al.*, 2014). This final differentiation step of keratinocytes results in the formation of a layer of cornified cells called corneocytes, which are essentially dead keratinocytes. The many layers of corneocytes compose the stratum corneum saturated with a high density of keratin protein filaments secured by corneodesmosomes, which are surrounded the lipids secreted from the

stratum granulosum. The stratum corneum, therefore, forms a physical and lipid barrier to protect the underlying live populations of keratinocytes.

The layers of the stratum corneum and structure between corneocytes is partly maintained by corneodesmosin tying the cornified protein envelope to the extracellular lipid envelope. Degradation of these corneodesmosin junctions causes desquamation and shedding of the outermost layers as a further protective measure and to allow for continual epidermal turnover without accumulation of many layers of dead keratinocytes. Degradation of corneodesmosomes is undertaken by the protease activity of kallikreins such as KLK5 and KLK7, which are pre-formed in the lamellar bodies of differentiating keratinocytes. Upon release and activation in the apical compartment, kallikreins target the structural proteins desmocollin and desmoglein for proteolytic cleavage resulting in breakdown of the intercellular junction and corneocyte shedding (Eckhart *et al.*, 2013).

The sequential differentiation of keratinocytes as they mature through the epidermis is heavily regulated by calcium and its gradient across the tissue. The calcium concentration is highest in the granular layer indicating it is essential for inducing and regulating keratinocyte differentiation (Menon, Grayson and Elias, 1985). *In vitro* keratinocyte cultures in low calcium conditions demonstrate proliferation but fail to differentiate, which is restored by concentration above 0.1 mM (Hennings *et al.*, 1980). Culturing above this so-called calcium switch concentration induces morphological and proteome changes in keratinocytes associated with initiation of differentiation such as desmosomal expression and a switch in keratin expression from KRT5/KRT14 to KRT1/KRT10.

1.1.3 Immunology of the epidermis

The function of the epidermis is extended beyond its structural aspect to include chemical, immunological and microbiome components. As the outermost layer of the skin, the epidermis is critical for barrier function, cutaneous immune activities and local microbiome homeostasis, for which a variety of cutaneous cell populations mount functional responses. In the epidermis, immune responses are initiated by resident cells such as keratinocytes or Langerhans cells and may lead to the recruitments of additional infiltrating populations to the skin. Keratinocytes, while crucial for establishing the stratum corneum and maintaining intercellular tight junctions, are also mediators of inflammation by their ability to sense bacterial products and activate subsequent downstream effector responses, express HLA class II molecules and act as non-professional antigen presenting cells, and their ability to secrete cytokines and chemokines (Banerjee *et al.*, 2004; Kim *et al.*, 2009).

The specialised immune populations of the epidermis, of which keratinocytes interact, with are Langerhans cells and resident $CD8^+$ memory T cells, which can be supplemented by infiltrating $CD4^+$ helper and $CD8^+$ cytotoxic populations during inflammation. The interactions and crosstalk between these three components are vital for maintenance of cutaneous health and understanding the role of inflammation in this tissue. Once activated, Langerhans cells migrate to the skin-draining lymph nodes. In the lymph nodes, Langerhans cells move to the T-cell-rich zone, inner paracortex of the lymph node for cellular interactions. Interactions between Langerhans cells and T cells instructs an adaptive response to mount effector responses to the presented antigen and its source (Banchereau and Steinman, 1998)

1.1.3.1 Keratinocytes as immune cells

Beyond their role in maintaining the physical barrier of the skin, keratinocytes also have innate immune functions involving antigen recognition, antigen presentation and the secretion of cytokines and AMPs (Nickoloff and Turka, 1994; Selsted and Ouellette, 2005; Lai and Gallo, 2009; Nestle *et al.*, 2009). Accumulating evidence suggests that keratinocytes have a critical role in restricting inappropriate responses from immune populations, such as, Langerhans cells, in order to maintain immune tolerance in what is an extremely immunogenic environment. An example of such restriction is the role of keratinocyte-derived E-cadherin in controlling the maturation of Langerhans cells (Mayumi *et al.*, 2013).

The two major components of innate immunity performed by keratinocytes are, firstly, cutaneous surveillance and, secondly, expression of specific molecules to recognise and regulate colonising microbial species. Immune surveillance is undertaken by a form of pattern recognition by keratinocytes in search for epidermal dysfunction by evidence of pathogen-associated molecular patterns (PAMPs) or from damaged cells (DAMPs) (Janeway, 1989). The keratinocyte-expressed pattern recognition receptors include Toll-like receptors (TLRs), NOD-like receptors (NLRs) and C-type lectin receptors (CLRs) (Pasparakis, Haase and Nestle, 2014). These receptors have evolved to recognise conserved microbial structures such as lipopolysaccharide (LPS), lipoteichoic acid (LTA), and peptidoglycan, as well as flagellin proteins, lipopeptides and viral nucleic acids (Bitschar *et al.*, 2017).

Recent genetic, biochemical, structural, cell biological and bioinformatics analysis of the TLRs has contributed to their place as the best understood pattern recognition receptors (Kawasaki and Kawai, 2014). The TLRs expressed by keratinocytes include TLR1, -2, -4, -5 and -6 which are located at the cell surface, and the endosomal TLRs 3 and 9 (**Figure 1.3**) (Baker *et al.*, 2003; Lebre *et al.*,

2007; Miller and Modlin, 2007). The outcome of TLR activation depends on the interacting molecule and the subsequent priming of molecular mediators or activation of downstream cellular effectors. There is, therefore, an array of ligands that can activate TLR signalling each modulating distinct downstream pathways and responses. Activation of TLR2 and TLR4 pathways in keratinocytes is critical for initiating antimicrobial responses through their respective ligands peptidoglycan and lipoteichoic acid, and lipopolysaccharide (**Figure 1.3**) (Schwandner *et al.*, 1999; Takeuchi *et al.*, 1999; Sumikawa *et al.*, 2006).

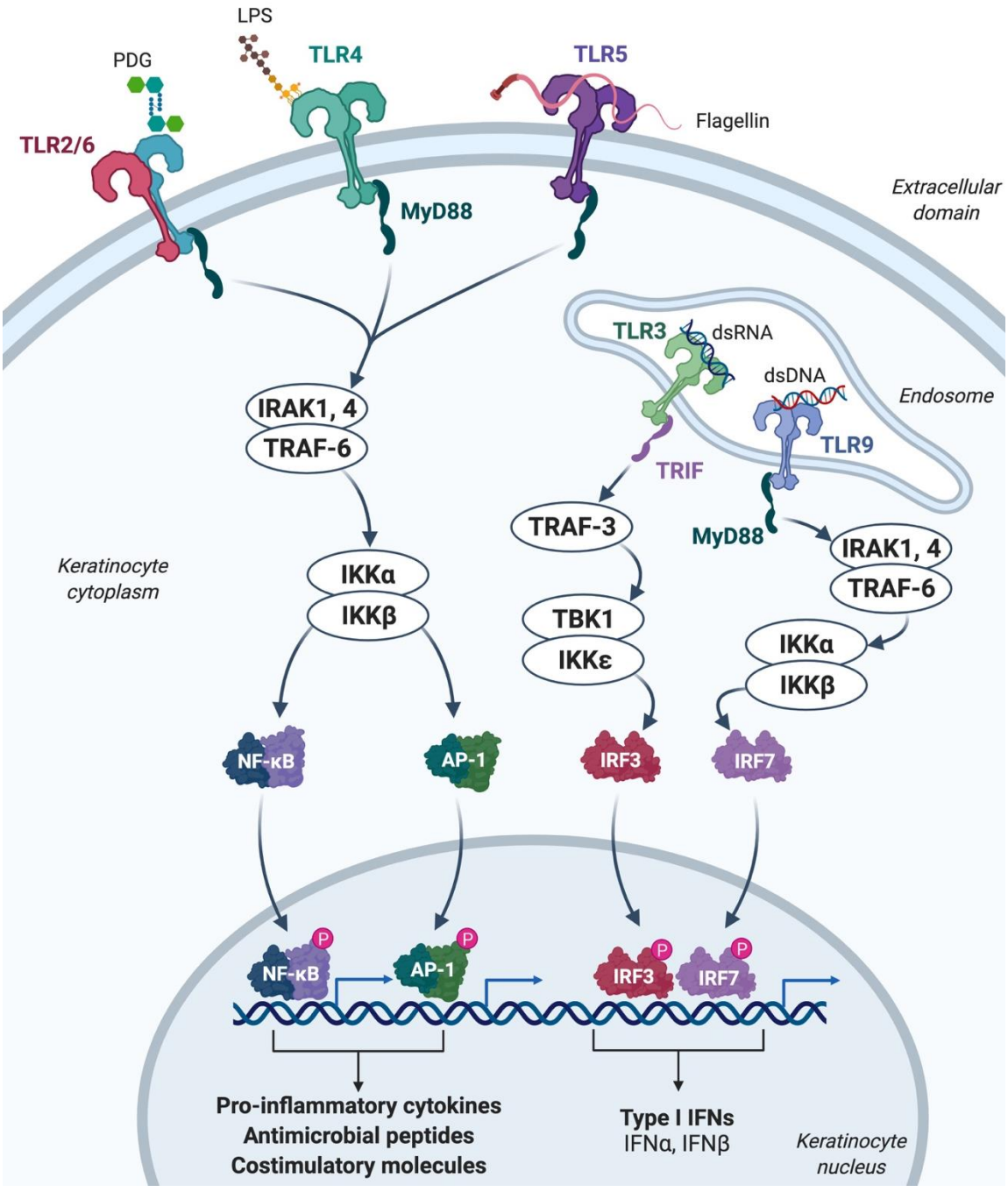


Figure 1.3. Toll-like Receptor (TLR) signalling cascade in keratinocytes. Created using biorender.com

TLR signalling is accomplished by activation of the intercellular domain of the TLR, shared with the IL-1 receptor, therefore known as the Toll/IL-1 receptor (TIR). Subsequently the activated TIR domain interacts with adapter proteins to initiate the signalling cascade. This signalling cascade occurs via either the myeloid differentiation factor 88 (MyD88)-dependent pathway or TIR-domain containing adapter-inducing interferon-β (TRIF)-dependent pathway (**Figure 1.3**). Therefore, the TIR domain can activate either the MyD88 pathways for inducing expression of pro-inflammatory cytokines, costimulatory molecules and anti-microbial peptides, or TRIF-dependent pathway production of type I interferons (Kobayashi *et al.*, 2015). The predominant response activated by keratinocyte TLR signalling is the production of type I interferons and a T_H1 immune response (Bitschar *et al.*, 2017).

Activation of keratinocytes induces the production of cytokines such as TNFα, the IL-1 family, an array of chemoattractant and immunomodulatory chemokines, and co-stimulatory molecules such as CD40, CD80 and CD86 (Denfeld *et al.*, 1996; Wakem *et al.*, 2000; Burns *et al.*, 2005; Romero-Tlalolini, Chávez Olmos and Garrido, 2013). The expression of these cytokines has important implications for keratinocyte function, including the modulation of vital proliferation and differentiation markers that can ultimately disturb epidermal homeostasis. For example, TNFα is a multifunctional proinflammatory cytokine reported to induce IL-6 and ICAM-1 expression by keratinocytes, and also important for inducing inflammation-activated expression of antimicrobial defensin peptides (Kondo and Sauder, 1997; Bernard *et al.*, 2012). Moreover, keratinocyte defensin production has also been reported in response to *S. aureus* protease damage via IL-1 cytokine family induction (Wang *et al.*, 2017). While IL-1α is constitutively expressed and acting as both a nuclear transcription factor and pro-inflammatory mediator, IL-1β is only activated after pathogen stimulation (Dinarello, 2018). This indicates that the role for both these IL-1 cytokines is as inflammatory alarmin molecules in response to epidermal PAMP/DAMP activation (Wang *et al.*, 2017).

Anti-microbial peptides are one of the key components of molecular epidermal defence. The main AMPs expressed by keratinocytes are the human β-defensins (hBD1, hBD2, hBD3 and hBD4), which are often produced in response to stimulation by bacterial products. For example, epidermal inflammation and expression of hBD3 is triggered by stimulation of TLR2 with lipoprotein agonist (Sumikawa *et al.*, 2006). Further human AMPs include cathelicidin (LL-37), RNase 7 and members of the S100 family of proteins such as S100A7, -A8 and -A9 and a family of serine protease inhibitors known as serpins (Wittersheim *et al.*, 2013; Wang, 2014). These molecules are a central component of host innate defence but also augment adaptive immune responses for more efficient regulation (Lai and Gallo, 2009).

In human epidermis, hBD1 mRNA has been reported to be constitutively expressed by keratinocytes indicating either continuous activity of this defensin in regulating the microbiome or its role during the rapid response to infection or injury (Selsted and Ouellette, 2005). In contrast, the other epidermal β-defensins are only expressed following either TLR stimulation or TNFα, IL-1β or IFNγ inflammation. Their induction by cutaneous inflammation often leads to changes in expression of these AMP during active disease, for example hBD2 is upregulated in atopic dermatitis by the aberrant T_H2 immune environment (Nomura *et al.*, 2003; Guilloteau *et al.*, 2010; Johnston *et al.*, 2011; Hönzke *et al.*, 2016).

1.1.3.2 Langerhans cells

The primary immune cells of the epidermis are Langerhans cells and memory T cells. Langerhans cells are the professional antigen-presenting cell of the epidermis that have the ability, once activated, to migrate from the skin to local draining lymph nodes (Romani *et al.*, no date; Stoitzner *et al.*, 2006). *In situ*, Langerhans cells form intimate connections with the many neighbouring keratinocytes and are important for initiating tolerogenic or immunogenic responses and bridging innate and adaptive responses once appropriately stimulated (Klechevsky *et al.*, 2008; Seneschal *et al.*, 2012; Van der Aar *et al.*, 2013; Polak and Singh, 2021). Despite their anatomical location and interaction with the external environment, their role as skin immune sentinels is called into question as, in murine models, their depletion results in neither increased skin infection nor dysregulated responses causing autoimmunity (West and Bennett, 2017). However, it has been reported that Langerhans cells are necessary for aggressive allergic responses in murine models (Bennett *et al.*, 2005).

Langerhans cell classification as either tissue resident dendritic cells or macrophages has further complicated understanding of their biology. Highly regulated excision through and reorganisation of keratinocyte tight junctions allows Langerhans cells to sample the local antigen environment without detriment to the epidermal barrier (Kubo *et al.*, 2009). This non-migratory antigen-presenting function closely resembles macrophage behaviour. In contrast, significant work on the function of Langerhans cells has shown that, like dendritic cells, migrate to the draining lymph nodes and present antigen to antigen-specific T cells (Shklovskaya, Roediger and Fazekas de St Groth, 2008; Bajaña *et al.*, 2012; Clayton, Vallejo, *et al.*, 2017; Ober-Blöbaum *et al.*, 2017). As such, definitively placing Langerhans cells within the dendritic cell-macrophage spectrum has proved controversial (Doebel, Voisin and Nagao, 2017).

Transcriptomic analysis of Langerhans cells has enabled new insight into resolving the identity of these cells as a dendritic cell or macrophage. Tissue environment has been reported as critical for shaping the phenotype of Langerhans cells, where dendritic cell localisation to the epidermal or dermal compartments corresponded with differences in the transcriptional underscoring the distinctions between the three antigen-presenting populations (Duluc *et al.*, 2014). Comparison of the transcriptomic profiles of Langerhans cells and $CD11c^+$ dermal dendritic cells showed significant differences between these two DC populations (Polak *et al.*, 2014). Previous reanalysis of a publicly-available dataset (GEO accession no.: GSE60317 (McGovern *et al.*, 2014)) using hierarchical clustering of the whole transcriptome of $CD14^+$ dermal dendritic cells, Langerhans cells and macrophages revealed that LCs formed a separate cluster from the other cell types (Clayton, Vallejo, *et al.*, 2017). This transcriptome-based analysis translated into Langerhans cells characterised as phenotypically and functionally distinct from both dendritic cells and macrophages. This separation remained true when considering monocytes, macrophages and other skin and blood dendritic cell populations (Carpentier *et al.*, 2016).

1.1.3.3 Skin resident T cells

Normal skin is the most abundant organ for T cells, having almost double that of the blood and three-times the number of central memory CLA^+ T cells (Clark *et al.*, 2006). Cutaneous T cells perform an important immunosurveillance function in response to chemokine signalling during inflammation and the presentation antigen by either MHC class I or II. Presentation of antigen by MHC class I activates $CD8^+$ T cells to induce cytotoxic immunity, while presentation by MHC class II activates immunomodulatory effects of $CD4^+$ T cells via differentiation to polarised T_H1, T_H2, T_H17 or regulatory T cell phenotypes (Nestle *et al.*, 2009).

The striking number of skin resident T cells points to their critical involvement in swift responses to stimulation by cutaneous antigen-presenting cells and inflammatory homeostasis. In addition, it has been demonstrated that these resident populations can potently activate cutaneous inflammation compared to circulating memory $CD4^+$ populations and, post-infection, a resident cytotoxic $CD8^+$ population in the epidermis and a circulating helper $CD4^+$ population are distinctly generated (Gebhardt *et al.*, 2009, 2011). Nevertheless, the cutaneous CLA^+ T cell population were majorly composed of $CD4^+$ populations (Clark *et al.*, 2006) indicating the immunogenic and inflammatory potential of this population. Inflammatory skin disease is often marked by pathogenic activation of T_H1, T_H2, T_H17 immune processes (Guttman-Yassky, Nograles and Krueger, 2011a; Dainichi, Hanakawa and Kabashima, 2014; Guttman-Yassky and Krueger, 2017).

1.2 Inflammatory skin diseases

The complex function and crosstalk between cutaneous cell populations coupled with dysregulation leads to a variety of disease phenotypes in the skin. This dysregulation often manifests as a malfunction of targeted immune responses, whereby an inflammatory disease phenotype in essence partly simulates the response to real cutaneous infection, injury or danger (Dainichi, Hanakawa and Kabashima, 2014). This response is inappropriately targeted and leads to disease.

The skin resident $CD4^+$ T cell population in normal skin is almost entirely skewed to a T_H1 polarisation (Clark *et al.*, 2006). Furthermore, the outcome of the TLR signalling cascade in keratinocytes is skewed to the production of interferon, which may explain the dominance of T_H1 axes in chronic inflammatory skin disorders (Bitschar *et al.*, 2017). It is conceivable that, these mechanisms are maintained in healthy tissue, albeit to a lesser extent, due to the observed sensitivity of keratinocytes to T_H1 over T_H2 cytokines (Albanesi *et al.*, 2001). Despite the strong preference for T_H1 responses in the skin, both T_H2 and T_H17; the other key $CD4^+$ types, have been shown to drive pathological cutaneous inflammation, for example, in atopic dermatitis and psoriasis, respectively. T_H17 responses in the skin are central for inducing AMP production by keratinocytes for host defence. Hence, adaptive T cell responses play a role in orchestrating host responses to skin pathogens, specifically by exerting action on epithelial cells to mount effective defence (Liang *et al.*, 2006).

Chronic skin disorders, often characterised by localised continuous and remitting cutaneous inflammation, are among the most common non-fatal diseases (Hollestein and Nijsten, 2014). Skin disorders can arise from any of the cutaneous compartments: barrier, innate or adaptive immunity, or microbiome dysregulation. Inflammation is a tightly controlled response of these cutaneous compartments to local homeostatic disruption, but aberrant inflammation leading to disease, manifests as a result of immunodeficiency, immunohyperactivity (i.e. atopy), and/or autoimmunity (Furman *et al.*, 2019). Atopic dermatitis is a characteristic immune hyperactivity disorder arising from aberrant atopic inflammation in the skin, while psoriasis is a characteristic autoimmune disorder arising from immune responses to keratinocyte hyper-proliferation (Guttman-Yassky and Krueger, 2017).

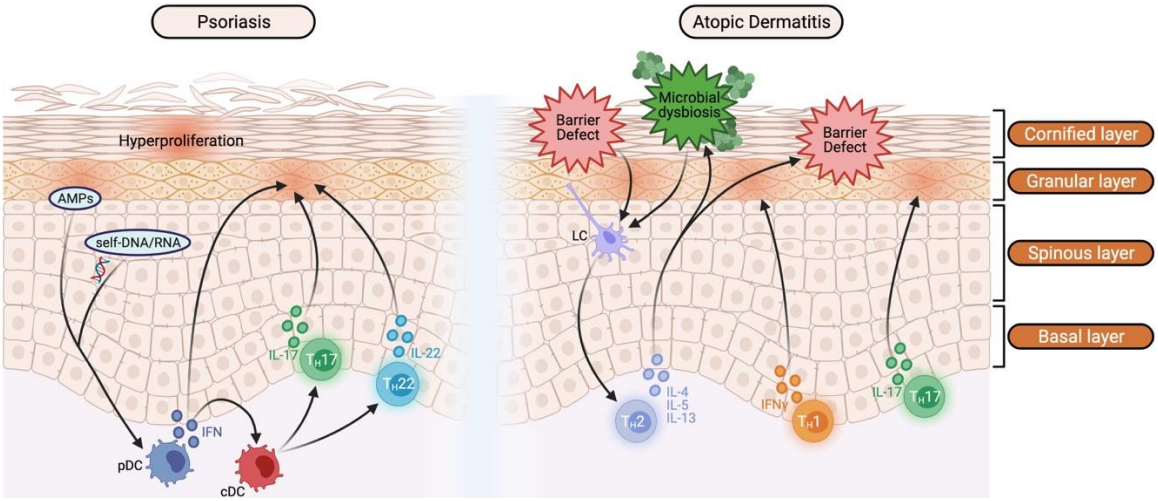


Figure 1.4. Epidermal and immune pathogenesis during psoriasis and atopic dermatitis lesions. Created using biorender.com

1.2.1 Atopic dermatitis

Atopic dermatitis is a relapsing, chronic inflammatory disease characterised by (1) abnormal immunity, (2) disrupted barrier and (3) cutaneous hyperactivity to otherwise innocuous triggers (**Figure 1.4**). This triad of pathology manifests in the skin as eczematous lesions with intense pruritis (Bieber, 2008). The incidence rate of atopic dermatitis has risen dramatically in the developed world over the last 30 years (Leung *et al.*, 2004). The significant prevalence rate among both children and adults (up to 20%, and 1-3%, respectively) and the rising incidence makes understanding the underlying drivers of disease critical for developing more effective treatment approaches (Odhiambo *et al.*, 2009). Developing these treatments is made harder by the extremely complex aetiology and pathogenesis of this disease, that is mediated by interactions of the epidermal barrier and its main resident cells, keratinocytes, the resident cutaneous immune population, the infiltrating immune responses and the local microbiome (Novak, Bieber and Leung, 2003; Wittmann and Werfel, 2006; Choy *et al.*, 2012; Hönzke *et al.*, 2016). All these factors contribute to atopic dermatitis pathogenesis but may manifest differently by proportion and temporal effect across the course of disease within an individual and across different patients.

Genomic studies have identified various susceptibility loci for development of allergic disease, which have revealed risk loci associated with immune abnormalities and skin barrier disruption (Paternoster *et al.*, 2015; Ferreira *et al.*, 2017). Genes critical to the function of the epidermal barrier and the cornified envelope have been identified as genetic susceptibility foci for atopic dermatitis such as *FLG*, *FLG2*, *LCE1D*, *LCE3D*, *SPRR3* (Irvine, McLean and Leung, 2011; Marenholz *et al.*, 2011; Margolis *et al.*, 2014). Therefore, disorders of the barrier are a well-described driver of atopic dermatitis (De Benedetto *et al.*, 2011; Hönzke *et al.*, 2016). Filaggrin classically sits at the

centre of this pathogenesis; functionally vital to barrier integrity, its loss of function is found in up to 50% of Northern European patients (Kim and Leung, 2012). Filaggrin is formed from cleaved profilaggrin in the granular layer of the epidermis and is vital for structuring keratin filaments in the cornified envelope (Kalinin, Marekov and Steinert, 2001). In terminally differentiated keratinocytes, corneocytes, filaggrin is degraded to generate natural moisturising factor (NMF), which is critical for maintaining the stratum corneum. Additionally, filaggrin breakdown products have been reported to have an inhibitory effect on *Staphylococcus aureus* growth (Miajlovic *et al.*, 2010).

Breakdown in proper barrier function in atopic dermatitis leads to a vicious cycle of persistence and lesioning of the skin. Concomitant dysregulation of T_H2 responses in atopic dermatitis skin adversely impairs keratinocyte differentiation, which exacerbates the broken barrier and leads to pathogenic alterations in the skin microbiome (**Figure 1.4**) (Hönzke *et al.*, 2016). Chronic pruritis and persistent scratching of affected skin adds to this cycle. Changes in the local microbiome occur, including the preferential colonisation away from commensal *Staphylococcus epidermidis* to pathogenic species such as *Staphylococcus aureus* (Hauser *et al.*, 1985; Guzik *et al.*, 2005; Kong *et al.*, 2012; Kobayashi *et al.*, 2015). Broken barrier, inappropriate T_H2 responses and pathogenic colonisation of the skin represent a sequence of events in atopic dermatitis of undefined chronology, contribution and cause. However, it is clearly understood that these three factors are required for the development of atopic dermatitis.

The immune response in AD is polarised to a T_H2/T_H22 environment at the acute stage. The T_H2-type cytokines, which include IL-4, IL-5 and IL-13, are chiefly associated with the production of IgE and eosinophilic responses marked of atopy (Berger, 2000). The immune process of AD intensifies at the chronic stage in both magnitude and heterogeneity of response, where a T_H1/T_H17 biology mixes with T_H2 atopy (**Figure 1.4**). The T_H-biology, however, involves the full spectrum of cytokine types: T_H2, T_H22, T_H1, T_H17 at all disease stages (Guttman-Yassky *et al.*, 2008; Koga *et al.*, 2008; Gittler *et al.*, 2012). This milieu of T_H axes has hampered understanding of the mechanisms that initiate and maintain lesions, with important therapeutic implications. The model of atopic dermatitis progression remains elusive as to whether the T_H biologies represent complimentary or competing processes that characterise acute and chronic stages as distinct immunological phases or define a continuous, progressive spectrum of inflammation.

Transcriptomic studies of atopic dermatitis lesions have begun to uncover the processes that initiate and sustain lesions. One report in particular undertook a comprehensive investigation of the molecular mechanisms that characterise the course of atopic dermatitis (Gittler *et al.*, 2012). This investigation compares clinically uninvolved, non-lesional atopic dermatitis skins to acute and

chronic lesions. Acute induction of lesions was found to be associated with increased epidermal thickness and proliferation. This hyperplasia continues as lesions became chronic, generally defined as beyond 72 hours, but was less significant than the initial acute induction. While epidermal hyperplasia was most closely associated with lesion onset, amplification of immune processes characterises chronic lesions. This amplification was evident for T_H2 and T_H22-related processes, showing that acute to chronic progression is not marked by an exclusive biphasic T_H2 to T_H1 switch. Indeed, this study found that the best indicators of disease severity are T_H2 at the acute stage but T_H2, T_H1 and T_H17 in chronic lesions.

The role of T_H17 processes in atopic dermatitis are underappreciated compared to the well understood pathogenic role of this axis in psoriasis. IL-17 is involved in bridging innate and adaptive responses and is particularly important for regulating anti-microbial responses. The persistent colonisation by *S. aureus* of atopic dermatitis skins points to a dysfunction of T_H17-mediated antimicrobial immunity. It had been reported that atopic dermatitis skins are not inherently abnormally unresponsive to T_H17 activation, but that the down-regulation of antimicrobial products is induced by the atopic T_H2 environment (Eyerich *et al.*, 2009). A later report found that atopic dermatitis keratinocytes do indeed respond normally to T_H17 stimulation and the classical T_H2 cytokines, IL-4 and IL-13, antagonise the expression of innate defence molecules (Nograles *et al.*, 2010).

The role of T_H2 cytokines on cutaneous AMP production, along with expression of structural and barrier proteins, was investigated in filaggrin-deficient skin models (Hönzke *et al.*, 2016). T_H2 cytokines were found to adversely impact barrier of the cornified envelope and reduce expression of key AMPs such as hBD-2 (Hönzke *et al.*, 2016). The hypercolonisation by *S. aureus* of skin of people with atopic dermatitis potentially induces lesion initiation and may induce downstream production of IL-17. The presence of microbial products is required for the full effector functions of IL-17-producing T cells (Eyerich *et al.*, 2009).

The role of these immune responses is also relevant to cutaneous cell populations, particularly keratinocytes. The role of keratinocytes in cutaneous immunity is being increasingly investigated. The effect of T_H2 and T_H17 cytokines on atopic skin is discussed above and is mainly mediated through keratinocytes and infiltrating immune populations. The crosstalk between keratinocytes and T cells is bidirectional. Keratinocytes play an active role in the pathology of atopic dermatitis through production of cytokines and surface marker expression that both leads and responds to inflammation and associated cellular and molecular infiltrates.

Effector cytokine responses from $T_H1/T_H2/T_H17/T_H22$ pathways modulate heterogeneous pleiotropic effects on keratinocytes during atopic dermatitis flares. These effects can either dominate, synergise or antagonise one another at different disease stages. It is well understood that the T_H2 effector cytokines, IL-4 and IL-13, are partly responsible for the dysregulation of the cornified envelope and anti-microbial proteins (Kabashima, 2013; Hönzke *et al.*, 2016). Recently, the role of T_H22 in acute and chronic atopic lesions has been explored, along with a potential synergy with T_H2 and T_H17 responses (Bernard *et al.*, 2012; Gittler *et al.*, 2012).

These inflammatory processes direct biological effect on keratinocytes. Indeed, it is increasingly understood that the inflammatory phenotype of the skin is mostly a consequence of keratinocyte responses to the local cytokine environment. The classical inflammatory cytokines IL-1, TNFα and IFNγ were first shown to directly affect keratinocyte function (Trefzer *et al.*, 1993; Rauschmayr, Groves and Kupper, 1997; Pastore *et al.*, 1998; Federici *et al.*, 2002; Yano *et al.*, 2008). Latterly, the role of IL-17 and IL-22 cytokines has been demonstrated as crucial for regulating the expression of genes associated with innate and anti-microbial immunity, namely the S100 proteins and β-defensins (Sa *et al.*, 2007; Guilloteau *et al.*, 2010; Bernard *et al.*, 2012). Modification of the keratinocyte transcriptome induced by the complex milieu of cytokines is of great interest to understanding cutaneous disease processes. Keratinocytes sit at the centre of the immunopathology and the remainder of the barrier and cutaneous host-microbe interaction triad of atopic dermatitis (Guilloteau *et al.*, 2010).

In summary, the immunobiology of atopic dermatitis has been demonstrated to include the full repertoire of helper T cell axes. T_H2 cytokines persist across the disease course and intensify as lesions progress. The roles in atopic dermatitis of T_H1, T_H17 and T_H22 processes are being increasingly unravelled, which presents new avenues for therapy. In particular, there is a complex interplay between T_H2 and the innate immunity and integrity of the cornified envelope. Additionally, the role of T_H17 in progression of atopic dermatitis is beginning to be elucidated, but the full effect of the opposing actions of T_H2 and T_H17 processes in disease and mediating innate-adaptive immunity yet remain to be fully explored. Finally, the effect of these immune processes on the epidermis and, in particular, keratinocytes is a critical focus of research to improve understanding of the pathology of cutaneous inflammation.

Major unanswered questions still remain to be researched, such as the global response on the keratinocyte transcriptome of the atopic environment, both *in vitro* and, perhaps more importantly, *in vivo*. Investigating the cellular and molecular response of keratinocytes within atopic skin during active lesions and from normal, non-lesional sites aims to identify populations and gene expression

profiles that are characteristically perturbed within the epidermis. Research should focus on the defining the effect on the keratinocyte compartment of the $CD4^+$ milieu, tracking the shift in these immune populations by their effect on the major epidermal cell type. Further investigation of cytokine and microbial interactions of keratinocytes should aim to distinguish the host-microbe factors underlying pathogenesis of inflammatory skin disease, such as the role of T_H17 immunity in atopic dermatitis.

1.2.2 Psoriasis

Psoriasis is another common inflammatory skin condition marked by lesions clearly demarcated with a characteristic silver-scaled appearance. Psoriasis affects up to three per cent of the European/North American Caucasian population, but has a reduced prevalence, estimated to be below one per cent, in other populations (Lowes, Bowcock and Krueger, 2007; Guttman-Yassky, Nograles and Krueger, 2011a, 2011b). The central pathogenesis of psoriasis arises from dysregulated hyper-proliferation of epidermal keratinocytes (**Figure 1.4**). This leads to the loss of normal cutaneous homeostasis and the activation of background immune circuits constitutive to normal function. The link between perturbed keratinocyte biology and immune activation leading to destructive inflammation arises because of the innate participation of keratinocytes in epidermal immunity and the active immunocompetency of the skin as a tissue.

The accelerated turnover of keratinocyte populations in the psoriatic epidermis is characteristic of the hyperproliferation of lesional tissue. This expansion affects the innate and adaptive immune responses mediated by keratinocytes, such as increased expression of AMP effector molecules, particularly S100A7 (psoriasin), S100A8 and S100A9, and LL-37 (Lowes, Suárez-Fariñas and Krueger, 2014). The causal driver of the hyperproliferation of keratinocytes is not fully understood. However, it is clear that there are distinct mechanisms underlying both initiation and maintenance of psoriasis (Lowes, Suárez-Fariñas and Krueger, 2014).

Investigations of the initiation of psoriasis have narrowed to understanding the autoimmune implications of self-DNA/RNA complexes activating populations of dendritic cells (Lande *et al.*, 2007; Gilliet and Lande, 2008) (**Figure 1.4**). It has been reported that such a mechanism of self-RNA or self-DNA forming complexes with LL37 activate both plasmacytoid and classical DCs via TLR7 and TLR8, respectively, leading to release of TNFα, IL-6 and IFNα/β (Ganguly *et al.*, 2009).

However, the difficulty in capturing the *in vivo* process underlying disease initiation has led to greater focusing and understanding of the maintenance of psoriasis. The major subsets of T cell and DC populations involved in sustaining psoriatic inflammation have been resolved (Cai, Fleming and

Yan, 2012). Psoriasis was first established as a T_H1-driven disease by observation that secretion of IFNγ from active $CD4^+$ T cells excised from psoriatic lesions could, *in vitro*, stimulate keratinocyte proliferation (Bata-Csorgo *et al.*, 1995). However, the classification of psoriasis as a purely T_H1 inflammatory disease does not satisfy all aspects of cellular and molecular events during maintenance of disease. For example, psoriatic lesions express high levels of non-T_H1-associated cytokines such as IL-17A, IL-17F, IL-22 and TNFα (K. Boniface *et al.*, 2007; Lowes *et al.*, 2008; Kagami *et al.*, 2010). Indeed, a reported high-fidelity murine model of psoriasis is based on IL-23-induced inflammation and not activated T_H1 axes, therefore underscoring the importance of these other immune processes in the pathogenesis of psoriasis (Suaŕez-Fariñas *et al.*, 2013). Consequently, a TNFα/T_H17/T_H22 inflammatory axis drives the epidermal inflammation underlying psoriatic lesions (Chiricozzi *et al.*, 2011).

The action of T_H17 immune pathways on keratinocyte biology is highly relevant to the perturbations observed in psoriatic lesions, such as increased proliferation, expression of AMPs and other effector molecules (Nograles *et al.*, 2008; Harper *et al.*, 2009). Therefore, the current understanding of the cytokines sustaining psoriatic lesions and the $CD4^+$ helper T cell populations from which these are derived centres on IFNγ, IL-17 and IL-22 from T_H1, T_H17 and T_H22 axes, respectively (**Figure 1.4**). In addition, activation of type I interferon responses has been implicated by MyD88-dependent activation in keratinocytes by IL-1B and IL-36 cytokines (Swindell *et al.*, 2018). Furthermore, IL-36 has been reported to be co-expressed alongside IL-17A and TNF is psoriatic lesions, leading to important consequences for this cytokine in the pathogenesis of psoriasis (Carrier *et al.*, 2011; Pfaff *et al.*, 2017)

The cellular alterations in psoriatic skin are recapitulated in the gene expression profiles of the tissue. Molecular transcriptomic investigation of the gene expression changes distinguishing lesions of psoriatic skin has resolved important further understanding of disease pathogenesis (Coda *et al.*, 2012; Bigler *et al.*, 2013; Swindell *et al.*, 2013, 2014). Whole transcriptome clustering of lesional against non-lesional samples recapitulated the molecular events of psoriatic inflammation demonstrating the near normal expression profile on non-lesional skin. It was reported that of the genes differentially expressed between psoriatic lesional and paired non-lesional tissue, the vast majority of genes (>80%) were attributable to processes involving keratinocyte function and infiltrating T cell activity (Swindell *et al.*, 2013). However, the most significantly elevated lesional-associated genes were related to epidermal or keratinocyte biology, while those less drastically elevated were associated with immune cell populations. *In vitro* culture of keratinocytes with psoriasis-implicated cytokines, IFNγ, TNFα, and IL-17, recapitulated the gene expression alterations observed from *in vivo* data, indicating that much of the keratinocyte perturbation derives from

activity of these immune processes and their related cytokines (Swindell *et al.*, 2012, 2013). The importance of T_H17 inflammation to sustaining psoriatic lesions was investigated by a clinical trial investigating the efficacy of TNFα blockade (Zaba *et al.*, 2009). Etanercept, an anti- TNFα biologic therapy, was shown to induce downregulation of IL-17 pathways responses leading to inactivation of the T_H17 axis. Hence, TNFα, while an early initiator of cutaneous psoriatic inflammation, can be put under blockade, leading to the cumulative downregulation of the T_H17 axis sustaining ongoing lesional inflammation and inducing clearance despite no such dampening of the T_H1 (IFNγ, TNFα) pathways.

In summary, although typically characterised by pathogenic disruption in keratinocyte homeostasis leading to hyperproliferation and increased expression of keratinocyte-derived effector molecules, psoriasis is driven by severe cutaneous inflammation. Inflammatory cytokine signalling and effector immune cells have been demonstrated to include IFNγ, TNFα, (T_H1 axis), IL-17 (T_H17 axis), and IL-22 (T_H22 axis). Further, the accumulation of these T_H1, T_H17, T_H22 immune axes in the skin of a psoriatic patient promotes addition to the cellular milieu of a lesion by other populations, including $CD8^+$ T cells, natural killer (NK) cells, macrophages and neutrophils, which has not been discussed here.

1.2.3 Role of keratinocytes in cutaneous health

Keratinocytes, owing to their position and role in the epidermis, are intrinsic to the cause and effect of epidermal inflammation. Communication from local environmental factors such as cytokine and chemokine production or shifting microbial interactions can lead to activation of pro-inflammatory signalling within the epidermis. Such communication is necessary for epidermal homeostasis but, as discussed above, also underscores the pathogenesis of both atopic dermatitis and psoriasis. Therefore, understanding the transcriptional responses of keratinocytes to such exogenous signals is critical for delineating how disturbance can trigger cutaneous inflammation. Once that keratinocyte responses are delineated these can be tracked in samples from *in vivo* patient data to understand individual inflammatory events. Consequently, this can lead to identification of optimal therapeutic regimens and, to better understand clinical resolution of inflammatory skin disease, how these therapies alter keratinocyte transcriptional phenotypes.

1.3 The cutaneous microbiome

While not host cells, the final component of the cutaneous system is the local microbiome. There are approximately 10^{12} cutaneous bacteria per square metre of skin, existing as a vast milieu of microorganisms, that closely interact with host cutaneous cells (Grice and Segre, 2011). The innate and adaptive immune systems can modulate the microbiome, while the microbiome, which is either

mostly harmless or indeed beneficial to the host, serves to educate immune responses to favour tolerance and dampen inflammatory responses (Jiang *et al.*, 2012; Naik *et al.*, 2012; Sanford and Gallo, 2013). Consequently, the role of the cutaneous microbiome in the maintenance of skin health and disease should be as fully appreciated as the role of the immune system. Indeed, the role of these two compartments is often so interlinked that uncoupling the pathogenic mechanisms driving cutaneous disease in relation to immune or microbiome dysregulation has so far proved complicated (Scharschmidt *et al.*, 2015).

The epidermis is colonised by diverse populations of microbiota that comprise the skin microbiome. While much focus has been given to the *Staphylococcal spp.* colonising of the skin due to their being readily cultured using standard microbiological protocols, the skin microbiome consists of a highly heterogeneous mixture of microbial species. Molecular approaches taking advantage of contemporary genomic sequencing and phylotyping revealed the diversity of the bacterial microbiome consisting of 19 detected phyla dominated by species from the Actinobacteria (51.8%), Firmicutes (24.4%), Proteobacteria (16.5%) and Bacteroidetes (6.3%) phyla (Grice *et al.*, 2009).

The first role of the skin microbiome is to prevent colonisation by otherwise pathogenic microorganisms. Thus, skin microorganisms provide commensal protection against potentially pathogenic species by competitive colonisation of available niches (Iwase *et al.*, 2010). This may involve a manifestation of interactions with host defences and immune responses, for example commensal species inducing cutaneous cells to secrete immune products against pathogenic species (Belkaid and Segre, 2014; Grice, 2015). Alternately, secreted products from preferentially colonising commensal species can help to train the host immune system and induce cutaneous tolerance to the wider microflora to prevent inflammation. In addition, cutaneous inflammation may alter the niches of the skin to encourage pathogen colonisation rather than beneficial symbiosis with commensal species.

Host-microbe interactions can play a critical role in shaping the cutaneous microbial community (Lai *et al.*, 2010). As such, disruption in the usually stable skin microbiome can result in and be perturbed by altered immune responses, which are a major contributing factor to inflammatory skin diseases (Leyden *et al.*, 1975; Kobayashi *et al.*, 2015; Nakatsuji *et al.*, 2017; Chang *et al.*, 2018). Dysbiosis of staphylococcal species on the skin is most apparent in atopic dermatitis. Typically, 5% of the healthy population are persistently colonised with *Staphylococcus aureus*, but this frequency raises to 90% of AD patients (Hauser *et al.*, 1985; Gong *et al.*, 2006). The host-microbe interaction of *Staphylococcus aureus* in atopic dermatitis is evident from the colonisation of lesional and non-lesional skins and the distinct correlation between colonisation and disease severity (Guzik *et al.*,

2005; Park *et al.*, 2013). Whether this dysbiosis is causative or resulting from other pathological mechanisms, and how these factors coalesce to lead to disease has yet to be fully delineated.

1.3.1 *Staphylococcus epidermidis*

Genetic phylotyping of colonising bacterial species on healthy human skin revealed that nearly two-thirds of the microbiome is composed of *Corynebacteria* (22.8%), *Propionibacteria* (23.0%) and *Staphylococci* (16.8%) (Grice *et al.*, 2009). These varied at different skin sites representative of distinct sebaceous, dry and moist niches, whereby *Staphylococcal* species were significantly associated with sebaceous and moist sites (Kong, 2011). *S. epidermidis* is historically regarded as the major bacterial species composing the commensal microbiome and is found to colonise the vast majority of the population (Percival *et al.*, 2012). Therefore, *S. epidermidis* is considered the archetypal cutaneous commensal.

The microbial mechanisms and host-microbe interactions that facilitate the benign symbiosis between *S. epidermidis* growth and cutaneous homeostasis are an important area of research for understanding dysbiosis. One important topic for research of *S. epidermidis* is understanding the model behind the almost universal colonisation of human skin but low virulence in providing benefits afforded to both microbe and host (Otto, 2009). Deciphering the mechanisms of cutaneous preference for this microbial species could help to uncover disturbed processes during pathogenic colonisation and the role of pathogens in contributing to tissue inflammation.

The beneficial role of *S. epidermidis* is targeted at multiple aspects of host epidermal homeostasis. A main active role is modulation of host innate immune responses. This extends to both immune priming for tolerance and promotion of wound healing. The mechanism of *S. epidermidis*-induced immune training and dampening has been reported to be conducted via TLR3 pattern recognition sensing by cutaneous immune populations of bacteria products such as lipoteichoic acid and LP78, a lipopeptide (Lai *et al.*, 2009; Li *et al.*, 2019). Furthermore, other *S. epidermidis* bacterial products, such as phenol soluble modulins (PSMs) have been indicated in augmenting keratinocyte AMP production to enact advantage against other AMP-susceptible microbes (Wanke *et al.*, 2011).

The secreted factors from *S. epidermidis* activate divergent signalling pathways in keratinocytes compared to competing pathogen. Presentation to TLR2 and subsequent activation of NFκB transcription pathways caused an elevated expression of hBD3 and RNase7 when challenged with commensal products (Lai *et al.*, 2010). hBD3 and RNase7 are critical AMPs required for the capacity of keratinocytes to effectively mount killing of pathogenic *S. aureus* (Kisich *et al.*, 2007; Bitschar *et al.*, 2017). Moreover, secreted products from *S. epidermidis* can directly inhibit colonisation by

pathogenic species, particularly *S. aureus*. *S. epidermidis* produces novel bacteriocins called lantibiotics that, alone or synergistically with host AMPS, protect against *S. aureus* (Nakatsuji *et al.*, 2017; Byrd, Belkaid and Segre, 2018). Finally, quorum crosstalk by colonising *S. epidermidis* is also proposed to passively inhibit *S. aureus* growth, however, the precise mechanism has not been fully elucidated, but research focuses around the shared *Staphylococcal* accessory gene regulator (*agr*) quorum-sensing pathway (Lina *et al.*, 2003; Otto, 2010).

1.3.2 *Staphylococcus aureus*

S. aureus is a Gram-positive, coagulase-positive pathogen. Structurally, the external components of the bacterium are composed of a plasma membrane, peptidoglycan cell wall and outer capsule (**Figure 1.5**). The peptidoglycan component of the Gram-positive cell wall accounts for 50% of the wall weight in *S. aureus* and is reported to be a putative endotoxin by causing *in vivo* platelet aggregation (Kessler, Nussbaum and Tuazon, 1991). Lipoteichoic acid is also part of the *S. aureus* cell wall and is a cytoplasmic membrane-bound glycerol phosphate monomer extending through the peptidoglycan scaffold. Lipoteichoic acid is an important agonist for host sensing via TLR signalling (Schwandner *et al.*, 1999; Takeuchi *et al.*, 1999). Other major components of the *S. aureus* cell wall include enzymes such as the membrane-bound penicillin-binding proteins that facilitate wall assembly, and β-lactamase responsible for hydrolysing β-lactam antibiotics (**Figure 1.5**).

The major virulence factors of *S. aureus* during exponential growth are the expressed surface proteins. Their ability to bind host adhesions molecules enables colonisation of infected tissue and thus, they form the group called microbial-surface components recognising adhesion matrix molecules (MSCRAMM). One such MSCRAMM molecule, coagulase, is used to characterise *Staphylococcal* species by their ability to enzymatically convert fibrinogen to fibrin. *S. aureus* is coagulase positive in contrast to *S. epidermidis*, which is coagulase negative. Clumping factor is another prothrombin coagulase virulence factor. Other MSCRAMM molecules that variably contribute to the virulence of different *S. aureus* strains are the fibronectin-binding, collagen-binding, and elastin-binding proteins. Protein A, another surface molecule, inhibits bacteria opsonophagocytosis by its ability to bind the Fc domain of IgG antibodies (**Figure 1.5**).

Once *S. aureus* growth reaches the stationary phase the bacterium begins producing various secreted toxins (**Figure 1.5**). Cytotoxins, such as α-toxin are responsible for pore formation and are pro-inflammatory molecules in host systems. Other toxins function as pyrogenic superantigens for binding to host MHC class II proteins. The consequences of these toxins are toxic shock syndrome caused by extensive T cell activation and cytokine storm; hence, such an example of these molecules is called toxic shock syndrome toxin 1 (TSST-1). Related to these are the enterotoxins,

such as Staphylococcal enterotoxin B (SEB). SEB is also a virulent superantigen similarly capable of causing major inflammatory responses. In the skin, $CD4^+$ T_H2 responses are potently induced by SEB and may contribute to the atopic activation of cutaneous immunity during flares of atopic dermatitis (Ardern-Jones *et al.*, 2007).

S. aureus is the most researched human pathogen, likely due to its common association with infection and disease (Otto, 2010). *S. aureus* is a mostly transient coloniser of human skin, found on 30% of the population at any given time (Eriksen *et al.*, 1995). However persistent carriage is associated with adverse infectious sequelae and inflammatory skin disease. For example, *S. aureus* is reported to colonise all patients with atopic dermatitis, whereby colonising density correlated with active disease severity, and lesion resolution associated with microbial clearance (Guzik *et al.*, 2005).

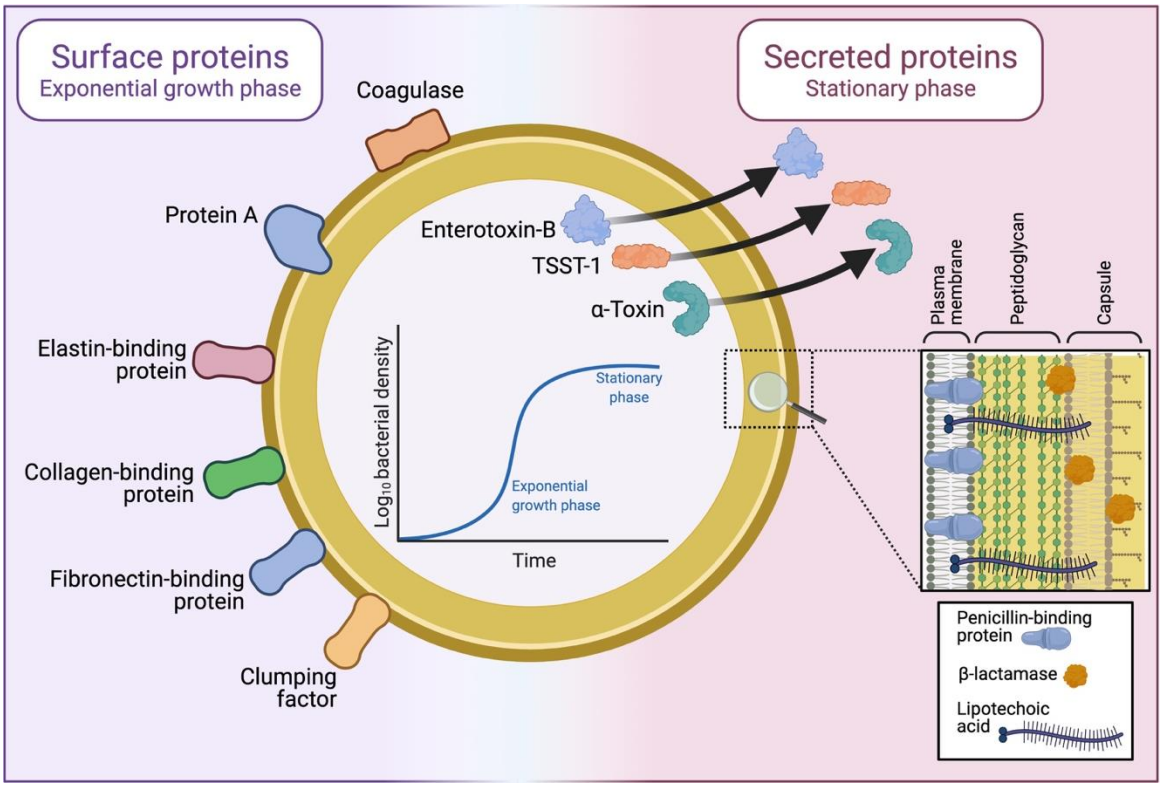


Figure 1.5. Structure of *Staphylococcus aureus* and protein expression during exponential and stationary growth phases. Adapted from Lowy, 1998. Created using biorender.com

Unlike *S. epidermidis*, and as discussed above, *S. aureus* possesses an arsenal of aggressive virulence factors that enable severe pathogenesis of infection (Dinges, Orwin and Schlievert, 2000). *S. aureus* causes a range of local and systemic infections in humans, including bacteraemia, meningitis, toxic-shock syndrome, skin and soft tissue infections such as impetigo, cellulitis and necrotising fasciitis

(Tong *et al.*, 2015). *S. aureus* biofilm formation on medical and prosthetic devices also causes severe infection (Percival *et al.*, 2012; Tong *et al.*, 2015).

Interactions between host and *S. aureus* in cutaneous tissues are of great interest to understanding inflammatory skin disease, especially the role of this microbe in the pathogenesis of atopic dermatitis. *S. aureus* sensing by keratinocyte PRRs, such as TLR2, but also TLR9, NOD1 and NOD2, induces inflammatory responses is altered in atopic skin (Bitschar *et al.*, 2017), while DAMPs elicit the immediate release of anti-*S. aureus* AMPs, hBD3 and RNase7, which are dysregulated by cutaneous inflammation (Kisich *et al.*, 2007). Further, hBD3 host-directed opposition to *S. aureus* is partly countered by microbial virulence factors (Bitschar *et al.*, 2017). Moreover, the burden of *S. aureus* infection disrupts the colonising microbiome, leading to the preferential overgrowth of *S. aureus* and the loss of cutaneous microbial diversity.

The crosstalk between bacterial sensing and activation of inflammatory responses in the epidermis is, therefore, multi-dimensional. The signalling pathways through which these aspects of cutaneous biology are conducted are well characterised, for example, keratinocyte TLR activation leading to NFκB-activated transcription of pro-inflammatory cytokines and AMPs, or activation of T_H1 responses via IRF7 transcription. Similarly, the immune pathways activated during cutaneous inflammation attributable to atopic dermatitis or psoriasis have been well elucidated. However, the role of cutaneous inflammation in disturbing host responses to the microbiome and contributing to circuits of pathogenesis is appreciated but not fully explored. A gap in knowledge remains in how these phenotypes overlap and alter the local epidermal gene expression profile.

Work to address this knowledge gap from both a cytokine and microbial perspective is underway but has largely been limited to simplified *in vitro* investigation involving monolayer keratinocyte cultures, single cytokine incubation experiments, or challenge with bacterial products not live inoculation, for example (Secor *et al.*, 2012; Swindell *et al.*, 2012). More complex *in vitro* modelling of three-dimensional epidermal responses to bacterial species, even by simulation of barrier breach, do not adequately resolve the complexity of these systems *in vivo* (Duckney *et al.*, 2013). Less well understood is the contribution and interplay of multiple populations of keratinocytes derived from the different stratified compartments, those of distinct immunophenotypes and those responding differentially to microbial interaction to cutaneous homeostasis and health, and how these apply to *in vivo* programmes underlying inflammatory disease in the skin. Development of a pipeline to understand immunophenotypes and bacterial sensing responses across multiple populations of keratinocytes for overlaying – or deconvolution – from *in vivo* biopsy data promises

to better characterise the crosstalk of these signatures, their role in maintaining cutaneous health and their contribution to local inflammation.

1.4 Transcriptomic cellular signatures

1.4.1 Transcriptomic in the era of public data

Analysis of gene expression has greatly improved research approaches to understanding tissue biology and disease. The accessible nature of the skin for biopsy has made this tissue a prime target for bulk transcriptomic profiling to address the limits of knowledge in inflammatory skin diseases. Two transcriptomic technologies remain contemporarily dominant: microarrays and RNAseq (Schena *et al.*, 1995; Miller and Tang, 2009; Wang, Gerstein and Snyder, 2009). Microarrays were the first widely used tool for large-scale transcriptomic research and have provided insight into the gene expression profiles of most human tissues and cells (Hoek *et al.*, 2004; She *et al.*, 2009; Swindell *et al.*, 2012; Thomas *et al.*, 2012; Li *et al.*, 2016) and major human diseases, including multiple cancers (Golub *et al.*, 1999; Singh *et al.*, 2002; Van 't Veer *et al.*, 2002), autoimmune diseases (Rozzo *et al.*, 2001; Lock *et al.*, 2002; Puccetti *et al.*, 2018), and infectious diseases (Berry *et al.*, 2010; Banchereau *et al.*, 2012; Parnell *et al.*, 2012). Indeed, one of the most popular microarray platforms, Affymetrix Human Genome U133 Plus 2.0 Array has over five thousand associated datasets comprising a total of over 140,000 samples ran since the array became available in 2003.

The popularity and potential of whole transcriptome studies led to the development of strict protocols for making this vast amount of data publicly available and useful for future re-analysis. Publications making use of microarray or RNAseq data must make the data available in repositories such as the Gene Expression Omnibus (GEO, NCBI). Datasets deposited in GEO must be compliant with the MIAME (or MINSEQE; for microarrays or RNAseq, respectively) standards (Brazma *et al.*, 2001; Edgar and Barrett, 2006). These guidelines are for making available the **m**inimum **i**nformation **a**bout a **m**icroarray **e**xperiment; known as the MIAME standards. There are six main components of MIAME: (1) raw data, (2) final processed, normalised data as used in the primary study, (3) appropriate sample annotation, (4) correct array annotation, (5) matched sample-expression data, (6) details of analysis pipeline used in the primary study. MIAME-compliance provides a level of confidence that re-analysis of publicly available data can be undertaken with sufficient annotation and supplementary information to make such analysis worthwhile, and that conclusions and findings are similarly valid.

While microarrays and RNAseq ultimately measure the gene expression profile of the whole transcriptome, the technologies differ quite significantly. Microarrays are based on the hybridisation of cDNA transcripts to an array composed of pre-designed complementary oligonucleotide probes. A commonly used Affymetrix platform, the Human Genome U133 Plus 2.0 Array, for example, has 54675 unique probes representing 20955 genes. Expression is then detected and quantified through the measurement of fluorophore or chemiluminescent labelling. The longevity, relative low cost for accurate quantification of the entire transcriptome, and comprehensive ancillary repositories, bioinformatic tools and pipelines has made microarrays the most successful and mature approach for genomic analysis (Trevino, Falciani and Barrera-Saldaña, 2007). RNAseq, in contrast, is the direct measuring of transcripts by high-throughput sequencing technologies (Wang, Gerstein and Snyder, 2009). Briefly, the total sample RNA is converted to a cDNA library, which then, with or without amplification, is sequenced using a high-throughput approach. The first advantage of RNAseq over microarrays is the need to no longer rely on limiting transcript detection corresponding to known, existing sequences. From this, RNAseq has the potential missing from microarray approaches to reveal sequences variations, such as SNPs and splice variants (Cloonan *et al.*, 2008; Morin *et al.*, 2008).

1.4.2 Transcriptomic analysis pipelines

The information obtained from microarrays and RNAseq is, at the most basic level, the same: gene expression across the whole transcriptome. As such, the downstream pipelines of both these technologies are often very similar. Indeed, the maturity of microarrays compared to RNAseq has meant the former led the way in establishing pipelines and developing bioinformatic tools to interpret and make use of vast amount of data these approaches deliver. Measuring total gene expression of a sample tissue across time, health and disease provides an amount of data that has revolutionised understanding of human disease. Such highly dimensional data has driven the innovation of integrating biological knowledge and clinical information with informatics disciplines of mathematical, computer and statistical sciences.

Transcriptomic analysis produces very highly dimensional data. As such, molecular phenotyping of the skin in this way promises to further explore the aetiology and pathogenesis of inflammatory disease. The skin is composed of complex networks of cellular crosstalk and communication that maintains cutaneous homeostasis. The globality of this signature can be captured from the expression pattern of the whole transcriptome using microarrays or RNAseq as described above. Many studies have assessed the perturbation of the investigated system by defining differentially expressed genes between disease states and healthy skin. This comparative expression analysis has provided novel insight into the transcriptomic and molecular mechanisms of disease. A wealth of

these studies have been conducted in skin owing to its ease of access for biopsy, and, for the most common inflammatory skin conditions, there exists list of differentially expressed genes compared to unaffected and/or healthy skins (Suárez-Fariñas *et al.*, 2011; Choy *et al.*, 2012; Coda *et al.*, 2012; Gittler *et al.*, 2012; Bigler *et al.*, 2013; Swindell *et al.*, 2013; Ghosh *et al.*, 2015; Martel *et al.*, 2016).

1.4.3 Deconvoluting cellular signatures from the heterogenous bulk tissue transcriptome

Using more advanced characterisation of the transcriptomic data it is possible to interrogate this whole tissue profile for signatures derived from cellular populations. It is possible to characterise the profiles of the cell subsets in a given tissue. As with differentially expressed genes, this higher-level analysis can be used to track cell-based transcriptional changes associated with disease phenotypes. Therefore, large-scale alteration of the local skin transcriptome can be dissected and deconvoluted for cellular meaning. Uncovering cellular interactions of the skin in the context of health, inflammation and homeostatic microflora responses will be key to understanding immune crosstalk in the skin, host-microbiome interactions and the pathology of skin disease.

An approach to characterising cellular signatures from the whole transcriptome is to identify co-expression within subsets, or clusters, of the expression data. Networks of genes can be constructed based on “guilt-by-association” co-expression patterns that can then be interrogated for biological meaning. This bioinformatic deconvolution of expression data characterises signatures in an unsupervised, non-hypothesis driven manner. Genes are first clustered by gene-gene Pearson correlation with a threshold set to retain the maximum number of genes with the lowest number of connections across the network. A clustering algorithm, such as Markov clustering (MCL), is then used to define refined co-expression patterns across gene groups(van Dongen, 2000; Doig *et al.*, 2013; Mabbott *et al.*, 2013).

Unsupervised co-expression analysis has previously been used to define a so-called *SkinSig* of 20 gene signatures(Shih *et al.*, 2016). These signatures were applied to 18 different skin diseases and was able classify these diseases into three distinct groups based on the collective expression patterns across the 20 signatures. Known biological processes associated with cutaneous aging such as decreased keratinisation and fatty acid metabolism were shown to be caused by loss of hair follicles and sebaceous glands. This demonstrated that within the whole transcriptomic data there remains discrete cell-based signatures that may be deconvoluted to provide biologically meaningful insight and in-context characterisation of cellular profiles.

Unsupervised deconvolution of cellular signatures has limitations, however. Deriving co-expression signatures directly from the analysis transcriptome requires careful implementation across

datasets. Large batch effects would render merging of datasets almost impossible due to current statistical limitations of modelling true batch effects compared to valuable biological artefacts. Most current methods for such batch correction often remove too much biological variation as an overzealous attempt to remove technical, artificially introduced variation. For example, merging of multi-disease datasets may force out intra-disease heterogeneity in order to provide smooth comparisons between diseases. As such, meta-analysis of cellular signatures is currently the most promising option for tracking cell-based signatures.

There can, however, be much discordancy between lists of differentially expressed genes that underscore the curation of these signatures for a given disease. Random background noise variation, within and across disease, biological heterogeneity, experimental and technical differences can compound efforts to derive a comparable co-expression profile of what may be assumed to be the same cellular signature. These confounding factors may amplify the technical variation between datasets to such an extent that meta-analysis of co-expression signatures proves highly challenging.

Another computational approach for characterising relative cell subsets from complex bulk transcriptomic data is to use a reference-based approach of pre-defined signatures that can then be traced across different samples and datasets. Advantageously, these tools do not rely on the integration of data for comprehensive comparison and analysis. One such tool for this called **C**ell-type **I**dentification **B**y **E**stimating **R**elative **S**ubsets **O**f **R**NA **T**ranscripts, CIBERSORT (Newman *et al.*, 2015). This tool has the advantage of both deconvoluting custom expression profiles (i.e. bulk samples of interest) and curation of custom cell reference signatures (i.e. expected cell populations found in the bulk sample of interest). When used with a comprehensive and properly curated set of reference signatures, characterising cell composition from bulk tissue has shown to be a powerful approach for deriving individual signatures from whole transcriptome data (Newman *et al.*, 2015; Chen *et al.*, 2018; Félix Garza *et al.*, 2019; Steen *et al.*, 2020; G. Wang *et al.*, 2020).

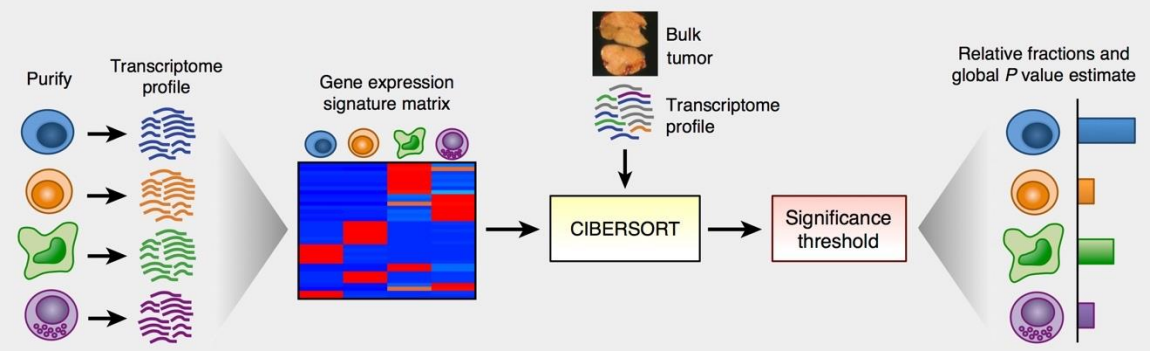


Figure 1.6. Schematic diagram of the CIBERSORT algorithm. From Newman *et al.* (2015).

Several deconvolution algorithms have been published that, with varying success, resolve bulk transcriptomic profiles into constituent signals representative of tissue-residing cell populations (Abbas *et al.*, 2009; Shen-Orr *et al.*, 2010; Clancy *et al.*, 2017). The basis of CIBERSORT is a supervised learning algorithm using a version of a support vector machine (SVM) approach called linear support vector regression (ν-SVR). This type of SVM approach was found to be highly robust to noise (Avila Cobos *et al.*, 2018). Further, CIBERSORT has been favourably and comprehensively compared to other deconvolution approaches: linear least-squares regression (LLSR), quadratic deconvolution (QD), perturbation model for gene expression deconvolution (PERT), robust linear regression (RLR), microarray microdissection with analysis of differences (MMAD) and digital sorting algorithm (Abbas *et al.*, 2009; Gong *et al.*, 2011; Qiao *et al.*, 2012; Zhong *et al.*, 2013; Liebner, Huang and Parvin, 2014). While all of these approaches are capable of achieving accurate results from idealised sample mixtures, CIBERSORT was shown to better deconvolute against noise, partially unknown sample composition and closely related cell types (Newman *et al.*, 2015).

It is these advantages that make CIBERSORT a highly versatile approach for *in silico* estimation of cell types from bulk gene profiles of any given tissue. This flexibility inherent to CIBERSORT means that robust deconvolution of a tissue of interest into enumerated constituent cell types can be obtained relatively simply by a reference-based approach. The reference cellular signatures can be curated and tailored to unique tissues, such as tumours, the blood, or skin to suit specific needs in order to address precise research questions. For example, the same samples can potentially be deconvoluted multiple times to, firstly, resolve heterogeneous populations of the same cell type and, secondly, these same populations redefined by separate transcriptomic reference signatures. This is extremely advantageous for investigating transcriptomic changes driven by multiple components, as might be expected of human cells. Such driver components for transcriptome alteration may include life cycle and differentiation status, activation status, cytokine modulation, immune interaction, host-microbiome interactions, longitudinal temporal changes and treatment effects. Each of these would require separate reference profiles to deconvolute the same samples but would offer the power to view the transcriptome of subsets of cells from different perspectives.

1.5 Hypothesis & Aims

1.5.1 Hypothesis

Pathogen sensing is altered between healthy and inflamed epidermis, and this is underpinned by differences in molecular crosstalk between structural and immune cutaneous populations.

1.5.2 Aims

- To identify immunophenotypes of keratinocyte populations within bulk skin transcriptome data using *in silico* deconvolution methods.

- To define, by means of single-cell DropSeq, signatures for bacterial sensing in keratinocytes in the context of healthy skin.

- To track pathogen and commensal sensing signatures in the context of healthy skin and cutaneous inflammation.

1.5.3 Objectives

- High-level whole tissue analysis using the *in-silico* pipeline and tools developed in this project will enable re-exploration of data to determine the contribution of pathogen-sensing pathways to inflammation in the skin.

- Initial bacteria sensing signatures will be characterised from available public data and bulk data of model epidermis before undertaking DropSeq scRNA-seq of *Staphylococcal* challenged model epidermis.

- Model epidermis of keratinocytes will be challenged with either commensal *S. epidermidis* or pathogenic *S. aureus* to identify transcriptomic alterations induced by challenge at the single-cell resolution.

- Utilise gene signatures identified from model challenge with *Staphylococcal spp.* for application as reference signatures to biopsy datasets of bulk skin transcriptome data for population subset deconvolution.

Chapter 2 Methods

2.1 Data analysis of microarray transcriptomic data

2.1.1 Raw expression data

Microarray were obtained from either the Gene Expression Omnibus (GEO, NCBI) or from in-house sources. All datasets used were analysed from raw data in R/Python and normalised according to the technologies used. For Affymetrix microarray platforms, classic microarray quality control was performed using the *arrayQualityMetrics* tool of Bioconductor. Data were normalised to obtain expression values by RMA/GCRMA methods. Illumina platforms and other technologies were quantile normalised using the *lumi* or *limma* Bioconductor packages. If datasets were to be merged, correction for the introduced batch effect using the *combat* function in the *sva* Bioconductor package was performed (Johnson, Li and Rabinovic, 2007; Leek *et al.*, 2012).

2.1.2 Reference skin cell populations datasets

To curate the skin signature composed of purified cells as a reference, datasets from GEO were collected. The datasets used for these signatures were: GSE36248 (keratinocytes stimulated with IFNα, IFNγ, IL-13, IL-17A, IL-4, TNFα, and unstimulated control), GSE7216 (keratinocytes stimulated with IL-1β, IL-22, IL-26, keratinocyte growth factor (KGF)), GSE34308 (primary dermal fibroblasts), GSE4570 (melanocytes), GSE74158 (skin resident $CD4^+$, $CD8^+$ and regulatory T cells), GSE49475 (activated Langerhans' cells, $CD11c^+$ dermal dendritic cells), and GSE23618 (steady-state Langerhans' cells). Datasets were normalised individually before combining into a single matrix for use in CIBERSORT as a cutaneous reference panel for deconvolution of bulk transcriptomic samples.

2.1.3 Bulk skin datasets

Datasets of skin biopsies from inflammatory skin diseases were also obtained from GEO.

2.1.3.1 Baseline inflammatory datasets

GSE32924 is a dataset of paired chronic lesional (CAL, n=11), non-lesional (ANL, n=11) atopic dermatitis samples, and healthy controls (HH, n=8). GSE34248 is a dataset of acute lesional (AAL, n= 7), chronic lesional (CAL, n=7), non-lesional (ANL, n=7) skin from atopic dermatitis patients and healthy controls (HH, n=6).

2.1.3.2 Inflammatory datasets in the context of treatment

GSE11903 is a longitudinal study of etanercept treatment (twice weekly, 50mg subcutaneous) of psoriasis patients with biopsies from both lesional and non-lesional sites at baseline, and the effect of treatment monitored from lesional biopsies taken at a further 1, 2, 4 and 12 weeks of treatment. Patients were grouped as responders based on histologic examination of decreased epidermal thickness and loss of Ki67, a proliferation marker, and keratin 16, a differentiation marker. There were 11 responders and four non-responders.

GSE58558 is a longitudinal study of cyclosporine treatment (5mg/kg daily) of atopic dermatitis patients with biopsies taken at baseline, 2 and 12 weeks of treatment from both lesional and non-lesional sites.

GSE130588 is a longitudinal study of dupilumab or placebo treatment (weekly, 200mg subcutaneous). Biopsies from both lesional and non-lesional sites was taken at week 0, 4 and 16, with healthy controls included for comparison. Atopic dermatitis patients were allocated to treatment or placebo cohorts randomly in equal ratio.

2.1.4 Quality control and filtering of microarray data

Raw microarray data were quality controlled and normalised according to platform-specific guidelines. CEL files from Affymetrix arrays were assessed using the *AffyPLM* in R/Bioconductor to compute probe-level relative log expression (RLE) and normalised unscaled standard error (NUSE). Subsequently expression data were background corrected and normalised using RMA/GRMA; depending on available computational power. Illumina arrays were variance stabilised and quantile normalised using the *lumi* R/Bioconductor package. Other platforms, such as Agilent SurePrint G3 Human G3 v3 array, raw data was processed using the *limma* R/Bioconductor package for background correction and quantile normalisation.

2.1.5 Dimension reduction interrogation

Following quality control and normalisation, samples were visually assessed for underlying biology or specific effects on the global transcriptome. Normalised microarray data was analysed using principal component analysis (PCA) and/or multi-dimensional scaling (MDS) for visualisation of similarity and sample grouping in two-dimensions. To provide linked clustering across the samples within a dataset, expression data were analysed by hierarchical clustering using the *hclust* package in R. The clustering parameters used were Euclidean distance and complete linkage to best recapitulate the clustering used by PCA and MDS analysis.

2.1.6 Differential gene expression for microarray data

Differential gene expression analysis was obtained using the *limma* package. Briefly, gene-wise log2 expression values were fit to a linear mixed-model. Comparisons were tested for significance using a moderated *t*-test corrected for multiple hypothesis testing by Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995). A gene was considered differentially expressed if the corrected p-value did not exceed 0.05 and the log fold-change was greater than 1.5 in either direction.

2.1.7 Barrier expression panel

Table 1. Panel of epidermal barrier-related genes.

Probe ID	Gene Symbol	Gene
218182_s_at	*CLDN1*	Claudin 1
222549_at	*CLDN1*	Claudin 1
228704_s_at[NOTE 1]	*CLDN23*	Claudin 23
228706_s_at[NOTE 1]	*CLDN23*	Claudin 23
228707_at	*CLDN23*	Claudin 23
206192_at	*CDSN*	Corneodesmosin
206193_s_at	*CDSN*	Corneodesmosin
207324_s_at	*DSC1*	Desmocollin
206642_at	*DSG1*	Desmoglein 1
215704_at	*FLG*	Filaggrin
205900_at	*KRT1*	Keratin 1
207023_x_at	*KRT10*	Keratin 10
210633_x_at	*KRT10*	Keratin 10
213287_s_at	*KRT10*	Keratin 10
207720_at	*LOR*	Loricrin

NOTE 1: Not included in final barrier panel

Barrier integrity was assessed by proxy of the expression a panel of marker genes. A total of nine genes were selected to assess barrier expression (**Table 1**). All atopic dermatitis samples and associated healthy controls were assessed using the same Affymetrix HGU133plus2.0 array technology, as such the panel of genes was selected from the probe identifiers shown in Table 1. Two probes, both representing *CLDN23*, were excluded from the panel, 228704_s_at and 228706_s_at, because they were observed to be lowly expressed across the samples and difficult to distinguish from background.

Across each probe, the highest expressing sample was determined and scaled to 100. For each probe within each sample, the expression was normalised according to its respective maximum.

Barrier expression was then interpreted for each sample as the average normalised expression across the panel of 13 probes representing 9 genes.

2.2 Unsupervised co-expression analysis and network clustering

The inflammatory skin disease datasets GSE32924, GSE36842 and GSE34248 were processed as above and subsequently merged and batch corrected using the COMBAT tool within the *sva* Bioconductor package. Differential gene expression analysis between healthy controls and lesional skin, within disease lesional and non-lesional, and across lesional comparisons was conducted as above also. The expression values of 4620 probeset-IDs, corresponding to 3058 unique genes were input into MIRU (now, GraphiaPro, Kajeka, Edinburgh, UK) for network analysis. A transcript-to-transcript correlation matrix using Pearson correlation coefficient of $r \geq 0.7$ was created. The resulting non-directional network graph was then clustered into group of genes using the Markov Clustering algorithm (MCL) at an inflation value of 3.1 and a minimum cluster size of 10 probesets, giving 50 clusters. These thresholds best defined the clusters by granularity, biological meaning and significance of gene annotation. The gene list for each cluster was interrogated for gene ontology using the web-based analysis tool ToppFun within the ToppGene suite, annotating the input genes lists for biological process terms with a Benjamini-Hochberg p-value significance lower than 0.05 (Chen *et al.*, 2009). To produce a single ontology term per cell-based cluster, the REVIGO online tool was used to summarise the top 200 or fewer statistically significant biological process terms output from ToppGene (Supek *et al.*, 2011). Terms were selected by both lowest Benjamini-Hochberg P-value and a term dispensability of zero. The most significant indispensable annotation was summarised by presenting the Benjamini-Hochberg p-value in the negative $\log_{10}$ space where significance is given by values above 1.3. Expression data for each of the seven cell-based clusters was plotted across the five tissue types within GraphPad Prism 8 (California, USA). Within Prism 8, normality testing performed using the D'Agostino-Pearson method indicated the use of nonparametric Kruskal-Wallis ANOVA with multiple comparison performed.

2.3 Deconvolution enrichment analysis using xCell

The inflammatory skin disease datasets GSE32924, and GSE34248 were processed as above and used independently for investigating the efficacy of enrichment deconvolution using the xCell tool (Aran, Hu and Butte, 2017). The tool, which can be run via a web-based application or locally in R by using the available code, includes a panel of 64 cell types to infer enrichment from supplied bulk data. In order to limit the panel to the cell types relevant to cutaneous analysis, deconvolution of the inflammatory samples was performed manually via the R code. The code was run using default arguments to establish the working pipeline before limiting the cell reference panel to eight

profiles. The cell types included for skin deconvolution are $CD4^+$ T cells, $CD8^+$ T cells, T_H1 cells, T_H2 cells, dendritic cells, sebocytes, fibroblasts and keratinocytes. Enrichment scores for each of the inferred cell types was plotted across the five tissue types within GraphPad Prism 8 (California, USA).

2.4 Deconvolution analysis using CIBERSORT

2.4.1 Preparation of reference samples

Reference panels for each use of CIBERSORT are described in the appropriate results section and accompanying figures. Preparation of reference samples followed the same initial processing steps unless otherwise stated. Replicate samples of reference cell types were averaged by mean expression for each gene, and matrices were subsequently merged on common genes preferentially selecting the gene with the highest expression across the samples.

2.4.2 Running the CIBERSORT algorithm

Datasets of bulk skin samples were deconvoluted against the reference signature sets using the online version of the CIBERSORT algorithm (https://cibersort.stanford.edu/) (Newman *et al.*, 2015). Reference signature files were provided as the *signature gene file*, while normalised expression data of bulk samples were provided as the *mixture file*. All run settings were kept at default, with the exception of the maximum number of barcode genes being set to 10227 to ensure whole transcriptomic signature derivation.

The output provided by CIBERSORT was downloaded as a .txt file, where relative abundance of each cellular signature was normalised to present as a per cent of the whole sample.

2.5 Cell culture

2.5.1 Primary keratinocyte monolayer culture

Primary human epidermal keratinocytes (single juvenile donor, HPEK, CellnTec, Switzerland) were cultured in 15ml KGM2 in a T75 flask and maintained at 37°C, 5% CO_2. Standard KGM2 culture media (**Table 2**) was changed three times weekly to a maximum passage of 70-80% confluency. The initial two passages were used for cell expansion and cryopreservation, while later passages used for experiments up to passage four for model generation.

Table 2. Cell culture media.

Culturing media	Supplements
Keratinocyte Growth Medium 2 (KGM2) *Standard culture media for monolayers*	Keratinocyte Growth Medium 2 kit (PromoCell, UK) supplemented with pre-prepared: • 0.004ml/ml bovine pituitary extract • 0.125ng/ml epidermal growth factor (recombinant human) • 5μg/ml insulin (recombinant human) • 0.33μg/ml hydrocortisone • 0.39μg/ml epinephrine • 10μg/ml transferrin (recombinant human) • 0.06mM $CaCl_2$
Keratinocyte Growth Medium 2 (KGM2) *Differentiation-optimised culture media*	Standard KGM2 culture media with added: • 2% heat-treated foetal bovine serum (Gibco) • 1.8mM $CaCl_2$ (Sigma)
Keratinocyte Growth Medium 2 (KGM2) *Model challenge-optimised culture media*	Standard KGM2 culture media with added: • 2% heat-treated foetal bovine serum (Gibco) • 1.8mM $CaCl_2$ (Sigma) • Without supplementation of hydrocortisone

2.5.1.1 Monolayer keratinocyte passage protocol

Keratinocyte monolayers were rinsed with room temperature phosphate buffered saline (PBS), before addition of 5ml TrypLE Express (Gibco) at 37°C, 5% CO_2 for a maximum of eight minutes until cells were observed to be rounded and detaching upon microscopic inspection. TrypLE Express was neutralised using 10ml 1:1 PBS KGM2 solution and transferred to a centrifuge tube. Cells were placed in a centrifuge (eight minutes at 160 x g) to form a pellet, after which cells were resuspended in fresh standard KGM2. Viable keratinocytes were quantified by trypan blue staining and manual counting using a haemocytometer. Further culture required a seeding density of 2-2.5x10^5 cells per T75 flask.

2.5.1.2 Cryopreservation and thawing of primary keratinocytes

Keratinocytes not used for immediately for experimental procedures were cryopreserved for storage. Cells were aliquoted to 5x10^5 per bank vial in 250μl FBS (Gibco, UK) which was supplemented to a volume of 500μl with 90% FBS, 10% DMSO (Sigma, UK) as a freezing mixture to protect cells from formation of ice crystals. Vials were placed in a Mr Frosty (Nalgene) container with isopropanol and frozen at minus 80°C for 24-72 hours before transfer for long term storage in the cold vapour phase of liquid nitrogen.

Cryopreserved aliquots of keratinocytes ($5x10^5$ cells) were thawed by gentle warming under running warm water and transferred into a centrifuge tube containing 20ml PBS. Cells were placed in a centrifuge (eight minutes at 160 x g) to form a pellet and remove diluted FBS-DMSO freeze solution, after which cells were resuspended in fresh 15ml standard KGM2 for culture in a T75. Culture media was replaced every second day until the desired confluency was achieved before passaging as above.

2.5.2 Reconstituted human epidermis model

The three-dimensional reconstituted human epidermis (RHE) model was originally developed and optimised by D. Holbrook (2019) and is illustrated in **Figure 2.1**.

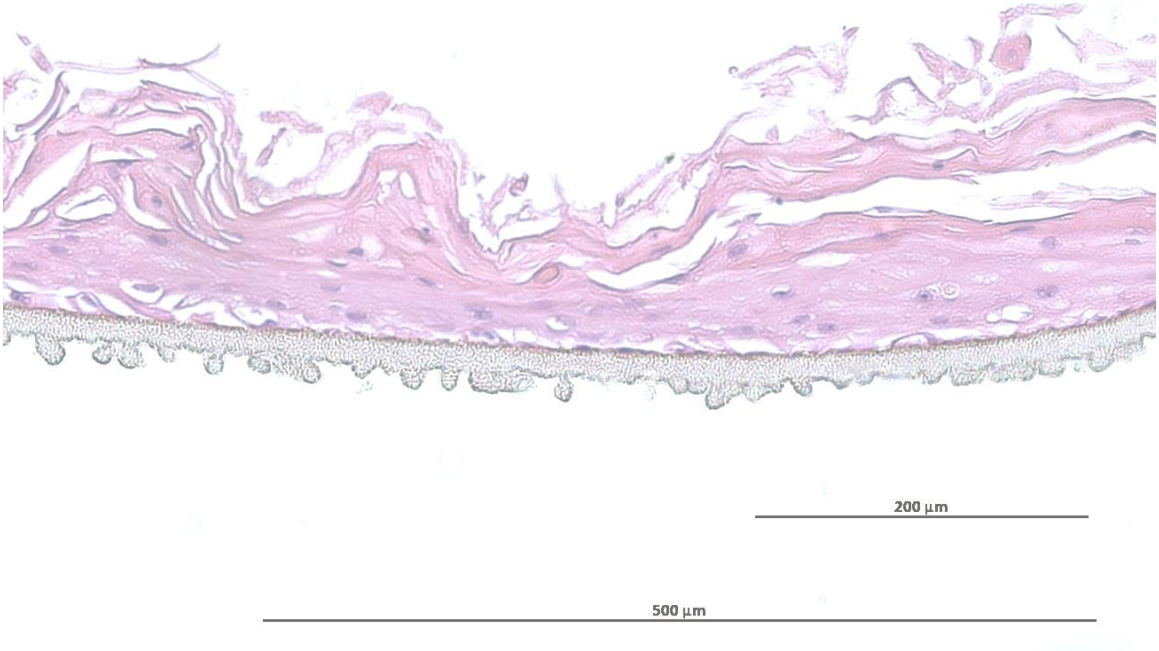


Figure 2.1. RHE model morphology. Representative RHE model morphology at day 14 of culture visualised by H&E staining.

2.5.2.1 Model generation

Primary keratinocytes from freshly detached monolayer culture were suspended in 400µl standard KGM2 culture media per required insert and seeded at a density of $2x10^5$ per insert onto polycarbonate membrane inserts (Millicell, 0.4µm pore size, 12mm diameter) (Merck Millipore, UK). Batches of up to 15 inserts per 10cm petri dish, with 25ml standard KGM2 culture media in the dish were grown in a semi-sealed humid container within a 37°C, 5% CO_2 incubator. Monolayers were allowed to establish for 2-3 days.

Following monolayer establishment of models, one insert was sacrificed to test for confluency. Confluency was confirmed using the standard protocol of the CnT-ST-100 stain kit (CellnTec, Switzerland). The stained insert was assessed for confluency by microscopic judgement by eye. If confluency was confirmed, the remaining batch of inserts was allowed to undergo the differentiation protocol. Lack of confluency resulted in further incubation of inserts monolayers for 24 hours. Subsequent lack of confluency indicating deficient proliferation would render the batch discarded.

To initiate model differentiation, culture media from within the inserts and within the dish were removed and replaced with KGM2 supplemented with 2% FBS and 1.8mM $CaCl_2$ (Sigma, UK) (differentiation optimised KGMS2, **Table 2**). Differentiation-supplemented culture media was added to the inserts at a volume of 400μl and 25ml to the petri dish. The dish was returned to semi-sealed humid container within a 37°C, 5% CO_2 incubator for 24 hours.

Following the 24 hours of differentiation protocol, culture media was removed from the inserts and petri dish. Fresh KGM2 with differentiation supplements was added only to the dish to a volume of 6ml and no media was returned to the inserts to lift the forming models to the air-liquid interface (ALI) and encourage formation of the cornified layer. The dish was returned to semi-sealed humid container within a 37°C, 5% CO_2 incubator for 24 hours.

ALI cultures were left to generate the full RHE model over 14- or 15-days subsequent days with replacement of the KGM2 differentiation-supplemented culture media every second day. From approximately day 5, models were visually inspected to assess for the forming cornified layer indicated by a switch from a shiny to a matt surface appearance.

2.5.2.2 Model dissociation

Models were intended for analysis by single-cell RNA-sequencing by the DropSeq protocol, thus requiring model dissociation and single-cell suspensions. A three-step washing followed by warm-cold TrypLE Express dissociation protocol was optimised for this work, requiring three wells of a 12-well plate for each insert to be dissociated. Inserts were placed in a well containing 5ml room temperature PBS for 15 minutes, followed by transfer to a second well containing 5ml room temperature PBS, 0.1% EDTA for another 15 minutes. Inserts were removed from the wells and wash solution removed by pipetting. Inserts were then placed in a third well containing 1ml pre-cooled TrypLE Express, and 500μl pre-warmed TrypLE Express added within the insert for 15 minutes. After this time 1ml cold PBS, 10% FBS was added within the insert followed by vigorous pipetting with the insert angled to allow pooling of dissociation solutions into the well. The

dissociation solution was filtered through a 70μm strainer and collected in a centrifuge tube. Cells were placed in a centrifuge (seven minutes at 350 x g) to form a pellet, after which cells were resuspended in fresh KGM2. Viable dissociated model keratinocytes were quantified by trypan blue staining and manual counting using a haemocytometer.

2.6 Bacterial culture

For challenge experiments of RHE models, bacteria suspensions were used. Bacterial suspensions and quantification were performed throughout using the same protocol.

2.6.1 *Staphylococcal* strains

The main *Staphylococcal* strains used in the work undertaken in this thesis were provided by *S. aureus* NCTC 8325-4 and *S. epidermidis* ATCC 12228 (Bæk *et al.*, 2013; MacLea and Trachtenberg, 2017). Reanalysis of bulk transcriptomic data of RHE models challenged with other *Staphylococcal spp.* undertaken by D. Holbrook (2019) was also performed including *S. aureus* strain ATCC 29213, and *S. capitis* strain ATCC 27840.

2.6.2 Enumerating *Staphylococcal* suspensions

Bacterial lawns from *S. aureus* NCTC 8325-4 and *S. epidermidis* ATCC 12228 were harvested and resuspending in 50ml PBS. These initial stocks were diluted to reach a minimum optical density measured at visible 600nm wavelength (OD_{600}) of approximately 1.500 using a spectrophotometer. Subsequent 1:2 dilutions were performed until a 1:16 dilution reached, with each dilution being quantified by spectrophotometer at OD_{600}. These solutions are quantifiable by spectrophotometry, but plated solutions would yield too many colonies for counting. Therefore, the 1:2 serial dilutions were further diluted serially 1:10 to reach a $1:10^5$ dilution. In triplicate, 100μl of the $1:10^5$ of each of the 1:2 serial dilutions were spread onto TSA plates and incubated overnight at 37°C.

The following day, the plates were counted for enumeration of colony-forming units (CFU) per 100μl, extrapolated to CFU/ml of the $1:10^5$ dilution, and finally CFU/ml of each 1:2 dilution. For each of the 1:2 dilutions of the CFU/ml were plotted against the quantified OD_{600} measured prior to the dilution protocol. Within GraphPad Prism 8, concentration curves were generated of OD_{600} by CFU/ml to interpolate the OD_{600} value of a $1x10^8$ CFU/ml suspension from both bacterial species. A non-linear semi-log standard curve was fit for interpolation. Goodness of fit (R-squared) was 0.9450 for the *S. aureus* standard curve, and 0.8331 for the *S. epidermidis* data, from which interpolated values were derived directly from the curve values provided by Prism. A $1x10^8$ CFU/ml solution of

S. aureus NCTC 8325-4 was quantified at OD_{600} 0.091 ($1.011x10^8$ CFU/ml, **Figure 2.2A**), and for *S. epidermidis* ATCC 12228 at OD_{600} 0.254 ($1.001x10^8$ CFU/ml, **Figure 2.2B**).

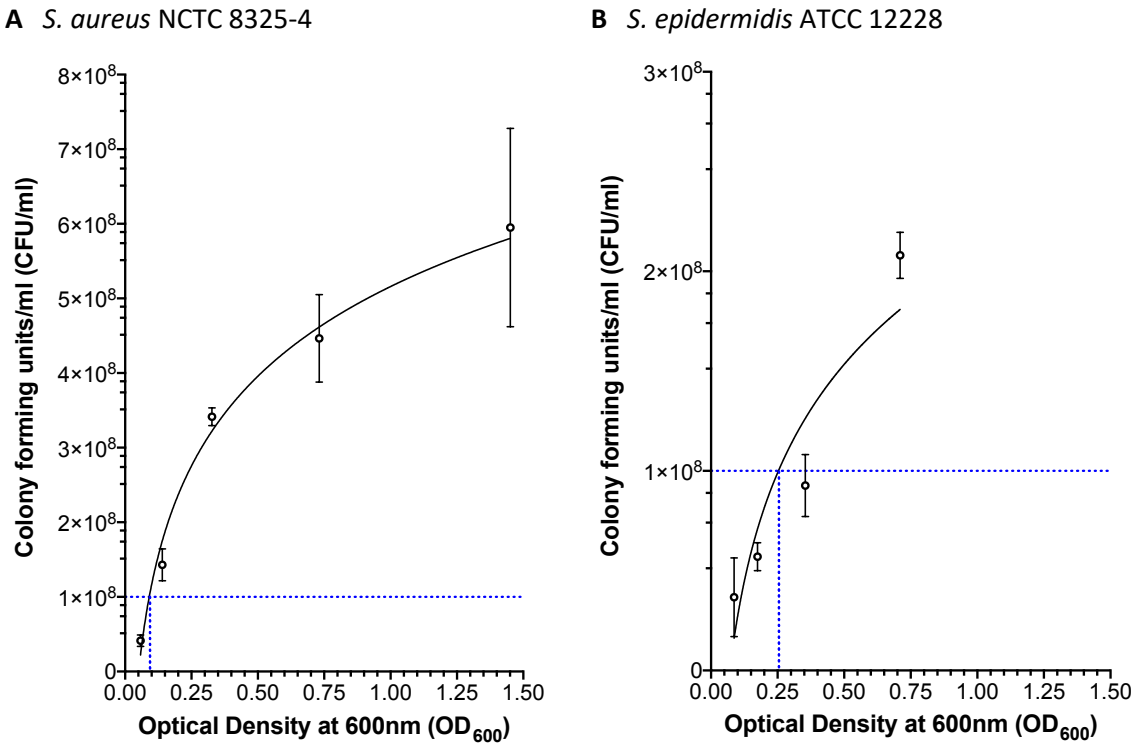


Figure 2.2. Concentration curves for bacterial suspensions. Bacterial suspensions from **[A]** *S. aureus* NCTC 8325-4, and **[B]** *S. epidermidis* ATCC 12228 were serially diluted 1:2 and OD_{600} measurements taken for 1:1 to 1:16 suspensions. Suspensions were further 1:10 serially diluted to $1:10^5$ for plating in triplicate and counting of colonies (CFU/ml). Counted CFU/ml were plot against respective OD_{600} measurements to allow interpolation of the OD_{600} value for a $1x10^8$ CFU/ml working suspension.

2.6.3 *Staphylococcal* culture and suspension production

The day prior to requiring bacteria suspensions, a small sample of frozen stock was streaked on a prepared plate of tryptic soy agar (TSA) (Sigma, UK). Streaked plates were incubated at 37°C overnight to establish stationary phase bacterial lawns. Approximately one-quarter of the plate lawn was harvested into 30ml PBS and gentle vortexing to break up the bacterial aggregate. Bacterial suspensions were diluted to reach an (±0.01) OD_{600} measured by a spectrophotometer to indicate a $1x10^8$ CFU/ml suspension for each species, as described above. A further working stock was then generating by a 1:10 dilution forming $1x10^7$ CFU/ml suspensions. For *Staphylococcal* challenge of RHE models, $1x10^6$ CFU of bacteria were required, therefore 100µl of the corresponding $1x10^7$ CFU/ml suspension was taken.

2.7 Challenge of model epidermis using *Staphylococcal spp.*

After 14 or 15 days of ALI culture, RHE models demonstrating matt surface texture indicative of formation of stratum corneum and barrier were taken for bacterial challenge experiments. Models that were observed to contain moisture or wetness inside the insert were discarded.

2.7.1 Preparation of models for challenge experiments

The day before challenge experiments, models were transferred to a new petri dish and 10ml of KGM2 containing differentiation additives but without hydrocortisone supplementation was added (challenge optimised KGM2 media, **Table 2**). The dish was returned to semi-sealed humid container within a 37°C, 5% CO_2 incubator for 24 hours.

2.7.2 Bacterial challenge protocol

Models were removed from the petri dish and placed, individually, in a well of a 24 well plate containing 300μl KGM2 containing differentiation additives but without hydrocortisone supplementation. Prepared PBS suspensions of *S. aureus* and *S. epidermis* containing $1x10^7$ CFU/ml were obtained as described above. To each insert, as per experimental design, 100ul of either PBS, *S. aureus* or *S. epidermidis* stock was added to the apical aspect of the model within the insert. Challenged models were therefore incubated with approximately $1x10^6$ CFU each. Models were incubated at 37°C, 5% CO_2 for three hours, after which each was washed three times with 400μl PBS and replaced to the incubator for a further 21 hours. Models were then dissociated to single cell suspensions as described in Section 2.5.2.2.

2.7.3 Culture undernatant harvest

The 300μl culture media used during the challenge experiment was removed from each of the wells of the 24-well plate and aliquoted into an Eppendorf tube per well. Undernatants were stored for further experiments, if required, at minus 80°C.

2.8 DropSeq

2.8.1 Antibody labelling for sample multiplexing

Multiplexing of samples to enable a single encapsulation was advantageous for compensating for the low cell numbers from model dissociation and achieve the 120 cells/μl required concentration for DropSeq encapsulation. Further, multiplexing removed the need to for dissociated samples to

be placed on ice, with the reduced cell viability this causes, due to the throughout limits of the encapsulation protocol. Sample multiplexing was achieved by tagging cells with barcode antibodies following the Cell Hashing protocol (Stoeckius *et al.*, 2018). Each sample after dissociation and prior to encapsulation was pelleted by centrifugation at 350 x g for seven minutes, and cells resuspended in 48μl Cell Staining Buffer (Biolegend, UK). Keratinocytes were incubated with addition of 5μl FcR blocker (Miltenyi, UK) for ten minutes at 4°C. Subsequently, 2μl of multiplexing Hashtag antibody (Biolegend, UK) was added to each sample; control samples were tagged with TotalSeq-A0251 (GTCAACTCTTTAGCG), *S. aureus* challenged samples were tagged with TotalSeq-A0252 (TGATGGCCTATTGGG), and *S. epidermidis* challenged samples were tagged with TotalSeq-A0253 (TTCCGCCTCTCTTTG). Samples were incubated with Hashtag antibodies for 30 minutes at 4°C. Hashtag antibodies recognise human CD298 and β2-microglobuin and are conjugated to sample-specific flanking oligonucleotide sequences for downstream demultiplexing by aligning to a dictionary of Hashtag sequences. Cells were prepared for DropSeq encapsulation following the Hashtag protocol after two washes using 500μl Cell Staining Buffer.

2.8.2 Microfluidic encapsulation of single cells and barcode microbeads

The single-cell co-encapsulation with primer-coated barcode micro-bead procedure was followed as developed for DropSeq (Macosko *et al.*, 2015). The microfluidic device channels two aqueous solutions of a cell suspension and a micro-bead suspension through an oil channel to form nanolitre-sized droplets. Co-encapsulation should aim for one cell and one bead per droplet. Droplet formation is monitored by placing the microfluidic device on the upturned stage of a 10x microscope (Olympus, UK). The cell suspension was prepared at a concentration of 120 cells/μl in PBS, 0.02% BSA and loaded into a 1ml syringe (Henke-Sass Wolf, HSW, Germany) and connected to the cell inlet of the microfluidics device. The micro-bead suspension was prepared by suspending the primer-coated and barcoded beads (ChemGenes, USA) at a concentration of 120 micro-beads/μl in 1ml lysis buffer (50μl 1M DTT (Sigma-Aldrich, UK), 40μl 0.5 EDTA (Fisher Scientific, UK), 100μl 2M Tris pH 7.5 (Sigma-Aldrich, UK), 10μl 20% sarkosyl (Sigma-Aldrich, UK), 300μl 20% Ficoll (Sigma-Aldrich, UK) and 500μl nuclease-free water) and connected to the cell inlet of the microfluidics device. The micro-bead suspension was loaded into a 3ml syringe (HSW, Germany) and connected to the bead inlet of the microfluidics device. Cell and bead suspensions were maintained by addition of a small stirrer magnet into each syringe and placement of magnetic plates near the pump-mounted syringes. The oil used for droplet generation (Bio-Rad, UK) was loaded into a 5ml syringe (HSW, Germany) and connected to the oil inlet of the microfluidics device. Inlet and outlet connections to the microfluidics device were made using a 0.38mm diameter polyethylene tubing (Smiths Medical, Fisher Scientific). A rig of three motors with attached pumps was used to hold and plunge the syringes at specified flow rates: oil at 14000 μl/hr, cell and micro-bead

suspensions at 4000 μl /hr. The outlet from the microfluidics device – the co-encapsulated cells and microbeads emulsion – was collected in a 50ml Falcon tube. Samples of outflow were collected to visually assess uniformity of droplet formation under a light microscope.

2.8.3 Microbead purification

During cell and micro-bead co-encapsulation, the cells are lysed and the cellular mRNA content coats the primer-coated micro-beads. The micro-beads remain in emulsion following the DropSeq procedure. The outlet from the encapsulation procedure was collected in a 50ml Falcon tube and allowed to settle on ice for 30 minutes, after which, the bottom oil layer was discarded. Droplet breakage was performed by adding 1ml perfluorooctanol (Sigma-Aldrich, UK) in 30ml room temperature SSC (Thermo Fisher, UK) before shaking up to five times. The released beads SSC mixture was vacuum filtered, followed by 30ml SSC wash of the Falcon tube to collect any remaining beads. The beads are retained on the outside of the disposable filter unit, which is further washed with 30ml SSC. The beads are then collected from the filter by resuspending in 1ml aliquots of SSC and apportioned to 1.5ml DNA low-bind tubes (Eppendorf, UK). A minimum of eight 1ml SSC washes of the filter unit was performed to reclaim the maximal number of beads.

The collected supernatant across the minimum of eight tubes was centrifuged at 1000rcf for one minute to pellet the micro-beads. The SSC supernatant was discarded, and the pellets sequentially collected in 1ml SSC and transferred to a single, fresh 1.5ml DNA low-bind tube (Eppendorf, UK). The collected beads were then pelleted by centrifugation at 1000rcf for one minute and washed twice using 1ml 6x SSC and once with 300μl 5x RT buffer.

2.8.4 cDNA library generation

To the pellet of collected micro-beads from one sample, 200μl of RT mix (75μl nuclease-free H_2O, 40μl Maxima 5x RT buffer, 40μl 20% Ficoll PM-400, 20μl 10mM dNTPs, 5μl RNase inhibitor, 10μl 50μM template switch oligo (TSO), and 10μl Maxima H- RTase) was added. The beads-RT mix was incubated with rotation at room temperature for 30 minutes, followed by a 90-minute incubation at 42°C with rotation. The collected beads were then pelleted by centrifugation at 1000rcf for one minute and the RT mix supernatant was removed. The micro-beads were then washed with 1ml x TE-SDS (10mM Tris pH 8.0, 1mM EDTA, 5% SDS) and then twice with 1ml TE-TW (10mM Tris pH 8.0, 1mM EDTA, 0.01% Tween-20). The micro-bead pellet was then resuspended in 200μl exonuclease mix (1ml 10mM Tris pH 8.0, 20μl 10x Exo I buffer, 10μl Exo I, 170μl ddH_2O) for incubation at 37°C for 45 minutes with rotation. The collected beads were then pelleted by centrifugation at 1000rcf for one minute and the exonuclease mix supernatant was removed. Finally, the micro-beads were

washed again with 1ml TE-SDS and then twice with 1ml TE-TW. The samples of micro-beads were counted using a Fuchs-Rosenthal cell counter and made into aliquots of 2000 micro-beads (equivalent of 100 STAMPS) ready for later PCR in a volume of 50μl PCR mix (25μl 2x Kapa HiFi Hotstart Readymix, 0.4μl 100 SMART PCR primer, 24.6μl H_20).

The aliquots were placed in a thermocycler following the PCR programme given in (**Table 3**). The second denaturing, annealing and extension protocol can be varied by number of cycles from 12-16 depending on the cell type used and to yield sufficient quantities of cDNA (>100pg/μl) using the fewest number of PCR cycles to avoid amplification bias of high abundant transcripts. cDNA libraries were purified using AMPure XP magnetic beads (Beckman Coulter, UK); 30μl of AMPure beads were added to the PCR samples and vortexed for five minutes followed by room temperature incubation for five minutes. The PCR tubes were then staged onto a 96-well magnetic micro-tube plate to hold the AMPure bead-bound cDNA. After two minutes on the magnetic plate, the supernatant was removed, and the tubes washed twice with 200μl freshly-made70% ethanol. The ethanol wash was dried by evaporation and the AMPure beads eluted in 10μl nuclease-free H_2O. The quantity and quality of the cDNA library was assessed on a Bioanalyzer DNA High-Sensitivity Chip (Agilent, UK).

Table 3. PCR protocol for cDNA library amplification.

Phase	Cycles	Temperature, °C	Time, mm:ss
Denaturation	1	95	03:00
Denaturation	4	98	00:20
Annealing		65	00:45
Extension		72	03:00
Denaturation	12-16	98	00:20
Annealing		67	00:20
Extension		72	00:30
Extension	1	72	05:00

2.8.5 Preparation of cDNA library for next-generation sequencing

The cDNA fragments were prepared for sequencing by preparing aliquots of 600pg cDNA in a volume of 5μl H_2O. Tagmentation using the Nextera XT DNA library preparation kit (Illumina, USA) was used to label the cDNA libraries with unique indices. To each sample, 10μl of Nextera TD buffer and 5μl of Amplicon Tagment enzyme was added then incubated at 55°C for five minutes. After incubation, 5μl neutralisation buffer was added and incubated at room temperature for five minutes. In the order of 15μl Nextera PCR mix (Illumina, USA), 8μl H_2O, 1μl 10μM New-P5-SMART

PCR hybrid oligo, and 1μl 10μM Nextera N70X oligo (Illumina, USA) was added to the sample for PCR tagmentation (**Table 5**). The PCR products were purified using AMPure XP magnetic beads and eluted in 10μl H_2O and quantified using a Bioanalyzer DNA High-Sensitivity Chip.

Table 4. PCR protocol used for cDNA Nextera XT tagmentation.

Phase	Cycles	Temperature, °C	Time, mm:ss
Denaturation	1	95	00:30
Denaturation	12	95	00:10
Annealing		55	00:30
Extension		72	00:30
Extension	1	72	05:00

Table 5. Barcoded primer bead and primer sequences used in the DropSeq protocol. (Macosko *et al.*, 2015)

Name	Sequence
Barcoded Bead SeqB	5'–Bead–Linker-TTTTTAAGCAGTGGTATCAACGCAGAGTACJJJJJJJJJJJJNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'
TSO	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG
SMART PCR primer	AAGCAGTGGTATCAACGCAGAGT
New-P5-SMART PCR hybrid oligo	AATGATACGGCGACCACCGAGATCTACACGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGT*A*C
Custom Read 1 primer	GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC

2.8.6 Next generation sequencing and transcript alignment

The Custom Read 1 primer was diluted using 6μl 100μM Custom Read 1 primer stock added to 1994μl HT1 buffer (Next-Seq 500/550 v2.5 kit, Illumina) to a working concentration of 0.3μM. Tagmented cDNA libraries were pooled to a concentration of 2.3nM. 10μl of 0.2N NaOH was added to 10μl of the sequencing pool, which was vortexed and briefly centrifuged. The pool-NaOH mix was incubated at room temperature for five minutes, after which 10μl 0.2M Tris was used to neutralise the NaOH. After neutralisation, 970μl ice-cold HT1 buffer was added to the sequencing pool and mixed, then 130μl was transferred to a tube containing prepared 1170μl HT1 buffer, on ice. The entire volume of the prepared library and working stock of Custom Read 1 primer were loaded onto an Illumina Next-Seq 500/550 v2.5 kit high output 75 cycles reagent cartridge. The reagent cartridge was loaded into a NextSeq sequencer for a paired-end run using 3.6-5.0x10^4 reads

per cell, 20bp for Read 1, 60bp for Read 2 and 8bp for Index 1. Sequencing was performed at the Wessex Investigational Sciences Hub laboratory, University of Southampton, UK.

Sequencing output files obtained and transferred to a user account in the Iridis High Performance Computing cluster (Iridis 4, University of Southampton, UK). Illumina-generated BCL (base call files) containing raw sequencing data were converted to fastq format by demultiplexing, removal of UMIs and transcript identification using the bcl2fastq tool. The resulting read files, read 1 and read 2 for each of the gene expression data and for the HashTag barcode tags, were aligned using Kallisto (v.0.46.1) and Bustools (v.0.39.3) to target sequences contained in an index FASTA formatted file built during the running of each alignment pipeline. Aligned count matrix data were generated and stored as clean annotated data (anndata, v0.7.1) as an h5ad file for input into scRNA-seq analysis pipeline.

2.9 Data analysis of RNA-seq single cell transcriptomic data

Transcriptomic data from scRNA-seq was analysed, unless otherwise stated, using the python-based Scanpy framework (Wolf, Angerer and Theis, 2018)(v1.4.6). Barcodes derived from DropSeq droplets containing background, ambient RNA were filtered without compromise to small cells containing little mRNA using EmptyDrops (Lun *et al.*, 2019). Count matrix data for each batch in the form of h5ad files were saved at this point for subsequent merging.

Count matrix data were loaded for each DropSeq experiment batch and were concatenated using the 'outer' joining argument to retain maximal number of genes. Quality control covariates were established such as number of counts per cell, number of genes per cell and per centage of genes being mitochondrial. Minimum and maximum values for these covariates were used to reduce barcode counts to the number of cells expected from the DropSeq experiments, while cells with a >15% mitochondrial gene count were assumed to be apoptotic so were filtered out. Expression data were then normalised by SCRAN method (Lun, Bach and Marioni, 2016). Data are first subclustered on the top 15 principal components using a Leiden resolution of 0.5, then counts are normalised per cell whilst excluding very highly expressed genes. Size factors are then scaled by SCRAN normalisation of the original raw data using the cluster annotations just previously identified by Leiden. Highly variable genes (HVGs) were identified using the Seurat dispersion-based method categorising genes by placing into bins defined by scaled mean and standard deviation and selecting the most highly variable genes per bin. To limit overfit of downstream batch integration, the number of highly variable genes was limited to 1500.

For visualising of reduced dimensions sufficiently explaining the variation across the data, a single-cell neighbourhood graph was computed using the top 50 principal components and 20 nearest neighbours. The DropSeq experiment batches were then integrated using the batch balanced k-nearest neighbour method (BBKNN) (Polański *et al.*, 2020), which calculates the top neighbours for each cell within each batch independently of the rest of the within-batch cells. The data were then visualised in the reduced dimensional space by computation of Uniform Manifold Approximation and Projection (UMAP) and observed by scatter plot. Localised topology was visualised by calculating embedded density for both DropSeq experiment batch origin and bacterial challenge condition to aid selection of clustering resolution for optimal granularity explaining cell biology. Clustering was performed by the Leiden algorithm using a resolution of 0.5 for the whole dataset and a resolution of 0.2 on a restricted subset to properly reflect observed difference observed from bacterial challenge (Traag, Waltman and van Eck, 2019).

Layer-defining epidermal makers were used to annotate the gross stratification of the single-cell data: basal markers were *KRT5*, *KRT14*, *COL17A1*; spinous makers were *KRT1*, *KRT10*; and granular markers were *KRT6A*, *KRT6B*, *KRT16*, *S100A7*, *S100A8*, and *S100A9*. Following identification of the basal, spinous and granular clusters, further markers demarcating between layer clusters were identified by ranking genes between these using a t-test method. Differential expression analysis between the more granular clusters was performed on a within-layer basis. This differential expression analysis was performed using MAST (BH p-value <0.01, log FC>±1) (Finak *et al.*, 2015). MAST is a robust tool for single-cell differential expressional analysis that takes into account the binomial nature of single-cell data using a two-part generalised linear model for discrete and continuous expression. Gene ontology analysis of the top 20 marker genes of the layer clusters or the entire list of DEGs for each comparison was undertaken using the web-based analysis tool ToppFun within the ToppGene suite, annotating the input genes lists for biological process terms with a Benjamini-Hochberg p-value significance lower than 0.05. Where relevant, the REVIGO online tool was used to summarise the top 200 or fewer statistically significant biological process terms output from ToppGene

UMAP cluster connectivity was visualised by computing partition-based graph abstraction (PAGA), which reduces the UMAP scatter plot to defined clusters as a single node each connected by edged weighted by connection confidence (Wolf *et al.*, 2019)

Chapter 3 Deconvolution of human skin into constituent cell populations using *in silico* methods

3.1 Introduction

The use of transcriptomics for understanding gene expression changes underlying cutaneous disease and inflammation has not only provided new insight into pathogenesis but has created a wealth of data yet to be fully exploited to maximum potential. One aspect of cutaneous research is to understand the interplay and crosstalk of the heterogeneous mixture of keratinocyte populations composing the tissue. So far, many investigations of the cutaneous transcriptome during inflammation have been limited to molecular-level analysis. Crucially keratinocyte heterogeneity is lost in such bulk analysis as gene expression is collapsed to a single value represented across the sample. The ability to derive, from the bulk transcriptome, the inferred contribution of different subsets and populations is, therefore, a powerful tool for expanding analysis from the molecular to the cellular. In atopic dermatitis and psoriasis, for example, the role of polarised helper T cell activation is predicted to impose characteristic immunophenotypes upon local keratinocytes. A computational method to deconvolute the relative enumeration of these cell types from bulk transcriptomic data of inflammatory skin samples is of great interest to maximising analysis from public microarray data.

Microarray and RNA sequencing technology allows us to survey the global transcriptome of interest to extract previously unknown processes and mechanisms of disease. Large-scale alteration of the local skin transcriptome can be dissected by the simultaneous sampling of tens of thousands of genes conducted by microarrays. Such molecular phenotyping of the skin promises to further explore the aetiology and pathogenesis of disease. The sampling of the whole transcriptome of many phenotypes and samples, produces very highly dimensional data, which presents significant challenges of its own, for example, data visualisation and application of appropriate statistical approaches. However, the wealth of data collated presents the valuable opportunity of applying deconvolutional approaches to these using an *in-silico* pipeline without necessity of single-cell studies to resolve heterogeneity.

Several studies (Suárez-Fariñas *et al.*, 2011; Choy *et al.*, 2012; Coda *et al.*, 2012; Gittler *et al.*, 2012; Bigler *et al.*, 2013; Swindell *et al.*, 2013; Ghosh *et al.*, 2015; Martel *et al.*, 2016) have comprehensively surveyed the transcriptome of skin in both psoriasis and AD to produce lists of differentially expressed genes that define lesional, involved skin from non-lesional, uninvolved skin,

and both from healthy skin. However, there can be much discordancy between the list of genes found in these studies. Random background noise variation, biological heterogeneity, clinical guidelines, experimental and technical differences, and variant bioinformatics analysis all play a part in producing this discordance and can even call into question the existence of a core disease phenotype (Cahan *et al.*, 2007). Increasingly, datasets are made available in public repositories such as the Gene Expression Omnibus (GEO, NCBI) (Barrett *et al.*, 2011). This allows researchers to conduct large-scale integrative analyses in an effort to increase the statistical power of differential expression analysis. Integration of such diverse and sometimes disparate for meta-analyses can prove difficult relying heavily on sample compositions and careful removal of batch effects while minimising effect on true biological variation. The aim of these merging analyses is to create a more robust list of differentially expressed genes that refines those produced by single studies in order to detect true variations between phenotypes and alleviate the problem of the conflicting results found from the separate studies before meta-analysis (Tian *et al.*, 2012).

Deconvolution of cellular signatures from bulk expression data offers another meta-analysis approach. This can be achieved by either unsupervised or hypothesis-driven machine learning methods. A common unsupervised approach is to identify co-expression within subsets, or clusters, of the expression data. One such approach that utilises both arbitrary Pearson correlation coefficient cut-offs and the Markov Clustering Algorithm (MCL) has been demonstrated to robustly interrogate gene expression data for biological meaning (van Dongen, 2000; Freeman *et al.*, 2007). Networks can then be constructed and visualised by connecting genes – as nodes – by edges that define co-expression based upon the assumption that genes linked in a common pathway or biological process have an analogous expression pattern (van Noort, Snel and Huynen, 2003). Hence, by virtue of identifying co-expressed genes, these algorithms allow identification of gene sets with coordinated expression involved in specific biological processes or specific cell types.

Cluster deconvolution by transcript-to-transcript co-expression analysis is a tool for identifying molecular signatures from a bulk transcriptome. Using this approach, a hypothesis develops that whole tissue skin samples can be deconvoluted *in silico* in order to identify cell-specific transcriptomic signatures that can be used to understand cell behaviour during cutaneous inflammation.

Hypothesis-driven deconvolution allows the resolution of pre-defined signatures of interest from the bulk transcriptome. For example, this approach allows the direct enumeration of inferred immunophenotypes or bacteria sensing subset of keratinocytes, which are the over-arching aims of this work. One such tool available for application of defined reference populations for resolving

from provided bulk skin samples is CIBERSORT (Newman *et al.*, 2015). CIBERSORT is a supervised learning algorithm for linear classification that uses a version of support vector machine called linear support vector regression (ν-SVR). CIBERSORT leverages a regression-based problem to define a regression line through the expression data using a subset of genes from the expression matrix as support vectors. Genes that fall within a defined boundary of the regression line, the ε-tube, are discounted as being within the limit for error. The advantage of this approach is the consequential feature selection as support vectors (genes) define the ε-tube by their outlier nature, satisfying the linear regression function. Critically, for its use in resolving multiple close-related keratinocyte phenotypes in the context of the remaining cutaneous cellular mixture, CIBERSORT was demonstrated to be optimal for such problems (Newman *et al.*, 2015).

The first part of this work aimed to identify distinct signatures from immune and keratinocyte populations from deconvoluted expression data. Given the importance of inflammatory cell biology underpinning atopic dermatitis and psoriasis, the representation of the identified signatures across the skin disease bulk transcriptome was evaluated. Such signatures would be derived from a curated merging-analysis of microarray studies in order to minimise technical variation and increase sample numbers for improved statistical interpretation. The aim is two-fold: first, use and refine an *in-silico* approach to identify cell-specific signatures in the whole skin transcriptome. Second, develop an understanding of how the disease-altered transcriptome varies in order to generate signatures that reflect a disease-specific phenotype.

This first aim can demonstrate the ability of computational biology to identify gene expression profiles in whole skin biopsies that are cell type specific and disease-specific, as an alternative to other approaches such as blood-transcriptome sampling or profiling single cells by laser capture microscopy or single-cell RNA-sequencing. While the practicality and cost of scRNA-seq continues to improve to make this kind of approach more commonplace, there exists a huge repository of legacy data of which can still be of great use for understanding the cellular and molecular mechanisms underpinning disease. In order for this analysis to be undertake, first the efficacy of *in silico* approaches must be demonstrated. Furthermore, developing understanding by computationally derived transcriptomic signatures can complement current understanding from a new perspective, generate data for the stratification of phenotypes or patients, provide new insight to the mechanisms leading to the perturbed tissue homeostasis, and lead to new areas of interest for treatment approaches.

3.1.1 Hypothesis

Whole tissue skin samples can be deconvoluted *in silico* in order to identify cell-specific transcriptomic signatures that can be used to understand cell behaviour during cutaneous inflammation.

3.1.2 Aims

- Identify datasets of atopic dermatitis and psoriasis for integration and curation of tissue defining differentially expressed genes.

- Use these differentially expressed genes to investigate networks of co-expressed genes for biological or cellular meaning.

- Investigate the strength of the immune and structural cell derived signatures from both non-hypothesis-driven and hypothesis-drive deconvolution approaches.

- Curate a panel of cellular signatures relevant to cutaneous resolution to replace the default LM22 panel used by CIBERSORT.

3.2 Results

3.2.1 Disease severity, but not lesion formation, defines the clustering of AD samples at the whole transcriptome level

To investigate the atopic dermatitis-associated transcriptomes at the global level, and in order to understand alterations in the gene expression caused by atopic dermatitis microarray data from whole skin biopsies of healthy and atopic dermatitis patients was investigated. The dataset chosen for the first analysis consisted of 11 paired lesional and non-lesional atopic dermatitis samples with eight healthy controls (GSE32924 (Suárez-Fariñas *et al.*, 2011)). It has been previously described that non-lesional atopic tissue is phenotypically distinct from both disease lesions and skin of healthy controls. However, in cases of severe disease it can often be difficult to distinguish skin as uninvolved; that is, lesional and non-lesional classifications may become irrelevant as most of the skin becomes affected.

The 11 paired atopic dermatitis samples recorded an average SCORAD (SCORing of Atopic Dermatitis index of lesion and disease severity) of 57.9 (range: 28 to 97.5), seven of which had a SCORAD above 50, classifying severe disease. The sample clustering of the Suarez-Farinas *et al.* dataset was confirmed by MDS plotting showing that healthy controls form a tight homogenous cluster separate to the atopic samples (**Figure 3.1A**). The atopic dermatitis sample clustering is more complex and as expected, given the generally severe disease of the patients, does not neatly separate by lesional or non-lesional tissue. In order to determine if lesional, non-lesional distinctions could be observed within sample sub-clusters, we used hierarchical clustering to produce dendrogram plots (**Figure 3.1B**).

Hierarchical clustering shows that at the highest level of similarity (i.e. nearest neighbour) four of the 11 samples (patients: 2, 4, 6 and 10) are found in their pairs of lesional and non-lesional samples (**Figure 3.1B**; yellow boxes). This pairing is independent of disease severity, with samples found across the full range of the SCORAD indicator. Overall, eight patients (patients 2, 4, 6, 10, 12, 13, 14) show remarkably close MDS clustering of their lesional and non-lesional sample. These samples span mild, moderate and severe disease, and demonstrates the non-lesional phenotype having a transcriptional profile very similar to lesional skins. This pairing indicates that the active disease within atopic lesions may be sub-clinical in non-lesions during an active disease flare. To distinguish between lesions and non-lesions the delineation is mostly found in the second component of the MDS plot. Here there is a scale from healthy samples to non-lesional samples to lesional samples along the PC2 axis.

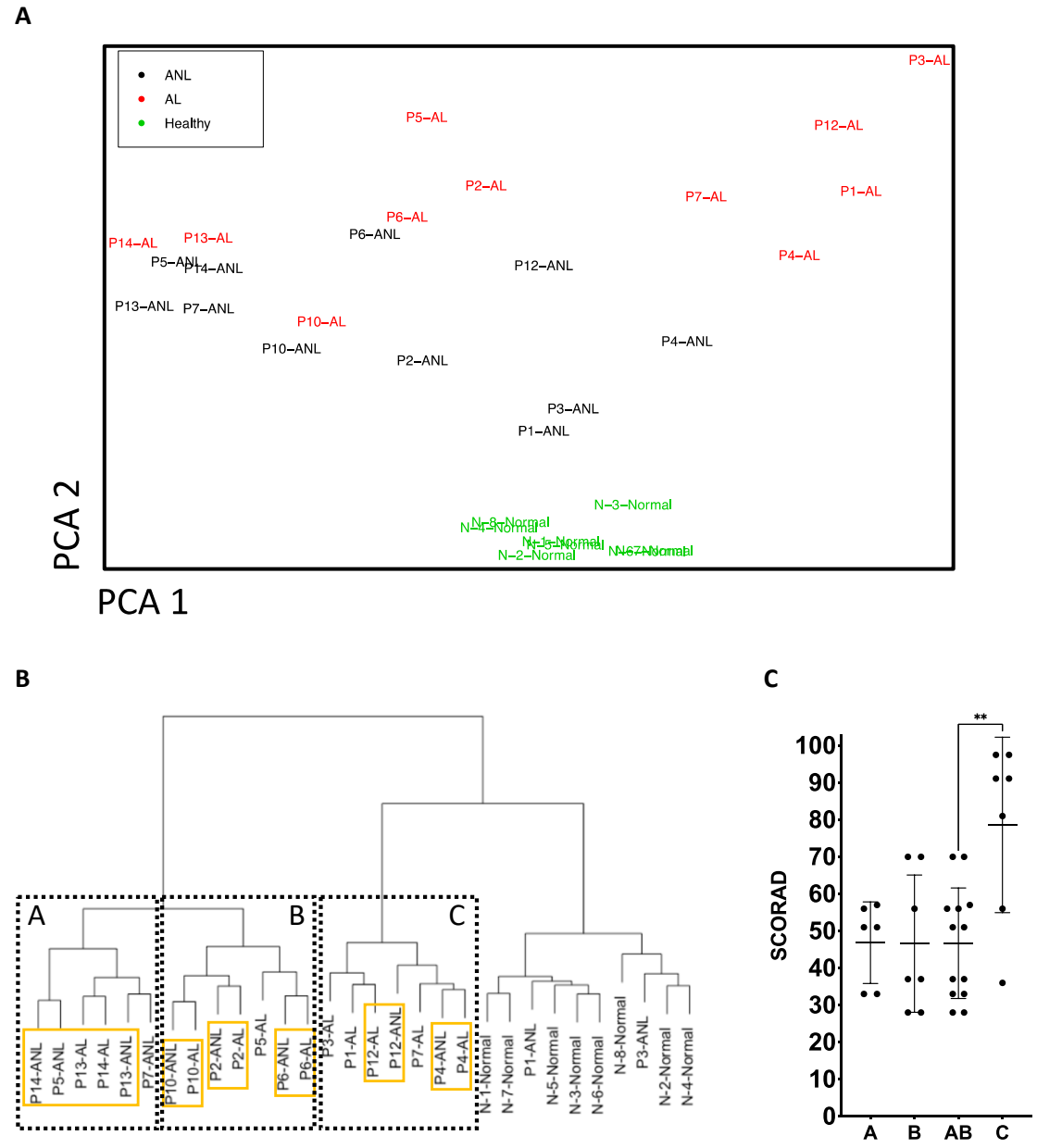


Figure 3.1. Dimensionality reduction of lesional and non-lesional AD samples compared to healthy controls from GSE32924. [A] MDS plot of the dataset showing. Chronic AD lesional (*n*=11, AL, red); AD non-lesional (*n*=11, ANL, black) and healthy controls (*n*=8, Normal, green). **[B]** Hierarchical clustering dendrogram of the dataset obtained using the *hclust* function in the R package *stats* using the "ward.D" agglomeration method, Euclidean distance and complete linkage. The branches of the dendrogram define the groups A, B and C. Yellow squares indicate close similarity between indicated paired lesional and non-lesional samples. **[C]** Disease severity of the patients represented in this dataset measured by the *SCOR*ing of *AD* index (SCORAD). Samples are grouped according to that observed by hierarchical clustering in B. Mean ±SD represented.

Hierarchical clustering splits the samples into three main groups: two groups (A, B) found off one major branch and the third group (C) found off the other major branch shared with healthy controls. Disease severity underscores this clustering (**Figure 3.1C**). Group A has an average SCORAD of 46.8 compared to group B, 46.6; defining both as moderate disease, while group C is defined as severe disease with a SCOARD of 78.6. The SCORAD difference between group A and B was not significant (p= 0.9764). Combining groups, A and B give samples with a mild-moderate SCORAD index of 46.7 compared to the severe SCORAD of group C; the difference in SCORAD between group AB against group C was significance (p=0.0016). Interestingly, the group with the highest average SCORAD, group C, clusters closest to the healthy controls.

3.2.2 Impaired barrier is characteristic of atopic dermatitis samples independent of lesion status

The primary purpose of the epidermis and keratinocytes is to maintain a barrier against the external environment. Perturbation of the barrier is characteristic of atopic dermatitis. The most important protein involved in this dysfunctional barrier is filaggrin (gene; *FLG*). Loss-of-function mutations and/or T_H2-mediated loss of expression contribute to the reduced filaggrin expression in atopic dermatitis (Kabashima, 2013). A number of other barrier-related components are also reported to be affected by atopic inflammation, these include keratin (*KRT1*, *KRT10*) intermediate filaments that mediate inter-cellular barrier with desmosomes and corneodesmosomes (*DSG1*, *DSC1*, *CDSN*) alongside cell-cell tight junction claudins (*CLDN1*, *CLDN23*), and finally loricrin (*LOR*) that helps stabilise the structures forming the barrier in the cornified envelope (De Benedetto *et al.*, 2011; Hönzke *et al.*, 2016; Bao *et al.*, 2017; Totsuka *et al.*, 2017).

Using a panel of these nine epidermal genes known to be affected in atopic dermatitis (Section 2.1.7), the epidermal barrier was assessed in health and across atopic dermatitis progression (**Figure 3.2**). The expression of the panel was normalised to account for inherent expression changes of these transcripts and to present a proxy of barrier expression inferred as a per centage.

Based on expression levels of genes within the barrier signatures, healthy barrier expression was determined to be 73.2$\pm$8.89 (SD, normalised units) (**Figure 3.2**). At all stages of atopic dermatitis, non-lesional, acute and chronic lesions, expression of the barrier panel was significantly reduced compared to healthy controls (all comparisons to healthy controls, p<0.0001). Most atopic samples had a barrier expression lower than even the lowest healthy samples. Barrier status did not significantly change between non-lesional and chronic lesions. Barrier expression was reduced in acute lesions compared to non-lesional skin and chronic lesions, but this did not reach significance. These data, therefore, recapitulate that atopic skin has reduced barrier compared to healthy skins.

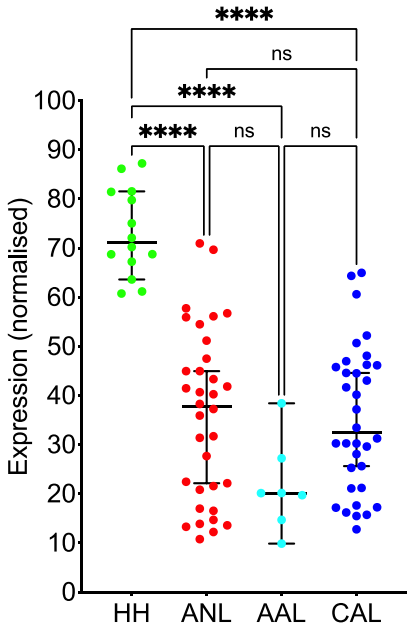


Figure 3.2. Barrier gene expression in healthy and AD skins. Normalised average expression of a panel of skin barrier genes known to be impaired in AD: claudin 1 (*CLDN1*), claudin 23 (*CLDN23*), corneodesmosin (*CDSN*), desmocollin (*DSC1*), desmoglein-1 (*DSG1*), filaggrin (*FLG*), keratin 1 (*KRT1*), keratin 10 (*KRT10*), loricrin (*LOR*). Samples from three datasets of healthy controls (n=14, HH), AD non-lesional (n=34, ANL), AD acute lesional skin (n=7, AAL), AD chronic lesional skin (n=34, CAL). *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show Kruskal-Wallis test with multiple comparisons adjusted p-value. Error bars show median ±95% confidence interval.

3.2.3 Unsupervised network clustering reveals keratinocyte-based inflammation status.

Investigation of cutaneous inflammation requires consideration of both the type of immunity perturbed and disease activity. For this reason, both lesional and non-lesional samples from atopic dermatitis patients were considered against healthy controls. While multiple immune axes may play varying roles in atopic dermatitis, it is nevertheless an archetypal T_H2 related disease. To extend analysis to other immune axes, another common cutaneous skin disease, psoriasis, was included. Importantly, including psoriasis samples allowed for the validation of signatures that may arise from the atopic analysis that are more common to generic skin inflammation, and shared across diseases. For the consideration of psoriasis, the dataset GSE34248 was used and is composed of 14 paired psoriatic lesional and non-lesional samples (Bigler *et al.*, 2013).

Lesional signals from skin biopsies of atopic dermatitis and psoriasis were characterised by comparing lesional against non-lesional transcriptomes. Batch effects from merging the atopic dermatitis and psoriasis datasets were successfully removed using the COMBAT function within the *sva* package of Bioconductor (**Figure 3.3**). The MDS of the transcriptomic data representing these

samples demonstrates clear clustering according to disease and lesion status (**Figure 3.3**). Healthy and psoriatic non-lesional samples cluster very close, while the latter cluster distinctly from their lesional counterparts. In contrast, there was less distinction between lesional and non-lesional samples from atopic patients, reflecting the underlying pathology common to unaffected lesional skin that is often described as sub-clinical. In this curated multi-disease dataset of both lesional and non-lesional biopsies from both conditions and healthy controls, the differences between healthy and inflamed skin, as well as different inflammatory condition were recapitulated.

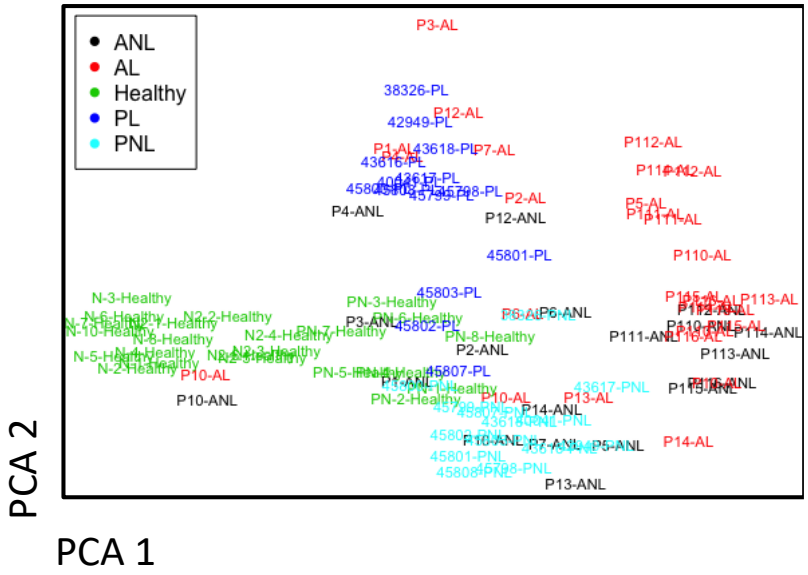


Figure 3.3. MDS plot of integrated AD and psoriasis datasets, GSE32924, GSE36842 and GSE34248. Three independent microarray datasets were obtained, following QC and normalisation were integrated and batch-corrected using the *sva* package function COMBAT. Healthy samples (*n*=23, green) were compared to AD lesions (*n*=27, red) and non-lesions (*n*=19, black), and psoriatic lesions (*n*=14, blue) and non-lesions (*n*=14, cyan).

In line with previous reports (Suárez-Fariñas *et al.*, 2011; Gittler *et al.*, 2012; Bigler *et al.*, 2013), unsupervised differential expression analysis using LIMMA identified 3058 genes. After applying DEG filtering, lesional atopic dermatitis samples were distinguished from healthy samples by 3289 probesets in the multi-disease dataset, representing 2237 unique genes. There was a steep drop in numbers of identified DEGs comparing atopic lesional skin to healthy, and then to non-lesional atopic skin. This lesional to non-lesional comparison found 22 probesets representing 18 unique genes. This suggests that atopic skin, regardless of active lesion sampling, is more similar across the whole tissue at the transcriptome level than non-lesional biopsies are to healthy skin. Indeed, non-

lesional skin of atopic dermatitis maintains specific features of lesional skin (Jensen *et al.*, 2004; Leung *et al.*, 2004), and has a clearly differentially expressed transcriptome from healthy skin.

While this DEG analysis is qualitatively similar to previous studies and demonstrates the reproducibility of such merging analyses as this, it offers limited use for addressing complex biological problems such as deconvolution of whole transcriptomic data. They can be used as a pre-filtered set of genes demonstrating most difference between important phenotypes for subsequent gene co-expression analysis to identify groups of genes that can be assumed to share similar biological processes.

To identify patterns of co-expressed genes, transcript-to-transcript clustering was performed on the 4620 differentially expressed probesets (representing 3058 unique genes) identified across all comparisons identified 50 clusters (GraphiaPro, Pearson r >0.7, MCL=3.1, >10 clustered probesets) (**Figure 3.4**). The number of unique genes per cluster ranged from 642 (cluster 1) to 3 (cluster 39). The REVIGO online tool was used to summarise the significant gene ontology terms for each cluster to unique mother terms only.

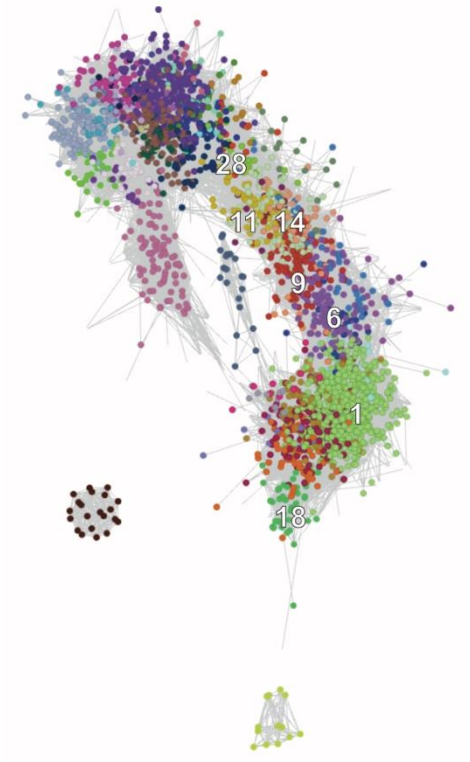
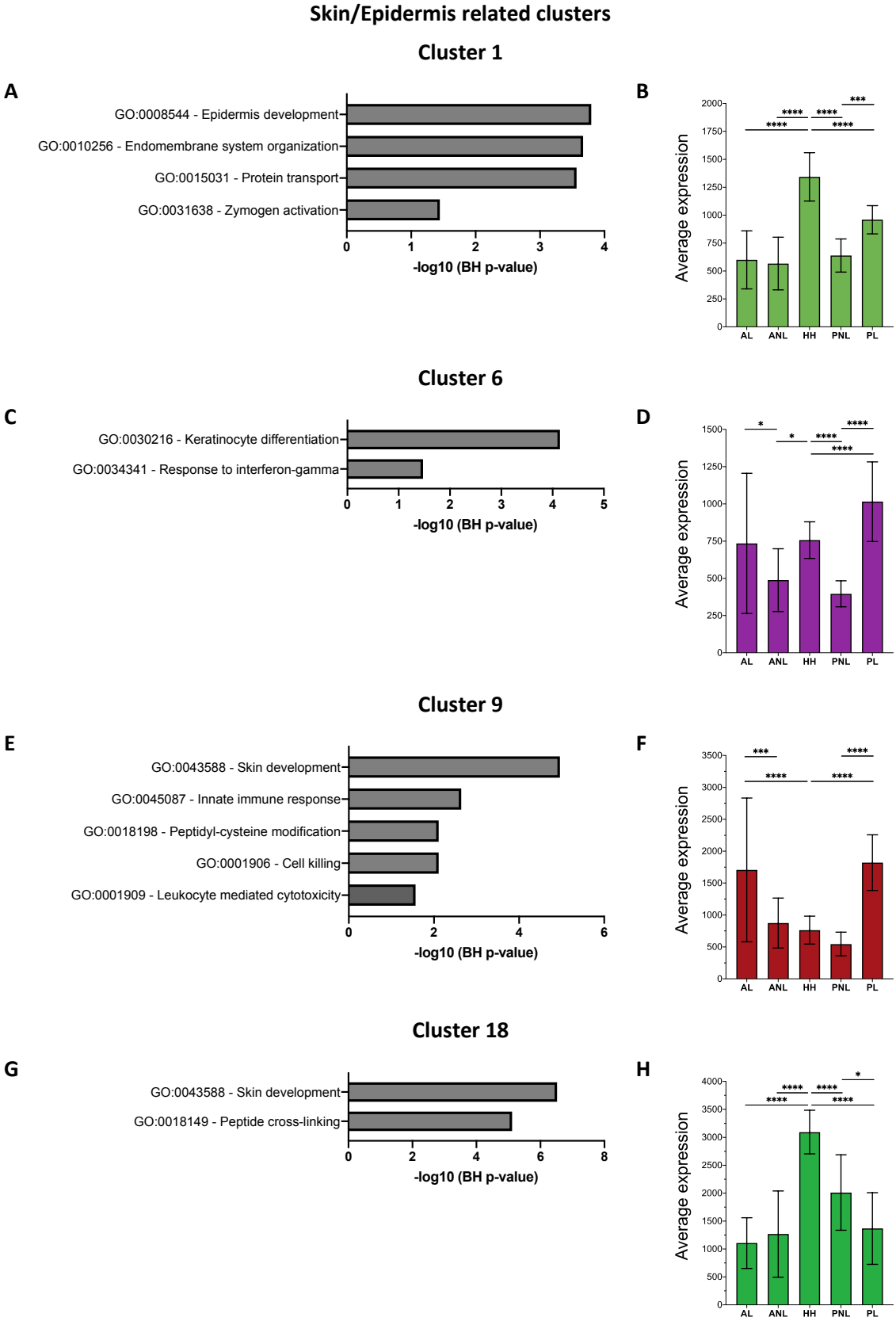


Figure 3.4. Unsupervised co-expression of whole skin from healthy, atopic and psoriasis skin. Transcript-to-transcript clustering of 4620 differentially expressed genes of AD lesions compared to AD non-lesions and healthy controls, and psoriatic lesions compared to psoriatic non-lesions and healthy controls. Initial co-expression filtering applied by a minimum threshold of 0.7 Pearson correlation with clusters identified using the Markov clustering algorithm (MCL) at an inflation value of 3.1. A total of 50 clusters identified (indicated by colours), of which seven cell-based clusters were identified by gene ontology analysis (numbered).

Clusters 1, 6, 9 and 18 were enriched in biological processes characteristic for the epidermis and/or keratinocytes (**Figure 3.5A,C,E,G**). Within the top five significant annotations, 'epidermis development' (GO: 0008544) was significant for all four clusters. Despite their initially similar biological function, these clusters revealed interesting expression patterns across the range of samples and tissue types. Cluster 1 was annotated for 'epidermis development' (GO: 0008544, BH p-value= $1.60E^{-4}$) (**Figure 3.5A**) and was highly associated with expression in healthy samples as indicated by genes such as *DMKN*, *DSG1*, and *OCLN*, which are critical for cell-cell barrier integrity (**Figure 3.5B**). 'Keratinocyte differentiation' processes (GO:0030216, BH p-value= $7.14E^{-5}$) described cluster 6 (**Figure 3.5C**), which was significantly associated with psoriatic lesional skin samples underscoring the hyper-proliferative phenotype of this tissue (**Figure 3.5D**). A 'skin development'

annotation (GO:0043588) was shared by both clusters 9 (BH p-value= $1.09E^{-5}$) and 18 (BH p-value= $3.02E^{-5}$) (**Figure 3.5E,G**). Strikingly, despite their similar annotation these clusters reflected differing expression in healthy versus inflamed samples. Cluster 9 included genes such as the inflammatory keratins (*KRT6A*, *KRT6B*) and S100 proteins (*S100A7*, *S100A8*, *S100A9*) and serpins (*SERPINB3*, *SERPINB4*), which reflects the association of this cluster with both atopic and psoriatic lesional samples (**Figure 3.5F**). In contrast, cluster 18 was significantly associated with healthy skin and included genes involved with functional epidermal barrier such as *FLG*, *FLG2*, *LOR* and the late-cornified envelope proteins (*LCE1B*, *LCE1E*, *LCE2B*) (**Figure 3.5H**).

Annotation revealed three clusters (11, 14 and 28) encoding immune-related processes such as lymphocyte activation, interferon and cytokine signalling. The relative expression pattern across these clusters showed similar changes in lesions, regardless of disease, and the least expression in healthy skin. Cluster 11 was underscored by processes involved in 'regulation of lymphocyte activation' (GO: 0051249, BH p-value= $1.51E^{-8}$) (**Figure 3.5**I) and included genes such as *CD2*, *CD28*, *CD3D*, *CTLA4*, *CXCR4*, *GZMA*, *GZMB*, *ICOS*, *IL12RB1*, *IL12RB2*, and *IL7R*, which are indicative of T cell biology and explain the high expression of this cluster in lesional samples from both diseases due to heightened immune activity (**Figure 3.5J**). However, the data did not deconvolute further to distinguish T_H-subtypes in order to effectively describe the T_H-specific polarisation of the samples. This is despite the analysis considering classic T_H-polarised diseases involving all of T_H1, T_H2, T_H17 and T_H22 inflammation. 'Type I interferon signalling pathway' (GO: 0060337, BH p-value= $8.30E^{-24}$) explained the biological annotation of cluster 14 (**Figure 3.5K**), which again showed heightened expression in lesional samples (**Figure 3.5L**). Finally, cluster 28 reflected processes involved in 'defence response' (GO: 006952, BH p-value= $7.^{68E-6}$) (**Figure 3.5M**). Interestingly, non-lesional atopic skin showed a prominent expression in this cluster, corroborating the previously reported subclinical immune alteration in these samples as compared to healthy skin (**Figure 3.5N**) (Suárez-Fariñas *et al.*, 2011; Tang, Bieber and Williams, 2014; Brunner *et al.*, 2017).



[*figure continues onto next page*]

[*figure continued from previous page*]

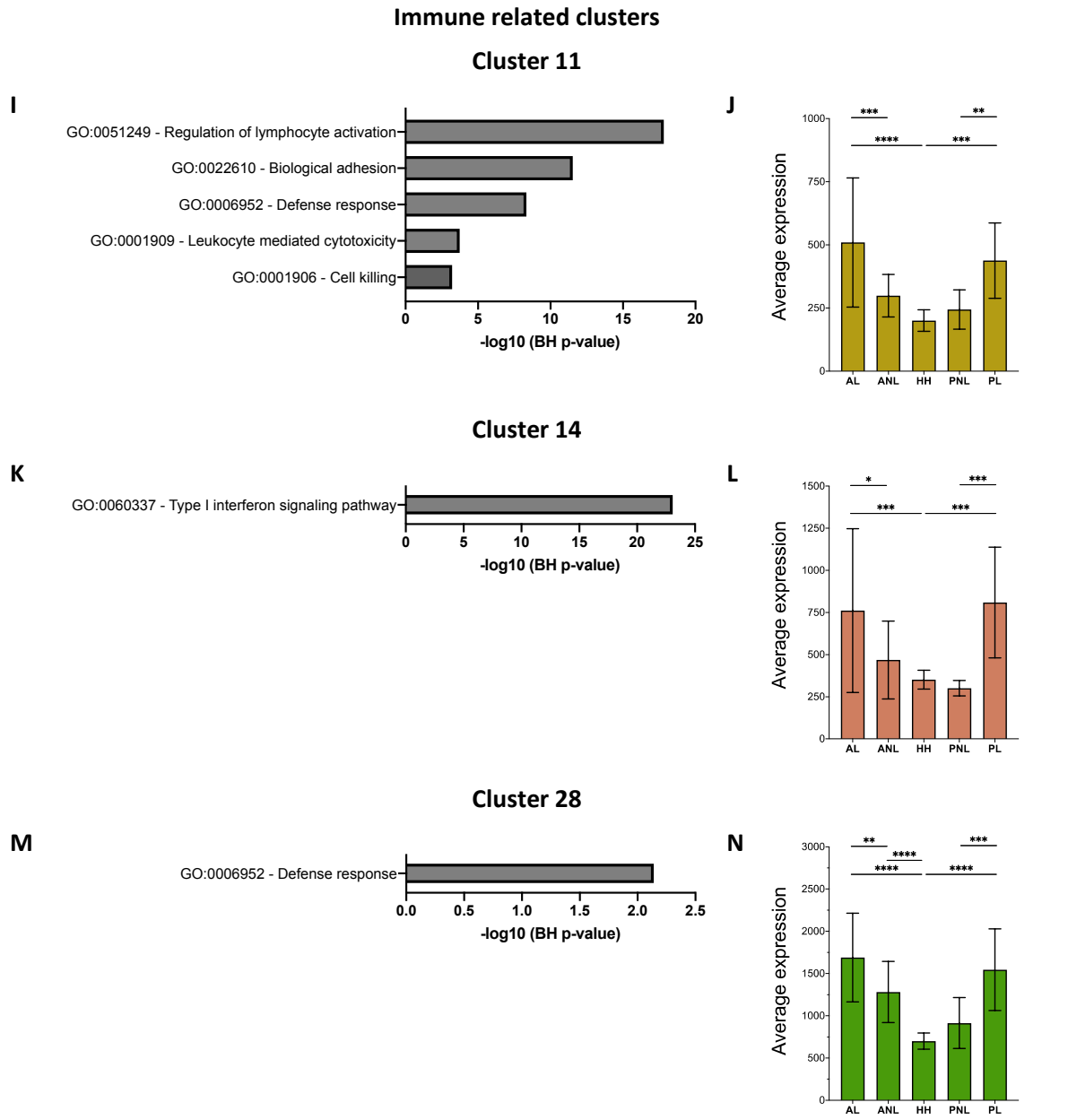


Figure 3.5. Co-expression clusters identified within the correlation network of disease defining genes. The top 50 clusters identified by correlation-MCL network analysis were annotated for biological meaning. **[Left panel: A, C, E, G, I, K, M]** Biological process ontology analysis performed by ToppGene and summarised by REVIGO in cell-based clusters 1, 6, 9, 11, 14, 18 and 28. B.H.-adjusted p-values represented in the negative log10 space. **[Right panel: B, D, F, H, J, L, N]** The average expression of all the genes in each cluster shown per phenotype: AD lesional, AL; AD non-lesional, ANL; Healthy controls, HH; psoriasis non-lesions, PNL; and psoriasis lesions, PL. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Normality testing performed using the D'Agostino-Pearson method indicated the use of nonparametric Kruskal-Wallis ANOVA with multiple comparison performed. Error bars show median ±95% confidence interval.

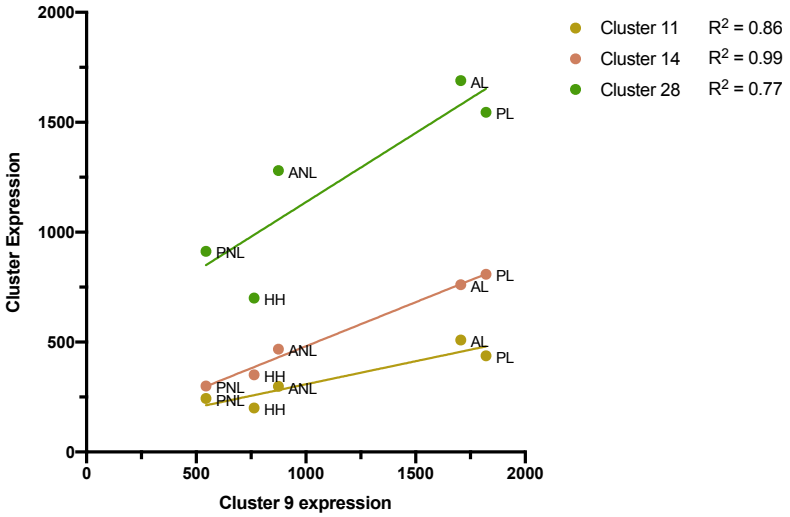


Figure 3.6. Correlation between cluster 9 and cluster 11, 14 and 28. Average expression of the genes in clusters 11, 14 and 28 representing immune/inflammatory activation collapsed by tissue type show strong positive correlation with cluster 9 representing perturbed skin development.

Cluster to cluster comparisons of gene expression by tissue showed of the immune-related clusters 11, 14 and 28 against the lesion-associated keratinocyte cluster (cluster 9) showed high correlation (Pearson *r* >0.85 (**Figure 3.6**)). It is notable that genes in cluster 9 include *KRT6A*/*B*, *KRT16*, and the *S100A7/8/9* and *SERPIN* encoding proteins which are known to be involved in epidermal perturbation from inflammation and hyper-proliferative barrier breach (Lessard *et al.*, 2013; Rorke *et al.*, 2015; Ungar *et al.*, 2017), indicating that the cross-talk between immune inflammation and keratinocyte function is important for lesion pathogenesis.

3.2.4 Profiling cutaneous cell heterogeneity using xCell

The previous section has demonstrated the ability to deconvolute skin expression data into pre-defined cellular populations. To begin computationally enumerating cellular heterogeneity from bulk skin, the tool xCell was used (Aran, Hu and Butte, 2017). xCell is a hybrid gene set enrichment and deconvolution tool constructed from a wide panel of 64 cell types from lymphoid, myeloid, stem, stromal and epithelial populations. The output of this tool is an independent enrichment score per cell type, which can be pre-filtered to tailor to a tissue of interest. The data obtained from this tool are not to be interpreted as a relative enumeration.

The panel of 64 cell types was limited to eight profiles for the deconvolution of bulk skin samples from GSE32924 and GSE34248. The cell types included for skin deconvolution are $CD4^+$ T cells, $CD8^+$ T cells, T_H1 cells, T_H2 cells, dendritic cells, sebocytes, fibroblasts and keratinocytes (**Figure 3.7**).

Applying xCell enrichment analysis to bulk skin data revealed that this tool did not perform as expected. Enrichment of $CD4^+$ T cells across the tissue types investigated shows an increase in this population in both AD and psoriatic lesions compared to their non-lesional counterparts and healthy controls (**Figure 3.7A**). This is expected during flares of cutaneous inflammation. However, when this is investigated further for T_H polarisation arising from this increased T cell profile in the skin, as indicated by the T_H1 and T_H2 scores, the results are in opposition to understanding of inflammation in atopic dermatitis and psoriasis. T_H1 enrichment was consistently similar across all disease samples, and not raised in psoriatic lesions as might be expected (**Figure 3.7D**). While the T_H2 profile of psoriatic lesions was raised compared to non-lesions and comparable to AD lesions, these were not different to healthy controls (**Figure 3.7E**).

Additionally, the keratinocyte results did show an elevated enrichment in psoriatic lesions compared to non-lesions, but this was not raised compared to healthy controls (**Figure 3.7F**). These results are difficult to interpret across cell types because enrichment score is not inferred as a cellular proportion. This is most obvious when comparing between keratinocytes and T_H1 cells, for example. The enrichment scale and results of these cell types are represented as similar, yet the samples from which they derive cannot be assumed to have equal fractions of keratinocytes and T_H1 cells within the biopsy.

The original description of xCell provided insight into the derivation of these signatures (Aran, Hu and Butte, 2017). For the 64 cell types includes in the tool, each was defined by an average of 42 genes from 28 replicates across at least two or three independent datasets. T_H1 and T_H2 profiles were only derived from five and six replicates, respectively each from a single dataset. This gave 36 genes defining a T_H1 signature and 22 genes define T_H2. Keratinocytes, in contrast, were defined by 121 genes from 31 replicates across three datasets.

Gene ontology analysis of these genes revealed the strength of these signatures in defining their supposed cell type (**Figure 3.8**). Ontology analysis for both T_H1 and T_H2 defining signatures did not reveal processes relevant to their cellular identity, while the annotations that were found were not highly significant (**Figure 3.8A, B**). In contrast, the keratinocyte signature proved much stronger and was well annotated for keratinocyte/epidermal biology (**Figure 3.8C**).

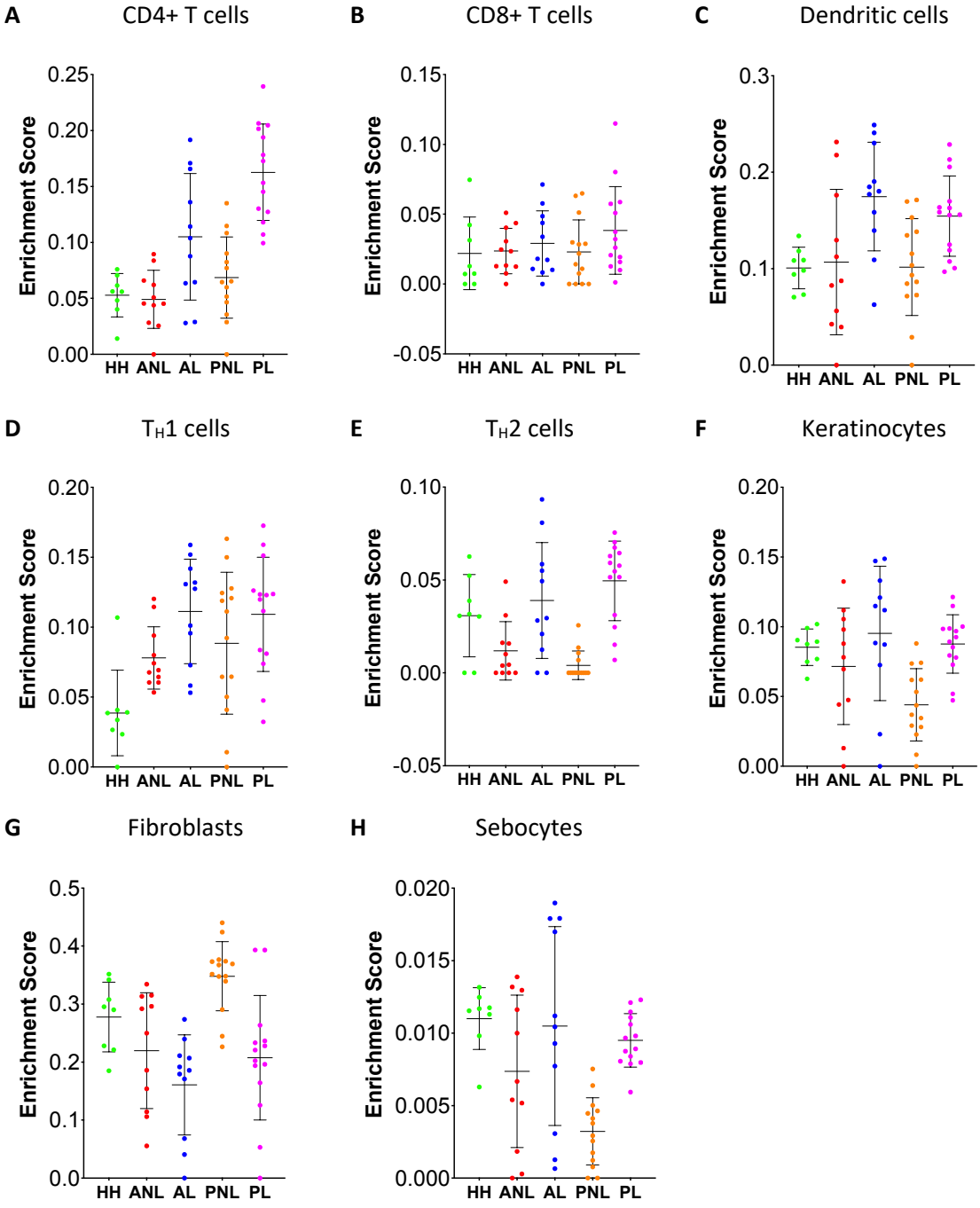


Figure 3.7. xCell enrichment deconvolution of bulk skin. Bulk skin samples were applied to xCell for deconvolution and enrichment scoring of eight cell types **[A]** CD4+ T cells, **[B]** CD8+ T cells, **[C]** Dendritic cells, **[D]** T_H1 cells, **[E]** T_H2 cells, **[F]** keratinocytes, **[G]** fibroblasts and **[H]** sebocytes. Error bars show mean ±SD.

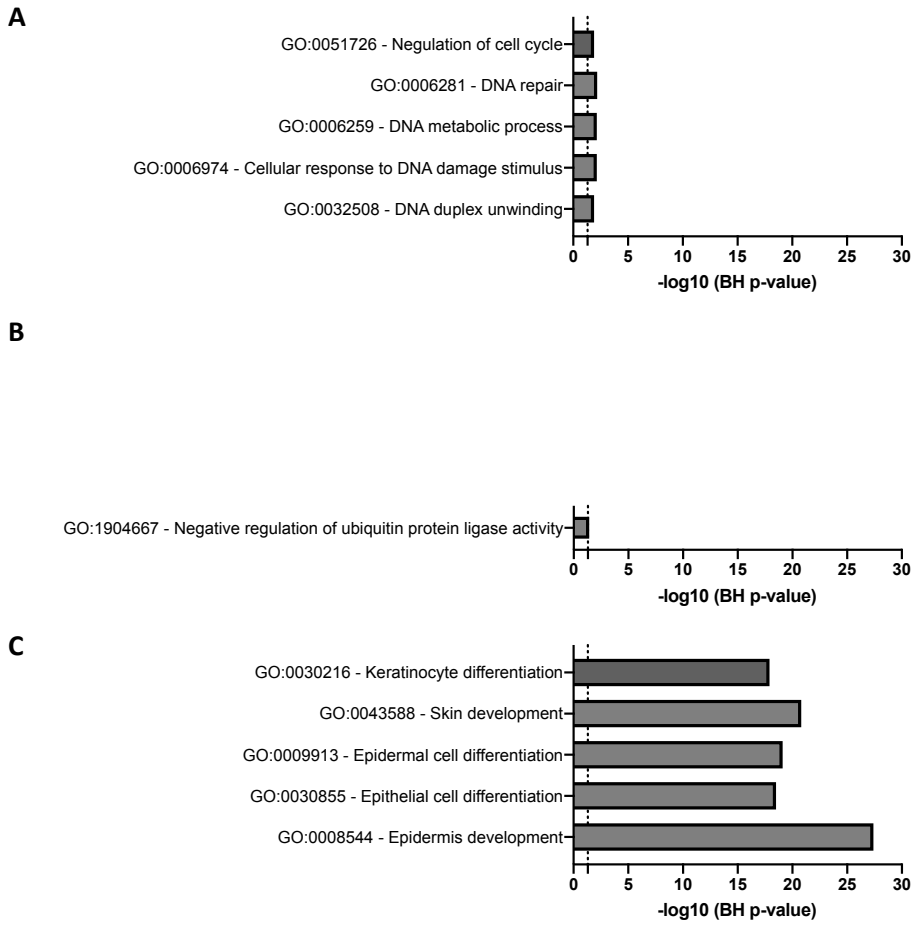


Figure 3.8. Biological process gene ontology analysis for xCell profiles. Genes defining three selected cell profiles from xCell were interrogated for biological process ontology analysis performed by ToppGene, showing the top five annotations for **[A]** T_H1, **[B]** T_H2 and **[C]** keratinocytes. B.H.-adjusted p-values represented in the negative log10 space. Dotted line indicates significance level of 0.05.

xCell has demonstrated the concept in using a gene-signature-based method for inferring cell types from complex bulk tissue such as skin. A particular strength of this tool is the fidelity of the keratinocyte signature, but simple enumeration of the keratinocyte compartment does not allow for the investigation of epidermal sub-populations or phenotypes. Furthermore, particularly for immune cell types, the included cell type signatures were not well-defined. This reflected the difficulty in separating T_H2 types from the total $CD4^+$ T cell compartment experienced in the unsupervised co-expression analysis. Although attempted, the T_H1 and T_H2 signatures in this tool are poorly defined but the *in-silico* immunophenotyping of T_H biology arising from cutaneous inflammation is critical for expanding computational analysis of skin disease. Therefore, these two elements should be brought together by *in-silico* investigation to resolve the effect of T_H biology on the keratinocyte transcriptome.

3.2.5 Machine learning can accurately identify dermis and epidermis using fibroblast versus keratinocyte signatures

Unsupervised co-expression analysis revealed the potential for bulk transcriptomic data to be deconvoluted for cellular meaning. This approach was expanded to consider the deconvolution of the diverse cell populations that compose the whole skin in a supervised fashion. This approach enables resolving predefined populations of interest from the complex gene expression profiles that make up whole tissue biopsies for relative enumeration. In this approach the linearity of the sum of the individual cellular profiles can be inferred with a mixture sample, and as such these can be imputed from whole tissue samples such as the skin (Shen-Orr *et al.*, 2010; Zhong and Liu, 2012). This approach to reference-based bulk deconvolution is the basis for CIBERSORT (Newman *et al.*, 2015). This tool uses a default deconvolution panel called LM22 in the context of a solid tumour microenvironment and is not optimised for skin deconvolution. Therefore, it must be re-optimised for analysis of skin data.

The anatomy of skin provides a potential challenge for *in silico* deconvolution. The dermis and epidermis are majorly composed of different cellular populations; fibroblasts and keratinocytes, respectively, with no cross-over between these two cutaneous layers. Simultaneous bulk deconvolution of such heterogenous and distinct layers requires careful optimisation to prevent overfitting of reference profiles. Therefore, to ensure that fibroblast references could correctly identify dermis, and keratinocyte signatures to identify epidermis, a two-profile panel was constructed for this optimisation.

From GEO microarray data of microdissected skin was obtained (GSE120721 (Esaki *et al.*, 2015)). This dataset is composed of samples of epidermis and dermis obtained by laser capture microdissection (LCM) from whole skin biopsies. Using this dataset, the ability of CIBERSORT to correctly identify epidermal and dermal sections composed of keratinocytes or fibroblasts, respectively was tested (**Figure 3.9A**). Also tested were the healthy and atopic sections, of both lesional and non-lesional tissues, against these reference fibroblast and keratinocyte signatures only to observe the effect, if any, of cutaneous inflammation in the correct resolution of samples (**Figure 3.9B**).

Dermal and epidermal compartments were correctly deconvoluted as being composed of their major constituent cells. Healthy dermal sections were the most correctly deconvoluted samples, found to be 95.6% fibroblast ($\pm$1.2%, SD). The atopic dermis samples display slightly lower composition of fibroblasts, but the overlapping error bars between the three tissues tests shows this to be marginal and not driven by biology or inflammation status (**Figure 3.9B**). Newman *et al.*

(2015) showed that fraction enumeration validation on purified populations by CIBERSORT results in approximately 2-15% error when inferred on B and T cell samples.

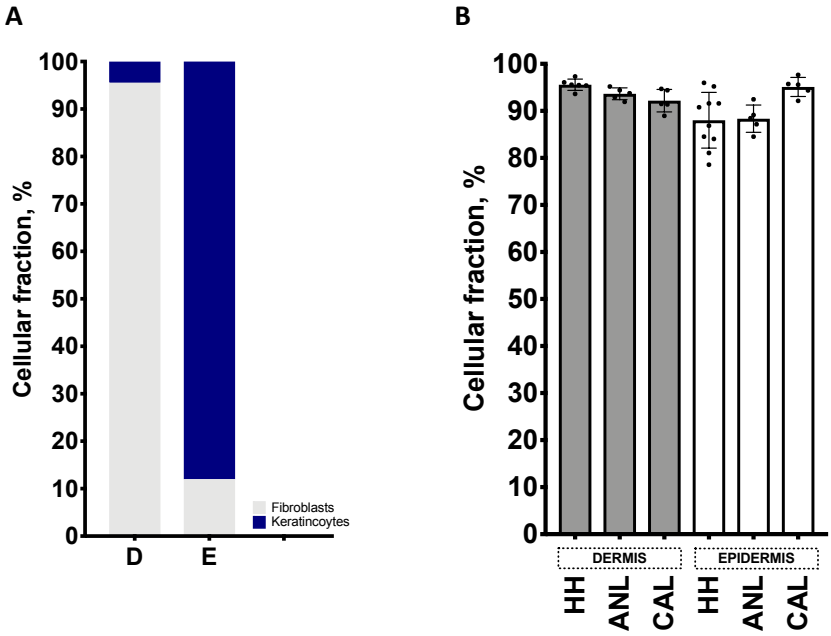


Figure 3.9. CIBERSORT-derived deconvolution of dermis and epidermis. [A] Healthy dermis ("D") and epidermis ("E") obtained from whole tissue by laser microdissection were deconvoluted in CIBERSORT against reference signatures of fibroblasts and keratinocytes to assess correct identification of these cutaneous layers. **[B]** Comparing the cellular fraction for healthy and AD skins as identified as the correct reference signature: dermis (grey), fibroblasts; epidermis (white), keratinocytes. Error bars show mean±SD.

Interestingly, there was much variation in keratinocyte composition healthy epidermal sections, ranging from 78.6 to 96.0%. This was considered acceptable and within the normal ranges of keratinocyte proportions of the epidermis. Conversely to the dermal samples, as atopic phenotype increased, the resolved correct keratinocyte fraction of epidermal sections increased. Again, the error bars between tissue types overlapped indicating this to not be driven by biology.

These results show the ability of this approach to correctly resolve the defined skin layers, and there is minimal overfit between keratinocyte and fibroblast reference signatures. Much of the minimal overfit may be explained by the other cell types within the dermis and epidermis that are not accounted for in the fibroblast-keratinocyte-only panel. In addition, this is not detrimentally affected or altered by inflammatory processes that otherwise dramatically alters the transcriptomic phenotype of the skin.

The analysis of the dermal and epidermal compartments of the laser microdissected biopsies was expanded to consider the morphological layers of keratinocytes within the epidermis. Data from a single cell dataset including healthy epidermis was obtained for this purpose (Cheng *et al.*, 2018). Basal, mitotic and spinous clusters comprised the major populations of the keratinocytes. Using a pseudobulk approach derived of these three clusters, the profiles were included with the fibroblast signature from before and used to deconvolute the LCM dermal and epidermal samples.

In the context of healthy skin, the fibroblasts were again found to correctly dominate the dermis compartment (**Figure 3.10**). The added complexity of the keratinocyte profiles slightly reduced the resolution of correct cell type identification in healthy epidermis to around 70%. However, using these additional reference profiles provides insight into the composition of keratinocyte sub-populations within bulk transcriptomic data. The ability to deconvolute from bulk data the proliferating and differentiating populations of keratinocytes is useful in addressing the over-arching hypothesis of this work; that these different layers respond differently to pathogens as they do cytokines.

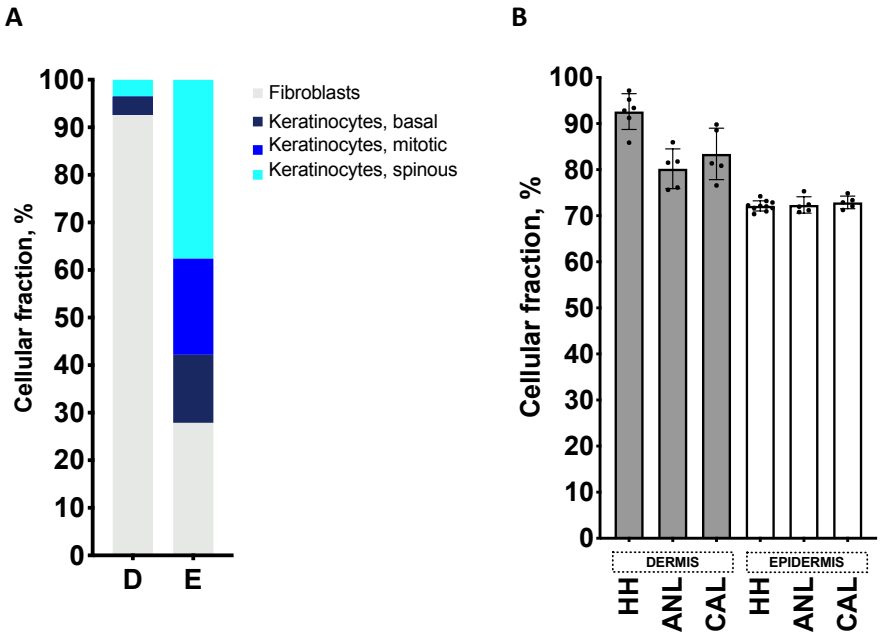


Figure 3.10. CIBERSORT-derived deconvolution of dermis and stratified epidermis. [A] Healthy dermis ("D") and epidermis ("E") obtained from whole tissue by laser microdissection were deconvoluted in CIBERSORT against reference signatures of fibroblasts and layer-defining keratinocytes (basal, mitotic and spinous populations) to assess correct identification of these cutaneous layers. **[B]** Comparing the cellular fraction for healthy and AD skins as identified as the correct reference signature: dermis (grey), fibroblasts; epidermis (white), keratinocytes. Error bars show mean±SD.

3.2.6 Whole transcriptome-based generation of reference profile increases accurate resolution of test samples

Unsupervised deconvolution of bulk skin data has revealed distinct immune cell and keratinocyte-based populations and the complexity of cellular programming during cutaneous inflammation. Reference-based deconvolution using xCell demonstrated the potential of pre-defined cellular signatures applied to the bulk transcriptome. Within CIBERSORT, estimation of the abundance of the reference signatures can be accomplished using the in-tool barcode genes protocol or providing a gene expression profile of any number of marker genes. The barcode approach is a method to define set of signature genes across the panel of reference cells provided that are distinct to each cell type. Barcode genes are defined within CIBERSORT across the whole panel and are assigned to each signature based on the cell type the gene is mostly highly expressed. This minimises the number of genes for deconvolution while attempting to retain the maximum number of genes for efficient identification of individual signatures.

Before attempting deconvolution of bulk skin using CIBERSORT, it was necessary to train the algorithm to resolve different transcriptomic signatures of keratinocyte cytokine responses. To apply this algorithm to understand skin inflammation and host-microbe interactions, the default 22 haematopoietic signatures needed to be replaced with a tailored set appropriate for cutaneous deconvolution. This required testing the ability of CIBERSORT to correctly identify gene expression profiles from keratinocytes when trained on reference profiles from these same conditions. A dataset of keratinocytes co-cultured with cytokines; IFNα, IL-13, IL-17A, IL-4, TNFα, and unstimulated control, was used to optimise CIBERSORT on cutaneous cell types (GSE36287). This dataset is composed of samples in triplicate, which was split into duplicate for providing the reference signatures and the remaining replicate was used for deconvolution as the test sample.

Using the barcode approach, CIBERSORT identified 342 marker genes of the six keratinocyte subsets provided as potential reference signatures (**Figure 3.11**)(**Table 7**) A whole transcriptome approach, of 19851 genes, to minimise inter-signature crossover and noise, and maximise the ability of the algorithm to distinguish very closely related cell types against more distinct populations was performed in parallel for comparison to the barcode approach.

Figure 3.11. Heatmap of CIBERSORT-derived cytokine stimulation barcode genes. Genes that were highly expressed in a given cell type were labelled as barcode genes by CIBERSORT. The number of enriched genes per reference sample: unstimulated control, 83; IFNα, 27; IL-13, 0; IL-17A, 36; IL-4,27; and TNFα, 169. Full gene list given in Table 7.

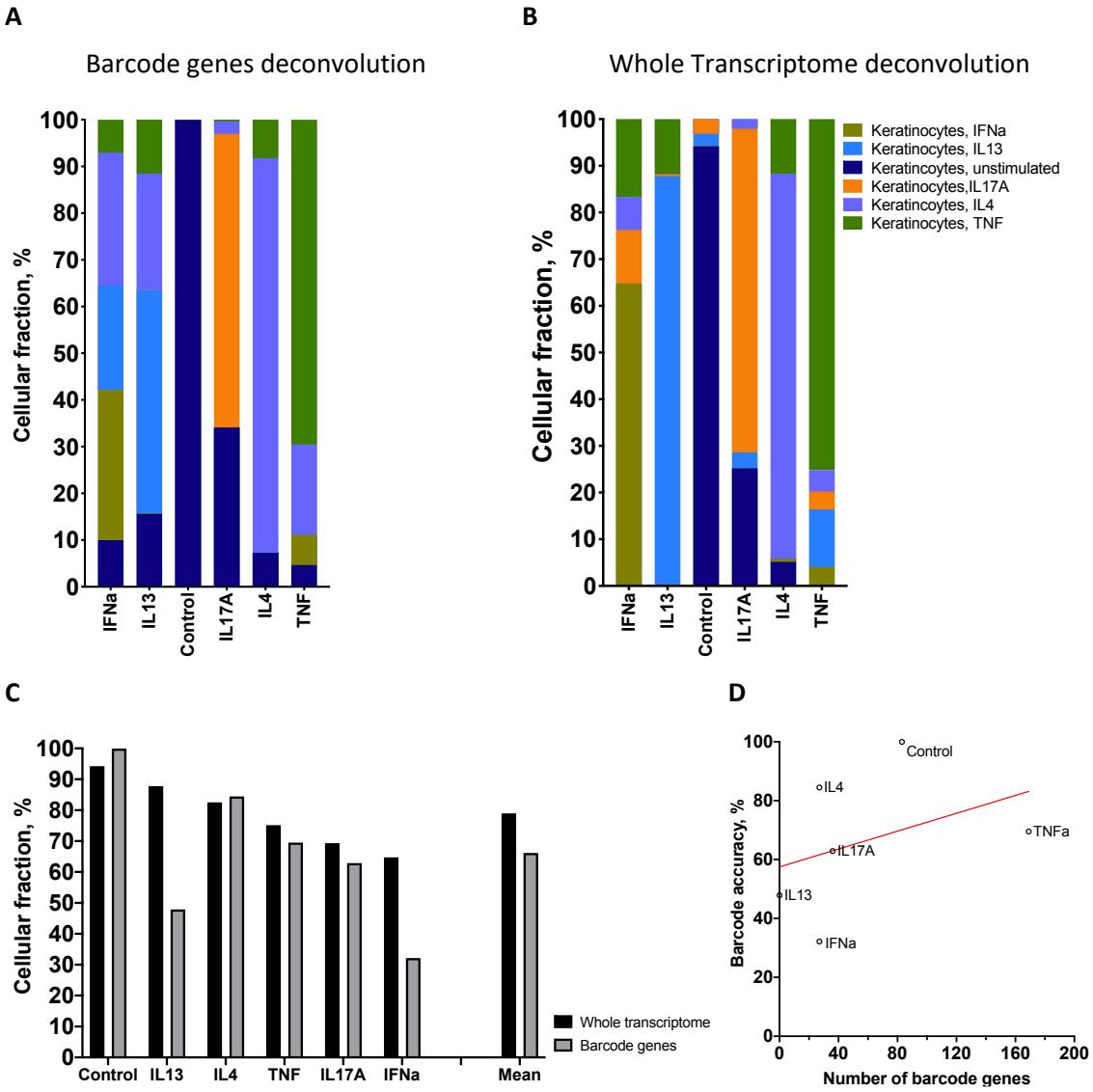


Figure 3.12. CIBERSORT-derived deconvolution of monolayer cultured keratinocytes with or without single cytokine stimulation. CIBERSORT was trained on two samples per condition from the dataset GSE36287 using either **[A]** barcode genes or **[B]** the whole transcriptome. **[C]** The remaining sample per condition was used to test the accuracy of deconvolution for correct cytokine condition identification and optimal reference gene list (whole transcriptome, black; barcode list, grey). **[D]** Number of barcode genes per condition and fraction per centage of the correctly labelled sample.

To compare the faithfulness of these gene lists in enabling the algorithm to properly identify keratinocyte samples, these were tested using a leave-one-out training approach where CIBERSORT reference profiles were defined on the mean expression across *n* minus one replicates of a sample. The remaining sample was therefore 'unseen' by the algorithm and tested for deconvolution accuracy.

Deconvolution of the six keratinocyte conditions revealed the ability of CIBERSORT to assign the correct signature is highly accurate (**Figure 3.12**). The average fraction of the correct signature for

each sample using the whole transcriptome was 79% (range: 64.8-94.2%) (**Figure 3.12C**). This was reduced to 66% using the barcode approach. In general, the barcode gene approach was less accurate at identifying the correct stimulation condition. The accuracy of the algorithm using only the barcode genes was independent of the number of signature genes for each condition (**Figure 3.12D**, slope= 0.1849, R^2 = 0.1439, p=0.4582).

Newman *et al.* (2015) demonstrated that between closely related cell types accurate identification of validation cell types could be as low as 50-75%. Using these data, in this work, an arbitrary 75% cuff-off marked a correctly identified test sample from its cognate reference profile. Using the whole transcriptome approach this was achieved in the control, IL-13 and IL-4 samples. The barcode approach was less successful, only correctly identifying control and IL-4 samples. IFNα and IL-13 samples were found to be composed of less than 50% of their correct signature using the barcode approach. Where the stimulated keratinocytes sample was identified as coming from less than 75% of its cognate reference profile, the leftover signature was generally composed of a mix of the remaining signatures (whole transcriptome example of IFNα and TNFα). The average accuracy of deconvolution passed the 75% threshold only for whole transcriptome input.

In order to estimate the effect of variation in the genes used by CIBERSORT, a set of graded gene matrices based on varying the fold change statistical filter for differentially expressed genes between the conditions was produced. Each keratinocyte cytokine-stimulated sample was deconvoluted against a selection of three reference samples. The three reference samples were derived from the merging of signatures based on the dendrogram structure of the six conditions considered. The groupings consisted of control and IL-17; IFNα, IL-4 and IL-13; and singularly TNFα. Grouping in this way is appropriate as the fidelity of the specific signatures is not being assessed, but the effect of varying the number of genes per signature.

To reduce the complexity and noise of this analysis, the number of test samples was reduced to three. Therefore, of the tested reference profiles there was one correct and two mismatched references. This was repeated at a varying number of genes: whole transcriptome (19850 genes), two-way DEGs at a logFC filter of ±1.0 (4936 genes), 1.5 (2749 genes), 2.0 (1520 genes), and 2.5 (854 genes) defined against the other two reference groups, and the CIBERSORT-defined 342 barcode genes. The ability across the range of genes in the reference profiles to correctly identify the sample deconvoluted, which we determined to be >70% in all cases (**Figure 3.13**).

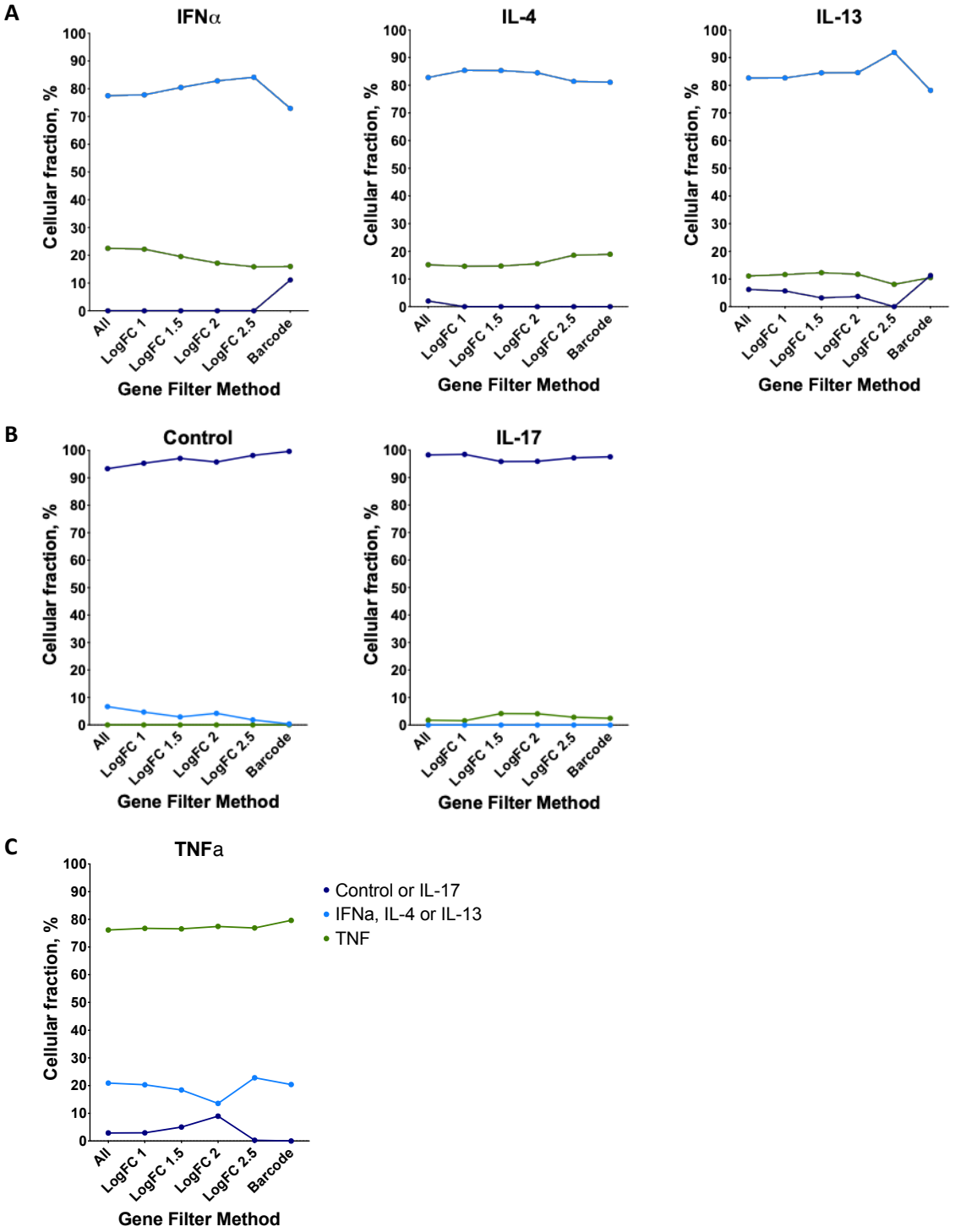


Figure 3.13. CIBERSORT-derived deconvolution of monolayer cultured keratinocytes with or without single cytokine stimulation profiled by panels of reference signatures with varying numbers of genes. The stability of the CIBERSORT algorithm to correctly identify samples across a range of input genes in the reference profiles was tested. For simplicity, samples were grouped into three based on total transcriptome clustering (not shown). The number of genes for each group was: All, 19850 (whole transcriptome); "LogFC1", 4936; "LogFC1.5", 2749 genes; "LogFC2", 1520 genes; "LogFC2.5", 854 genes, and by CIBERSORT-defined 'barcode genes', 342 genes.

The stability of the algorithm, therefore, to identify the correct composition of the test samples was not altered by the number of genes in the reference matrix. However, a whole transcriptome approach over selective barcode genes proved best at deconvoluting both the difference between major cutaneous cell populations and the more subtle difference between the keratinocyte subpopulations of interest.

3.2.7 Profiling of the epidermal compartment within bulk skin for cytokine immunophenotypes

In line with validation work performed by Newman *et al.* (2015) on purified leukocyte populations, signatures of fibroblasts and keratinocytes are almost uniquely deconvoluted from their native skin compartment. Additionally, the ability of CIBERSORT to further sub-divide the epidermis fraction from laser dissected whole skin into basal and suprabasal epidermal layers representing the proliferating and differentiating compartments is demonstrated. Finally, the power of CIBERSORT to properly deconvolute closely related transcriptomic signatures from cytokine-stimulated keratinocytes in an experiment involving only one cell type was explored.

A major aim of this project and the *in-silico* approach was to deconvolute, from bulk transcriptomic profiles, multiple fractions derived from the same cell populations, for example, altered keratinocyte expression patterns from differentiation status or inflammation status, or potentially both. The results presented in this chapter demonstrate that this aim is achievable and robust in addressing the hypothesis that the various structural populations in the skin have altered abilities to pathogen sensing in the context of inflammation.

A set of reference signatures curated for this work, composed of a total 19 profiles including 11 keratinocyte profiles, was then validated for its ability to not overfit to dermal and epidermal sections of the microdissection dataset. This provides this work's first full cutaneous reference panel akin to LM22 (Newman *et al.*, 2015) or DerM22 (Félix Garza *et al.*, 2019).

Interestingly, a universal overrepresentation of steady-state Langerhans cells was found (**Figure 3.14A**). This signature was markedly larger in the deconvoluted epidermis and whole skin than the expected 1% frequency *in vivo*. This signature was also present in dermal skin despite not being observed biologically. As there proved an absence of inferred fraction of dermal dendritic cells, it is possible that the entire populations of cutaneous dendritic and other immune cells have been deconvoluted as one fraction, here given as steady-state Langerhans cells. This highlights a potential limitation of the CIBERSORT approach to accurately deconvolute this important immune subset of cutaneous populations.

Reflecting the single keratinocyte phenotype results presented above, the known major cell component of either dermis and epidermis; fibroblasts or keratinocytes, were found to compose the majority the expected deconstructed skin sections (**Figure 3.14B**). FACS analysis and unbiased single-cell expression studies have showed fibroblasts to constitute 60-65% of all dermal cells and approximately one-third of the whole skin (Philippeos *et al.*, 2018; Tabib *et al.*, 2018; Korosec *et al.*, 2019). These deconvolution results of healthy dermis and whole skin reflect these expected proportions.

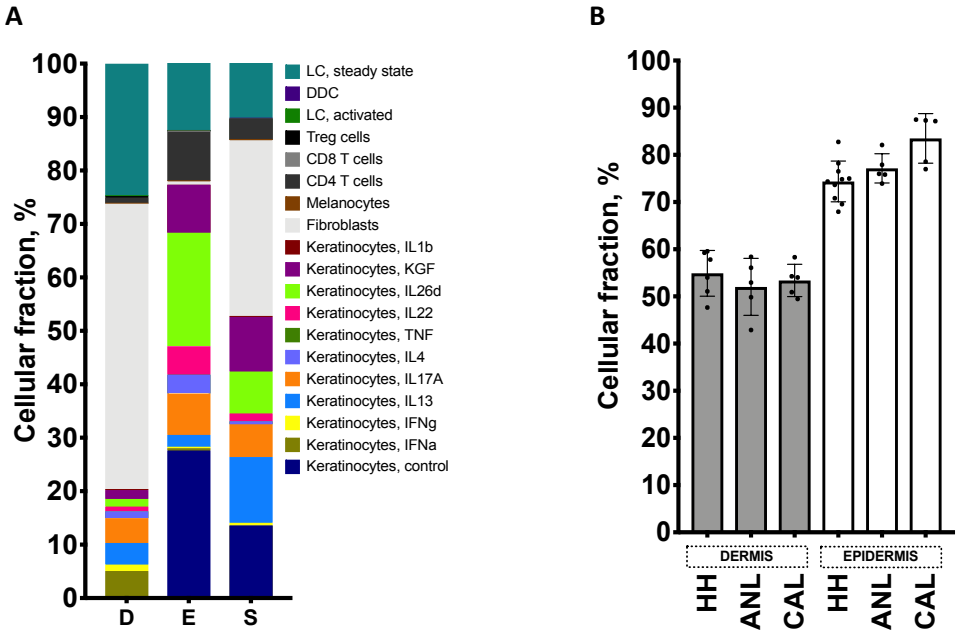


Figure 3.14. CIBERSORT-derived deconvolution of dermis and epidermal immunophenotypes. [A] Healthy dermis ("D"), epidermis ("E") and whole skin ("S") obtained from whole tissue by laser microdissection were deconvoluted in CIBERSORT against reference signatures of fibroblasts and layer-defining keratinocytes (basal, mitotic and spinous populations) to assess correct identification of these cutaneous layers. **[B]** Comparing the cellular fraction for healthy and AD skins as identified as the correct reference signature: dermis (grey), fibroblasts; epidermis (white), keratinocytes. Error bars show mean±SD.

3.3 Discussion

3.3.1 *In silico* approaches for cutaneous deconvolution

This chapter establishes the use of deconvolution of bulk skin data to find cell-related signatures and processes. The initial unsupervised analysis revealed the core transcriptomic signatures of atopic dermatitis and psoriasis derive not only from the immune subset, but these are mimicked by the keratinocyte compartment, indicating the extensive crosstalk between cutaneous immune and structural populations. Despite our understanding of the role of inflammatory cells in the pathogenesis of both atopic dermatitis and psoriasis, little attention has been paid to characterising the keratinocyte response to this local context in detail.

It is not currently fully understood how the interaction between the immune and keratinocyte populations affects the immunophenotype of the latter. Thus, to characterise the keratinocyte immunophenotype *in-situ* it is necessary to resolve cellular signatures from the bulk cutaneous transcriptome. Previous investigations have relied on statistical bioinformatic approaches to identify lists of genes differentially expressed by active disease followed by functional annotation by gene-set enrichment analysis to infer cellular processes perturbed by localised inflammation (Suárez-Fariñas *et al.*, 2011; Gittler *et al.*, 2012; Bigler *et al.*, 2013; Guttman-Yassky *et al.*, 2019). However, such a global view of the skin transcriptome, especially during inflammation, prevents the characterisation of the various cellular populations within the tissue and the phenotyping of potentially important sub-populations.

The hypothesis throughout this chapter was that these gene expression profiles could be deconvoluted using a machine learning approach based on purified cell populations. Resolving these fractions across health and disease would enable tracking of these specific populations across different inflammatory states and treatment responses. This approach was first attempted by use of xCell, an approach somewhere between gene-set enrichment and full deconvolution. The results provided from xCell confirmed the strength of the signal arising from the keratinocyte compartment of the bulk transcriptome. However, the resolution of the immune signatures proved to be weak and not accurately defined from the understood biology of the samples. The T_H1 and T_H2 signatures defined within the xCell tool were derived from a small number of marker genes and ontology analysis of these did not correspond to any specific biological process related to these cell types.

While the reference T_H-profiles were weak from the outset, this may also be a problem of the sample provided for deconvolution. Optimisation of xCell was done on tissues derived from tumour microenvironments, which are richer for infiltrating immune cells than the skin. There would be a

bias in the optimised panel in inferring cell types on which the tool is extensively validated. This is further compounded by the fact that tumour cells often display non-conventional expression which may alter the reference signature away from a standard applicable to all tissue types to being only reliable to the context of a tumour.

The outcome of immune and inflammatory processes in the skin is characterised by its effect in the epidermal microenvironment. Previous approaches to deconvolute $CD4^+$-specific signatures describing T_H1, T_H2, or T_H17 inflammation have demonstrated the difficulty in elucidating these (Inkeles *et al.*, 2015; Shih *et al.*, 2016). In line with the co-expression analysis described in Section 3.2.3, the skin deconvolution signature, '*SkinSig*,' developed by Shih *et al.* (2016) also proved unable to describe T cell signatures specific to $CD4^+$ axes. Instead, a generic T cell signature was found to be upregulated in infectious but not hyper-proliferative/neoplastic or allergic skin conditions. Additionally, the molecular signature from T cell immunopathogenesis mechanisms in both psoriasis and atopic dermatitis was reported as notably absent compared to other T-cell dominated skin conditions such as mycosis fungoides and Stevens Johnson syndrome (Inkeles *et al.*, 2015).

The keratinocyte signature is well defined in xCell and the unsupervised co-expression analysis showed a mirroring of the inflammatory context of the epidermis in this cell type. This puts the resolution of keratinocyte abnormalities at the centre of analysis where they also sit at the centre of disease pathogenesis by focus of their epidermal structural role, anti-microbial activity and microbiome homeostasis. An initial reference-based machine learning approach to deconvolution of the bulk skin transcriptome would therefore have to centre around the keratinocyte expression profile. CIBERSORT was utilised for this purpose. CIBERSORT is a supervised learning algorithm centred on ν-support vector regression and the biological assumption that gene expression mixtures are linear in composition (Shen-Orr *et al.*, 2010; Zhong and Liu, 2012; Newman *et al.*, 2015). The key advantages of CIBERSORT are its robustness to noise, its applicability to tissues beyond its original design and the ability to simultaneously resolve closely-related cell types against other well-defined populations (Xiong *et al.*, 2018; Félix Garza *et al.*, 2019).

3.3.2 Validating CIBERSORT for cutaneous deconvolution

The skin is an attractive tissue for deconvolution; easily accessible for biopsy, well-defined by the major constituent cell types including keratinocytes, comprehensively characterised by a wealth of publicly available data and relatively easy to culture/model *in vitro*. These factors have led to the increasing application of deconvolution tools from their original neoplastic tissues to cutaneous investigation (Félix Garza *et al.*, 2019; Clayton *et al.*, 2020; G. Wang *et al.*, 2020; Gong and Wang, 2020). However, many of these have yet to utilise skin-specific panels of reference cells for accurate

cutaneous. Relying on the original LM22 panel described by Newman *et al.* could lead to both the potentially erroneous inferring of cell types in the skin, or many cell types falling below the approximately one-per cent detection limit determined for CIBERSORT.

Felix-Garza *et al.* (2019) were the first to construct a panel of cell types relevant to bulk skin deconvolution when applied to the CIBERSORT algorithm. Their DerM22 panel was optimised in a similar way to the set of profiles used in this work. By utilising the clear distinction of the dermal and epidermal layers, deconvolution by fibroblasts and keratinocytes, respectively, was tested against their native cutaneous region to determine accuracy. Interestingly, Garza *et al.* showed a general overfit of adipocyte fractions to all layers tested. This included the epidermis samples where the adipocyte fraction could not be biologically true. The authors attribute this to a computational overfit of the adipose signature to the keratinocyte transcriptome based on a report describing trans-differentiation of adipose stem cells into keratinocytes. This did not prove a significant problem in the analysis carried out by Garza *et al.* because only a baseline keratinocyte fraction was investigated. However, it would introduce unwanted artefacts in the ability to infer the composition of the keratinocyte compartment if the aim were to determine sub-populations within this compartment.

In this chapter the optimisation of CIBERSORT for analysis of bulk skin data is undertaken. The curation of appropriate reference cell types for cutaneous tissue is carefully addressed, showing the ability of *in silico* approaches to delineate accurately dermis and epidermis. Newman *et al.*, using their default LM22 panel of reference cells, showed thar the enumeration error of profiling purified leukocyte subsets was approximately 2-15% (Newman *et al.*, 2015). This range is true for sample replicates and not limited to within a specific subset.

The validation of the fibroblast and keratinocyte signatures using the LCM samples shows similar accuracy and is, crucially, not affected by local inflammation or lesion status. There was a fraction, approximately 20%, of the dermal tissue found to be related to a keratinocyte transcriptomic profile. This may simply be an overfit of the signatures caused by an error in the deconvolution, however it may also be experimental or biological. Hair follicles protruding into dermal sections may introduce keratinocytes into the microdissected sections (Philippeos *et al.*, 2018). Given the more cellular dense nature of epidermis than dermis, a contaminating fraction of a magnitude deconvoluted in this compartment is not unreasonable. Additionally, the results may come from the cytokine-related phenotype of the signatures rather than strictly from the cellular identity of the keratinocytes within the deconvolution panel. This keratinocyte fraction within dermis may also be an artefact of cutaneous biology. Fibroblast and keratinocytes at the dermal-epidermal interface

are known to communicate and coordinate tissue homeostasis. As such a transcriptional programme experienced by keratinocytes may, too, be translated in fibroblasts (Schultz *et al.*, 2011; Wojtowicz *et al.*, 2014). Indeed, papillary fibroblasts have been shown to express genes involved in the immune response compared to deeper reticular fibroblasts (Janson *et al.*, 2012). Essentially, this approximate 20% fraction in dermal sections is possibly a cytokine-response fraction regardless of cellular identity as either keratinocyte or fibroblast.

A fibroblast signature is almost completely absent in the epidermal sections. The main cell type deconvoluted from these sections are keratinocytes, of unstimulated and cytokine-stimulated phenotype. The total keratinocyte fraction varies across the sections analysed from approximately 70-90%. This is slightly lower than the documented 90-95% population by keratinocytes of the epidermis. This may be explained by the fraction of epidermal cells being corneocytes. While strictly keratinocytes, these enucleated cells are terminally differentiated and transcriptomically inactive, but may account for 20% of cells in the epidermis (Hoath and Leahy, 2003; Eckhart *et al.*, 2013). A transcriptome-based deconvolution would not detect these corneocytes of the stratum corneum and as such, the results have accurately enumerated the proportion of live keratinocytes in the epidermis.

Further, single-cell data was leveraged to resolve from bulk skin the proliferating and differentiating keratinocyte compartments within the epidermis. This computational deconvolution of such heterogenous samples comes without the disadvantages of requiring prior tissue disaggregation and digestion protocols that affect tissue composition and population proportions.

Validating the ability of CIBERSORT to work with closely related keratinocyte populations derived from various cytokine stimulation conditions is crucial to approaching the *in-silico* deconvolution of skin. Newman *et al.* also demonstrated that closely similar cell types, in their case naïve and memory B cells, or gamma-delta T cells and CD4 memory or follicular helper T cells, generally have large overfit from one another when examining purified populations. A similar overlap was observed in some of the cytokine-response keratinocytes. However, by using the whole transcriptome for feature selection, validation work showed the smallest correctly identified fraction was approximately 65%. There was also minimal domination from any other single signature in the residual fraction. This is an improvement upon LM22 and the various leukocyte profiles, likely demonstrating the transcriptional shift in keratinocytes upon cytokine stimulation is quantifiable and distinct enough to provide gene expression profiles for computational deconvolution.

3.3.3 Expanding the deconvolution pipeline to whole skin to understand cutaneous inflammation

The final part of the results presented in this chapter introduces a full panel of cutaneous cell references for bulk skin deconvolution in the context of investigating keratinocyte sub-populations. The sub-populations of interest in this initial analysis are derived from keratinocytes under stimulation from cytokine exposure. This signature is composed of a total of 19 signatures of which keratinocytes profiles, providing 11 signatures, are dominant. This is built on the work showing that the inflammatory context of the epidermis is paralleled or mimicked by the local keratinocytes. In addition, the total keratinocyte-derived profile is strong enough to be separated into derivative profiles relevant to particular research hypotheses.

This mimicking of immune status of the local structural cells was recapitulated by the breakdown of the keratinocyte compartment into the 11 cytokine exposed profiles. The tissues from different disease states had altered fractions of these cell types, which should enable tracking of these profiles to the known immunobiology of inflammatory skin diseases. Therefore, the next stage of work involves comprehensive deconvolution of bulk skin biopsies from samples of normal, atopic dermatitis and psoriatic skin. The aim of this is to understand the behaviour of the cytokine-stimulated keratinocyte fractions in the context of diseases with well-known immune involvement.

Chapter 4 Defining cell-specific biology in skin disease samples

4.1 Introduction

Inflammatory skin disease is mediated by both epidermal perturbation and heterogenous immune infiltration. Psoriasis is an archetypal T_H1/T_H17 condition, while the pathogenesis of atopic dermatitis is largely driven by T_H2 inflammation (Guttman-Yassky, Nograles and Krueger, 2011b; Tsoi *et al.*, 2019). The effects of these differential pro-inflammatory signals upon keratinocyte biology yet to be fully explored and as such are a major unknown in the complex aetiology of inflammatory skin conditions.

Global analysis of atopic lesions versus non-lesional tissue has demonstrated these tissues share very similar transcriptomic profiles. Indeed, approximately 70% of the differentially expressed genes that distinguish non-lesional (unaffected) skin of atopic dermatitis patients from healthy tissue are found to also distinguish lesional tissue of the same patients from healthy tissue (Suárez-Fariñas *et al.*, 2011). The molecular mechanisms that explain why such phenotypically different tissues can be transcriptomically similar are not fully understood. It is conceivable, that the disease phenotype is driven by differences within a transcriptome of a specific cell population, masked by the similarities of more abundant constituents in bulk tissue data. The pathogenesis of cutaneous inflammation links both immune perturbation and epidermal biology, for example the role of type 2 cytokines in impacting keratinocyte function and barrier function has been previously reported (De Benedetto *et al.*, 2011; Hönzke *et al.*, 2016; Bao *et al.*, 2017; Totsuka *et al.*, 2017). In order to investigate this, one approach is to deconvolute the bulk transcriptome of the lesional and non-lesional atopic dermatitis samples into cell type-specific signatures.

Local expression of cytokines driving cutaneous inflammation are derived from infiltrating $CD4^+$ immune populations and the inflammatory networks they sustain. Expressed cytokines act a effector molecules onto local keratinocytes. It has been demonstrated that *in vitro* culture of monolayer keratinocytes with key psoriasis-associated cytokines such as IL-1β, IL-17A and IFNγ alters the transcriptional response of cells to a more *in vivo* lesional-like immunophenotype (Swindell *et al.*, 2012). Further, patient biopsy data could be clustered according to patterns of transcriptional responses derived from *in vitro* experiments. This demonstrated the utility of applying *in vitro* derived transcriptomic data to stratification of patient samples and idiosyncratic lesional phenotypes.

Functional enrichment analysis of the inflamed transcriptome and clusters of differentially expressed genes in atopic dermatitis and psoriasis has indicated that immune processes are evident in the molecular signature comprising the keratinocyte compartment (Ghosh *et al.*, 2015; Fyhrquist *et al.*, 2019; Clayton *et al.*, 2020). This has important downstream biological consequences for keratinocyte function and cutaneous homeostasis. For example, T_H2 cytokines are reported to disturb skin barrier function through multiple mechanisms including downregulation of cornified envelope proteins, decreased expression of granular-layer tight junction claudins, and dysregulated cutaneous antimicrobial defences by disturbed expression of anti-microbial peptides (Hönzke *et al.*, 2016). Consequently, this impacts the anti-microbial function of the epidermis by impairing multiple mechanisms by which the epidermis prevents pathogen colonisation and subsequent inflammation (Bitschar *et al.*, 2017). Therefore, resolution of the role of $CD4^+$ T cell-derived cytokines on the transcriptomic programmes of keratinocytes and its contribution to altering keratinocyte biology in skin inflammation is critical for understanding host skin biology, disease pathogenesis and, later, host-microbiome crosstalk and dysbiosis.

Similarly, the molecular consequences of T_H17 inflammation on keratinocytes in psoriatic lesions are not completely known. It has been reported that T_H17 inflammation is powerfully amplified by IL-17-activation of keratinocyte expression of TNFα (Chiricozzi *et al.*, 2011). This synergy is a direct example of the crosstalk between immune and keratinocyte interactions, which regulates further keratinocyte inflammatory-related gene expression, for example *IL1B*, *MMP9*, *IL-6*, *IL-8*, *S100A7*, *S100A8*, *S100A9*, *DEFB4* (Chiricozzi *et al.*, 2011). Therefore, during psoriatic inflammation, as during atopic dermatitis, while the driver of disease is cytokines from polarised $CD4^+$ T cells, the epidermal consequences of this are significantly mediated by keratinocyte responses to these cytokines.

To address this aim, CIBERSORT has been shown to be optimal (Newman *et al.*, 2015, 2017). The tool is robust to unknown content, which, while a comprehensive cutaneous panel is used, the overfit of signatures to rare and unaccounted populations cannot be discounted. Most importantly, CIBERSORT robustly deconvolutes closely related populations. When reoptimized for deconvolution of skin biopsy data, the ability of CIBERSORT to distinguish close-related keratinocyte sub-populations in the context of a wider cutaneous panel was maintained (Sections 3.2.5-3.2.7). To study cellular composition that is computationally resolved from bulk mRNA expression the panel of signatures developed in Section 3.2.7 should be applied to whole transcriptome data of patient samples. This aims to confirm T_H2 and T_H1/T_H17 inflammation effect on keratinocyte biology that underscores atopic dermatitis and psoriasis, respectively.

Furthermore, the tracking of specific keratinocyte immunophenotypes in skin biopsies from patients undergoing treatment for inflammatory skin disease (e.g., ciclosporin for atopic dermatitis) allows for the elucidation of the link between immune and keratinocyte biology. While treatment of atopic dermatitis and psoriasis, as investigated in this work by ciclosporin and dupilumab, or etanercept, respectively, is clinically effective at resolving immune abnormalities driving cutaneous inflammation, the effect of treatment or continuing effect of previous immune dysfunction on keratinocyte phenotypes is unknown. Investigation of whether keratinocyte immunophenotypes are resolved or dampened by treatment, or specific therapies but not others prove effective, may help to elucidate new avenues of analysis for understanding inflammatory disease. Additionally, this may help to stratify patients according to treatment efficacy, appropriate time course or exclusion of ineffective therapies for given phenotypes.

4.1.1 Hypothesis

Cutaneous T cell driven inflammation results in alteration of keratinocyte transcriptomic programmes, and these can be used to distinguish skin diseases and understand treatment responses.

4.1.2 Aims

- Resolve cytokine-induced immunophenotypes in keratinocytes, first from healthy skin to understand baseline fractions, and then track these phenotypes in samples from atopic dermatitis and psoriasis.

- Establish key disease related keratinocyte immunophenotypes in atopic dermatitis and psoriasis.

- Understand how these key disease related keratinocyte immunophenotypes respond to treatment of psoriasis with etanercept, and atopic dermatitis with ciclosporin and dupilumab.

4.2 Results

4.2.1 Stimulation with pro-inflammatory cytokines drives distinctive transcriptional changes in keratinocytes

Before deconvolution of keratinocytes into cytokine-induced specific immunophenotypes can be undertaken, it is necessary to characterise the transcriptomic response of the keratinocyte profiles in response to cytokine exposure. Expression data from primary human keratinocyte cultures stimulated *in vitro* with cytokines demonstrates an alteration in the transcriptomic response on these cells. A total of 11 cytokine-stimulated keratinocyte conditions, including an unstimulated control, from two datasets was considered (GEO accession no.: GSE36287 (Swindell *et al.*, 2012), GSE7216 (Sa *et al.*, 2007)). Whole transcriptome analysis for both datasets by MDS shows the relationship of the cytokine-stimulated conditions with each other (**Figure 4.1**). Analysis of the transcriptomes in order to identify keratinocyte sub-populations underscores heterogeneous cell responses. Overall, there is very good clustering by cytokine replicates demonstrating repeatable transcriptomic alterations induced by the same cytokine in keratinocytes.

The clustering of replicates is maintained at the whole transcriptome level after the two datasets are integrated and corrected to remove technical batch effects by COMBAT (**Figure 4.2A**). The two datasets were merged on the similarity of the unstimulated control samples. MDS plots of the merged dataset showed a core group of samples closely associated with these control samples, namely IL-13, IL-26d and IFNα. Hierarchical clustering better shows the closeness of the replicate samples and the neat clustering of these with only a few samples losing this granularity (**Figure 4.2B**).

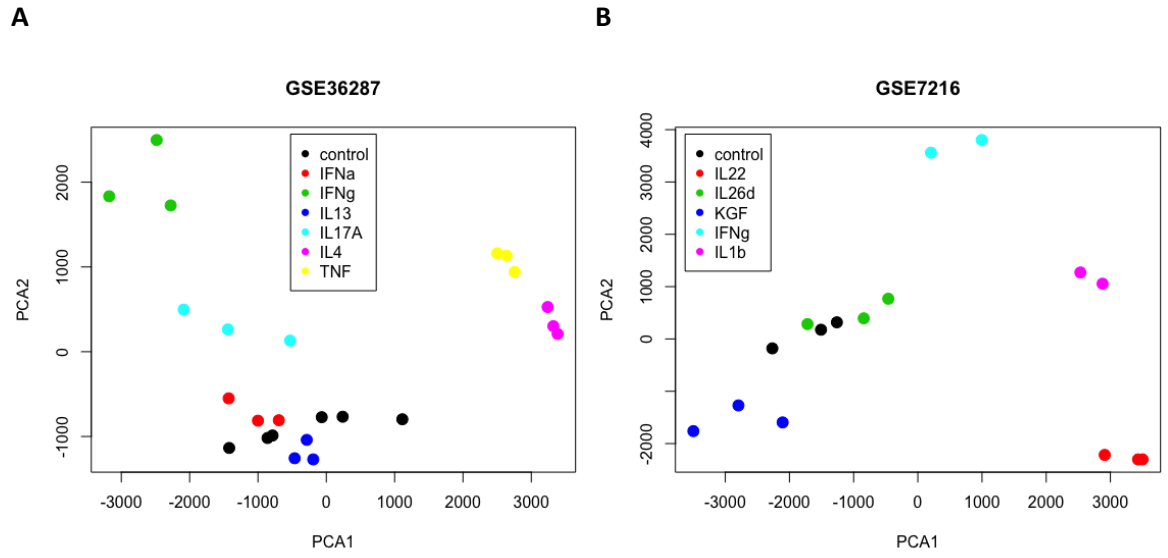
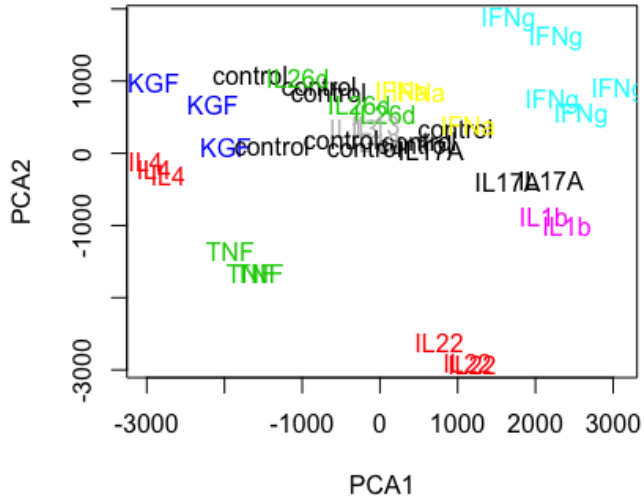


Figure 4.1. Dimensionality reduction of monolayer keratinocytes stimulated *in vitro* with various cytokines. [A] Multi-dimensional scaling (MDS) plot of cytokine-stimulated keratinocytes dataset GSE36287. Primary keratinocytes were cultured untreated or exposed in triplicate to cytokines: IL-4, IL-13, IL-17A, IFNα, IFNγ and TNFα. The IL-17A samples were obtained separately from six donors: in triplicate each of untreated controls and IL-17A stimulation. The batch effect this caused was removed using combat from the GCRMA-normalised expression data. **[B]** MDS plot of cytokine-stimulated keratinocytes dataset GSE7216. Neonatal foreskin obtained NHEKs were cultured untreated or exposed in triplicate to cytokines: IL-1b, IL-22, IL-26d, IFNγ, KGF. The whole transcriptome of the microarray dataset has been reduced to two dimensions representing the first two principal components: x-axis, PC1; y-axis, PC2.

A



B

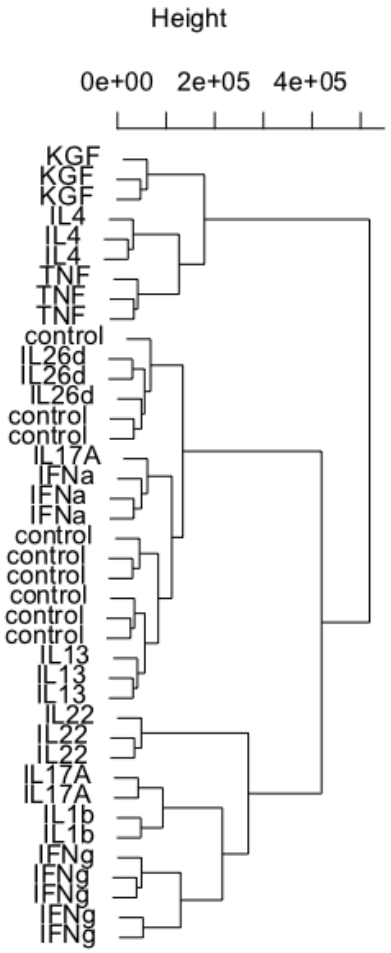


Figure 4.2. Dimensionality reduction of two integrated monolayer keratinocyte datasets stimulated in vitro with various cytokines. Monolayer datasets as presented in Figure 4.1 were integrated: **[A]** MDS plot prior to batch correction by COMBAT showing separation of GSE7216 (black) from GSE36287 (red), **[B]** MDS plot after successful batch removal by COMBAT, **[C]** Hierarchical clustering dendrogram of the integrated and batch corrected dataset using the "ward.D" agglomeration method.

In the previous chapter it was determined that a whole transcriptome input provides the best signature for bulk skin deconvolution. Although this determination will be maintained in further analysis, it is useful to understand the behaviour of this integrated keratinocyte dataset when provided to CIBERSORT for characterising barcode genes. The number of barcode genes ranged from one (IL-13) to 57 (IFNγ) (**Table 8**). With an average of 14 genes characterising the cytokine conditions, it was imagined that these would not be highly specific to their defining sample. However, heatmap analysis of these genes indicated that the barcode genes were indeed reasonably unique to the sample they describe (**Figure 4.3**).

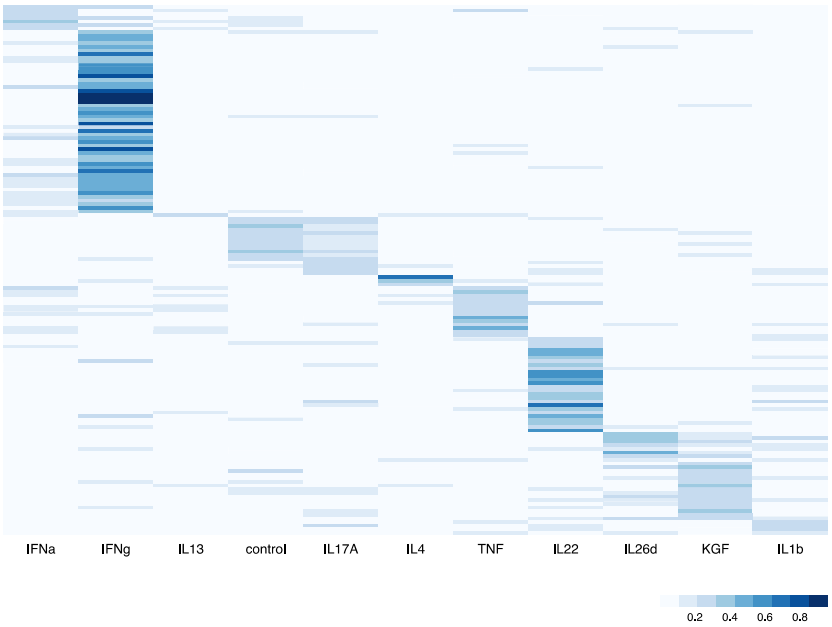


Figure 4.3. Heatmap of CIBERSORT-derived cytokine stimulation barcode genes for the integrated dataset. Genes that were highly expressed in a given cell type were labelled as barcode genes by CIBERSORT. The number of enriched genes per reference sample: IFNα, 7; IFNγ, 57; IL-13, 1; unstimulated control, 11; IL-17A, 5; IL-4,3; TNFα, 14; IL-22, 26; KGF, 16; and IL-1β, 5. Full gene list given in Table 8.

Use of the whole transcriptome of the reference profiles was maintained in order to maximise profile characterisation, minimise overfit between keratinocyte sub-populations and prevent bias to keratinocytes at the detriment of the other cutaneous cell reference signatures. By curating this panel of reference signatures for use by CIBERSORT, the transcriptomic profiles of cytokine-exposed keratinocytes can be used to assess the cytokine phenotype of healthy and inflamed skin. In order to investigate this, it is necessary to computationally resolve the bulk transcriptomic microarray data into constituent cellular compartments.

4.2.2 A curated reference skin signature can be used to deconvolute cellular fractions from whole skin

A reference signature composed of 19 cell types and conditions to reflect the composition of whole skin forms the basis of the deconvolution results presented in this chapter. The cell signatures were sourced from transcriptomes of purified cell types from six public GEO datasets: GSE36287 (keratinocytes: unstimulated, IL-17A, IL-13, IL-4, IFNα, IFNγ, TNFα); GSE7216 (keratinocytes: IL-22, IL-1β, IL-26, KGF); GSE34308 (fibroblasts), GSE4750 (skin-resident $CD4^+$ and $CD8^+$ T cells); GSE49475 (activated Langerhans cells, $CD11c^+$ dermal DCs); GSE23618 (steady-state Langerhans cells). Once all datasets of the purified cell populations were made into a single matrix and taking microarray platform into account with respect to gene probe duplicates, there were 10227 genes defining this reference panel. Using this reference panel, the CIBERSORT algorithm deconvoluted bulk skin samples from publicly available datasets.

Healthy skins were first deconvoluted to establish the baseline profile of the tissue. Deconvolution of healthy skin shows the majority of the cellular profile is accounted for by the sum keratinocytes signature and fibroblasts (**Figure 4.4A**). This is expected from skin biopsies as keratinocytes make up the majority of the cells found in the epidermis and, indeed whole skin, and fibroblasts are the major cell constituent of the dermis.

Focussing on distinct cytokine exposure keratinocyte signatures it is possible to elucidate changes in the keratinocyte context of analysed skin samples (**Figure 4.4B**). Therefore, it is possible to assess from bulk skin microarray data the effect of the cytokine environment and milieu of the skin on keratinocyte biology. Applying *in-silico* deconvolution revealed a major portion of the keratinocyte signature from healthy samples is homeostatic in nature and lacking in hallmarks of disease-related inflammatory influence. Keratinocyte growth factor (KGF) is a fibroblast growth factor considered to be an important paracrine mediator of epithelial proliferation and differentiation (Rubin *et al.*, 1989; Danilenko, 1999). The IL-10 family cytokines: IL-22 and IL-26, may also be considered homeostatic in this context too. In the absence of additional proinflammatory signatures, crucially IL-17; IL-22 and IL-26 are potentially essential cytokines for bridging T cell and epithelial responses, implicated in antimicrobial immunity, and steady-state keratinocyte cellular processes (Wolk *et al.*, 2004, 2006; Braum, Pirzer and Fickenscher, 2012; Tengvall, Che and Lindén, 2016; Itoh *et al.*, 2018).

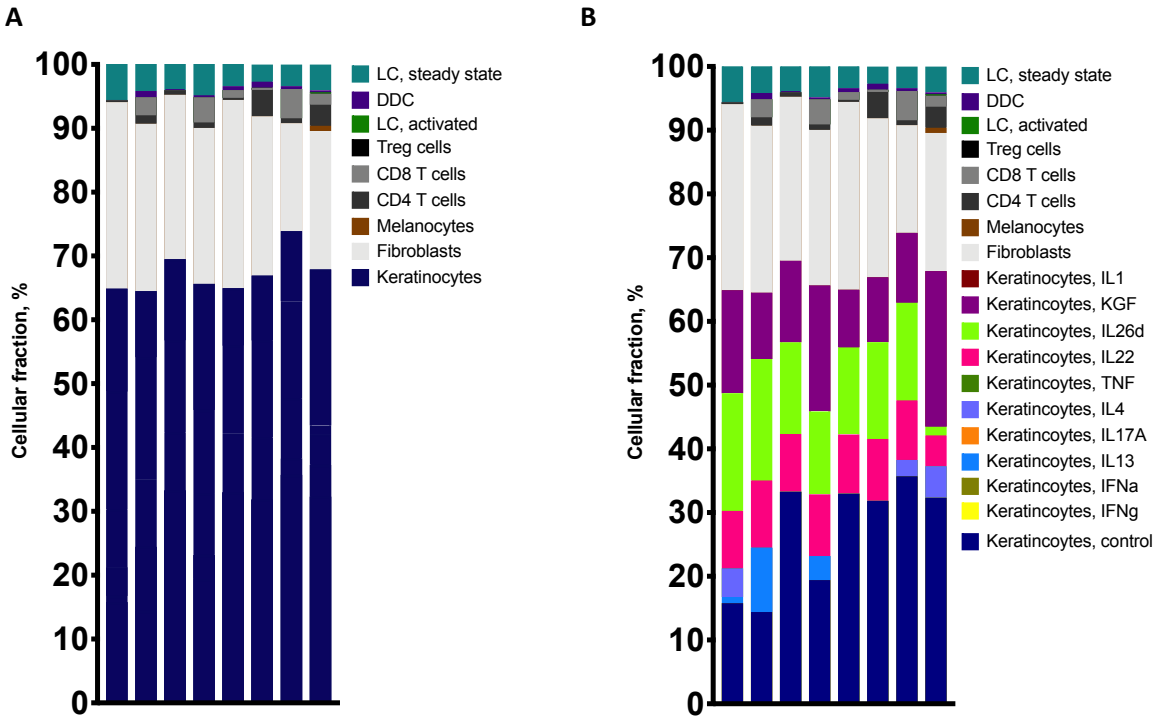


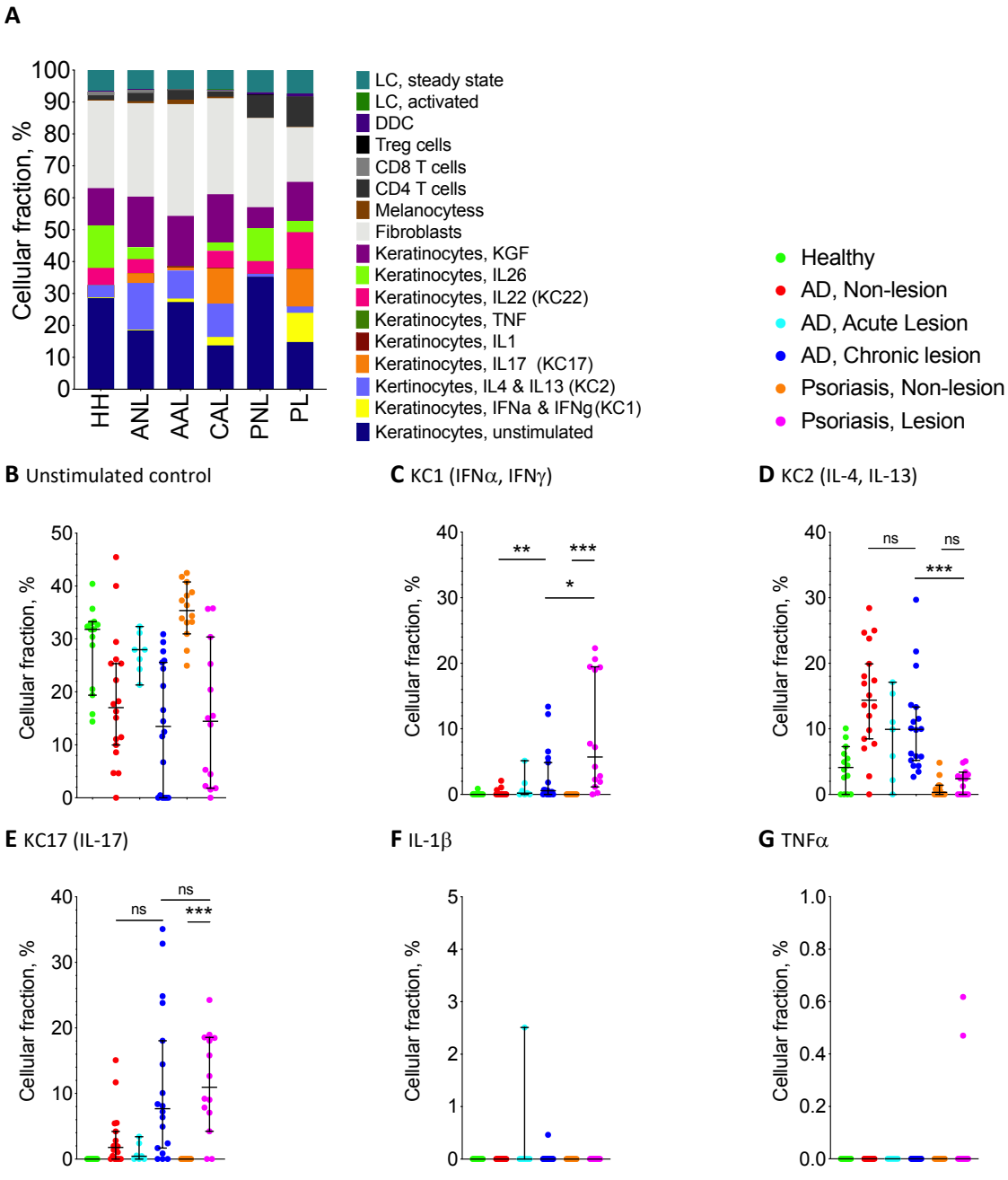
Figure 4.4. CIBERSORT-derived deconvolution of healthy skin. Microarray expression data from eight healthy skins (GSE32924) deconvoluted into reference cellular fractions by CIBERSORT. A total 19 reference cellular signatures were used for deconvolution: *in situ* Langerhans cell, post-migration Langerhans cell, dermal dendritic cell, regulatory T cells, CD8 T cells, CD4 T cells, melanocytes, fibroblasts; and 11 cytokine-stimulated keratinocyte signatures: IL-1β, KGF, IL-26d, IL-22, TNFα, IL-4, IL-17A, IL-13, IFNα, IFNγ, and unstimulated control. **[A]** Breakdown of the samples showing the main cell-based signatures, with the 11 keratinocyte fractions represented as a single, total compartment. **[B]** The same data as [A] but the keratinocyte fractions are coloured according to their proper reference signatures.

4.2.3 Atopic dermatitis and psoriasis keratinocytes can be differentiated by KC1, KC2 and KC17 immunophenotype

To investigate the alterations in cell specific signatures in skin diseases the CIBERSORT deconvolution of healthy skin samples was expanded to include samples from atopic dermatitis and psoriasis patients. Combining the CIBERSORT-derived deconvolution of GSE32924 and GSE36842 allowed comparison not only lesional and non-lesional skins, but also to make comparisons between acute and chronic lesions. Incorporation of the lesional and non-lesional samples from the psoriasis dataset (GSE34248) into the interrogated sample collection enabled cross-comparison to another inflammatory skin disease. Furthermore, the increase in sample number increases the statistical power of analysis. This resulted in a dataset of 14 healthy control (HH), 18 atopic

dermatitis non-lesional samples (ANL), 18 atopic dermatitis chronic lesional samples (CAL), 7 atopic dermatitis acute samples (AAL) and 14 each of psoriatic lesions (PL) and non-lesions (PNL).

Compared to the healthy controls, a significant disease-related change occurred in keratinocyte profiles, namely in interferon alpha, IL-4 and IL-13, IL-17, IL-22 and response signatures (**Figure 4.5A**). As these signatures identify subpopulations of keratinocytes displaying a molecular response to a specific disease-related inflammatory axis, these can be termed according to an immunophenotype as KC1 (interferon response), KC2 (T_H2 response), KC17 (IL-17 response) and KC22 (IL-22 response). Statistical analysis was therefore only conducted on these disease relevant fractions.



[*figure continues onto next page*]

[*figure continued from previous page*]

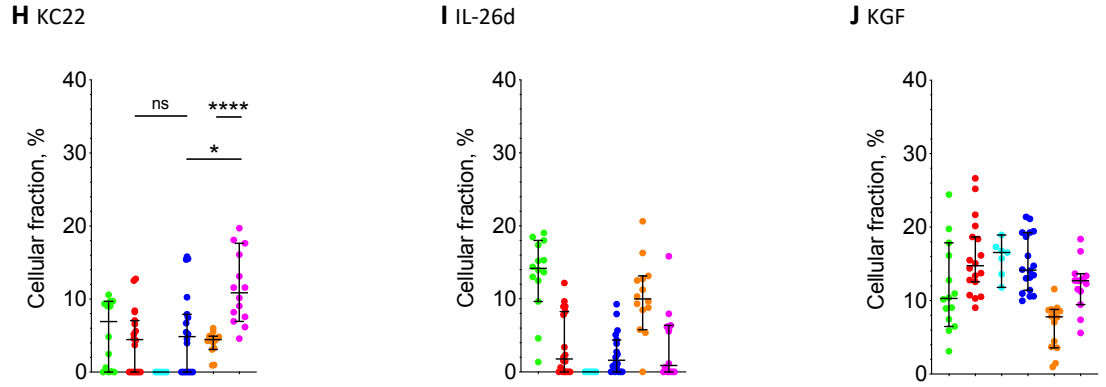


Figure 4.5. Resolution of whole skin samples into constituent cellular profiles by CIBERSORT. [A] Average deconvolution of bulk skin transcriptomes across six skin phenotypes: healthy (n=14, green, HH), AD non-lesional (n=18, red, ANL), AD acute lesional (n=7, cyan, AAL), AD chronic lesional (n=18, blue, CAL) psoriasis non-lesional (n=14, orange, PNL) and psoriasis lesional (n=14, magenta, PL). **[B-G]** Individual keratinocyte response fractions across the skin phenotypes: [B] unstimulated control; [C] KC1 (IFNα, and IFNγ); [D] KC2 (IL-4 and IL-13); [E] KC17 (IL-17); [F] IL-1β; [G] TNFα; [H] KC22 (IL-22); [I] IL26d; and [J] Keratinocyte growth factor, KGF. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Normality testing by D'Agostino-Pearson method: non-normal data significance tested by unpaired nonparametric Mann-Whitney tests, normal data significance tested by unpaired t-test, all Bonferroni corrected for multiple testing. Error bars show median ±95% confidence interval.

The keratinocyte interferon programme, KC1, was prominent in psoriatic lesions but absent from non-lesional counterparts (Bonferroni p<0.0001) (**Figure 4.5C**). A smaller KC1 immunophenotype was also resolved from chronic atopic lesions showing a trend of increase from non-lesional and acute comparisons. While the KC1 fraction was smaller in chronic atopic lesions that in psoriatic lesions (Bonferroni p=0.024), it was significantly raised compared to non-lesional skin from patients with atopic dermatitis (Bonferroni p=0.0048).

In tissues from atopic dermatitis, the main disease related signatures were KC2 and KC17. The KC2 fraction of the atopic samples, revealed this to be a T_H2-driven disease (**Figure 4.5D**). This fraction was present in both lesional and non-lesional skins, which matches our understanding of this disease; that non-lesional skin has sub-clinical T_H2 pathology. There proved no statistical difference in comparing the size of the KC2 fractions between chronic lesional and non-lesional atopic samples. This common signature may underlie the observation that non-lesional atopic dermatitis skins have a much more similar transcriptomic profile to lesional skins than true healthy

comparators, despite non-lesional skins appearing clinically normal. KC2 was remarkably specific to chronic atopic lesions when compared to psoriatic lesions (Bonferroni p<0.001).

As expected, the KC17 signal was significantly elevated in psoriatic lesions compared with healthy controls and paired non-lesional samples (Bonferroni p<0.001) (**Figure 4.5E**). The elucidation of this IL-17 responding fraction from psoriasis lesions confirmed the provenance of this immunophenotype. In the blood, a mixed T_H2/T_H17 cytokine biology has been detected in atopic dermatitis, and this cytokine milieu could be expected to be found in the skin (Roesner *et al.*, 2015). Chronic atopic lesions revealed a similar KC17 fraction to psoriatic lesions, which was not statistically different in magnitude between these disease types. However, non-lesional atopic samples also displayed a slight KC17 phenotype, which proved the comparison between atopic non-lesional and chronic lesional statistically not significant (Bonferroni p=0.0555). This, therefore, suggested an intensification of the KC17 programme over the course of an atopic lesion, and its magnitude in chronic lesions akin to a psoriatic lesion, suggested an evolution in T_H17 inflammation over lesion duration.

4.2.4 A KC17 immunophenotype is indicative of a higher atopic programme

A transcriptomic programme in keratinocytes from IL-17 stimulation (KC17) has been resolved from psoriatic lesions and was mirrored in size in chronic atopic lesions, and present at smaller background levels in non-lesional and acute lesional tissues. In contrast, this signature was entirely absent from psoriatic non-lesions and control healthy skin. This formed the assumption, therefore, that this fraction of keratinocytes is linked to severe cutaneous inflammation.

The analysed samples were grouped according to the presence or absence of the KC17 fraction. Non-lesional atopic dermatitis samples were split almost evenly by this distinction; 19 samples were negative, and 15 samples were KC17 positive, of any fraction size. There were double the number of KC17 positive lesional samples (n=23) than there were KC17 negatives (n=11), underscoring the lesion association of this response fraction. For both non-lesional and lesional KC17 comparisons, disease severity as measured by SCORAD was not significantly different (**Figure 4.6A, D**). Although, all of the lesional samples from patients extremely severe disease (SCORAD >80, n=5) were KC17 positive. Utilising the same barrier expression panel as described earlier, both KC17 groups in both lesional and non-lesional samples showed significantly reduced expression of barrier-related genes (all comparisons, p<0.0001) (**Figure 4.6B, E**). For lesional samples the comparison between KC17 groups was not significant (p=0.5731), but, interestingly, between non-lesional KC17-negative and KC17-positive samples the latter samples displayed reduced barrier expression (p=0.0169).

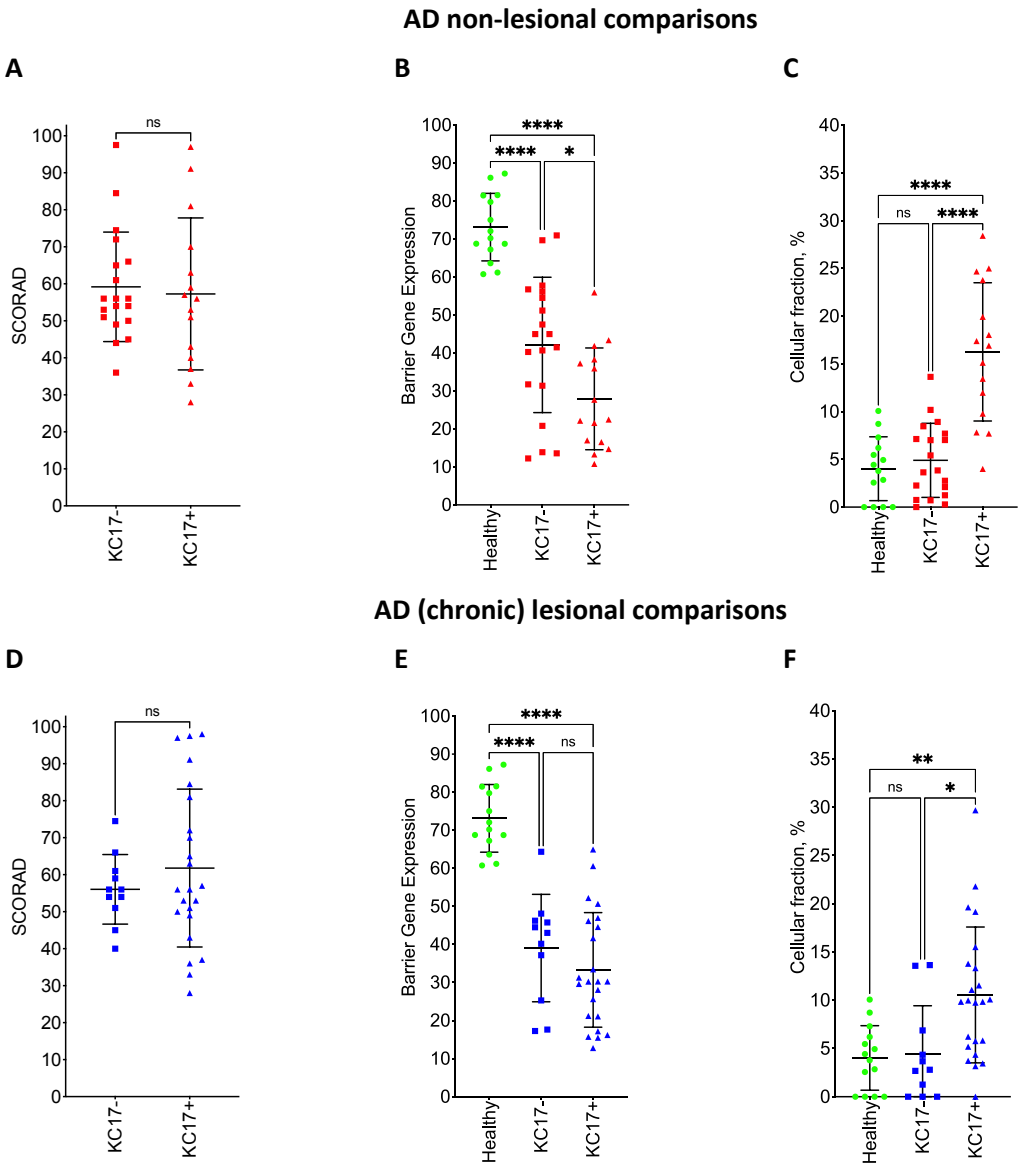


Figure 4.6. Association between AD KC17 status of skin samples and disease severity, barrier expression and KC2 fraction size. AD non-lesional (top panel, red) and AD lesional (bottom panel, blue) samples were grouped according to KC17 status: positive (triangles) or negative (squares) and compared for differences in **[A, D]** SCORAD; **[B, E]** Expression of a panel of barrier-related genes; **[C, F]** KC2 immunophenotype fraction. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean ±SD.

KC17 classification showed remarkable delineation of atopic dermatitis samples according to KC2 immunophenotype (**Figure 4.6C, F**). Although absence of KC17 correlated with a healthy skin phenotype, it was not expected that this would coincide with atopic samples having a low KC2 fraction. For both non-lesional and lesional samples that were negative for KC17, the KC2 fraction was indistinguishable to healthy controls (p=0.8754 and 0.9971, respectively). Only those atopic

samples that were positive for KC17 displayed the KC2 immunophenotype expected from atopic skin. Indeed, the KC2 fraction in the KC17-positive samples was significantly higher than in KC17-negative samples for both non-lesions ($p<0.0001$) and lesions ($p=0.0170$).

4.2.5 Treatment-specific modification of computationally identified keratinocyte immunophenotypes

Computational deconvolution of bulk skins from atopic dermatitis and psoriasis patients has allowed the inferring of complex $T_H1/T_H2/T_H17$ inflammation experienced by lesional epidermis which was apparent by alteration of keratinocyte transcriptomic programmes. To validate these keratinocyte programmes as disease-specific, their modification or amelioration by treatment should further be demonstrated. Datasets involving the treatment of psoriasis using etanercept and treatment of atopic dermatitis with ciclosporin or dupilumab were obtained for this investigation.

4.2.5.1 Etanercept treatment of psoriasis

Further validation of the KC17 immunophenotype is derived from deconvolution of the samples from the etanercept treatment of psoriasis undertaken in the dataset GSE11903 (Zaba *et al.*, 2009). Samples at week-0 comparing psoriatic lesional to non-lesional recapitulated the KC1 and KC17 immunophenotypes were significantly resolved from lesional skin (**Figure 4.7A, C, D, F**). This was true for cohorts of patients that respond or not to etanercept. Indeed, KC1 and KC17 averages between responders and non-responders were not qualitatively different. The 12-week course of etanercept treatment resolved the KC1 (i.e., keratinocyte responses to IFN) immunophenotype to zero in patients responding to treatment (**Figure 4.7A**) but did not in non-responders (**Figure 4.7D**). Although, the non-responder group may have been affected by an outlier at the end of treatment. At four weeks of treatment, many of the responder group had lost the KC1 immunophenotype while the non-responding group had not. This may indicate a longer time course might completely resolve this fraction.

It was described that although etanercept mechanism of action is the inhibition of TNFα, response to treatment in psoriatic patients was dependent on downregulation of myeloid dendritic cell genes and, crucially, T_H17 pathways (Zaba *et al.*, 2009). Etanercept treatment of psoriasis over the course of 12 weeks had a remarkable dampening effect on KC17 fractions (**Figure 4.7C**). Longitudinal abrogation of this fraction in responders further underscores the role of this keratinocyte programme in psoriasis pathogenesis ($p<0.0001$). In contrast, however, patients who did not respond clinically to etanercept treatment were found to have an unaltered KC17

immunophenotype over the time course, which was not significantly different between pre-treatment and week-12 (**Figure 4.7F**).

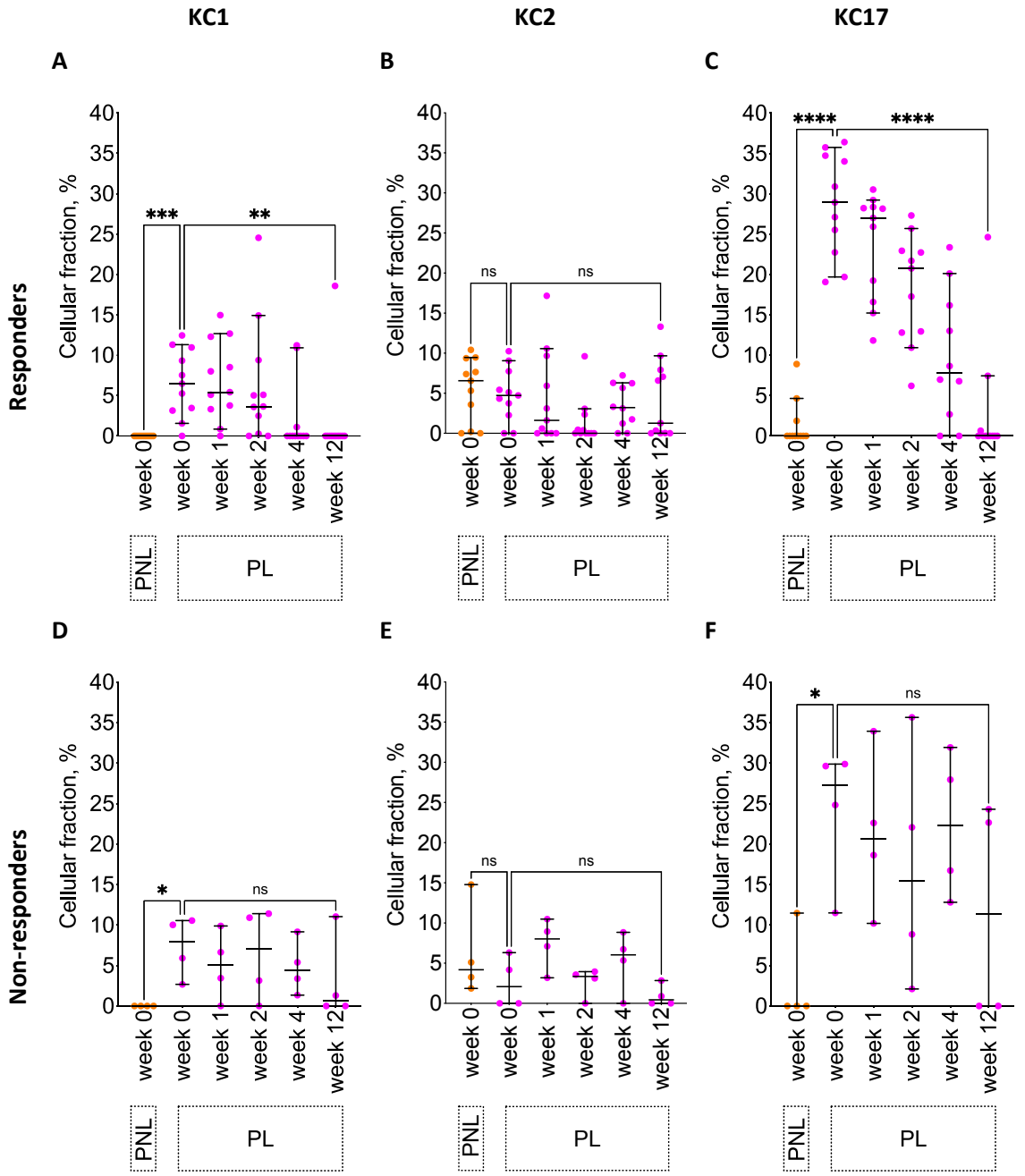


Figure 4.7. Disease-related immunophenotypes during etanercept treatment of psoriasis. KC1 (left), KC2 (middle) and KC17 (right) immunophenotypes were resolved from bulk skin transcriptomes of psoriatic patients (lesional, PL, magenta; non-lesional, PNL, orange) undergoing 12 weeks of etanercept treatment of which 11 were classified as responders [A-C] and four as non-responders [D-F]. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show Kruskal-Wallis test with multiple comparisons adjusted p-value. Error bars show median ±95% confidence interval.

4.2.5.2 Ciclosporin treatment of atopic dermatitis

It has been previously reported that the potent immunosuppressant ciclosporin supresses production of T cell cytokines and can be effective in treatment of atopic dermatitis (Khattri *et al.*, 2014). Additionally, production of IL-17 by T_H17 T cells has been shown to be sensitive to ciclosporin treatment (Ziolkowska *et al.*, 2000; Cho *et al.*, 2007; Zhang *et al.*, 2008; Saitoh *et al.*, 2016). From finding that IL17-respondng KCs are generally increased in chronic lesions compared to other atopic disease stages, the effect on keratinocyte signatures in atopic dermatitis on ciclosporin treatment was assessed (**Figure 4.8**). A longitudinal microarray dataset of 12-week ciclosporin treatment of AD patients was obtained from GEO (GSE58558 (Khattri *et al.*, 2014)). This dataset has samples from both lesional and non-lesional samples at pre-treatment baseline, 2-weeks mid-treatment, and 12-weeks at the end of treatment. The dataset presented six patients who did not respond to treatment of a total cohort of 19 study patients, but they were not present at all timepoints or tissue types do not provide statistical power for investigating these keratinocyte immunophenotypes in non-responders.

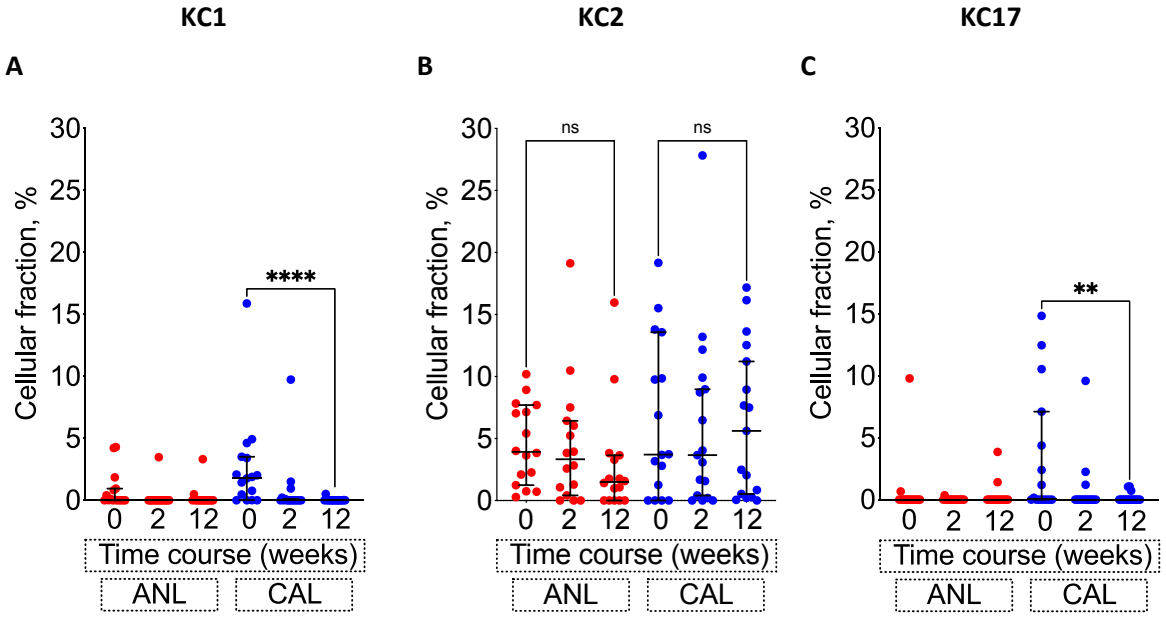


Figure 4.8. Disease-related immunophenotypes during ciclosporin treatment of AD. [A] KC1; **[B]** KC2 and **[C]** KC17 immunophenotypes were resolved from bulk skin transcriptomes of AD patients (n=16) undergoing 12 weeks of ciclosporin treatment (chronic lesional, CAL; non-lesional, ANL). Samples were taken from non-lesional (red) and chronic lesional (blue) skin at baseline, week-2, and week-12. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show Kruskal-Wallis test with multiple comparisons adjusted p-value. Error bars show median ±95% confidence interval.

From this dataset, KC1 fractions were confirmed as lesion associated and were resolved by ciclosporin treatment ($p<0.0001$) (**Figure 4.8A**). Further, the constitutive KC2 response of atopic keratinocytes regardless of lesion status was recapitulated (lesional versus non-lesional at pre-treatment baseline, $p=0.298$) (**Figure 4.8B**). Surprisingly, and despite the described resolution of active lesions in response to treatment and the improvement in disease severity scores, ciclosporin treatment did not reduce the atopic transcriptomic response of keratinocytes from atopic dermatitis patients. In lesional samples the KC2 fraction remained remarkably stable across the course of treatment. There was a downward trend for this fraction in non-lesional biopsies, however, this did not prove significant in comparing pre-treatment and 12-week results.

The nuanced treatment of atopic dermatitis with ciclosporin administration was captured by the reduction in the KC17 fraction in lesional tissues to almost zero; a healthy phenotype ($p=0.0085$) (**Figure 4.8C**). The improvement of this fraction was observed very early in the treatment time course at two weeks indicating resolution of this immunophenotype was highly targeted and effective by ciclosporin.

This deconvolution work has demonstrated a keratinocyte immunophenotype characteristic of atopic skin and responsive to T_H2 cytokines. This fraction was unaltered by broad-spectrum ciclosporin treatment despite clinical efficacy.

4.2.5.3 Dupilumab treatment of atopic dermatitis

To follow the effect of anti-T_H2 treatment on keratinocyte transcriptomic responses, the dupilumab treatment dataset GSE130588 was investigated (Guttman-Yassky *et al.*, 2019). This placebo-controlled double-blind trial investigated the effect of subcutaneous dosage of 200mg dupilumab to cellular and molecular phenotypes of moderate-to-severe AD over 16 weeks. The KC1 (**Figure 4.9A,D**) and KC2 (**Figure 4.9B,E**) fractions were again recapitulated as disease related fractions in both treatment and placebo groups.

The KC1 fraction was ameliorated in the dupilumab treated cohort ($p=0.0137$) (**Figure 4.9A**) but not in those that received the placebo ($p=0.3083$) (**Figure 4.9D**). This is particularly interesting because Guttman-Yassky *et al.* (2019) in their reporting of dupilumab treatment of this study cohort did not find downregulation of T_H1/IFNγ markers in response to IL-4Rα blockade. This may indicate a lack of sensitivity at the molecular level whereas a computational approach is able to identify this at an inferred cellular level.

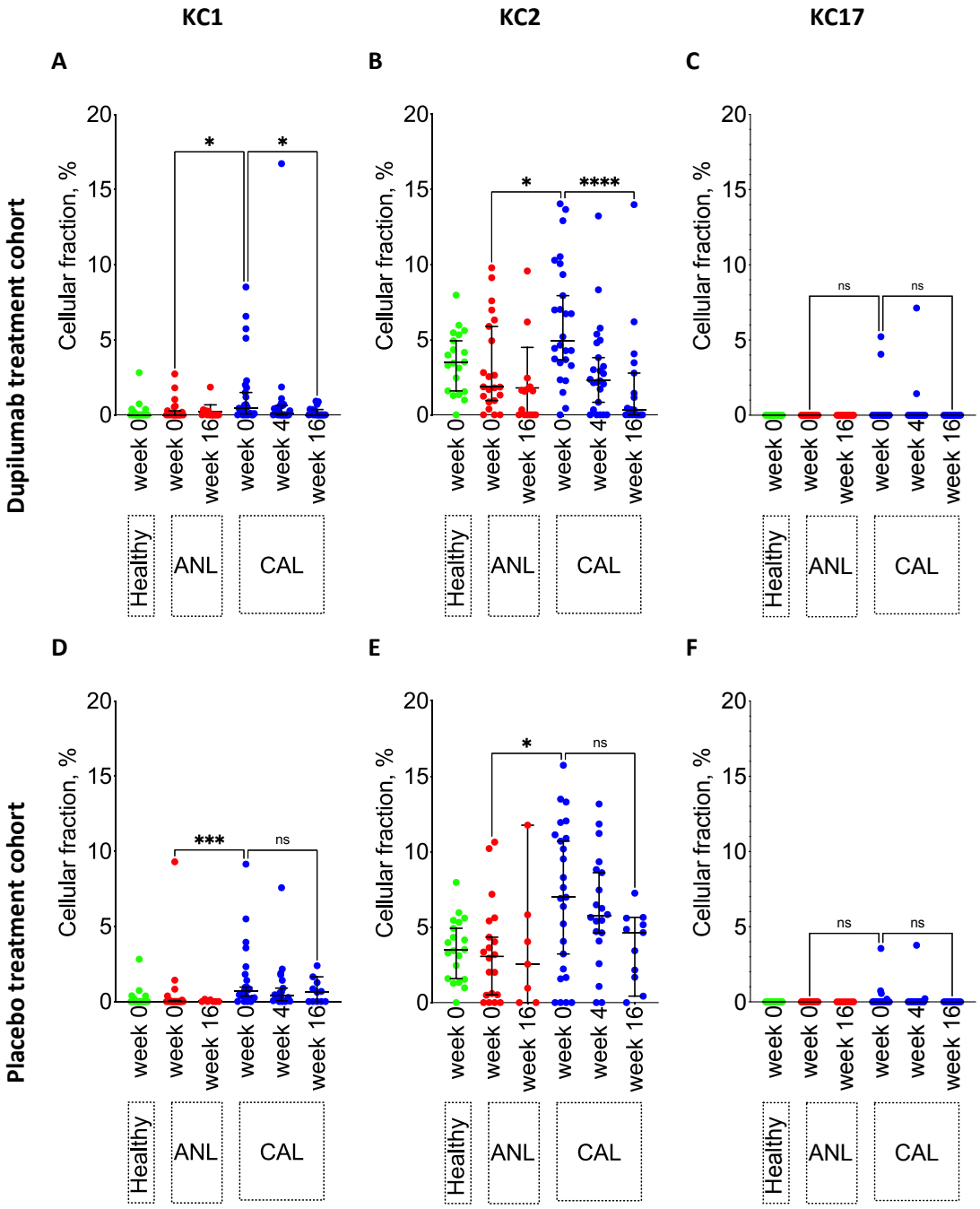


Figure 4.9. Disease-related immunophenotypes during dupilumab treatment of AD. [A, D] KC1; **[B, E]** KC2 and **[C, F]** KC17 immunophenotypes were resolved from bulk skin transcriptomes of AD patients undergoing 16 weeks of dupilumab treatment (top penal) or placebo control (bottom panel) (chronic lesional, CAL; non-lesional, ANL). Samples were taken from non-lesional (red) and chronic lesional (blue) skin at baseline and week-16, with an additional lesional samples taken mid-treatment at week-4. Healthy controls (green) are also included and shown in both treated and placebo graphs. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show Kruskal-Wallis test with multiple comparisons adjusted p-value. Error bars show median ±95% confidence interval.

KC2 responses in this cohort, even at pre-treatment baseline were generally lower than observed in previous datasets (**Figure 4.9B, E**). Previously, chronic lesions showed an averaged enumerated KC2 fraction of 10%. In dupilumab treated patients, the pre-treatment KC2 fraction was an average of 5%, only slightly raised compared to healthy controls. Indeed, unlike previous comparisons, KC2 fractions were significantly raised in pre-treatment lesional samples compared to non-lesional samples in both dupilumab and placebo groups (p=0.0139 and 0.0102, respectively). Indicative of the reduced disease burden on keratinocyte immunophenotypes, the KC17 fraction was not higher in chronic lesional biopsies of atopic dermatitis (**Figure 4.9C, F**).

The aim of interrogating this dataset was to predict that KC2 fractions would be diminished upon targeted anti-T_H2 treatment. As the lesional samples still presented raised KC2 this remained possible. Over the course of 16 weeks, subcutaneous dupilumab significantly reduced KC2 immunophenotype in lesions compared to baseline (p<0.0001) (**Figure 4.9B**). Indeed, the majority of the decrease occurs by week 4. The placebo control arm did not show a significant decrease in KC2 in lesions over the course of treatment (**Figure 4.9E**). While there was a trend of decrease between week zero and week 16, the dramatic decrease seen in the treated cohort between early timepoints was not similarly observed. Therefore, it is likely this trend of decrease is an artefact of the long duration of the study reflecting normal but slow resolution of lesion pathology.

4.2.6 Dupilumab but not ciclosporin improve barrier expression in atopic dermatitis

In Section 3.2.2 barrier integrity was inferred using a panel of barrier marker genes, with expression normalised to present an indicator of barrier status in the skin. The role of atopic inflammation, and indeed *S. aureus* colonisation of epidermis, on barrier integrity is well understood (Ardern-Jones *et al.*, 2007; Brandt and Sivaprasad, 2011; Esaki *et al.*, 2015; Hönzke *et al.*, 2016; Danso *et al.*, 2017). Treatment with ciclosporin did not improve KC2 atopic dermatitis immunophenotype, while dupilumab proved effective in diminishing this fraction. Therefore, it was considered important to investigate the effect of these treatment modalities on the expression of the barrier marker panel.

Ciclosporin treatment did not improve lesional barrier expression (**Figure 4.10A**). Although the disease severity improved with ciclosporin, overall, the KC2 immunophenotype persisted. This stability of the keratinocyte phenotype over 16 weeks possibly explains the ongoing reduced barrier gene expression over the course of treatment. In contrast, specific inhibition of T_H2 by dupilumab effectively reduced the KC2 immunophenotype, possibly suggesting a superior anti-TH2 mechanism mediated by dupilumab than ciclosporin. In addition, with effective TH2 inhibition by dupilumab there was progressive improvement in barrier expression (p=0.0302) (**Figure 4.10B**) in contrast to that seen with ciclosporin, linking the known counter regulation of barrier function by T_H2

inflammation. The placebo control cohort experience no improvement in barrier gene expression (p=0.7872) (**Figure 4.10C**).

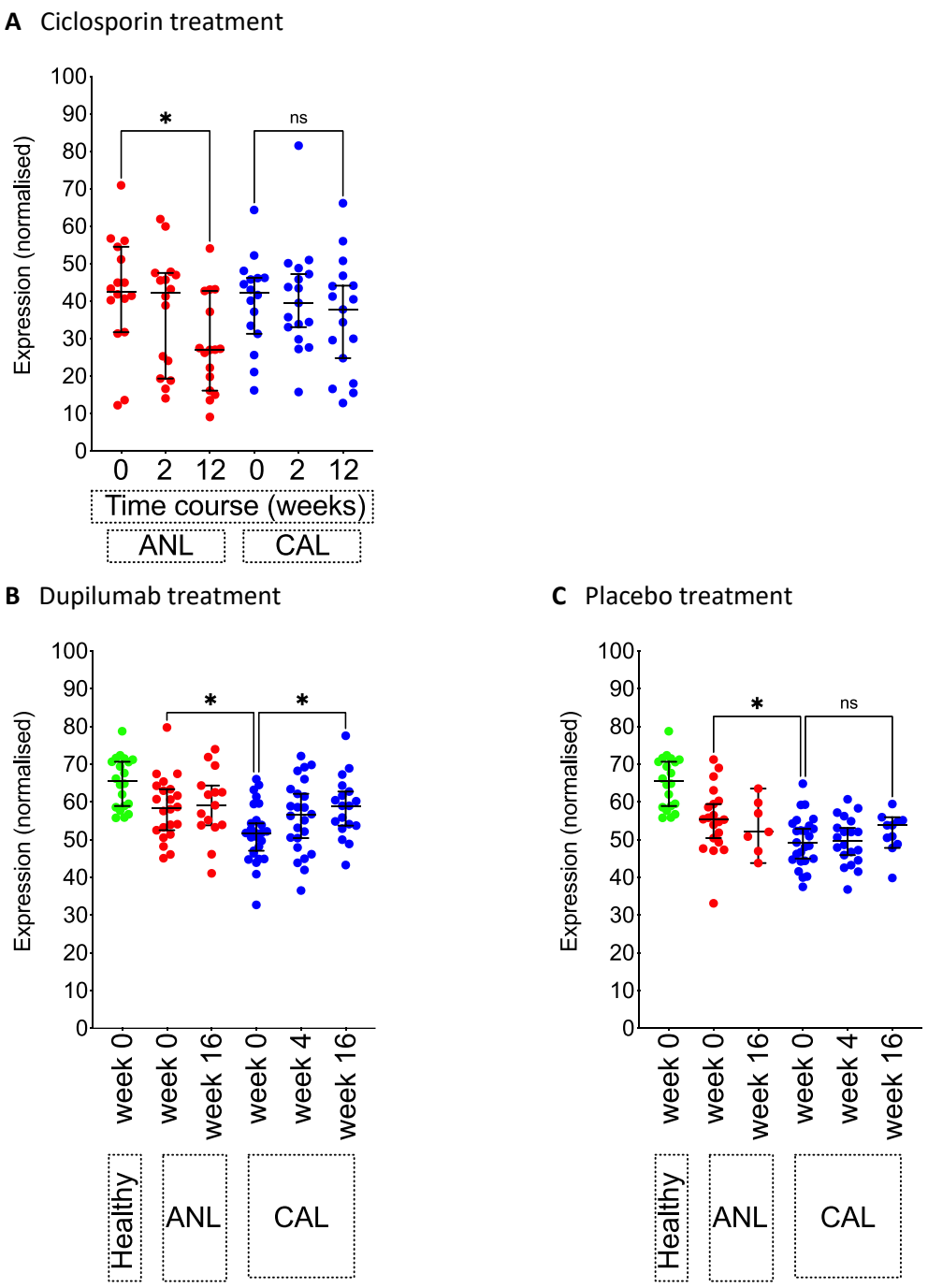


Figure 4.10. Barrier gene expression in response to treatment of atopic dermatitis. A panel of barrier genes was used to infer barrier integrity and is shown in healthy (green), AD non-lesional samples (ANL, red) and AD chronic lesional samples (CAL, blue) in response to **[A]** ciclosporin treatment; **[B]** dupilumab treatment; and **[C]** placebo treatment related to dupilumab treatment. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show Kruskal-Wallis test with multiple comparisons adjusted p-value. Error bars show median ±95% confidence interval.

4.3 Discussion

Understanding of the role of inflammatory cells and the milieu of inflammatory signals in inflammatory skin disease has vastly improved over recent years. Despite this, changes in growth and differentiation of keratinocytes in response to infiltrating T cell populations and associated immune mediators remains to be fully elucidated. This is perhaps, in part, been due to technical challenges associated with addressing this question. Atopic dermatitis and psoriasis offer an opportunity to understand how altered cytokine signalling can shape keratinocyte immunophenotypes.

A global analysis of the whole skin transcriptome precludes the resolution required to elucidate individual responses of the various cutaneous cell populations. Current and widely applied approaches for this such as flow cytometry and immunohistochemistry, provide insight, valuable though it may be, for only a narrow set of pre-defined markers. Therefore, insight is constrained by the low dimensions this type of investigation can yield.

Non-hypothesis-driven approaches allow for unbiased investigation of molecular and cellular changes in perturbed systems. One such approach is single-cell sequencing of whole skin or of flow-sorted populations. This enables detailed characterisation of whole skin samples unimpeded by limitations of current understanding or canonical insight. However, as yet, single-cell sequencing is limited by factors of cost, practicality and experimental artefacts. Tissue dissociation and population marker definition, along with appropriate number of cells sequenced often confound potential sampling biases. Nevertheless, single-cell investigation of cutaneous inflammation has drastically improved understanding of skin biology and disease (Cheng *et al.*, 2018; S. Wang *et al.*, 2020; Reynolds *et al.*, 2021). Despite these works, large single cell investigation of the skin and cutaneous inflammation remains the exception not the norm.

This project sought to address cellular analysis within tissue by a different approach. Using machine learning, based on the CIBERSORT algorithm, different skin cell types, and subsequently cellular response phenotypes, could be reliably identified in healthy skin and samples from both atopic dermatitis and psoriasis. This method was tested and validated on transcriptomic studies of microdissected healthy skin and bulk psoriatic skin that is known to display immune pathway dependence. This shows that this *in silico* approach is a powerful method for investigating transcriptomic signatures in skin samples of complex disease.

This chapter details the *in-silico* deconvolution of healthy and inflamed whole skin transcriptomic data. Leading this working is the hypothesis that whole tissue expression could be resolved into

distinct and characteristic cell-based signals that add to current understanding of cutaneous inflammation. Unsupervised co-expression analysis underscored this hypothesis by showing that not only could the bulk transcriptome be resolved into inferred cellular compartments or processes, but also there were multiple keratinocyte profiles, some of which mirrored the expression of the immune-based profiles.

To further address this hypothesis, an expanded CIBERSORT computational approach, optimised for deconvolution of human skin samples was utilised. This analysis allowed extraction, from bulk transcriptomic data, of specific keratinocyte signals, reflecting the steady-state and those associated with cytokine stimulation. The analysis demonstrated that, from a keratinocyte perspective, the biology of the epidermis is skewed to a T_H2 phenotype in all atopic dermatitis skin tissues regardless of the lesion presence or status. This is in concordance with previous reports, which have demonstrated that atopic skin is preferentially and abnormally polarised to a T_H2 programme compared to a T_H1 biology as expected of healthy skin (Albanesi *et al.*, 2001; Dhingra, Gulati and Guttman-Yassky, 2013; Newell *et al.*, 2013).

Additionally, this confirms the classification of non-lesional and clinically normal skin from patients with atopic dermatitis as abnormal by most researchers and clinicians (Jensen *et al.*, 2004; Sugiura *et al.*, 2014; Danso *et al.*, 2017). Interestingly, despite the accepted T_H2 pathology of the atopic dermatitis paradigm, independent meta-analysis of two of the datasets analysed (GSE32924, GSE36842) found that upregulation of the classical T_H2 cytokines, IL-4 and IL-13, was not consistently observed (Ghosh *et al.*, 2015). The detection of low expression genes by microarray has long been a reported obstacle (Zhao *et al.*, 2014; Esaki *et al.*, 2015). Despite the lack of detection of the cytokine transcripts, using a cell-based classifier of the whole transcriptome, the *in-silico* approach used in this work was able to identify IL-4 and IL-13 transcriptomic gene expression profiles specific to keratinocytes. This atopic, T_H2 transcriptomic programme imposed upon keratinocytes was termed KC2.

The uniformity of the KC2 transcriptional response in non-lesional skin samples can be a result of disease severity. As described in the supplementary information associated with the transcriptomic datasets, all the samples were obtained from patients that all had active lesions. During such active disease phases, it may be that the skin is only temporarily and, across the whole tissue, primed to an atopic phenotype, while localised but specific insults result in a lesion. One such possible trigger for lesions could be from microbiome factors, specifically *Staphylococcal* exposure. Furthermore, deconvolution analysis of independent cohorts revealed that not all non-lesional samples display a lesion-like KC2 profile. These less atopic non-lesional samples were derived from patients involved

in a dupilumab clinical trial where disease severity may have limited involvement either by the patient or clinician-led exclusion.

However, this work proposes that the KC2 immunophenotype of keratinocytes from atopic dermatitis patients is not a transient feature of active disease but a permanent phenotype of these cells. Considerable and convincing work has to date described that patients with atopic dermatitis have an abnormal polarisation to T_H2 immunity and there is no reason why this would not be extended or apparent in the epidermis. Indeed, reports have demonstrated that atopic skin responds in a T_H2 manner unlike healthy skins that typically mount T_H1 responses (Ardern-Jones *et al.*, 2007; Newell *et al.*, 2013).

In contrast, T_H1 and T_H17 axes were dominant in psoriasis (Kagami *et al.*, 2010), and induced specific keratinocyte immunophenotypes, KC1 and KC17, respectively, in psoriatic lesions. Critically, these disease-associated fractions were absent from non-lesional skin from psoriatic patients, underscoring the healthy phenotype of this tissue. IL-17 driven responses by keratinocytes induces and synergises with keratinocyte-derived TNFα (Zaba *et al.*, 2007, 2009; Chiricozzi *et al.*, 2011), the phenotype of which is inferred from the KC17 immunophenotype in psoriatic lesional samples. Further, IL-17 action on psoriatic keratinocytes is reported to affect differentiation and barrier function, such as impaired filaggrin expression, contributing to the perturbed phenotype of keratinocytes in this tissue (Gutowska-Owsiak *et al.*, 2012; Pfaff *et al.*, 2017; Swindell *et al.*, 2017).

The deconvolution data presented in this chapter further support usefulness of this approach for tracking the effect of treatment in inflammatory skin disease. Ciclosporin is a potent immunosuppressant of T cell inflammation via inhibition of calcineurin and induces rapid clinical improvement of atopic dermatitis. Over the course of ciclosporin treatment lesional keratinocytes became more non-lesional in phenotype as disease severity was ameliorated. This phenotype change was caused by lesional keratinocytes losing KC17 fractions while completely retaining KC2 fractions. Systemic treatment of atopic dermatitis did not make atopic skins similar to healthy skin because the effect of T_H2 immunity upon keratinocytes was maintained. As such treatment only reverted lesional skin to a non-lesional phenotype. At the end of treatment, lesions were resolved but the abnormal cutaneous immunity, measured by transcriptomic responses by these processes in keratinocytes, persisted.

Indeed, the extent of epidermal barrier function correlation with clinical improvement during ciclosporin administration is not yet fully known. Markers for epidermal thickness/proliferation (e.g. KRT16, Ki67), epidermal hyperplasia (e.g. IL-19, IL-22, FGF, VEGF) and differentiation genes

(e.g. *FLG*, *LOR*) were improved to a healthy phenotype by ciclosporin and culminating in the restoration of the granular layer and downregulation of hyperplasia (Khattri *et al.*, 2014). Data from the normalised expression of a panel of barrier genes suggested that barrier remains impaired over the course of treatment.

Barrier integrity was inferred by proxy of a range of epidermal markers expected to be relevant to proper barrier function throughout the epidermis. The panel includes early differentiation markers (*KRT1*, *KRT10*), terminal differentiation markers (*FLG*, *LOR*), epidermal tight junction proteins (claudins 1 and 23), inter-keratinocyte junctional proteins (desmocollin, desmoglein and corneodesmosin). The persistent downregulation of these barrier indicators suggests that while some markers may be improved, a generalised perturbation still affects the epidermal barrier during and after ciclosporin treatment. Therefore, these results are not able to definitively link immune suppression with reversal of barrier abnormalities. In contrast, highly targeted treatment of the atopic dysregulation using dupilumab in patients with atopic dermatitis demonstrated an improvement in both a reduction of KC2 and upregulation of barrier markers.

The IL17-stimulated keratinocyte response in the atopic samples, especially in the lesional skins, is particularly interesting. In atopic lesional skin, this KC17 fraction was of a similar magnitude to that observed in psoriatic lesions. It has been previously observed that atopic keratinocytes respond to IL-17 in much the same manner as healthy keratinocytes (Nograles *et al.*, 2010). Although atopic keratinocytes do not have an abnormal IL-17 response, there is an increased number of IL-17-producing T cells in atopic lesions (Toda *et al.*, 2003). While the infiltration of polarised T cell populations may be histologically relevant, lesional expression of IL-17A, particularly, has been shown to be remarkably similar to healthy skin (Wilson *et al.*, 2007; Bernard *et al.*, 2012). Furthermore, it has been suggested that IL-17-driven skin inflammation may be dependent on synergistic action of other cytokines such as IL-22 and TNFα (Guilloteau *et al.*, 2010; Bernard *et al.*, 2012). Although not found to be statistically significant between non-lesional and chronic lesions from atopic dermatitis samples, KC17 is generally increased and distinguishes chronic lesional keratinocytes. This correlates well with understanding that progress of atopic lesions begins with a T_H2 programme that shifts to a mixed and complex $T_H1/T_H2/T_H17$ biology over the longitudinal course of a chronic lesion (Thepen *et al.*, 1996; Oyoshi *et al.*, 2009; Brandt and Sivaprasad, 2011).

KC17 fractions were not found in all atopic lesions analysed. Such samples without this KC17 fraction were found to have KC2 atopic keratinocyte responses similar to healthy samples. Despite their keratinocyte cytokine responses appearing healthy, the IL-17-negative samples still demonstrated reduced expression of a panel of barrier genes. As such, these patients potentially

have a barrier-driven, rather than an immune-driven, disease. Further complexity to this system may derive from the microbiome, where the well-described association of *S. aureus* colonisation may be an important factor in lesion formation and pathogenesis in low KC2/KC17 patients, and amplification of T_H2 immunity (Ardern-Jones *et al.*, 2007).

These results demonstrate, from a keratinocyte perspective, the cytokine shift characteristic of progression during atopic dermatitis and the underlying effect this has on pathophysiology. At early stages, including the pre-lesional state, an abnormal atopic environment exists, where non-lesional skin shows lower expression of differentiation and cornified envelope genes and desmosomal junction markers. As disease progresses to form a lesion then becoming chronic, keratinocytes experience an increase in transcriptomic IL-17 response. There is a clear heterogeneity in patients centred on KC2/KC17 fractions. Those who present with clinical lesions, but deconvolution analysis reveals low KC2 and absence of KC17 allude to as yet unexplored aspects of pathogenesis. These patients do not, however, show worse disease scoring indices than the high KC2, KC17 positive cohort.

Etanercept treatment of psoriasis targets the TNF component of the TNF-IL-17 synergising inflammation, and was shown to induce drastic downregulation in downstream T_H17 responses rather than immediate keratinocyte-derived TNFα responses (Zaba *et al.*, 2009). The longitudinal amelioration of the KC17 immunophenotype deconvoluted from lesional patient samples during etanercept treatment corroborates this finding. These deconvolution data indicate the utility of examining keratinocyte responses and immunophenotypes to the local environment, in this case, the cytokine milieu released by disease specific $CD4^+$ populations. This aims to identify the molecular contributions by keratinocytes, and related sub-populations, to pathogenesis that are induced by innate or exogenous factors to further understand cutaneous inflammation – ultimately with the goal of improved patient stratification and treatment targets.

Chapter 5 Defining bulk cutaneous microbial responses

5.1 Introduction

The microbiome is a critical component of cutaneous health. There are approximately 10^{12} cutaneous bacteria per square metre of skin, composed of a diverse microflora that closely interact with host cutaneous cells (Grice and Segre, 2011). The homeostasis of this exogenous and diverse community is managed at the interface of the skin. It involves the balancing of tolerance to colonising commensal species to take advantage the immune training they offer, and avoidance of inappropriate inflammatory responses to beneficial microbial species that inhabit niches otherwise available to pathogenic species (Otto, 2010). In addition, clearance of potential pathogens without eliciting severe cutaneous inflammation is a key function of epidermal homeostasis (Lai and Gallo, 2008). Dysfunction of these homeostatic processes leads to imbalance of the microbiome resulting in dysbiosis and the resulting potential infections and inflammation.

As the major resident population of the epidermis and at the interface with the cutaneous microbiome, keratinocytes play a central role in both homeostasis and dysbiosis. The aim of this chapter is to define keratinocyte molecular responses to *in vitro* exposure to archetypal cutaneous commensal and pathogenic species, *S. epidermidis* and *S. aureus*, respectively. This will identify the host-microbe interactions and host cell crosstalk key to microbiome homeostasis and extend to delineation of these in the context of commensal or pathogenic colonisation.

The first active cutaneous defence against pathogens is the secretion by keratinocytes of AMPs, such as hBD2 and the S100 proteins (Fulton *et al.*, 1997; Liang *et al.*, 2006). Keratinocyte detection of pathogens occurs through expression of cell surface pattern recognition receptors such as TLRs leading to the production of proinflammatory cytokines and downstream effector immune responses. *S. aureus* hypercolonisation is a marker of atopic dermatitis epidermis, and so too is overexpression of hBD2 in this disease (Kobayashi *et al.*, 2015; Hönzke *et al.*, 2016). This led to the hypothesis that bacterial sensing by keratinocytes is altered by cutaneous inflammation.

Monolayer keratinocyte cultures are useful for investigating broad molecular transcriptional responses in response to bacterial exposure in a simplified system (Secor *et al.*, 2011, 2012). However, these model systems are limited to using soluble bacteria factors, lysates or radiation inactivated bacteria to enable *in vitro* challenge without risk to the cell culture. Reconstituted epidermal models differ from monolayer keratinocyte cultures because they form layered

organisation exhibiting recapitulation of the basal proliferation and suprabasal differentiation of *in vivo* human epidermis (Duckney *et al.*, 2013; Hennies and Poumay, 2021).

Previous reports investigating maintenance of cutaneous health in the context of microbiome handling implicated the epidermal barrier as critical for this homeostasis (Duckney *et al.*, 2013). Topical application of both commensal and pathogenic *Staphylococcal spp.* did not drastically induce an inflammatory response in an *in vitro* epidermis model, whereas bacterial inoculation of culture medium simulating defective barrier resulted in widespread model cell death. Interestingly, bacterial cytotoxicity during barrier breach was associated with immediate IL-1α release but models were destroyed before later TNFα and IL-8 expression and responses could be monitored (Duckney *et al.*, 2013). Alternatively, in the barrier breach model, challenge with commensal *S. epidermidis* species indicated effects on keratinocyte differentiation and proliferation status that were not apparent from topical application indicating divergent epidermal responses to pathogen and commensal colonisation. During disrupted barrier, commensal colonisation induced barrier repair and keratinocyte differentiation responses, possibly due to the low virulence of *S. epidermidis* allowing for host-microbial accommodation after reformation of the stratum corneum; in contrast, pathogen colonisation induced cytotoxic inflammation that proved destructive to the epidermal model.

Similar topical bacterial challenge experiments were performed previously (Holbrook, 2019). This work again demonstrated the IL-1 alarmin production of *S. aureus* challenged models by presence of soluble inflammatory mediators in culture undernatant. Additionally, expression of TNFα was detected after 24 hours of culture. The expression of IL-1 family cytokines from *S. aureus* challenged models indicated the stimulation of IL-1/TLR signalling by keratinocytes resulting in the expression of pro-inflammatory cytokines, such as TNFα, and AMPs, such as hBD2, hBD3 and RNase 7 (Cohen, 2014; Holbrook, 2019). Furthermore, pathway analysis indicated keratinocyte activation of T_H17 immunity in the absence of $CD4^+$-derived IL-17, which has important implications for enhancing host defence but also potentially triggering epidermal inflammation.

The RHE models used in this project are cultured on an inert polycarbonate membrane within a free-standing insert. Initially cells are allowed to fully proliferate by submersion in culture media to achieve confluency across the membrane. The physical structure and topological differentiation achieved by this model aims to recapitulate that of *in vivo* epidermis. This allows for investigation of transcriptomic signatures that are more biologically relevant and free of artefacts compared to monolayer cultures. The advantage of the differentiated stratification of the models is the formation of the stratum corneum, which permits infection challenge with live *Staphylococcal spp.*

Therefore, the bacterial co-culture of RHE models enable us to develop keratinocyte-centric reference signatures to live *Staphylococcal spp.* interaction. These can be used to deconvolute the bulk expression profiles from the skin of healthy and inflamed donors in a similar manner to the cytokine related analysis.

5.1.1 Hypothesis

Keratinocytes have a distinct transcriptomic response to bacterial challenge that can be defined from bulk mRNA expression of *in vitro* cultures. Further, the cutaneous microbiome can regulate the cutaneous inflammatory response in the skin and, therefore, different cutaneous microbes differentially modulate gene expression of inflammatory skin disease.

5.1.2 Aims

- Investigate publicly available datasets for potential in deriving *Staphylococcal* responses by keratinocyte transcriptome.

- Re-analysis of the bulk transcriptome of a 3D model of keratinocyte challenged with live *Staphylococcal spp*.

- Derive, from this re-analysis, a bacterial sensing signature from stratified keratinocytes to apply for deconvolution of bulk skin.

- Compare the bacteria sensing profile from bulk human skin biopsies to the KC2/KC17 immunophenotypes previously described to understand cutaneous inflammation.

5.3 Results

5.3.1 Analysis of keratinocytes cultured with conditioned media from *S. aureus* demonstrate distinct transcriptomic changes

To delineate basic keratinocyte signature underpinning bacterial sensing, existing publicly available datasets of keratinocytes cultured in monolayer with conditioned media from *S. aureus* secreted factors were investigated (GSE24118, GSE24370, and GSE32920) of monolayer keratinocytes cultured with conditioned media from *S. aureus* secreted factors. These secreted factors are derived from *S. aureus* grown in either a planktonic (PCM) or biofilm state (BCM).

To assure direct comparison between the datasets was possible, datasets from Affymetrix Genome U133A 2.0 Array were used. Quality control, including reduced dimension (MDS) plots of the datasets, confirmed that both primary keratinocytes and HaCaTs are significantly altered by culture in conditioned media from both planktonic and biofilm derived secreted factors (**Figure 5.1A, B**). Importantly, while the samples from the two media conditioning types clustered away from the control samples, in both datasets, they did not cluster together. This indicates that the growth of *S. aureus* in planktonic versus biofilm conditions produces different secreted factors that induce different transcriptomic programmes in keratinocytes.

The first dimension of the HaCaTs MDS plot (**Figure 5.1C, D**) explains approximately 50% of the variance among the samples and represents the difference between sterile control media and conditioned media. Along this dimension the PCM and BCM HaCaT samples are indistinguishable, where the difference in media type is only explained at the second dimension, which represents 20% of the principal component variance. In this second dimension, sterile controls sit between the two conditioned media types.

The difference between the media conditioning treatments of primary keratinocyte cultures is explained in the first principal component, which accounts for approximately 40% of the variance (**Figure 5.1E, F**). Similar to the HaCaT culture, considered along this dimension, the control samples sit between the PCM and BCM samples. The second principal component represents approximately 25% of the total variance and demonstrates the distinction between sterile media and condition media samples.

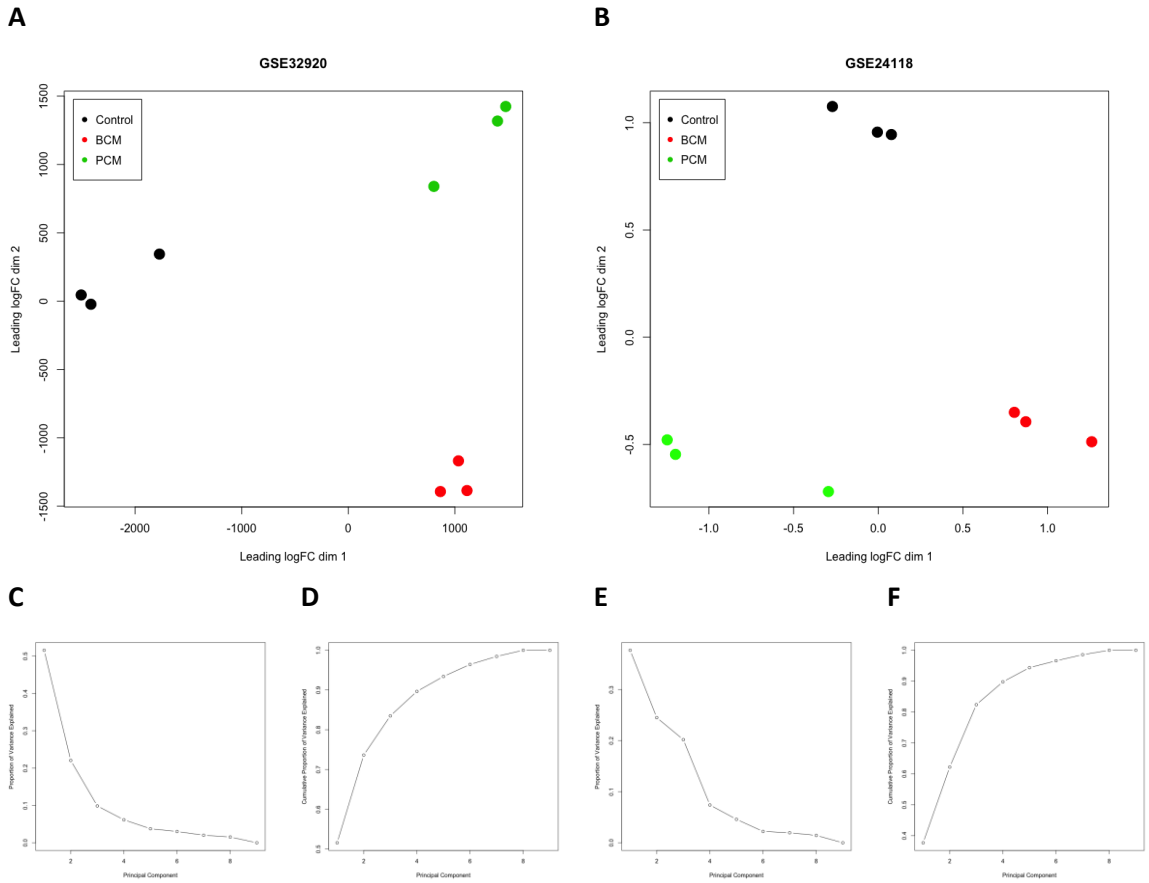


Figure 5.1. Multi-dimensional scaling of monolayer keratinocyte datasets cultured in conditioned media. Two publicly available datasets of monolayer keratinocytes cocultured with *S. aureus* biofilm conditioned media (BCM, red) or planktonic *S. aureus* conditioned media (PCM, green) or control media (black). **[A, B]** Multi-dimensional scaling (MDS) of the HaCaTs samples [A] and primary keratinocytes [B]. **[C-F]** Variance analysis explained by the top nine principal components of the HaCaTs dataset [C, D] and primary keratinocytes dataset [E, F]: individual [C, E], cumulative [D, F].

To investigate the transcriptomic changes induced by PCM and BCM in a directly comparable way, the datasets were normalised, integrated and batch-corrected using COMBAT (Leek *et al.*, 2012). Initial analysis of the changes was observed by MDS plotting (**Figure 5.2A**). In the integrated dataset, the HaCaT samples demonstrated the most divergent responses to conditioned media (red samples, **Figure 5.2A**). The difference between samples cultured in planktonic or biofilm conditioned media remained apparent. In this integrated view, the primary keratinocytes were less altered by bacteria conditioned media, despite being observed when considered as a single dataset.

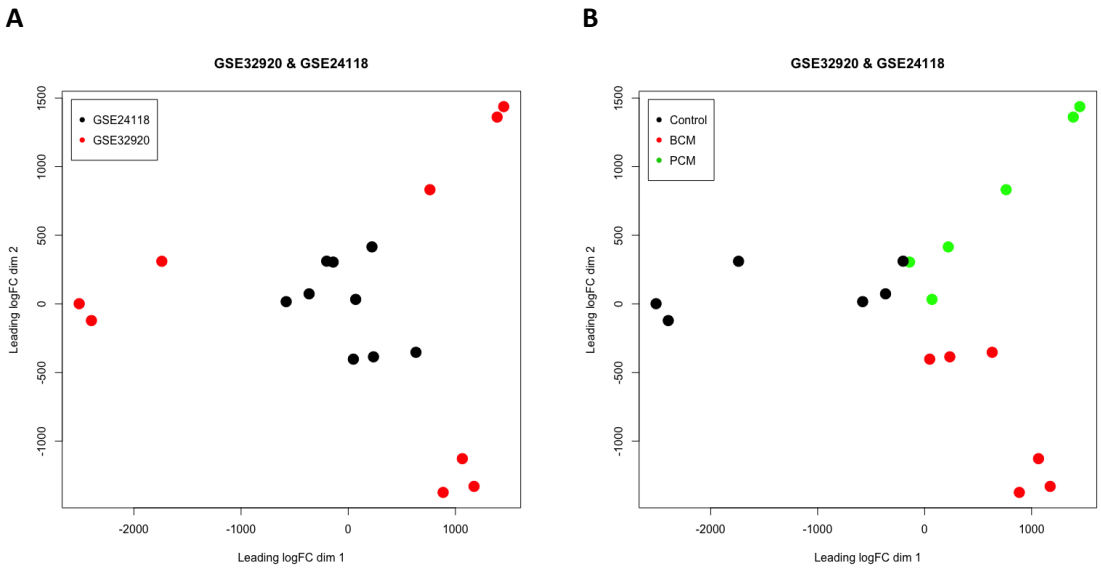


Figure 5.2. Multi-dimensional scaling of batch-corrected and integrated monolayer keratinocyte datasets cultured in conditioned media. [A] MDS plot of the two monolayer keratinocyte datasets GSE32920 (red) and GSE24118 (black). **[B]** The same MDS plot recoloured to show the media conditioning; control (black); BCM (red) and PCM (green).

From the integrated dataset by using CIBERSORT-based deconvolution, it was tested whether whole transcriptome reference signatures from HaCaT control, biofilm and planktonic media could identify those from the primary keratinocyte dataset, and vice versa. The ability of CIBERSORT to correctly profile the samples according to which media type the keratinocytes were cultured was independent of dataset, and hence the cell type providing the reference signature (**Figure 5.3**). The difference in global transcriptome of primary cells or immortalised cell lines did not affect the ability of correct identification.

Cross cell-type comparisons were considered important in defining the optimal signature source as it would be expected that references from primary cells would perform better in deconvoluting human tissue samples in later analysis. Primary keratinocytes cultures in control media were better at identifying control HaCaTs than vice versa (**Figure 5.3C**, means 86.5% versus 66.1%). Both iterations of deconvolution accurately identified BCM samples at around 75% (**Figure 5.3E**). For both datasets, the PCM samples failed to be accurately profiled. No deconvolution identified any PCM samples as composing more than 10% of its correct profile (**Figure 5.3D**). Therefore, planktonic conditioned media samples were excluded from further analysis. Furthermore, the biofilm form of bacteria is more relevant to this investigation in the context of normal cutaneous microbial interactions because the microbiome exists as a biofilm not a planktonic state.

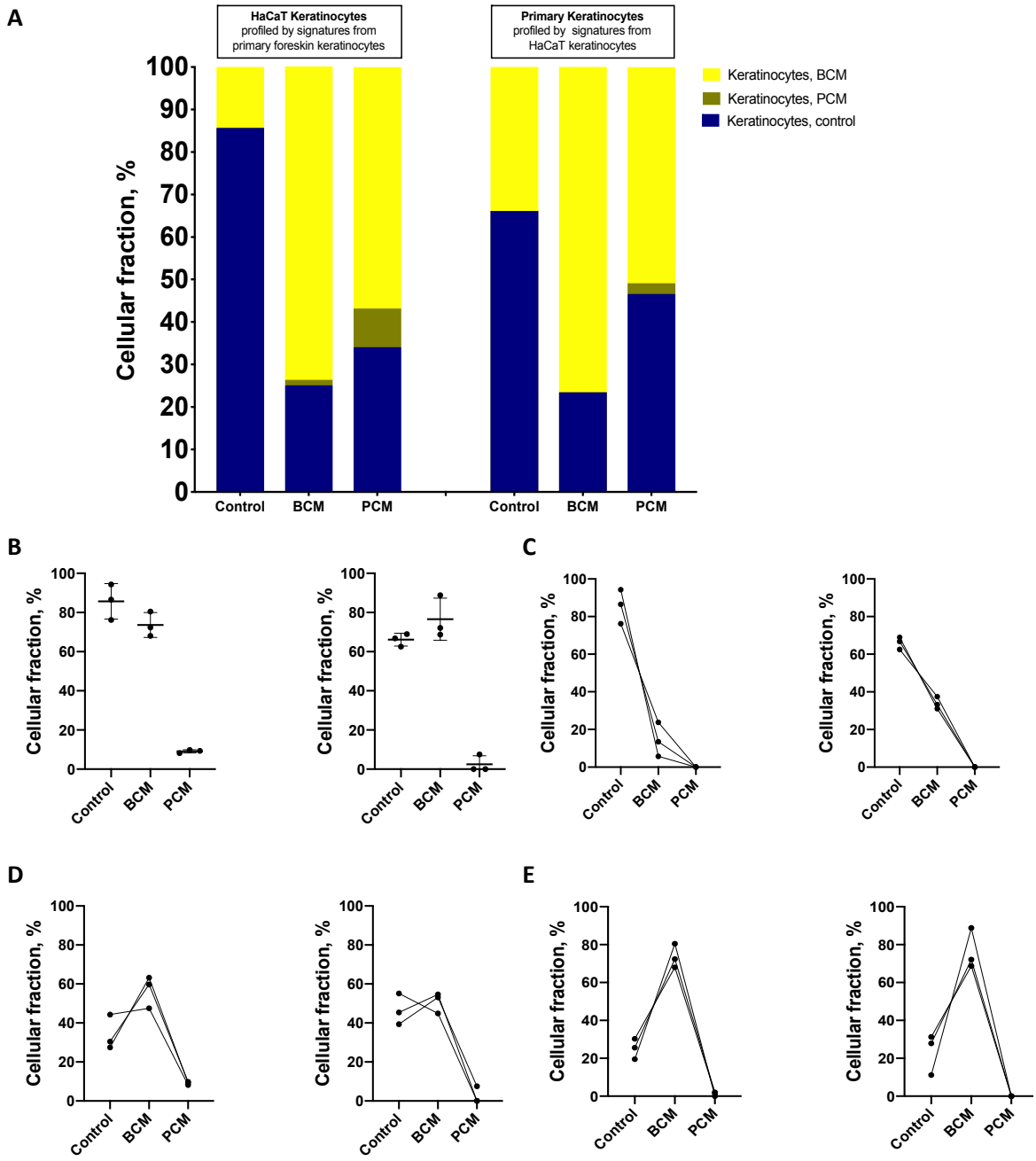


Figure 5.3. Deconvolution of HaCaT and primary keratinocyte monolayer datasets by their analogous samples from the opposite dataset. [A] Average cellular fractions output by CIBERSORT of conditions in both datasets; (left bars) HaCaTs deconvoluted by profiles from primary keratinocyte data; (right bars) Primary keratinocyte deconvoluted by profiles from HaCaT data. **[B-E]** Individual deconvolution results from HaCaT (left panel) or primary keratinocytes (right panel) **[B]** Per centage of each sample by its correct, cognate identifying signature from the analogous dataset. **[C]** Per centage of the control signature from primary keratinocytes found in HaCaT samples (left) and found in primary keratinocytes from HaCaT references. **[D]** As C, showing the planktonic resolved profiles. **[E]** As C, showing the biofilm resolved profiles.

Having identified which cell culture conditions induced the strongest transcriptomic change in keratinocytes to be reliably used for *in silico* analysis, the next aim was to define the genes associated with these changes. For this, a third dataset, GSE24370, involving media conditioning but performed on a different array technology as the other datasets was also considered. As media conditions were compared only within each dataset and then subsequent genes lists compared between datasets, this approach does not require integration of the three datasets. Standard differential expression analysis using LIMMA was performed independently for each dataset, comparing the control to the BCM samples. Between 308 and 888 genes were found to be up- or down-regulated above the threshold of statistical significance (FDR BH p-value <0.05, logFC >±1.5) in keratinocytes grown in BCM. Gene ontology analysis on these gene lists revealed relevant biological processes, which were summarised to remove redundant terms using REVIGO (**Figure 5.4**).

REVIGO identified three main biological process terms associated with upregulated genes from BCM grown primary keratinocytes: response to lipopolysaccharide (GO:0032496, FDR BH p-value = $2.47E^{-12}$); negative regulation of nucleotide-containing compound metabolism (GO:0045934, FDR BH p-value = $2.44E^{-9}$); haemopoiesis (GO:0030097, FDR BH p-value = $9.74E^{-7}$) (**Figure 5.4A**). In addition, other top GO terms were highly related to microbial responses: response to molecule of bacterial origin (GO:0002237, FDR BH p-value = $2.47E^{-12}$); inflammatory response (GO: 0006954, FDR BH p-value = $3.82E^{-10}$); response to bacterium (GO:0009617, FDR BH p-value = $1.29E^{-9}$).

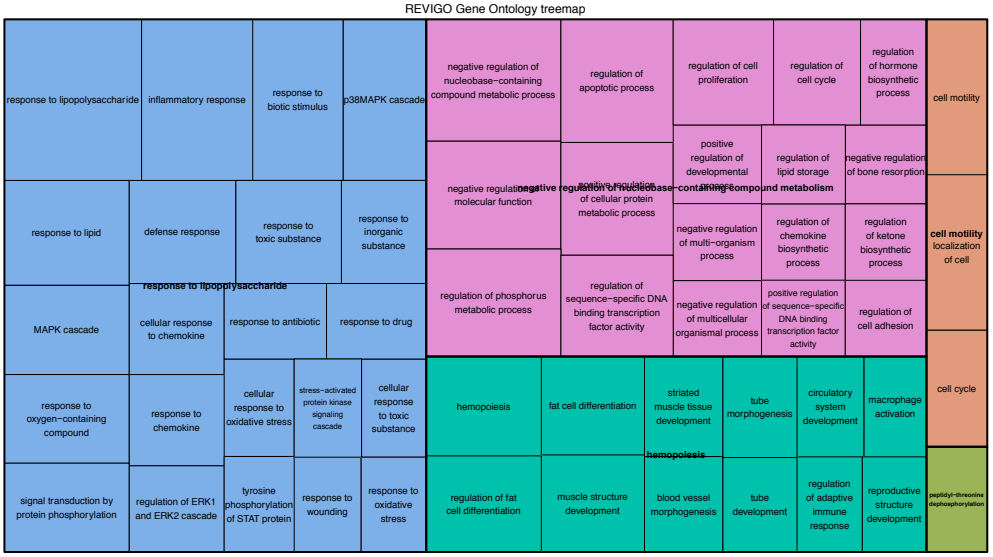
The interrogation of the HaCaT dataset, which yield 469 genes upregulated in the BCM samples, showed five main summarised terms from REVIGO (**Figure 5.4B**). These were not immediately related to bacterial responses but generalised keratinocyte responses to microbial stimulation. Generalised annotations for epithelial or skin processes and immune responses such as epithelium development, skin development, response to cytokine, inflammatory response, response to lipopolysaccharide, lymphocyte activation and T cell activation were found among the top 200 annotations. These indicate a weaker but potential bacterial response profile in these cells among the larger DEG list in this comparison.

The third microarray dataset (GSE24370), also from primary keratinocytes, but performed on a two-colour array, confirmed the findings of the other primary cell experiment. This dataset only included a *S. aureus* biofilm arm along with control and did not include a planktonic condition. The SAB grown samples had 133 genes upregulated compared to controls. Response to lipopolysaccharide was again the top REVIGO summarised biological process annotation (GO:0032496, FDR BH p-value = $2.16E^{-16}$). This was followed by a further two main summary terms: localisation of cell (GO: 0051674,

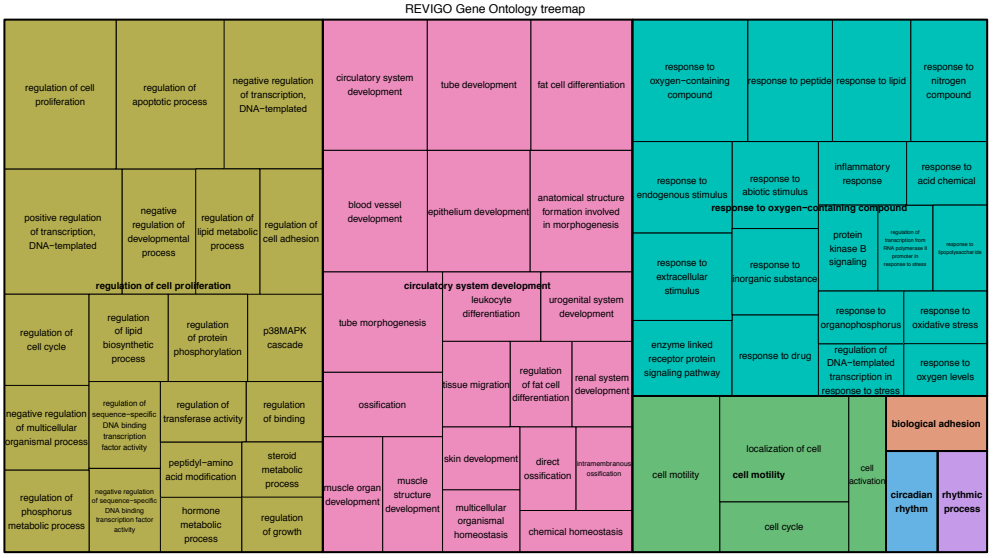
FDR BH p-value = $5.96E^{-10}$) and regulation of cell proliferation (GO: 0042127, FDR BH p-value = $2.21E^{-9}$). The high-level term 'response to lipopolysaccharide' as summarised by REVIGO, in this case included many annotations highly relevant to bacterial responses: cytokine-mediated signalling pathway (GO: 0019221, FDR BH p-value = $2.16E^{-16}$); cellular response to lipopolysaccharide (GO: 0071222, FDR BH p-value = $2.16E^{-16}$); response to molecule of bacterial origin (GO: 0002237, FDR BH p-value = $2.21E^{-16}$); cellular response to molecule of bacterial origin (GO: 0071219, FDR BH p-value = $2.68E^{-16}$) (**Figure 5.4C**).

All three monolayer datasets presented convincing bacterial response signatures from keratinocytes when interrogating the biofilm upregulated genes. However, despite the overlap in the gene ontologies for biological processes, comparing the gene lists, identified only 13 genes were common to all three analyses: *ATF3*, *BMP2*, *CXCL1*, *CXCL2*, *CXCL3*, *ERCC6L*, *GEM*, *HBEGF*, *HES1*, *IL6*, *PPP1R15A*, *TNFAIP3*, *VEGFA*. A further six genes were also common between the primary keratinocyte datasets: *CCL20*, *CDKN1C*, *CSF2*, *DTL*, *IFIT1*, *S100P* (**Figure 5.5**). Comparing these to the top 20 DEGs generated from investigating the RHE models infected by SA83 or SE12 at 3 and 24 hours ((Holbrook, 2019)), *HBEGF* was the only common gene to the monolayer and RHE *S. aureus* samples. Seven of the 19 mutual DEGs from the monolayer datasets were found among the top 20 DEGs from the *S. epidermidis* infected models: *CCL20*, *CSF2*, *CXCL1*, *CXCL2*, *CXCL3*, *IL6* and *TNFAIP3*. *S. epidermidis* was found to be a strong inducer of *TNFAIP3* in the RHE models and is upregulated by *S. aureus* secreted factors in the context of monolayer keratinocyte cultures.

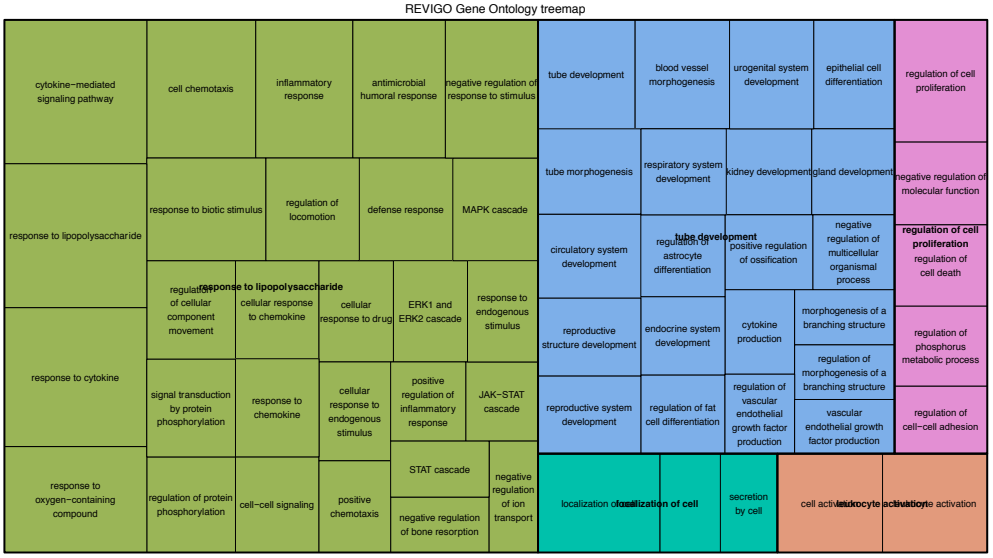
A



B



C



[*Legend follows on next page*]

[*Legend for figure on previous page*]

Figure 5.4 REVIGO tree map summarising the top significant biological process gene ontology terms in the three public datasets. [A] Summary of the top terms associated with the genes upregulated in BCM samples in primary keratinocytes, GSE24118. **[B]** Summary of the top terms associated with the genes upregulated in BCM samples in HaCaT keratinocytes, GSE32920. **[C]** Summary of the top terms associated with the genes upregulated in BCM samples in primary keratinocytes, GSE24370.

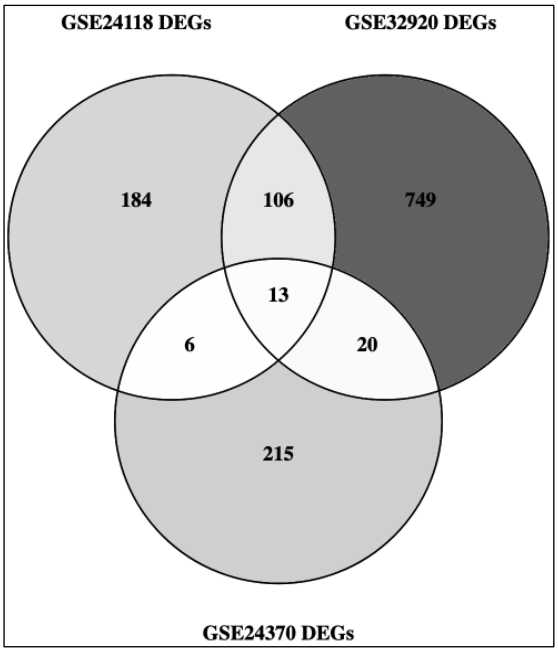


Figure 5.5. Venn diagram of differentially expressed genes found in the public monolayer datasets. DEGs comparing control and BCM samples across the three public datasets of monolayer keratinocytes (GSE24118, GSE32920 and GSE24370) were compared for gene list overlap(s).

5.3.2 Optimising bacteria sensing signatures from monolayer keratinocytes for computational deconvolution

The three monolayer keratinocyte microarray datasets displayed varied lists of differentially expressed genes in response to culture with secreted factors from *S. aureus* biofilm. Global transcriptome analysis of the samples revealed a preserved but different in magnitude response to bacterial product challenge when observed by MDS plotting (**Figure 5.2B**). Computational analysis, relying on the either the biofilm conditioned media upregulated genes or all differentially expressed

genes in CIBERSORT, was used to resolve the samples against the defined reference signatures. This analysis would determine whether the identified differentially expressed genes and the expression of these performs well enough to identify like-for-like bacteria challenged samples.

Each dataset was used to create two different deconvolution panels, each composed of a sterile control and a biofilm stimulated signature. The first panel included genes only found to be upregulated by *S. aureus* biofilm conditions, while the second signature was composed of all genes found to be differentially expressed. A total of six deconvolution panels was created. To perform an internal control, the ability of these panels to correctly identify the samples from the same datasets was assessed. Hence, this internal validation is performing a deconvolution of the samples using the reference panel derived from the same dataset. The rationale of this is the quantification of the fidelity of the signatures against cognate samples, to determine which gene list provides the optimal signature and to add robust validation when used against unseen samples from the other datasets.

The internal validation for genes upregulated from a *S. aureus* media environment (**Figure 5.6A, E, & I**) show that the signatures accurately identify their derivative signatures. Control samples in all three datasets were identified to an accuracy exceeding 95%. The BCM samples were less strictly defined. Of the ten samples across the three datasets, only three samples are identified with an accuracy above 95%, while two samples were found to be composed of less than 80% of their correct BCM signature. The best performing internal validation was the signatures and mixtures derived from the HaCaT dataset, GSE32920, where the reference signatures correctly identified all samples with an accuracy above 95% (**Figure 5.6E**).

The internal validation of the *S. aureus* upregulated differentially expressed genes was expanded to cross-deconvolute the other samples from the remaining two analogous datasets. The signatures derived from the primary keratinocytes of GSE24118 (Signature A) was unable to identify the unseen BCM samples of the other two datasets (**Figure 5.6B, C**).

The HaCaT derived signatures (Signature B) performed better in deconvoluting the unseen samples of the two primary cell datasets (**Figure 5.6D, F**). The signatures correctly identified the control samples (>80%) of GSE24118 and the BCM samples of GSE24370 (100%). Deconvolution accuracy of the BCM samples and sterile control samples of GSE24118 and GSE24370, respectively, was less accurate. In these cases, the incorrect signature composed between 20-45% of the samples.

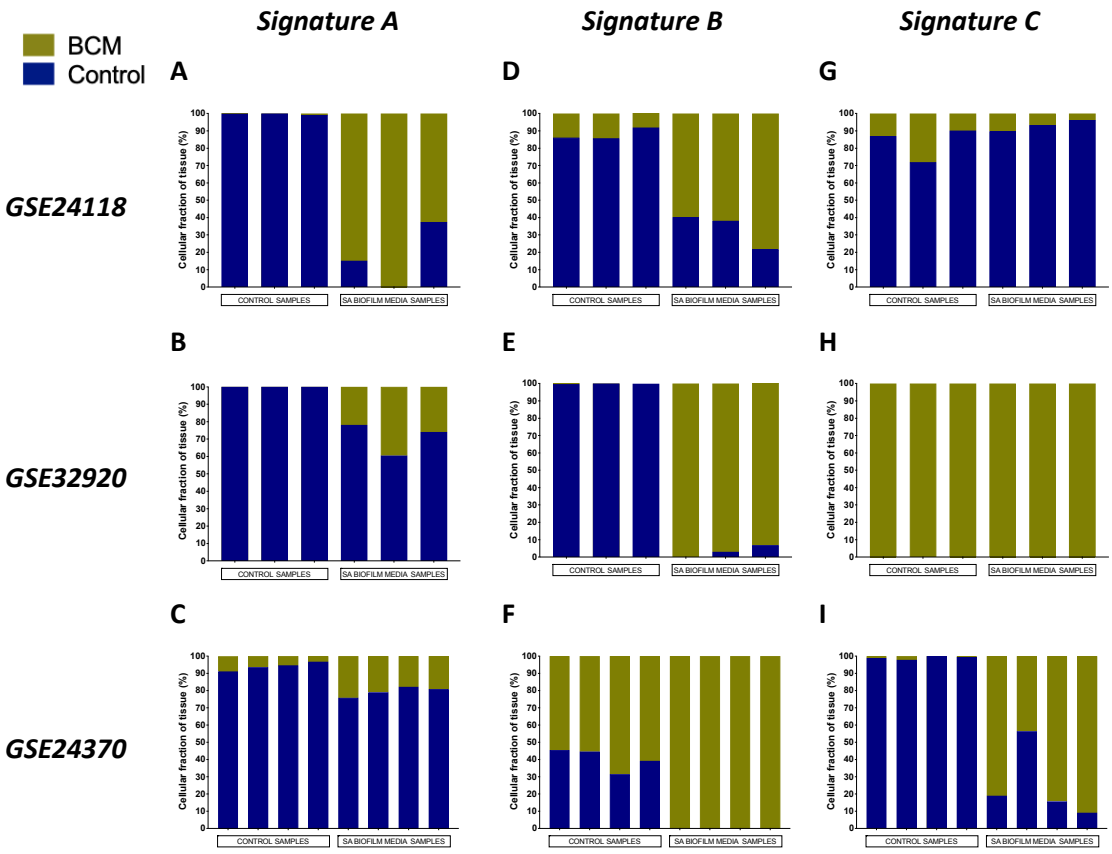


Figure 5.6. Deconvolution results from validating the signatures derived from the three separate BCM challenge datasets using biofilm upregulated genes. Each of the samples from GSE24118 (top panel), GSE32920 (middle panel) and GSE24370 (bottom panel) were deconvoluted against control and BCM signatures curated from DEGs upregulated in BCM samples of GSE24118 (Signature A – 68 genes), GSE32920 (Signature B – 469 genes) and GSE24370 (Signature C – 133 genes).

Similar to the other primary keratinocyte-based signatures, the signatures sourced from GSE24370 (Signature C) did not identify the BCM samples from GSE24118 (**Figure 5.6G**), nor identify the sterile control samples from the HaCaT dataset, GSE32920 (**Figure 5.6H**).

Expanding this validation to include all differentially expressed genes of the sterile control and BCM samples comparison would be expected improve deconvolution resolution. Therefore, the genes upregulated in the sterile control samples were added to the deconvolution reference panels. The samples from all three datasets were again deconvoluted for accurate identification of media condition using this expanded gene profile (**Figure 5.7**).

For the signatures derived from GSE24118, increasing the number of genes included greatly improved the deconvolution accuracy for BCM cultured samples (Signature A). The internal validation of this signature proved near 100% deconvolution accuracy (**Figure 5.7A**). Previously, the unseen BCM samples from the other datasets were highly composed of the control signature, but this was improved to an accuracy of above 70% in the HaCaT samples (**Figure 5.7B**) and above 80% in the other primary cell samples (**Figure 5.7C**). Across the ten BCM samples over the three datasets, deconvolution accuracy improved from a mean of 40% to a mean accuracy of 85%. There was a slight reduction in accuracy identifying the sterile control samples from above 95% to above 80%. However, this was affected by the poor deconvolution of the control samples from GSE24370, which were found to only be up to 60% composed of their correct profile (**Figure 5.7C**). The control samples from the other two datasets were identified accurately.

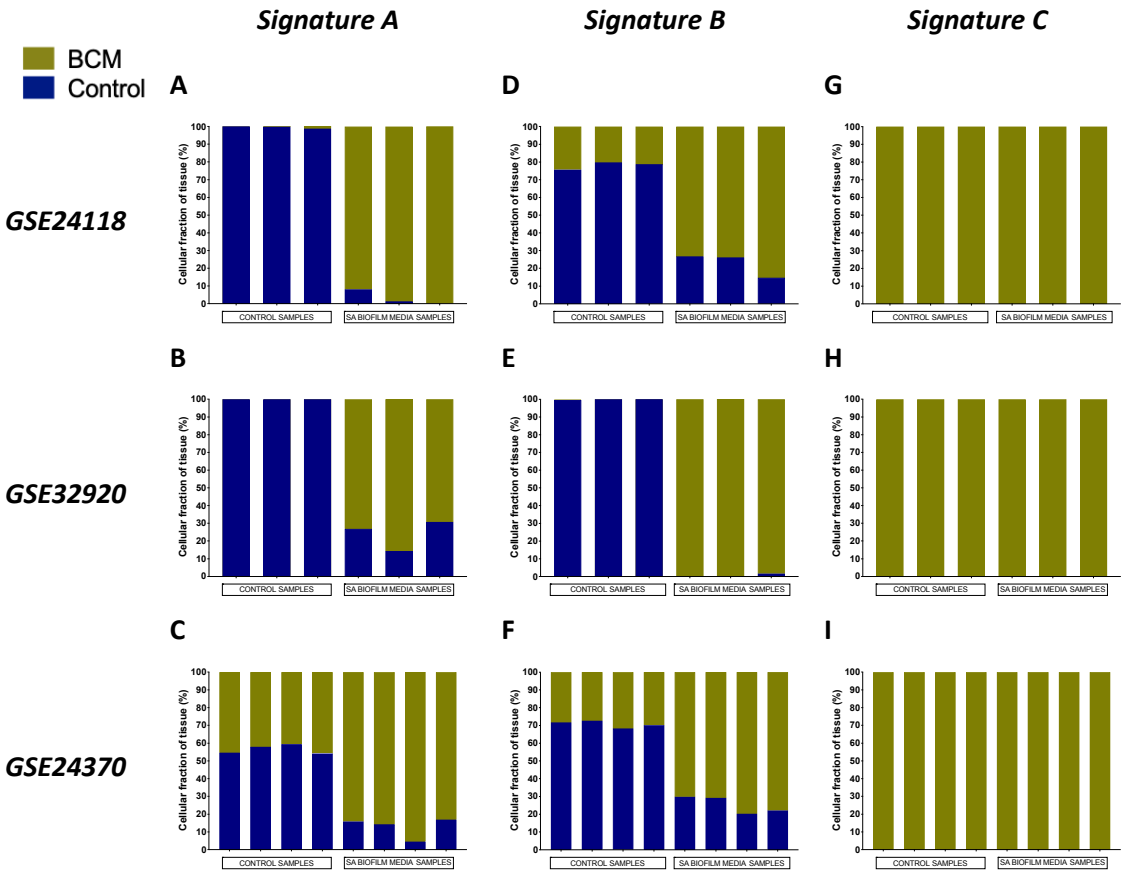


Figure 5.7. Deconvolution results from validating the signatures derived from the three separate BCM challenge datasets using all differentially expressed genes. Each of the samples from GSE24118 (top panel), GSE32920 (middle panel) and GSE24370 (bottom panel) were deconvoluted against control and BCM signatures curated from DEGs upregulated in BCM and control samples of GSE24118 (Signature A – 309 genes), GSE32920 (Signature B – 888 genes) and GSE24370 (Signature C – 254 genes).

The major changes of the second, expanded reference panels by inclusion of all differentially expressed genes occurred in slightly improving the accurate identification of the BCM samples of GSE24370 using the reference panel derived from GSE24118 (**Figure 5.7 C**). Alternatively, this decreased the resolution for the sterile samples within the same dataset, which fell below the 70% accuracy rate determined to be acceptable.

Using all differentially expressed genes for curating a signature from GSE32920 samples (Signature B) did not alter deconvolution results for primary cells (**Figure 5.7D**). The internal validation samples for this signature were deconvoluted almost perfectly (**Figure 5.7E**). A marginal improvement was noticed in the identification of the control samples of GSE24370 using this signature (**Figure 5.7F**). This came at the cost of decreasing identification accuracy of the BCM samples, but crucially this did not fall below 70%.

Interestingly, the expanded gene signature derived from GSE24370 (Signature C) completely overfit the BCM signature to all the samples, including the sterile controls from all three datasets being composed of 100% BCM exposed keratinocytes (**Figure 5.7G, H, I**). This was also true for the internal control of this signature set (**Figure 5.7I**) and, therefore, these signatures were discounted as unreliable for further bacteria sensing deconvolution.

The efficacy of the deconvolution approach is confirmed by the high deconvolution accuracy for identifying control and BCM samples when performing internal validation. Signatures derived from GSE24370 samples showed a tendency to over-fit, which was not explained by use of a smaller signature or specific to certain samples. As such, signatures from this dataset were not to be considered further. The enlarged signature composed of all differentially expressed genes improved deconvolution accuracy for both primary cell and HaCaT samples, but the primary signature proved marginally better in identifying HaCaT controls than vice versa. For this reason, and because later samples for deconvolution would be primary cells, the primary keratinocyte signatures were selected for downstream analysis.

5.3.3 Monolayer keratinocytes cultured with *S. aureus* biofilm conditioned media as a reference to deconvolute cellular fractions from epidermal models

In the previous section monolayer keratinocyte cultures grown with media supplemented with the extracellular proteome of *S. aureus* biofilm were investigated for strength and fidelity of a bacteria sensing signature. Genes differentially expressed from biofilm factors supplementation and their gene ontologies revealed annotations related to infection and inflammatory processes. This was

true for both cultures of primary cells and immortalised HaCaT lines. The expression of the differentially expressed genes was used to curate cell type reference signatures for input into the deconvolution pipeline.

The next aim was to compare the monolayer primary keratinocyte derived infection signature to the bulk expression data from RHE models infected with either *S. aureus* or *S. epidermidis.* This analysis was restricted to the RHE data from control and single *Staphylococcus* species challenge only but across the whole 24-hour time course; with samples taken at time zero, three hours and 24 hours. These selected samples were deconvoluted against the sterile control and BCM cultured monolayer cell profiles using CIBERSORT.

Surprisingly, the sterile media control RHE sample comprised both control and BCM monolayer signatures (**Figure 5.8**). This can reflect the increased cellular complexity of the RHE models, the recapitulation of proliferation and differentiation is these models and the potential impact of bacterial challenge to induce differentiation. All six *S. epidermidis* infected samples across the three and 24-hour timepoints were found to be composed entirely of the BCM profile. Within *Staphylococcal sp.* challenge, there was no significant difference in the average size of the BCM fraction when comparing between the three hour and 24-hour timepoint.

The 24-hour samples defined the timepoint at which persistent infection is modelled, and the transcriptomic changes induced by bacterial challenge become stable. The stabilisation of transcriptome after challenge over 24 hours was previously reported in the context of IL-1 stimulation and MyD88-dependent signalling (Swindell *et al.*, 2018). Comparing the three *Staphylococcal spp.* challenges of the RHE models at 24 hours to control models at the same timepoint showed *S. aureus* and *S. capitis* infected models were not significantly altered in composition of the deconvolution panel compared to control. Samples infected with *S. epidermidis* were significantly different to controls in size of BCM fraction at both three hours and 24 hours (p-values = 0.0117 and 0.0236).

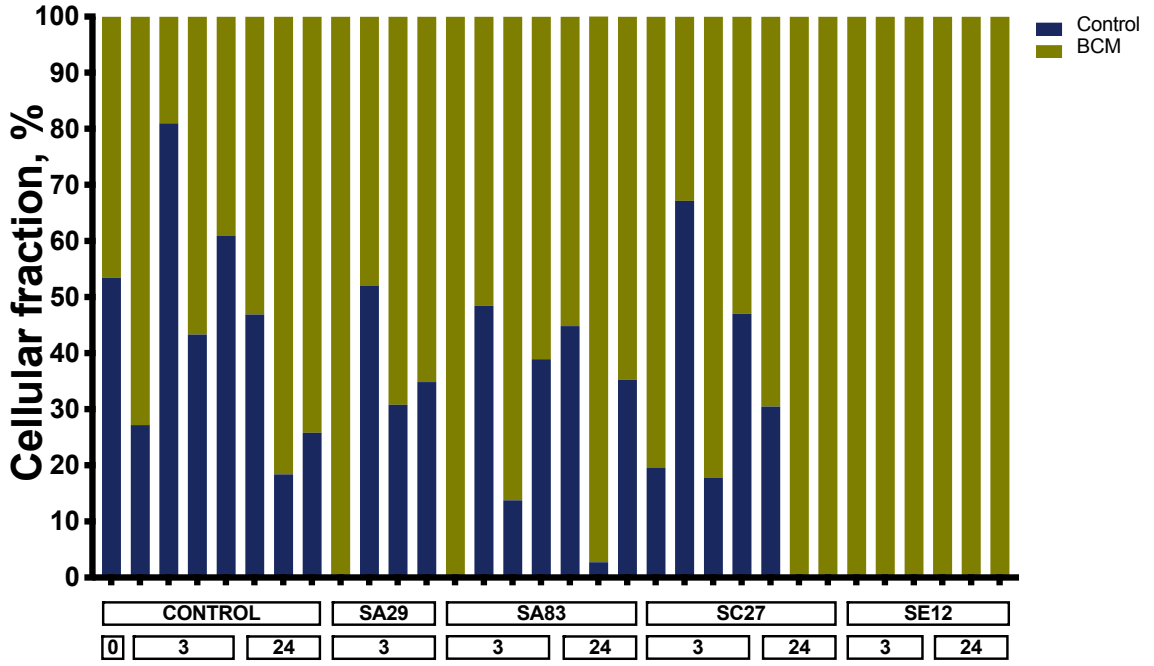


Figure 5.8. Deconvolution of bulk transcriptome of stratified reconstituted human epidermis (RHE) models using monolayer keratinocyte signatures. The mature RHE samples were cultured for three or 24 hours with single species challenge (*S. aureus*, SA29 and SA83; *S. capitis*, SC27; *S. epidermidis*, SE12) and deconvoluted by CIBERSORT using monolater control keratinocyte (blue) and BCM (olive green) signatures.

5.3.4 Keratinocyte cytokine-induced signatures can be used to deconvolute immunophenotype from RHE models

The optimised deconvolution panel for determining keratinocyte immunophenotypes from Section 4.2.2 can be used to compare and contrast the transcriptomic alterations induced by challenge with *Staphylococcus spp*. The aim of this is to identify shared and signal-specific transcriptomic programmes between cytokine induced and microbial induced responses in keratinocytes. In order to investigate the system in the context of pathogenic and commensal species, only the RHE samples challenged with a single staphylococcal species, *S. aureus* or *S. epidermidis,* were considered.

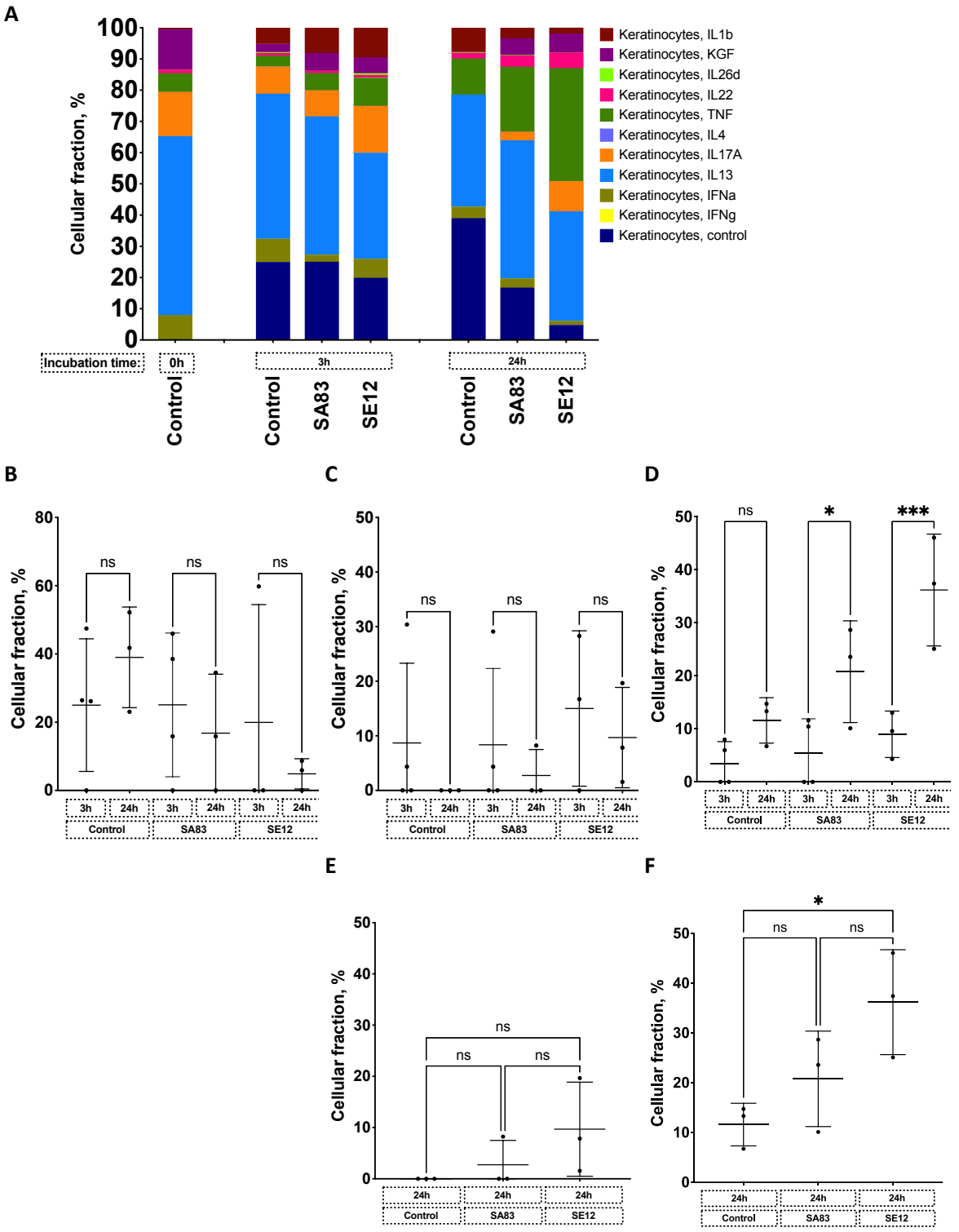
The cytokine response profiles in the RHE models infected with *S. aureus* or *S. epidermidis* were investigated next. This deconvolution demonstrates the changing immunophenotypes induced by different bacterial challenge and their shift over the time course (**Figure 5.9A**). The control, TNFα and IL-17 signatures were observed to be the most changed by challenge species and time course. While there was no unstimulated control fraction in the uninfected control RHE sample at zero

hours, this profile was consistent across the three-hour samples, regardless of bacterial challenge (**Figure 5.9B**). However, the change in the control fraction for each challenge species did not alter when comparing the three-hour time point with 24 hours (**Figure 5.9B**).

Similarly, KC17 fractions did not change between time points within challenge conditions (**Figure 5.9C**). The IL-17 responding fractions across timepoints was impacted by high variation. As such the average IL-17 fraction seen in Figure 5.9A is not representative of the change in IL-17 fractions either over time or bacterial challenge. Many of the samples indeed resolved no IL-17 responding fraction and of those that did, the expression proved high enough to impact the group average.

Comparing across the three to 24-hour timepoints by bacterial challenge, the only cytokine response fraction to increase markedly over time was TNFα (**Figure 5.9D**). While there was a general trend of increase in the TNFα fractions over the time course, this was not significant for control (p-values = 0.3696). The TNFα fraction was significantly increased by both *S. aureus* and *S. epidermidis* infections at 24 hours compared to three hours (p-value = 0.0328 and 0.0007, respectively) (**Figure 5.9D**).

The outlier effect on the KC17 fraction in the models is again visible when comparing samples only at 24-hours. There was no significant difference in comparing this fraction at this time point (**Figure 5.9E**). The TNFα responding fraction in *S. epidermidis* challenged models at 24 hours remained significant when compared to control models at the same timepoint (p=0.0295) (**Figure 5.9F**). There was no significant difference comparing *S. aureus* challenged models and control and comparing *S. aureus* to *S. epidermidis* challenged models indicating a weak TNFα response. This high fraction of TNFα response in *S. epidermis* challenged models at 24-hours indicates that the stimulation of this immunophenotype in keratinocytes is dominated by *S. epidermidis* challenge rather than model culture or experimental procedure between baseline and end timepoint.



[*Legend follows on next page*]

[*Legend for figure on previous page*]

Figure 5.9. CIBERSORT deconvolution of RHE models challenged with *Staphylococcal spp.* Keratinocyte insert cultures were incubated with *S. aureus* (SA83), *S. capitis* (SC27), or *S. epidermidis* (SE12) for 3h or 24h. Bulk microarray analysis was then deconvoluted into reference cellular fractions by CIBERSORT using the 11 cyctokine-stimulate keratinocyte profiles: IL-1β, KGF, IL-26d, TNFα, IL-4, IL-17A, IL-13, IFNα, IFNγ. **[A]** The average of each fraction in each challenge condition and timepoint. **[B-D]** Comparing temporal change in fractions by bacterial challenge: **[B]** unstimulated profile, **[C]** IL-17 profile , **[D]** TNF profile. Comparing size of **[E]** IL-17 and **[F]** TNF fractions between sterile models and SA, SE infected models at 24 hours. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean ±SD.

5.3.5 Keratinocytes challenged with live *S. epidermidis* for 24 hours have a distinct transcriptomic response to bacteria

The *Staphylococcal* challenged stratified-model keratinocyte data also provides the opportunity to derive signatures of keratinocyte transcriptome responses to specific bacterial stimulation. The models have the advantage over monolayer cultures by recapitulating the stratification characteristic of human epidermis. This would enable the deriving of reference signatures of bacteria-stimulated keratinocytes accounting for the different epidermal compartments, which could then be used as a set of reference signatures for CIBERSORT-led deconvolution of bulk skin.
The response of keratinocytes to commensal *S. epidermidis,* and pathogenic *S. aureus* is a major focus of this work. As done for the cytokine-stimulated keratinocyte data, the aim was to derive these bacteria-stimulated signatures using the whole transcriptome. Plotting the samples using MDS analysis shows that there is no timepoint nor species-linked clustering in this data, that is, the samples do not cluster by bacterial species co-culture and/or timepoint (**Figure 5.10**).

Plotting the samples by MDS showed two clusters of samples, being separated by the second principal component (**Figure 5.10**, black dotted line). The vast majority of the samples form one heterogeneous grouping. There is a separate cluster of four samples mostly derived from *S. epidermidis* 24-hour co-culture. These two clusters were compared by differential expression analysis using LIMMA. In order to retain only genes expected to share biological function and therefore maximise gene ontology significance, transcript-to-transcript correlation analysis was

used to identify co-expressed differentially expressed genes. This identified 3668 genes contrasting the two groups filtered down to 2457 genes using co-expression analysis at a Pearson R coefficient of 0.9. Filtering in this way strengthens significance for the gene ontology terms due to the reduction of input gene lists while retaining co-expressed genes.

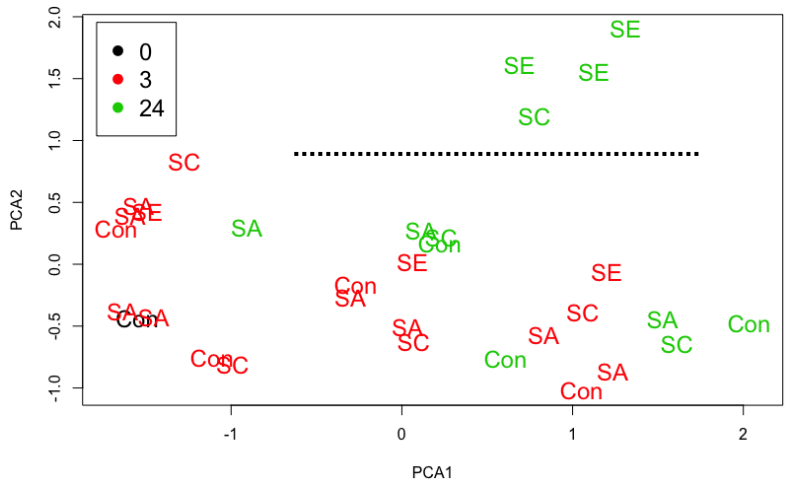


Figure 5.10. MDS plot of RHE bacterial challenge models. Multi-dimensional scaling (MDS) plot of the *Staphylococcal sp.* co-culture reconstituted human epidermis models: *S. aureus* (SA), *S. capitis* (SC), *S. epidermidis* (SE) across timepoints of three (red) and 24 hours (green). The whole transcriptome of the microarray dataset has been reduced to two dimensions representing the first two principal components: x-axis, PC1; y-axis, PC2. The black dotted line indicates two clusters of samples, being separated by the second principal component.

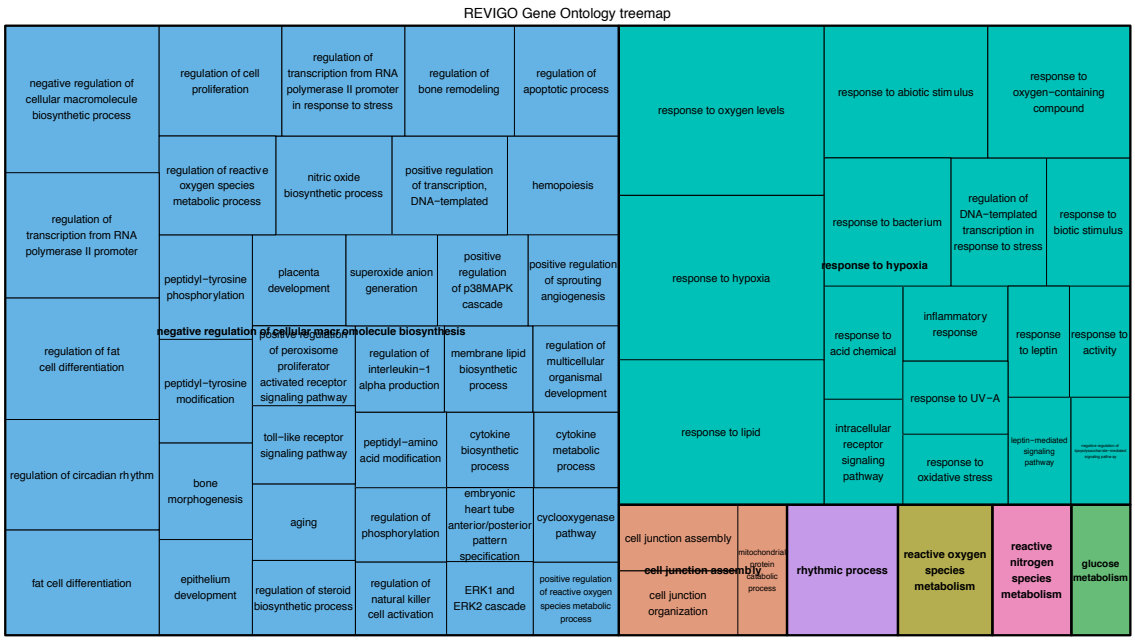


Figure 5.11. REVIGO tree map summarising the top significant biological process gene ontology terms for genes upregulated in the SE24h cluster versus the rest of the samples.

Co-expression analysis identified a cluster of 678 genes upregulated in RHE samples co-cultured with *S. epidermidis* for 24 hours (SE24h). Gene ontology annotation of the top 200 associated processes was collapsed to summarise the biological process results in a sematic similarity-based REVIGO treemaps (Supek *et al.*, 2011) (**Figure 5.11**). The visualisation tree map of the biological process gene ontology terms produced from ToppGene and collapsed by REVIGO gave a total of seven summarised ontologies and showed over-representation of two ontologies: negative regulation of cellular macromolecule biosynthesis (GO:2000113, FDR BH p-value = $1.75E^{-5}$), and response to hypoxia (GO:0001666, FDR BH p-value = $8.13E^{-8}$). Interestingly, 1% of the weighted ontologies related to cell junction assembly (GO: 0034329, FDR BH p-value = $2.22E^{-2}$). It has previously been shown that improving keratinocyte cell-to-cell contacts to tighten barrier function is upregulated with bacterial treatment, including on *S. epidermidis* culture (Duckney *et al.*, 2013).

Response to lipopolysaccharide (GO:0032496, FDR BH p-value = $1.32E^{-4}$) was the second ranked biological process term during gene ontology investigation. This annotated came from 27 of the genes upregulated in the SE24h cluster. Among these genes, many are related to cutaneous inflammation or activation, such as *CD86*, *CSF2*, *CXCL1*, *CXCL2*, *ICAM1*, *IL24*, *IRAK2*, *PPARD*, *TNF*, *TNFAIP3* (Tan *et al.*, 2007). The signature derived from this dataset had both a descriptive bacterial sensing annotation and encompassed the epidermis-like stratification making it applicable for deconvolution of bulk skin samples.

5.3.6 Deconvolution of whole skin datasets to investigate microbial responses and inflammation status

The bulk skin atopic dermatitis datasets were re-analysed using the new set of reference cutaneous signatures reflecting keratinocyte responses to *Staphylococcal* species. The nine keratinocyte signatures derived from cytokine stimulation were replaced with two signatures based on the clustering shown in Figure 5.10 and described in section 5.3.5. These clusters produced signatures defined as “unstimulated” and “SE24h”. The remaining references were retained such that this second reference set contained a total of 10 signatures to allow for whole skin analysis (**Figure 5.12A**).

Across the tissue types assessed, the cumulative proportion of the two keratinocyte fractions was found to be a mean of 58.8% of the whole tissue. The minimum keratinocyte compartment was 35.8% while the maximum was 74.3%. Therefore, by using different deconvolution panels on the same samples, the same total relative proportion of keratinocytes was estimated. This corresponds to the total proportion cytokine stimulated keratinocytes deconvoluted from the same samples in Section 4.2.3, (**Figure 5.13**). However, the correlation between keratinocyte compartments

identified from healthy skin samples using the two deconvolution panels was poor, likely indicating the inclusion of a phenotype in one panel that is absent and unaccounted for in the other panel leading to underfit of the keratinocyte compartment. Including fibroblasts, the fraction of the tissue account for is 80-90%, which is what is expected of the composition of cells from whole skin (Philippeos *et al.*, 2018; Tabib *et al.*, 2018; Korosec *et al.*, 2019).

The fractions of the two keratinocyte signatures remains remarkably stable across all tissue types, regardless of disease involvement (**Figure 5.12B, C**). However, there was a significant range in the bacteria sensing fraction within each tissue type; between the lowest and highest fraction for each tissue there was an average range of 23% (**Figure 5.12B**). The heterogeneity of this fraction within tissue types instead of between tissue types could indicate underlying patterns with other phenotypes. Therefore, correlations were investigated between the bacterial signature within tissue types against the cytokine response fractions, particularly KC2 (IL-4/IL13) and KC17 (IL-17); and against the expression of the panel of defined barrier-related genes.

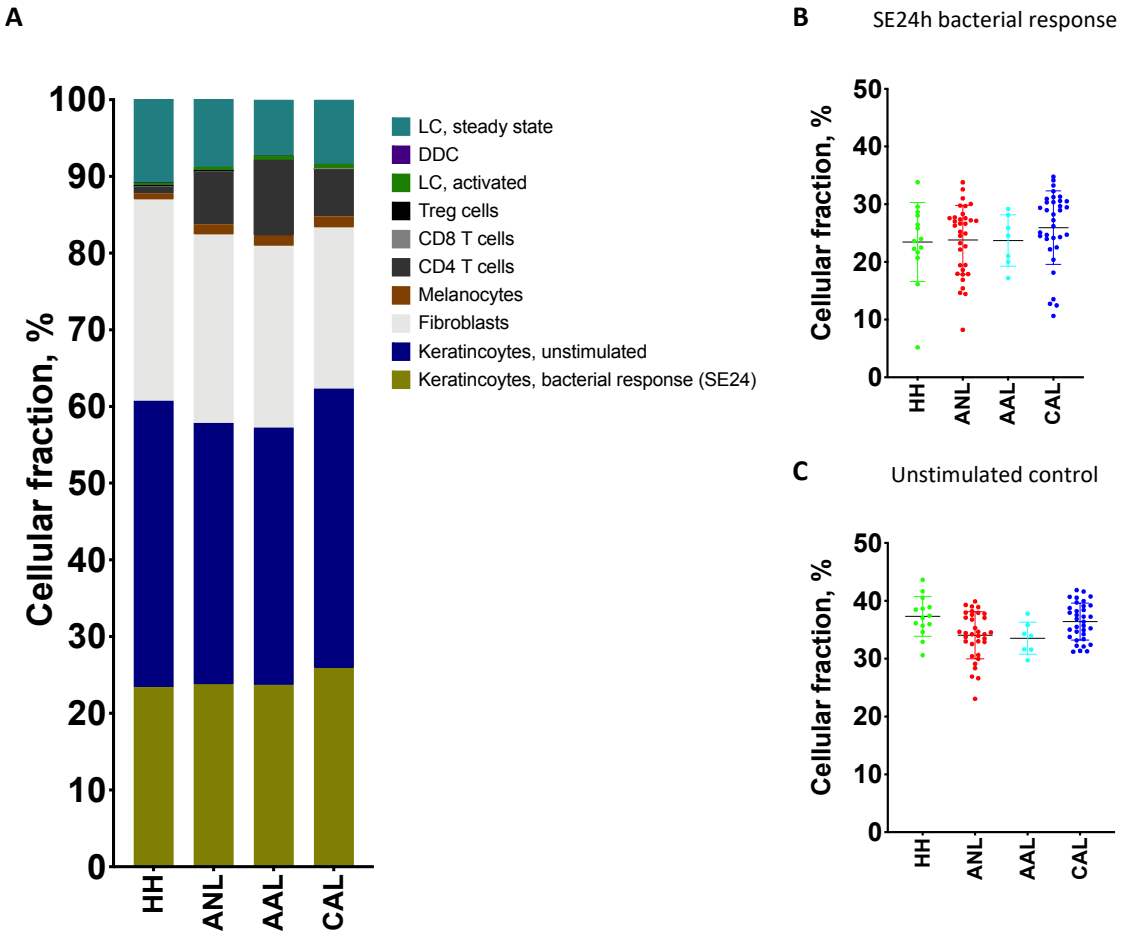


Figure 5.12. CIBERSORT deconvolution of healthy and AD skin resolving keratinocyte bacterial sensing fractions. Microarray expression data from 14 healthy skins; and 34 AD non-lesional (ANL, red), 7 AD acute (AAL, cyan) and 34 AD chronic lesional (CAL, blue) skins deconvoluted into reference cellular fractions by CIBERSORT. **[A]** Ten reference cellular signatures were used for deconvolution: *in situ* steady-state Langerhans cells, post-migration activated Langerhans cells, dermal dendritic cells, regulatory T cells, $CD8^+$ T cells, $CD4^+$ T cells, melanocytes, fibroblasts, and two bacterial co-culture keratinocyte model signatures: **[B]** *S. epidermidis* response at 24-hours (SE24h, olive green) and **[C]** unstimulated control. Error bars indicate mean ± SD.

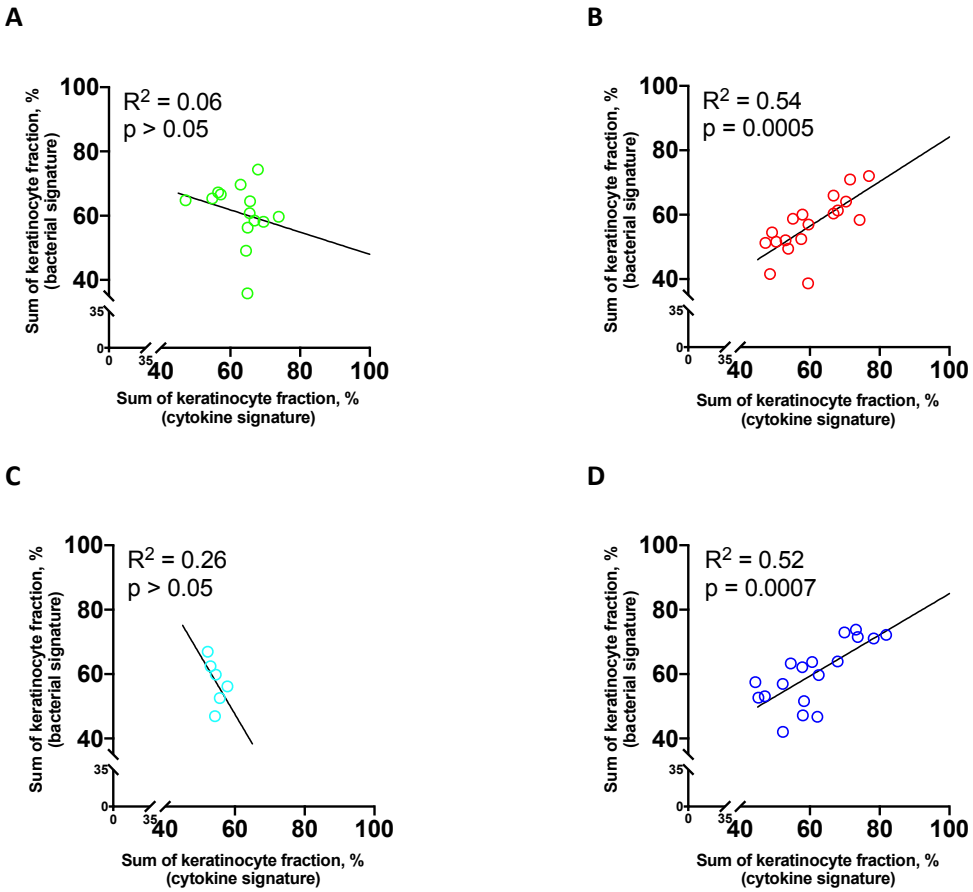


Figure 5.13. Correlation of cumulative keratinocyte signatures from both cytokine and bacteria sensing deconvolution panels. The total fraction of the 11 cytokine-stimulated keratinocyte profiles represented the total keratinocyte compartment resolved by the first whole skin deconvolution panel (x-axis). In the second whole skin deconvolution panel the keratinocyte compartment was composed of the two bacteria response profiles (y-axis). For each sample these two total keratinocyte fractions were correlated in **[A]** healthy, **[B]** AD non-lesional skin, **[C]** AD acute lesional skin, and **[D]**, and AD chronic lesional skin.

5.3.7 Cytokine responses and barrier expression in keratinocytes correlate with response to *S. epidermidis* challenge

Healthy samples showed no correlation between the bacterial signature and KC2 (IL-4, IL-13 fractions), KC17 (IL-17) immunophenotypes or barrier expression (**Figure 5.14**). The behaviour of KC2, KC17 and barrier expression was distinctly different between healthy controls and samples from atopic dermatitis. While the bacteria sensing SE24h fraction in healthy skin is not grossly different to inflamed skin, the fraction of KC2 does not exceed 10% nor is there any resolved KC17 fraction, while barrier expression does not fall below 60%. Therefore, KC2 fractions higher than this, resolving of a KC17 fraction and barrier expression below 60% are characteristic but not definitive of atopic skin regardless of lesion status.

Within the atopic dermatitis samples, both non-lesional and chronic lesions demonstrated a weak but significant negative correlation between KC2 immunophenotype fractions and bacterial, SE24h response fractions (**Figure 5.14A**, R-squared=0.2084, p=0.007; R-squared=0.235, p=0.004, respectively). Given that the KC2 transcriptomic response of keratinocytes from atopic dermatitis samples was earlier found to be constitutive regardless of disease state, the mirrored relationship of the response to infection and atopic cytokines in lesional and non-lesional skins is unsurprising.

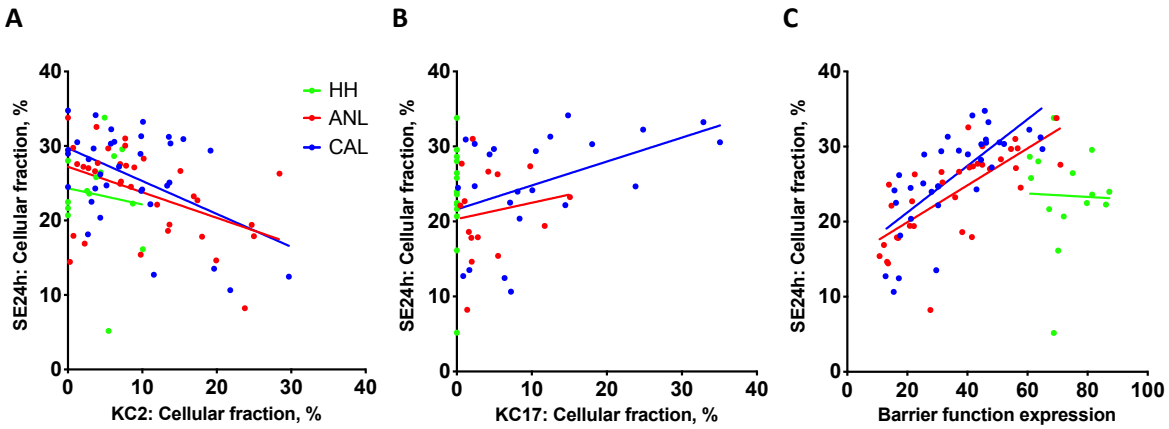


Figure 5.14. Keratinocyte responses to *S. epidermidis* (SE24h) correlated with KC2, KC17 immunophenotype and barrier expression. *S. epidermidis* response in RHE model keratinocytes (SE24h) correlated with the cellular fraction of: **[A]** KC2 immunophenotype; **[B]** KC17 immunophenotype and **[C]** expression of a panel of barrier-related genes in healthy skin (HH, green), AD non-lesional skins (ANL, red) and AD chronic lesional skins (CAL, blue).

Much of the KC17 fraction size in non-lesional atopic samples was found to be limited to below 15% and there was no correlation to bacteria sensing fractions (**Figure 5.14B**). Chronic lesions, where KC17 was found to be a defining fraction per previous CIBERSORT-based deconvolution (Section 4.2.3), did display a slight positive and significant correlation with bacteria sensing fractions (**Figure 5.14B**, R-squared=0.202, p=0.03). Although weak, this correlation in chronic lesional fractions indicates a potential inter-play in keratinocyte involvement with the microbiome and IL-17 cytokine responses. It remains unclear whether bacterial interaction is driving the keratinocyte signature via local recruitment of T_H17 cells for example, or alternatively, if chronic inflammation driven by IL-17, triggers keratinocyte AMP expression against bacteria colonisation.

The effect of T_H2 cytokines, namely IL-4 and IL-13, on the expression of vital epidermal barrier genes has been reported, and so has the downregulation of barrier genes in atopic dermatitis (Brandt and Sivaprasad, 2011; Hönzke *et al.*, 2016). These two factors form two-thirds of the atopic dermatitis

disease triad: atopic immune dysregulation and barrier abnormalities. The third component to this triad is the cutaneous microbiome. Therefore, it is interesting to consider the correlation between barrier gene expression and keratinocyte responses to commensal *S. epidermidis* (**Figure 5.14C**). These correlations are only moderately positive but highly significant in non-lesional and chronic lesional skins (R-squared=0.51, p<0.0001; R-squared=0.52, p<0.0001, respectively). Hence, an intact barrier appears to facilitate keratinocyte responses to commensal *S. epidermidis*, or alternatively that loss of commensal microbiome occurs alongside barrier dysfunction in AD skins.

5.4 Discussion

5.4.1 Monolayer challenge with exogenous bacterial factors on epithelial homeostasis

The main aim of this phase of work was to derive transcriptomic profile from keratinocytes during staphylococcal challenge. To investigate the epidermal response and inflammatory profile linked to this challenge, the available public datasets of keratinocyte *Staphylococcus spp.* challenge were initially explored. Live bacterial challenge of keratinocyte monolayer culture severely impacts the viability of the cells, which is experimentally restrictive over a duration of co-culture modelling persistent infection response. Secor *et al.* approached this problem by supplementing keratinocyte culture medium with the secreted extracellular proteome of *S. aureus* grown under planktonic or biofilm conditions (Secor *et al.*, 2011, 2012). Functional annotation of the *Staphylococcal*-treated keratinocyte gene expression revealed processes important for immune and inflammatory processes. These included reduced MAPK phosphorylation, production of cytokines IL-1β, IL-6, TNFα and GM-CSF, and chemokines CXCL-1 and CXCL-8.

IL-1 cytokines and GM-CSF have been implicated in the crosstalk between epidermal and dermal compartments involving paracrine regulation of keratinocyte proliferation. IL-1 released during keratinocyte stress stimulates KGF and GM-CSF expression from dermal fibroblasts , which in turn stimulates proliferation of the epidermal compartment in a paracrine manner (Werner and Smola, 2001). Furthermore, IL-6 expression is mediated by keratinocyte-derived IL-1 and TNFα, and is produced in a wounding context and downregulates the expression of the differentiation keratins (KRT1 and KRT10) (Sugawara *et al.*, 2001; Hernández-Quintero *et al.*, 2006). The expression of these genes by keratinocytes in the context of *Staphylococcal* challenge, indicates the potential role of keratinocytes outside the stratum corneum in maintaining epidermal homeostasis as opposed to strict anti-microbial activity. Their inclusion in investigations of monolayer keratinocytes strengthens this hypothesis.

IL-6 and TNFα have an important role in cutaneous adaptive immunity. These, along with IL-23, are critical for the T_H17 polarisation of naïve $CD4^+$ T cells (Iwakura and Ishigame, 2006). This is necessary for invoking the anti-microbial immunity offered by the T_H17 axis but may also elucidate the role of hyper-proliferation/lichenification associated with atopic dermatitis and psoriasis, and particularly the aggregation of the former by *S. aureus*. The deconvolution results investigating keratinocyte immunophenotype did not look for IL-6-induced responses due to this not being included in samples available from public data. However, TNFα immunophenotypes were significantly upregulated by microbial challenge in stratified model keratinocytes, while IL-17 responses were raised but not significant, partly due to low number of replicates. The resolution of these immunophenotypes from microbial challenged cultures indicates a link between TNFα, IL-17 adaptive immunity and a potentially proliferative phenotype. In the context of a wound or microbial invasion enhanced keratinocyte proliferation is required for epidermal maintenance and activation of T_H17 immunity (Liang *et al.*, 2006), however, aberrant proliferation would invoke these same processes manifesting as hyperproliferation and inflammatory disease (Zaba *et al.*, 2007; Chiricozzi *et al.*, 2011).

5.4.2 Validating challenged gene signatures from primary and HaCaT monolayer keratinocytes

Before curating gene lists associated with secreted *Staphylococcal* challenge, the ability of analogous samples across parallel datasets to correctly cross-identify was tested within the deconvolution pipeline. The purpose of this was to test the robustness of gene changes in keratinocytes induced by *S. aureus* secreted factors and, thus, determine if the associated signature is strong enough for use as a reference profile to deconvolute unrelated keratinocyte transcriptomes. This showed that planktonic associated changes were insufficient compared to biofilm changes, whereby PCM samples were incorrectly deconvoluted as composed of a BCM profile. Interestingly, sterile control samples were found to have a small BCM proportion. This may allude to the strengthened proliferative processes in challenged cultures which are nonetheless still present in sterile cultures but in lesser magnitude.

Both the two primary and HaCaT keratinocyte datasets were reanalysed for differentially expressed genes specific to growth with BCM. These results were similar to the original reports and included infection related processes such as ‘response to lipopolysaccharide’, ‘inflammatory response’, and ‘defence response’ (Secor *et al.*, 2011, 2012). For both primary cell bases experiments, ‘response to lipopolysaccharide’ was the dominant gene ontology term describing the upregulated biological processes from BCM culture. Although, *S. aureus*, as a Gram-positive bacterium, does not produce lipopolysaccharide, this gene ontology term is a reflection of the fact that many of the host genes

activated in response to bacterial challenge are similar in their downstream effector processes and subsequent biological phenotype.

There was no core overlapping gene list in comparing genes altered by BCM conditions between the three complimentary datasets. Even between both primary cell datasets, the overlap was approximately only one-tenth of their respective lists. Despite the lack of core overlap, signatures derived from primary keratinocytes of both sterile control and BCM condition were able to correctly annotate seen and unseen samples during CIBERSORT-led deconvolution. This was also true for HaCaT-derived signatures, but primary cell signatures proved more accurate. While HaCaTs are often used in place of primary keratinocyte cultures, they demonstrate reduced lamellar bodies resultant of abnormal lipid production and deficiency compared to primary cells (Boukamp *et al.*, 1988; Schoop, Mirancea and Fusenig, 1999). Additionally, key inflammatory processes are abnormal in HaCaTs such as low expression of TLRs and NFκB subunits (Qin *et al.*, 1999; Seo *et al.*, 2012). For investigations involving bacterial challenge, HaCaTs may not be appropriate as model cells due to impacted PAMP-activation of protective keratinocyte responses.

5.4.3 Developing bacteria response signature from commensal challenged epidermal models

The bacterial response signature derived from reconstituted model keratinocytes co-cultured for 24-hours with *S. epidermidis* has been difficult to place on the commensal-pathogenic spectrum. The CIBERSORT deconvolution of the *S. epidermidis* samples at 24-hours showed a significant TNFα response from keratinocytes. This confirmed previous reports showing that topical application of *S. epidermidis* to a similar RHE model system resulted in the increased media concentration of pro-inflammatory cytokines such as TNFα as well as IL-1α, IL-6 and IL-8 (Duckney *et al.*, 2013). Furthermore, the synergistic link between TNFα and IL-17 has been reported as an important cutaneous inflammatory circuit (Zaba *et al.*, 2007; Guilloteau *et al.*, 2010; Chiricozzi *et al.*, 2011). The complete reversal of the IL-17 response at 24-hours in the control models, and near-complete removal of this signature in the other co-culture models may be relevant to *S. epidermidis* inducing or maintaining an IL-17 response in keratinocytes at persistent infection. This TNFα and IL-17 response by keratinocytes is characteristic of the commensal activity of *S. epidermidis* to exclude *S. aureus* colonisation through the production of anti-*S. aureus* products, such as hBD2 (Liang *et al.*, 2006; Hönzke *et al.*, 2016). Under normal circumstances, *S. epidermidis* and its products and secreted factors effect a highly specific immune response to *S. aureus* by keratinocytes (Li *et al.*, 2013; Wang *et al.*, 2017). This protective commensal role is dependent upon the integrity of the epidermal barrier, which once compromised, amplifies the pathogenicity of *S. aureus* and makes *S.*

epidermidis an opportunistic pathogen (Duckney *et al.*, 2013; Burian *et al.*, 2017; Dong, Speer and Glaser, 2018).

The *S. epidermidis* samples at 24-hours are reported to have a marked increase in the protein expression of the AMPs S100A7 (log_2FC 0.93), S100A12 (log_2FC 0.67), and hBD2 (log_2FC 2.12) (Holbrook, 2019). These key keratinocyte AMPs are commensally induced and point to *S. epidermidis* in this context acting as a cutaneous commensal. Furthermore, while *S. epidermidis* challenge increased mRNA expression of pro-inflammatory markers, these were not increased at the protein level, indicating a level of post-transcriptional regulation to commensal challenge not observed by pathogenic challenge (Holbrook, 2019).

Similar work by Duckney and colleagues where *S. epidermidis* was cultured topically on RHE models corroborates these results (Duckney *et al.*, 2013). Topical application of *S. epidermidis* did not elicit an inflammatory response by keratinocytes and was considered in this context to be a commensal. However, when *S. epidermidis* was added to the culture medium in order to simulate barrier breach, this bacterium was shown to be exceptionally inflammatory and cytotoxic to keratinocytes in the model. In this experimental context, *S. epidermidis* is opportunistically pathogenic. In contrast, *S. aureus* was shown to be pathogenic through induction of inflammatory response, even when applied topically. When added to culture medium *S. aureus* was devastatingly cytotoxic resulting in complete keratinocyte cell death and destruction of the reconstituted epidermis model. In all, our CIBERSORT-derived results, the AMP expression analysis from D.H., and work on the location-dependent pathogenic-inducement of *S. epidermidis* point to the keratinocyte signature derived from co-culture of this *Staphylococcal sp.* being, indeed, commensal in nature.

The identity of the 24h *S. epidermidis* co-culture signature as commensal or pathogenic is complicated by ontology analysis of the DEGs between reference samples and the rest of the samples in the dataset. The 678 genes found to be upregulated in the bacterial signature had bacterial process functional enrichment that are characteristic of pathogenic *Staphylococcal* responses. The response to hypoxia annotation, which accounts for one-quarter of the redundant but significant ontology terms returned by ToppGene, is noted when keratinocytes are infected. Hypoxia as a result of metabolic stress experienced by infected keratinocytes induces the expression of HIF-1α, a transcription factor that regulates responses to this hypoxia (Wickersham *et al.*, 2017). This promotes increased glycolysis in keratinocytes in an effort to compete with the infecting bacteria for their metabolic needs. Indeed, upregulated glucose metabolism is represented in the collapsed visualisation of the gene ontology terms associated with our *S.*

epidermidis signature. This analysis could point to our *S. epidermidis*-derived signature in keratinocytes being pathogenic in nature.

The reconstituted epidermis model used with the *S. epidermidis* co-culture has presented signatures reflective of both commensal and pathogenic impact on keratinocyte responses. The commensal signature is derived from whole transcriptome analysis and reference-based deconvolution. The altered transcriptomic profile of the SE24h group, compared to the unstimulated group is driven by keratinocyte responses to, mainly, TNFα and, partly, IL-17. This cytokine response from keratinocytes is classical to commensal action by *S. epidermidis* to prevent pathogenic *S. aureus* colonisation through the production of AMPs, specifically hBD2 (Lai *et al.*, 2010; Otto, 2010; Li *et al.*, 2013). Surprisingly, *S. epidermidis* is also capable of driving inflammatory gene expression responses. This may be reflective of the experimental system utilising keratinocytes without other immune regulating cellular populations, timing, inoculation density, favourable growth conditions, or any other experimental factors not native to *in vivo* experience of host responses to *S. epidermidis*.

However, as the commensal signature can be determined from the whole transcriptome, i.e., >10,000 analysed genes, it may be assumed to dominate the potential pathogenic signal also observed. Indeed, protein expression analysis performed on previous investigation of *S. epidermidis* challenge of the same RHE model system indicated pro-inflammatory transcripts are not translated to the protein level whereas anti-microbial factors moderating but not opposing colonisation were (Holbrook, 2019).

The cause of this duality of the SE24h signature could be due to any number of experimental factors. It is possible that either the epidermal barrier and/or keratinocyte tight junctions in the RHE models are not as robust as in healthy human epidermis. For example, some *S. aureus* strains are highly immunogenic to the RHE models, even when applied topically, however, in healthy human skin colonisation by this bacterium does not necessarily lead to adverse sequelae. Any breach of the topical barrier in models by *S. epidermidis* would induce a pathogenic signature, which may not reflect the events or severity in human skin. Additionally, the models only comprise of keratinocytes, without the other cutaneous cell populations. As such, it is difficult to determine if the observed responses of bacteria in RHE models are reflective of whole skin, particularly how these missing populations would mediate immunogenic or tolerogenic responses bacteria.

Additionally, this bulk analysis may only capture the strongest signatures from the keratinocytes in response to bacterial challenge. In this sense, it may not allow for the appreciation of signatures

that are layer specific as might be expected from a highly stratified tissue such as the epidermis. While bacterial response signatures in the upper layers of the epidermis may be similar in the context of commensal and pathogenic *Staphylococcal spp.*, it may indeed be the subtle responses of the underlying basal populations that influence the global response. To investigate this would require the resolution of the epidermis at much more complexity than monolayer experiments and much higher granularity than the bulk transcriptomic analysis. The comprehensive delineation of layer specific and bacteria specific signatures would require transcriptomic analysis at the single cell level, using DropSeq analysis of the established epidermal model bacterial challenge protocol.

Chapter 6 Defining gene signatures of keratinocyte microbial responses at a single cell level

6.1 Introduction

Transcriptomic analysis of bulk reconstituted models carried out in this work, and previously by D. Holbrook (2019), demonstrated the induction of new transcriptional keratinocyte programmes upon live bacterial challenge. The models allow for the investigation of the effect of live *Staphylococcal spp.* colonisation of stratified differentiated keratinocytes by benefiting from the reconstruction of the stratum corneum. The advantage of reproducing these terminally differentiated keratinocytes is by protecting underneath proliferating and differentiating populations from the cytotoxic effect of live bacterial challenge and modelling the responsive transcriptomic programmes of these deeper populations. The three-dimensional models permit investigation within the stratified epidermis. Previously, the models were used to derive a bulk transcriptomic signature from 24-hour challenge and showed this signature correlated differently with atopic phenotype and barrier marker expression (Section 5.3.7), providing proof of concept for the current study.

The epidermis is a highly stratified and structured tissue with cell populations performing different cutaneous functions (Werner and Smola, 2001). These include proliferation for maintenance of epidermal turnover; differentiation to produce the skin barrier physically by tight junction formation and production of AMPs; and terminal differentiation to protect against dehydration and infection (Madison, 2003; Duckney *et al.*, 2013; Eckhart *et al.*, 2013; Sumigray and Lechler, 2015). It is expected that these different populations of keratinocytes would be differentially altered by contact or the indirect effects of microbial challenge (Duckney *et al.*, 2013). The limitation of bulk transcriptomic analysis of the RHE models is the projection of these putative individual cell type differentiated epidermis signatures into the whole tissue phenotype.

A stratified morphology describes the classical epidermal layers but also compartmentalises proliferation/differentiation functions, tissue homeostasis and inflammation. Although this morphology is well described, the contribution of this cellular heterogeneity to epidermal homeostasis and function is not yet fully understood (Cheng *et al.*, 2018). It has recently been reported, using a large single-cell investigation of approximately 100,000 cells from healthy and inflamed epidermis, that almost one-eighth of the epidermal transcriptome aligns with classical differentiation patterns yet their cutaneous function and biological mechanisms are unknown

(Cheng *et al.*, 2018). Other studies have used single cell analysis to investigate this question, however, these have almost exclusively focussed on the roles of inflammation and immune activity underscoring epidermal heterogeneity (Cheng *et al.*, 2018; He *et al.*, 2020; Reynolds *et al.*, 2021).

The role of the microbiome is predicted to be as important as host immune crosstalk in mediating and altering cutaneous cell populations (Duckney *et al.*, 2013; Grice, 2015; Chang *et al.*, 2018; Woo and Sibley, 2020). Indeed, the role of *S. aureus* in modulating T_H2 and T_H17 immune responses in the skin and, critically, the intimate involvement of microbial signatures with keratinocyte differentiation and extracellular matrix organisation were recently reported (Fyhrquist *et al.*, 2019). The elucidation of high-resolution host-microbe interactions in the epidermis, therefore, presents a valuable topic for determining cellular crosstalk and its role in cutaneous health and inflammation.

To further characterise the epidermal response to *Staphylococcal spp.* challenge, single-cell analysis of the reconstructed models was undertaken. The aim of this was to investigate the sterile keratinocyte signature in comparison to commensal signatures derived from *S. epidermidis* co-culture and pathogenic signatures derived from *S. aureus* infection. The development, culturing and challenge protocol was previous established by D. Holbrook (Holbrook, 2019), which was followed in order to provide comparison to previous work. In order to expand the analysis to the single-cell resolution, optimisation of model dissociation and subsequent keratinocyte encapsulation for DropSeq was required.

6.1.1 Hypothesis

Bacterial challenge with either commensal *S. epidermidis* or pathogenic *S. aureus* alters the keratinocyte transcriptome to induce different subsets of proliferating and differentiating populations.

6.1.2 Aims

- Computationally compare the RHE models to healthy epidermis to identify specific stratified epidermal layers and understand their transcriptomic profile.

- Optimise the dissociation of the RHE models for single cell analysis by DropSeq for the creation of a scRNA-seq dataset of *Staphylococcal* challenged keratinocytes.

- Identify stratified populations of keratinocytes from the single cell transcriptomic data.

- Investigate the effect of *Staphylococcal* challenge on the transcriptome of layer sub-populations to determine effect on proliferation and differentiation.

6.2 Results

6.2.1 Reconstituted human epidermis models mimic the structure and transcriptional programmes of the *in vivo* keratinocyte compartment

Unlike whole skin or epidermis samples, the reconstituted epidermis models are composed only of keratinocytes. For cell-based reference deconvolution by CIBERSORT of the epidermal models, all the non-keratinocyte signatures were removed. The models are intended to display proliferating and differentiating compartments of human epidermis.

While the previous work was able to determine the general morphological consistency between and within batches, this is done by sacrificing a single sample from each batch for H&E staining. A computational approach using the data derived from experimental samples provides the advantage of understanding the composition the samples taken for analysis and is directly relatable to the data produced. For this purpose, single cell data of healthy keratinocyte basal, mitotic and spinous populations were identified from the Cheng *et al.* (2018) study (**Figure 6.1A**). The three populations of keratinocytes were identified by expression of KRT5 and KRT14 (basal), CCND1 and PCNA (mitotic), and KRT1 and KRT10 (spinous) as reported by the authors (**Figure 6.1B**). Other populations such as channel and follicular keratinocytes, and other epidermal populations were excluded, and the dataset did not report a well-defined cluster forming the granular layer of keratinocytes. The gene expression profiles of these population clusters were collapsed into a pseudobulk profile and provided as reference signatures to CIBERSORT for deconvolution of the bulk RHE samples. Therefore, using CIBERSORT and a panel representing keratinocyte layers it is possible to assess the defined layers within the RHE models. This approach allowed investigation of any potential batch effect arising from either culture date or microbial inoculation, and the morphological complexity of the models.

For each of the three signatures the number of cells and proportion of total keratinocytes as extracted from whole human epidermis is shown (**Figure 6.1C**). As the RHE models are composed of solely keratinocytes they were compared only to the total keratinocyte fraction of human epidermis, which showed remarkable consistency between the three populations of keratinocyte between human epidermis and model epidermis (**Figure 6.1C, D**). Approximately 90% of human epidermis is captured by defining the cells into the three populations considered here and demonstrates that the models greatly recapitulate human epidermis. This comparison to human epidermal keratinocytes in relation to the proportions of the basal, mitotic and spinous populations also demonstrates that the models have share very similar morphological complexity.

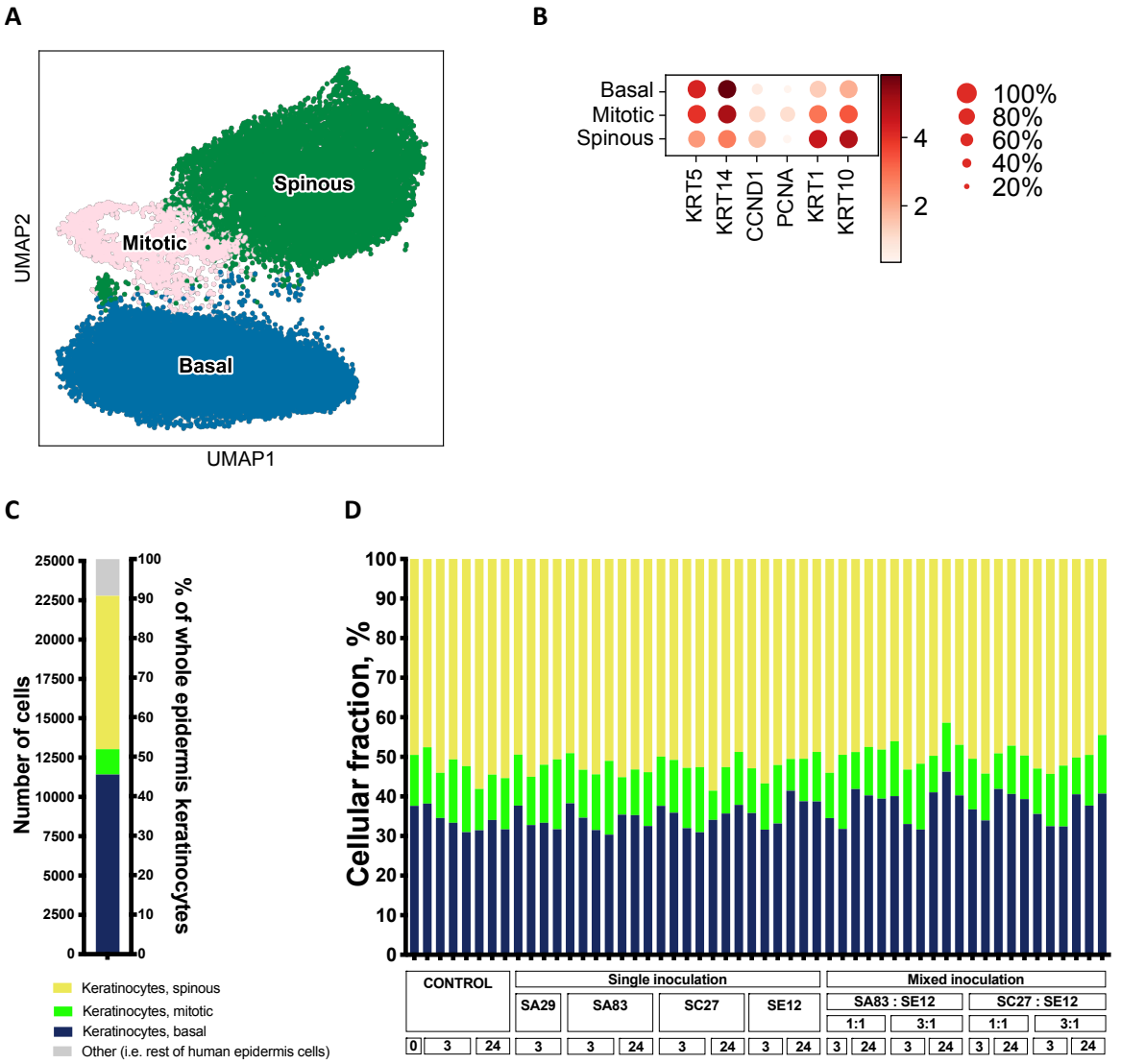


Figure 6.1. CIBERSORT deconvolution of the bulk RHE samples investigated for epidermal layers. **[A]** Basal (11,424 cells), mitotic (1,610 cells) and spinous (9,752 cells) keratinocytes subset from the healthy epidermis samples from the Cheng *et al.* (2018) study represented by UMAP. **[B]** Expression of layer specific markers *KRT5*, *KRT14*, *CCND1*, *PCNA*, *KRT1* and *KRT10* used to annotate these clusters by Cheng *et al.* **[C]** Breakdown of the defined populations in the single cell epidermis data showing the absolute number of cells (left y-axis) or per centage of epidermis (right y-axis) for comparison to the RHE samples. **[D]** The 54 RHE samples investigated by bulk microarray analysis by D.H. (iCASE Unilever PhD) were resolved into epidermal compartments as defined by analysis from scRNA-seq of human epidermis (Cheng *et al.*, 2018) defining populations of basal (blue), mitotic (green) and spinous (yellow) cells.

The samples in the RHE bulk dataset profiled by CIBERSORT into the three keratinocyte layer signatures (**Figure 6.1D**). The composition of the dataset included samples at time zero, three hours and 24 hours from models challenged with either *S. aureus* (SA 8325-4, SA 29213), *S. epidermidis* (SE 12228) or *S. capitis* (SC 27840), or mixed inoculations thereof, or left sterile as a control. The samples showed very consistent proportions of the three signatures across batch dates, inoculum and culture condition. The basal compartment represented approximately 30-45% of the whole sample. Together with the mitotic fraction, forming the proliferating compartment, made up around half of the of the models. The remainder, 45-65%, was comprised of the spinous fraction representing the differentiating compartment of the epidermal model.

6.2.2 Optimisation of DropSeq protocol and creation of reconstituted human epidermis model singe cell transcriptomic data

To undertake single cell DropSeq analysis of the *Staphylococcal spp.* challenged model epidermis, the dissociation on the RHE models was optimised and described in Methods (Section 2.5.2.2). The effect of size and shape variability of the dissociated differentiated keratinocytes on efficient encapsulation was visualised by CSFE-staining of cells before droplet encapsulation (**Figure 6.2**). Dissociated keratinocytes were encapsulated and subsequent formed droplets were assessed visually by fluorescence microscopy for uniformity and proportion of encapsulated cells. The reported encapsulation efficiency of the DropSeq protocol using the standard flow rates for aqueous (cell and bead suspensions) inlets, and oil inlets (4000μl/hr, 14000μl/hr, respectively) is around 5% (Macosko *et al.*, 2015). Figure 6.2 shows 314 highly uniform fully visible oil droplets in the field with 25 CFSE-stained keratinocytes, giving an encapsulation efficiency of 7.3%.

Parallel encapsulation of dissociated keratinocytes from each challenge condition was not possible due to limitations of laboratory equipment. Further, concerns about the viability of dissociated keratinocytes, especially those post-challenge, while awaiting encapsulation was the main justification for multiplexing the three conditions by use of barcoded antibodies. Antibodies directed against β2-microglobulin and CD298 with conjugated oligonucleotide sequences allow sample-specific labelling to allow multiplexing and, therefore, only require one single encapsulation run (Stoeckius *et al.*, 2018). An additional benefit of this is the elimination of the batch effect between encapsulation runs of each condition. Samples are identified by de-multiplexing after sequencing by reference to a set of hashtag oligonucleotide library.

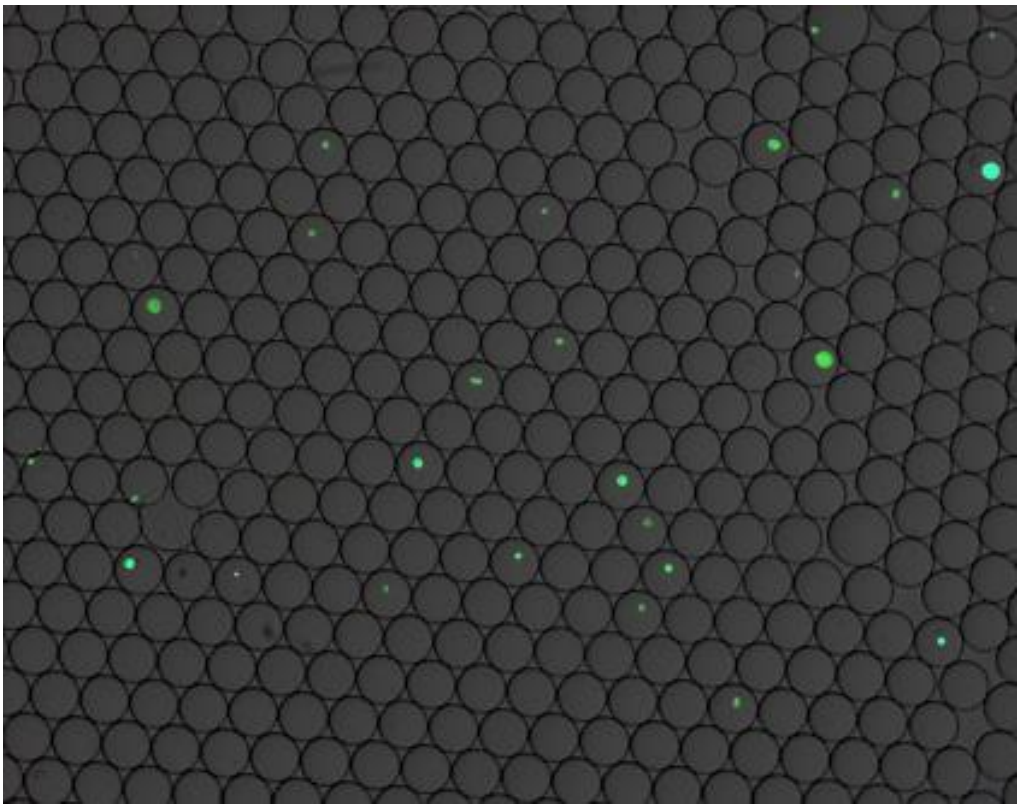


Figure 6.2. Optimisation of DropSeq encapsulation. RHE keratinocyte models were dissociated and stained with CFSE. Encapsulation efficiency and uniform droplet quality was visualised using fluorescence microscopy (10X). Shown here are 341 full droplets with 25 encapsulated keratinocytes giving an encapsulation efficiency of 7.3%.

Dissociated and multiplexed keratinocyte samples were processed through the DropSeq encapsulation protocol (Macosko *et al.*, 2015) and the produced STAMPs (single transcriptome attached to microparticle) used to create cDNA libraries from both the sample gene expression and HTO barcode. For each of the three batches, cDNA libraries were then tagmented, indexed and checked for size and quality on an Agilent Bioanalyzer before sequencing using the Illumina NextSeq system (**Figure 6.3**).

After sequencing output files were demultiplexed by aligning hashtag oligonucleotide reads to the barcode reference dictionary. Barcodes were plotted for each batch independently using uniform manifold approximation and projection (UMAP) calculation of STAMPs sequenced (**Figure 6.4**). The UMAP plots for each batch were clustered using the highest Leiden resolution and mapped to the regions of highest detection of hashtag barcode for each sample of sterile control, *S. aureus* challenge, and *S. epidermidis* challenge (**Figure 6.5**). The STAMP barcodes along with challenge annotation were transferred to the gene expression data for normal scRNA-seq quality control and filtering pipeline.

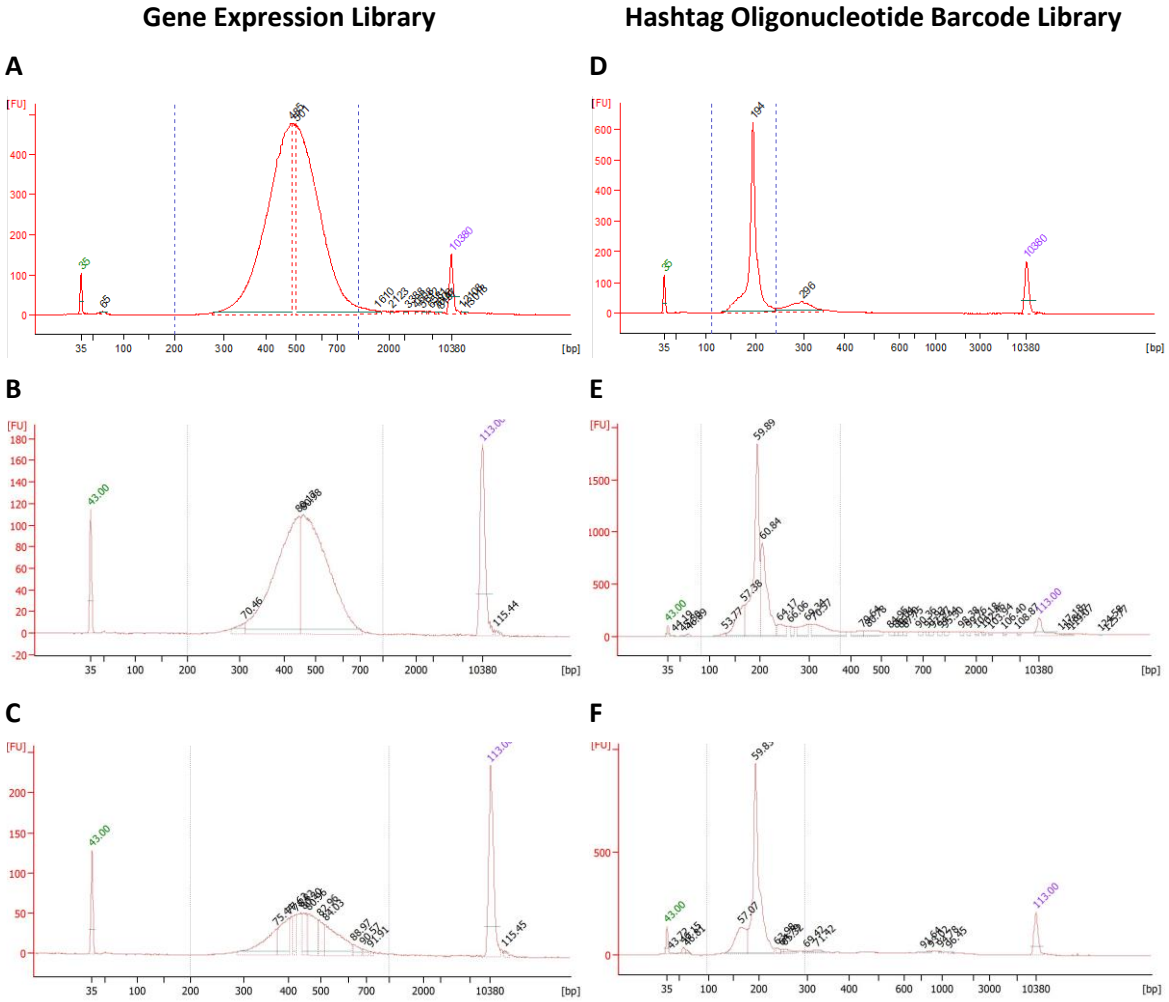


Figure 6.3. cDNA libraries from three independent co-culture batches. Amplified, tagmented and indexed keratinocyte gene expression (A-C) and multiplexing Hashtag oligonucleotide (HTO) libraries (D-F) were assessed for size and quality using Agilent Bioanalyzer system prior to sequencing. The major peak between the green and purple indicated peaks represents the library of interest. Batch 1 indicated a gene expression concentration of 6109.60pg/μl with average fragment size of 500bp [A]; and a HTO library size of 1063.49pg/μl, average fragment size of 192bp. [D] Batch 2 indicated a gene expression concentration of a 1-in-10 dilution of 1004.34/μl with average fragment size of 459bp [B]; and a HTO library size of 5456.41pg/μl, average fragment size of 214bp. [E] Batch 3 indicated a gene expression concentration of a 1-in-10 dilution of 438.02/μl with average fragment size of 464bp [C]; and a HTO library size of 2547.62pg/μl, average fragment size of 195bp. [F]

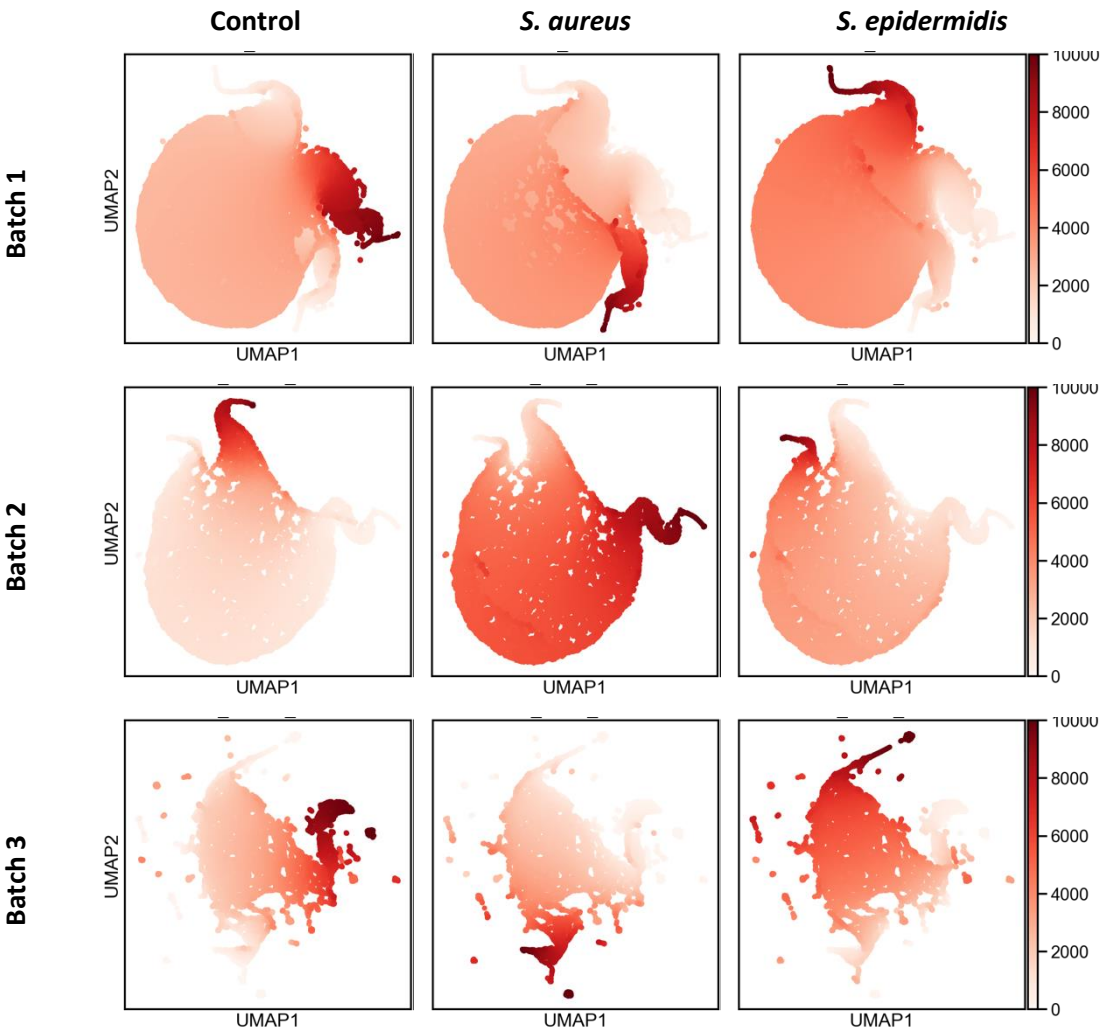


Figure 6.4. HashTag oligonucleotide (HTO) detection by UMAP plotting. Samples within each of the three batches were multiplexed using HTO antibody protocol. Detection of the sample-specific oligonucleotide was performed by plotting on a UMAP projection.

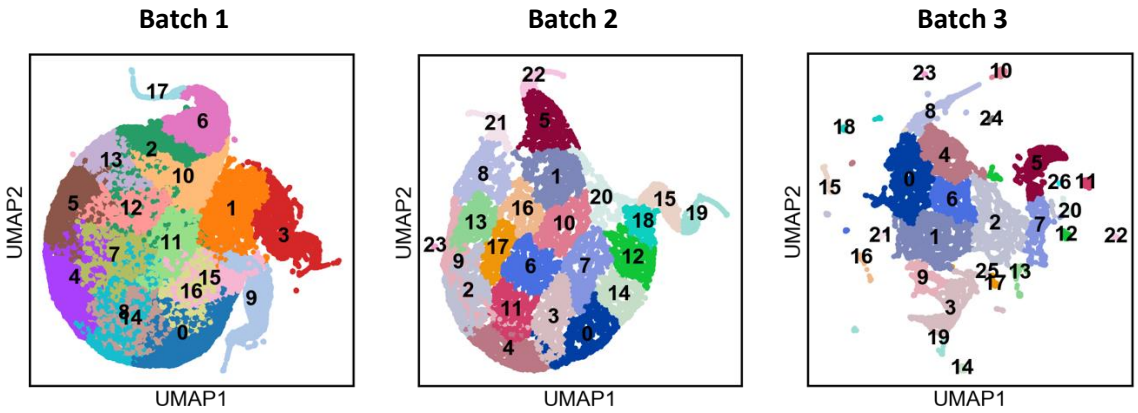


Figure 6.5. Mapping HTO detected STAMPs to Leiden clusters for annotation and filtering. The UMAP plots of STAMPs were clustered using the maximum Leiden resolution and regions mapping to detected HTO samples for each batch were annotated to retain only HTO-labelled cells. Retained in batch 1 for control were clusters 1, 3; for *S. aureus* was cluster 9; for *S. epidermidis* were clusters 6, 17. Retained in batch 2 for control were clusters 1, 5 22; for *S. aureus* was clusters 12, 15, 18, 19, 20; for *S. epidermidis* were clusters 8, 21. Retained in batch 3 for control were clusters 5, 7, 11, 12, 13, 22, 26; for *S. aureus* was clusters 3, 9, 14, 17, 19; for *S. epidermidis* were clusters 0, 4, 8, 10, 15, 18, 23, 24.

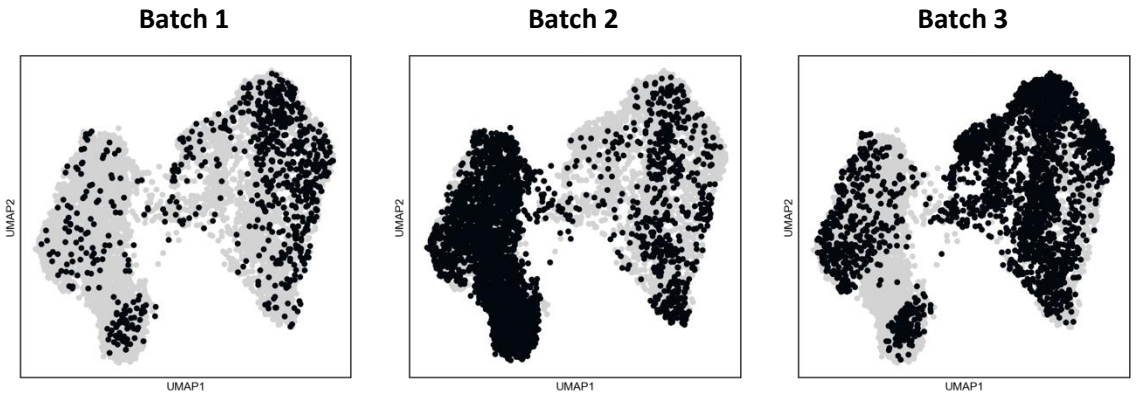


Figure 6.6. Post-integration UMAP scatter plot of the three independent batches across the whole dataset.

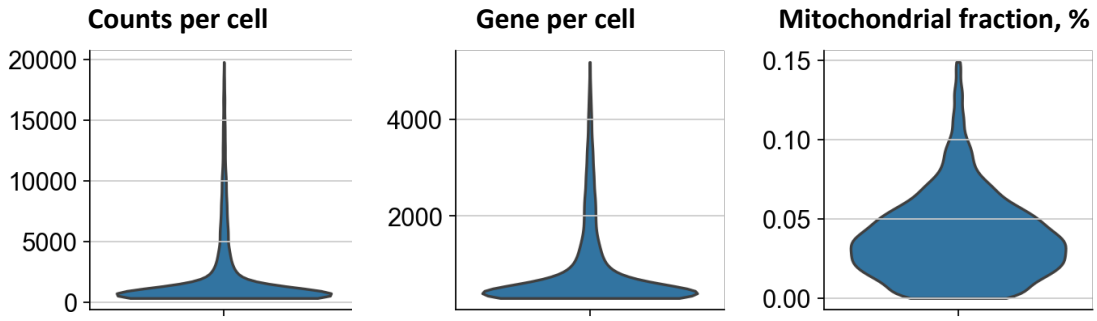


Figure 6.7. Quality control violin plots of the integrated RHE scRNA-seq data. The integrated three batches of challenged models were filtered according to a quality control pipeline using minimum (300) and maximum (20,000) counts, minimum (300) genes and maximum (15%) mitochondrial genes per cell.

Table 6. Quality control processing and filtering of the integrated scRNA-seq data.

Filter step	**Number of cells retained:**	**Number of cells filtered:**
Pre-QC/Filtering	11,532	*na*
Minimum counts per cell (300)	8,773	2,759
Maximum counts per cell (20,000)	8,772	1
Maximum mitochondrial genes (15%)	8,271	501
Minimum genes per cell (300)	6,391	1,880
Batch breakdown of the filtered 6391 cells		
Batch 1	762	
Batch 2	3,129	
Batch 3	2,500	
Challenge breakdown of the filtered 6391 cells		
Sterile control	2,044	
S. aureus	2,319	
S. epidermidis	2,028	

Raw, unnormalized read counts were imported into the SCANPY pipeline for pro-processing , cell and gene filtering (Wolf, Angerer and Theis, 2018). Using the SCANPY toolkit to analyse the single-cell data from three independent batched were first concatenated and merged by BBKNN integration. The dispersion over the UMAP space of each batch indicates that all populations were captured within each batch (**Figure 6.6**). The data were then filtered by a minimum of genes per cell, a minimum of cells per gene, removal of encapsulation doublets by SCRUBLET and removal of cells with high proportion of mitochondrial genes indicating apoptotic cells (**Figure 6.7**). The filtering process is shown in **Table 6**.

6.2.3 Layer defining keratinocytes can be identified from the populations of single-cell transcriptomes

Three batches of models each of the same experimental design comprising RHEs challenged with *S. aureus* or *S. epidermidis* or a sterile control were dissociated and encapsulated for DropSeq single-cell analysis. The batches after quality control filtering had 762, 2500 and 3129 cells comprising the total dataset of 6391 cells. After sample integration all three conditions were approximately equally represented: control contributed 2044 cells; *S. aureus* challenge, 2319 cells; *S. epidermidis* challenge, 2028 cells. The BBKNN package was used to remove batch effects on the integration of separate datasets. Batch annotation overlaid the UMAP projection show that while some cells are more dominant in some batches over other, all batches were included across the dataset (**Figure 6.6**).

Unbiased clustering using the Leiden algorithm at a resolution 0.2 identified three clusters in the UMAP plot (**Figure 6.8A**). Gene characteristic for basal, spinous and granular keratinocytes were used as markers to identify layer specific populations within the UMAP projection (**Figure 6.8B**). Leiden clusters mapped generally to the local expression of selected basal, spinous and granular markers (Traag, Waltman and van Eck, 2019). Basal cells made up 40.0% (2559 cells) of the dataset, spinous and granular cells composed 9.7% (621 cells), and 50.2% (3211 cells), respectively.

The top 20 genes characterising each layer cluster against the other two groups was derived by a rank-genes t-test method and used for gene ontology analysis to confirm layer-specific keratinocyte biology (

Table 9). The main biological process associated with these genes was hemidesmosome activity (GO: 0031581, B.H. adjusted p-value = $3.36E^{-13}$) (**Figure 6.9A**). Ontology analysis of the genes characterising basal cells for cell compartment enrichment indicated they were related to the basement membrane, basal plasma membrane and basal part of the cell (**Figure 6.9B**). Biological process interrogation of the genes characteristic of the spinous cells was dominated by “membrane assembly” (GO: 0071709, B.H. adjusted p-value = $5.32E^{-17}$) and “epidermis development” (GO: 0008544, B.H. adjusted p-value = $1.46E^{-16}$) (**Figure 6.9C**). These processes indicated that the cells in this cluster are undergoing the characteristic cell junction and cytoskeleton organisation characteristic of keratinocyte differentiation and epidermis development. The gene ontology summary of the biological processes involved in the top 20 genes characterising the granular cluster confirmed the activated state of these cells and the complex biology displayed by these cells (**Figure 6.9E**). The top ontology terms summarising this cluster were equally weighted and were “epithelial cell differentiation” (GO:0030855, B.H. adjusted p-value = $2.77E^{-05}$) and “antimicrobial humoral response” (GO: 0019730, B.H. adjusted p-value = $9.64E^{-07}$). Cellular compartment summarisation of

the genes indicated the apical, secretory nature of these cells forming the upper layers of the reconstituted models (**Figure 6.9F**).

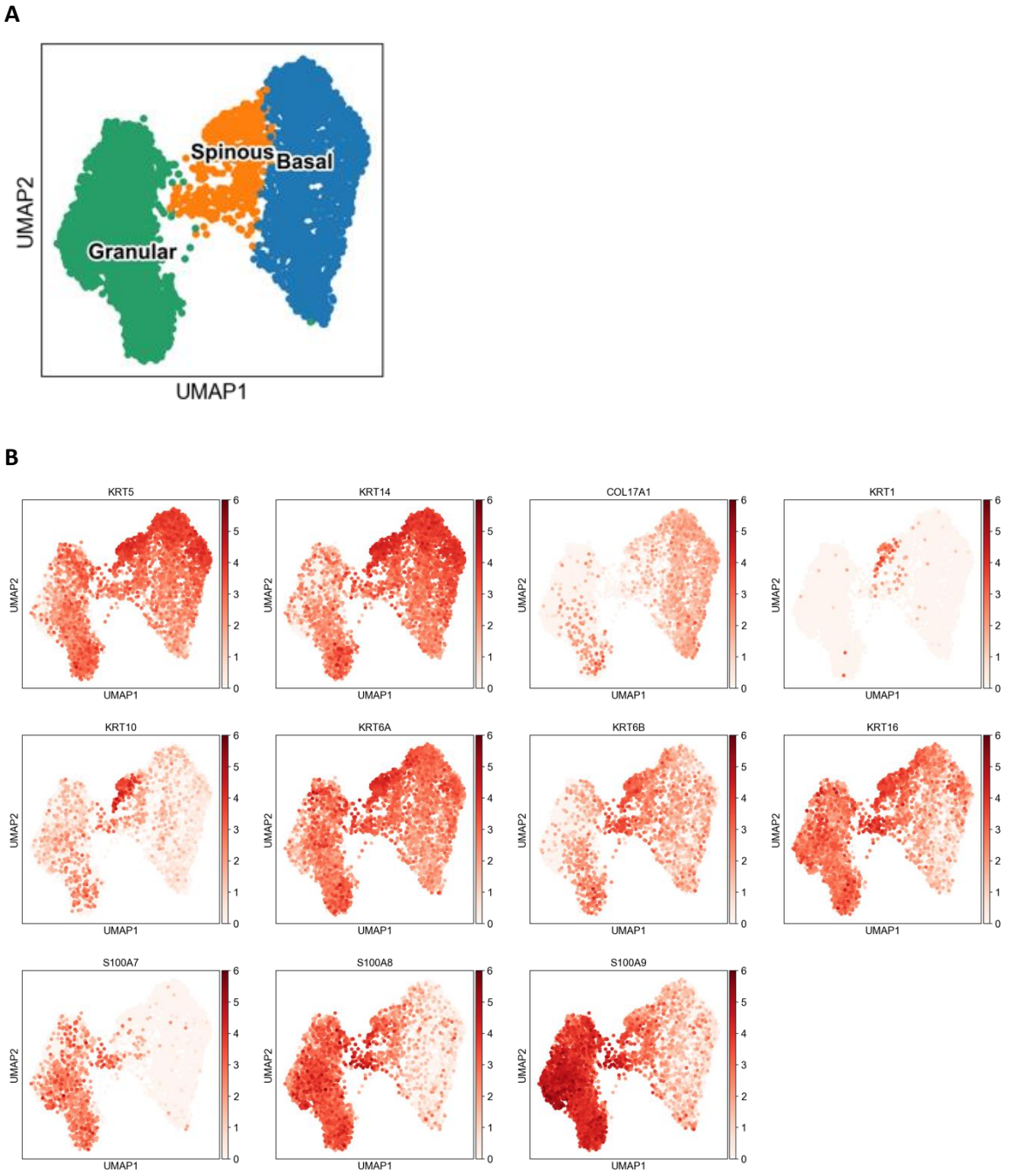
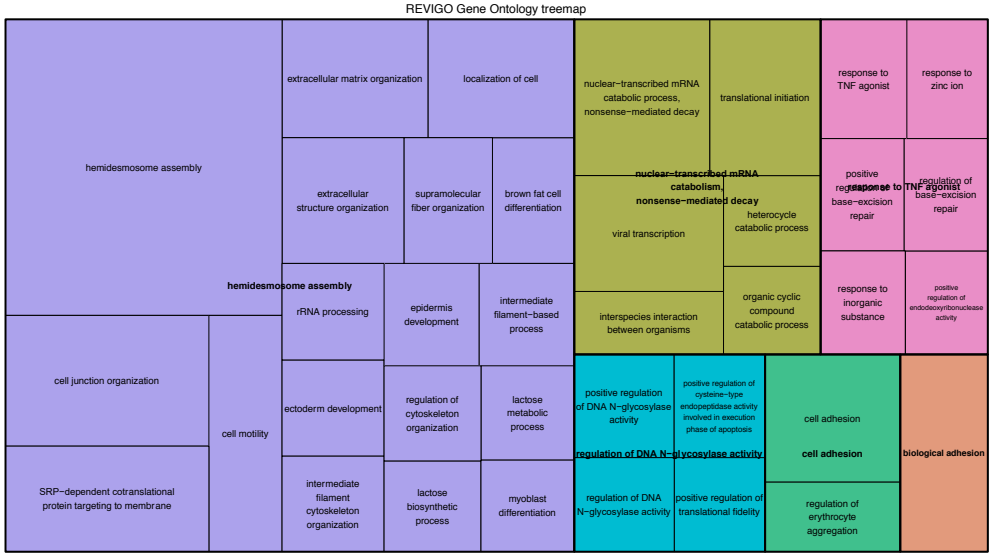
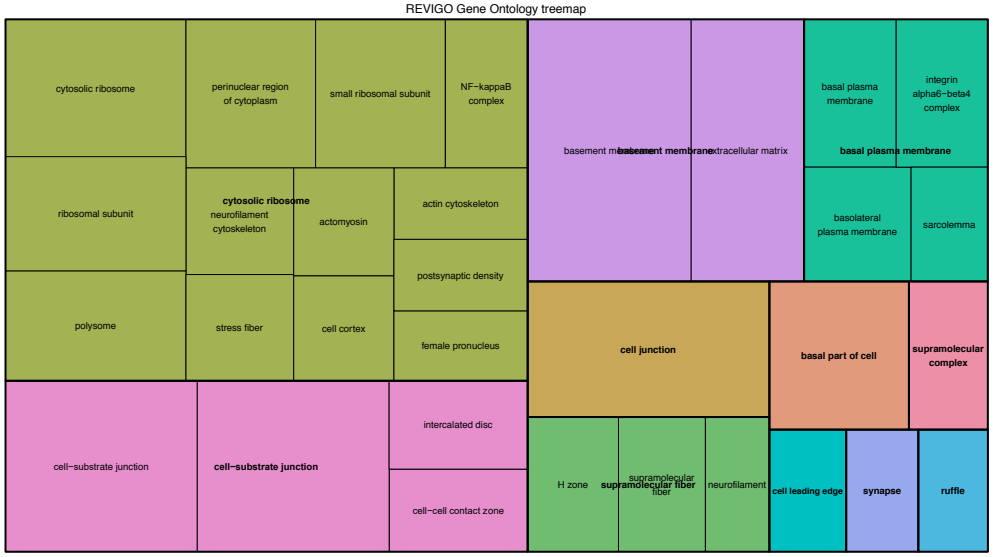


Figure 6.8. Epidermal markers plotted for expression across the whole dataset to aid unbiased cluster granularity. [A] Clusters defined by unbiased Leiden clustering at a resolution of 0.2 matched approximate expression of layer markers for annotation. **[B]** Markers for basal keratinocytes (*KRT5*, *KRT14* and *COL17A1*), spinous keratinocytes (*KRT1*, *KRT10*) and granular/differentiating markers (*KRT6A*, *KRT6B*, *KRT16*, *S100A7*, *S100A8*, *S100A9*) were used to aid annotation of regions of the UMAP plot.

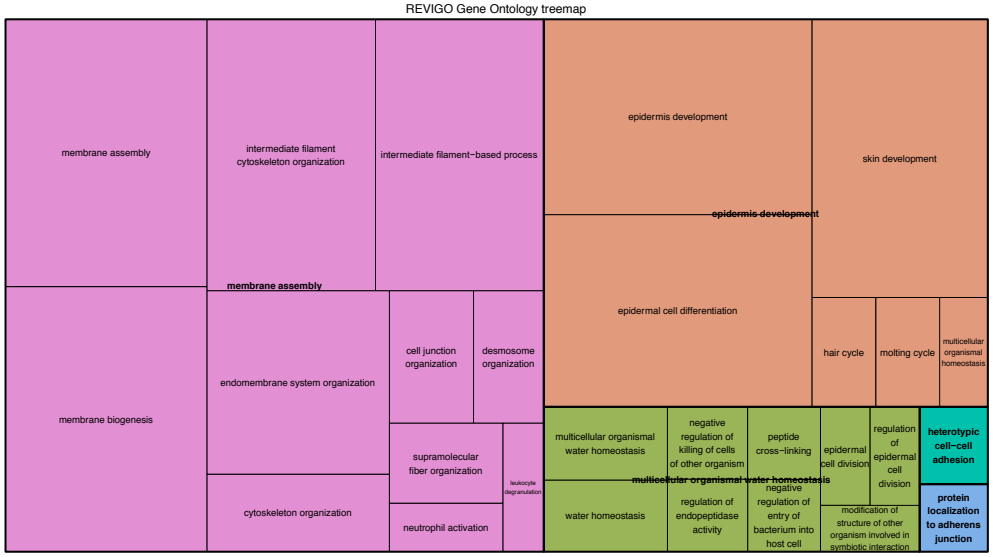
A



B



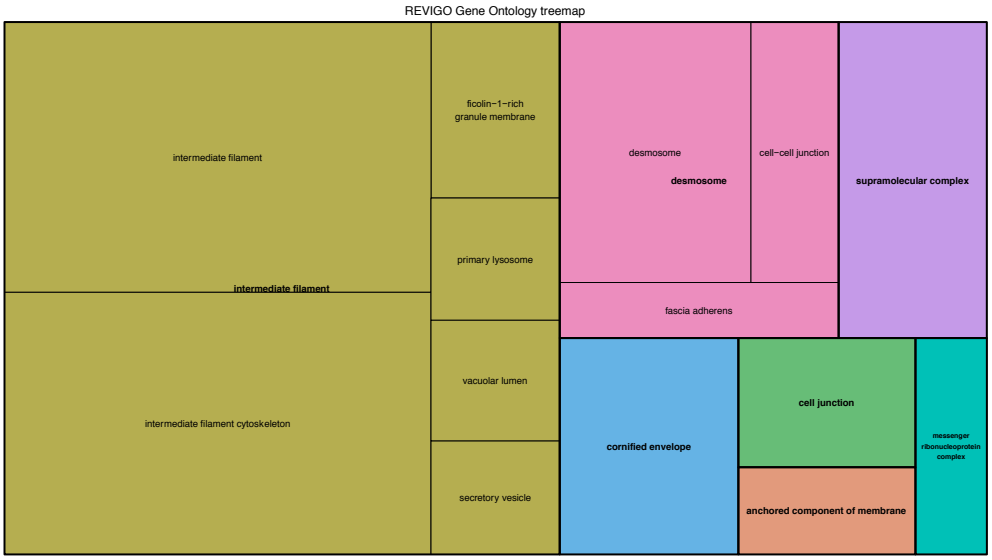
C



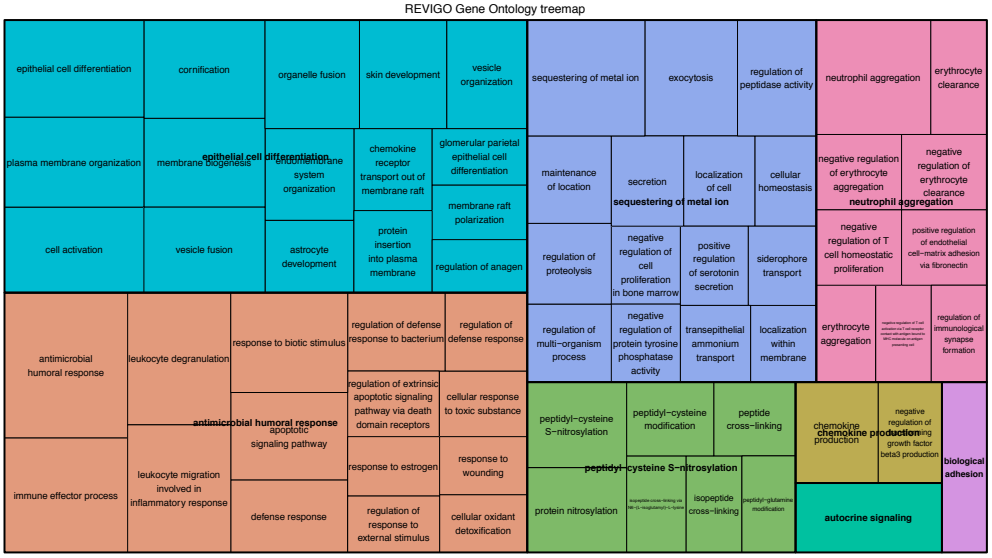
[Figure continues onto next page]

[Figure continued from previous page]

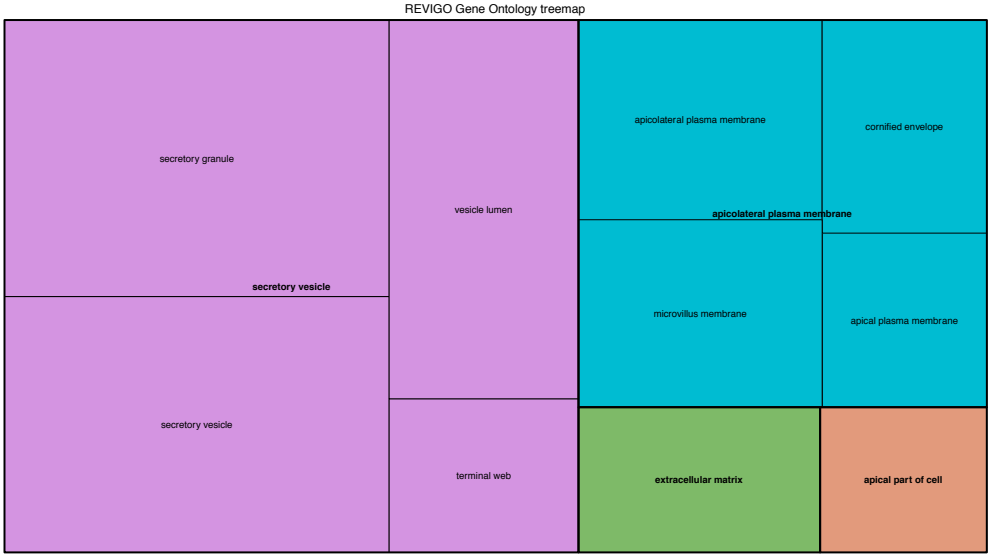
D



E



F



[Legend follows on next page]

[*Legend for figure on previous page*]

Figure 6.9. Gene ontology summary of the biological processes involved in the top 20 genes characterising each layer. Gene lists available in

Table **9**.REVIGO summarised statistically significant biological process (A, C, E) and cellular compartment (B, D, F) annotations of the top 20 t-test ranked genes of **[A, B]** basal versus spinous and granular clusters, **[C, D]** spinous versus basal and granular clusters, and **[E, F]** granular versus basal and spinous clusters.

6.2.4 Keratinocytes challenged with live bacteria reveal single cell changes in transcriptome across all layers of the epidermal model

ScRNA-seq analysis has shown that a mixture of proliferating basal cells, differentiating upper layer cells, and upper epidermis pre-terminal differentiation keratinocytes are present in the RHE model system. Crucially these populations are not lost by the potentially highly cytotoxic bacterial co-culture process nor lost by the disruptive tissue dissociation protocol to obtain single cells for DropSeq encapsulation. Thus, it creates a useful experimental system for investigations of the effect of live *Staphylococcal spp.* challenge to the keratinocyte transcriptional programmes.

To validate the presence of bacteria-induced transcriptional programmes across epidermal layers, embedded sample density for each experimental challenge was projected onto the UMAP plots (**Figure 6.10A**). Further Leiden sub-clustering of the basal and granular clusters at a resolution of 0.5 and spinous sub-clustering at a resolution of 0.2 separated these layer clusters into smaller populations that broadly matched the topography of the density plotting (**Figure 6.10B**).

Separating the clusters by layer annotation type: basal, spinous or granular, and plotting the same challenge density heatmap illustrates layer-specific changes in transcriptome by co-culture with *S. epidermidis* or *S. aureus*. The basal cluster consists of 2559 cells, of which 1061 are from control, 716 are from *S. epidermidis* challenge and 782 are from *S. aureus* challenge. The density plot of the conditions over this layer demonstrate that *S. aureus* infection induces a total change in transcriptome of basal cells compared to sterile control because there is very little overlap in the cells from each of these conditions. Interestingly, basal cells from *S. epidermidis* co-culture show a degree of overlap with sterile control cells (Basal-1) and an overlap with *S. aureus* cells (Basal-3) (**Figure 6.10C, D**). There is a portion of basal keratinocytes that are only *S. aureus* related (Basal-4). Likewise, there is a sub-cluster of basal cells only from the sterile condition (Basal-2). The spinous cluster is composed of 621 cells: 175 sterile control, 328 from SE-challenge, and 118 from SA-

challenge. A shared sterile control and *S. epidermidis* cluster (Spinous-1), an *S. epidermidis* only cluster (Spinous-2) and an *S. aureus* only cluster (Spinous-3) (**Figure 6.10E, F**).

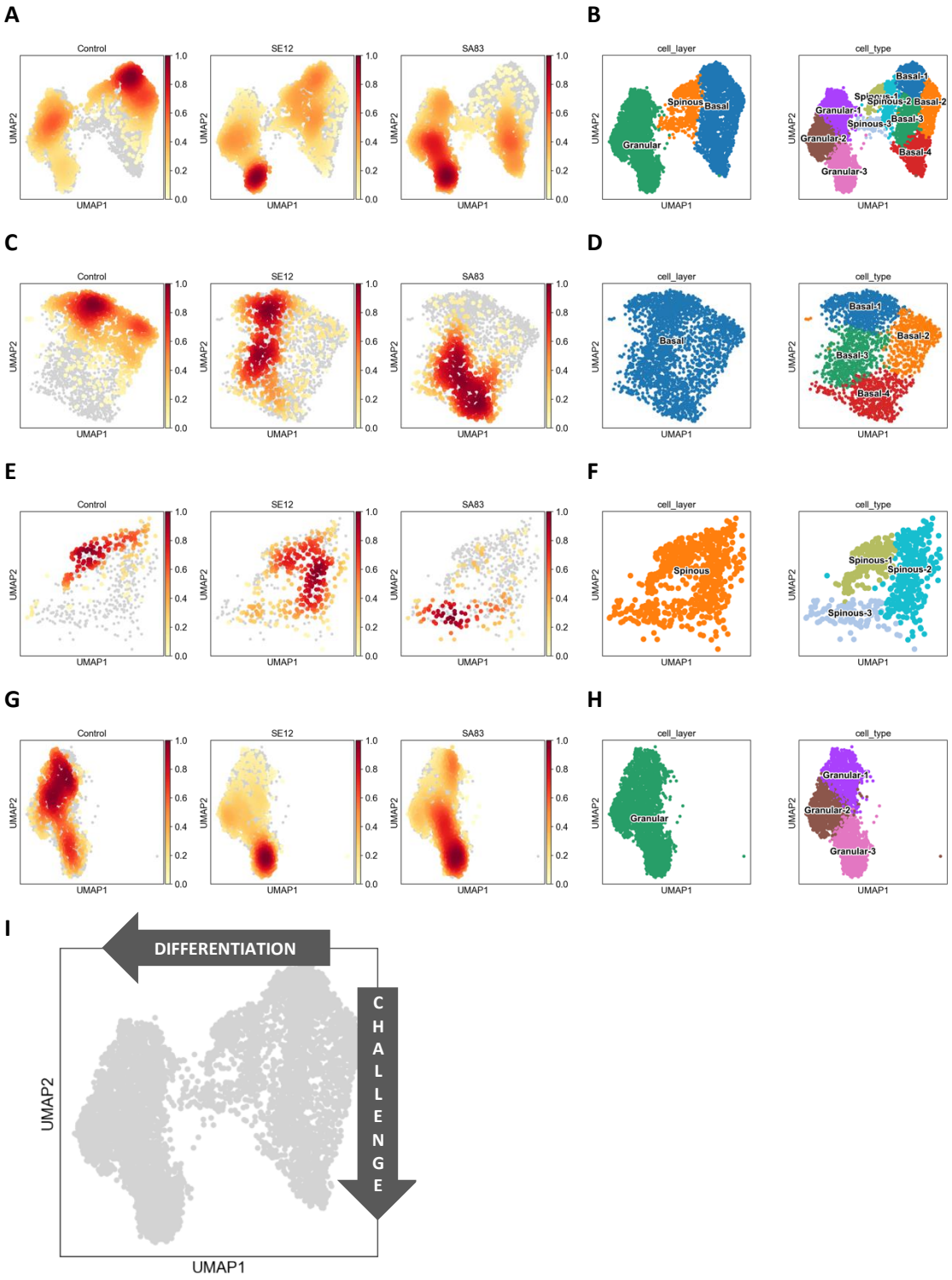


Figure 6.10. Sub-clustering of the epidermal layers by challenge. The embedded density of the challenge conditions was plotted for **[A]** all the cells, **[C]**, basal cells, **[E]** spinous cells and **[G]** granular cells showing control (left), *S. epidermidis* (SE12, middle) and *S. aureus* (SA, right) regions of sub-clusters. These regions matched unbiased clustering for **[B, D]** basal and **[B, H]** granular cells as defined by Leiden at a resolution of 0.5; **[B, F]** spinous cells were subset for further unbiased clustering at a resolution of 0.2. **[I]** Observed trajectories of differentiation and bacterial challenge in order sterile, commensal and pathogenic along the axes of UMAP1 and UMAP2, respectively.

The granular cluster is made up of 3211 cells, approximately half of all the cells presented. The majority of the granular cells, 1419, are from SA-infected samples, while 984 cells are from SE-infected samples, with the remaining 808 cells from sterile control. Like the basal cluster, this granular cluster was clustered by higher Leiden resolution that mirrored the pattern of the density heatmap (**Figure 6.10G, H**). Cells from sterile control appeared relatively uniform across all three granular sub-clusters, with a slight skew to Granular-1 and Granular-2. *S. epidermidis* challenged cells in this cluster were highly associated with Granular-3, while cells challenged with *S. aureus* were split between Granular-2 and Granular-3, with a skew to Granular-3. In sum, Granular-3 appears to be associated with bacterial challenge, regardless of commensal or pathogenic species.

The RHE models have shown to recapitulate the layered biology of human epidermis (**Figure 6.10I**, UMAP1). Furthermore, challenge with commensal or pathogenic *Staphylococcal* species indicated change in the transcriptional programmes of the layers to uncover distinct sub-populations (**Figure 6.10I**, UMAP2). These sub-populations potentially reflect specific biology induced by bacterial challenge, which is forms the remaining investigation of this chapter.

6.2.5 *S. aureus* challenge of epidermis models drives early differentiation in keratinocytes

The challenge protocol involved topical challenge of the models with *Staphylococcal spp.* and, therefore, the transcriptional responses to challenge were investigated in a top-down fashion. The *Staphylococcal spp.* challenge distribution across the granular clusters indicated that the upper-layer keratinocytes are responding differently to bacterial co-culture. This potentially induces divergent differentiation pathways for models infected or left sterile.

The gene ontology annotation of the top 20 genes between the main granular cluster and other two layer-defining clusters was split between epithelial cell differentiation and antimicrobial humoral response (**Figure 6.9E**). Epidermis development related to six of those genes (*IVL*, *KLK7*, *KRT19*, *PLAAT4*, *SPINK5*, *SPRR3*). Antimicrobial humoral response related to also six genes: *KLK7*, *LCN2*, *S100A8*, *S100A9*, *SLPI* and *SPINK5*. The expression of these genes, and therefore, their related process, did not vastly alter between the three granular subclusters (**Figure 6.11**). Only SPINK5 and SPRR3 expression was noticeably expressed in more cells and with a higher expression in Granular-2.

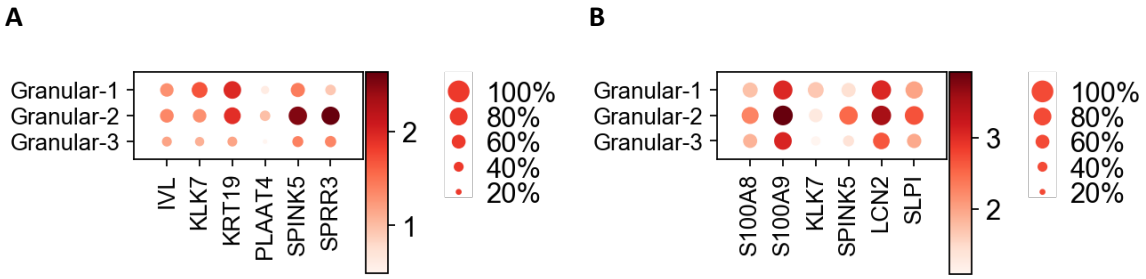


Figure 6.11. Granular layer expression of genes related to epidermis development and antimicrobial response. [A] Expression of genes associated with biological process term 'epidermis development', GO: 0008544 across the three granular clusters; **[B]** Expression of genes associated with biological process 'antimicrobial humoral response', GO: 0019730; found to be upregulated in granular layer compared to a t-test comparison against basal and spinous layers.

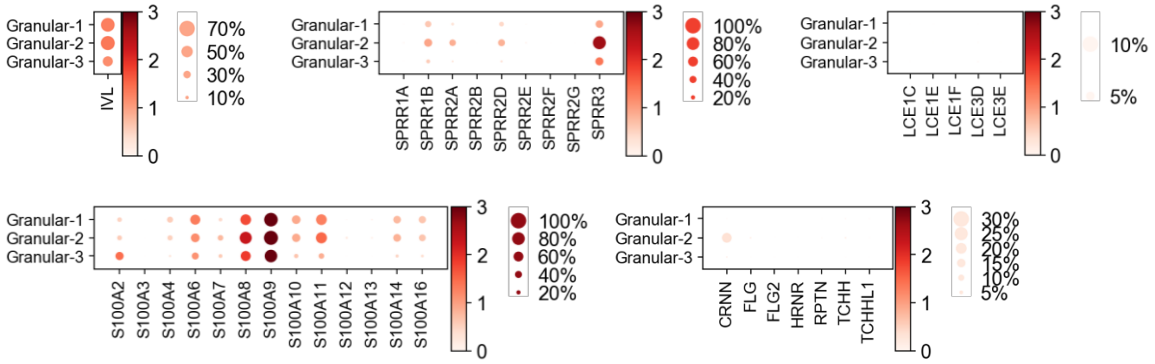


Figure 6.12. Expression of the Epidermal Differentiation Complex of genes across the granular sub-clusters. The list was sub-divided into cornified envelope precursor genes (*IVL*, *LOR*); small proline-rich (SPRR) proteins (*SPRR1A*, *SPRR1B*, *SPRR2A*, *SPRR2B*, *SPRR2C*, *SPRR2D*, *SPRR2E*, *SPRR2F*, *SPRR2G*, *SPRR3*, *SPRR4*; late cornified envelope (LCE) proteins: *LCE1A*, *LCE1B*, *LCE1C*, *LCE1D*, *LCE1E*, *LCE1F*, *LCE2A*, *LCE2B*, *LCE2C*, *LCE2D*, *LCE3A*, *LCE3B*, *LCE3C*, *LCE3D*, *LCE3E*, *LCE4A*, *LCE5A*, *LCE6A*, *LCE7*; the S100 family: *S100A1*, *S100A2*, *S100A3*, *S100A4*, *S100A5*, *S100A6*, *S100A7*, *S100A8*, *S100A9*, *S100A10*, *S100A11*, *S100A12*, *S100A13*, *S100A14*, *S100A15*, *S100A16*; and the S100 fused-type protein (SFTP) family: *CRNN*, *FLG*, *FLG2*, *HRNR*, *RPTN*, *TCHH*, *TCHHL1*. Not all EDC genes were expressed in the DropSeq data and these are not shown on the dot plots.

Many of these genes from the above lists are part of the wider epidermal differentiation complex (EDC) (Volz *et al.*, 1993; Mischke *et al.*, 1996). To further investigate the differentiation status of the granular clusters, the expression of the genes comprising the epidermal differentiation complex was observed (**Figure 6.12**).

To observe any such change in the expression of the EDC panel of genes across the granular clusters and in context of the rest of the RHE cells, the average expression of the panel was compared to the expression of a randomly selected panel of genes as described by Satija *et al.* (2015). As expected, the expression of the EDC panel in the granular subset was high and uniform (**Figure 6.13A**), in order to find areas of relatively high score, the lower bound of the plotting scale was set to 0.5 (upper bound set to 1.3 for both plots) (**Figure 6.13B**). This revealed an area of modestly higher score for the panel of differentiation genes, roughly corresponding to Granular-2. Interestingly, and against expectation, this did not coincide with the area of highest density of the *Staphylococcal*-challenged cells (Granular-3), perhaps indicating that bacterial co-culture is not the main driver of differentiation or cornification.

To investigate the effect of bacterial challenge upon keratinocyte transcriptomic differentiation programmes the expression score analysis for the panel of EDC genes was expanded to the whole dataset of cells (**Figure 6.13C**). This showed a second cluster of cells with a high score for this panel, which aligned with the annotated Spinous-3 cluster. Within the spinous population of keratinocytes, *S. aureus* challenged samples were almost exclusively found in Spinous-3. This phenotype was not shared by the sterile control keratinocytes or *S. epidermidis* challenged in the alike spinous populations 1 and 2.

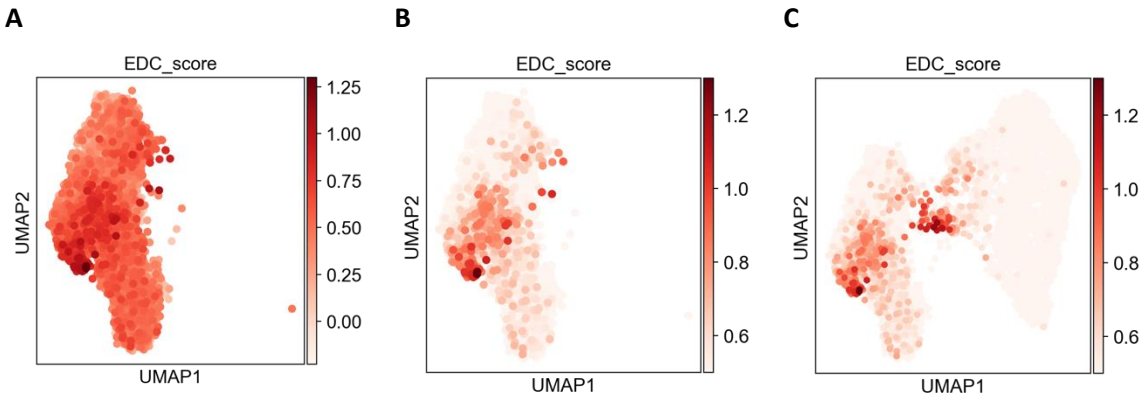


Figure 6.13. Expression score for the panel of EDC genes. For each cell the average expression of the EDC genes is compared to a random selection; **[A]** raw score in granular cells; **[B]** legend lower bound set to 0.5 for improved interpretations in granular cells; **[C]** as B but shown across all cells.

6.2.6 Bacterial challenge of epidermis modes corresponds to transcriptional changes in deeper layers

Investigating differential gene expression using the MAST algorithm, that relies on linear models accounting for the bimodal nature of single cell data, was performed for within layer comparisons (Finak *et al.*, 2015). As described above, the distribution of keratinocytes from sterile control or *Staphylococcal* challenge within each layer was revealed by density plotting and recapitulated by unsupervised clustering using Leiden. This demonstrated that these co-culture conditions altered the keratinocyte transcriptome in a layer-specific manner, which, by using MAST differential expression analysis, could be formally investigated (**Table 10**)

Differential expression analysis between the three granular clusters did not reveal any qualitative change in the transcriptomes of these clusters relating to a biological change (**Figure 6.14**). Comparing Granular-1 with Granular-2 produced no up-regulated genes in the former and only six upregulated genes in the latter (**Figure 6.14A**). Two of these genes, *S100A9* and *SPRR3* are part of the EDC, while *ECM1*, an extracellular matrix encoding gene, is found near to the chromosomal locus of the EDC and is also involved in keratinocyte differentiation (Smits *et al.*, 2000). *SPINK5*, a serine protease inhibitor, is expressed more highly in Granular-2 compared to Granular-1, and its expression regulates the desquamation of stratified epithelial tissues through degradation of corneodesmosomes cadherins (Wang *et al.*, 2014). These revealed a slight increase in the differentiation phenotype of Granular-2 as discussed above.

The comparison of Granular-2 against Granular-3 also revealed a similar pattern of increased expression of epidermal differentiation (**Figure 6.14C**). The five genes upregulated in Granular-2 compared to Granular-1, were also upregulated in comparing Granular-2 to Granular-3. Additionally, a further six genes were upregulated: *CEACAM6*, *KRT19*, *LCN2*, *LGALS3*, *RHCG* and *SLPI*. *CEACAM6* is a member of the carcinoembryonic antigen family of immunoglobulin glycoprotein cell-adhesion molecules reported to be overexpressed in a number of squamous cell carcinomas, but its role in epidermal differentiation is unknown (Cameron *et al.*, 2012). *LCN2* is a bacterial siderophore and its overexpression in Granular-2 over Granular-3 potentially indicates dysregulation of anti-microbial activity of the *Staphylococcal*-challenged keratinocytes in Granular-3 (Yang *et al.*, 2002). *LGALS3* is reported to promote epithelial inflammation via IL-1β and negatively associates with psoriatic lesions (Shi *et al.*, 2018; Uchino *et al.*, 2018). The increased expression of *LGALS3* in Granular-2 compared to the infected cluster could indicate a dysregulation of innate immunity within challenged terminally differentiated keratinocytes, or microbiome interactions in this cluster that promote a psoriatic lesional phenotype. *RHCG* is an ammonia transporter protein previously found to be expressed in differentiated granular epidermis by single-cell analysis (S.

Wang *et al.*, 2020). *SLPI* is a protease inhibitor reported to be expressed in human epidermis, with activity in the cornified layer against LPS (Jin *et al.*, 1997; Wingens *et al.*, 1998).

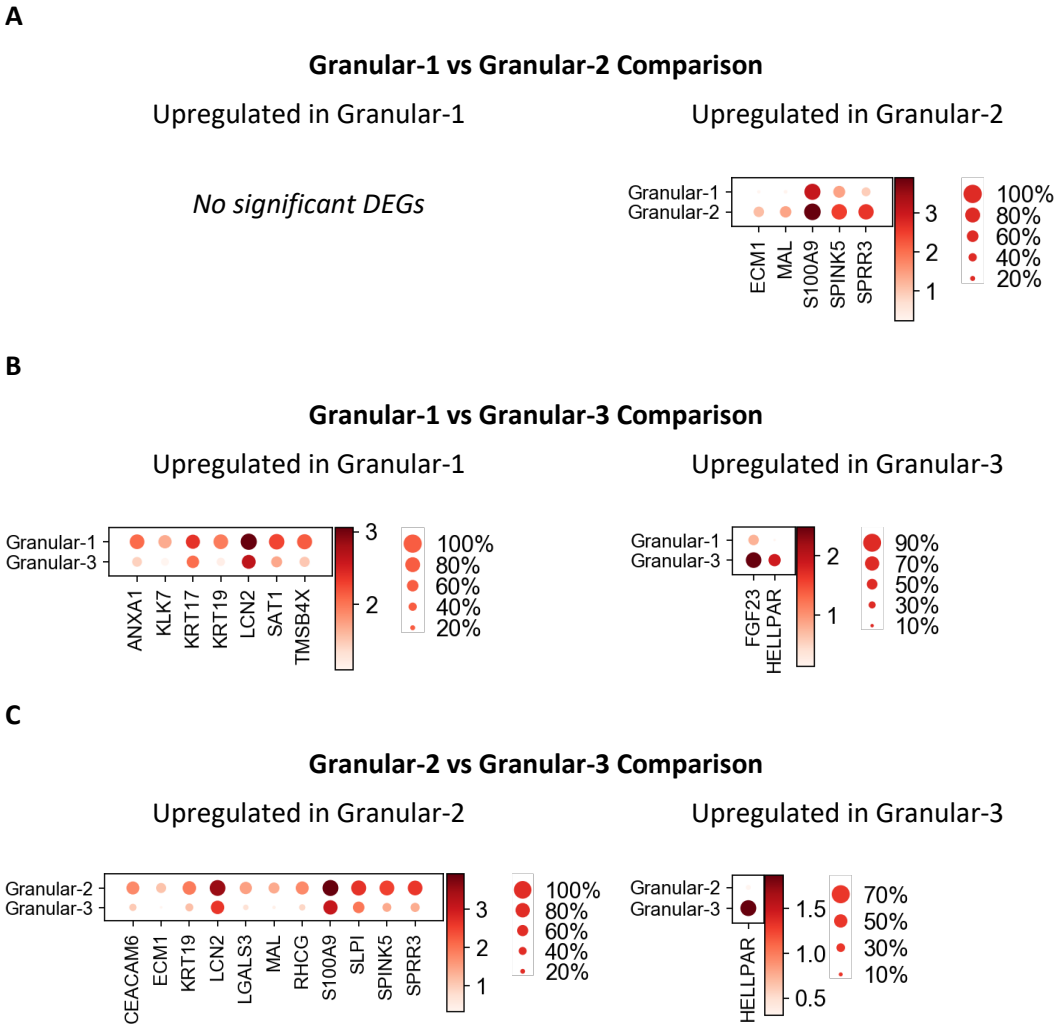


Figure 6.14. Differentially expressed genes between granular sub-clusters. MAST was used to define genes differentially expressed between **[A]** Granular-1 and Granular-2, **[B]** Granular-1 and Granular-3, **[C]** Granular-2 and Granular-3. Statistical significance for differential expression was BH p-value <0.01 and logFC >±1.

Genes found to be upregulated in Granular-3 compared to Granular-1 (**Figure 6.14B**) or Granular-2 (**Figure 6.14C**) are *FGF23* and *HELLPAR*. The former is a fibroblast growth factor protein shown to induce late differentiation in HaCaT keratinocytes, while the role of *HELLPAR* in the epidermis remains unknown (Wu *et al.*, 2018).

The spinous subclusters were phenotypically distinguished broadly by association with sterile control or *S. epidermidis* (Spinous-1), *S. epidermidis* alone (Spinous-2) and *S. aureus* alone (Spinous-3). This could be interpreted as forming clusters of sterile, commensal and pathogenic challenge,

respectively. Comparisons of the upregulated genes in the sterile Spinous-1 cluster versus both Spinous-2 and Spinous-3 revealed a general maintenance of normalise keratinization expression in these cells (**Figure 6.15A, B**). There was very little difference in the gene expression upregulated in Spinous-2 compared to Spinous-1, likely as a result of the commonality of *S. epidermidis* challenge between these clusters. Only three genes were significantly upregulated in Spinous-2 over Spinous-1: *FTL*, *MMP1* and *SAT1*.

Sterile versus pathogen and commensal versus pathogen associated spinous cluster comparisons proved very similar. Of the 27 genes upregulated in the pathogenic Spinous-3 cluster versus control (**Figure 6.15A**), 19 were also upregulated in commensal versus pathogen comparison. This indicates that the genes upregulated by *S. aureus* co-culture of the models are unique to the pathogenic niche of this bacterium compared to its more commensal counterpart, *S. epidermidis*. The top gene ontology terms for biological process describing these 19 genes are antimicrobial humoral response (GO: 0019730, BH p-value = $8.711E^{-7}$) and epidermis development (GO:0008544, BH p-value =$2.539E^{-6}$), indicating a microbe-induced differentiation.

The other genes upregulated in Spinous-3 compared to Spinous-2 were also highly related to epidermal barrier formation and anti-microbial homeostasis (**Figure 6.15C**). These included EDC genes such as *IVL* and *SPRR1B*, and associated barrier genes such as *DMKN*, *KRT6B*, and *SBSN*, and the anti-microbial serpin, *SERPINB3*. Indeed, the top biological process ontology term for these genes; and *LYPD*, whose epidermal function is not understood, is cornified envelope assembly (GO:1903575, BH p-value= $5.764E^{-6}$). The high expression of *SERPINB3* in the *S. aureus* infected cells is interesting because it is a known contributor to early epidermal dysfunction and inflammation of atopic dermatitis (Sivaprasad *et al.*, 2015).

The remaining eight genes that were unique to the sterile against pathogenic comparison, being overexpressed in the latter, were associated with inflammatory processes: *CD24*, *IL1RN*, *SERPINB1* and *NEAT1* (Hui *et al.*, 2014; Lundberg *et al.*, 2015; Tan *et al.*, 2016; Zhang *et al.*, 2019; Martin *et al.*, 2020). *UUP1* was found to be upregulated in epidermis following UV-B exposure and allergic contact dermatitis (Alameda *et al.*, 2016). *MMP1*, expressed here in *S. aureus* challenged cells, is known to be expressed by wound keratinocytes in response to exposure to lipoteichoic acid (Brauweiler *et al.*, 2015).

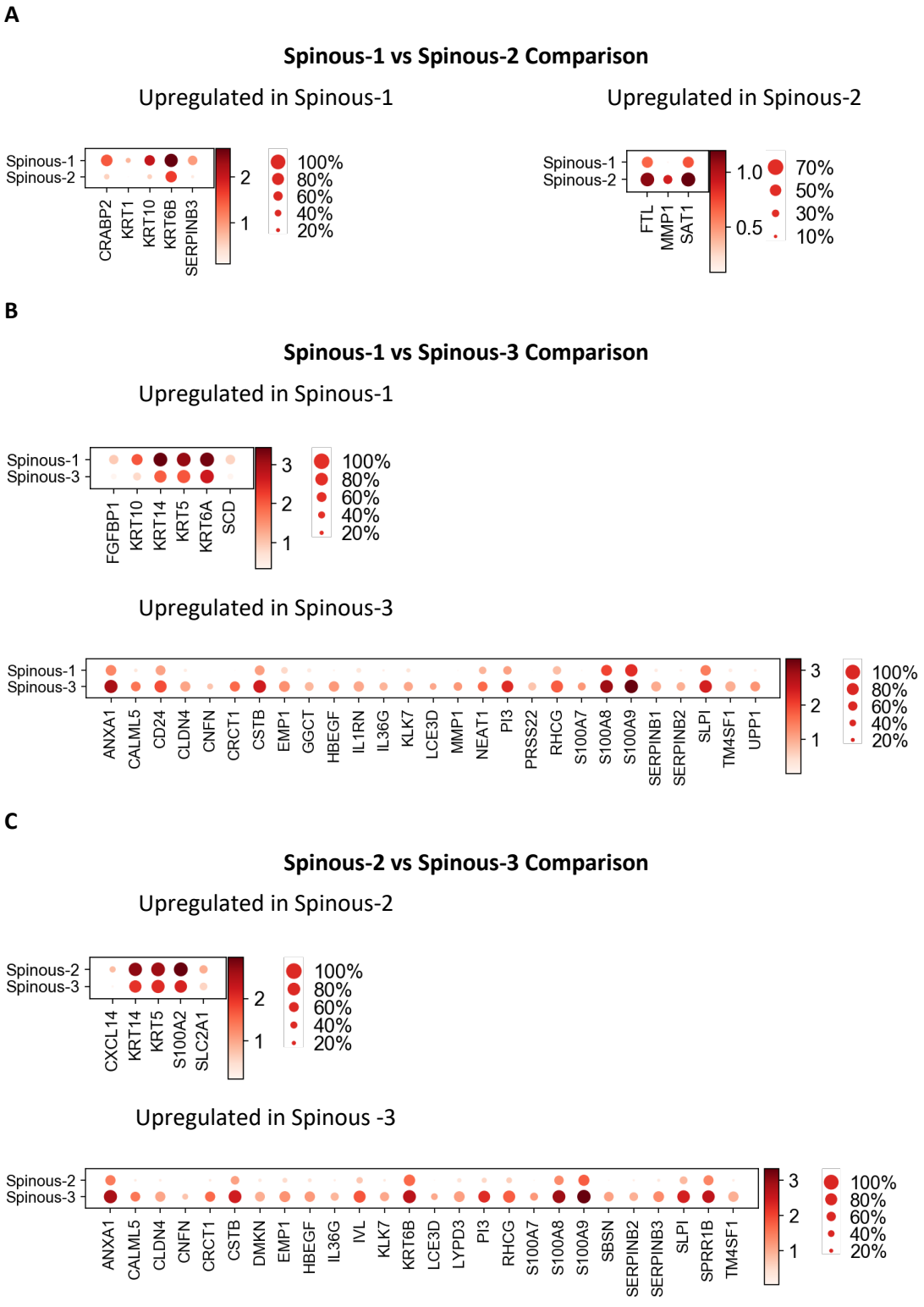


Figure 6.15. Differentially expressed genes between spinous sub-clusters. MAST was used to define genes differentially expressed between **[A]** Spinous-1 and Spinous-2, **[B]** Spinous-1 and Spinous-3, **[C]** Spinous-2 and Spinous-3. Statistical significance for differential expression was BH p-value <0.01 and logFC >±1.

The phenotype of basal keratinocytes as revealed by unsupervised Leiden clustering was reflective of their bacteria challenge condition. The experimental procedure involved bacterial co-culture on the apical aspect of the stratified models, where the greatest change in the keratinocyte phenotype was predicted in those cells closest to the cornified layer. Instead, the basal compartment shows the clearest alteration in transcriptome arising from microbial challenge.

The control condition strongly associated with the Basal-2 cluster, giving it a putative sterile phenotype. Control keratinocytes were also found in Basal-1, along with a proportion of the *S. epidermidis* challenged cells. Only four genes were found to statistically upregulated in either cluster: *MALAT1*, *MT-RNR2*, *NEAT1* and *CXCL14* (**Figure 6.16A**). Their role does not indicate any major perturbation in biology between the two clusters, despite the association of Basal-1 with commensal infection.

There was no cluster that uniquely associated with *S. epidermidis* alone. Basal-3 was composed of cells exposed to both *Staphylococcal spp.* The genes upregulated in the sterile cluster compared to this shared infection-related basal cluster were associated with typical basal biology, such as *KRT5* and *CXCL14* (**Figure 6.16B**). *DST* is a basal keratinocyte related protein by interaction with *COL17A1*, stabilising hemidesmosome anchoring to the basal lamina connecting to the dermis (Hopkinson and Jones, 2000). According to gene ontology analysis, the genes upregulated in the shared infection cluster, Basal-3, compared to Basal-2, were enriched for a response to wounding (GO:0009611, BH p-value = $2.626E^{-2}$). The genes responsible for this annotation were *ANXA1*, *HBEGF*, *LGALS1*, *PLAU* and *ZFP36L1*.

Basal-4 was highly associated with samples challenged with *S.* aureus. Genes statistically significantly upregulated in the sterile Basal-2 cluster against this cluster were again related to normal basal processes and intermediate filament formation of the basal cytoskeleton: *KRT5*, *KRT14*, *KRT17*, *KRT6A*, *CXCL14* and *DST* (**Figure 6.16C**). This indicates that the normal basal programmes are highly downregulated or perturbed pathogenic infection.

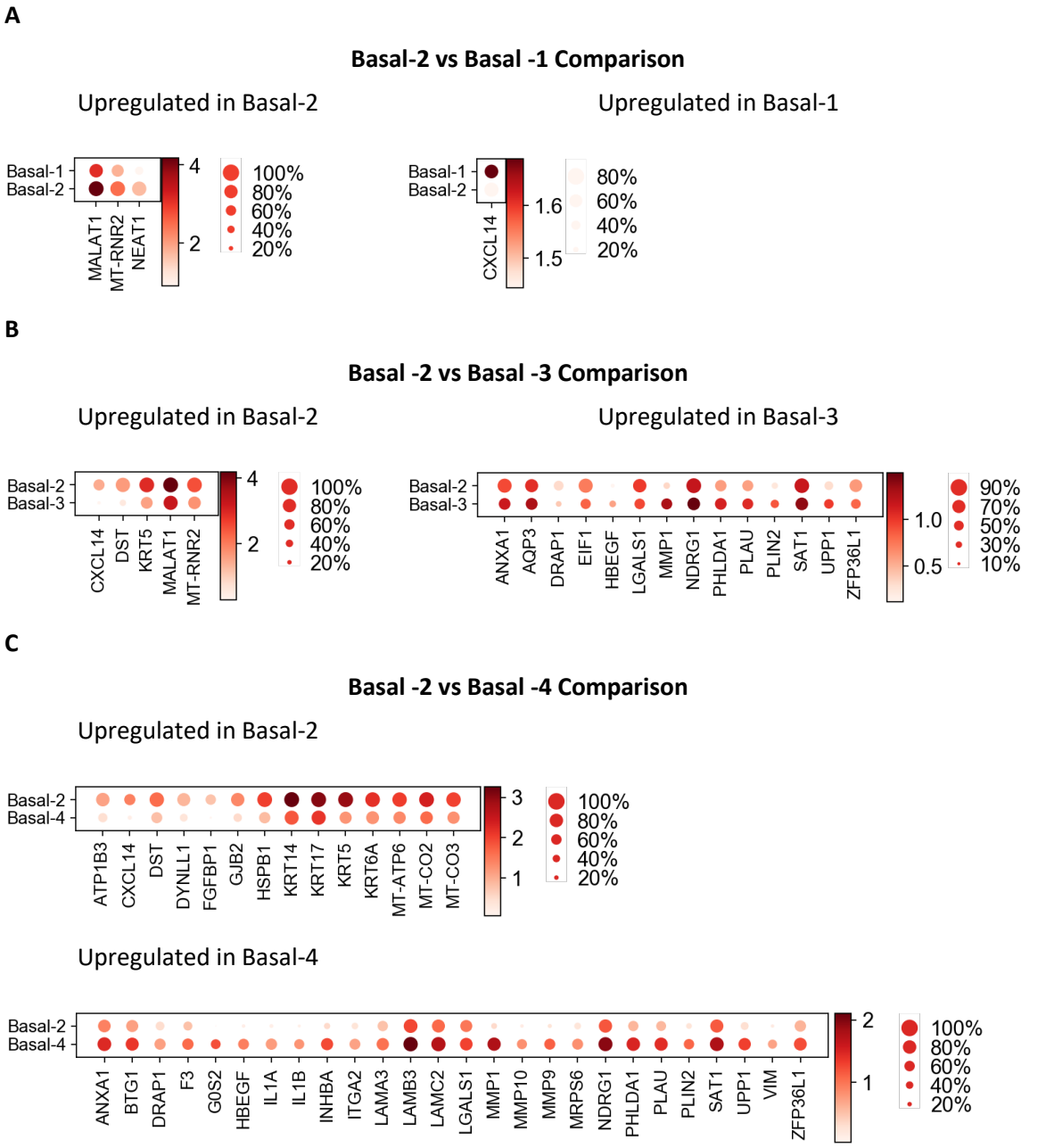


Figure 6.16. Differentially expressed genes between basal sub-clusters. MAST was used to define genes differentially expressed between **[A]** Basal-2 and Basal-1, **[B]** Basal-2 and Basal-3, **[C]** Basal-2 and Basal-4. Statistical significance for differential expression was BH p-value <0.01 and logFC >±1.

The genes upregulated in Basal-4 compared to Basal-2 form the bulk of the transcriptomic response to pathogen exposure (**Figure 6.16C**). In all, 26 genes were upregulated in this condition. Gene ontology analysis revealed biological processes invoked by these genes are highly associated with tissue development ("anatomical structure formation involved in morphogenesis", GO:0048646, BH p-value= $8.17E^{-07}$), and homeostasis ("regulation of cell proliferation", GO:0042127, BH p-value = $6.35E^{-07}$; "response to wounding", GO:0009611, BH p-value = $2.49E^{-05}$) (**Figure 6.17**). Indeed, the

almost exclusive co-expression of a number of these genes in Basal-4, for example the matrix metalloproteinases-1, -9, and -10 and laminin genes A3, B3 and C2 highlights the epidermal proliferation and re-epithelialisation expected of models infected by *S. aureus.*

This response by basal keratinocytes to events occurring in the upper layers of the model would require either bacterial invasion of the stratified epidermis, sensing of bacterial secreted factors and/or keratinocyte paracrine signalling. The data suggests that bacterial contact drives keratinocyte differentiation to a barrier-maintaining corneocyte phenotype, and the response by basal keratinocytes is a result of host cell signalling. The data have shown in the spinous cells from models infected with *S. aureus* that the expression of the panel of EDC genes is upregulated and alike granular cells. In basal cells from the same condition, there is an overexpression of IL-1α, and IL-1β. These cytokines are rapidly released from keratinocytes in response to environmental stress and activated cutaneous regeneration and pro-inflammatory pathways (Werner and Smola, 2001). In this sense, IL-1α, and IL-1β act as epidermal alarmins to generate the protective proliferative phenotype of this basal compartment. Interestingly, IL1RN, an IL-1 receptor antagonist, is highly expressed in spinous cells from models infected with *S.* aureus. Its function may be in potentially blocking the IL-1 alarmin signalling from basal keratinocytes and switching spinous cells from a proliferating to differentiating program (Martin *et al.*, 2020).

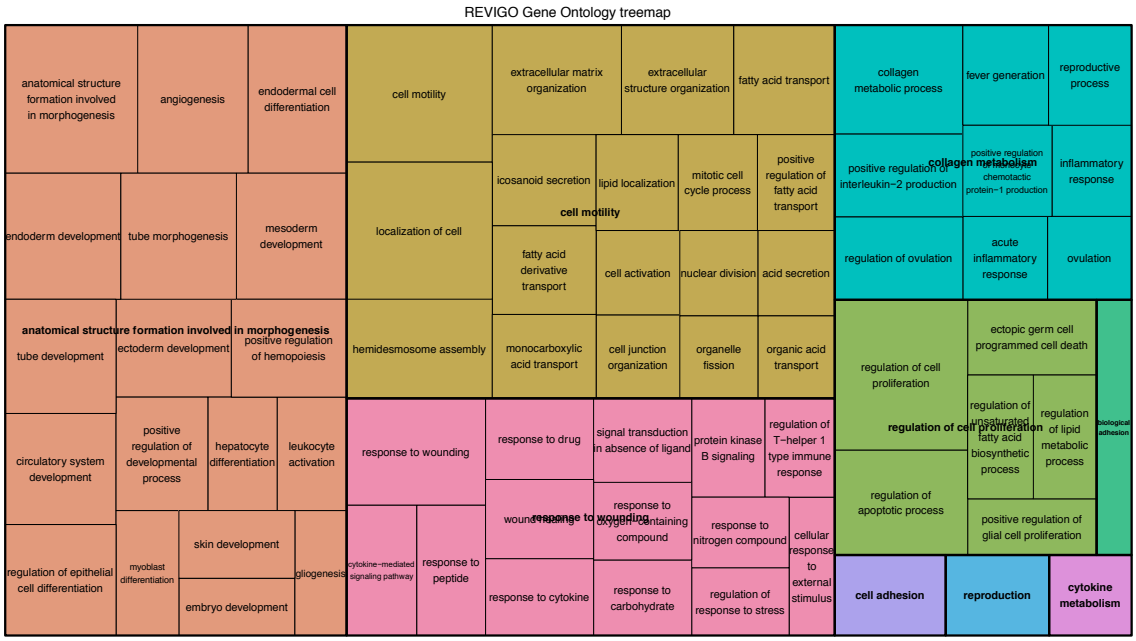


Figure 6.17. Gene ontology summary of the biological processes involved in the upregulated genes of Basal-4 compared to Basal-2. REVIGO summarised statistically significant biological process invoked by the 26 genes upregulated in *S. aureus* challenge associated Basal-4 cluster compared to sterile Basal-2 cluster.

6.2.7 Bacterial challenge expands keratinocyte development

Partition-based graph abstraction (PAGA) is used to predict linear trajectories in single-cell data while preserving the overall manifold structure of the reduced dimension plotting (Wolf *et al.*, 2019). The single-cell data was separated according to challenge condition for PAGA analysis. Only those keratinocyte clusters associated with the challenge, as observed by embedded density plotting, were retained for each analysis (**Figure 6.18**). This removed from analysis clusters that were not highly associated with each challenge to simplify the graph abstraction of the connectivity landscape.

PAGA analysis does not give direction to the connectivity map, but direction can be inferred from known biology to give reasonable interpretation of the pseudo-temporal order. The sterile control analysis is useful for determining the starting point of the connectivity and the route of the connections in an unperturbed context (**Figure 6.18A**). Basal-1 is inferred as the starting point from which the map arises, while previous investigation of the EDC genes indicated that the terminus of the graph lies in Granular-2. The control PAGA plot shows a directional lineage from Basal-1 to Basal-2, reflecting a proliferating compartment. From there a split occurs forming the spinous layer and early granular layer, Granular-3. Spinous-2 connects to Granular-1 and Granular-3 indicating the early formation of this layer. Both Granular-1 and Granular-3 connect heavily to Granular-2 reflecting the strong transition of these cells to a well-defined granular and late-stage differentiation phenotype. In an undisturbed setting the major connectivity of the graph represents differentiation rather than proliferation as the main process of the models.

The graph abstraction for inferring trajectory in *S. epidermidis* challenged models demonstrates a qualitative shift in connectivity weighting from formation of the granular layer to stronger connections between basal and spinous layers (**Figure 6.18B**). The strong connection between Spinous-1 and Spinous-2 reflects the crosstalk between the normal spinous phenotype of Spinous-1 cells, as marked by their higher expression of KRT1 and KRT10, and the remodeling/wound phenotype of cells in Spinous-2 as evidenced by their increased expression of matrix metalloproteinases. The lack of a connection of the healthy Spinous-1 cluster to basal cells potentially indicates that a shift to remodeling spinous phenotype is driven by instruction from the underlying basal cells, where both Basal-1 and Basal-3 connect to Spinous-2. The molecular basis for this instruction remains to be elucidated.

The trajectory inference of the *S. aureus* challenged cells reveals a two-part process (**Figure 6.18C**). The EDC of genes was shown to be upregulated in Spinous-3, indicating the differentiation of these cells that is comparable to Granular-2. The weighting of the connection between spinous and

granular parts of the graph indicates an overall increase in transitions, which can be interpreted as an increase in differentiation of the keratinocytes in these compartments. It is biologically reasonable that pathogenic challenge of the models would result in upper layers increasing differentiation in an effort to maintain functioning epidermal barrier. The connection of Basal-3 to Basal-4 is also heavily weighted and reflects the increase proliferation of the basal compartment, where Basal-4 cells are highly overexpressing tissue proliferative/remodeling genes such as MMPs and laminins, and IL1 alarmins. Basal-4 is highly associated with *S. aureus* infection, while Basal-3 is associated with both infection of either *Staphylococcal spp.* challenge, which indicates a hierarchy of reaction to challenge in favour of Basal-4. The single connection arising from Basal-3 points to this cluster as being the starting cluster for keratinocyte proliferation and may explain its slightly weaker association with pathogenic co-culture. There is an overall increase in connectivity weighting in the *S. aureus* plot compared to sterile control, indicating an increase in both proliferation and differentiation in response to pathogenic challenge.

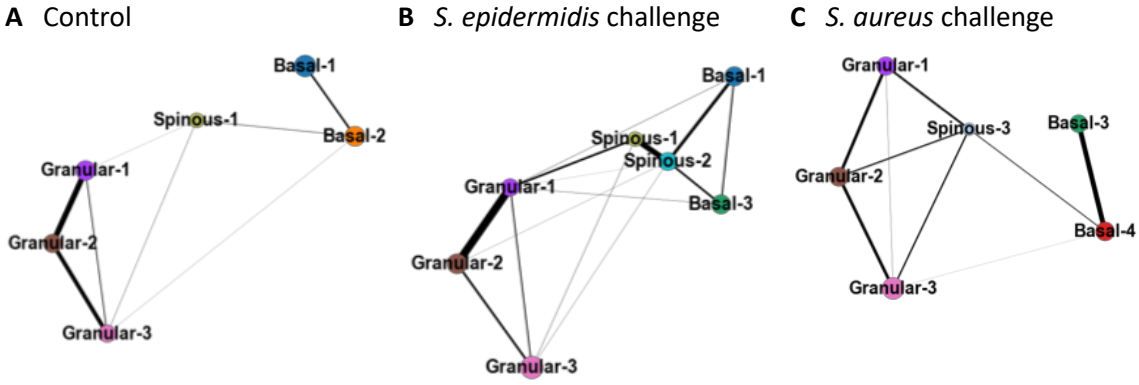


Figure 6.18. Partition-based graph abstraction (PAGA) of RHE sc-RNA-seq data by challenge condition. The integrated single-cell data was restricted by filtering out clusters not associated with **[A]** sterile control, **[B]** *S. epidermidis* challenge or **[C]** *S. aureus* challenge to reduce complexity of the connectivity map and observe only relevant trajectories.

6.3 Discussion

6.3.1 Single cell analysis of a reconstituted epidermal model challenged with live bacteria

Keratinocytes almost exclusively form the stratified and squamous epithelium of the epidermis that is continually exposure to the outside environment. Consequently, the epidermis is a unique niche for facultative and mutually beneficial commensal colonisation while maintaining effective defence against pathogen colonisation (Bitschar *et al.*, 2017). Keratinocyte transcriptomic responses differentiating between commensal or pathogen exposure remain to be elucidated. Added complexity comes from the proliferation to differentiation process experienced by keratinocytes as they mature upwardly through the epidermis. These background transcriptomic responses, and the increasing likelihood of upper layer cells encountering microbes and their molecules is predicted to contribute to host orchestration of microbial activity (Fyhrquist *et al.*, 2019).

In this chapter, the alteration of keratinocyte transcriptional programmes by interaction with microbiota has been investigated in the context of reconstituted epidermis models. This experimental system recapitulates epidermal development, where different populations of keratinocytes undertaking proliferation, differentiation, and indeed terminal differentiation to form the outermost cornified layer, forming the protective barrier against the environment. Computational analysis alongside single-cell resolution of the models has demonstrated the layered composition of our studied model, and crucially indicates that either bacterial challenge and/or dissociation procedure has not eliminated populations of interest.

A total of 6391 cells were analysed from three challenge conditions each contributing approximately 2000 cells. This proved to be a sufficient number of cells to investigate the layer-specific changes upon live bacterial challenge of the models. Bacterial challenge was applied topically to the model to mimic natural colonization of the epidermis, and for a duration of 24 hours to imitate persistent infection.

Whilst a change in keratinocyte transcriptional programmes upon commensal or pathogenic challenge was expected, it was unclear whether such a change would dominate the differences between keratinocytes isolated from different epidermal layers. The crosstalk between these, potentially competing, phenotypes of keratinocytes was observed from the gross structure of the UMAP landscape and unsupervised community clustering. The main regions of the UMAP plot correspond to the well described basal, spinous and granular morphologies of keratinocytes, while, within these, sub-populations emerged that were distinct to challenge experiments. The distinct clustering of layer keratinocytes to regions of sterile-only, sterile/commensal, commensal-only

commensal/pathogen or pathogen-only associations revealed microbial interactions altered keratinocyte gene expression through the whole tissue and not just in the upper layers experiencing contact with live bacteria.

The lack of a transcriptomic differentiation programme attributable to bacterial challenge in the upper layers of the models was surprising, but differentiation profiles were increased in deeper spinous layers of models exposed to *S. aureus*. This indicated that there may be differentiation processes within keratinocytes that is independent of direct contact or exposure to the live bacteria. This would involve exposed keratinocytes signalling to deeper, undifferentiated keratinocytes to: (1) differentiate early; or (2) proliferate to maintain a population of keratinocytes for wound healing and maintenance of the epidermis.

6.3.2 Disturbance of the spinous transcriptome of *S. aureus* challenged models

The results presented in this chapter demonstrate that pathogenic challenge did not significantly impact differentiation status or progress in the granular layer. Consequently, very little difference attributable to microbial challenge was found in comparing the three clusters of upper layer granular cells. The only major difference to distinguish the three clusters was the increased expression of EDC mRNAs in one of the granular clusters. An investigation by others of topically applied or media supplemented infection – the latter to model barrier breach by bacteria – found that topical application of *S. aureus* did not alter expression of a small panel of differentiation markers of *CLD*, *FLG*, *OCCL* and *TG1* (Duckney *et al.*, 2013), in line with the findings presented above.

Interestingly, this current analysis discovered instead a population of spinous cells that share a differentiation expression phenotype of the defined EDC panel with granular cells. This small spinous population may have been lost in previous bulk transcriptomic studies because the differentiation phenotype may have been assumed to come from the granular compartment Additionally, pathway inference analysis using PAGA showed increased connectivity of the spinous layer to the granular layers in response to both commensal and pathogen challenge reflecting the more closely related transcriptome of these layers compared to sterile models.

Alongside this increased differentiation phenotype, the spinous population from *S. aureus* challenged models downregulated expression of normal spinous markers *KRT1* and *KRT10*. Previous studies on the effect of *S. aureus* challenge of epidermal models demonstrated reduced differentiation of keratinocytes as observed by decreased expression of differentiation markers *KRT1* and *KRT10* (Son *et al.*, 2014). A similar decrease in *KRT1* and *KRT10* mRNA was detected in spinous cells obtained from models infected with *S. aureus* compared to sterile counterparts. The

findings that this pathogen challenge associated spinous population is both highly differentiated but *KRT1*/*KRT10* are downregulated is an interesting finding given that it is well understood that *S. aureus* impairs epidermal barrier function, but its effect on keratinocyte differentiation is not well understood (Miedzobrodzki *et al.*, 2002; Otto, 2004; Cork *et al.*, 2009).

6.3.3 Basal keratinocyte responses to bacterial challenge indicate regulation of responses of other epidermal layers via IL-1 alarmin signaling

Previous work investigating bulk transcriptome of the RHE models under the same infection conditions investigated in this work demonstrated that *S. aureus* and *S. epidermidis* had similar growth curves when cultured with NHEKs compared to differentiation incompetent HaCaTs (Holbrook, 2019). These data suggest that undifferentiated keratinocytes do not actively inhibit bacterial growth. Therefore, the microbial response of basal keratinocytes is potentially unrelated to direct anti-microbial activity but mediating crosstalk between different epidermal layers.

The response observed by monolayer keratinocytes to exogenous factors secreted by *S. aureus* is a good model for this basal challenge response because monolayer cultures are not fully differentiated, and they lack highly granular and cornified layers central to inhibiting bacterial growth. The cytokines IL-1β, IL-6, TNFα and GM-CSF were found to be overexpressed by monolayer keratinocytes in response to exposure to secreted *S. aureus* factors (Secor *et al.*, 2011). In the RHE model system these cytokines were also all shown to be upregulated at protein level by live *S. aureus* challenge at higher levels of inoculation (Holbrook, 2019).

IL-1α and IL-1β was upregulated by basal keratinocytes in the single-cell analysis presented in this work. While IL-6, TNFα nor GM-CSF were not found to be differentially expressed in the single-cell data, the stringency of MAST analysis may have limited the detection of these. Additionally, the number of cells in this current study might not provide sufficient power to detect these changes. The association of these cytokines in the role of epidermal homeostasis, co-expression in keratinocyte monolayers cultured with conditioned media, and significant protein expression in bulk RHE models challenged with *S. aureus* point to a role in basal/undifferentiated regulation of proliferation.

Basal cells expressing IL-1 cytokines also express genes significant for wounding and tissue remodeling (MMPs, laminins) and are associated with pathogen challenge. One of the downstream effects of IL-1 cytokines is the epidermal production of IL-6, which, in turn, effects a wounding phenotype in expressed tissue and downregulates differentiation keratins. Spinous layer keratinocytes from either commensal or pathogenic *Staphylococcal* challenge show decreased

expression of KRT1 and KRT10 indicative of IL-6 action (Sugawara *et al.*, 2001; Hernández-Quintero *et al.*, 2006). However, while the *S. aureus* associated spinous cluster downregulates expression of keratins associated with differentiation, to link this to a paracrine blockade of differentiation by basal alarmin expression is difficult. This cluster also displays expression of other differentiation markers similar to granular layers, and also expresses an IL1-receptor antagonist that would block this paracrine signaling.

6.3.4 Epidermal paracrine signaling by keratinocytes in response to bacterial challenge

The over-expression of IL-1α and IL-1β in basal keratinocytes from *S. aureus* infected models, and the expression of the IL1 receptor antagonist, IL1RN in corresponding spinous cells presents an interesting paradigm in the role of bacterial-induced inflammation in the epidermis. Cutaneous IL-1 responses have been documented as pivotal in orchestrating keratinocyte AMP production, such as the β-defensins, LL-37 and RNase-7 (Yano *et al.*, 2008; Krishna and Miller, 2012; Cohen, 2014; Wang *et al.*, 2017; Swindell *et al.*, 2018). The response by IL-1 cytokines is activated through keratinocyte sensing of bacterial infection via TLR2 and NOD2 (Hruz *et al.*, 2009; Krishna and Miller, 2012). The expression of the alarmins IL-1α and IL-1β in the basal layer but the expression of their antagonist in the supra-basal spinous layer reveals specific and targeted intra-compartment signaling and inflammation.

As IL-1 over-expression is associated with psoriatic lesions, the expression of its antagonist may indicate the control of this pro-inflammatory cytokine in differentiating layers to prevent aberrant inflammation. It has been reported that deficiency of this receptor antagonist causes uncontrolled inflammation in affected tissues, including skin, due to the unopposed action of IL-1 cytokines (Aksentijevich *et al.*, 2009). In addition, previous reports have indicated that expression of IL36G mRNA occurs in keratinocytes destined to undergo cornification (Lachner *et al.*, 2017), which correlates with the increased differentiated phenotype of this spinous cluster compared to others. Consistent with the findings presented here, expression of IL-1β and IL-36 cytokines by stimulated keratinocytes was found to regulate epithelium morphogenesis and development, and links to their role during psoriatic inflammation arising from a pathological hyperproliferation of keratinocyte populations (Swindell *et al.*, 2018).

The analysis of single-cell data derived from the stratified epidermis-bacteria challenge system presents an opposing hypothesis for the phenotype of spinous cells from *S. aureus* challenged models. Their differentiation may be hindered by basal-derived IL-1 inducing IL-6 responses, while differentiation is promoted by contact with live bacteria. This conflict may arise from the dual exposure of these spinous 'in-between' cells: differentiating in the face of live microbial contact but

physically close enough to basal cells to experience pro-proliferation instructions. The reduced stratification of the model as compared to human skin may have amplified this cross-signalling and presented a population of cells with this complicated trans-phenotype.

Chapter 7 *In-silico* deconvolution of cutaneous microbial responses using signatures from a 3D co-culture model

7.1 Introduction

Previous chapters established a deconvolution pipeline based on the CIBERSORT algorithm for enumerating cellular profiles from skin based on reference gene expression profiles. This pipeline has, so far, been applied to elucidate keratinocyte immunophenotypes and bulk bacteria sensing from healthy and inflamed skin. However, single-cell analysis of the epidermal transcriptome sharply highlights the classic stratified anatomy of this tissue. This stratification is critical to understanding cutaneous functions such as cellular crosstalk, proliferation versus differentiation signals, inflammation and microbiome homeostasis (Joost *et al.*, 2016; S. Wang *et al.*, 2020). However, bulk skin or epidermis analysis loses this granularity by the fact of creating a single gene expression mixture of all the cells in the sample. Alternatively single cell expression studies resolve this problem but there is currently a paucity of such studies (Joost *et al.*, 2016; Cheng *et al.*, 2018; Reynolds *et al.*, 2021). Further, even fewer studies explore the role of the microbiome in altering the transcriptional programmes of the stratified compartments of the epidermis.

The data from Chapter 6 defining clustered populations characterised by epidermal layers and association with commensal or pathogenic challenge will be used as reference signatures for bulk deconvolution of skin biopsies. The same samples of baseline healthy, atopic dermatitis and psoriasis lesional and non-lesional tissue, in addition to the treatment datasets already investigated will be re-investigated in the context of these new profiles.

The previous chapter investigated the molecular events during the differentiation of spatial signatures in relation to bacterial challenge. However, the mechanisms by which *S. aureus* or *S. epidermidis* trigger inflammation, or even keratinocyte evasion/tolerance, in real skin is not fully understood. The epidermis must facilitate colonisation by commensals and microbial symbionts, which together form an intricate and intimate relationship for preventing pathogenic colonisation and the destructive inflammation that arises from this event. Microbiome sensing by keratinocytes is achieved by expression of a variety of pattern recognition receptors including toll-like receptors (TLRs), NOD-like receptors (NLRs), C-type lectin receptors (CLRs) (Bitschar *et al.*, 2017). TLRs, particularly, are critical for keratinocyte interactions with *Staphylococcal* species (Takeuchi *et al.*, 1999; Miller and Modlin, 2007; Lai and Gallo, 2008). TLR2 is the main TLR associated with sensing

S. aureus colonisation and expression of this receptor and TLR2-dependent signalling are reported to be perturbed in atopic dermatitis (Kuo *et al.*, 2013; Kaesler *et al.*, 2014; Skabytska *et al.*, 2016).

Activation of microbe-sensing stimuli induces keratinocytes to express an array of anti-microbial peptides (AMP). AMP expression is induced by exposure of the epidermis to inflammatory cytokines such as IL-1β, TNFα and IL-17 and microbe interaction. In addition, AMPs are associated with epidermal differentiation and demonstrate layer-specific expression (Bitschar *et al.*, 2017). This forms a critical intersection for understanding transcriptomic programs of keratinocyte crosstalk in response to microbial co-culture in the context of cutaneous inflammation and the stratified anatomy of the epidermis.

The interaction of the microbiome with the epidermis extends beyond colonisation of the outer cornified layers. For example, *S. aureus* has been shown to not only colonise the stratum corneum but also penetrate in the nasal epidermis to both the spinous and basal layers (Hanssen *et al.*, 2017). In addition, sensing and regulation of the microbiome and the role of commensal and pathogenic species is altered by cutaneous inflammation (Kong *et al.*, 2012; Zeeuwen *et al.*, 2012; Scharschmidt *et al.*, 2015; Chang *et al.*, 2018). This alteration is further moderated by the stratified nature of the epidermis resulting in layer-specific phenotypes as demonstrated by analysis of single-cell data from a 3D co-culture model in the previous chapter. Consequently, this chapter aimed to dissect the effect of microbial and inflammatory signals onto the specific layers of keratinocytes within bulk disease skin biopsies. By deconvoluting microbe-associated transcriptional signatures from patient data this work sought to progress the investigation of the cellular events underpinning atopic or psoriatic inflammation and responses to treatment. The aim of this is to identify key molecular or cellular switches that appear causal to pathogenesis, differentially response to therapy or delineate patient sub-groups.

7.1.1 Hypothesis

Specific layer-defined subpopulations of keratinocytes expressing transcriptomic programmes attributed to *Staphylococcal* challenge can be tracked by deconvolution in bulk human skin to understand the microbial component of inflammatory skin disease.

7.1.2 Aims

- To identify keratinocyte immunophenotypes in RHE model single cell transcriptomes.

- To deconvolute the effect of microbial signalling from bulk human epidermis using single-cell-derived bacteria challenged RHE model populations.

- To track changes in keratinocyte responses to microbial signalling during atopic dermatitis and psoriatic inflammation and in response to treatment.

- To understand crosstalk between the microbiome and keratinocyte layers underpinning cutaneous homeostasis.

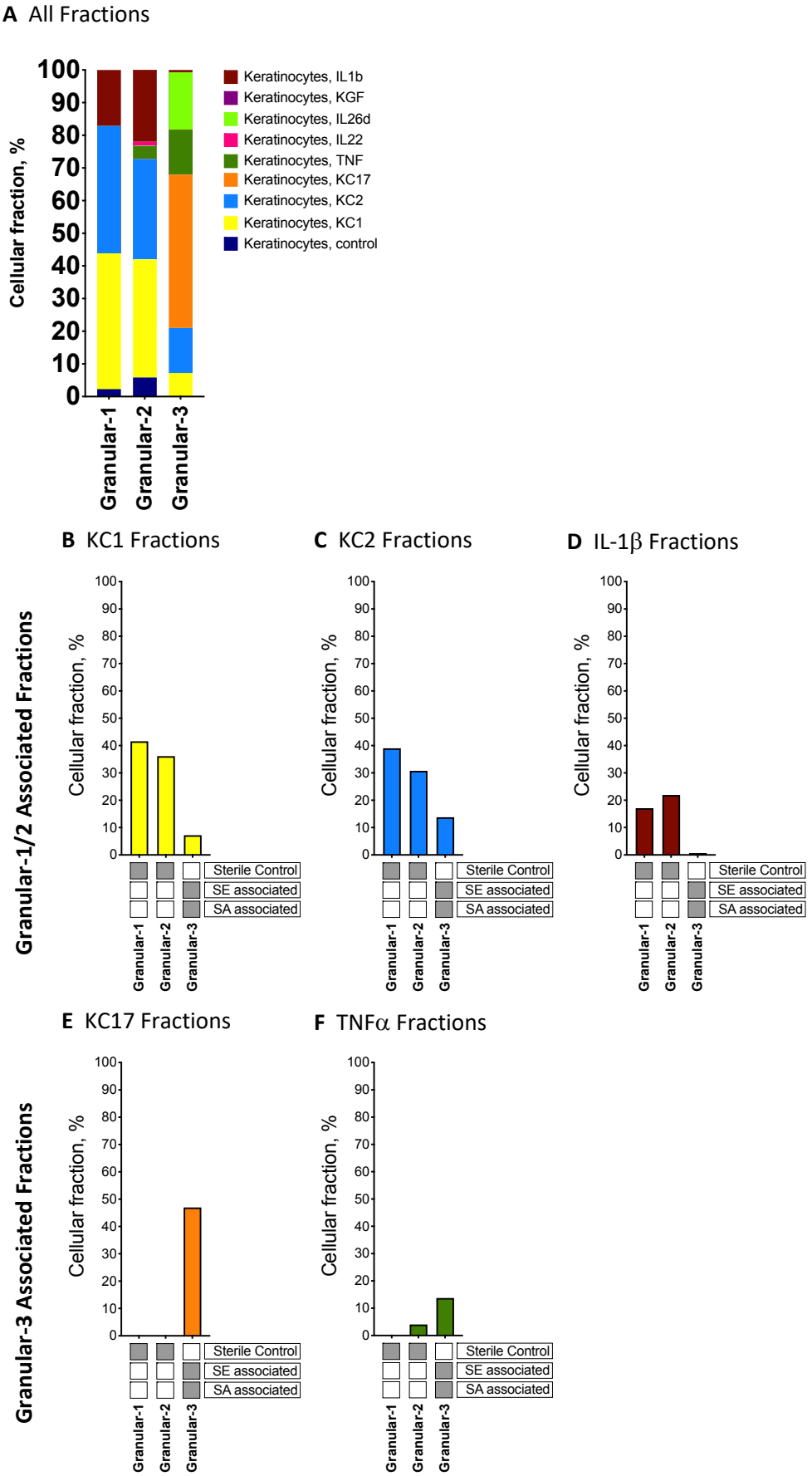
7.2 Results

7.2.1 Granular keratinocytes challenged with *Staphylococcal* species display a TNFα and IL-17 immunophenotype

Previously the bulk transcriptomic data of RHE models was deconvoluted against a panel of keratinocyte cytokine-stimulated signatures (Section 5.3.4). This revealed the immunophenotype fractions relevant to the model conditions, including bacterial exposure and the effect of challenge time course. The aim of this analysis was two-fold: (1) validate the signatures altered by *Staphylococcal* challenge; (2) determine the epidermal layers contributing these altered signatures.

The added granularity of the single-cell model data enabled the determination of layer-specific elucidation of the cytokine immunophenotypes. For this analysis, the identified single cell clusters were made into pseudobulk to enable deconvolution via CIBERSORT using the same keratinocyte immunophenotypes panel as previously. As there were ten clusters in the single cell data, ten pseudobulk samples sharing the same naming convention were generated.

The single-cell data indicated three distinct granular populations. Granular keratinocytes from models challenged with either *S. epidermidis* or *S. aureus* were found to cluster together in Granular-3. In contrast, the Granular-1 and -2 clusters were mainly composed of keratinocytes from sterile models. This association impacted the immunophenotype responses between Granular-1/2 and Granular-3 populations, with the former populations shown to be very similar with distinct changes in the latter (**Figure 7.1A**). The distinctive transcriptomic immunophenotype programmes of sterile Granular-1/2 populations was derived from IL-1β, KC1 (IFNα, IFNγ) and KC2 (IL-4, IL-13) (**Figure 7.1B,C,D**). In bulk immunophenotype analysis of the epidermal models, KC2 was also found to be a major transcriptomic programme (Section 5.3.4). In these sterile, unchallenged granular keratinocytes, the KC1 fraction accounted for approximately 40%, with IL-1β responses further contributing 20-25%.



[*Legend follows on next page*]

[*Legend for figure on previous page*]

Figure 7.1. CIBERSORT deconvolution of pseudobulk granular keratinocytes from RHE models challenged with *Staphylococcal spp.* Keratinocyte insert cultures were incubated with *S. aureus*, or *S. epidermidis* for 24 hours. Models were then dissciated for single-cell RNA-seq by DropSeq. Clusters from the single-cell data analysis were made into pseudobulk data for deconvolution into reference cellular fractions by CIBERSORT using the 11 cyctokine-stimulate keratinocyte profiles: IL-1β, KGF, IL-26d, TNFα, IL-4, IL-17A, IL-13, IFNα, IFNγ. **[A]** The average of each fraction in each granular population. **[B]** KC1 fractions composed of IFNα and IFNγ fractions; **[C]** KC2 fractions composed of IL-4 and IL-13; **[D]** IL-1β fractions; **[E]** KC17 (IL-17) fractions; **[F]** TNFα fractions. Grey boxes indicate bacterial association of each population.

This KC1, KC2 and IL-1β phenotype was drastically reduced in *Staphylococcal* associated granular keratinocytes. The Granular-3 population displayed a combined cellular fraction of less than 20% for these three immunophenotypes. These were replaced by KC17 (**Figure 7.1E**) and TNFα (**Figure 7.1F**) fractions composing 47.0% and 13.7% of this population, respectively. A TNFα fraction indicated an activated phenotype in these keratinocytes. The KC17 fraction in the Granular-3 population was the highest deconvoluted of any population at 47%. The high fraction of KC17 in those outmost keratinocyte populations exposed to bacteria indicates the activation of an anti-microbial transcriptomic programme in these cells.

7.2.2 A TNFα immunophenotype dominates spinous keratinocytes from *S. aureus* challenged models

Analysis of single-cell populations of epidermal models indicated that commensal or pathogen challenge resulted in development of divergent spinous layer phenotypes. Immunophenotype deconvolution analysis of the three described spinous populations recapitulated the undisturbed spinous transcriptomes of Spinous-1 and -2 populations and inflammatory phenotype of the Spinous-3 population (**Figure 7.2A**).

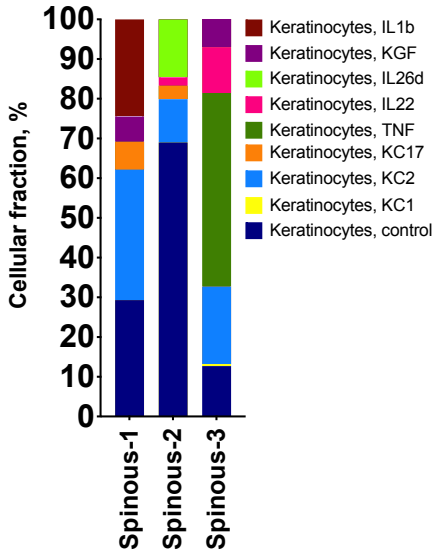
Interestingly, despite being derived from unstimulated monolayer keratinocyte, the control profile was found in all three spinous populations (**Figure 7.2B**). The unstimulated signature composed approximately 30% of the Spinous-1 population made up of cells from sterile and *S. epidermidis*

challenged models. Surprisingly, this unstimulated signature composed an even larger proportion of the population associated uniquely with *S. epidermidis* challenge, Spinous-2.

A small KC17 fraction was also associated with spinous populations (**Figure 7.2C**). The KC17 was highest in the spinous population associated with the least bacterial challenge burden, Spinous-1, followed by the commensal challenged Spinous-2 population, while being completely absent from the S. aureus challenged population, potentially indicating this response profile is linked to effective keratinocyte handling of the microbiome. The anti-microbial effectors of IL-17 responses may also account for the high IL-1β response in the Spinous-1 population (**Figure 7.2D**), which might reflect TLR2/NOD2 sensing by these keratinocytes of bacteria in a commensal context leading to IL-1 transcriptomic responses (Krishna and Miller, 2012).

The activated inflammatory phenotype of the spinous cells from *S. aureus* challenged models, Spinous-3, described earlier was corroborated by the immunophenotype deconvolution by demonstration of 50% of the cellular transcriptome accounted for by a TNFα response signature (**Figure 7.2E**). It was notable that this signature was completely absent from the other two spinous populations that were not associated with pathogenic *Staphylococcal* challenge.

A All Fractions



[*Figure continues onto next page*]

[*Figure continued from previous page*]

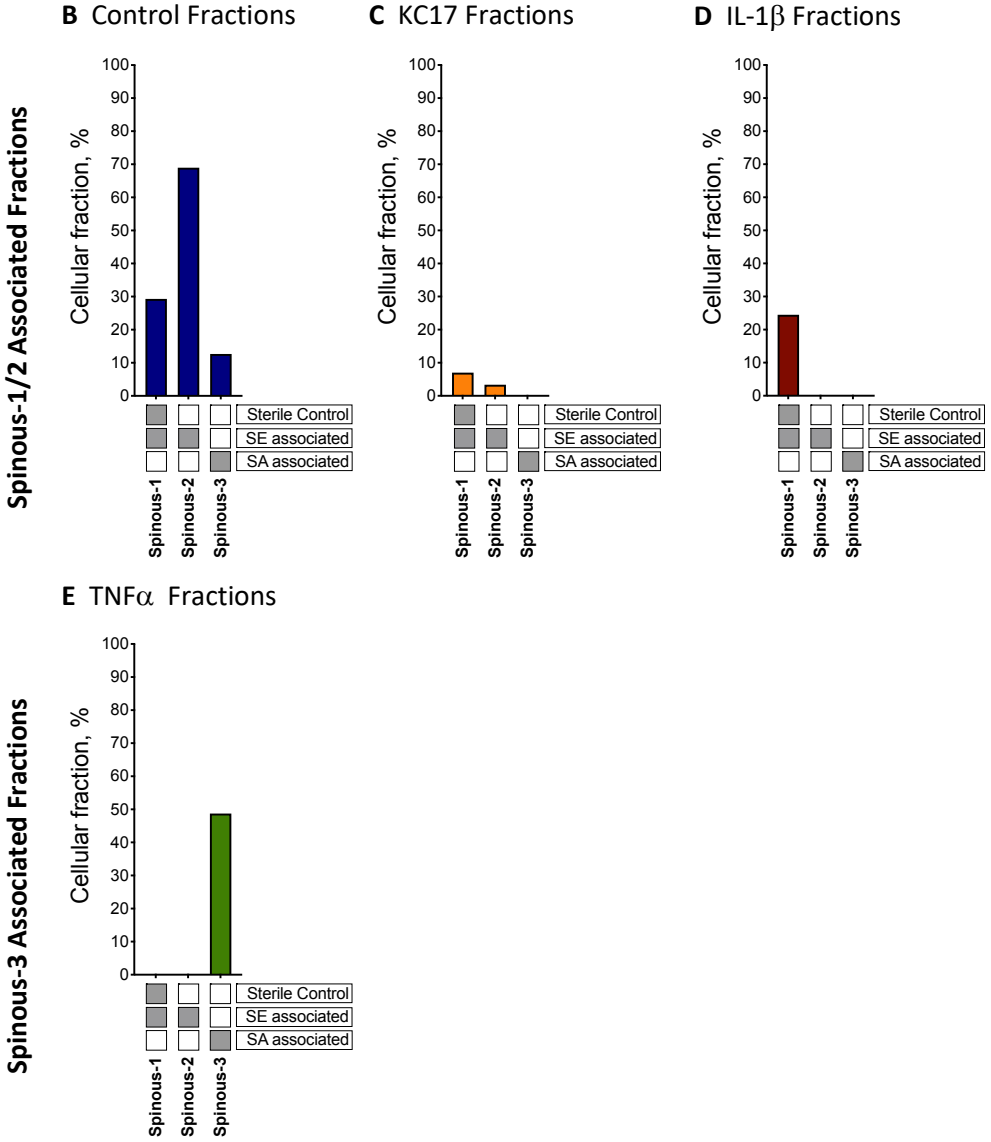


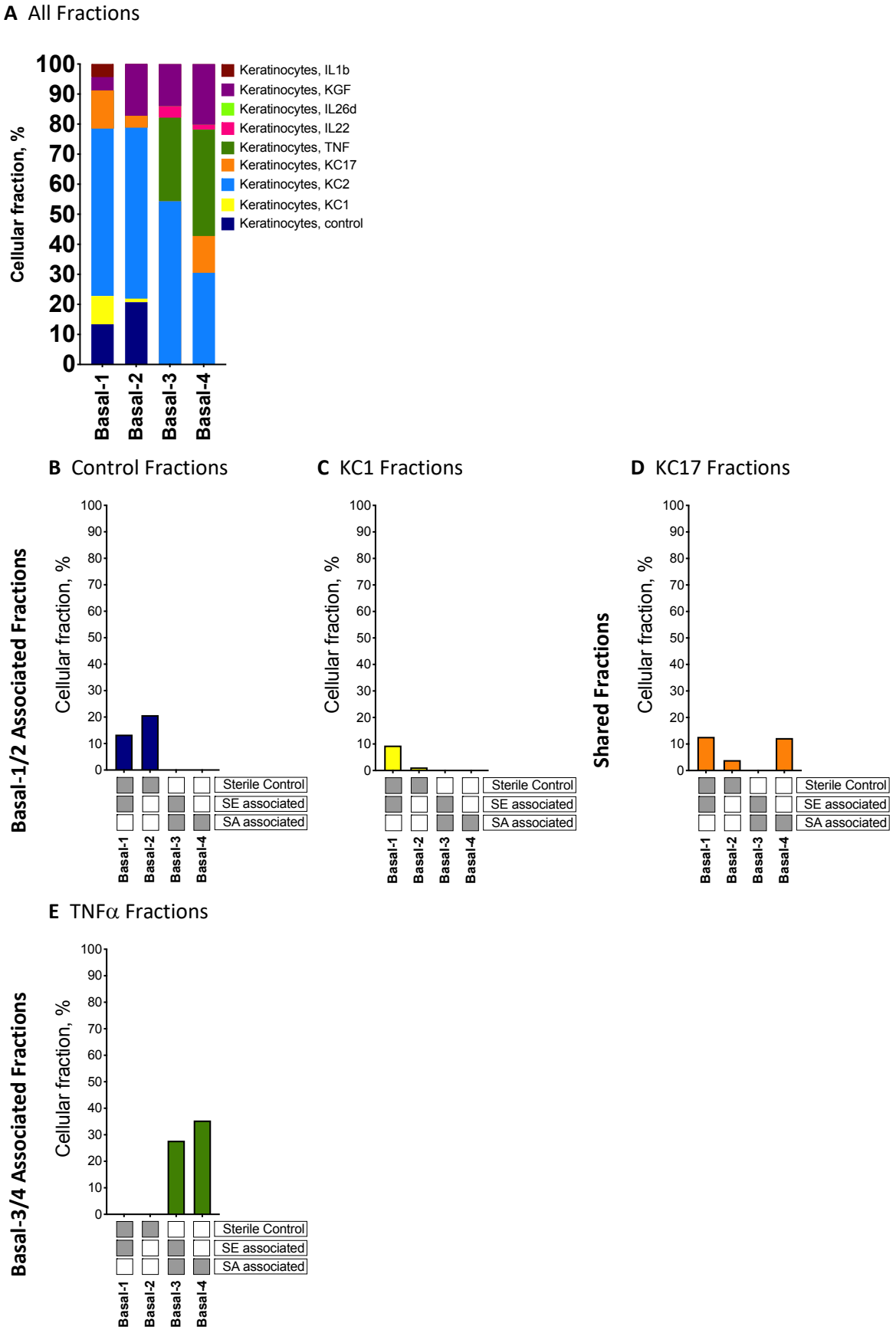
Figure 7.2. CIBERSORT deconvolution of pseudobulk spinous keratinocytes from RHE models challenged with *Staphylococcal spp.* Keratinocyte insert cultures were incubated with *S. aureus*, or *S. epidermidis* for 24 hours. Models were then dissciated for single-cell RNA-seq by DropSeq. Clusters from the single-cell data analysis were made into pseudobulk data for deconvolution into reference cellular fractions by CIBERSORT using the 11 cyctokine-stimulate keratinocyte profiles: IL-1β, KGF, IL-26d, TNFα, IL-4, IL-17A, IL-13, IFNα, IFNγ. **[A]** The average of each fraction in each spinous population. **[B]** Control; **[C]** KC17 (IL-17) fractions; **[D]** IL-1β fractions; **[E]** TNFα fractions. Grey boxes indicate bacterial association of each population.

7.2.3 Basal cells from *S. aureus* challenged epidermal models activate TNFα , but co-expression of a KC17 immunophenotype is required for an inflammatory phenotype

The basal cells from challenged models formed four distinct populations. Approximately half of the basal cells from each condition were split between two populations, which was also shared with another challenge condition. Simply, Basal-1 and Basal-2 populations versus Basal-3 and Basal-4 populations can be described as commensal and pathogen associated responses, respectively. The changes in immunophenotype responses between these populations are shown in **Figure 7.3A**. Deconvolution of basal populations pseudobulk single-cell data again confirmed the high association of KC2 immunophenotype with these models (**Figure 7.3A**). Three of the four basal samples indicated a KC2 fraction greater than 50%. Interestingly, the SA associated basal samples, Basal-4, showed approximately half the KC2 fraction compared to the other basal clusters. Basal-4 KC2 fraction size is reported as 30.5% compared to an average of 55.7±1.2% for the other three basal clusters.

Despite their basal biology, the control signature derived from unstimulated monolayer reference keratinocytes did not compose more than 20% of the basal populations (**Figure 7.3B**). Indeed, this signature was only found in populations that did not include basal keratinocytes from *S. aureus* challenged models, Basal-1 and Basal-2, indicating a loss of this programme during cutaneous inflammation from the microbiome. An interferon programme was deconvoluted from populations composed of sterile keratinocytes alone (Basal-2) and that shared between sterile and *S. epidermidis* challenge (Basal-1) (**Figure 7.3C**). Interestingly, this was only slight in the sterile basal population (1.2%) but was much higher (9.4%) in the dual sterile/*S. epidermidis* population indicating the inducing of this programme during commensal but not disruptive cutaneous colonization.

Resolution of a KC17 fraction across the four populations did not associate with any one population or challenge association (**Figure 7.3D**). The sterile basal keratinocyte population resolved a 4.0% fraction of this signature. This increased to 12.7% in the population composed of sterile and *S. epidermidis* keratinocytes, which was maintained (12.3%) in the population of cells from *S. aureus* only condition. This indicated a bacterial-induced switching on of this immunophenotype programme as might be expected from the anti-microbial effects of IL-17 responses in keratinocytes. However, the basal population composed of commensal and pathogen challenged models (Basal-3) did not resolve this fraction at all. This may indicate diverse KC17 responses in keratinocytes modulated and synergizing with other immunophenotypes to effect either microbiome homeostasis (e.g. in absence of TNFα response fraction) or cutaneous inflammation (e.g. in presence of TNFα response fraction).



[*Legend follows on next page*]

[*Legend for figure on previous page*]

Figure 7.3. CIBERSORT deconvolution of pseudobulk basal keratinocytes from RHE models challenged with *Staphylococcal spp.* Keratinocyte insert cultures were incubated with *S. aureus*, or *S. epidermidis* for 24 hours. Models were then dissciated for single-cell RNA-seq by DropSeq. Clusters from the single-cell data analysis were made into pseudobulk data for deconvolution into reference cellular fractions by CIBERSORT using the 11 cyctokine-stimulate keratinocyte profiles: IL-1β, KGF, IL-26d, TNFα, IL-4, IL-17A, IL-13, IFNα, IFNγ. **[A]** The average of each fraction in each basal population. **[B]** Control fractions; **[C]** KC1 composed of IFNα and IFNγ fractions; **[D]** KC17 (IL-17) fractions; **[E]** TNFα fractions. Grey boxes indicate bacterial association of each population.

The TNFα profile was much increased in basal (Basal-3, Basal-4) populations derived from models associated with *S. aureus* challenge with or without *S. epidermidis* (**Figure 7.3E**). Basal-3 enumerated the TNFα fraction at 27.8%, while Basal-4 indicated a fraction size of 35.4%. No TNFα fraction was resolved in samples where *S. aureus* did not associate, indicating that the TNFα immunophenotype is a pathogen associated response.

7.2.4 Layer-defining keratinocyte populations computationally resolve epidermal stratification from bulk human skin transcriptome

The ten single-cell populations of keratinocytes describing specific keratinocyte layers from model epidermis bacterial challenge were utlised as signatures for bulk skin deconvolution. The expression data were collapsed to pseudobulk for integration with the other cutaneous signatures previously used. This generated the third full cutaneous deconvolution panel alongside those presented in Section 3.2.7 and Section 5.3.6.

Before considering within-layer sub-populations and to observe total stratification across the biopsy samples, the three summed epidermal layers showed fair consistency across healthy, atopic dermatitis and psoriasis tissue samples (**Figure 7.4A**). Keratinocytes composed approximately 40-50% of the whole skin. This observation is slightly lower than against previous deconvolution attempts of the same samples. An enlargement to 15-20% of the steady-state Langerhans cell fraction was common using this panel, which is not observed biologically and likely represents an underfit attributable another signature. It likely reflects a keratinocyte profile not fully captured by

the keratinocyte signatures derived from the stratified model system, potentially encoding basic cell biology shared and common with Langerhans cells also.

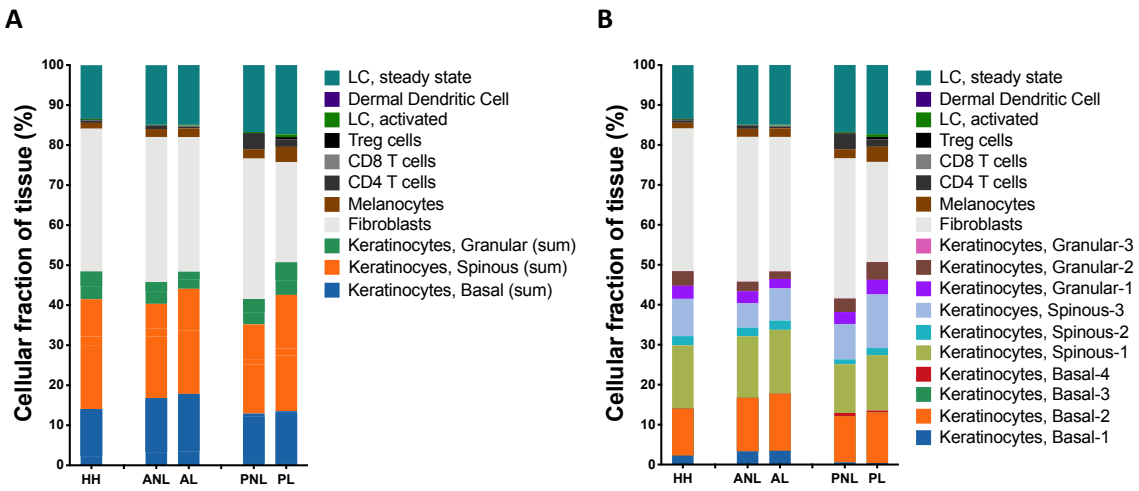


Figure 7.4. CIBERSORT deconvolution of bulk skin samples profiled by single-cell-derived model keratinocyte signatures challenged with *Staphylococcal spp*. [A] The keratinocyte epidermal compartment represented by anatomical basal (blue), spinous (orange) and granular (green) layers alongside the remaining cutaneous populations deconvoluted previously **[B]** The anatomic stratified epidermis layers sub-divided according to the cluster annotation from single-cell analysis of models challenged with either *S. aureus* or *S. epidermidis*.

Basal fractions composed approximately 15% of the whole skin, compared to approximately 30% for the spinous fractions and 4-8% for granular fractions (**Figure 7.4A**). Considering the keratinocyte-only compartment, basal cells composed 30%, spinous cells composed 50% and granular cells 20%. This compares favourably to the epidermal proportions described in single cell analysis of healthy epidermis previously (Cheng *et al.*, 2018). In this analysis by Cheng *et al.*, basal cells composed 20-40% of the dissociated epidermis of trunk or scalp regions. Likewise, spinous populations represented 25% of the epidermis, with granular cells absent due to exclusion by DAPI-negative FACS selection. Here, bulk deconvolution analysis using model epidermis populations as reference signatures has enabled the investigation of the upper, highly differentiated populations which are usually lost in single-cell analysis.

The aim of this stage of deconvolution of bulk skin samples was to investigate the role of bacterial sensing or exposure to keratinocyte populations. This is particularly relevant to the highly stratified

anatomy of epidermis and the distinct keratinocytes residing in these layers. The model epidermis single-cell data enables the tracking of the gross anatomy of human skin as described above, but, further, the added granularity provided by the layer sub-populations identified in Section 6.2.4 opens up the possibility of tracking bacteria sensing programmes in specific layers.

7.2.5 Basal cells are increased in atopic dermatitis

Single cell analysis of the reconstituted epidermal models revealed four distinct populations comprising the basal compartment. These populations, along with the other defined spinous and granular populations were resolved from bulk skins of healthy and inflamed phenotypes (**Figure 7.5**). When applied to resolve the basal compartment from bulk human skin, the cumulative fractions of these four profiles characterizes the total enumeration of the basal population of the sample (**Figure 7.6A**). Compared to healthy controls, the total basal compartment in atopic dermatitis biopsies is increased in both non-lesional and lesional tissues (lesional, p=0.001; non-lesional, p=0.0324). There was no significant difference comparing between lesional and non-lesional samples indicating this expansion was attributable to the underlying disturbed atopic phenotype and not specific to lesion formation.

Basal-1 and Basal-2 defined populations both contributed to this increase for lesional atopic dermatitis samples, whereas Basal-1 was responsible for the expansion of this layer in non-lesional skin (**Figure 7.6B, C**). Increased keratinocyte proliferation in lesional skin, and less so of non-lesional skin, of atopic patients is well described (Jensen *et al.*, 2004). Basal-1 was significantly upregulated in AD samples but significantly down-regulated in psoriatic samples (**Figure 7.6B**). This is despite the known hyperplasia and hyperproliferation present in these diseases. Basal-1 is a *Staphylococcal* challenge associated cluster indicating a microbiome-driven process or underlying environmental factors us causal to this increased basal phenotype in atopic dermatitis but not psoriasis.

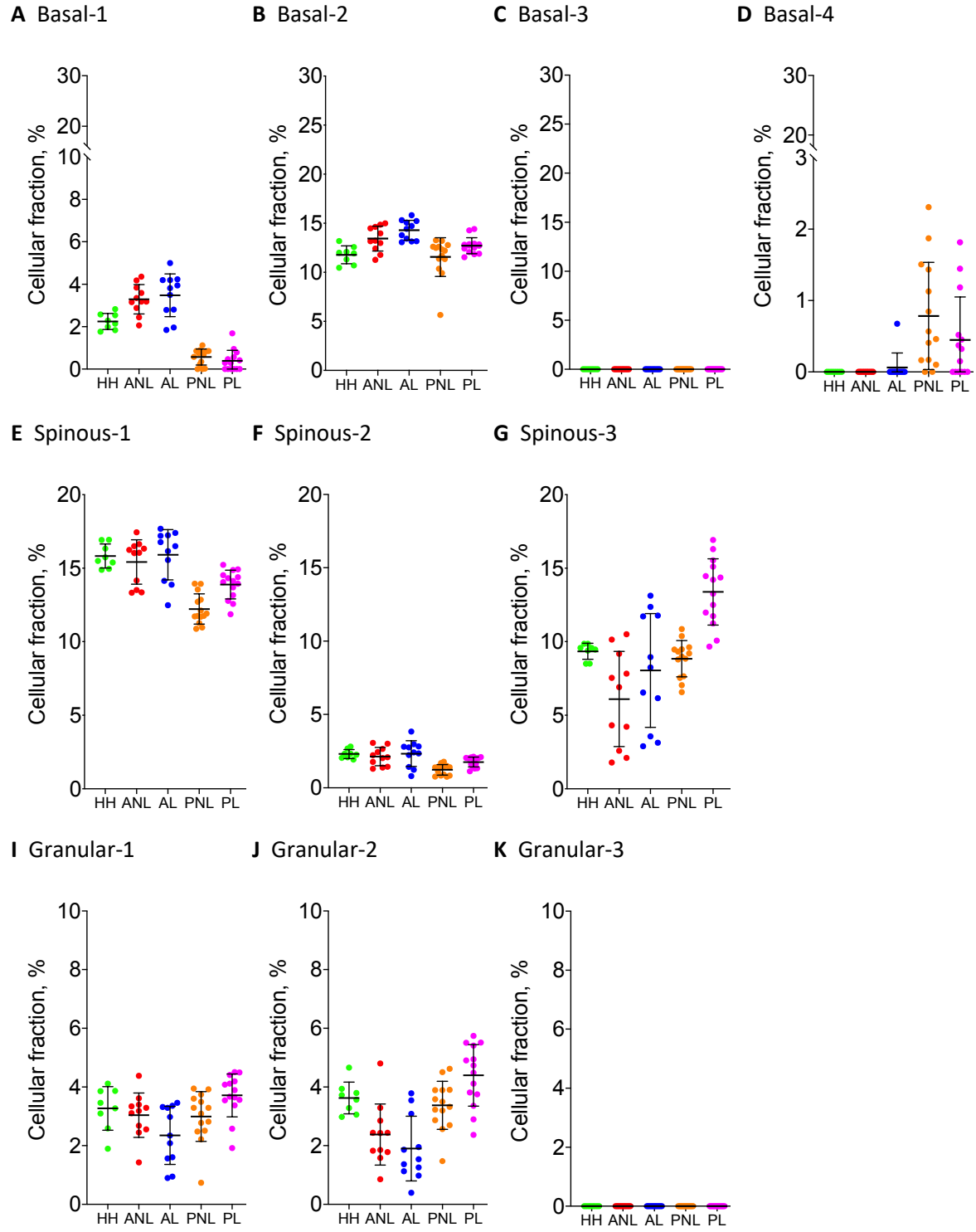
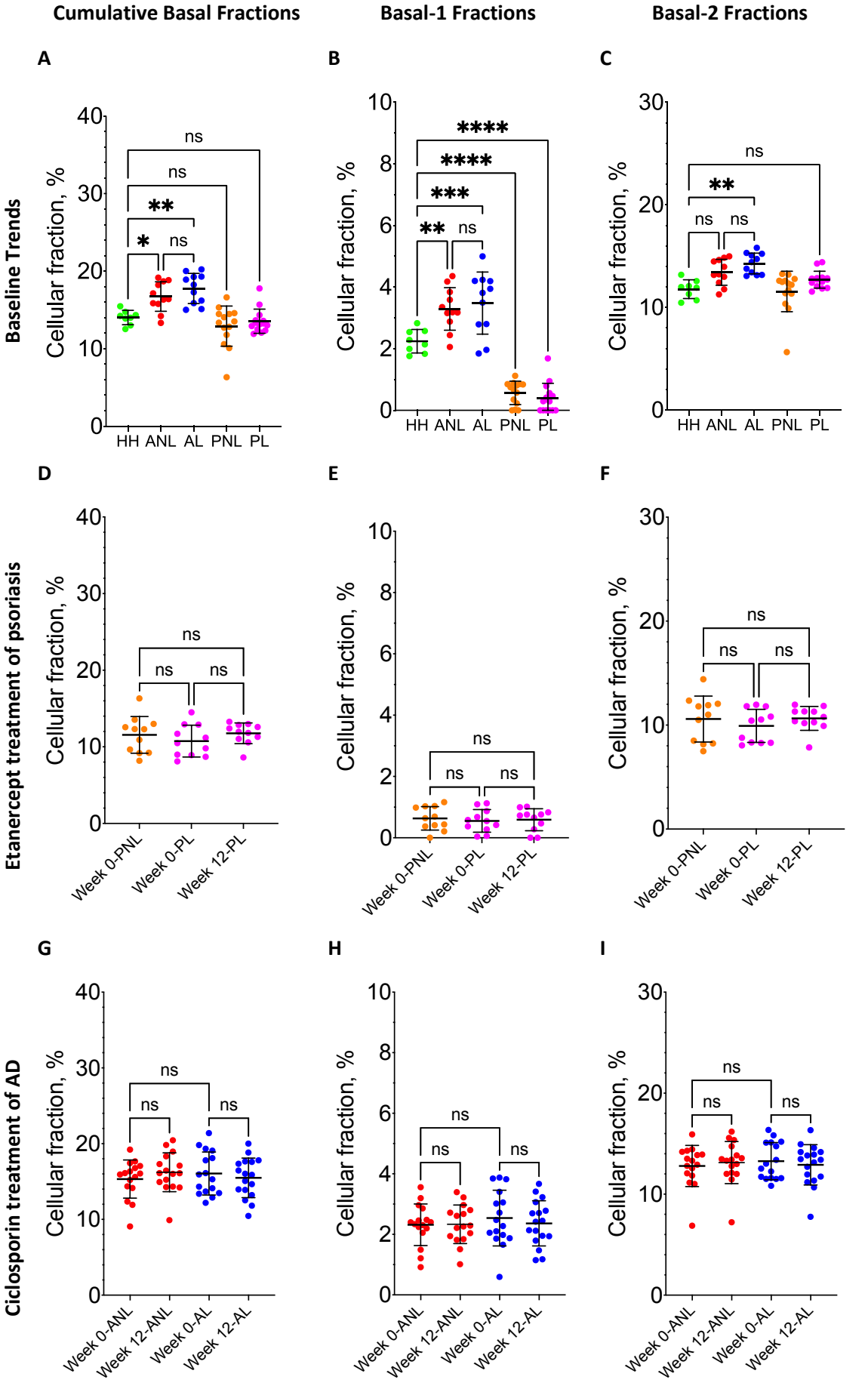


Figure 7.5. Stratified and *Staphylococcal* challenge sub-populations of keratinocytes resolved from bulk skin samples. Healthy (n=8, green), AD lesional (n=11, red) and non-lesional (n=11, blue), and psoriatic lesional (n=14, orange) and non-lesional (n=14, magenta) samples were deconvoluted against a cutaneous panel as shown in Figure 7.2.4B. **[A-D]** Basal populations subdivided by **[A]** Basal-1; **[B]** Basal-2; **[C]** Basal-3; **[D]** and Basal-4. **[E-G]** Spinous populations subdivided by **[E]** Spinous-1; **[F]** Spinous-2; and **[G]** Spinous-3. **[I-K]** Granular populations subdivided by **[I]** Granular-1; **[J]** Spinous-2; and **[K]** Granular-3. Error bars show mean±SD.



[*figure continues onto next page*]

[*figure continued from previous page*]

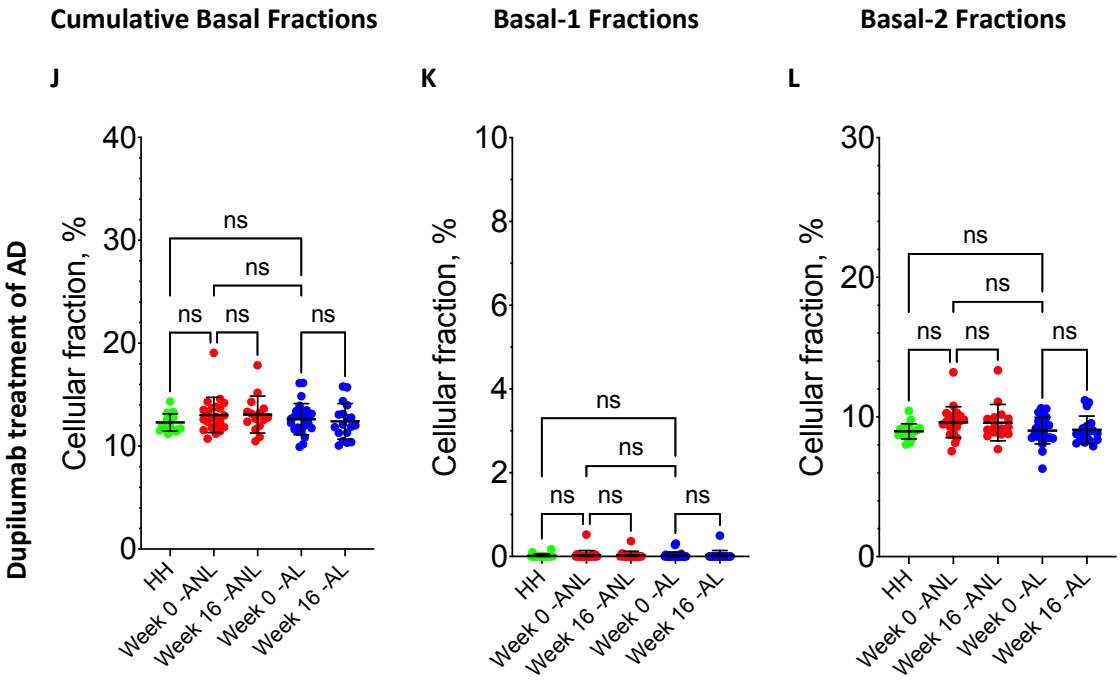


Figure 7.6. Deconvoluted basal fractions from bulk skin transcriptomes. Healthy (green), AD lesional (red) and non-lesional (blue), and psoriatic lesional (orange) and non-lesional (magenta) samples showing cumulative basal fractions [left panel]; Basal-1 fractions [middle panel]; and Basal-2 fractions [right panel]. Basal-3 fractions were not resolved from any samples and average Basal-4 populations were minimal across all tissues (<1%) (see Figure 7.5C, D). **[A-C]** Baseline trends establishing fraction phenotypes across the tissue types investigated. **[D-F]** Responses in psoriatic lesions and non-lesions in response to etanercept treatment. **[G-I]** Responses in AD lesions and non-lesions in response to ciclosporin treatment. **[J-L]** Responses in AD lesions and non-lesions in response dupilumab treatment, including within-dataset healthy controls. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean±SD.

Even in healthy skin, Basal-1 is a small fraction, composing around 2% of the whole skin. However, this fraction is almost entirely absent, <1%, from psoriatic skin (**Figure 7.6B**). Its absence from non-lesional psoriatic samples, similar to lesional comparators, indicates constitutive dysregulation, an observation generally at odds with canonical understanding of non-lesional psoriatic skin as being considered clinically normal. The etanercept treatment dataset of psoriasis confirmed that Basal-1 is a reduced fraction in psoriatic skin regardless of lesion type and is not improved by treatment that is known to clinically improve disease severity (**Figure 7.6E**).

The fraction size of the Basal-1 or Basal-2, or cumulatively, was similar for both lesional and non-lesional samples from patients with atopic dermatitis when compared to healthy controls. The magnitude of these basal fractions was confirmed in the cohort studied in the ciclosporin treatment dataset (**Figure 7.6G,H,I**), as was the statistical insignificance comparing between lesions and non-lesions prior to treatment. In addition, ciclosporin therapy did not alter these fractions.

The analysis of dupilumab treatment dataset did not support the observed increase in basal fractions in atopic dermatitis skin samples compared to healthy skin by comparison to a different cohort of healthy controls. The cumulative fraction of basal signatures was not significantly different in atopic dermatitis, lesional nor non-lesional, in this comparison (**Figure 7.6J**). Likewise, the same healthy versus atopic comparisons for Basal-2 previously shown to be significant for both lesional and non-lesional samples was not repeated in this dataset (**Figure 7.6L**). Furthermore, this dataset did not resolve any Basal-1 fraction from the majority of the samples analysed, which may be an artefact of this dataset specifically. No dupilumab treatment response was indicated from the analysis involving these basal fractions.

7.2.6 A normal spinous population in psoriasis is diminished in favour of an inflammatory population

A spinous population (Spinous-1) defined by normal expression of layer-defining differentiation keratins 1 and 10 composed approximately 16% of bulk healthy skin (**Figure 7.7A**). Compared to the healthy baseline, this population was significantly reduced in psoriatic samples of both non-lesional ($p<0.0001$) and lesional types ($p=0.0078$). Conversely, a spinous population previously described (Spinous-3) as early differentiating and highly expressing inflammatory markers including *CALML5*, *IL1RN*, *IL36G*, *KLK7*, *S100A7*, *S100A8*, *S100A9*, *SERPINB1*, *SERPINB2* was unsurprisingly highly increased in psoriatic lesions compared to healthy controls ($p=0.0063$) and non-lesional comparators ($p=0.0002$) (**Figure 7.7B**). In agreement with the understanding that non-lesional skin from psoriasis patients being otherwise phenotypically and transcriptomically healthy, the Spinous-3 population was unchanged between healthy and psoriatic non-lesional samples ($p=0.9914$).

The deconvolution of samples from the etanercept treatment dataset of psoriasis did not show alteration in the normal, Spinous-1 population (**Figure 7.7C**) but confirmed the increased size of the inflammatory Spinous-3 population in lesions compared to non-lesions ($p<0.0001$) (**Figure 7.7D**). Further, validation of this population expressing a psoriatic lesion-like transcriptome is provided by its reduction over the course of an etanercept treatment regime ($p=0.0019$) to a non-lesional phenotype ($p>0.05$, ns).

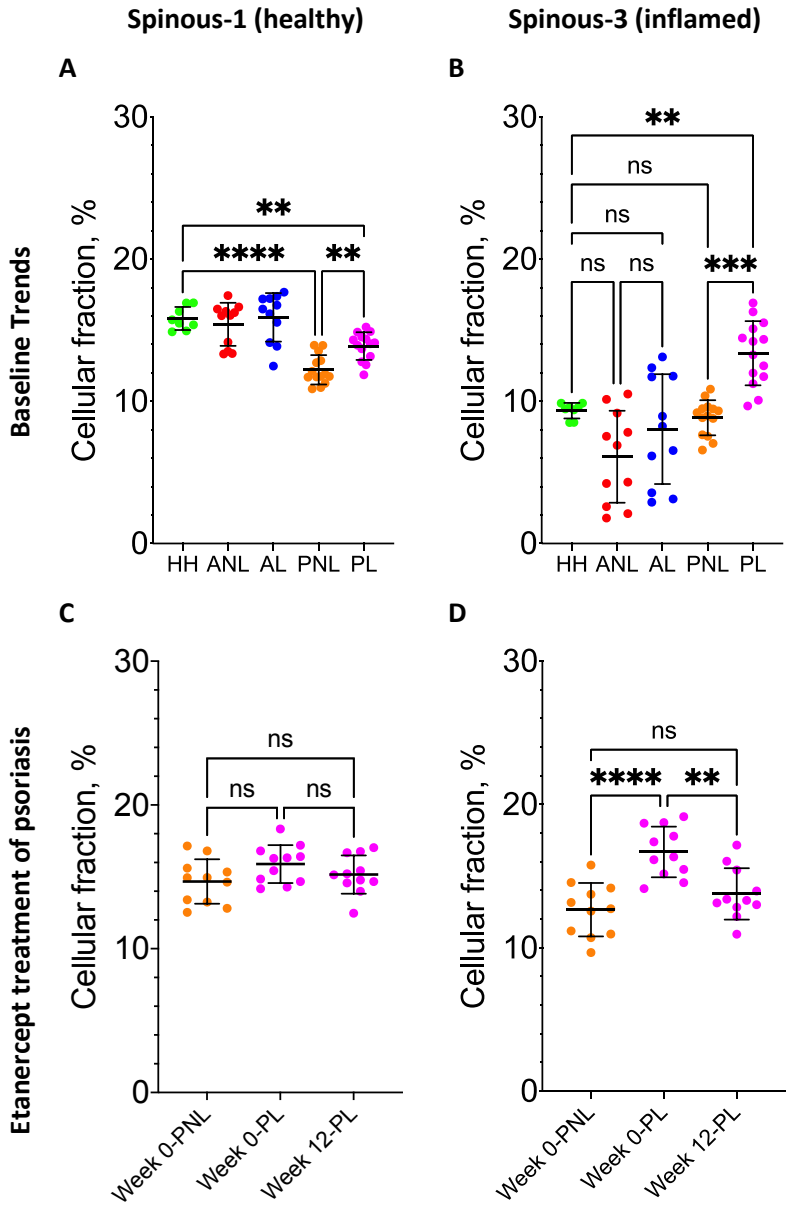


Figure 7.7. Deconvoluted spinous fractions from bulk skin transcriptomes. Healthy (green), AD lesional (red) and non-lesional (blue), and psoriatic lesional (orange) and non-lesional (magenta) samples showing Spinous-1 fractions [left panel]; and Spinous-3 fractions [right panel]. **[A, B]** Baseline trends establishing fraction phenotypes across the tissue types investigated. **[C, D]** Responses in psoriatic lesions and non-lesions in response to etanercept treatment. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean±SD.

7.2.7 Atopic dermatitis populations correlate with bulk *Staphylococcal* challenge

The Spinous-3 population was derived from epidermal models experimentally challenged with live *S. aureus* and *S. epidermidis*. Notably, all *S. aureus* challenged spinous cells were included in this population, whereas only a small fraction of *S. epidermidis* exposed spinous cells composed this population. Most spinous cells from *S. epidermidis* challenged models formed a distinct, separate population in Spinous-2. Despite this bacterial challenge association, Spinous-3 fractions were generally reduced in samples from atopic dermatitis compared to healthy controls (Figure 7.7B). However, this observation did not prove significant for neither non-lesional nor lesional atopic samples.

Furthermore, the resolved fraction size of the Spinous-3 population was heterogenous in atopic dermatitis compared to healthy skin, indicating a potentially variable process regulating or impacting this population not inherent to healthy skin. Healthy samples had a mean Spinous-3 fraction size of 9.3% with a range of 1.4% and a standard deviation of 0.5%. In comparison, lesional atopic skin had a mean fraction size of 8.0% (range, 10.23%; SD, 3.9%), and non-lesional samples had an average fraction size of 6.1% (range, 8.7%; SD, 3.2%).

The fraction of this inflammatory Spinous-3 population positively correlated with the healthy spinous cluster, Spinous-1, in healthy samples (r=0.83, p<0.05) (**Figure 7.8A**). This correlation was weaker in non-lesional atopic skin (r=0.6, p<0.001) (**Figure 7.8B**) and weaker still in lesional atopic skin (r=0.38, p<0.05) (**Figure 7.8C**). This alludes to a regulated expansion of the spinous populations in healthy skin that is lost in atopic skin, particularly within a lesion, in favour of expanding inflammatory spinous populations.

A previous signature of bacteria sensing, and complementary unstimulated signature, was described in Section 5.3.5. The signature was applied to bulk skin biopsy deconvolution in Section 5.3.6. The resolved Spinous-3 fraction sizes correlated with the bacteria sensing signature ("Bulk-SE24") very strongly in healthy skin (r=0.87, p<0.01), atopic non-lesional samples (r=0.86, p<0.001), and atopic lesional samples (r=0.91, p<0.001) (**Figure 7.8**). This underscores the inflammatory/microbial provenance of both signatures. However, the unstimulated signature strongly correlated only with Spinous-3 in healthy skin (r=0.92, p<0.01) (non-lesional r=0.51; lesional r=0.47) (**Figure 7.8A**). This indicates that the Spinous-3 population in healthy skin derives from both endogenous and exogenous microbiome factors, whereas the generation of Spinous-3 in atopic skin likely occurs almost exclusively from microbiome factors such as *Staphylococcal* exposure.

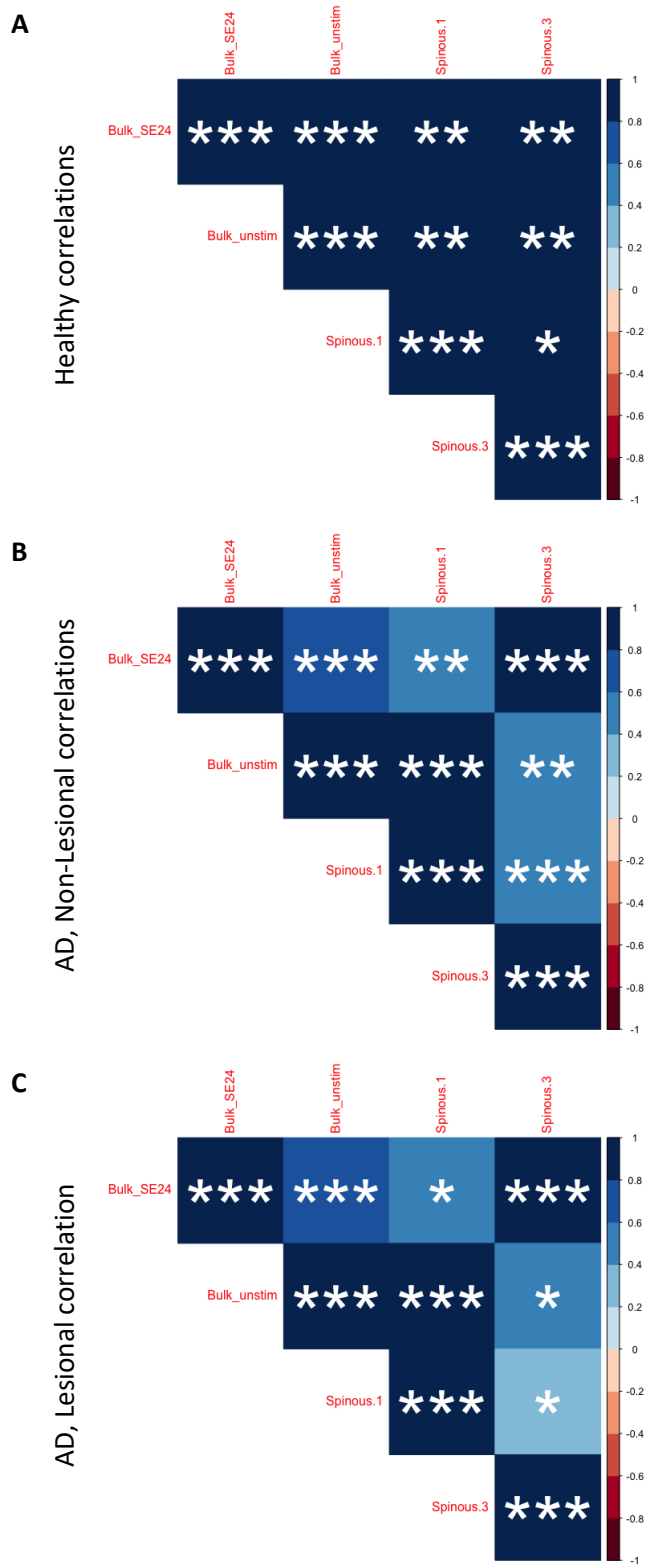


Figure 7.8. Correlation analysis between healthy and inflammatory spinous clusters with bulk bacteria sensing profiles. Healthy (Spinous-1) and inflamed (Spinous-3) fractions were assessed for correlation using the R package *Corrplot* against bulk-defined bacteria sensing signatures "SE24h" and "unstimulated" (described in Section 5.3.5) in healthy **[A]**, AD non-lesional **[B]** and AD lesional **[C]** samples. Pearson correlation *r* value indicated by colour according to key. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001.

7.2.8 The inflammatory spinous population in atopic dermatitis is ameliorated by both ciclosporin and dupilumab treatment

Two atopic dermatitis treatment datasets were deconvoluted using the single-cell-derived stratified keratinocyte populations. The ciclosporin treatment dataset at baseline showed that Spinous-3 fractions are significantly higher in lesional than non-lesional samples (p=0.0103) (**Figure 7.9A**). This was confirmed in the dupilumab dataset (p=0.0401), which additionally indicated that Spinous-3 fractions are significantly raised in lesional (p<0.0001) but not non-lesional samples (p=0.1432) compared to within-dataset healthy controls (**Figure 7.9B**). Indeed, a proportion of the samples in the originally Spinous-3 deconvoluted atopic lesional samples displayed increased fractions than healthy controls (Figure 7.7B). The Spinous-3 results from atopic lesions in the dupilumab dataset may be a continuation of the heterogeneity early observed in this population that appears unique to atopic skin. Treatment with either ciclosporin or dupilumab significantly reduced the size of the Spinous-3 fraction in lesions comparing pre-treated lesions with those at the end of the trial (ciclosporin p=0.0006; dupilumab p<0.0001).

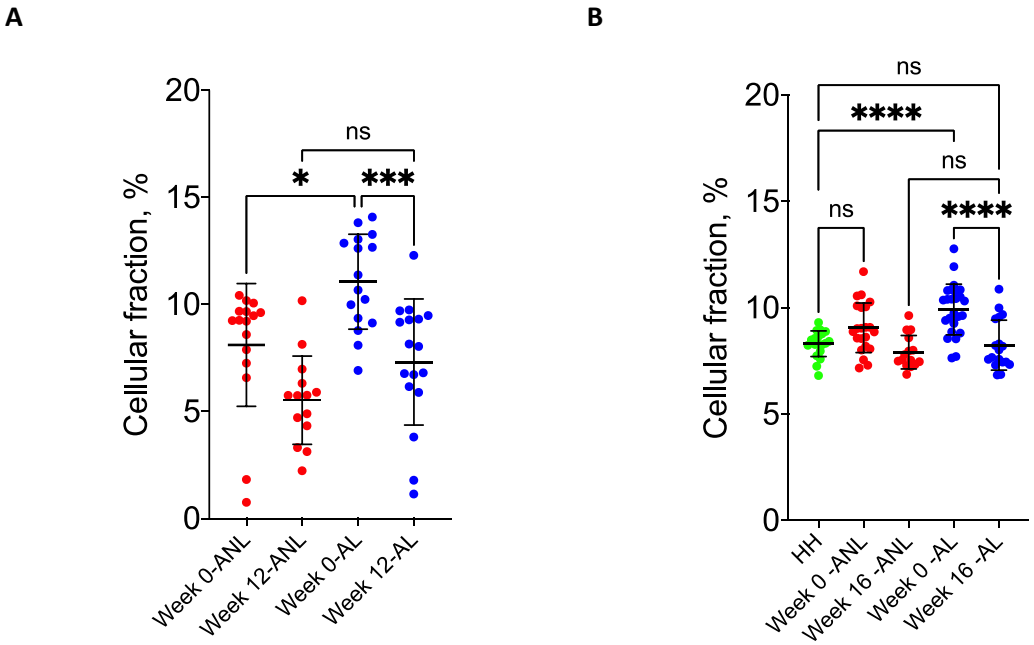


Figure 7.9. Deconvoluted inflammatory Spinous-3 fractions from bulk skin transcriptomes of AD samples. Healthy (green), AD lesional (red) and non-lesional (blue) samples showing Spinous-3 fractions in: **[A]** a ciclosporin treatment dataset over a 12-week time course; and **[B]** a dupilumab treatment dataset over a 16-week time course. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean ±95% confidence interval.

7.2.9 The total granular layer is reduced in lesions from atopic dermatitis patients

Transcriptome investigation of the identified three granular clusters from the scRNA-seq model data did not reveal any specific bacterial challenge-mediated associations. The major difference found was the upregulation of epidermal differentiation markers in the Granular-2 population. For this reason, investigation of the granular layers computationally enumerated from bulk skin biopsies focuses on the cumulative sum of these three fractions.

The total granular layer was found to be only significantly changed in atopic lesions. Atopic lesions showed significant reduction in size of total granular fraction compared to healthy (p=0.0129), but not atopic non-lesional samples (p=0.5002) (**Figure 7.10A**). Similarly, atopic non-lesional samples were not significantly different to healthy (p=0.3486), indicating a subtle diminishing of the granular layer in atopic skin over the course of lesion formation. Psoriatic samples were not distinguishable from healthy (lesional, p=0.4983; non-lesional, p=0.9562) or from one another (p=0.0666).

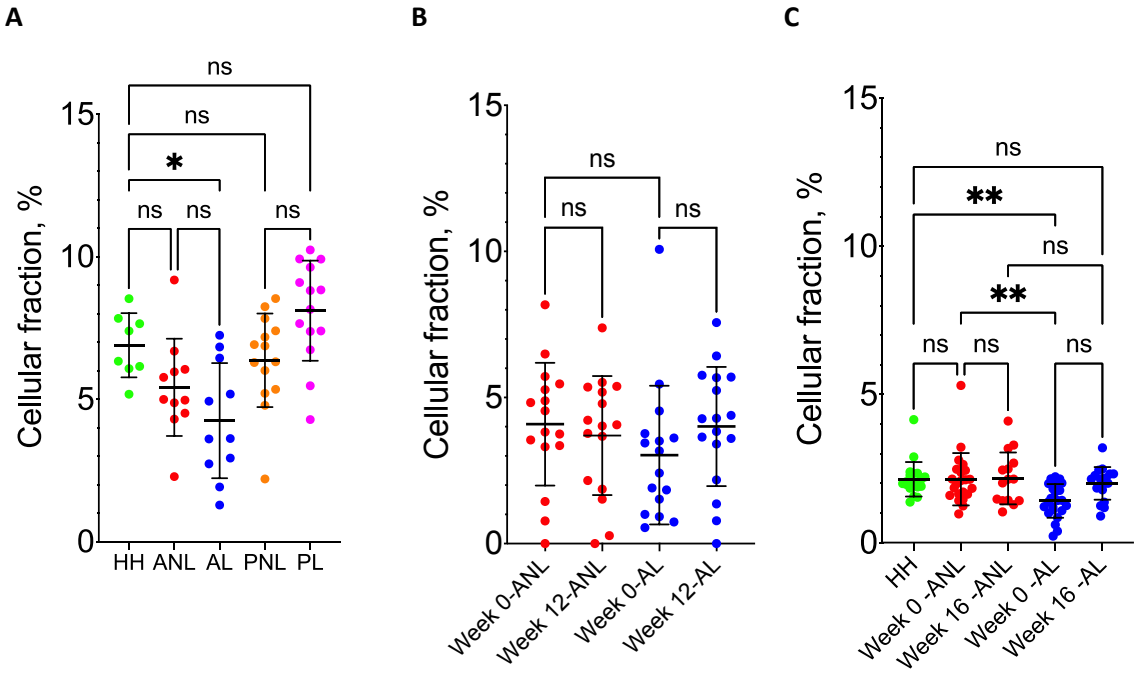


Figure 7.10. Deconvoluted cumulative granular fractions from bulk skin transcriptomes of AD samples. Healthy (green), AD lesional (red) and non-lesional (blue), and psoriatic lesional (orange) and non-lesional (magenta) samples showing cumulative granular fractions in: **[A]** baseline trends of AD and psoriatic lesional and non-lesional samples; **[B]** a ciclosporin treatment dataset over a 12-week time course; and **[C]** a dupilumab treatment dataset over a 16-week time course. *P <0.05, **P<0.01, ***P<0.001, ****P<0.0001. Statistical tests show ordinary one-way ANOVA with multiple comparison performed using Tukey adjusted p-value. Error bars show mean±SD.

In the ciclosporin treatment dataset, cumulative granular fractions from atopic dermatitis samples were generally two per cent smaller than in the baseline dataset (**Figure 7.10B**). There remained no significant difference in comparing granular fractions between atopic lesional and non-lesional samples. The dupilumab treatment dataset presented a healthy cohort with a much-reduced fraction than previously observed, which might be an indication of sample preparation. Despite this reduced healthy granular phenotype, atopic lesional samples prior to dupilumab treatment had a significantly lower cumulative granular fraction than both healthy controls and non-lesional samples ($p=0.0087$, $p=0.0067$, respectively) (**Figure 7.10C**). While dupilumab treatment of atopic dermatitis did not increase the granular fraction in lesions comparing pre- and post-treatment samples, there is a clear upward trend approaching significance ($p=0.0675$). This trend is reflected by comparison to healthy and non-lesional samples. Prior to treatment the reduced granular fraction in lesional samples was distinct to healthy and non-lesional samples, but after 16 weeks of treatment this distinction had been lost.

7.3 Discussion

7.3.1 Comparing bulk and single-cell RHE model transcriptomes for deconvoluted immunophenotypes

This work has presented the immunophenotype deconvolution of scRNA-seq data from a reconstituted human epidermis model previously optimised and investigated by bulk microarray analysis ((Holbrook, 2019)). The bacterial challenge protocol for these datasets remained the same, involving the topical inoculation of $1x10^6$ CFU of either live *S. aureus* or *S. epidermidis* to the models for 24 hours. This enabled the direct comparison of the bulk transcriptome and the single-cell transcriptome. The benefit of this comparison is to validate the results from *Staphylococcal* challenge and utilise the added granularity of single-cell analysis to determine epidermal layer-specific responses.

Although the way the populations derived from the single cell data, and subsequently collapsed to pseudobulk samples, did not entirely correspond to specific bacterial challenge and mirror the bulk dataset exactly, the underlying biology discriminating these clusters was considered more important. This was recapitulated by the Leiden clustering that delineated these clusters in an unbiased, unsupervised way. For that reason, the single cell-derived clusters were retained over separating the data into three pseudobulk samples of each bacterial challenge condition. This approach would enable understanding of why clusters associated with sterile or *S. epidermidis* conditions such as Basal-1, Basal-2, Spinous-1 and Spinous-2; why some are dual to both *S. epidermidis* and *S. aureus* such as Basal-3, or some such as Spinous-3 are highly associated with *S. aureus* only. It is likely that these distinctions would be masked by grouping by bacterial challenge. Crucially, these subtle phenotypes that cause differences in same layer keratinocytes to respond differentially to the same bacterial condition must be elucidated to fully understand cutaneous inflammation.

This analysis confirmed the findings from the bulk immunophenotype deconvolution (see Section 5.3.4). Similar to the bulk transcriptomic analysis, TNFα and IL-17 response fractions are the main signatures altered by *Staphylococcal* exposure at the single-cell level. Importantly, the single cell analysis allowed for the added granularity to determine which epidermal layer these were most prominent. In addition, by separating the model transcriptome into layers and populations within these, the previously slight increase in IL-17 immunophenotype observed in bulk data in response to *Staphylococcal* challenge was shown to almost exclusively and strongly originate from the granular layer. The transcriptome of the Granular-3 population, unlike any other population, strongly indicated a IL-17 response from the keratinocytes within the population. This population

was also very highly associated with bacterial challenge, unspecific to either *S. aureus* or *S. epidermidis*. IL-17 is essential in host defence against cutaneous pathogens, therefore evidence of this transcriptomic programme in the outermost layers of challenged epidermal models is unsurprising (Archer *et al.*, 2016; Dixon *et al.*, 2016). Single cell analysis of pathogen challenged models presented in this work demonstrated basal layer alarmin release and IL-36 expression in upper spinous populations. Previous work corroborates this finding and has demonstrated that this keratinocyte response is specifically induced by *S. aureus* virulence factors and damage to host cells, and triggers epidermal IL-17 inflammation (Nakagawa *et al.*, 2017).

7.3.2 Applications for a single cell derived panel of keratinocyte profiles

An *in vitro* model system for deconvolution of bulk *in vivo* data has enabled the addressing of specific research questions that may not have been possible from either a bulk reference perspective or from single-cell analysis of biopsy samples. A keratinocyte model recapitulating epidermal stratification and challenged with either pathogen or commensal species defines a panel of signatures for deconvolution from bulk skin to understand the crosstalk within the epidermis regulating or responding to local inflammation. The model approach allowed the characterisation of highly differentiated keratinocyte populations that in the bulk transcriptome are lost within the complexity and are often lost by *in vivo* single-cell analysis (Cheng *et al.*, 2018; He *et al.*, 2020). Single cell analysis can often fail to capture these differentiated or cornified cell populations either due to sample processing and dissociation removing these cells or by the fact that many of these cells are terminally differentiated involving nuclear and organelle loss rendering them transcriptomically inactive (Eckhart *et al.*, 2013).

While a differentiated granular population was computationally inferred from bulk skin of patient biopsies, the granular populations highly associated with bacterial challenge were not resolved. The Granular-3 population of highly differentiated cells derived from *S. aureus* or *S. epidermidis* challenge, was not resolved from any samples at all. This likely indicates that this population is an artefact of the model system. The transcriptomic change induced in a population of keratinocytes by a single species of bacteria, and in such high dosage, is not reflective of the diverse and well-regulated epidermal microbiome. Further, MAST differential expression analysis indicated that very few genes were differentially expressed between the granular populations. The highly differentiated phenotype of these cells may explain these subtle expressed differences as cells undertake terminal differentiation rather than active anti-microbial responses.

The sub-granular epidermal layers were resolved from bulk skin biopsies into populations defined from RHE model investigation. Cutaneous inflammation showed to be highly disruptive to a normal

spinous population. Lesional psoriasis samples indicated a reduction of unperturbed spinous population in favour of an increase in inflammatory, early differentiating phenotype. This was successfully ameliorated by during anti-TNF etanercept treatment, further validating the identified TNFα immunophenotype of this inflammatory population.

The inflammatory spinous population arose from epidermal models challenged with *S. aureus*. Despite this, there was a general decrease in this population in samples from atopic dermatitis patients. Interestingly, this population was further diminished in atopic lesions during the course of both systemic ciclosporin and cutaneous dupilumab treatment. Deconvolution of healthy skin indicated that this inflammatory population comprised approximately ten per cent of the whole skin without causing tissue perturbation or dysfunction. Therefore, presence of this fraction in atopic skin is not primarily causal to the inflammatory phenotype but may synergise with an as-yet unidentified trigger to drive pathogenesis.

The effective diminishing in atopic lesions of the inflammatory spinous fraction in response to both broad immunosuppressive and targeted anti-T_H2 therapy and the association of this cluster with pathogen exposure in the *in vitro* models is a promising result for current treatment paradigms. However, it does indicate that while the downstream dysfunction of microbiome handling by the epidermis can be removed from the disease axis in order to resolve lesions, the underlying pathology is not treated as indicated by the presence of this fraction in higher proportions, and without causing inflammation, in healthy skins. Consequently, this fraction is not essential for skin to appear clinically healthy, but removal of this population from atopic lesions correlates well with lesion clearance.

Four separate basal populations were defined from stratified model single cell data but only two were found in transcriptomic data from skin biopsies. One of the basal sub-populations, Basal-2, was increased in atopic skin independent of lesion status and conversely downregulated in psoriatic skin. Unusually, this decrease in psoriatic skin was common to both non-lesional and lesional samples indicating an underlying disturbance in epidermal biology. Healthy skin indicated a fraction of this population at around two to three per cent, which was reduced to approximately one per cent in psoriasis. Coupled with the very slight expression difference between this and other basal populations, correctly resolving this fraction and interpreting such changes in such a small population is difficult. Indeed, only one gene was found using MAST analysis to be differentially expressed in Basal-1 versus Basal-2 (*CXCL14* was upregulated in Basal-1 versus Basal-2). This gene is thought to have anti-microbiocidal function but its pleiotropic effects have confused accurate investigation of its epidermal role (Lu *et al.*, 2016).

In sum, this chapter has investigated the immunophenotypes of epidermal layers during bacterial challenge. These data confirmed previous analysis from bulk models indicating a TNF and KC17 programme induced by *Staphylococcal* model colonisation. This was further distilled to resolve these fractions across the epidermal strata for subsequent application to *in vivo* transcriptomic data to understand the crosstalk of cutaneous inflammation and microbiome handling.

KC17 responses, indicative of anti-microbial T_H17 responses in the skin, were found in challenged outer granular layers alongside an inflammatory TNFα signal. This TNFα inflammatory programme also characterised a spinous population highly associated with *S. aureus* colonisation and psoriatic lesions. Further, anti-TNF treatment showed a dampening effect on this population in etanercept treated psoriatic patients. Interestingly, this population also existed in the skin of healthy and atopic dermatitis donors. However, treatment of atopic dermatitis with either ciclosporin or dupilumab resulted in the diminishing of this fraction away from a healthy phenotype. Removal of this fraction resulted in clinical resolution of lesions but may associate microbiome mishandling and dysbiosis with this population in atopic but not healthy skin.

KC17 and TNFα transcriptomic programmes also underscored a basal signature derived from some, but not all, keratinocytes from *S. aureus* challenged models (Basal-4). Absence of this KC17 fraction but retention of the TNFα programme shifted a population of *S. aureus* challenged basal keratinocytes to cluster with keratinocytes from commensal *S. epidermidis* challenge (Basal-3). Likewise, a TNFα programme was absent from other basal populations (Basal-1, Basal-2) while a small KC17 fraction was observed indicating activity of microbial homeostasis independent of inflammation.

Chapter 8 Final discussion

At least 80% of the epidermis is composed of keratinocytes. These cells undergo continual renewal involving proliferation of basal populations, suprabasal migration and differentiation into mature populations (Eckhart *et al.*, 2013). For maintenance of cutaneous health, the regulation of epidermal proliferation and differentiation must be properly controlled. This major homeostatic function of keratinocytes is important for normal turnover, regulating cell numbers, wound repair and microbial defence. Therefore, disruption of this critical homeostasis can disturb the balance between cellular populations resulting in aberrant inflammation and microbiome dysbiosis.

Keratinocytes, alongside their cutaneous barrier homeostasis role, play a critical role in managing the skin microbiome and subsequent inflammation from dysbiosis. Cutaneous inflammation can disturb the proliferation-differentiation equilibrium and the epidermal barrier leading to infection and disease (Yano *et al.*, 2008; Duckney *et al.*, 2013; Newell *et al.*, 2013; Kobayashi *et al.*, 2015; Swindell *et al.*, 2017). Work presented in this thesis demonstrates that epidermal dysfunction from pathogenic challenge can arise from the basal layers as the potential basis for whole tissue disruption. Therefore, the stratified nature of the epidermis further adds to this paradigm, requiring the investigation of inflammation and microbial interaction in the context of layer-specific responses.

The role and impact of the microbiome is important to the pathogenesis of inflammatory skin disease. The nature of this computational analysis is that it readily lends itself to re-analysis data from the same samples to overlay multiple perspectives. So far, the deconvolution approach investigated immune and cytokine involvement underpinning cutaneous inflammation (G. Wang *et al.*, 2020; Gong and Wang, 2020). The implication of these studies is the better stratification of patient-level data to understand individualised responses to inflammation to effect better treatment options for these patients. For example, Wang *et al.* (2020) developed a novel biometric score for characterising the lesional microenvironment of psoriatic biopsies with potential utility for observing transcriptomic improvements prior to clinically manifest improvement. While these studies have addressed the deconvolution of *in vivo* data to improve understanding of pathological changes associated with inflammatory skin disease, they have not considered the role of microbiome components in driving cutaneous disease. Bacterial exposure is predicted to be just as important in delineating keratinocyte phenotypes. The derivation of a different set of keratinocyte response profiles derived from exposure to microbial species would allow the overlaying of cytokine, barrier and microbiome aspects to fully consider the role of this triad in cutaneous inflammation.

The ultimate aim of tissue deconvolution in a disease context, as presented and investigated throughout this work, is the potential identification of cellular mechanisms of disease, and subsequently therapeutic targets. Biopsies taken from inflamed skin demonstrate marked change in the cutaneous transcriptome and altered cellular heterogeneity. Previous work to understand the gross alterations in the bulk transcriptome have revealed important immune and cutaneous perturbations for further study (Guttman-Yassky, Nograles and Krueger, 2011a; Suárez-Fariñas *et al.*, 2011; Gittler *et al.*, 2012; Bigler *et al.*, 2013; Khattri *et al.*, 2014; Correa da Rosa *et al.*, 2017). Basic ontology analysis of differentially expressed genes between comparisons of interest provides a useful starting point for linking molecular changes to higher level biological or cellular functions (Clayton, Polak, *et al.*, 2017). However, ontology analysis is only as good as the annotations available in the databases used.

Application of co-expression analysis to cutaneous inflammation revealed key transcriptomic profiles from bulk skin involved in cellular processes including lymphocyte activation, interferon signalling and defence responses. Critically, these immune related processes were mirrored in independent clusters describing keratinocyte biology, while other keratinocyte clusters retained expression profiles favoured by healthy skin. This led to the development of the hypothesis that these keratinocyte sub-populations could be deconvoluted using a more direct reference-based hypothesis-driven solution. Previous work performing reference-based deconvolution of the whole skin transcriptome also constructed a skin-specific panel for use in CIBERSORT but did not investigate the transcriptomic effect on keratinocyte subsets under different inflammatory environments (Félix Garza *et al.*, 2019). Therefore, this thesis contributes a panel of cutaneous signatures for deconvolution of bulk skin to investigate relative changes of keratinocyte immunophenotypes in response to local inflammation and treatment. In addition, these signatures are publicly available for use by other researchers for deconvoluting skin transcriptomic data (Clayton *et al.*, 2020).

Understanding of cutaneous inflammation is improved by utilisation of deconvolution signatures leveraging the granularity of the keratinocyte compartment into defined sub-populations. First, cytokine response profiles were deconvoluted from the keratinocyte compartment to establish the validity of this approach in recapitulating known inflammatory axes of atopic dermatitis and psoriasis and ensure fraction changes were biologically relevant in the context of ongoing inflammation and in response to treatment. Secondly, it enables the curation of multiple panels of signatures for repeated deconvolution of the same samples. The benefit of this is the extraction of multiple transcriptomic programmes of interest.

This work later focussed on the role of the microbiome in altering the crosstalk between immune and structural populations and its effect on inflammation. The was achieved by investigation of keratinocyte transcriptomic responses to *in vitro* challenge of bacteria in both the bulk and single-cell context. This established keratinocyte signatures from stratified subsets responding differentially to bacterial stimulation for application to bulk patient biopsy data. A major advantage of the pipeline developed by this work is the ability to apply experimentally derived cellular signatures to *in vivo* biopsy data to progress understanding of inflammatory skin disease. Furthermore, the ease of application of deconvolution panels, once curated and validated, lends their use to re-analysis of the many publicly available datasets and the potentially thousands of archived samples for highly powered studies. This second advantage helpfully avoids the necessity of repeating patient biopsy studies already addressed by bulk analysis using single-cell technologies, without, at least, undertaking exploratory analysis using *in silico* deconvoluted signatures.

8.1 Keratinocyte immunophenotypes recapitulating disease-causing immune axes

In-silico deconvolution of bulk transcriptomic data from healthy, atopic and psoriatic patients demonstrated inflammatory-driven changes in keratinocyte gene expression profiles of lesional and non-lesional tissues. The finding that that non-lesional skins from atopic dermatitis patients, while clinically normal are in, fact, distinct from healthy skin was recapitulated in this work. This distinction is underscored by the effect of local T_H2 inflammation upon keratinocytes, which was termed KC2 and was found to be universal to atopic samples, regardless of lesion status. Barrier integrity, as measured by gene expression, also showed this atopic-healthy split rather than a lesional-non-lesional paradigm. Whether the atopic immune response and impaired barrier in atopic patients are causative of one another, or confounding factors remains an unanswered question of the field (Brandt and Sivaprasad, 2011; Kabashima, 2013; Hönzke *et al.*, 2016).

Further deconvolution results indicate that IL-17 transcriptomic response in atopic dermatitis keratinocytes (termed KC17) are prominent in driving different endotypes. T_H17 responses are pro-inflammatory, which causes increased burden of disease. IL-17, however, mediates pleiotropic effects; it appears to cause more severe disease-related inflammation, while improving barrier and tight junction expression (Onishi and Gaffen, 2010; Maxwell *et al.*, 2015). This work showed that in those atopic patients with a resolved KC17 population, regardless of skin lesioning, this signature was associated with higher KC2 burden.

The role of IL-17 is also critical for innate and antimicrobial immunity and potentially links the role of the microbiome in atopic responses. Colonisation by commensal *Staphylococcal sp.*, such as *S. epidermidis*, induces the homing of IL-17-prodcuing T_H17 cells to the skin (Naik *et al.*, 2015). This may, in part, account for increased KC17 transcriptional profile in chronic atopic dermatitis lesion keratinocytes. This confluence of immunity and microbial commensalism is potentially a response to hypercolonisation by *S. aureus*, itself caused by dysfunction in the epidermal barrier and abnormal cutaneous immune environment. The exact contribution and chronology of these processes is extremely complex and not yet fully understood.

The process of exacerbation and lesion formation in atopic dermatitis is complex, and the exact sequence of events is not understood. This research points to the critical role of IL-17 signalling in lesion formation. Development of a lesion is also dependent on impaired epidermal barrier. Demonstrated in this work is that atopic epidermis has markedly reduced expression of barrier related genes compared to healthy skin. It is proposed, in keratinocytes of patients with atopic dermatitis, an underlying and persistent atopic environment primes the skin, which is later triggered by an unknown insult to cause a lesion. The interplay and roles of the dysregulated barrier and atopic cytokine environment is complex, including both cytokine signalling and barrier reduction, with one driving the other and vice versa. The balance of these pathological mechanisms is likely to vary between patients and between initiation of different lesions within an individual.

The effect of IL-17 in chronic lesions, especially if induced by *Staphylococcal* colonisation is intriguing. By their pleiotropic nature, T_H17 immune and anti-microbial processes would counter and intensify the highly skewed atopic processes driving pre-lesion and early lesion pathology. In order to be effective in the T_H2 altered cutaneous cytokine environment, the IL-17 process would be much amplified. This intense cytokine burden in the skin could be expected to inflect severe disease as evidenced by the coincidence of KC17 with higher KC2. In addition, ciclosporin treatment, which is clinically effective in inducing lesion remission, is associated with a clearing of the KC17 fractions but not KC2 fractions, indicating the importance of IL-17 to lesion formation and maintenance.

A comprehensive study by Fyhrquist *et al.* (2019) investigated the role of the microbiome in atopic and psoriatic patients by parallel 16S amplicon microbiome sequencing and host transcriptomic analysis. This highlights a limitation of the similar co-expression analysis conducted in this current work (Section 3.2.3). Using publicly available data, analysis is limited to the patient information and details provided from each published study. As microbiome sequencing is not routinely undertaken, the meta-analysis performed in this work lacked that level of desired detail.

Microbial communities in atopic dermatitis were heavily dominated by *S. aureus* (Fyhrquist *et al.*, 2019). Host gene expression alterations by this bacterium revealed an enrichment for keratinisation and T_H17 processes, with a prediction from ingenuity pathway analysis of IL-1β, TNFα and IFNγ as upstream regulators. Regulation of T_H17 inflammation by upstream TNFα pathways was also shown in the treatment of psoriasis using the anti-TNF agent etanercept, where suppression of TNFα activity mediated lesion clearance by downregulation of T_H17 pathway genes (Zaba *et al.*, 2009). The data showing the diminishing of the KC17 fraction in etanercept treated psoriasis patients and also ciclosporin treated atopic dermatitis patients adds to this understanding by indicating that responses by keratinocytes to this inflammatory milieu are critical for driving the inflammatory phenotype.

8.2 *S. aureus* host responses and KC17 profiles in atopic dermatitis

Interestingly, atopic patients could be stratified by high or low colonisation by *S. aureus* (Fyhrquist *et al.*, 2019). In this study, low colonisation was actually defined devoid/absence of *S. aureus* within the microflora. Between the high and devoid colonisation cohorts, and comparing the lesional transcriptomes of these groups, differentially expressed genes were identified as *S. aureus* associated. Co-expression analysis, similar to the work presented in Section 3.2.3, projecting these genes onto the correlation network further revealed *S. aureus* associated host modules that were annotated for keratinocyte differentiation and extracellular matrix organisation. Surprisingly, very few of the *S.* aureus related signature co-localised within the modules describing regulation of T_H2 immune responses.

Immunophenotype analysis of atopic keratinocytes indicated a strong KC17 profile in lesional samples. While this signature was also found in psoriatic lesions, and at a similar magnitude, the origin of these signatures between the two diseases may differ (Onishi and Gaffen, 2010). The KC17 profile in psoriasis derives from T_H17-associated autoimmunity (Nograles *et al.*, 2008; Zaba *et al.*, 2009; Onishi and Gaffen, 2010), whereas, as indicated by Fyhrquist and colleagues, the IL-17 profile in atopic lesional epidermis is associated with anti-microbial functions. IL-17-regulated genes such as *S100A9*, *DEFB4*/*DEFB4B*, were significantly higher in atopic patients colonised with *S. aureus*. Further, T_H2 cytokines IL-4, IL-5 and IL-13 were significantly elevated in the *S. aureus* high, IL-17 active cohort. This strongly parallels with the observation in Section 4.2.4 that the presence of a KC17 fractions coincides with a higher KC2 fraction. These data indicate that the KC17 fraction resolved in atopic lesions is anti-microbial in provenance. In addition, its depletion upon ciclosporin treatment indicates that removal of this disease axis related to the microbiome dysbiosis is sufficient to clinically resolve active disease.

This raises the important point of delineating molecular and cellular perturbations in the view of patient and clinical definition of disease activity and burden. The pathogenesis of atopic dermatitis exists as spectra of abnormal background atopic inflammation, genetic susceptibility for impaired barrier and T_H2 skew, and dysfunctional handling of the cutaneous microbiome. This triad of disease pathogenesis intersects at any given and unknown point to cause active disease and lesion formation. Ameliorating some but not all of these axes may reduce lesion intensity and cause remission, but the underlying cause of disease remains. Clearance of lesions does not necessarily indicate restoration of healthy cutaneous biology, as indicated by the KC2 fraction in ciclosporin-treated patients. This may be sufficient for long-term disease control, but it remains vital to further understand the roles, mechanisms and importance of these disease axes. For example, ineffective initial therapy increases the likelihood of further relapse (Tang, Bieber and Williams, 2014). Understanding disease pathogenies in atopic dermatitis means experimental elucidation of background abnormal immune environment, epidermal barrier and microbial dysbiosis to better understand the contributions of these at the patient level for best chance of inducing remission and effective maintenance therapy.

8.3 Limitations of the hypothesis-driven signature approach

The complex environment and interplay of the cutaneous cytokine milieu and local microbiome makes uncovering and understanding disease processes highly challenging. Expression of a given cytokine from healthy to inflamed skin may remain constant but will drive pathology in the latter based on the synergy of other cytokines, while in some processes and disease stages, commensal and pathogenic bacterial species may elicit aberrant and dysregulated immune responses. Analysis of keratinocytes has shown that most cytokines delivering transcriptomic changes on these cells share overlapping effect and often alone will only generate limited alterations in expression profiles (Hamid, Boguniewicz and Leung, 1994; Katia Boniface *et al.*, 2007; Nograles *et al.*, 2008; Guilloteau *et al.*, 2010).

This highlights a potential limitation in this deconvolution approach that centres on using reference signatures derived from single cytokine-stimulated keratinocyte cultures. For this analysis, reference profiles were restricted to the limited datasets available from the public domain; of which the great majority of the keratinocyte cytokine-stimulated microarray data are single cytokine experiments. These single cytokine culture of keratinocytes, therefore, overlook the potential pleiotropic and synergistic effects of the complex, multi-cytokine environment normally encountered by cells.

By using the whole transcriptome for these reference signatures, this limitation is hopefully mitigated. Indeed, despite IL-17 expression found to be unvarying between healthy and atopic skins and requiring synergy with at least IL-22 to drive pathological inflammation, the reference signatures of single cytokine IL-17 keratinocytes were able to resolve significant expression in chronic AD lesions, where expression of this cellular signature was found to be completely absent in healthy skins.

This limitation also extends to the single bacterial co-culture experiments. Human skin does not exist with a mono-species microflora, and the single species model experiments in this work do not capture the bacterial interactions occurring on the skin. These interactions, for example, include the direct inhibition by commensal bacteria of pathogenic species, the priming effect of commensal species of host cells again pathogenic species, or the change in dominant species by alteration of niches in healthy and atopic skin (Otto, 2010; Kong *et al.*, 2012). While these factors would almost certainly be relevant to atopic disease initiation and progression, they remain beyond the scope of this work in first defining stratified keratinocyte responses in the context of archetypal commensal or pathogen interactions.

The microbial influence on the epidermal transcriptome was much more difficult to resolve than the immunophenotype fractions. This was confounded by a few major and overlapping issues. Firstly, defining the bacteria signatures of interest and confidence in their biological annotation. For example, only an *S. epidermidis* signature could be determined from bulk model data but data revealing this signature to be commensal or pathogenic in nature was largely inconclusive. This is further complicated by the fact that a cutaneous commensal such as *S. epidermidis* can act as an opportunistic pathogen and could be such in the reconstituted model system used to define its impact on host cells (Otto, 2009).

A second issue proved to be modelling the chosen bacteria signatures in the appropriate host system. The three-dimensional model system used here effectively recapitulates epidermal stratification and allows live bacterial challenge by formation of a protective cornified layer. However, lack of other cellular populations such as dendritic cells and T cells limits full *in vivo* comparison. The modification and crosstalk between keratinocytes and these missing populations is removed from the keratinocyte only models, therefore precluding investigation of the immunosurveillance, immune tolerance and activation native to human epidermis (Hennies and Poumay, 2021). Other systems have been developed to recapitulate the multi-cellular complexity of human skin such a keratinocyte-fibroblast full-thickness models, Langerhans cell or melanocyte inclusion for improved biological relevance (Bell *et al.*, 1981; Mathes, Ruffner and Graf-Hausner,

2014). However, increased *in vitro* complexity increases the risk of culture failure, reduces reproducibility and makes deciphering the cell-type-specific biological responses difficult. For this research, the main focus of this research is the keratinocyte response to simplified microbiome exposure, therefore a keratinocyte-centric three-dimensional model is desired, and is optimal for its relatively quick culture period and reproducibility.

8.4 Modelling bacteria sensing in the epidermis

The cutaneous microbiome is increasingly viewed as critical for driving pathogenesis in the skin alongside perturbed immune processes (Fyhrquist *et al.*, 2019; Carrieri *et al.*, 2021). Understanding how the microbiome affects keratinocyte biology and the crosstalk between host epidermal cells will prove vital in uncovering the mechanisms behind maintenance of cutaneous health and the role of inflammation. A few studies have attempted to address the alteration of keratinocyte behaviour by microbial exposure (Secor *et al.*, 2011, 2012; Duckney *et al.*, 2013). However, *in vitro* models are limited by the necessity to reduce complexity and move away from *in vivo* reality. For instance, monolayer keratinocyte cultures can only be exposed to irradiated bacteria (live but replication deficient) or bacterial products, either lysate products or secreted proteome factors. Three-dimensional cultures that model the stratified differentiation of the epidermis and form the apical cornified layer show more promise for modelling cutaneous microbiome interactions.

Bulk transcriptomic analysis of a reconstituted human epidermis model topically exposed to live *S. aureus* or *S. epidermidis* did not produce distinct and separate challenge-associated signatures for use a deconvolution reference profiles. Indeed, only a commensal *S. epidermidis* signature was found. This data aligns with previous studies indicating that topical application of bacterial species drives subtle alterations in the transcriptome, while media infection – simulating a barrier breach scenario – induces much cytotoxicity and inflammation (Duckney *et al.*, 2013). It is possible that the apical cornified layer of the models is well formed enough to protect against the most pathogenic and disruptive aspects of infection. Previous work optimising the bacterial strains for a RHE co-culture system determined that the virulence of some *S. aureus* strains, i.e., SA-ATCC-29213, proved catastrophic to model integrity over the desired 24-hour time course, while other strains, i.e., SA-NCTC-8325-4, demonstrated more mild virulence (Holbrook, 2019). Balancing the virulence of bacterial strains experienced by a normal skin microbiome with the viability of the epidermal model is an important consideration for generating bacterial sensing signatures in keratinocyte over 24 hours.

Furthermore, alterations in the transcriptome in response to bacterial challenge are predicted to be layer-specific, which may be lost within the expression admixture in the context of bulk

transcriptomic investigation. A DropSeq approach resolved the keratinocyte signatures at the single cell level and allowed for the assessment of heterogeneity from within the same co-culture experimental protocol. The improved resolution of cellular signatures and subpopulations could then be leveraged to gain greater insight into the bulk gene expression samples already analysed.

In agreement with the observation that topical application of *Staphylococcal* bacteria to well-formed models is tolerated, single-cell analysis did not find a major transcriptome alteration in granular cells from challenged models compared to sterile models. The major changes were shifts in the phenotype of spinous and basal populations. A spinous population from *S. aureus* challenged models showed reduction of normal *KRT1* and *KRT10* spinous markers, and an increase in differentiation alike higher granular cells. The implication of this for bacterial challenge is the driving of lower epidermal populations to increase differentiation, either to induce active anti-microbial activity or towards the terminal cornified phenotype in order to maintain physical barrier.

8.5 Microbiome alteration of basal keratinocyte populations

Basal populations are affected by both commensal and pathogenic topical challenge. Each of the challenge conditions; sterile, *S. epidermidis*, and *S. aureus*, form two distinct populations, one shared each with another condition. The underlying mechanism separating these binary populations is unclear. It is unlikely that live bacterial penetration has occurred because such a barrier breach would cause widespread cytotoxicity and model disruption as found by Duckney and colleagues (2013). Indeed, the ability to still observe well-defined stratification of challenged models indicates that barrier breach has not occurred. Therefore, microbiome signalling must either arise from bacterial products penetrating the model to the basal layer (i.e., PAMP signalling), upper layer cytotoxicity signalling host damage to deeper unaffected cells (i.e., DAMP signalling), or otherwise paracrine signalling between keratinocytes.

The alarmins IL-1α and IL-1β were predicated to be functionally associated with *S. aureus* colonisation of atopic dermatitis patients using ingenuity pathway analysis (Fyhrquist *et al.*, 2019). Single cell analysis performed in this current work demonstrated that these transcripts were significantly expressed by basal keratinocytes from *S. aureus* challenged models indicating that alarmin production arises from the cells most distal to infection. In addition, using the same model system and challenge protocol utilised in this work, previous challenge experiments confirmed the protein expression of IL-1α and IL-1β by detection in model culture undernatants (Holbrook, 2019).

The *S. aureus* associated basal population called Basal-4 in this analysis is highly active for tissue remodelling and IL-1 alarmin production. The protective role of these IL-1 alarmins is proposed to

be via inducing production of defensins, such as hBD2, to counter *S. aureus* proteases (Wang *et al.*, 2017). Indeed, hBD2 was also increased at the protein level in models challenged with *S. aureus* confirming this association (Holbrook, 2019). *DEFB4*, the gene encoding hBD2, was not detected in the single-cell model data so the layer expressing this gene could not be determined.

Another basal sub-population associated with both *S. aureus* and *S. epidermidis* challenge did not show this alarmin and remodelling response. The response of this second population was more associated with proliferation homeostasis indicative of barrier reparation in response to microbial presence. This indicated that not all basal keratinocytes respond to *S. aureus* in the same manner; some respond like *S. aureus* is an inconsequential commensal, while others activate a pro-inflammatory and anti-microbial response.

The differential response by basal populations to the same exogenous stimulation has important implications for understanding epidermal crosstalk and inflammation. *S. aureus* is highly associated with atopic dermatitis and is present on most atopic lesions (Guzik *et al.*, 2005; Gong *et al.*, 2006). However, in healthy skin, colonisation by *S. aureus* is often cleared without eliciting cutaneous inflammation or adverse events. It may be that these populations are altered in their ratio to one another in atopic skin compared to healthy skin. Furthermore, the known association of *S. aureus* released factors enhancing T_H2 inflammation in atopy-susceptible patients may further drive the ratio in favour of the inflammatory population.

The topological context of these populations is also important to consider because it may help to elucidate if these populations are distinct and anatomically mixed or form a lineage. A stratified, “double-decker” model of these two basal populations each forming distinct basal sub-layers might indicate that a constitutive signal is causing this population to arise. It may also indicate that one population will differentiate into the other population because of the outward turn-over of keratinocytes within the epidermis. In contrast, a dispersed, three-dimensional “chequer-board” pattern indicates a sporadic, potentially stochastic, inducer of the population phenotypes.

Single-cell analysis of model epidermis challenged with *Staphylococcal spp.* has revealed key changes in keratinocyte transcriptome that is altered by both the stratified anatomy of the host cells and the pathogen-commensal niche of the microbial populations. The populations derived from this work, in much the same way as the keratinocyte cytokine immunophenotype signatures, were applied to bulk patient biopsies to understand bacteria sensing in the role of host cutaneous inflammation.

8.6 Microbiome-associated alteration in whole skin biopsies

The populations identified from model bacterial challenge were applied, using the CIBERSORT pipeline, to deconvolute these profiles from bulk skin biopsies. Interestingly, a small granular population was able to be defined from bulk data whereas previous analysis of human skin or epidermis has struggled to identify these from direct single-cell analysis of the tissue (Cheng *et al.*, 2018; He *et al.*, 2020). Granular populations were shown to be reduced in atopic lesions, and this was unaltered by treatment with either ciclosporin or dupilumab.

S. aureus challenge induced inflammation, particularly in the spinous layer, with a molecular signature very similar to psoriatic lesions. Expression of IL1R and IL36γ are highly associated with lesions from psoriatic biopsies but were also found to be expressed at the mRNA level by spinous model cells challenged with *S. aureus* (Carrier *et al.*, 2011; Pfaff *et al.*, 2017). Protein expression from bulk models, analysed from media undernatants, confirmed the presence of IL-1α, IL-1β, and IL-36γ (Holbrook, 2019). Their association with *S. aureus* challenge likely arises from the stimulation of IL-1/TLR signalling to mediate microbial defence, and a shared MyD88-dependent signalling pathway (Cohen, 2014; Swindell *et al.*, 2018). This confuses the distinction between their expression arising from psoriatic inflammation or in response to adverse microbial infection (Johnston *et al.*, 2011; Li *et al.*, 2014). This crossover indicates some degree of sharing of activated pathways but leading to generation of distinct phenotypes whereby *S. aureus* activation of the pathway contributes to downstream atopic inflammation whereas IL-1 activation drives psoriatic-like inflammation.

Indeed, these models also resolved significant IL-17 and TNFα-related immunophenotypes, further mirroring psoriatic lesions. As psoriasis is mainly driven by keratinocyte perturbations and hyperproliferation, with cutaneous inflammation working to sustain this pathogenesis, whereas atopic dermatitis is much more highly associated with significant T cell infiltration and inflammation, it is unsurprising in a keratinocyte-only model that inflammation is alike psoriasis rather than atopic dermatitis. The lack of a T cell component to the model system is a disadvantage to this data. It is proposed that both a perturbed T_H2 infiltrate and microbiome dysbiosis are required for pathogenesis of atopic dermatitis. Therefore, effective deconvolution of these causal disease axes using this RHE model alone was not sufficient in elucidating aberrant microbial-keratinocyte crosstalk leading atopic inflammation.

When applied to bulk skin deconvolution, the basal populations of high interest (Basal-3 and Basal-4) were very lowly enumerated. In fact, Basal-3 was not identified in any samples analysed. Basal-

4 was very lowly identified in most samples below one per cent, except for the dupilumab data where fractions across all samples, including healthy, were homogeneous at around 5% composition. The Basal-4 population is a putatively *S. aureus* induced population that was expected to be computational found in atopic samples. However, the data did not reflect this.

In the model system, the proportion of the described binary basal populations in each challenge condition was approximately equal. It is, therefore, surprising that these basal populations were not strongly resolved in human skin. It may be that in the context of the four basal signatures utilised in the deconvolution panel, the distinction between these is insufficient to allow for proper deconvolution. Indeed, differential expression analysis performed by MAST identified very few genes characteristic of each population in relation to the others. Future deconvolution analysis of these should carefully consider which of the signatures to retain without introducing a forced bias.

8.7 Improving deconvolution accuracy

To minimise deconvolution bias, the whole panel of signatures identified by single-cell analysis of challenged models was utilised. However, this may have proven sub-optimal due to such closely related transcriptomes of the layer sub-populations. For the immunophenotype analysis this was not an issue because cytokine stimulation so drastically alters the transcriptome. Additionally, these cytokine signatures were extensively validated by a train-two-test-one approach (Section 3.2.6). Further, the transcript dropouts common to single-cell technologies may have confounded this problem by mixing sparse single-cell reference signatures with bulk transcriptome signatures. This was necessary in order to retain the previous cutaneous cell types established within the deconvolution pipeline. This may also explain the expansion of the resting Langerhans cell fraction at the expense of the keratinocyte compartment. Individually, due to the problem of dropouts, the keratinocyte profiles were difficult to define as reference signatures. Consequently, compared to the other cutaneous signatures they were underpowered, thus allowing the overfit of the Langerhans cell signature into what was previously identified as the keratinocyte compartment, for example.

The simplest approach for tackling this problem is refining the keratinocyte signatures of interest. This approach should take careful consideration of not biasing the deconvolution panel to deliver preferred, idealised results. That is, a signature should not be retained purely for its presumed presence in bulk tissue. As the populations derive from a model system, contemplating which signatures may be artefactual and not expected *in vivo* could improve signature definition. For example, clusters associated with solely sterile conditions, for example, Basal-2, may be eliminated as not biologically relevant. This would define three basal clusters of commensal-only, commensal-

pathogen, and pathogen-only background available for accurate *in vivo* computational deconvolution.

Another approach may be to target populations for merging based on commensal/pathogen association. Sterile basal keratinocytes form populations Basal-2 and Basal-1; *S. epidermidis* challenged keratinocytes share the Basal-1 population with sterile conditions, and have an additional population, Basal-3. The Basal-3 population is shared between a fraction of *S. epidermidis* and *S. aureus* challenged models, with the remaining basal keratinocytes from *S. aureus* models forming a distinct population in Basal-4. A commensal microbiome signature may be defined from merging Basal-1 and Basal-2, while a pathogenic microbiome signature may be defined from merging Basal-3 and Basal-4, or Basal-4 alone. Repeating the deconvolution of the skin samples using a considered and refined panel should elucidate improve fractions in terms of bacterial challenge background but may risk losing nuanced anti-microbial biology between the populations, especially in terms of comparing Basal-3 and Basal-4.

8.8 Summary

- Functional annotation of clusters from correlation networks of co-expressed atopic and psoriatic defining genes revealed biological processes relevant to cellular populations such as T cells and keratinocytes.
- The expression of these clusters was descriptive of tissue samples by lesion and inflammation status, including multiple keratinocyte-based clusters more highly associated with inflammation or more highly expressed in healthy skin.
- Hypothesis-driven machine learning deconvoluted the immune programmes from the keratinocyte transcriptome revealed interferon, T_H2, T_H17 and T_H22 inflammation is mirrored by KC1, KC2, KC17 and KC22 immunophenotypes, respectively.
- KC2 was descriptive of atopic samples regardless of lesion status, whereas KC17 was more highly associated with lesional skin and in similar magnitude to the fraction in psoriatic lesions.
- Broad spectrum treatment of atopic dermatitis did not reduce the KC2 immunophenotype but specifically targeted KC17 fractions for reduction. Dupilumab, as expected, did ameliorate the KC2 immunophenotype compared to placebo controls.
- A signature from monolayer keratinocyte cultures exposed to biofilm related *S. aureus* proteome did not correlate well with reconstituted model signatures.
- Deconvolution of reconstituted epidermis models indicated TNFα and KC17 profiles in response to challenge with *S. aureus* or *S. epidermidis.*

- In a bulk context, *S. epidermidis* displayed the most prominent bacteria sensing signature from models after challenge for 24 hours. Other species and timepoints were difficult to distinguish from sterile controls at all timepoints.
- Deconvolution of this bacteria sensing signature from atopic skin biopsies indicated a negative correlation with the KC2 immunophenotype but a positive relationship with KC17 in lesions, and barrier expression in both lesions and non-lesions.
- Single-cell transcriptome analysis of *Staphylococcal* challenged epidermis models demonstrated *in vitro* layered stratification and layer-specific sub-populations by bacterial challenge.
- Bacterial challenge associated with a population of granular keratinocytes underscored by a KC17 and TNF transcriptome programme indicative of inflammatory anti-microbial T_H17 induced responses.
- *S. aureus* infection of models formed a specific population of spinous layer keratinocytes that shared differentiation phenotype with granular cells and expressed inflammatory markers such as *IL-36G*, S100 protein and serpin-encoding genes.
- An inflammatory TNF immunophenotype was also resolved from this population., implicating its amelioration during anti-TNF etanercept treatment of psoriasis.
- This inflammatory spinous population was also present in healthy skin without disruption to cutaneous homeostasis. In contrast, diminishing of this population from atopic skin associated with lesion clearance alluding to the role of this population in microbiome dysbiosis.
- Basal keratinocytes were strongly affected by apical bacteria challenge forming binary populations demonstrating an increasing gradient of inflammation and remodelling in commensal-only, shared commensal-pathogen and pathogen-only populations.
- Basal keratinocytes from *S. aureus* challenged models either had a transcriptome of mild remodelling shared with some *S. epidermidis* infected models, or displayed a strong IL-1 alarmin, MMP-mediated wounding phenotype.

8.9 Future work

Deconvolution of immune and bacteria sensing fractions presented the issue of the dominance of immune influence on the keratinocyte transcriptome. This is demonstrated by the fact that an infection drives immune activation, but immune-mediated alteration of the microbiome is difficult to observe in the host transcriptome and more difficult still to disentangle from any background inflammation. In this context, based on data presented in this thesis and work by others (Fyhrquist *et al.*, 2019), it is further hypothesised that the KC17 fraction in atopic lesions, and oppositely to

that present in psoriatic lesions, is anti-microbial related. It may be that finding the microbiome effect on cutaneous inflammation necessitates the deconvolution of the immune signatures first, followed by investigation of these for causal programmes such as autoimmune or anti-microbial initiation. This could be investigated within the unsupervised clustering data in Section 3.2.3 by applying the *S. aureus* associated atopic lesion genes from the Fyhrquist *et al.* study for further sub-clustering and observation of expression across the tissue types. A more involved analysis might include limiting hypothesis-driven deconvolution to the transcriptome captured within a single cluster of the unsupervised analysis.

The *S. aureus* signature coincides within co-expression clusters for functional annotations involving keratinocyte differentiation and extracellular matrix organisation (Fyhrquist *et al.*, 2019). Those clusters could be easily mapped to the epidermis development (cluster 1), keratinocyte differentiation (cluster 6) and skin development (clusters 9 and 18) identified in this work (Section 3.2.3). Taking only these clusters and their constituent genes for hypothesis-driven deconvolution would effectively remove the immune-related background compartment. Additionally, as the clusters can be assumed to be an *in silico* isolated keratinocyte compartment, the deconvolution panel can be limited to solely keratinocyte signatures. Therefore, problems involving mixing single-cell and bulk reference signatures would no longer arise as only the single-cell data may be used. Consequently, the single-cell signatures can then be used within tools designed specifically for single-cell reference deconvolution of bulk data, such as Bseq-sc, CIBERSORTx, and MuSic (Baron *et al.*, 2016; Wang *et al.*, 2019; Steen *et al.*, 2020).

Alternatively, new tools such as CIBERSORTx can infer gene expression profiles of the input reference cell types alongside the estimated tissue fraction (Steen *et al.*, 2020). This new development of the algorithm would enable reanalysis of the skin samples already deconvoluted to obtain the imputed expression profile of the KC17 fraction from atopic and psoriatic lesional samples. Consequently, these inferred gene expression profiles could then be investigated for comparing *S. aureus* or auto-inflammatory programmes in each sample, either by simple expression comparison or a second round of *in silico* deconvolution using distinct anti-microbial and autoimmune defining keratinocyte KC17 profiles.

This research has investigated the importance of understanding the role of local cutaneous environment in the phenotypic programming of keratinocytes. The data derived from this analysis has shown the transcriptomic alteration of keratinocyte populations from either an immunophenotype perspective or host-microbe interactions. Valuable understanding has been derived from this, such as the stratified immune profiles of keratinocytes in the context of

commensal or pathogenic *Staphylococcal* challenge. However, the pipeline of *in vitro* modelling followed by *in vivo* deconvolution as established here is necessarily hypothesis driven. Well-defined but otherwise limited panels of signatures are used, while signatures from as-yet-unknown or complex axes are unfortunately excluded.

An attempt to address this is consideration of more complex *in vitro* systems by widening research questions. For example, epidermal model challenge with dual species of bacteria, either concurrently or sequentially. The hypothesis arising from this being that commensal challenge induces epidermal tolerance and competitively occupies important cutaneous niches to counter pathogen-associated inflammation in keratinocytes. The added advantage of this is capturing, in part, the complexity of the true cutaneous microbiome, which is composed of a vast array of colonising and interacting bacterial species.

Furthermore, the model culture system used so far directly investigates only immune-mediated or microbe-mediated epidermal responses in a separated way. This has ignored the role of these axes in synergising, amplifying or opposing their downstream effects in the epidermis. For example, deconvolution analysis indicated the importance of atopic inflammation in underscoring the epidermal phenotype of skin from patients with atopic dermatitis, therefore supplementing the RHE model media with T_H2 cytokines before bacterial challenge promises to more properly recapitulate this environment and host-microbiome interactions. Consequently, a further hypothesis forms that keratinocyte programmes and epidermal homeostasis in an inflammatory environment have different responses to bacterial than in a healthy, unperturbed context. To investigate this, epidermal models should be cultured with key disease-related cytokines (such as IL-4/IL-13, or IL-17) to establish an *in vitro* recapitulation of inflamed skin, followed by challenge with *S. epidermidis*, *S. aureus* or left sterile. Subsequent scRNA-seq analysis of these additional models can then be compared to the stratified populations described in this work to understand the direct crosstalk between endogenous inflammation and the exogenous microbiome in altering these populations. This would also increase the panel of signatures available for *in vivo* deconvolution. This potentially allows for more granular deconvolution results to stratify patients based on cutaneous perturbation or treatment responses arising from either endogenous inflammation, mishandled microbiome or a both.

Appendix A Tables

Table 7. CIBERSORT-derived barcode genes for GSE36287 samples.

IFNα	IL-13	Control	IL-17A	IL-4	TNFα
n=27	**n=0**	**n=83**	**n=36**	**n=27**	**n=169**
BST2		AADAC	ACKR4	ALOX15B	ABCG4
BTBD16		AKR1B10	APOBEC3A	ANKFN1	ADAMTS1
CFB		ALDH3A1	ARG1	CISH	ADAMTS6
CMPK2		ARRDC3	BDNF	CTSC	ADAMTSL4
DDX60L		BCHE	CCL20	FAM26E	ALDH1A3
DKK1		BCKDHB	CD36	GAST	ANKRD30BP3
HERC5		C10orf99	CHAC1	GCNT3	APCDD1L-AS1
IFI35		CABLES1	COL21A1	HSD3B1	AQP9
IFI44L		CASP14	CTH	KAL1	ASPRV1
IFIT1		CBX5	CYP26B1	LBH	ATP12A
IFIT2		CCDC88C	DDIT4	LOC100288152	ATP6V1C2
IFIT3		CDHR1	DOCK8	MAP3K14	AZGP1
IFITM1		CDKN3	ENDOU	MYRFL	BAMBI
ISG15		CFD	ID1	NABP1	BCO2
LGALS3BP		CHP2	KYNU	NELL2	BIRC3
LY6E		CLCA4	LINC00673	NNMT	C10orf10
MX2		COL5A3	MBNL1-AS1	NTRK1	C15orf48
OAS2		CYP4F12	NAMPT	PAPSS2	C2orf54
PLAC8		DAPL1	NEFL	PMP22	CDH11
PSG5		DLGAP5	NEFM	PPP1R3C	CDSN
PSMB9		DNASE1L3	NFKBIZ	RASGRP1	CEACAM6
RGCC		ECM2	PAPPA	RPTN	CHST2
RSAD2		EEF2K	PDE4B	RUNX2	CLDN17
SERPING1		EFHC2	PLAT	SLC7A4	CNFN
TLR3		EGFL6	S100A12	SOCS1	CRCT1
USP18		ELAC1	S100A7A	ST6GAL1	CSGALNACT1
XAF1		ENTPD3	SLC7A11	TNNT2	CST6
		EPHX2	SOX7		CTGF
		FAM83D	TCN1		CTSH
		FCF1	THAP2		CXCL1
		FGFR3	TNC		CXCL2
		FIBIN	TNFSF15		CYP27B1
		FOS	VLDLR		CYP4F22
		GBP2	VNN3		DCUN1D3
		GBP6	ZC3H12A		DHRS9
		HIBCH	ZC3H12C		DIRC3
		HMGN5			DLX3

IFNα	IL-13	Control	IL-17A	IL-4	TNFα
		HMMR			DNASE1L2
		IGFBP2			DRAM1
		IGFBP5			ECM1
		IL33			EPS8L1
		KIF20A			ERVH-6
		KIF2C			ESYT3
		KIF4A			FA2H
		KLRG2			FAM49A
		KRT13			FCHSD1
		KRT15			FLG2
		KRT2			FN1
		KRT8			G0S2
		LINC00273			GDA
		LOC100505664			GDPD3
		LOC100507303			GLB1L2
		LOC100507316			GLB1L3
		LOC338667			GLRX
		METTL7A			GRB7
		MT1X			HDAC9
		MYCL			HMOX1
		NDUFA4L2			HPGD
		NEK2			HSPB8
		NEURL1B			IL11
		NMU			IL15
		NREP			IL1B
		OLFML2A			IL1F10
		PART1			IL1R2
		PBX1			IL32
		PDK1			IL36G
		PKIA			IL36RN
		PPM1L			IL37
		PTGS1			IL8
		RBMS3			INHBA
		SELENBP1			INPP4B
		SLC16A10			IRAK2
		SOX6			ISM1
		SRD5A3			KCTD4
		STC2			KHDRBS3
		SULF1			KLK6
		THBD			KRT23
		TOP2A			KRT78
		TSC22D3			LCE1B

IFNα	IL-13	Control	IL-17A	IL-4	TNFα
		UNC5B-AS1			LCE1E
		WNT4			LCE2B
		WWOX			LCE3D
		ZBED3			LCN2
					LINC00302
					LINC00675
					LNX1
					LOC284454
					LPXN
					LRRC20
					LYPD5
					MAP3K9
					MIR205HG
					MIR31HG
					MMP10
					MMP9
					MUCL1
					MYCT1
					Mar-03
					NCF2
					NKPD1
					NLRP10
					NPR2
					NT5E
					OASL
					OSMR
					PLAUR
					PLCXD1
					PPIF
					PPP1R15A
					PRB1
					PRDM1
					PRR9
					PSAPL1
					PSG1
					PSG2
					PSG3
					PSG4
					PSG6
					PSG7
					PSG9
					PSORS1C2

IFNα	IL-13	Control	IL-17A	IL-4	TNFα
					PTGER3
					RAB11FIP1
					RAB3B
					RHCG
					RNASE7
					RNF39
					RRAD
					S100A7
					SCG5
					SCNN1B
					SDR9C7
					SERPINA12
					SERPINA3
					SH3GL3
					SHROOM3
					SLC15A1
					SLC26A9
					SLC5A1
					SLC6A14
					SLCO4A1
					SMPD3
					SORBS1
					SPINK13
					SPINK7
					SPNS2
					SPRR2G
					SPRR4
					STEAP1B
					STXBP6
					TEX12
					TGFB2
					TIMP2
					TMEM255A
					TMEM86A
					TNFAIP2
					TNFAIP3
					TREX2
					TRIM47
					UBE2J1
					USP2
					VGLL1
					VSIG10L

IFNα	IL-13	Control	IL-17A	IL-4	TNFα
					WFDC12
					XDH
					XKRX
					YOD1
					YPEL4
					ZNF662

Table 8. CIBERSORT-derived barcode genes for integrated dataset of GSE36287 and GSE7216 samples.

IFNα	IFNγ	IL-13	Control	IL-17A	IL-4
n=7	n=57	n=1	n=11	n=5	n=3
BST2	ACKR4	LINC00113	AKR1B10	LINC00273	HSD3B1
DKK1	AIM2		DEPDC1	NFKBIZ	LBH
HERC5	APOBEC3G		EGFL6	TNC	RPTN
IFI44L	APOL1		IGFBP5	VNN3	
IFIT1	APOL3		IL33	ZC3H12A	
ISG15	APOL6		KRT15		
RGCC	BATF2		LOC100505664		
	BTN3A2		NMU		
	BTN3A3		PBX1		
	C1R		STC2		
	C1S		TOP2A		
	CD274				
	CD74				
	CFH				
	CLIC2				
	CMPK2				
	CX3CL1				
	CXCL10				
	CXCL11				
	CXCL9				
	EDNRA				
	ERAP2				
	GBP1				
	GBP2				
	GBP4				
	GBP5				
	HCP5				
	HLA-B				
	HLA-DMA				
	HLA-DMB				

IFNα	IFNγ	IL-13	Control	IL-17A	IL-4
	HLA-DPA1				
	HLA-DPB1				
	HLA-DQB1				
	HLA-DRA				
	ICAM1				
	IFI35				
	IFIT3				
	IL18BP				
	IL32				
	IRF1				
	ISG20				
	NLRC5				
	PSMB8				
	PSMB9				
	RARRES1				
	RARRES3				
	RSAD2				
	RTP4				
	SAMD9L				
	SAMHD1				
	SECTM1				
	SERPING1				
	TAP1				
	TLR3				
	UBE2L6				
	WARS				
	XAF1				

TNF	IL22	IL26d	KGF	IL1b
N=14	**N=26**	**N=8**	**N=16**	**N=5**
ADAMTS1	CDH26	CA6	ACOX2	BIRC3
ADAMTS6	CH25H	CGA	BTC	CCL20
ATP12A	CHI3L1	DCN	CLCA4	KLK12
C10orf10	CHI3L2	KRT2	DOCK11	KRTAP19-1
C15orf48	CRISP3	MMP1	ISL1	SPINK7
CTGF	CSF2RB	MSMB	KRT13	
FN1	DPYD	PDZRN4	KRT77	
HDAC9	DPYSL3	PRR9	NDUFA4L2	
INHBA	ELF5		NEFL	
LPXN	EPB41L3		PITX2	
RHCG	FGA		RPS4Y1	
SCG5	FGG		SERPINB12	

TNF	IL22	IL26d	KGF	IL1b
TGFB2	HTR3A		SGPP2	
TNFAIP3	KLK13		SLC16A6	
	KLK6		SLC16A9	
	LRG1		UPK1A	
	PRSS27			
	S100A12			
	S100A7A			
	SERPINA1			
	SNX10			
	SOCS3			
	SYTL5			
	TCHH			
	TLR2			
	TMC5			

Table 9. Top 20 genes by rank gene t-test between basal, spinous and granular layers.

Basal	Spinous	Granular
S100A2	FABP5	LCN2
KRT14	KRT6B	S100A9
COL17A1	KRT6A	SLPI
ITGA6	KRT16	KRT19
DST	KRT17	SPRR3
RPS3	SPRR1B	CEACAM6
LAMB3	KRT5	SPINK5
LGALS1	KRT14	KLK7
LAMC2	SFN	S100A8
TMSB10	AQP3	S100P
SLC2A1	PERP	MT-RNR2
NDRG1	DSP	C15orf48
RPS6	GLTP	RHCG
RPS9	S100A14	CD24
LIMA1	KRT6C	IVL
MT2A	LY6D	LGALS3
RPL10A	SERPINB3	CLDN7
S100A14	LYPD3	PLAAT4
CXCL14	PKP1	MAL
TGFBI	SERPINB5	MT-RNR1

Table 10. Differentially expressed genes between layer clusters in the RHE scRNA-seq data as identified by MAST.

Within BASAL comparison					
Basal 2 vs Basal 1		**Basal 2 vs Basal 3**		**Basal 2 vs Basal 4**	
UP Basal-2	**UP Basal-1**	**UP Basal-2**	**UP Basal-3**	**UP Basal-2**	**UP Basal-4**
n=3	**n=1**	**n=5**	**n=14**	**n=14**	**n=26**
MALAT1	CXCL14	CXCL14	ANXA1	ATP1B3	ANXA1
MT-RNR2		DST	AQP3	CXCL14	BTG1
NEAT1		KRT5	DRAP1	DST	DRAP1
		MALAT1	EIF1	DYNLL1	F3
		MT-RNR2	HBEGF	FGFBP1	G0S2
			LGALS1	GJB2	HBEGF
			MMP1	HSPB1	IL1A
			NDRG1	KRT14	IL1B
			PHLDA1	KRT17	INHBA
			PLAU	KRT5	ITGA2
			PLIN2	KRT6A	LAMA3
			SAT1	MT-ATP6	LAMB3
			UPP1	MT-CO2	LAMC2
			ZFP36L1	MT-CO3	LGALS1
					MMP1
					MMP10
					MMP9
					MRPS6
					NDRG1
					PHLDA1
					PLAU
					PLIN2
					SAT1
					UPP1
					VIM
					ZFP36L1
Within SPINOUS comparison					
Spinous 1 vs Spinous 2		**Spinous 1 vs Spinous 3**		**Spinous 2 vs Spinous 3**	
UP Spinous-1	**UP Spinous-2**	**UP Spinous-1**	**UP Spinous-3**	**UP Spinous-2**	**UP Spinous-3**
n=5	**n=3**	**n=6**	**n=27**	**n=5**	**n=26**
CRABP2	FTL	FGFBP1	ANXA1	CXCL14	ANXA1
KRT1	MMP1	KRT10	CALML5	KRT14	CALML5
KRT10	SAT1	KRT14	CD24	KRT5	CLDN4
KRT6B		KRT5	CLDN4	S100A2	CNFN
SERPINB3		KRT6A	CNFN	SLC2A1	CRCT1
		SCD	CRCT1		CSTB

			CSTB		DMKN
			EMP1		EMP1
			GGCT		HBEGF
			HBEGF		IL36G
			IL1RN		IVL
			IL36G		KLK7
			KLK7		KRT6B
			LCE3D		LCE3D
			MMP1		LYPD3
			NEAT1		PI3
			PI3		RHCG
			PRSS22		S100A7
			RHCG		S100A8
			S100A7		S100A9
			S100A8		SBSN
			S100A9		SERPINB2
			SERPINB1		SERPINB3
			SERPINB2		SLPI
			SLPI		SPRR1B
			TM4SF1		TM4SF1
			UPP1		
Within GRANULAR comparison					
Granular 1 vs Granular 2		**Granular 1 vs Granular 3**		**Granular 2 vs Granular 3**	
UP Granular-1	**UP Granular-2**	**UP Granular-1**	**UP Granular-3**	**UP Granular-2**	**UP Granular-3**
n=0	**n=5**	**n=7**	**n=2**	**n=11**	**n=1**
	ECM1	ANXA1	FGF23	CEACAM6	HELLPAR
	MAL	KLK7	HELLPAR	ECM1	
	S100A9	KRT17		KRT19	
	SPINK5	KRT19		LCN2	
	SPRR3	LCN2		LGALS3	
		SAT1		MAL	
		TMSB4X		RHCG	
				S100A9	
				SLPI	
				SPINK5	
				SPRR3	

List of References

Van ’t Veer, L. J. *et al.* (2002) ‘Gene expression profiling predicts clinical outcome of breast cancer.’, *Nature*, 415(6871), pp. 530–6. doi: 10.1038/415530a.

Van der Aar, A. M. G. *et al.* (2013) ‘Langerhans cells favor skin flora tolerance through limited presentation of bacterial antigens and induction of regulatory T cells.’, *The Journal of investigative dermatology*. Elsevier, 133(5), pp. 1240–9. doi: 10.1038/jid.2012.500.

Abbas, A. R. *et al.* (2009) ‘Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus’, *PLoS ONE*. Edited by P. Tan, 4(7), p. e6098. doi: 10.1371/journal.pone.0006098.

Aksentijevich, I. *et al.* (2009) ‘An Autoinflammatory Disease with Deficiency of the Interleukin-1–Receptor Antagonist’, *New England Journal of Medicine*. Massachusetts Medical Society, 360(23), pp. 2426–2437. doi: 10.1056/nejmoa0807865.

Alameda, J. P. *et al.* (2016) ‘IKKα regulates the stratification and differentiation of the epidermis: Implications for skin cancer development’, *Oncotarget*. Impact Journals LLC, 7(47), pp. 76779–76792. doi: 10.18632/oncotarget.12527.

Albanesi, C. *et al.* (2001) ‘A cytokine-to-chemokine axis between T lymphocytes and keratinocytes can favor Th1 cell accumulation in chronic inflammatory skin diseases.’, *Journal of leukocyte biology*, 70(4), pp. 617–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11590199 (Accessed: 8 November 2018).

Aran, D., Hu, Z. and Butte, A. J. (2017) ‘xCell: digitally portraying the tissue cellular heterogeneity landscape’, *Genome Biology*. BioMed Central Ltd., 18(1), p. 220. doi: 10.1186/s13059-017-1349-1.

Archer, N. K. *et al.* (2016) ‘Interleukin-17A (IL-17A) and IL-17F Are Critical for Antimicrobial Peptide Production and Clearance of Staphylococcus aureus Nasal Colonization’, *Infection and Immunity*. Edited by N. E. Freitag, 84(12), pp. 3575–3583. doi: 10.1128/IAI.00596-16.

Ardern-Jones, M. R. *et al.* (2007) ‘Bacterial superantigen facilitates epithelial presentation of allergen to T helper 2 cells’, *Proceedings of the National Academy of Sciences*, 104(13), pp. 5557–5562. doi: 10.1073/pnas.0700733104.

Avila Cobos, F. *et al.* (2018) ‘Computational deconvolution of transcriptomics data from mixed cell populations’, *Bioinformatics*, 34(11), pp. 1969–1979. doi: 10.1093/bioinformatics/bty019.

Bæk, K. T. *et al.* (2013) ‘Genetic Variation in the Staphylococcus aureus 8325 Strain Lineage Revealed by Whole-Genome Sequencing’, *PLoS ONE*. PLoS One, 8(9). doi: 10.1371/journal.pone.0077122.

Bajaña, S. *et al.* (2012) ‘IRF4 promotes cutaneous dendritic cell migration to lymph nodes during homeostasis and inflammation.’, *Journal of immunology (Baltimore, Md. : 1950)*. NIH Public Access, 189(7), pp. 3368–77. doi: 10.4049/jimmunol.1102613.

Baker, B. S. *et al.* (2003) ‘Normal keratinocytes express Toll-like receptors (TLRs) 1, 2 and 5: modulation of TLR expression in chronic plaque psoriasis.’, *The British journal of dermatology*, 148(4), pp. 670–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12752123 (Accessed: 8 November 2018).

Banchereau, J. and Steinman, R. M. (1998) ‘Dendritic cells and the control of immunity.’, *Nature*, 392(6673), pp. 245–52. doi: 10.1038/32588.

Banchereau, R. *et al.* (2012) 'Host immune transcriptional profiles reflect the variability in clinical disease manifestations in patients with Staphylococcus aureus infections.', *PloS one*. Public Library of Science, 7(4), p. e34390. doi: 10.1371/journal.pone.0034390.

Banerjee, G. *et al.* (2004) 'Role of Keratinocytes in Antigen Presentation and Polarization of Human T Lymphocytes', *Scandinavian Journal of Immunology*, 59(4), pp. 385–394. doi: 10.1111/j.0300-9475.2004.01394.x.

Bao, L. *et al.* (2017) 'A molecular mechanism for IL-4 suppression of loricrin transcription in epidermal keratinocytes: implication for atopic dermatitis pathogenesis', *Innate Immunity*. SAGE Publications Ltd, 23(8), pp. 641–647. doi: 10.1177/1753425917732823.

Baron, M. *et al.* (2016) 'A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure', *Cell Systems*, 3(4), pp. 346-360.e4. doi: 10.1016/j.cels.2016.08.011.

Barrett, T. *et al.* (2011) 'NCBI GEO: archive for functional genomics data sets--10 years on.', *Nucleic acids research*, 39(Database issue), pp. D1005-10. doi: 10.1093/nar/gkq1184.

Bata-Csorgo, Z. *et al.* (1995) 'Kinetics and regulation of human keratinocyte stem cell growth in short-term primary ex vivo culture: Cooperative growth factors from psoriatic lesional T lymphocytes stimulate proliferation among psoriatic uninvolved, but not normal, stem keratinocytes', *Journal of Clinical Investigation*. The American Society for Clinical Investigation, 95(1), pp. 317–327. doi: 10.1172/JCI117659.

Belkaid, Y. and Segre, J. A. (2014) 'Dialogue between skin microbiota and immunity.', *Science (New York, N.Y.)*, 346(6212), pp. 954–9. doi: 10.1126/science.1260144.

Bell, E. *et al.* (1981) 'Living tissue formed in vitro and accepted as skin-equivalent tissue of full thickness', *Science*. Science, 211(4486), pp. 1052–1054. doi: 10.1126/science.7008197.

De Benedetto, A. *et al.* (2011) 'Tight junction defects in patients with atopic dermatitis', *Journal of Allergy and Clinical Immunology*. Mosby Inc., 127(3), pp. 773-786.e7. doi: 10.1016/j.jaci.2010.10.018.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple', *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp. 289–300.

Bennett, C. L. *et al.* (2005) 'Inducible ablation of mouse Langerhans cells diminishes but fails to abrogate contact hypersensitivity.', *The Journal of cell biology*. The Rockefeller University Press, 169(4), pp. 569–76. doi: 10.1083/jcb.200501071.

Berger, A. (2000) 'Th1 and Th2 responses: what are they?', *BMJ (Clinical research ed.)*, 321(7258), p. 424.

Bernard, F.-X. *et al.* (2012) 'Keratinocytes under Fire of Proinflammatory Cytokines: Bona Fide Innate Immune Cells Involved in the Physiopathology of Chronic Atopic Dermatitis and Psoriasis.', *Journal of allergy*. Hindawi Limited, 2012, p. 718725. doi: 10.1155/2012/718725.

Berry, M. P. R. *et al.* (2010) 'An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.', *Nature*. Europe PMC Funders, 466(7309), pp. 973–7. doi: 10.1038/nature09247.

Bieber, T. (2008) 'Atopic Dermatitis', *New England Journal of Medicine*, 358(14), pp. 1483–1494. doi: 10.1056/NEJMra074081.

Bigler, J. *et al.* (2013) 'Cross-study homogeneity of psoriasis gene expression in skin across a large

expression range.', *PloS one*. Edited by J. M. Brandner, 8(1), p. e52242. doi: 10.1371/journal.pone.0052242.

Bikle, D. D., Xie, Z. and Tu, C. L. (2012) 'Calcium regulation of keratinocyte differentiation', *Expert Review of Endocrinology and Metabolism*. Expert Rev Endocrinol Metab, pp. 461–472. doi: 10.1586/eem.12.34.

Bitschar, K. *et al.* (2017) 'Keratinocytes as sensors and central players in the immune defense against Staphylococcus aureus in the skin', *Journal of Dermatological Science*. Elsevier Ireland Ltd, pp. 215–220. doi: 10.1016/j.jdermsci.2017.06.003.

Boniface, K. *et al.* (2007) 'A role for T cell-derived interleukin 22 in psoriatic skin inflammation', *Clinical and Experimental Immunology*. Clin Exp Immunol, 150(3), pp. 407–415. doi: 10.1111/j.1365-2249.2007.03511.x.

Boniface, Katia *et al.* (2007) 'Oncostatin M secreted by skin infiltrating T lymphocytes is a potent keratinocyte activator involved in skin inflammation.', *Journal of immunology (Baltimore, Md. : 1950)*, 178(7), pp. 4615–22.

Boukamp, P. *et al.* (1988) 'Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line', *Journal of Cell Biology*. The Rockefeller University Press, 106(3), pp. 761–771. doi: 10.1083/jcb.106.3.761.

Brandt, E. B. and Sivaprasad, U. (2011) 'Th2 Cytokines and Atopic Dermatitis.', *Journal of clinical & cellular immunology*. NIH Public Access, 2(3). doi: 10.4172/2155-9899.1000110.

Braum, O., Pirzer, H. and Fickenscher, H. (2012) 'Interleukin-26, a highly cationic T-cell cytokine targeting epithelial cells.', *Anti-inflammatory & anti-allergy agents in medicinal chemistry*, 11(3), pp. 221–9.

Brauweiler, A. M. *et al.* (2015) 'Th2 Cytokines Suppress Lipoteichoic Acid-Induced Matrix Metalloproteinase Expression and Keratinocyte Migration in Response to Wounding', *Journal of Investigative Dermatology*. Nature Publishing Group, pp. 2550–2553. doi: 10.1038/jid.2015.181.

Brazma, A. *et al.* (2001) 'Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.', *Nature genetics*, 29(4), pp. 365–71. doi: 10.1038/ng1201-365.

Brunner, P. M. *et al.* (2017) 'Nonlesional atopic dermatitis skin shares similar T-cell clones with lesional tissues', *Allergy*, 72(12), pp. 2017–2025. doi: 10.1111/all.13223.

Burian, M. *et al.* (2017) 'The Protective Effect of Microbiota on S. aureus Skin Colonization Depends on the Integrity of the Epithelial Barrier.', *The Journal of investigative dermatology*, 137(4), pp. 976–979. doi: 10.1016/j.jid.2016.11.024.

Burns, R. *et al.* (2005) 'Keratinocyte-derived, CD80-mediated costimulation is associated with hapten-specific IgE production during contact hypersensitivity to TH1 haptens', *Journal of Allergy and Clinical Immunology*, 115(2), pp. 383–390. doi: 10.1016/j.jaci.2004.11.019.

Byrd, A. L., Belkaid, Y. and Segre, J. A. (2018) 'The human skin microbiome', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 143–155. doi: 10.1038/nrmicro.2017.157.

Cahan, P. *et al.* (2007) 'Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization.', *Gene*, 401(1–2), pp. 12–8.

Cai, Y., Fleming, C. and Yan, J. (2012) 'New insights of T cells in the pathogenesis of psoriasis', *Cellular and Molecular Immunology*. Cell Mol Immunol, pp. 302–309. doi: 10.1038/cmi.2012.15.

Cameron, S. *et al.* (2012) 'Focal overexpression of CEACAM6 contributes to enhanced

tumourigenesis in head and neck cancer via suppression of apoptosis', *Molecular Cancer*. Mol Cancer, 11. doi: 10.1186/1476-4598-11-74.

Carpentier, S. *et al.* (2016) 'Comparative genomics analysis of mononuclear phagocyte subsets confirms homology between lymphoid tissue-resident and dermal XCR1+ DCs in mouse and human and distinguishes them from Langerhans cells', *Journal of Immunological Methods*, 432, pp. 35–49. doi: 10.1016/j.jim.2016.02.023.

Carrier, Y. *et al.* (2011) 'Inter-regulation of Th17 cytokines and the IL-36 cytokines in vitro and in vivo: Implications in psoriasis pathogenesis', *Journal of Investigative Dermatology*. Nature Publishing Group, 131(12), pp. 2428–2437. doi: 10.1038/jid.2011.234.

Carrieri, A. P. *et al.* (2021) 'Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences', *Scientific Reports*. Nature Publishing Group, 11(1), p. 4565. doi: 10.1038/s41598-021-83922-6.

Chang, H.-W. *et al.* (2018) 'Alteration of the cutaneous microbiome in psoriasis and potential role in Th17 polarization.', *Microbiome*, 6(1), p. 154. doi: 10.1186/s40168-018-0533-1.

Chen, B. *et al.* (2018) 'Profiling Tumor Infiltrating Immune Cells with CIBERSORT.', *Methods in molecular biology (Clifton, N.J.)*, 1711, pp. 243–259. doi: 10.1007/978-1-4939-7493-1_12.

Chen, J. *et al.* (2009) 'ToppGene Suite for gene list enrichment analysis and candidate gene prioritization', *Nucleic Acids Research*, 37(Web Server), pp. W305–W311. doi: 10.1093/nar/gkp427.

Cheng, J. B. *et al.* (2018) 'Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution.', *Cell reports*, 25(4), pp. 871–883. doi: 10.1016/j.celrep.2018.09.006.

Chiricozzi, A. *et al.* (2011) 'Integrative responses to IL-17 and TNF-α in human keratinocytes account for key inflammatory pathogenic circuits in psoriasis.', *The Journal of investigative dermatology*. Elsevier, 131(3), pp. 677–87. doi: 10.1038/jid.2010.340.

Cho, M.-L. *et al.* (2007) 'Cyclosporine A inhibits IL-15-induced IL-17 production in CD4+ T cells via down-regulation of PI3K/Akt and NF-kappaB.', *Immunology letters*, 108(1), pp. 88–96. doi: 10.1016/j.imlet.2006.11.001.

Choy, D. F. *et al.* (2012) 'Comparative transcriptomic analyses of atopic dermatitis and psoriasis reveal shared neutrophilic inflammation.', *The Journal of allergy and clinical immunology*, 130(6), pp. 1335–43.e5. doi: 10.1016/j.jaci.2012.06.044.

Chuong, C. M. *et al.* (2002) 'What is the "true" function of skin?', *Experimental dermatology*, 11(2), pp. 159–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11994143 (Accessed: 7 November 2018).

Clancy, T. *et al.* (2017) 'Bioinformatics Approaches to Profile the Tumor Microenvironment for Immunotherapeutic Discovery.', *Current pharmaceutical design*, 23(32), pp. 4716–4725. doi: 10.2174/1381612823666170710154936.

Clark, R. A. *et al.* (2006) 'The vast majority of CLA+ T cells are resident in normal skin.', *Journal of immunology (Baltimore, Md. : 1950)*, 176(7), pp. 4431–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16547281 (Accessed: 8 November 2018).

Clayton, K., Polak, M. E., *et al.* (2017) 'Gene expression signatures in tuberculosis have greater overlap with autoimmune diseases than with infectious diseases', *American Journal of Respiratory and Critical Care Medicine*, 196(5). doi: 10.1164/rccm.201706-1248LE.

Clayton, K., Vallejo, A. F., *et al.* (2017) 'Langerhans cells-programmed by the epidermis', *Frontiers in Immunology*, 8(NOV). doi: 10.3389/fimmu.2017.01676.

Clayton, K. *et al.* (2020) 'Machine learning applied to atopic dermatitis transcriptome reveals distinct therapy-dependent modification of the keratinocyte immunophenotype', *British Journal of Dermatology*, p. bjd.19431. doi: 10.1111/bjd.19431.

Cloonan, N. *et al.* (2008) 'Stem cell transcriptome profiling via massive-scale mRNA sequencing', *Nature Methods*, 5(7), pp. 613–619. doi: 10.1038/nmeth.1223.

Coda, A. B. *et al.* (2012) 'Global transcriptional analysis of psoriatic skin and blood confirms known disease-associated pathways and highlights novel genomic "hot spots" for differentially expressed genes.', *Genomics*, 100(1), pp. 18–26. doi: 10.1016/j.ygeno.2012.05.004.

Cohen, P. (2014) 'The TLR and IL-1 signalling network at a glance', *Journal of Cell Science*. Company of Biologists Ltd, 127(11), pp. 2383–2390. doi: 10.1242/jcs.149831.

Cork, M. J. *et al.* (2009) 'Epidermal barrier dysfunction in atopic dermatitis', *Journal of Investigative Dermatology*. Nature Publishing Group, 129(8), pp. 1892–1908. doi: 10.1038/jid.2009.133.

Correa da Rosa, J. *et al.* (2017) 'Shrinking the Psoriasis Assessment Gap: Early Gene-Expression Profiling Accurately Predicts Response to Long-Term Treatment', *Journal of Investigative Dermatology*, 137(2), pp. 305–312. doi: 10.1016/j.jid.2016.09.015.

Dainichi, T., Hanakawa, S. and Kabashima, K. (2014) 'Classification of inflammatory skin diseases: a proposal based on the disorders of the three-layered defense systems, barrier, innate immunity and acquired immunity.', *Journal of dermatological science*. Elsevier, 76(2), pp. 81–9. doi: 10.1016/j.jdermsci.2014.08.010.

Danilenko, D. M. (1999) 'Preclinical and Early Clinical Development of Keratinocyte Growth Factor, an Epithelial- Specific Tissue Growth Factor', *Toxicologic Pathology*, 27(1), pp. 64–71. doi: 10.1177/019262339902700113.

Danso, M. *et al.* (2017) 'Altered expression of epidermal lipid bio-synthesis enzymes in atopic dermatitis skin is accompanied by changes in stratum corneum lipid composition', *Journal of Dermatological Science*, 88(1), pp. 57–66. doi: 10.1016/j.jdermsci.2017.05.005.

Denfeld, R. W. *et al.* (1996) 'CD40 is functionally expressed on human keratinocytes', *European Journal of Immunology*. Eur J Immunol, 26(10), pp. 2329–2334. doi: 10.1002/eji.1830261009.

Dhingra, N., Gulati, N. and Guttman-Yassky, E. (2013) 'Mechanisms of contact sensitization offer insights into the role of barrier defects vs. intrinsic immune abnormalities as drivers of atopic dermatitis.', *The Journal of investigative dermatology*, 133(10), pp. 2311–2314. doi: 10.1038/jid.2013.239.

Dinarello, C. A. (2018) 'Overview of the IL-1 family in innate inflammation and acquired immunity', *Immunological Reviews*. Blackwell Publishing Ltd, pp. 8–27. doi: 10.1111/imr.12621.

Dinges, M. M., Orwin, P. M. and Schlievert, P. M. (2000) 'Exotoxins of Staphylococcus aureus', *Clinical Microbiology Reviews*. American Society for Microbiology, 13(1), pp. 16–34. doi: 10.1128/CMR.13.1.16-34.2000.

Dixon, B. R. E. A. *et al.* (2016) 'IL-17a and IL-22 Induce Expression of Antimicrobials in Gastrointestinal Epithelial Cells and May Contribute to Epithelial Cell Defense against Helicobacter pylori', *PLOS ONE*. Edited by P. L. Ho, 11(2), p. e0148514. doi: 10.1371/journal.pone.0148514.

Doebel, T., Voisin, B. and Nagao, K. (2017) 'Langerhans Cells – The Macrophage in Dendritic Cell Clothing', *Trends in Immunology*, 38(11), pp. 817–828. doi: 10.1016/j.it.2017.06.008.

Doig, T. N. *et al.* (2013) 'Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment.', *BMC genomics*, 14(1), p. 469. doi:

10.1186/1471-2164-14-469.

Dong, Y., Speer, C. P. and Glaser, K. (2018) 'Beyond sepsis: Staphylococcus epidermidis is an underestimated but significant contributor to neonatal morbidity.', *Virulence*, 9(1), pp. 621–633. doi: 10.1080/21505594.2017.1419117.

van Dongen, S. (2000) *Graph Clustering by Flow Simulation*. University of Utrecht.

Duckney, P. *et al.* (2013) 'The role of the skin barrier in modulating the effects of common skin microbial species on the inflammation, differentiation and proliferation status of epidermal keratinocytes.', *BMC research notes*, 6(1), p. 474. doi: 10.1186/1756-0500-6-474.

Duluc, D. *et al.* (2014) 'Transcriptional fingerprints of antigen-presenting cell subsets in the human vaginal mucosa and skin reflect tissue-specific immune microenvironments.', *Genome medicine*. BioMed Central, 6(11), p. 98. doi: 10.1186/s13073-014-0098-y.

Eckhart, L. *et al.* (2013) 'Cell death by cornification', *Biochimica et Biophysica Acta - Molecular Cell Research*. Biochim Biophys Acta, pp. 3471–3480. doi: 10.1016/j.bbamcr.2013.06.010.

Edgar, R. and Barrett, T. (2006) 'NCBI GEO standards and services for microarray data.', *Nature biotechnology*, 24(12), pp. 1471–2. doi: 10.1038/nbt1206-1471.

Eichner, R., Sun, T. T. and Aebi, U. (1986) 'The role of keratin subfamilies and keratin pairs in the formation of human epidermal intermediate filaments', *Journal of Cell Biology*. J Cell Biol, 102(5), pp. 1767–1777. doi: 10.1083/jcb.102.5.1767.

Elias, P. M. *et al.* (2014) 'Formation and functions of the corneocyte lipid envelope (CLE)', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. Biochim Biophys Acta, pp. 314–318. doi: 10.1016/j.bbalip.2013.09.011.

Eriksen, N. H. R. *et al.* (1995) 'Carriage of Staphylococcus aureus among 104 healthy persons during a 19-month period', *Epidemiology and Infection*. Epidemiol Infect, 115(1), pp. 51–60. doi: 10.1017/S0950268800058118.

Esaki, H. *et al.* (2015) 'Identification of novel immune and barrier genes in atopic dermatitis by means of laser capture microdissection.', *The Journal of allergy and clinical immunology*. NIH Public Access, 135(1), pp. 153–63. doi: 10.1016/j.jaci.2014.10.037.

Eyerich, K. *et al.* (2009) 'IL-17 in atopic eczema: linking allergen-specific adaptive and microbial-triggered innate immune response.', *The Journal of allergy and clinical immunology*, 123(1), pp. 59-66.e4. doi: 10.1016/j.jaci.2008.10.031.

Federici, M. *et al.* (2002) 'Impaired IFN-gamma-dependent inflammatory responses in human keratinocytes overexpressing the suppressor of cytokine signaling 1.', *Journal of immunology (Baltimore, Md. : 1950)*, 169(1), pp. 434–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12077274 (Accessed: 17 December 2018).

Félix Garza, Z. C. *et al.* (2019) 'Characterization of disease-specific cellular abundance profiles of chronic inflammatory skin conditions from deconvolution of biopsy samples', *BMC Medical Genomics*. BioMed Central Ltd., 12(1), p. 121. doi: 10.1186/s12920-019-0567-7.

Ferreira, M. A. *et al.* (2017) 'Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology', *Nature Genetics*. Nature Publishing Group, 49(12), pp. 1752–1757. doi: 10.1038/ng.3985.

Finak, G. *et al.* (2015) 'MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data', *Genome Biology*. BioMed Central Ltd., 16(1). doi: 10.1186/s13059-015-0844-5.

Freeman, T. C. *et al.* (2007) 'Construction, visualisation, and clustering of transcription networks from microarray expression data.', *PLoS computational biology*. Public Library of Science, 3(10), pp. 2032–42. doi: 10.1371/journal.pcbi.0030206.

Fulton, C. *et al.* (1997) 'Expression of natural peptide antibiotics in human skin', *The Lancet*, 350(9093), pp. 1750–1751. doi: 10.1016/S0140-6736(05)63574-X.

Furman, D. *et al.* (2019) 'Chronic inflammation in the etiology of disease across the life span', *Nature Medicine*. Nature Research, 25(12), pp. 1822–1832. doi: 10.1038/s41591-019-0675-0.

Fyhrquist, N. *et al.* (2019) 'Microbe-host interplay in atopic dermatitis and psoriasis', *Nature Communications*. Nature Publishing Group, 10(1). doi: 10.1038/s41467-019-12253-y.

Ganguly, D. *et al.* (2009) 'Self-RNA-antimicrobial peptide complexes activate human dendritic cells through TLR7 and TLR8', *Journal of Experimental Medicine*. J Exp Med, 206(9), pp. 1983–1994. doi: 10.1084/jem.20090480.

Gebhardt, T. *et al.* (2009) 'Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus', *Nature Immunology*. Nat Immunol, 10(5), pp. 524–530. doi: 10.1038/ni.1718.

Gebhardt, T. *et al.* (2011) 'Different patterns of peripheral migration by memory CD4+ and CD8+ T cells', *Nature*. Nature, 477(7363), pp. 216–219. doi: 10.1038/nature10339.

Ghosh, D. *et al.* (2015) 'Multiple Transcriptome Data Analysis Reveals Biologically Relevant Atopic Dermatitis Signature Genes and Pathways.', *PloS one*. Edited by M. Simon, 10(12), p. e0144316. doi: 10.1371/journal.pone.0144316.

Gilliet, M. and Lande, R. (2008) 'Antimicrobial peptides and self-DNA in autoimmune skin inflammation', *Current Opinion in Immunology*. Curr Opin Immunol, pp. 401–407. doi: 10.1016/j.coi.2008.06.008.

Gittler, J. K. *et al.* (2012) 'Progressive activation of T(H)2/T(H)22 cytokines and selective epidermal proteins characterizes acute and chronic atopic dermatitis.', *The Journal of allergy and clinical immunology*, 130(6), pp. 1344–54. doi: 10.1016/j.jaci.2012.07.012.

Golub, T. R. *et al.* (1999) 'Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.', *Science (New York, N.Y.)*, 286(5439), pp. 531–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10521349 (Accessed: 9 November 2018).

Gong, J. Q. *et al.* (2006) 'Skin colonization by Staphylococcus aureus in patients with eczema and atopic dermatitis and relevant combined topical therapy: a double-blind multicentre randomized controlled trial', *British Journal of Dermatology*, 155(4), pp. 680–687. doi: 10.1111/j.1365-2133.2006.07410.x.

Gong, T. *et al.* (2011) 'Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples.', *PloS one*. Edited by M. Rattray, 6(11), p. e27156. doi: 10.1371/journal.pone.0027156.

Gong, X. and Wang, W. (2020) 'Profiles of Innate Immune Cell Infiltration and Related Core Genes in Psoriasis'. doi: 10.21203/RS.3.RS-63594/V1.

Grice, E. A. *et al.* (2009) 'Topographical and temporal diversity of the human skin microbiome', *Science*. Science, 324(5931), pp. 1190–1192. doi: 10.1126/science.1171700.

Grice, E. A. (2015) 'The intersection of microbiome and host at the skin interface: genomic- and metagenomic-based insights.', *Genome research*, 25(10), pp. 1514–20. doi: 10.1101/gr.191320.115.

Grice, E. A. and Segre, J. A. (2011) 'The skin microbiome', *Nat Rev Microbiol*, 9(4), pp. 244–253. doi: 10.1038/nrmicro2537.

Guilloteau, K. *et al.* (2010) 'Skin Inflammation Induced by the Synergistic Action of IL-17A, IL-22, Oncostatin M, IL-1{alpha}, and TNF-{alpha} Recapitulates Some Features of Psoriasis.', *Journal of immunology (Baltimore, Md. : 1950)*, 184(9), pp. 5263–5270. doi: 10.4049/jimmunol.0902464.

Gutowska-Owsiak, D. *et al.* (2012) 'IL-17 downregulates filaggrin and affects keratinocyte expression of genes associated with cellular adhesion.', *Experimental dermatology*, 21(2), pp. 104–10. doi: 10.1111/j.1600-0625.2011.01412.x.

Guttman-Yassky, E. *et al.* (2008) 'Low expression of the IL-23/Th17 pathway in atopic dermatitis compared to psoriasis.', *Journal of immunology (Baltimore, Md. : 1950)*, 181(10), pp. 7420–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18981165 (Accessed: 17 December 2018).

Guttman-Yassky, E. *et al.* (2019) 'Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis', *Journal of Allergy and Clinical Immunology*. Mosby, 143(1), pp. 155–172. doi: 10.1016/J.JACI.2018.08.022.

Guttman-Yassky, E. and Krueger, J. G. (2017) 'Atopic dermatitis and psoriasis: two different immune diseases or one spectrum?', *Current Opinion in Immunology*. Elsevier Current Trends, 48, pp. 68–73. doi: 10.1016/J.COI.2017.08.008.

Guttman-Yassky, E., Nograles, K. E. and Krueger, J. G. (2011a) 'Contrasting pathogenesis of atopic dermatitis and psoriasis--part I: clinical and pathologic concepts.', *The Journal of allergy and clinical immunology*, 127(5), pp. 1110–8. doi: 10.1016/j.jaci.2011.01.053.

Guttman-Yassky, E., Nograles, K. E. and Krueger, J. G. (2011b) 'Contrasting pathogenesis of atopic dermatitis and psoriasis--part II: immune cell subsets and therapeutic concepts.', *The Journal of allergy and clinical immunology*, 127(6), pp. 1420–32. doi: 10.1016/j.jaci.2011.01.054.

Guzik, T. J. *et al.* (2005) 'Persistent skin colonization with Staphylococcus aureus in atopic dermatitis: relationship to clinical and immunological parameters.', *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, 35(4), pp. 448–55. doi: 10.1111/j.1365-2222.2005.02210.x.

Hamid, Q., Boguniewicz, M. and Leung, D. Y. (1994) 'Differential in situ cytokine gene expression in acute versus chronic atopic dermatitis.', *Journal of Clinical Investigation*, 94(2), pp. 870–876. doi: 10.1172/JCI117408.

Hanssen, A. M. *et al.* (2017) 'Localization of Staphylococcus aureus in tissue from the nasal vestibule in healthy carriers', *BMC Microbiology*. BioMed Central Ltd., 17(1). doi: 10.1186/s12866-017-0997-3.

Harper, E. G. *et al.* (2009) 'Th17 cytokines stimulate CCL20 expression in keratinocytes in vitro and in vivo: Implications for psoriasis pathogenesis', *Journal of Investigative Dermatology*. J Invest Dermatol, 129(9), pp. 2175–2183. doi: 10.1038/jid.2009.65.

Hauser, C. *et al.* (1985) 'Staphylococcus aureus skin colonization in atopic dermatitis patients.', *Dermatologica*, 170(1), pp. 35–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3972149 (Accessed: 22 February 2019).

He, H. *et al.* (2020) 'Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis', *Journal of Allergy and Clinical Immunology*. Mosby Inc., 145(6), pp. 1615–1628. doi: 10.1016/j.jaci.2020.01.042.

Hennies, H. C. and Poumay, Y. (2021) 'Skin Disease Models In Vitro and Inflammatory Mechanisms: Predictability for Drug Development', in *Handbook of experimental pharmacology*. Handb Exp

Pharmacol. doi: 10.1007/164_2020_428.

Hennings, H. *et al.* (1980) 'Calcium regulation of growth and differentiation of mouse epidermal cells in culture', *Cell*. Cell, 19(1), pp. 245–254. doi: 10.1016/0092-8674(80)90406-7.

Hernández-Quintero, M. *et al.* (2006) 'Interleukin-6 promotes human epidermal keratinocyte proliferation and keratin cytoskeleton reorganization in culture', *Cell and Tissue Research*. Cell Tissue Res, 325(1), pp. 77–90. doi: 10.1007/s00441-006-0173-9.

Hoath, S. B. and Leahy, D. G. (2003) 'The Organization of Human Epidermis: Functional Epidermal Units and Phi Proportionality', *Journal of Investigative Dermatology*, 121(6), pp. 1440–1446. doi: 10.1046/j.1523-1747.2003.12606.x.

Hoek, K. *et al.* (2004) 'Expression profiling reveals novel pathways in the transformation of melanocytes to melanomas.', *Cancer research*, 64(15), pp. 5270–82. doi: 10.1158/0008-5472.CAN-04-0731.

Holbrook, D. J. (2019) *Host cellular and immune responses in models of inflammatory skin conditions*. University of Southampton.

Hollestein, L. M. and Nijsten, T. (2014) 'An insight into the global burden of skin diseases.', *The Journal of investigative dermatology*, 134(6), pp. 1499–1501. doi: 10.1038/jid.2013.513.

Hönzke, S. *et al.* (2016) 'Influence of Th2 Cytokines on the Cornified Envelope, Tight Junction Proteins, and ß-Defensins in Filaggrin-Deficient Skin Equivalents.', *The Journal of investigative dermatology*, 136(3), pp. 631–639. doi: 10.1016/j.jid.2015.11.007.

Hopkinson, S. B. and Jones, J. C. R. (2000) 'The N terminus of the transmembrane protein BP180 interacts with the N- terminal domain of BP230, thereby mediating keratin cytoskeleton anchorage to the cell surface at the site of the hemidesmosome', *Molecular Biology of the Cell*. American Society for Cell Biology, 11(1), pp. 277–286. doi: 10.1091/mbc.11.1.277.

Hruz, P. *et al.* (2009) 'NOD2 contributes to cutaneous defense against Staphylococcus aureus through α-toxin-dependent innate immune activation', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 106(31), pp. 12873–12878. doi: 10.1073/pnas.0904958106.

Hui, L. *et al.* (2014) 'Calcitriol inhibits keratinocyte proliferation by upregulating leukocyte elastase inhibitor (serpin B1)', *Journal of Dermatology*. Blackwell Publishing Ltd, 41(5), pp. 393–398. doi: 10.1111/1346-8138.12434.

Inkeles, M. S. *et al.* (2015) 'Comparison of molecular signatures from multiple skin diseases identifies mechanisms of immunopathogenesis.', *The Journal of investigative dermatology*, 135(1), pp. 151–9. doi: 10.1038/jid.2014.352.

Irvine, A. D., McLean, W. H. I. and Leung, D. Y. M. (2011) 'Filaggrin Mutations Associated with Skin and Allergic Diseases', *New England Journal of Medicine*. Massachusetts Medical Society, 365(14), pp. 1315–1327. doi: 10.1056/nejmra1011040.

Itoh, T. *et al.* (2018) 'Biological Effects of IL-26 on T Cell–Mediated Skin Inflammation, Including Psoriasis', *Journal of Investigative Dermatology*. doi: 10.1016/j.jid.2018.09.037.

Iwakura, Y. and Ishigame, H. (2006) 'The IL-23/IL-17 axis in inflammation', *Journal of Clinical Investigation*. J Clin Invest, pp. 1218–1222. doi: 10.1172/JCI28508.

Iwase, T. *et al.* (2010) 'Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization.', *Nature*, 465(7296), pp. 346–9. doi: 10.1038/nature09074.

Janeway, C. A. (1989) 'Approaching the asymptote? Evolution and revolution in immunology', in *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harb Symp Quant Biol, pp. 1–13. doi: 10.1101/sqb.1989.054.01.003.

Janson, D. G. *et al.* (2012) 'Different gene expression patterns in human papillary and reticular fibroblasts.', *The Journal of investigative dermatology*, 132(11), pp. 2565–72. doi: 10.1038/jid.2012.192.

Jensen, J.-M. *et al.* (2004) 'Impaired Sphingomyelinase Activity and Epidermal Differentiation in Atopic Dermatitis', *Journal of Investigative Dermatology*, 122(6), pp. 1423–1431. doi: 10.1111/j.0022-202X.2004.22621.x.

Jiang, X. *et al.* (2012) 'Skin infection generates non-migratory memory CD8+ T(RM) cells providing global skin immunity.', *Nature*, 483(7388), pp. 227–31. doi: 10.1038/nature10851.

Jin, F. Y. *et al.* (1997) 'Secretory leukocyte protease inhibitor: A macrophage product induced by and antagonistic to bacterial lipopolysaccharide', *Cell*. Cell Press, 88(3), pp. 417–426. doi: 10.1016/S0092-8674(00)81880-2.

Johnson, W. E., Li, C. and Rabinovic, A. (2007) 'Adjusting batch effects in microarray expression data using empirical Bayes methods.', *Biostatistics (Oxford, England)*, 8(1), pp. 118–27. doi: 10.1093/biostatistics/kxj037.

Johnston, A. *et al.* (2011) 'IL-1F5, -F6, -F8, and -F9: A Novel IL-1 Family Signaling System That Is Active in Psoriasis and Promotes Keratinocyte Antimicrobial Peptide Expression', *The Journal of Immunology*. The American Association of Immunologists, 186(4), pp. 2613–2622. doi: 10.4049/jimmunol.1003162.

Joost, S. *et al.* (2016) 'Single-Cell Transcriptomics Reveals that Differentiation and Spatial Signatures Shape Epidermal and Hair Follicle Heterogeneity', *Cell Systems*. Cell Press, 3(3), pp. 221-237.e9. doi: 10.1016/j.cels.2016.08.010.

Kabashima, K. (2013) 'New concept of the pathogenesis of atopic dermatitis: interplay among the barrier, allergy, and pruritus as a trinity.', *Journal of dermatological science*, 70(1), pp. 3–11. doi: 10.1016/j.jdermsci.2013.02.001.

Kaesler, S. *et al.* (2014) 'Toll-like receptor 2 ligands promote chronic atopic dermatitis through IL-4-mediated suppression of IL-10', *Journal of Allergy and Clinical Immunology*. Mosby Inc., 134(1). doi: 10.1016/j.jaci.2014.02.017.

Kagami, S. *et al.* (2010) 'Circulating Th17, Th22, and Th1 cells are increased in psoriasis', *Journal of Investigative Dermatology*. J Invest Dermatol, 130(5), pp. 1373–1383. doi: 10.1038/jid.2009.399.

Kalinin, A., Marekov, L. N. and Steinert, P. M. (2001) 'Assembly of the epidermal cornified cell envelope.', *Journal of cell science*, 114(Pt 17), pp. 3069–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11590230 (Accessed: 8 November 2018).

Kanitakis, J. (2002) 'Anatomy, histology and immunohistochemistry of normal human skin.', *European journal of dermatology : EJD*, 12(4), pp. 390–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12095893 (Accessed: 7 November 2018).

Kawasaki, T. and Kawai, T. (2014) 'Toll-like receptor signaling pathways', *Frontiers in Immunology*. Frontiers Media S.A. doi: 10.3389/fimmu.2014.00461.

Kessler, C. M., Nussbaum, E. and Tuazon, C. U. (1991) 'Disseminated intravascular coagulation associated with staphylococcus aureus septicemia is mediated by peptidoglycan-induced platelet aggregation', *Journal of Infectious Diseases*. J Infect Dis, 164(1), pp. 101–107. doi: 10.1093/infdis/164.1.101.

Khattri, S. *et al.* (2014) 'Cyclosporine in patients with atopic dermatitis modulates activated inflammatory pathways and reverses epidermal pathology.', *The Journal of allergy and clinical immunology*. NIH Public Access, 133(6), pp. 1626–34. doi: 10.1016/j.jaci.2014.03.003.

Kim, B. E. and Leung, D. Y. (2012) 'Epidermal barrier in atopic dermatitis.', *Allergy, asthma & immunology research*, 4(1), pp. 12–6. doi: 10.4168/aair.2012.4.1.12.

Kim, B. S. *et al.* (2009) 'Keratinocytes function as accessory cells for presentation of endogenous antigen expressed in the epidermis.', *The Journal of investigative dermatology*. NIH Public Access, 129(12), pp. 2805–17. doi: 10.1038/jid.2009.176.

Kisich, K. O. *et al.* (2007) 'The constitutive capacity of human keratinocytes to kill Staphylococcus aureus is dependent on β-defensin 3', *Journal of Investigative Dermatology*. J Invest Dermatol, 127(10), pp. 2368–2380. doi: 10.1038/sj.jid.5700861.

Klechevsky, E. *et al.* (2008) 'Functional Specializations of Human Epidermal Langerhans Cells and CD14+ Dermal Dendritic Cells', *Immunity*, 29(3), pp. 497–510. doi: 10.1016/j.immuni.2008.07.013.

Kobayashi, T. *et al.* (2015) 'Dysbiosis and Staphylococcus aureus Colonization Drives Inflammation in Atopic Dermatitis.', *Immunity*, 42(4), pp. 756–66. doi: 10.1016/j.immuni.2015.03.014.

Koga, C. *et al.* (2008) 'Possible pathogenic role of Th17 cells for atopic dermatitis.', *The Journal of investigative dermatology*, 128(11), pp. 2625–30. doi: 10.1038/jid.2008.111.

Kondo, S. and Sauder, D. N. (1997) 'Tumor necrosis factor (TNF) receptor type 1 (p55) is a main mediator for TNF-α-induced skin inflammation', *European Journal of Immunology*. Eur J Immunol, 27(7), pp. 1713–1718. doi: 10.1002/eji.1830270718.

Kong, H. H. (2011) 'Skin microbiome: Genomics-based insights into the diversity and role of skin microbes', *Trends in Molecular Medicine*. Trends Mol Med, pp. 320–328. doi: 10.1016/j.molmed.2011.01.013.

Kong, H. H. *et al.* (2012) 'Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis', *Genome Research*. Cold Spring Harbor Laboratory Press, 22(5), pp. 850–859. doi: 10.1101/gr.131029.111.

Korosec, A. *et al.* (2019) 'Lineage Identity and Location within the Dermis Determine the Function of Papillary and Reticular Fibroblasts in Human Skin.', *The Journal of investigative dermatology*, 139(2), pp. 342–351. doi: 10.1016/j.jid.2018.07.033.

Krishna, S. and Miller, L. S. (2012) 'Host-pathogen interactions between the skin and Staphylococcus aureus', *Current Opinion in Microbiology*. NIH Public Access, pp. 28–35. doi: 10.1016/j.mib.2011.11.003.

Kubo, A. *et al.* (2009) 'External antigen uptake by Langerhans cells with reorganization of epidermal tight junction barriers.', *The Journal of experimental medicine*, 206(13), pp. 2937–46. doi: 10.1084/jem.20091527.

Kuo, I. H. *et al.* (2013) 'Activation of epidermal toll-like receptor 2 enhances tight junction function: Implications for atopic dermatitis and skin barrier repair', *Journal of Investigative Dermatology*. Nature Publishing Group, 133(4), pp. 988–998. doi: 10.1038/jid.2012.437.

Lachner, J. *et al.* (2017) 'Epidermal cornification is preceded by the expression of a keratinocyte-specific set of pyroptosis-related genes', *Scientific Reports*. Nature Publishing Group, 7(1). doi: 10.1038/s41598-017-17782-4.

Lai, Y. *et al.* (2009) 'Commensal bacteria regulate toll-like receptor 3-dependent inflammation after skin injury', *Nature Medicine*. Nat Med, 15(12), pp. 1377–1382. doi: 10.1038/nm.2062.

Lai, Y. *et al.* (2010) 'Activation of TLR2 by a small molecule produced by staphylococcus epidermidis increases antimicrobial defense against bacterial skin infections', *Journal of Investigative Dermatology*. Nature Publishing Group, 130(9), pp. 2211–2221. doi: 10.1038/jid.2010.123.

Lai, Y. and Gallo, R. L. (2008) 'Toll-like receptors in skin infections and inflammatory diseases.', *Infectious disorders drug targets*. NIH Public Access, 8(3), pp. 144–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18782031 (Accessed: 9 August 2017).

Lai, Y. and Gallo, R. L. (2009) 'AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense', *Trends in Immunology*. Trends Immunol, pp. 131–141. doi: 10.1016/j.it.2008.12.003.

Lande, R. *et al.* (2007) 'Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide', *Nature*. Nature Publishing Group, 449(7162), pp. 564–569. doi: 10.1038/nature06116.

Lebre, M. C. *et al.* (2007) 'Human keratinocytes express functional Toll-like receptor 3, 4, 5, and 9.', *The Journal of investigative dermatology*, 127(2), pp. 331–41. doi: 10.1038/sj.jid.5700530.

Lee, S. E. and Lee, S. H. (2018) 'Skin barrier and calcium', *Annals of Dermatology*. Korean Dermatological Association, pp. 265–275. doi: 10.5021/ad.2018.30.3.265.

Leek, J. T. *et al.* (2012) 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinformatics*, 28(6), pp. 882–883. doi: 10.1093/bioinformatics/bts034.

Lessard, J. C. *et al.* (2013) 'Keratin 16 regulates innate immunity in response to epidermal barrier breach.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(48), pp. 19537–42. doi: 10.1073/pnas.1309576110.

Leung, D. Y. M. *et al.* (2004) 'New insights into atopic dermatitis.', *The Journal of clinical investigation*, 113(5), pp. 651–7. doi: 10.1172/JCI21060.

Leyden, J. J. *et al.* (1975) 'Propionibacterium levels in patients with and without acne vulgaris.', *The Journal of investigative dermatology*, 65(4), pp. 382–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/126263 (Accessed: 7 November 2018).

Li, D. *et al.* (2013) 'A novel lipopeptide from skin commensal activates TLR2/CD36-p38 MAPK signaling to increase antibacterial defense against bacterial infection.', *PloS one*. Edited by B. Ryffel, 8(3), p. e58288. doi: 10.1371/journal.pone.0058288.

Li, D. *et al.* (2019) ' Lipopeptide 78 from Staphylococcus epidermidis Activates β-Catenin To Inhibit Skin Inflammation ', *The Journal of Immunology*. The American Association of Immunologists, 202(4), pp. 1219–1228. doi: 10.4049/jimmunol.1800813.

Li, J. *et al.* (2016) 'Transcriptional Analysis of T Cells Resident in Human Skin.', *PloS one*. Edited by R. Arens, 11(1), p. e0148351. doi: 10.1371/journal.pone.0148351.

Li, N. *et al.* (2014) 'Alarmin Function of Cathelicidin Antimicrobial Peptide LL37 through IL-36γ Induction in Human Epidermal Keratinocytes', *The Journal of Immunology*. The American Association of Immunologists, 193(10), pp. 5140–5148. doi: 10.4049/jimmunol.1302574.

Liang, S. C. *et al.* (2006) 'Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides.', *The Journal of experimental medicine*. The Rockefeller University Press, 203(10), pp. 2271–9. doi: 10.1084/jem.20061308.

Liebner, D. A., Huang, K. and Parvin, J. D. (2014) 'MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples.', *Bioinformatics (Oxford, England)*, 30(5), pp. 682–9. doi: 10.1093/bioinformatics/btt566.

Lina, G. *et al.* (2003) 'Bacterial competition for human nasal cavity colonization: Role of Staphylococcal agr alleles', *Applied and Environmental Microbiology*. Appl Environ Microbiol, 69(1), pp. 18–23. doi: 10.1128/AEM.69.1.18-23.2003.

Lock, C. *et al.* (2002) 'Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis', *Nature Medicine*, 8(5), pp. 500–508. doi: 10.1038/nm0502-500.

Lowes, M. A. *et al.* (2008) 'Psoriasis vulgaris lesions contain discrete populations of Th1 and Th17 T cells', *Journal of Investigative Dermatology*. Nature Publishing Group, 128(5), pp. 1207–1211. doi: 10.1038/sj.jid.5701213.

Lowes, M. A., Bowcock, A. M. and Krueger, J. G. (2007) 'Pathogenesis and therapy of psoriasis', *Nature*. Nature Publishing Group, pp. 866–873. doi: 10.1038/nature05663.

Lowes, M. A., Suárez-Fariñas, M. and Krueger, J. G. (2014) 'Immunology of psoriasis', *Annual Review of Immunology*. Annual Reviews Inc., pp. 227–255. doi: 10.1146/annurev-immunol-032713-120225.

Lowy, F. D. (1998) ' Staphylococcus aureus Infections ', *New England Journal of Medicine*. Massachusetts Medical Society, 339(8), pp. 520–532. doi: 10.1056/nejm199808203390806.

Lu, J. *et al.* (2016) 'CXCL14 as an emerging immune and inflammatory modulator', *Journal of Inflammation (United Kingdom)*. BioMed Central Ltd. doi: 10.1186/s12950-015-0109-9.

Lun, A. T. L. *et al.* (2019) 'EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data', *Genome Biology*. BioMed Central Ltd., 20(1), p. 63. doi: 10.1186/s13059-019-1662-y.

Lun, A. T. L., Bach, K. and Marioni, J. C. (2016) 'Pooling across cells to normalize single-cell RNA sequencing data with many zero counts', *Genome Biology*. BioMed Central Ltd., 17(1), p. 75. doi: 10.1186/s13059-016-0947-7.

Lundberg, K. C. *et al.* (2015) 'Proteomics of skin proteins in psoriasis: From discovery and verification in a mouse model to confirmation in humans', *Molecular and Cellular Proteomics*. American Society for Biochemistry and Molecular Biology Inc., 14(1), pp. 109–119. doi: 10.1074/mcp.M114.042242.

Mabbott, N. A. *et al.* (2013) 'An expression atlas of human primary cells: inference of gene function from coexpression networks.', *BMC genomics*, 14(1), p. 632. doi: 10.1186/1471-2164-14-632.

MacLea, K. S. and Trachtenberg, A. M. (2017) 'Complete genome sequence of Staphylococcus epidermidis ATCC 12228 chromosome and plasmids, generated by long-read sequencing', *Genome Announcements*. American Society for Microbiology, 5(36). doi: 10.1128/genomeA.00954-17.

Macosko, E. Z. *et al.* (2015) 'Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.', *Cell*, 161(5), pp. 1202–14. doi: 10.1016/j.cell.2015.05.002.

Madison, K. C. (2003) 'Barrier function of the skin: "la raison d'être" of the epidermis.', *The Journal of investigative dermatology*, 121(2), pp. 231–41. doi: 10.1046/j.1523-1747.2003.12359.x.

Marenholz, I. *et al.* (2011) 'Association screening in the epidermal differentiation complex (EDC) identifies an SPRR3 repeat number variant as a risk factor for eczema', *Journal of Investigative Dermatology*. Nature Publishing Group, 131(8), pp. 1644–1649. doi: 10.1038/jid.2011.90.

Margolis, D. J. *et al.* (2014) 'Filaggrin-2 variation is associated with more persistent atopic dermatitis in African American subjects', *Journal of Allergy and Clinical Immunology*. J Allergy Clin Immunol, 133(3), pp. 784–789. doi: 10.1016/j.jaci.2013.09.015.

Martel, B. C. *et al.* (2016) 'Distinct molecular signatures of mild extrinsic and intrinsic atopic dermatitis.', *Experimental dermatology*, 25(6), pp. 453–9. doi: 10.1111/exd.12967.

Martin, P. *et al.* (2020) 'Intracellular IL-1 Receptor Antagonist Isoform 1 Released from Keratinocytes upon Cell Death Acts as an Inhibitor for the Alarmin IL-1α', *The Journal of Immunology*. The American Association of Immunologists, 204(4), pp. 967–979. doi: 10.4049/jimmunol.1901074.

Mathes, S. H., Ruffner, H. and Graf-Hausner, U. (2014) 'The use of skin models in drug development', *Advanced Drug Delivery Reviews*. Elsevier, pp. 81–102. doi: 10.1016/j.addr.2013.12.006.

Maxwell, J. R. *et al.* (2015) 'Differential Roles for Interleukin-23 and Interleukin-17 in Intestinal Immunoregulation.', *Immunity*, 43(4), pp. 739–50. doi: 10.1016/j.immuni.2015.08.019.

Mayumi, N. *et al.* (2013) 'E-cadherin interactions are required for Langerhans cell differentiation', *European Journal of Immunology*, 43(1), pp. 270–280. doi: 10.1002/eji.201242654.

McGovern, N. *et al.* (2014) 'Human dermal $CD14^+$ cells are a transient population of monocyte-derived macrophages.', *Immunity*. Elsevier, 41(3), pp. 465–477. doi: 10.1016/j.immuni.2014.08.006.

Di Meglio, P., Perera, G. K. and Nestle, F. O. (2011) 'The multitasking organ: recent insights into skin immune function.', *Immunity*, 35(6), pp. 857–69. doi: 10.1016/j.immuni.2011.12.003.

Menon, G. K., Grayson, S. and Elias, P. M. (1985) 'Ionic calcium reservoirs in mammalian epidermis: Ultrastructural localization by ion-capture cytochemistry', *Journal of Investigative Dermatology*. J Invest Dermatol, 84(6), pp. 508–512. doi: 10.1111/1523-1747.ep12273485.

Miajlovic, H. *et al.* (2010) 'Effect of filaggrin breakdown products on growth of and protein expression by Staphylococcus aureus.', *The Journal of allergy and clinical immunology*. Elsevier, 126(6), pp. 1184–90.e3. doi: 10.1016/j.jaci.2010.09.015.

Miedzobrodzki, J. *et al.* (2002) 'Proteolytic activity of Staphylococcus aureus strains isolated from the colonized skin of patients with acute-phase atopic dermatitis', *European Journal of Clinical Microbiology and Infectious Diseases*. Eur J Clin Microbiol Infect Dis, 21(4), pp. 269–276. doi: 10.1007/s10096-002-0706-4.

Miller, L. S. and Modlin, R. L. (2007) 'Human keratinocyte Toll-like receptors promote distinct immune responses.', *The Journal of investigative dermatology*, 127(2), pp. 262–3. doi: 10.1038/sj.jid.5700559.

Miller, M. B. and Tang, Y.-W. (2009) 'Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology', *Clinical Microbiology Reviews*, 22(4), pp. 611–633. doi: 10.1128/CMR.00019-09.

Mischke, D. *et al.* (1996) 'Genes Encoding Structural Proteins of Epidermal Cornification and S100 Calcium-Binding Proteins Form a Gene Complex ("Epidermal Differentiation Complex") on Human Chromosome 1q21', *Journal of Investigative Dermatology*. Blackwell Publishing Inc., 106(5), pp. 989–992. doi: 10.1111/1523-1747.ep12338501.

Mizutani, Y. *et al.* (2009) 'Ceramide biosynthesis in keratinocyte and its role in skin function.', *Biochimie*, 91(6), pp. 784–90. doi: 10.1016/j.biochi.2009.04.001.

Moll, R. *et al.* (1982) 'The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells', *Cell*. Cell, pp. 11–24. doi: 10.1016/0092-8674(82)90400-7.

Morin, R. D. *et al.* (2008) 'Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing', *BioTechniques*, 45(1), pp. 81–94. doi:

10.2144/000112900.

Naik, S. *et al.* (2012) 'Compartmentalized control of skin immunity by resident commensals.', *Science (New York, N.Y.)*, 337(6098), pp. 1115–9. doi: 10.1126/science.1225152.

Naik, S. *et al.* (2015) 'Commensal–dendritic-cell interaction specifies a unique protective skin immune signature', *Nature*, 520(7545), pp. 104–108. doi: 10.1038/nature14052.

Nakagawa, S. *et al.* (2017) 'Staphylococcus aureus Virulent PSMα Peptides Induce Keratinocyte Alarmin Release to Orchestrate IL-17-Dependent Skin Inflammation', *Cell Host & Microbe*, 22(5), pp. 667-677.e5. doi: 10.1016/j.chom.2017.10.008.

Nakatsuji, T. *et al.* (2017) 'Antimicrobials from human skin commensal bacteria protect against Staphylococcus aureus and are deficient in atopic dermatitis.', *Science translational medicine*, 9(378), p. eaah4680. doi: 10.1126/scitranslmed.aah4680.

Nestle, F. O. *et al.* (2009) 'Skin immune sentinels in health and disease.', *Nature reviews. Immunology*, 9(10), pp. 679–91. doi: 10.1038/nri2622.

Newell, L. *et al.* (2013) 'Sensitization via Healthy Skin Programs Th2 Responses in Individuals with Atopic Dermatitis', *Journal of Investigative Dermatology*. Elsevier, 133(10), pp. 2372–2380. doi: 10.1038/JID.2013.148.

Newman, A. M. *et al.* (2015) 'Robust enumeration of cell subsets from tissue expression profiles.', *Nature methods*, 12(5), pp. 453–7. doi: 10.1038/nmeth.3337.

Newman, A. M. *et al.* (2017) 'Data normalization considerations for digital tumor dissection', *Genome Biology*. BioMed Central Ltd. doi: 10.1186/s13059-017-1257-4.

Nickoloff, B. J. and Turka, L. A. (1994) 'Immunological functions of non-professional antigen-presenting cells: new insights from studies of T-cell interactions with keratinocytes.', *Immunology today*, 15(10), pp. 464–9. doi: 10.1016/0167-5699(94)90190-2.

Nograles, K. E. *et al.* (2008) 'Th17 cytokines interleukin (IL)-17 and IL-22 modulate distinct inflammatory and keratinocyte-response pathways.', *The British journal of dermatology*, 159(5), pp. 1092–102. doi: 10.1111/j.1365-2133.2008.08769.x.

Nograles, K. E. *et al.* (2010) 'Atopic dermatitis keratinocytes exhibit normal T(H)17 cytokine responses.', *The Journal of allergy and clinical immunology*. NIH Public Access, 125(3), pp. 744–6, 746.e1-746.e2. doi: 10.1016/j.jaci.2009.12.934.

Nomura, I. *et al.* (2003) 'Distinct patterns of gene expression in the skin lesions of atopic dermatitis and psoriasis', *Journal of Allergy and Clinical Immunology*, 112(6), pp. 1195–1202. doi: 10.1016/j.jaci.2003.08.049.

van Noort, V., Snel, B. and Huynen, M. A. (2003) 'Predicting gene function by conserved co-expression.', *Trends in genetics : TIG*, 19(5), pp. 238–42. doi: 10.1016/S0168-9525(03)00056-8.

Novak, N., Bieber, T. and Leung, D. Y. M. (2003) 'Immune mechanisms leading to atopic dermatitis.', *The Journal of allergy and clinical immunology*, 112(6 Suppl), pp. S128-39. doi: 10.1016/j.jaci.2003.09.032.

Ober-Blöbaum, J. L. *et al.* (2017) 'Monitoring Skin Dendritic Cells in Steady State and Inflammation by Immunofluorescence Microscopy and Flow Cytometry', in *Methods in molecular biology (Clifton, N.J.)*, pp. 37–52. doi: 10.1007/978-1-4939-6786-5_3.

Odhiambo, J. A. *et al.* (2009) 'Global variations in prevalence of eczema symptoms in children from ISAAC Phase Three.', *The Journal of allergy and clinical immunology*, 124(6), pp. 1251–8.e23. doi:

10.1016/j.jaci.2009.10.009.

Onishi, R. M. and Gaffen, S. L. (2010) 'Interleukin-17 and its target genes: Mechanisms of interleukin-17 function in disease', *Immunology*, pp. 311–321. doi: 10.1111/j.1365-2567.2009.03240.x.

Otto, M. (2004) 'Virulence factors of the coagulase-negative staphylococci', *Frontiers in Bioscience*. Frontiers in Bioscience, pp. 841–863. doi: 10.2741/1295.

Otto, M. (2009) 'Staphylococcus epidermidis - The "accidental" pathogen', *Nature Reviews Microbiology*. Nat Rev Microbiol, pp. 555–567. doi: 10.1038/nrmicro2182.

Otto, M. (2010) 'Staphylococcus colonization of the skin and antimicrobial peptides', *Expert Review of Dermatology*. Expert Rev Dermatol, pp. 183–195. doi: 10.1586/edm.10.6.

Oyoshi, M. K. *et al.* (2009) 'Cellular and Molecular Mechanisms in Atopic Dermatitis', *Advances in Immunology*, 102, pp. 135–226. doi: 10.1016/S0065-2776(09)01203-6.

Park, H.-Y. *et al.* (2013) 'Staphylococcus aureus Colonization in Acute and Chronic Skin Lesions of Patients with Atopic Dermatitis.', *Annals of dermatology*, 25(4), pp. 410–6. doi: 10.5021/ad.2013.25.4.410.

Parnell, G. P. *et al.* (2012) 'A distinct influenza infection signature in the blood transcriptome of patients with severe community-acquired pneumonia', *Critical Care*, 16(4), p. R157. doi: 10.1186/cc11477.

Pasparakis, M., Haase, I. and Nestle, F. O. (2014) 'Mechanisms regulating skin immunity and inflammation', *Nature Reviews Immunology*. Nature Publishing Group, pp. 289–301. doi: 10.1038/nri3646.

Pastore, S. *et al.* (1998) 'Interferon-gamma promotes exaggerated cytokine production in keratinocytes cultured from patients with atopic dermatitis.', *The Journal of allergy and clinical immunology*, 101(4 Pt 1), pp. 538–44. doi: 10.1016/S0091-6749(98)70361-6.

Paternoster, L. *et al.* (2015) 'Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis', *Nature Genetics*. Nature Publishing Group, 47(12), pp. 1449–1456. doi: 10.1038/ng.3424.

Percival, S. L. *et al.* (2012) 'Microbiology of the skin and the role of biofilms in infection', *International Wound Journal*. Int Wound J, 9(1), pp. 14–32. doi: 10.1111/j.1742-481X.2011.00836.x.

Pfaff, C. M. *et al.* (2017) 'The psoriasis-associated IL-17A induces and cooperates with IL-36 cytokines to control keratinocyte differentiation and function', *Scientific Reports*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/s41598-017-15892-7.

Philippeos, C. *et al.* (2018) 'Spatial and Single-Cell Transcriptional Profiling Identifies Functionally Distinct Human Dermal Fibroblast Subpopulations.', *The Journal of investigative dermatology*, 138(4), pp. 811–825. doi: 10.1016/j.jid.2018.01.016.

Polak, M. E. *et al.* (2014) 'Distinct molecular signature of human skin Langerhans cells denotes critical differences in cutaneous dendritic cell immune regulation.', *The Journal of investigative dermatology*, 134(3), pp. 695–703. doi: 10.1038/jid.2013.375.

Polak, M. E. and Singh, H. (2021) 'Tolerogenic and immunogenic states of Langerhans cells are orchestrated by epidermal signals acting on a core maturation gene module', *BioEssays*. John Wiley and Sons Inc. doi: 10.1002/bies.202000182.

Polański, K. *et al.* (2020) 'BBKNN: Fast batch alignment of single cell transcriptomes', *Bioinformatics*.

Oxford University Press, 36(3), pp. 964–965. doi: 10.1093/bioinformatics/btz625.

Puccetti, A. *et al.* (2018) 'Gene Expression Profiling in Behcet's Disease Indicates an Autoimmune Component in the Pathogenesis of the Disease and Opens New Avenues for Targeted Therapy.', *Journal of immunology research*. Hindawi Limited, 2018, p. 4246965. doi: 10.1155/2018/4246965.

Qiao, W. *et al.* (2012) 'PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions.', *PLoS computational biology*. Edited by R. Bonneau, 8(12), p. e1002838. doi: 10.1371/journal.pcbi.1002838.

Qin, J. Z. *et al.* (1999) 'Role of NF-κB in the apoptotic-resistant phenotype of keratinocytes', *Journal of Biological Chemistry*. J Biol Chem, 274(53), pp. 37957–37964. doi: 10.1074/jbc.274.53.37957.

Rauschmayr, T., Groves, R. W. and Kupper, T. S. (1997) 'Keratinocyte expression of the type 2 interleukin 1 receptor mediates local and specific inhibition of interleukin 1-mediated inflammation.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(11), pp. 5814–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9159157 (Accessed: 17 December 2018).

Reynolds, G. *et al.* (2021) 'Developmental cell programs are co-opted in inflammatory skin disease', *Science*. American Association for the Advancement of Science, 371(6527). doi: 10.1126/science.aba6500.

Roesner, L. M. *et al.* (2015) 'Der p1 and Der p2-Specific T Cells Display a Th2, Th17, and Th2/Th17 Phenotype in Atopic Dermatitis', *Journal of Investigative Dermatology*. Elsevier, 135(9), pp. 2324–2327. doi: 10.1038/JID.2015.162.

Romani, N. *et al.* (no date) 'Langerhans cells - dendritic cells of the epidermis.', *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, 111(7–8), pp. 725–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12974775 (Accessed: 8 November 2018).

Romero-Tlalolini, M. A., Chávez Olmos, P. and Garrido, E. (2013) 'Differential DNA methylation patterns in the CD86 gene controls its constitutive expression in keratinocytes', *Biochemical and Biophysical Research Communications*, 438(1), pp. 54–60. doi: 10.1016/j.bbrc.2013.07.023.

Rorke, E. A. *et al.* (2015) 'Structural and biochemical changes underlying a keratoderma-like phenotype in mice lacking suprabasal AP1 transcription factor function.', *Cell death & disease*. Nature Publishing Group, 6(2), p. e1647. doi: 10.1038/cddis.2015.21.

Rozzo, S. J. *et al.* (2001) 'Evidence for an interferon-inducible gene, Ifi202, in the susceptibility to systemic lupus.', *Immunity*, 15(3), pp. 435–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11567633 (Accessed: 9 November 2018).

Rubin, J. S. *et al.* (1989) 'Purification and characterization of a newly identified growth factor specific for epithelial cells.', *Proceedings of the National Academy of Sciences of the United States of America*, 86(3), pp. 802–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2915979 (Accessed: 19 February 2019).

Sa, S. M. *et al.* (2007) 'The effects of IL-20 subfamily cytokines on reconstituted human epidermis suggest potential roles in cutaneous innate defense and pathogenic adaptive immunity in psoriasis.', *Journal of immunology (Baltimore, Md. : 1950)*, 178(4), pp. 2229–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17277128 (Accessed: 17 December 2018).

Saitoh, K. *et al.* (2016) 'Anti-IL-17A blocking antibody reduces cyclosporin A-induced relapse in experimental autoimmune encephalomyelitis mice.', *Biochemistry and biophysics reports*, 8, pp. 139–145. doi: 10.1016/j.bbrep.2016.08.021.

Sanford, J. A. and Gallo, R. L. (2013) 'Functions of the skin microbiota in health and disease.',

Seminars in immunology, 25(5), pp. 370–7. doi: 10.1016/j.smim.2013.09.005.

Satija, R. *et al.* (2015) 'Spatial reconstruction of single-cell gene expression data', *Nature Biotechnology*. Nature Publishing Group, 33(5), pp. 495–502. doi: 10.1038/nbt.3192.

Scharschmidt, T. C. *et al.* (2015) 'A Wave of Regulatory T Cells into Neonatal Skin Mediates Tolerance to Commensal Microbes.', *Immunity*, 43(5), pp. 1011–21. doi: 10.1016/j.immuni.2015.10.016.

Schena, M. *et al.* (1995) 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray.', *Science (New York, N.Y.)*, 270(5235), pp. 467–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7569999 (Accessed: 9 November 2018).

Schoop, V. M., Mirancea, N. and Fusenig, N. E. (1999) 'Epidermal organization and differentiation of HaCat keratinocytes in organotypic coculture with human dermal fibroblasts', *Journal of Investigative Dermatology*. Blackwell Publishing Inc., 112(3), pp. 343–353. doi: 10.1046/j.1523-1747.1999.00524.x.

Schultz, G. S. *et al.* (2011) 'Dynamic reciprocity in the wound microenvironment.', *Wound repair and regeneration : official publication of the Wound Healing Society [and] the European Tissue Repair Society*, 19(2), pp. 134–48. doi: 10.1111/j.1524-475X.2011.00673.x.

Schwandner, R. *et al.* (1999) 'Peptidoglycan- and lipoteichoic acid-induced cell activation is mediated by Toll-like receptor 2', *Journal of Biological Chemistry*. J Biol Chem, 274(25), pp. 17406–17409. doi: 10.1074/jbc.274.25.17406.

Secor, P. R. *et al.* (2011) 'Staphylococcus aureus Biofilm and Planktonic cultures differentially impact gene expression, mapk phosphorylation, and cytokine production in human keratinocytes', *BMC Microbiology*. BioMed Central, 11, p. 143. doi: 10.1186/1471-2180-11-143.

Secor, P. R. *et al.* (2012) 'Phevalin (aureusimine B)Production by Staphylococcus aureus Biofilm and Impacts on Human Keratinocyte Gene Expression', *PLoS ONE*, 7(7), p. 40973. doi: 10.1371/journal.pone.0040973.

Selsted, M. E. and Ouellette, A. J. (2005) 'Mammalian defensins in the antimicrobial immune response', *Nature immunology*, 6(6), pp. 551–7. doi: 10.1038/ni1206.

Seneschal, J. *et al.* (2012) 'Human epidermal Langerhans cells maintain immune homeostasis in skin by activating skin resident regulatory T cells.', *Immunity*. NIH Public Access, 36(5), pp. 873–84. doi: 10.1016/j.immuni.2012.03.018.

Seo, M. D. *et al.* (2012) 'HaCa T keratinocytes and primary epidermal keratinocytes have different transcriptional profiles of cornified envelope-associated genes to T helper cell cytokines', *Biomolecules and Therapeutics*. Korean Society of Applied Pharmacology, 20(2), pp. 171–176. doi: 10.4062/biomolther.2012.20.2.171.

She, X. *et al.* (2009) 'Definition, conservation and epigenetics of housekeeping and tissue-enriched genes.', *BMC genomics*, 10(1), p. 269. doi: 10.1186/1471-2164-10-269.

Shen-Orr, S. S. *et al.* (2010) 'Cell type-specific gene expression differences in complex tissues.', *Nature methods*, 7(4), pp. 287–9. doi: 10.1038/nmeth.1439.

Shi, Z. rui *et al.* (2018) 'Decrease of galectin-3 in keratinocytes: A potential diagnostic marker and a critical contributor to the pathogenesis of psoriasis', *Journal of Autoimmunity*. Academic Press, 89, pp. 30–40. doi: 10.1016/j.jaut.2017.11.002.

Shih, B. B.-J. *et al.* (2016) 'Derivation of marker gene signatures from human skin and their use in the interpretation trancriptional changes associated with dermatological disorders', *The Journal of*

Pathology. John Wiley & Sons, Ltd. doi: 10.1002/path.4864.

Shklovskaya, E., Roediger, B. and Fazekas de St Groth, B. (2008) 'Epidermal and dermal dendritic cells display differential activation and migratory behavior while sharing the ability to stimulate CD4+ T cell proliferation in vivo.', *Journal of immunology (Baltimore, Md. : 1950)*, 181(1), pp. 418–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18566408 (Accessed: 15 February 2019).

Singh, D. *et al.* (2002) 'Gene expression correlates of clinical prostate cancer behavior.', *Cancer cell*, 1(2), pp. 203–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12086878 (Accessed: 9 November 2018).

Sivaprasad, U. *et al.* (2015) 'SERPINB3/B4 Contributes to Early Inflammation and Barrier Dysfunction in an Experimental Murine Model of Atopic Dermatitis', *Journal of Investigative Dermatology*. Nature Publishing Group, 135(1), pp. 160–169. doi: 10.1038/jid.2014.353.

Skabytska, Y. *et al.* (2016) 'The role of innate immune signaling in the pathogenesis of atopic dermatitis and consequences for treatments', *Seminars in Immunopathology*. Springer Verlag, pp. 29–43. doi: 10.1007/s00281-015-0544-y.

Smits, P. *et al.* (2000) 'Differentiation-dependent alternative splicing and expression of the extracellular matrix protein 1 gene in human keratinocytes', *Journal of Investigative Dermatology*. Blackwell Publishing Inc., 114(4), pp. 718–724. doi: 10.1046/j.1523-1747.2000.00916.x.

Son, E. D. *et al.* (2014) 'Staphylococcus aureus inhibits terminal differentiation of normal human keratinocytes by stimulating interleukin-6 secretion', *Journal of Dermatological Science*. Elsevier Ireland Ltd, 74(1), pp. 64–71. doi: 10.1016/j.jdermsci.2013.12.004.

Steen, C. B. *et al.* (2020) 'Profiling cell type abundance and expression in bulk tissues with CIBERSORTx', in *Methods in Molecular Biology*. Humana Press Inc., pp. 135–157. doi: 10.1007/978-1-0716-0301-7_7.

Stoeckius, M. *et al.* (2018) 'Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics', *Genome Biology*. BioMed Central Ltd., 19(1). doi: 10.1186/s13059-018-1603-1.

Stoitzner, P. *et al.* (2006) 'Langerhans cells cross-present antigen derived from skin.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), pp. 7783–8. doi: 10.1073/pnas.0509307103.

Suaŕez-Fariñas, M. *et al.* (2013) 'Suppression of molecular inflammatory pathways by Toll-Like receptor 7, 8, and 9 antagonists in a model of IL-23-induced skin inflammation', *PLoS ONE*. PLoS One, 8(12). doi: 10.1371/journal.pone.0084634.

Suárez-Fariñas, M. *et al.* (2011) 'Nonlesional atopic dermatitis skin is characterized by broad terminal differentiation defects and variable immune abnormalities.', *The Journal of allergy and clinical immunology*, 127(4), pp. 954-64.e1–4. doi: 10.1016/j.jaci.2010.12.1124.

Sugawara, T. *et al.* (2001) 'Regulation and role of interleukin 6 in wounded human epithelial keratinocytes', *Cytokine*. Academic Press, 15(6), pp. 328–336. doi: 10.1006/cyto.2001.0946.

Sugiura, A. *et al.* (2014) 'Reevaluation of the non-lesional dry skin in atopic dermatitis by acute barrier disruption: an abnormal permeability barrier homeostasis with defective processing to generate ceramide', *Archives of Dermatological Research*, 306(5), pp. 427–440. doi: 10.1007/s00403-013-1430-x.

Sumigray, K. D. and Lechler, T. (2015) 'Cell adhesion in epidermal development and barrier formation', in *Current Topics in Developmental Biology*. Academic Press Inc., pp. 383–414. doi: 10.1016/bs.ctdb.2014.11.027.

Sumikawa, Y. *et al.* (2006) 'Induction of beta-defensin 3 in keratinocytes stimulated by bacterial lipopeptides through toll-like receptor 2.', *Microbes and infection*, 8(6), pp. 1513–21. doi: 10.1016/j.micinf.2006.01.008.

Supek, F. *et al.* (2011) 'REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms', *PLoS ONE*. Edited by C. Gibas, 6(7), p. e21800. doi: 10.1371/journal.pone.0021800.

Swindell, W. R. *et al.* (2012) 'Heterogeneity of inflammatory and cytokine networks in chronic plaque psoriasis.', *PloS one*. Edited by D. L. Kontoyiannis, 7(3), p. e34594. doi: 10.1371/journal.pone.0034594.

Swindell, W. R. *et al.* (2013) 'Dissecting the psoriasis transcriptome: inflammatory- and cytokine-driven gene expression in lesions from 163 patients.', *BMC genomics*, 14(1), p. 527. doi: 10.1186/1471-2164-14-527.

Swindell, W. R. *et al.* (2014) 'Cellular dissection of psoriasis for transcriptome analyses and the post-GWAS era.', *BMC medical genomics*, 7(1), p. 27. doi: 10.1186/1755-8794-7-27.

Swindell, W. R. *et al.* (2017) 'RNA-seq identifies a diminished differentiation gene signature in primary monolayer keratinocytes grown from lesional and uninvolved psoriatic skin', *Scientific Reports*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/s41598-017-18404-9.

Swindell, W. R. *et al.* (2018) 'RNA-seq analysis of IL-1B and IL-36 responses in epidermal keratinocytes identifies a shared MyD88-dependent gene signature', *Frontiers in Immunology*. Frontiers Media S.A., 9(JAN). doi: 10.3389/fimmu.2018.00080.

Tabib, T. *et al.* (2018) 'SFRP2/DPP4 and FMO1/LSP1 Define Major Fibroblast Populations in Human Skin.', *The Journal of investigative dermatology*, 138(4), pp. 802–810. doi: 10.1016/j.jid.2017.09.045.

Takeuchi, O. *et al.* (1999) 'Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components.', *Immunity*, 11(4), pp. 443–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10549626 (Accessed: 15 February 2019).

Tan, N. S. *et al.* (2007) 'The Nuclear Hormone Receptor Peroxisome Proliferator-Activated Receptor β/δ Potentiates Cell Chemotactism, Polarization, and Migration', *Molecular and Cellular Biology*. American Society for Microbiology, 27(20), pp. 7161–7175. doi: 10.1128/mcb.00436-07.

Tan, Y. *et al.* (2016) 'CD24: from a Hematopoietic Differentiation Antigen to a Genetic Risk Factor for Multiple Autoimmune Diseases', *Clinical Reviews in Allergy and Immunology*. Humana Press Inc., pp. 70–83. doi: 10.1007/s12016-015-8470-2.

Tang, T. S., Bieber, T. and Williams, H. C. (2014) 'Are the concepts of induction of remission and treatment of subclinical inflammation in atopic dermatitis clinically useful?', *Journal of Allergy and Clinical Immunology*, 133(6), pp. 1615-1625.e1. doi: 10.1016/j.jaci.2013.12.1079.

Tengvall, S., Che, K. F. and Lindén, A. (2016) 'Interleukin-26: An Emerging Player in Host Defense and Inflammation', *Journal of Innate Immunity*, 8(1), pp. 15–22. doi: 10.1159/000434646.

Thepen, T. *et al.* (1996) 'Biphasic response against aeroallergen in atopic dermatitis showing a switch from an initial TH2 response to a TH1 response in situ: an immunocytochemical study.', *The Journal of allergy and clinical immunology*, 97(3), pp. 828–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8613640 (Accessed: 23 January 2017).

Thomas, E. *et al.* (2012) 'HCV infection induces a unique hepatic innate immune response associated with robust production of type III interferons.', *Gastroenterology*, 142(4), pp. 978–88. doi: 10.1053/j.gastro.2011.12.055.

Tian, S. *et al.* (2012) 'Meta-Analysis Derived (MAD) Transcriptome of Psoriasis Defines the "Core" Pathogenesis of Disease', *PLoS ONE*. Edited by H. P. Soyer, 7(9), p. e44274. doi: 10.1371/journal.pone.0044274.

Toda, M. *et al.* (2003) 'Polarized in vivo expression of IL-11 and IL-17 between acute and chronic skin lesions.', *The Journal of allergy and clinical immunology*. Elsevier, 111(4), pp. 875–81. doi: 10.1067/MAI.2003.1414.

Tong, S. Y. C. *et al.* (2015) 'Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management', *Clinical Microbiology Reviews*. American Society for Microbiology, 28(3), pp. 603–661. doi: 10.1128/CMR.00134-14.

Totsuka, A. *et al.* (2017) 'Expression of keratin 1, keratin 10, desmoglein 1 and desmocollin 1 in the epidermis: Possible downregulation by interleukin-4 and interleukin-13 in atopic dermatitis', *European Journal of Dermatology*. John Libbey Eurotext, 27(3), pp. 247–253. doi: 10.1684/ejd.2017.2985.

Traag, V. A., Waltman, L. and van Eck, N. J. (2019) 'From Louvain to Leiden: guaranteeing well-connected communities', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–12. doi: 10.1038/s41598-019-41695-z.

Trefzer, U. *et al.* (1993) 'The 55-kD tumor necrosis factor receptor on human keratinocytes is regulated by tumor necrosis factor-alpha and by ultraviolet B radiation.', *The Journal of clinical investigation*, 92(1), pp. 462–70. doi: 10.1172/JCI116589.

Trevino, V., Falciani, F. and Barrera-Saldaña, H. A. (2007) 'DNA microarrays: a powerful genomic tool for biomedical and clinical research.', *Molecular medicine (Cambridge, Mass.)*. The Feinstein Institute for Medical Research, 13(9–10), pp. 527–41. doi: 10.2119/2006-00107.Trevino.

Tsoi, L. C. *et al.* (2019) 'Atopic Dermatitis Is an IL-13–Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis', *Journal of Investigative Dermatology*, 139(7), pp. 1480–1489. doi: 10.1016/j.jid.2018.12.018.

Uchino, Y. *et al.* (2018) 'Galectin-3 is an amplifier of the interleukin-1β-mediated inflammatory response in corneal keratinocytes', *Immunology*. Blackwell Publishing Ltd, 154(3), pp. 490–499. doi: 10.1111/imm.12899.

Ungar, B. *et al.* (2017) 'An Integrated Model of Atopic Dermatitis Biomarkers Highlights the Systemic Nature of the Disease', *Journal of Investigative Dermatology*, 137(3), pp. 603–613. doi: 10.1016/j.jid.2016.09.037.

Volz, A. *et al.* (1993) 'Physical mapping of a functional cluster of epidermal differentiation genes on chromosome 1q21', *Genomics*. Genomics, 18(1), pp. 92–99. doi: 10.1006/geno.1993.1430.

Wakem, P. *et al.* (2000) 'Allergens and Irritants Transcriptionally Upregulate CD80 Gene Expression in Human Keratinocytes', *Journal of Investigative Dermatology*, 114(6), pp. 1085–1092. doi: 10.1046/j.1523-1747.2000.00997.x.

Wang, B. *et al.* (2017) 'IL-1β-Induced Protection of Keratinocytes against Staphylococcus aureus-Secreted Proteases Is Mediated by Human β-Defensin 2.', *The Journal of investigative dermatology*, 137(1), pp. 95–105. doi: 10.1016/j.jid.2016.08.025.

Wang, G. (2014) 'Human antimicrobial peptides and proteins', *Pharmaceuticals*. MDPI AG, pp. 545–594. doi: 10.3390/ph7050545.

Wang, G. *et al.* (2020) 'Association of the Psoriatic Microenvironment with Treatment Response', *JAMA Dermatology*. American Medical Association, 156(10), pp. 1057–1065. doi: 10.1001/jamadermatol.2020.2118.

Wang, S. *et al.* (2014) 'SPINK5 knockdown in organotypic human skin culture as a model system for Netherton syndrome: Effect of genetic inhibition of serine proteases kallikrein 5 and kallikrein 7', *Experimental Dermatology*. Blackwell Publishing Ltd, 23(7), pp. 524–526. doi: 10.1111/exd.12451.

Wang, S. *et al.* (2020) 'Single cell transcriptomics of human epidermis identifies basal stem cell transition states', *Nature Communications*. Nature Research, 11(1). doi: 10.1038/s41467-020-18075-7.

Wang, X. *et al.* (2019) 'Bulk tissue cell type deconvolution with multi-subject single-cell expression reference.', *Nature communications*, 10(1), p. 380. doi: 10.1038/s41467-018-08023-x.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics*, 10(1), pp. 57–63. doi: 10.1038/nrg2484.

Wanke, I. *et al.* (2011) 'Skin commensals amplify the innate immune response to pathogens by activation of distinct signaling pathways', *Journal of Investigative Dermatology*. Nature Publishing Group, 131(2), pp. 382–390. doi: 10.1038/jid.2010.328.

Werner, S. and Smola, H. (2001) 'Paracrine regulation of keratinocyte proliferation and differentiation', *Trends in Cell Biology*. Elsevier Ltd, pp. 143–146. doi: 10.1016/S0962-8924(01)01955-9.

West, H. C. and Bennett, C. L. (2017) 'Redefining the Role of Langerhans Cells As Immune Regulators within the Skin.', *Frontiers in immunology*. Frontiers Media SA, 8, p. 1941. doi: 10.3389/fimmu.2017.01941.

Wickersham, M. *et al.* (2017) 'Metabolic Stress Drives Keratinocyte Defenses against Staphylococcus aureus Infection.', *Cell reports*, 18(11), pp. 2742–2751. doi: 10.1016/j.celrep.2017.02.055.

Wilson, N. J. *et al.* (2007) 'Development, cytokine profile and function of human interleukin 17–producing helper T cells', *Nature Immunology*, 8(9), pp. 950–957. doi: 10.1038/ni1497.

Wingens, M. *et al.* (1998) 'Induction of SLPI (ALP/HUSI-I) in epidermal keratinocytes', *Journal of Investigative Dermatology*. Blackwell Publishing Inc., 111(6), pp. 996–1002. doi: 10.1046/j.1523-1747.1998.00425.x.

Wittersheim, M. *et al.* (2013) 'Differential expression and in vivo secretion of the antimicrobial peptides psoriasin (S100A7), RNase 7, human beta-defensin-2 and -3 in healthy human skin', *Experimental Dermatology*. Exp Dermatol, pp. 364–366. doi: 10.1111/exd.12133.

Wittmann, M. and Werfel, T. (2006) 'Interaction of keratinocytes with infiltrating lymphocytes in allergic eczematous skin diseases.', *Current opinion in allergy and clinical immunology*, 6(5), pp. 329–34. doi: 10.1097/01.all.0000244792.95615.3a.

Wojtowicz, A. M. *et al.* (2014) 'The importance of both fibroblasts and keratinocytes in a bilayered living cellular construct used in wound healing.', *Wound repair and regeneration : official publication of the Wound Healing Society [and] the European Tissue Repair Society*, 22(2), pp. 246–55. doi: 10.1111/wrr.12154.

Wolf, F. A. *et al.* (2019) 'PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells', *Genome Biology*. BioMed Central, 20(1), pp. 1–9. doi: 10.1186/s13059-019-1663-x.

Wolf, F. A., Angerer, P. and Theis, F. J. (2018) 'SCANPY: large-scale single-cell gene expression data analysis', *Genome Biology*, 19(1), p. 15. doi: 10.1186/s13059-017-1382-0.

Wolk, K. *et al.* (2004) 'IL-22 Increases the Innate Immunity of Tissues', *Immunity*, 21(2), pp. 241–

254. doi: 10.1016/j.immuni.2004.07.007.

Wolk, K. *et al.* (2006) 'IL-22 regulates the expression of genes responsible for antimicrobial defense, cellular differentiation, and mobility in keratinocytes: a potential role in psoriasis', *European Journal of Immunology*, 36(5), pp. 1309–1323. doi: 10.1002/eji.200535503.

Woo, T. E. and Sibley, C. D. (2020) 'The emerging utility of the cutaneous microbiome in the treatment of acne and atopic dermatitis', *Journal of the American Academy of Dermatology*. Mosby Inc., pp. 222–228. doi: 10.1016/j.jaad.2019.08.078.

Wu, W. *et al.* (2018) 'Regulation of 25-hydroxyvitamin D-1-hydroxylase and 24-hydroxylase in keratinocytes by PTH and FGF23', *Experimental Dermatology*. Blackwell Publishing Ltd, 27(11), pp. 1201–1209. doi: 10.1111/exd.13760.

Xiong, Y. *et al.* (2018) 'Profiles of immune infiltration in colorectal cancer and their clinical significant: A gene expression-based study', *Cancer Medicine*. Blackwell Publishing Ltd, 7(9), pp. 4496–4508. doi: 10.1002/cam4.1745.

Yang, J. *et al.* (2002) 'An iron delivery pathway mediated by a lipocalin', *Molecular Cell*. Mol Cell, 10(5), pp. 1045–1056. doi: 10.1016/S1097-2765(02)00710-4.

Yano, S. *et al.* (2008) 'Transcriptional responses of human epidermal keratinocytes to cytokine interleukin-1.', *Journal of cellular physiology*, 214(1), pp. 1–13. doi: 10.1002/jcp.21300.

Zaba, L. C. *et al.* (2007) 'Amelioration of epidermal hyperplasia by TNF inhibition is associated with reduced Th17 responses.', *The Journal of experimental medicine*. The Rockefeller University Press, 204(13), pp. 3183–94. doi: 10.1084/jem.20071094.

Zaba, L. C. *et al.* (2009) 'Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not immediate response TNF genes.', *The Journal of allergy and clinical immunology*. NIH Public Access, 124(5), pp. 1022-10.e1-395. doi: 10.1016/j.jaci.2009.08.046.

Zeeuwen, P. L. *et al.* (2012) 'Microbiome dynamics of human epidermis following skin barrier disruption', *Genome Biology*, 13(11), p. R101. doi: 10.1186/gb-2012-13-11-r101.

Zhang, C. *et al.* (2008) 'Cyclosporin A inhibits the production of IL-17 by memory Th17 cells from healthy individuals and patients with rheumatoid arthritis.', *Cytokine*, 42(3), pp. 345–52. doi: 10.1016/j.cyto.2008.03.006.

Zhang, P. *et al.* (2019) 'The lncRNA Neat1 promotes activation of inflammasomes in macrophages', *Nature Communications*. Nature Publishing Group, 10(1). doi: 10.1038/s41467-019-09482-6.

Zhao, S. *et al.* (2014) 'Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells.', *PloS one*. Edited by S.-D. Zhang, 9(1), p. e78644. doi: 10.1371/journal.pone.0078644.

Zhong, Y. *et al.* (2013) 'Digital sorting of complex tissues for cell type-specific gene expression profiles.', *BMC bioinformatics*, 14(1), p. 89. doi: 10.1186/1471-2105-14-89.

Zhong, Y. and Liu, Z. (2012) 'Gene expression deconvolution in linear space', *Nature Methods*. Nature Publishing Group, pp. 8–9. doi: 10.1038/nmeth.1830.

Ziolkowska, M. *et al.* (2000) 'High levels of IL-17 in rheumatoid arthritis patients: IL-15 triggers in vitro IL-17 production via cyclosporin A-sensitive mechanism.', *Journal of immunology (Baltimore, Md. : 1950)*, 164(5), pp. 2832–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10679127 (Accessed: 15 November 2018).